Coordination Action

Harmonising population-based biobanks and cohort studies to strengthen the foundation of European biomedical science in the post-genome era

POPULATION BIOBANKS

Date of preparation: 13.11.2003

List of participants:

Norwegian Institute of Public Health
University of Bristol, Prof. George Davey Smith
Rheinishe Friedrich-Wilhelms University of Bonn, Prof. Dr. Max P. Baur
National Public Health Institute Helsinki, Prof. Leena Peltonen
Mc Gill University, Dr. Thomas Hudson
University of Leicester, Prof. Paul Burton/Dr. Mary Dixon-Woods
Free University of Amsterdam, Prof. Dr. Dorret Boomsma
University of Montreal, Prof. Dr. Bartha Maria Knoppers
Estonian Genome project Foundation, Dr. Jaanus Pikani
Karolinska Institutet, Prof. Jan-Eric Litton
Second University of Naples, Prof. Paolo Gasparini
International Agency for Research on Cancer, Dr. Elio Riboli
The UK Biobank, Prof. John Norman Newton
Norwegian Institute of Public Health, Dr. Jennifer R. Harris/Prof. Per Magnus

Name of the coordinating person: Dr. Camilla Stoltenberg
e-mail: camilla.stoltenberg@fhi.no
fax: +47 23 40 82 47

Table of contents

RESEARCH TOPIC ADDRESSED **3**

PROPOSAL ABSTRACT **3**

B.1 SCIENTIFIC AND TECHNOLOGICAL OBJECTIVES OF THE PROJECT AND STATE OF THE ART **4**

B. 1.1 CURRENT STATE-OF-THE-ART 5
B.1.2 HOW THE PROPOSED PROJECT WILL ENHANCE THE CURRENT STATE-OF- THE-ART 5
B.1.3 IS A COORDINATED ACTION THE CORRECT STRATEGIC APPROACH? 6

B.2. RELEVANCE TO THE OBJECTIVES OF THE LIFESCIHEALTH PRIORITY **7**

B.3 POTENTIAL IMPACT **10**

WHAT DOES THE COORDINATION ACTION ADD TO ALREADY ESTABLISHED BIOBANK PROJECTS? 11
PLANS FOR EXPLOITATION AND DISSEMINATION 12
B.3.1 CONTRIBUTIONS TO STANDARDS 12

B.4 THE CONSORTIUM AND PROJECT RESOURCES **14**

Collaborators *18*

B.5 PROJECT MANAGEMENT **21**

THE ESTONIAN GENOME PROJECT 23
INTERNATIONAL AGENCY FOR RESEARCH ON CANCER/EPIC 23

B.6 WORKPLAN **24**

B.6.1 STRUCTURE OF THE WORKPLAN 24
WORKPACKAGE 1 ..... *26*
AIMS AND OBJECTIVES 26
DELIVERABLES 27
WORKPACKAGE 2 ..... *28*
WORKPACKAGE 3 ..... *29*
DATABASES AND BIOBANK INFORMATION MANAGEMENT SYSTEMS 29
WORKPACKAGE 4 ..... *31*
STRATEGIES FOR GENOTYPING IN LARGE SCALE BIOBANKS 31
WORKPACKAGE 5 ..... *33*
HARMONIZING THE ASSESSMENT OF PHENOTYPES AND LIFE-STYLE EXPOSURES 33
WORKPACKAGE 6 ..... *36*
ETHICAL, LEGAL, AND SOCIAL ISSUES IN POPULATION BIOBANKS 36
B.6.2 WORKPLANNING 38
Deliverables *42*

OBJECTIVES **46**

Deliverables *47*

B.7 ETHICAL, SAFETY AND OTHER EC-POLICY RELATED ISSUES **47**

ETHICAL, LEGAL, SOCIAL AND SAFETY ISSUES 47
OTHER EC-POLICY RELATED ISSUES 48
B.7.1 ETHICAL ASPECTS 49

B.8 GENDER ISSUES **50**

B.8.1 PARTICIPATION OF WOMEN 50
B.8.2 GENDER ASPECTS IN RESEARCH 51

Proposal summary page

Harmonising population-based biobanks and cohorts studies to strengthen the foundation of European biomedical science in the post-genome era

POPULATION BIOBANKS

Research topic addressed

LSH-2003-1.1.0-2

Proposal abstract

This Coordinated Action aims to establish a collaborative research network that will identify and explore some of the key issues that will help ensure that Europe is able to make best use of its rich array of population-based biobanks and longitudinal cohort studies. These include world-leading cohorts that already exist and exciting new initiatives that are just beginning. The ultimate long-term aim is to harmonise those features that are common to many biobanks and cohort studies and that, when implemented in a complementary manner, can act to: (1) promote communication between major biobanking initiatives; (2) enhance the effective sharing and synthesis of information; and (3) avoid the expensive mistakes and inefficiencies that can arise when individual initiatives repeatedly "re-invent the wheel". Because of its large and varied population-base, the population-centred nature of many of its health systems, and its past emphasis on the importance of population-based cohort studies, Europe is well placed to capitalise on many of the exciting biomedical research opportunities that now exist in the post-genome era. But, if we can ensure that our biobanks and large cohort studies are able to work together to address pivotal research questions that fall outside the scope of a large project funded within one nation, or even of a single large cohort distributed across several nations, then we can ensure that Europe remains right at the cutting-edge of biomedical research internationally. Our long-term aim is unashamedly ambitious, but it can undoubtedly be realised if it is approached in manageable stages. Given the pre-existing thought that has already been put into the harmonisation of biobanks under important initiatives such as the EU funded COGENE project and under P3G (with Canada), and within EU funded studies such as GenomEUtwin, this proposal for a Coordination Action represents the next logical step towards the ultimate aim.

B.1 Scientific and technological objectives of the project and state of the art

Our **ultimate aim** is to establish and maintain a cost-effective and "harmonised" network of population-based biobanks and longitudinal cohort studies across Europe and in Canada. This harmonisation aims to optimise the ability of biobanks to: communicate with one another, share ideas, information and data, and collaborate effectively in a complex world where laws and ethical guidelines are necessarily rigorous, often differ between nations and alter with time as biomedical science advances and societal interests and concerns change and mature.

This **long-term aim** is very ambitious and we believe that it must be approached in manageable stages. Consequently, building upon pre-existing experience under the COGENE and P3G (Public Population Project in Genomics) initiatives and within the GenomEUtwin study, we believe that the next logical steps to take are encapsulated within this application for a Coordination Action (CA) that has the following objectives:

To identify and describe, in a standardized form, large population-based biobanks and longitudinal cohort studies in Europe. Particular emphasis is placed on studies that can contribute substantially to coordinated investigations of the genetic and environmental determinants of complex diseases.

To identify new biobanking opportunities within Europe. This will include a particular focus on genetically isolated populations, and we will establish standard criteria for selection and collection of data and samples from these populations.

To review current best practice for Biobank Information Management Systems. Key issues of harmonisation in relation to the management of large and complex databases for biobanks will be explored with a focus on efficient technologies, high level programming and the development of flexible communication engines that support reliable, efficient and secure communication between biobanks.

To create an operational infrastructure for the evaluation of ongoing large-scale genotyping efforts in population cohorts. This will provide a natural forum for expert opinions regarding marker selection, genotyping methods, quality assessment steps, database structure and analysis of the produced genotypes.

To lay the groundwork for a harmonised approach to the assessment of a wide range of complex phenotypes and life-style exposures. One complex phenotype, major (unipolar) depression, will be worked up in full, as an exemplar of what will ultimately be required for all phenotypes and exposures.

To establish ethical-legal and governance criteria consistent with the international norms and European practices and will enable data and sample sharing for research purposes.

To integrate collected European (and Canadian) expertise in study design, genotyping, database structures and statistical analyses and make this expertise accessible via the web-placed SOP:s or guidelines for study design.

A cost-effective and "harmonised" network of population-based biobanks and longitudinal cohort studies across Europe and in Canada would support the study of a wide range of disease-environment-lifestyle-genetic relationships with much greater power than previously possible. It would enable studies of the genetic and non-genetic determinants of the onset and natural history of established diseases. This infrastructure, and new information generated from it, would create invaluable opportunities for research into prevention, diagnosis and management of a wide range of important complex diseases with major implications for public health throughout the world.

Extensive discussions and investigations are required before such an ambitious project can sensibly designed and implemented. The primary aim of this Coordination Action is to properly lay the foundation for a definitive harmonisation programme that can later be implemented in an economically viable manner and has a high probability of success. The

preparatory groundwork undertaken within the CA will be detailed in a series of reports, from the individual workpackages (WPs), the critical elements of which will be made accessible on our web site. This information will provide specific guidelines and recommendations pertaining to the most effective way to take forward a harmonisation programme for population based biobanks and cohort studies across Europe and in Canada.

B. 1.1 Current state-of-the-art

The strength and depth of European bioscience is founded, in no small part, on its rich collection of large population-based cohort studies. Such studies permit a comprehensive investigation of the relationships between disease outcomes and causative exposures, including genes, in a manner that avoids many of the limitations inherent to retrospective case-control data. However, financial realities mean that large prospective studies are generally limited in scope. With its scientific focus on dissecting individual biological pathways modern biomedicine demands additional and detailed assessment of both disease status and, potentially, of exposure. This can be prohibitively expensive in the setting of a traditional cohort study. Consequently, the investigation of the relationships between disease, environment, life-style and genes, upon which many of the future advances in biomedicine will depend, must be based on case-control studies nested within a cohort framework. That is, diseased and non-diseased individuals will be sampled from large prospective population-based cohorts. This hybrid design combines the rigour of prospective information collection inherent to a traditional cohort study with the ability to undertake a detailed assessment of outcome and exposure that is only possible in a case-control study.

The relevance of these design considerations is highlighted by the fact that most contemporary genetic epidemiology data sets are woefully limited in statistical power. It is not at all unusual for a formal *post hoc* power calculation to demonstrate that the sample size underpinning a pivotal analysis is too small by a factor of ten or more. In consequence, positive results are unusual, very rarely replicated, and may be spurious. This is costly in terms of wasted time and money and impedes the broader development of bioscience throughout the world. . Large multinational studies, maximizing the use of different types of populations and collective expertise of epidemiologists, geneticists and molecular geneticists mastering the genome-wide genotyping technologies are all needed to gain optimal benefits from the existing cohorts. The harmonisation efforts, shared database structures and intellectual expertise we describe form the platforms for pan-European studies and basically provide a proof of principle to facilitate integrated efforts in data analysis across the world.

B.1.2 How the proposed project will enhance the current state-of- the-art

A number of European nations, including Iceland, Estonia, Canada and the UK have embarked upon a series of major, *de novo* biobanking initiatives for storing information and tissue samples (including blood) on a very large number of people that can then be linked to routine health information systems to identify disease end-points. This approach is extremely expensive. For example, the UK Biobank initiative (500,000 middle aged recruits) is a total of €89M (£61M) just to set up the infrastructure. However, much of the detailed analyses will be based upon nested case-control methods as described above. Power calculations reveal that very large samples, with at least 5,000 new cases of a disease, are needed to detect interactions between a genetic effect and an environmental determinant in the presence of realistic levels of biological complexity. The opportunity to address questions of this nature is greatly enhanced through internationally coordinated biobank procedures and resources. For example, pooling the equivalent of five UK Biobanks across Europe (total sample size ≈2.5M) with 30 years follow-up would yield the required number of cases to study more than 30 important complex diseases. Such a resource would place Europe at the forefront of attempts to analyze aetiological architecture of these conditions. Given the array of pre-

existing and proposed initiatives that we can bring together, this aim is very realistic, provided we can successfully undertake the harmonisation program articulated within this CA.
Although the information available for analysis varies markedly from cohort study to cohort study and from biobank to biobank, there are important classes of data items that will be common to numerous studies. We therefore propose to undertake a detailed review, with the aim of evaluating the current information content of major European biobanks and cohorts, and studies in isolated population. This will inform us of the potential for creating a Europe-wide network of studies which share a common minimum data set and meet appropriate criteria for the quality and type of their design and conduct. This CA will also pave the way for genetic analysis in a number of Europe's most important cohort studies that already have first rate phenotypic and/or exposure information but currently have no DNA extracted for genetic or genomic research. This review will demand the identification of studies that could join the initiative, negotiation with the responsible investigators to determine whether they would be prepared to collaborate in such an initiative, and that there are limited legal and ethical barriers to such collaboration. We will then need to determine a common minimum data set, to identify where each current study falls short of that minimum data set and to determine the feasibility and cost of correcting each shortfall. We will also evaluate the extent to which each study exceeds the minimum requirement, as there may be sub-groups of studies that can be pooled with additional objectives in mind. The expertise already collected in the ongoing population cohort based studies and studies in isolated populations could greatly benefit the new initiatives and maximize their benefits.

B.1.3 Is a Coordinated Action the correct strategic approach?

We do not underestimate the difficulties of this ambitious program, which represents a natural extension P3G , which represents a collaboration between Canada and Europe, and of the activity of the Working Group on Population Based Cohorts and Biobanks that has been one of the invaluable outcomes of the successful EU funded COGENE initiative. There are recognised scientific barriers to pooling information between different studies and the potential ethical, legal and personal problems are legion. This is why we propose a preliminary CA based approach to explore the problems and opportunities and to bring more potential collaborators formally together. We consider a CA to be a safe ("low risk") way of preparing for actual harmonisation of one or more networks of biobanks and integrated research projects in the future. The risk of not reaching a critical step is small because the main outcome of the CA is focused meetings of experts, reports, a web page, and one or more networks. However, go into this with our eyes open. We realise that there may be conflicting scientific views and legal barriers that hinder full harmonisation; in other words the CA may conclude collaborative opportunities between biobanks are limited, but the funds invested will have yielded essential information regarding feasibility and potentials for harmonisation. However, given the evident good will of the scientists and funding agencies who have so far been involved, we are optimistic that this will not be the case, and the scientific and social rewards are so great if harmonisation is possible that we believe that this small investment in a CA can be well justified. We realise that the total budget of the CA is higher than most other CAs that have been awarded. But we would argue that a CA is the correct funding mechanism in this case, that any further paring down of the budget would start to damage the science, and that it is far more strategic to spend €700,000 on this CA than to invest larger amounts on "high risk" projects that merely represent individual components of the overall "big picture".

B.2. Relevance to the objectives of the LifeSciHealth Priority

The aims of this CA coincide with the overall objectives of the *Life Sciences Health Priority, 1: Life sciences, Genomics and Biotechnology for Health* and directly address the content solicited in the second call of Sixth Framework Programme (LSH-2003-1.1.0-2). In the wake of the sequencing of the human genome attention has turned towards elucidation of the relationships between genetic factors, environments and complex disease. Most common human diseases, such as hypertension, coronary artery disease, and psychiatric disorders, develop through a culmination of lifelong interactions between our genome and the environment. Predicting the contribution of genes to such complex disorders is still a challenge, and determining the interactions between genes and the environment during any disease process is a daunting task. Many experts agree that a particularly valuable strategy to dissect the relationship between genetic factors and complex disease will be to collect large epidemiological study samples from many different populations. Another important approach focuses on extensive studies of genetically isolated populations from many different European and Canadian populations (Peltonen and Mc Kusick 2001, Peltonen et al 2000). These ideas are congruent with plans for developing genomics research at the National Human Genome Research Institute, which emphasizes the importance of "obtaining unbiased assessments of the relative disease risk that particular gene variants contribute, a large longitudinal population-based cohort study, with collection of extensive clinical information and ongoing follow-up, would be profoundly valuable to the study of all common diseases " (Collins et al, Nature 2003). Although establishing or finding such populations is difficult in countries such as the USA, ample opportunities to build upon diverse European population present themselves today. Mining this array of population studies for the purpose of understanding disease etiology and pathogenesis requires multidisciplinary effort across multiple areas of coordination.

Our project has a powerful European dimension and builds upon data resources in population studies across the ERA, including associated candidate countries (Estonia). The harmonisation of assessments, methodological approaches, protocols and guidelines that we propose have been facilitated by previous European support towards most of the involved partners under FP5 and previous Framework Programmes. Participants in this CA are invited based on direct involvement in epidemiological cohorts or studies of genetic isolates in Europe and Canada. Estonia is an *associated candidate country* already involved as a participant in the CA. Representatives from other associated candidate countries will be invited as collaborators to the workshops and other activities described in the work packages. For instance, we will invite Judith Sandor, Central European University, Budapest, Hungary to collaborate on the work package activities concerning ethical and legal aspects, and Valdis Pirags from COGENE and the Latvian Genome Project will also be involved. Participation from other EU member countries and associated candidate countries is critical to enhance the competence of the CA organising group, to disseminate information and awareness of the research opportunities in population biobanks, to promote candidate countries competencies, and provide support to researchers in these countries to participate in workshops, and to establish and reinforce networks encompassing countries that are not participants in the CA. This international composition adds to the community added value and helps to ensure cost effective use of national and European resources.

Involvement of researchers, teams and institutions from third countries having an S&T agreement with the European Union, i.e. Canada, will facilitate access to existing knowledge

and expertise and help to ensure Europe's strong and coherent participation in the research initiatives conducted at international level in order "*to push back the boundaries of knowledge"* in the field. This will represent a strong "*contribution to a European Research Area open to the world"*.

Our goal, to lay the groundwork for a harmonised network of human population samples based on large epidemiological cohorts and isolated populations, will contribute to the *integrating and strengthening the European Research Area* in several ways.

First, integration of European and Canadian expertise in population genetic studies provide numerous competitive advantages. Among these are: a) access to multiple populations with different history and different level of isolation, b) access to data of health and disease information that has been collected using relatively similar health care systems c) access to vast amount of life style and environmental information collected in epidemiological studies across Europe, 4) possibilities to address (on a multinational scale) critical issues concerning ethical and legal challenges of such studies. Efficient networks must be forged to handle issues of data sharing and seamless collaborations among clinicians, epidemiologists, geneticists, mathematicians, and computer experts will be needed to solve the genetic underpinnings of complex diseases that affect the lives of millions. This coordinated exchange of ideas between scientists, clinical researchers, healthcare providers and physicians, ethicists, industry and other stakeholders directly strengthens the ERA. It will promote rapid evaluation and integration into project protocols of advances (e.g. genotyping methods or statistical methods to analyse genetic variation and gene-environment dynamics) and also provides a forum to discuss decisions about design strategies. For instance, the debate over gene identification strategies continues: Which populations should be studied? What are the optimal strategies for statistical analysis that can take advantage of both vast genetic data sets and quantitative phenotype information? How will associations between a genetic marker and a clinical phenotype be validated?

Second, the project will stimulate economic growth and enhance competitiveness by promoting the involvement of small and medium sized enterprises (SMEs). Furthermore, reports from the CA will provide guidelines regarding the activities (i.e. large scale DNA extraction, genotyping, data management and information systems) which are particularly suited for being conducted by SMEs, or where there is an interest in exploring different ways of organising the activities (i.e. data collection) as enterprises. The participant projects and researchers in the CA are responsible for projects with a variety of organisational structures and funding mechanisms. The Estonian genome project (EGP), for example, is organised in public/privet partnerhip with non-profit Estonan Genome Project Foundation (EGPF) and SME AS EGeen (EG) . EGPF receives a minor contribution from the Estonian government, while the major source of funding will be private investors through EG. The UK BioBank is organised as a publicly funded company. BioHealth Norway is publicly organised and funded but will consider organising sub-activities in companies or buying services from private enterprises.

Third, this CA will help integrate the ethical, social, legal and wider issues in the study of large population cohorts by involving professional ethicists, and other stakeholders in the project. The project approach will include the development and monitoring of procedures for compliance with ethical principles and European laws for research, including special attention to consent to participation in human research projects and the use of human DNA collections. Through WP6 we will take an active role in dissemination and public discussion of the ethical issues raised by populations studies and genetic research in complex trait diseases. We will

also seek input from the wider community. These activities help to foster awareness for ethical issues in biobanking research.

Fourth, it is important to involve associated candidate countries in the Community's research policy and in the European Research Area. The project participants already have strong working links with partners from several candidate countries (i.e. Poland, Romania, Croatia). The project will encourage further active participation through involving their representatives in the discussion of the different topics and in stimulating new initiatives in these countries. There will be open access to internet based tools, databases and training programs.

Fifth, the project's information dissemination strategy will include an internet platform to provide European researchers with guidelines, common criteria, general rules, etc. and a network of symposia, conferences and workshops to disseminate knowledge and skills. The project will also envisage a structured system for dissemination to researchers through e-newsletters, e-learning tools and meetings, and will target dissemination specifically to candidate countries.

Finally, mobility of researchers will be an indirect benefit resulting from building collaborations and networks.

References:
Collins FS, Green ED, Guttmacher AE, Guyer MS on behalf of the US National Human Genome Research Institute. A vision for the future of genomics research. A blueprint for the genomic era. Nature 2003; 422: 1-13.

Peltonen L, Mc Kusick. Genomics and medicine. Dissecting human disease in the postgenomic era. Science. 291(5507):1224-9, 2001 Feb 16.

Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. Nat Rev Genet. 2000 1(3):182-90

B.3 Potential impact

The impact of the CA in the thematic priority area 1: Life sciences, genomics and biotechnology for health.

The potential impact of our harmonisation goals will be to:
Promote communication within and between major biobanking initiatives thereby helping to overcome existing fragmentation of European population genomic research
Enhance the effective sharing and synthesis of information, thereby addressing the need for very large sample sizes and helping to promote collaborative international genetic epidemiological research
Avoid the expensive mistakes and inefficiencies that can arise when individual initiatives repeatedly "re-invent the wheel", thereby saving funders and researchers a lot of time and money.

This CA will lay the groundwork for:
Capitalizing on European strengths in epidemiology, human genetics and biomedicine to launch a comprehensive programme with long-term aims to harmonise pre-existing and planned population-based biobanks and cohort studies and thereby to contribute greatly to the capacity of European bioscientists to characterise the role of genes and environment in the aetiology of complex diseases.
Harmonisation of data collections common to different projects so that there are some common nodes of identity in the medical, demographic and social data collected from participants
Maximum standardisation of common biological phenotypes collected by each partner as well as agreement on methods and quality control
Compatible genotype databanks and common nomenclature of data and genomic variations
Systematic evaluation of the most cost-beneficial genotyping efforts using well defined statistical parameters
Collection of expertise in complex statistical analyses of the quantitative parameters and longitudinal events collected from these cohorts
Compatible phenotype databanks and common nomenclature of data and phenotypic variations
Coordination of access to each other's databanks
Protecting confidentiality
Development of security measures for the protection of genetic data and banks
Coordinated development of bioinformatics for compatible data mining and for clarity on ownership/copyright issues
Sharing of approaches to public engagement, governance and intellectual property issues
Exchange of experts and young researchers
Support for the transfer of knowledge and technology to other developed or developing countries
International leadership in the ethics of genetic research involving populations
Comparative evaluation and validation of research results and/or hypotheses on health and disease

There are several advantages to establishing a strong network of population based biobanks and cohorts in Europe and Canada. We believe that European and Canadian strengths include:

The advantage of having population cohorts with well defined parameters and collected information of genetic data as well as environmental and life style exposures prior to disease development
Large amounts of public money have already been invested in these existing and planned data collections and biobanks. Therefore it is essential to harmonize these databases to maximize the potential. EU has been crucial in the initiation of Genomeutwin, the large cohort of over 1.6 million twin pairs, representing a unique population cohort collected from eight countries.
There is sufficient environmental and population diversity
It is possible to identify numerous genetic isolates within the European and Canadian populations
There are good epidemiological and population records
Equal educational systems makes launching of such studies feasible and creates a good infrastructure for information of study participants of the progress and findings made.
One payer health care system creates a situation with relatively easy feedback of new knowledge to the health care systems and translation into health policy
Existing genealogies in several countries (Iceland, Sweden, Norway, Finland, Italy, Netherlands)
The existing large epidemiological studies have already been used for exploring the role of genetic and life style risk factors in common diseases and they have even influenced some health care decisions in European countries, so we know that collection of reliable data form population cohorts really works.
Stable sub-populations with different histories make various genetic and statistical analyses productive.
Biotechnology and informatics industries are eager to follow up on scientific discoveries
Mostly socialized health care systems with good records, good demographic data and possibilities for collection of "end-points".

What does the Coordination Action add to already established biobank projects?
This CA builds on major pre-existing initiatives and projects, including the COGENE and P3G initiatives and the GenomEUtwin Project. The multidisciplinary collaborative research group submitting this CA has arisen as a merger between a *working party on population-based biobanks and cohort studies* that was set up as one of the key outcomes of the EU funded COGENE initiative and a group of research leaders of large population-based biobanks that was brought together under the joint European/Canadian P3G initiative. The projects subsumed by P3G include: GenomEUtwin, UK Biobank, the Estonian Biobank and Cartagene in Canada. Both COGENE and P3G have aims that extend beyond the remit of this CA alone, but in agreeing to come together to submit this CA, we have committed ourselves to forming a merged initiative (name to be formally agreed upon when we next meet in Manchester in December 2003) which should be seen as the umbrella initiative underpinning this application. This joint initiative will have a broad focus, but one of its central concerns will be the conduct of this CA and the translation of its findings into the next steps in our long-term aim to achieve effective harmonisation of European population-based Biobanks and cohort studies. As a technical detail, it is important to note that many of the broader interests of the original COGENE initiative are unrelated to this CA (for example, some relate to disease-based rather than population-based biobanks) and that these aspects are being subsumed within a different application (for a STREP to continue COGENE) to the current call for FP6, by a group from Germany led by, amongst others, Christina Schroder from Frankfurt. Their application and ours should not be seen as competing scientifically with one another, because we are dealing with different aspects of the remit of the original COGENE initiative.

An particularly important contribution of this CA above already funded work is that it will expand the focus from twin studies to population based studies in general, encompassing the genomic strategies, methodologies and approaches involved in epidemiological cohorts, twin registries and studies of isolated populations. This CA and GenomEUtwin overlap because the inclusion of genotyping, phenotyping, data management and ethical issues is a prerequisite for any effort in population genetics. However, we will not duplicate what GenomEUtwin does, but rather develop it further into new kinds of biobanks and cohorts. The GenomEUtwin project has access to a very large collection of twin families who have been carefully phenotyped on a wide range of traits, usually including longitudinal phenotyping. These phenotyped twins and their families form the basis for selection of subjects and families who are informative for genetic linkage and association studies of complex traits. DNA and other biological samples will be collected in the selected families. The biobanking initiative in this CA aims to contribute to the collection of biological samples in populations that are unselected for any specific disease or phenotype and put one of its foci on the phenotyping that can be done in the biological samples themselves.

Plans for exploitation and dissemination
All reports and documents from the CA will be publicly available on a web page and knowledge about the web page and its contents will be actively disseminated far beyond the participants in the CA. We plan to publish selected papers or special issues of *Human Genetics* and/or the *International Journal of Epidemiology* based on the CA deliverables. Participants in the CA, Thomas Hudson and George Davey Smith respectively, edit the two journals. We also plan "proof of principle projects" among the work packages. We are also considering other new proposals. For example, we are going to investigate the possibility of introducing a scheme allowing pre-registration of analyses within the collaborating consortium which will come with the requirement that all registered results are submitted to the web site (regardless whether they are positive or negative), thereby producing a results reservoir that will be free from the severe distortion arising from publication bias.

B.3.1 Contributions to standards

The outcomes of this CA will have substantial impact on national and international procedures for biobanking and coordinated projects. Specifically, the harmonization tasks articulated herein will contribute substantially to set standards for genotyping, database management, data collection related to questions of phenotyping and genetic study design, genotyping, and ethical-legal governance issues; each an essential building block for collaborative biobank initiatives. Plans to make the findings and recommendations easily accessible will help promote commonality of practices and allow for financial and scientific resources to focus on next steps rather than repeatedly tackling the same problems again and again. The standards generated from the work in this CA will build upon and extend those established in the base initiatives GenomEUtwin and COGENE upons which this CA is built. These projects are described below.
GenomEUtwin (www.genomeutwin.org), was funded by the EU in October, 2002 and is based upon merging twin cohorts from 8 countries, as well as a population cohort of MORGAM from 9 countries. It combines the power of the twin study design with genomewide tools of modern genetics to circumvent common problems that hinder genetic studies aiming to identify genetic and life style risk factors of common diseases. The participating eight twin cohorts form an amazing collection of over 0.8 million pairs of twins. Over 30 000 DNA samples with informed consents for genetic studies of common diseases have already been stored from these population-based twin cohorts. Studies targeted to

cardiovascular traits are now being undertaken in MORGAM, a prospective case-cohort study based upon data from several European countries. MORGAM cohorts include approximately 6000 individuals, drawn from population-based cohorts consisting of more than 80 000 participants who have donated DNA samples. These unique study samples will be analysed using the platform of four intellectual core facilities of the project with accumulated expertise by partners in genetics, epidemiology, biostatistics and ethical issues. GenomEUtwin aims to create a solid infrastructure and research environment for European-wide studies using representative population cohorts and population-based twin registries that are particularly beneficial for genetic research when compared to other types of cohorts. GenomEUtwin works to harmonise and standardize the epidemiological, phenotypic and genotypic databases in participating centres to facilitate pooled analyses of genetic, phenotypic and epidemiological data. The final goal of GenomEUtwin is to build the intellectual and computational infrastructure and expertise to produce new information of "genetic profiles", predisposing to common diseases. The involved projects create training opportunities for tens of students and fellows in genetic epidemiology in Europe.

COGENE (Coordination of Genomes Research Across Europe)
COGENE is an EU-funded strategic accompanying measure aimed at promoting the development of synergies between national genome research programmes related to human health in Europe (http://forum.europa.eu.int/irc/rtd/cogene/info/data/pub/home.htm). It acts on behalf of the Forum of Genomes Programme Managers with representatives from 25 European countries. The inaugural COGENE workshop, entitled *Co-ordination of European Population Databases for Investigating Genetic Susceptibility to Disease*, was held in Cyprus in October 2002. Following from that meeting, a working party was set up, under the chairmanship of Paul Burton, to explore a number of the issues arising from the workshop, particularly those that pertained to population-based biobanks. Assisted with some further funding from the EU and COGENE, the working party met regularly by teleconference, and face-to-face on three further occasions: (a) at the Biobanks for Health workshop led by Camilla Stoltenberg, Oslo, January, 2003; (b) at the second COGENE workshop on Pharmacogenetics, Frankfurt, September, 2003; and (c) in Oslo, October, 2003. The working party explored the most effective ways to further the ultimate aim of harmonising European biobanks. The main elements of the work undertaken were: (1) to gather the evidence that led to the early recognition that the most appropriate way to proceed was to submit a CA bid; (2) to undertake the extensive networking activity that has brought together the collaborative CA submission group; (3) the writing of the CA bid itself; and (4) the link up with the P3G initiative (Leena Peltonen and Jaanus Pikani were leaders in both initiatives) which has now been ratified in the decision to formally merge.

B.4 The consortium and project resources

As this is a Coordination Action, our principal resource is the reservoir of expertise and knowledge amongst the senior researchers who comprise our collaborative team. Therefore this section focuses mainly on these individuals and their biosketches. We start, however, by briefly considering the obvious strength in depth of the collaborative team as a whole.

Our research consortium spans eight European nations. The group is extremely multidisciplinary and contains leading researchers in epidemiology, biostatistics, genetics, information technology, public health and in social, ethical and legal issues applied to bioscience. It contains a number of distinguished genetic epidemiologists, with specific expertise and experience both across many areas of methodology and in applied studies spanning most classes of design including: out-bred individuals; out-bred families; specialized designs based on twins; and studies based on genetic isolates. The group also includes scientists with a leading expertise in genome technologies and genome mapping. In specific relation to the aims of this particular CA, we also derive great strength from our collective involvement in some of the most significant population-based bioscience projects in the world. These include UK Biobank, the Estonian Genome Project, GenomEUtwin, EPIC, Biohealth Norway, ALSPAC and CARTaGENE. Because of a lack of space in this section, these projects are described in greater detail in section B5. In summary, there really can be no question but that our collaborative group is ideally constituted to undertake the programme of work outlined in the workpackages we describe. The researchers comprising this consortium are well published in international journals, however, space restrictions here have compelled us to omit each participant's list of 5 selected references in this section.

Participating researchers (listed in alphabetic order):

Max P. Baur is Professor and Chairman of the Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn. He is a current member of the German National Genomics Research Network (NGFN) and a speaker of the platform "Centers of Excellence for Genetic Epidemiological Methods (GEM) since 2002. From 1994-1997 he was a member of the Board of Directors in the International Genetic Epidemiology Research Network (IGES) and he the President of IGES from 2001-2002, and received the IGES Leadership Award in 2002. Since 1995 he has been a member of The Advisory Board of The Genetic Analysis Workshop (GAW), and was the President of the German Region of the International Biometric Society from 1992-1994.

Dorret Boomsma is Professor of Biological Psychology at the Vrije Universiteit in Amsterdam, The Netherlands. Her research interests are in behaviour genetics, psychophysiology and genetics and on molecular genetics of cardiovascular risk factors, depression and twinning. Most of the research is carried out in twin families, who have been registered with the Netherlands Twin Register (www.tweelingenregister.org). The group's applied research focuses on the design, conduct and analysis of large, longitudinal twin and genetic epidemiological studies aimed at dissecting the aetiological architecture of complex traits. The research conducted in her lab is funded by: Netherlands Heart Foundation, NWO, NIH, BPDRF and Sophia Foundation. The group participates in the EU funded GenomEUtwin project. Large scale collection of biological samples will begin in 2004 as part of the Dutch Centre for Medical Systems Biology (CMSB). Recent relevant publications:

Paul Burton is Professor of Genetic Epidemiology at the University of Leicester. His research involves methodological and applied work at the interface between traditional and genetic epidemiology. The main focus of the methods work undertaken by his research group

is the construction of mathematical models for complex traits, particularly when analysis is inherently difficult (*e.g.* complex pedigrees, non-random ascertainment, longitudinal phenotypes or traits mitigated by the effect of a treatment). The group's applied research focuses on the design, conduct and analysis of large epidemiological studies aimed at dissecting the aetiological architecture of specific complex diseases: particularly in cardiovascular and respiratory medicine. The group is involved in projects funded by: MRC, Wellcome Trust, British Heart Foundation, NH&MRC (Australia) and the Leverhulme Trust. Following the inaugural COGENE workshop (Cyprus, 2002) Paul Burton, with others, took a leading role in progressing some of the key ideas from the meeting. He is also extensively involved with UK Biobank: he was a member of the original Protocol Development Committee and now sits on the Scientific Committee and on the subcommittee charged with addressing "Questionnaires and Measurement". He leads one of six Regional Collaborating Centres (RCCs) that will form the collecting arm of UK Biobank. He sits on the "Health Services and Public Health Research Board" of MRC.

George Davey Smith is Professor of Clinical Epidemiology at the University of Bristol. He has extensive experience with epidemiological cohort studies and has written on methodological issues in genetic epidemiology, particularly with respect to Mendelian Randomisation and power/replication in association studies. His department is involved with a Regional Collaborating Centre in the UK Biobank project. Recent publications:

Mary Dixon-Woods, Senior Lecturer in Social Science and Health at the University of Leicester, is interested in social science applications to health and healthcare. She conducts methodological and empirical research in a range of areas. (1,2) She has a particular interest in the ethics and governance issues that arise in people's participation in research. (3, 4) She has just completed a major study of informed consent funded by the NHS, and a study of women's experiences of consenting to trial participation during pregnancy. Dr Dixon-Woods has recently been awarded a large grant from the UK Economic and Social Research Council to investigate the area of donations to children's cancer tissue banks. This project will, as well as conducting a review of ethics and governance issues in relation to tissue banks, draw on the disciplines of sociology and anthropology to understand the perspectives of clinicians involved in seeking donations and the perspectives of families of are requested to make donations.

Paolo Gasparini, Professor of Medical Genetics at Second University of Naples is Responsible for the Linkage Unit of TIGEM (Telethon Institute of Genetics and Medicine, Naples, Italy). He has expertise in mapping and cloning disease genes for several inherited disorder such as hereditary hemochromatosis, genetics of hearing loss, cystinuria, neurological genetic disorders. He has contributed to the identification of more than 10 disease genes including the most important ones for hearing loss. He is editor of the Connexin-Deafness homepage at www.crg.es/deafness. He is a member of the Forum of Genomes Programme Managers, one of the measures towards the European Research Area, the Commission has called to find synergies between national research activities in the field of genomes research in Europe. The Forum includes members from 25 countries in Europe. The Commission supports the Forum through the accompanying measure "COGENE". He is Author of more than 200 full-articles on peer reviewed international journals. Recent publications:

Jennifer R. Harris, Ph.D. is a senior researcher with Professor Competence at the Norwegian Institute of Public Health, she is the founder and director of the NIPH twin program of research in genetic epidemiology. She is also a special expert at the National Institute of

Aging working to develop genetic research directions for the behavioral and social sciences. She is the leader of the Ethics Core in GenomEUtwin and has extensive experience with research that integrates genetics and developmental approaches for the purpose of studying disease etiology, obesity, asthma, health behaviors, health perceptions and psychological well-being.

Thomas Hudson is Director of the McGill University and Genome Quebec Innovation Centre, and past assistant Direrector of the Whitehead/MIT Center for Genome Research. In addition to Dr. Hudson's leadership role in genome technologies and human and mouse genome mapping, Dr. Hudson is currently leading one of the genotyping centres involved in the HapMap project. Dr. Hudson's interests include complex trait mapping (for asthma, inflammatory bowel disease, etc.).

Bartha-Maria Knoppers, Canada Research Chair in Law and Medicine, is Professor at the Faculté de droit, Université de Montréal and Senior Researcher at the Centre for Public Law (C.R.D.P.). Currently, Chair of the International Ethics Committee of the Human Genome Organization (HUGO), she was a member of the International Bioethics Committee of the United Nations, Educational, Scientific and Cultural Organization (UNESCO) which drafted the *Universal Declaration on the Human Genome and Human Rights* (1993-97). She is Co-Founder of the International Institute of Research in Ethics and Biomedicine (IIREB) and a Co-Director of the Quebec Network of Applied Genetic Medicine (RMGA). Former member of the Canadian Biotechnology Advisory Committee and of the Standing Committee on Ethics of the Canadian Institutes of Health Research, she was named to the Board of Genome Canada in 2000. In 2002, she was elected *Fellow* of the American Association for the Advancement of Science, named *Officer* of the Order of Canada and received the Queen's Jubilee Medal. In 2003, she became Director of the international Public Population Project in Genomics (P^3G).

Jan-Eric Litton is Professor of Biomedical Computer Technology at the Karolinska Institutet, he is the director of Informatics for the Karolinska Institutet Biobank, and head of the database core in GenomEUtwin. His research team includes: P. Lindqvist (MSc Biomedicine, MSc Computer Science) who is the system architect of KI Biobank and is responsible for constructing the IT infrastructure at the biobank, with focus on data communication and storage; G.Petersson, Fil Kand. and B.Sc, Programmer; J. Söderberg, UNIX administrator; T. Mellin, IT administrator; O. Hansson, PC and UNIX technician; J. Carlsson, B.Sc., Database administrator for Oracle and the Swedish Twin Registry; A. Björklund B.Sc., Database software manager ; L. Jacobson B.Sc, Database manager; A. Ekman, M.Sc., doctorial student; A. Olsen, Oracle database manager, B. Rydh, Fil Kand, O. Bälter, Ass.prof., Tekn.Dr., Royal Institute of Technology.

Per Magnus, M.D., Ph.D. is a Professor in Community Medicine at the University of Oslo and the Head of the Department of Genes and Environment in the Division of Epidemiology at the Norwegian Institute of Public Health. His background is in perinatal and genetic epidemiology. He is the project leader for the Norwegian Mother and Child Study, -Child Study (see section B.5). The departmental resources include a modern laboratory for molecular epidemiology, with several biobanks and genotyping activities. The research group in his department is highly cross-disciplinary with training in biology, medicine, nutrition, statistics and public health.

John Newton is the Project Director, CEO and Principal Investigator of the UK Biobank. He is currently Honorary Professor of Public Health and Epidemiology at the University of

Manchester. He spent 10 years as a medical epidemiologist in the University of Oxford. He has also held significant research management posts for example as Director of Research and Development for two major teaching hospitals in the UK (Oxford and Southampton) and as leader of the research support group for the UK Government's Clinical Standards Advisory Group. His personal research interests are based on the use of largescale routine data sets for research, in particular the use of linked data from primary and secondary care for epidemiology and health service evaluation. He has recently completed an important national review of the use of disease registers in England and Wales which has formed the basis of UK Government policy in this area (www.ihs.ox.ac.uk/Register.pdf). He has been advisor to the Department of Health on the impact of legislation on confidentiality on research. He was until recently a member of the Scientific and Ethical Advisory Group for the GPRD database.

Leena Peltonen, M.D., Ph,D is a Professor and Head of the Department of Molecular Medicine at the National Public Health Institute in Finland and is a Professor in Medical Genetics at the University of Helsinki. She is among the leading molecular geneticists world-wide, and is one of the pioneers in the use of genetically isolated populations in the genetics-based identification of disease genes. She has coordinated molecular genetic studies in many international scientific networks including ESF Network of Neuropsychiatric Diseases, EU funded MS consortium and established two new Departments, one at the National Public Health Institute of Finland and one in the Medical School of UCLA, USA. She is the Director of the Center of Excellence in Disease Genetics in Finland, co-ordinator of the GenomEUtwin project, and has been actively involved in the national and international committees on the legal and social issues of genetics. She is a member of the Scientific Advisory Board, Marshfield Medical Research and Education Fdn's Medicine Program, and Chair of the Scientific Advisory Board Netherlands, Genome Initiative. She was the member of the international HUGO Council (2001-2002) and is the member of the UNESCO Bioethics Committee. Dr. Peltonen has received several many awards, and has produced 365 publications and 40 review articles.

Jaanus Pikani, M.D. has an extensive experience in public and healthcare management. He has been one of the initiators of the EstonianGenome Project (www.geenivaramu.ee) while leading the drafting of Estonian special law on population based genomic database (Human Genome Research Act).Janus has also been among the group to design the project and setting up private-public structure. Currenlty he is active as the chairman of the coordination committe of the Genome Project.

Elio Riboli has an M.D. and a Master of Science in Epidemiology . Between 1978 and 1983 he was a researcher at the National Institute for Research on Cancer in Milan. In 1983 he moved to IARC-WHO in Lyon, where he undertook the task of developing new research projects in the area of nutrition, nutritional status and cancer. In 1989 he initiated the European Prospective Investigation into Cancer and Nutrition (EPIC), which eventually included 26 centres in 10 European countries. Questionnaire date on diet and lifestyle have been obtained from about 500,000 study subjects, and blood samples from most of them. He has contributed to the development of major international cancer research activities based on the integration of epidemiology and laboratory research. In 1995 he was appointed Chief of the Unit of Nutrition and Cancer of IARC, whose main object will be follow-up of EPIC over the next decade and research into the role of nutrition, lifestyle, environment, genetics and metabolic and hormonal factors in cancer etiology. Selected publications:

Camilla Stoltenberg is an MD and has a PhD in epidemiology/genetic epidemiology. She is the director of the Division of Epidemiology at the Norwegian Institute of Public Health. Her

main research interest is perinatal epidemiology, particularly on effects of consanguinity on birth defects and infant death.She is PI of Biohealth Norway, which has received a grant of 6 million Euros for 2002-2007 from the Norwegian research council, and the Norwegian PI of a collaborative study with Columbia university, New York, on gene-environment interactions in an autism birth cohort. The autism cohort is based on the Norwegian mother and child cohort study (MoBa) and has recently been awarded a 13 million US dollar grant from NIH (NINDS).

Collaborators

In addition, there is great potential for collaboration from representatives from other major biobanks and epidemiological studies of outbred and isolated populations in Europe. We will remain open to new cohorts, initiatives and projects that can make a substantive contribution to the CA. One of the principal aims of WP 1 is to determine (early on in the CA) the criteria for inviting additional studies to join the consortium. Already during the development of this CA we have been approached by scientists with similar interests and needs for developing their biobank initiatives. For example, one such collaborating project will be **Centre for Medical Systems Biology** (CMSB), headed by Professor G.J. van Ommen. It is one of the two medical genomics centres of excellence recognized in 2002 by the Netherlands Genomics Initiative. It builds on long-standing and productive research networks that span six Universities and Research Institutes in Leiden, Amsterdam and Rotterdam. Genetic epidemiological cohort studies and high-throughput experimental approaches are used to discover and validate biological factors underlying complex diseases.

CA Project Effort Form
Full duration of project

	University of Bristol	University of Bonn	National Public Health Institute Helsinki	Mc Gill University	University of Leicester	Free University of Amsterdam	University of Montreal
Coordination activities							
WP1	1-2	1-6					
WP2			1-3				
WP3							
WP4			1-9	1-3			
WP5					1-3	1-3	
WP6					1-2		1-2
Total coordination activities	1-2	1-6	2-12	1-3	2-5	1-3	1-2
Management activities							
WP1							
WP2							
WP 3							
WP4							
WP5							
WP6							
Total management activities							
TOTAL ACTIVITIES	1-2	1-6	2-12	1-3	2-5	1-3	1-2

CA Project Effort Form
Full duration of project

	Estonian Genome Project Foundation	Karolinska Institutet	Second University of Naples	International Agency for Research on Cancer	UK Biobank	Norwegian Institute of Public Health	TOTAL PARTNERS
Coordination activities							
WP1							2
WP2			1-16	1-3			3
WP3		1-5			1-3	1-5	3
WP4							2
WP5						1-3	3
WP6	1-2					1-2	4
Total coordination activities	1-2	1-5	1-16	1-3	1-3	3-10	17
Management activities							
WP 1							
WP2							
WP3							
WP4							
WP5							
WP6							
Total management activities							
TOTAL ACTIVITIES	1-2	1-5	1-16	1-3	1-3	3-10	17-72

B.5 Project management

The project will be *co-ordinated* by the Norwegian Institute of Public Health, Oslo. We propose a project organisation based on the original P3G initiative and COGENE working group, and the work packages. A *steering committee* has been established whose 15 members are the leaders and co-leaders of the work packages plus the co-ordinator. The steering committee will meet twice a year and will organise additional telephone conferences when needed. In addition, we have established an *executive committee* consisting of Leena Peltonen, Paul Burton and Camilla Stoltenberg. The executive committee is responsible for implementing the project. The steering committee will act as a board and direct and delegate decisions to the executive committee, and will be responsible for preparing a list of publications (scientific and reports) generated under this CA. The executive committee will meet in person or on conference calls/video meetings at least once a month during the tenure of the project. Major strategic decisions will be made in the steering committee, while day-to-day decisions and preparation of decision-making in the steering committee is the responsibility of the executive committee. Updated and comprehensive information on the management of the project will be provided through a closed and an open web page. The CA builds on major pre-existing initiatives and projects, each with their own well-established management structures, infrastructure and routines for information dissemination as described previously in B.3.1 for GenomEUtwin and COGENE. Other projects in this CA are:

UK Biobank is the largest genome epidemiology initiative in the UK. Its set up is to be funded at £61M (≈ €89M) by MRC, Wellcome Trust, and the Department of Health. It will enrol a population-based sample of 500,000 participants aged 45-69 years from primary care. Baseline questionnaire data and standard clinincal measures are collected at enrolment. Blood samples will be analysed and/or stored to provide baseline biochemical measures and as a future source of material for genomics, transcriptomics, proteomics and metabolomics. Once enrolled, subjects will be tracked using routine information systems (primary-care, hospital and national registration systems) to identify key binary endpoints such as a diagnosis of cancer or death from myocardial infarction. Primary analysis will be based on nested case-control studies.

CARTaGENE is a population genomic project in Quebec, Canada that will recruit 65,000 adults between 25-74 years of age, representing 1.5% of the population in that age group. The demography of the population of Quebec is well-documented, and shows evidence of large diversity in centres such as Montreal, and founder effects in several regions. The CARTaGENE resource will be used by several research groups for different purposes, including the estimation of disease allele frequencies in various regions of Quebec. CARTaGENE will also allow for carefully selected case-control sub-cohorts of sufficient size that can be used to analyze genetic variants that predispose to or protect for common diseases. CARTaGENE will conduct research activities related to ethical and legal issues in population genetic research. Ultimately, a framework will be developed to guide the assessment of advances in genetics for the purpose of public health policy. A public Institute for Populations and Genetics (IPEG) was created as steward of CARTaGENE, for the benefit of the participating population.

The German National Genome Research Network – NGFN - was established in 2001 by the Federal Ministry of Education and Research (BMBF) with the objective to find genes involved in the aetiology of genetically complex diseases. Initial funding for a 3-year (180 million euros) allows for the establishment of five disease oriented networks (Cancer Network, Cardiovascular Network, Environmental Network, Infection/Inflammation

Network, Nervous System Network) and also support technological platforms (e.g. high throughput genotyping, expression profiling, sequencing, animal models etc). In addition, seven centers for Genetic Epidemiological Methods were established as a methodological platform to support the disease oriented networks in several areas, including: study design, data logistics, statistical analyses, development of common standards and quality management tools for data communication between clinical partners and genotyping facilities, monitoring of clinical projects (recruitment, sampling, workflow), establishment of common database structures, and establishment of training programs for young scientists in the field of genetic epidemiology. The call for NGFN2 has been published Nov 1st, 2003 for the next funding period of three years. The five well established disease networks will be complemented by two large population based cohorts within the GEM platform. KORA (Cooperative Health Research in the Region of Augsburg – PI: H.-E. Wichmann) is a cohort of 20,000 adults phenotyped for a broad spectrum of endpoints, with a biological specimen repository, genomic DNA of the 20,000 participants, and EBV immortalized cell lines of 1600 subjects. PopGen (PI: M. Krawczak) is a newly established regional (North-Schleswig Holstein) project. It's goals are: to identify and recruit all individuals newly classified as being affected for a relevant disease, to phenotype these individuals using standard criteria, and to create a repository for DNA and biological materials from this well defined base population of 1.1 million inhabitants. In addition a control group of 10,000 randomly selected probands will be sampled. Patients, controls, material and data will be accessible for the disease networks of the NGFN and any other scientific cooperation approved by the boards of these studies.

Biobanks for Health in Norway (BioHealth Norway: http://www.fhi.no/tema/biobank). The Norwegian Network of Human Research Biobanks and Health Studies is a large population based cohort established for genetic epidemiological research. The collection of samples and information is ongoing. When completed in 2006, the cohort will comprise blood samples and standardised health and exposure data from 450 000 Norwegian individuals of all ages, which is 1/10 of the Norwegian population. The population cohorts include two major projects: I. *MOBA:* The Norwegian mother and child study is a cohort of 100 000 pregnant women, 100000 children and 70 000 fathers. Approximately 30 000 mothers, 20 000 fathers and 25 000 infants are included in the biobank to date (Oct 2003). II. *CONOR:* Cohort of Norway. This cohort aims to include 200 000 adults, comprising eight different cohorts from various regions of Norway (including the HUNT (North Trøndelag), Tromsø, Hordaland and other health studies). 167 000 participants are included in the biobank to date (Oct 2003). The Norwegian Institute of Public Health (NIPH), Division of Epidemiology, is responsible for the network of biobanks funded by the Norwegian research council and organised in collaboration with the regional health studies at the four universities in Norway. The project is currently establishing a genealogy database for the total population of Norway, modelled after the Swedish generation database.

Public Population Project in Genomics (P3G) is an international consortium initiative in population genomics. Although the project was founded by members of CARTaGENE, GenomEUtwin, The Estonia Genome Project and U.K. Biobank, the project aims to link many other countries in order to catalyse the integration of resources, tools and know-how to perfect data management for improved methods of transfer and sharing. The motto is transparency and collaboration. Following a preliminary meeting between its proposed leader, Dr. Bartha Maria Knoppers, representatives from the four founding groups met in London in February 2003, in order to set the stage for the creation of P3G. A workshop was held in Montreal on July 2-4, 2003, which created 5 working groups and generated the following P3G Objectives: 1) To connect the leading public population genomics projects; 2) To provide necessary coordination, harmonization and standardization so the combined results be used for the advancement of science around the world; 3) To develop common understanding of

the socio-ethical and legal issues; 4) To foster a deeper understanding of the relative contribution of genetic and non-genetic determinants to health and disease; and 5) To transfer this knowledge to the international community so as to optimize benefits for public health. The next P3G workshop will be held in Manchester in December 2003. In addition to the high level of partnership that is already evident between P3G and this European collaborative project, there will be sessions dedicated to the integration of working groups, and division of labor in regards to the ambitious goals set forward by our groups.

The Estonian Genome Project

The Estonian Genome Project was to establish the Gene Bank, i.e. the database containing phenotype and genotype data of the majority of the Estonian population. The purpose for this Gene Bank is toconduct scientific academic and commercial research, to eludicate genetic causes of common diseases. The project aims to: create a collection of health care status descriptions encompassing the majority of the Estonian population; collect tissue samples of the donors, create LD maps of the donors; develop software that would enhance data analysis and findings generated from the project. The long-term goal of the Project is to improve public health using new knowledge and discoveries of genomic research.

The Estonian Genome Project is a Public-Private Partnership endeavor. The Project is managed by special purpose non-profit public organization Estonian Genome Project Foundation (EGPF) (www.geenivaramu.ee) in conjunction with privately held company AS EGeen (EG). EGPF is responsible for data and sample collection, storage and personal data security. In return for financial means EG has granted 25 years exclusive right to use the database for commercial research. The database is available for non-commercial research to public research institutions. The project has a Coordination Committee for strategic management comprising representatives from both parties – EGPF and EG. The Estonian parliament has approved special laws to regulate the genetic population database that also stipulates governance of the Estonian Genome project including supervising board that has representatives named by parliament, government and academy of science as well as an ethics committee.

International Agency for Research on Cancer/EPIC

The main goal of the International Agency for Research on Cancer is the identification of causes of cancer, so that preventive measures may be adopted against them, with emphasis on epidemiology, environmental carcinogenesis and research training. The main study being coordinated by the Unit is the European Prospective Investigation into Cancer and Nutrition (EPIC), was initiated in 1990 and supported by the Europe Against Cancer Programme of the EC. This is a multi-centre prospective cohort study investigating the relation between diet, nutritional status, various lifestyle and environmental factors and the incidence of different forms of cancer and other chronic diseases. The study includes 25 research centres in ten countries: France, Germany, Denmark, Greece, Italy, Netherlands, Norway, Spain, Sweden and the UK. Detailed data on diet and lifestyle, as well as biological samples (plasma, serum, lymphocytes and erythrocytes), are collected from all study subjects in an unprecedentedly large cohort of 520 000 healthy individuals. The subjects will be followed for the next 15 years, and subjects who develop cancer will be identified, and their personal habits (diet, lifestyle) and biological characteristics will be compared to those remaining in good health. The Unit is thus contributing to the development of major international cancer research activities based on the integration of epidemiology and laboratory research.

B.6 Workplan

B.6.1 Structure of the workplan

The workplan for this Coordination Action has been crafted around six workpackages (WPs), which will be overviewed strategically by a Steering Committee involving all WP leaders (see B5). They will also be subject to executive scrutiny by an Executive Committee (see B5).
Aside from managing the technical issues within each WP the main management activities will be related to overall integration of the WP results and the coordination and dissemination of information. Decisions regarding these activities will be the responsibility of the Coordinator and Executive Committee (taking account of strategic input from the Steering Committee). Other management activities will support individual WPs and the overall CA to reach stated goals. These will include organizing meetings, monitoring day-to-day needs of the projects, establishing the web-page procedures, and the centralization and support of functions common to the various WPs (*e.g.* the employment of a meeting rapporteur).

Each WP focuses on a specific coordination activity that will actively push forward the agenda of LSH-2003-1.1.0-2 to optimize and harmonize the management and research use of population-based biobanks and large cohort studies. The various workpackages are distinct but they interrelate in numerous ways. By integrating the individual workpackages within one CA we are able to take on what is an unashamedly ambitious task in a highly effective manner. By focusing the work within specialized work packages, we are able to concentrate a critical mass of expert scientists on topics that not only attract their interest, but also are highly pertinent to the development of their personal research strategies. At the same time, by combining workpackages with one another within a CA, and in particular by ensuring that all workpackages meet together at two large conferences, we also benefit from being able to make direct use of the links between WPs. For example, there will be cross-representation of many scientists on more than one workpackage. There will also be major areas of shared interest between workpackages. Thus, WP1, WP2, WP3 and WP4 will all work together to produce a list of major population-based biobank and cohort studies in Europe. Each WP needs this list for a different reason, and different WPs need different subsets of the list, nevertheless they can all work together on the time-consuming problem of identifying the relevant studies. Similarly, WP5 and WP2 are both interested in the harmonization of phenotypic assessment. However, WP5 will generally consider common complex diseases, while WP2 will particularly focus on potentially rare complex traits that happen to be at high prevalence in particular genetic isolates: the two WPs will share information emerging from their respective expert groups and can save work by avoiding duplicated effort. The same is also true for WP6 and WP2 in relation to social, ethical and legal issues in general populations and in genetic isolates respectively and for WP3 and WP4 in relation to information management systems for biobanks in general versus issues pertaining specifically to the management of DNA and genotype data. Complementarities such as these will accelerate the progress of the overall CA by synchronizing the time lines for developing key elements and by building the multidisciplinary networks needed to carry out future studies.

The specific themes of the workpackages have evolved from extensive discussions held at a series of international meetings on biobanks. These have included a number of workshops organized by the EU funded COGENE initiative and by the Canadian/European funded P3G initiative. One of the very positive outcomes of the COGENE initiative was the creation of a working party with a special interest in population-based biobanks and cohort studies. This working party has now fused with the leaders of the P3G initiative and it is this combined group, with some new members, that forms the core of the group submitting this CA. The

extensive discussions that have taken place have also highlighted the huge advantage that comes from the ability to build upon the intellectual foundations already laid down by the EU funded GenomEUtwin project. Several members of the GenomEUtwin research team are also part of the CA submission group, and it is no coincidence that the structure of the successful workpackages that were set up by GenomEUtwin with the particular aim of harmonizing research activity around large twin registries are very similar to those that we now propose for our CA relating to biobanks.

The individual workpackages, which are described in detail below, are:
WP1 – Identification of Potentially Informative Population Biobanks in Europe
WP2 – Opportunities for Future Biobanking in Europe
WP3 – Databases and Biobank Information Management Systems
WP4 – Strategies for Genotyping in Large Scale Biobanks
WP5 – Harmonizing the assessment of phenotypes and life-style exposures
WP6 - Ethical, Legal and Social Issues in Population Biobanks and in Data Sharing

As is quite evident from their titles, these WPs cover the fundamental topics that must, of necessity, be tackled if we are to promote biobanking from being an activity that is based primarily at the national level to an activity that is integrated at a European-level. A biobanking function at the "European-level" means identifying candidate biobanks and establishing optimal procedures for harmonization between them. Harmonization implies the use, where possible, of *complementary* protocols, this is not the same as the scientifically restrictive (and therefore unattractive) demand for *identical* protocols. The protocols that are most important to harmonize, and consequently warrant the greatest attention, are those pertaining to data management, genotyping, phenotyping, and ethical-legal constructs. These provide the hard-core of our proposal.

The individual WPs are congruous in structure. Each one relies on the calling together of a group of experts (many of whom will not have had the opportunity to work together before), the identification and assembly of up-to-data information and data, and the down-stream synthesis of recommendations. Within each WP, conclusions and recommendations will be summarized by one or more reports that will ultimately be posted on the CA web site. The major reports will also be presented and discussed at the concluding conference at the end the CA. These activities fall mainly in the realm of "coordination activities".

Crucially, the significant risks of the CA strategy are minimal. There is no dependence on novel methods, or new technologies; rather the work involves collecting, systematizing, standardizing and synthesizing a large amount of complex information that already exists. Furthermore, if any particular WP is unable to meet its specified deliverables this will not affect the progress of other WPs; indeed it may well highlight potential problems, that when identified and solved may move our agenda forward. This lack of risk is in stark contrast to the marked risk that would be incurred if one elected to miss out the CA step and instead went straight to projects that purported to produce definitive solutions to some of the same problems without exploring the problems themselves first. Genetic epidemiology is littered with good ideas that ought to have worked, but unfortunately didn't (almost always because things were more complicated than they first seemed). Many of these failures have been expensive and this has given genetic epidemiology a checkered reputation. We would like to step back from this image and, instead, spend a moderate amount of time and money thinking properly about the strategically important research problems. Once this thinking has been done, the way forward should be greatly clarified. It is true that it is still likely to be expensive. It will almost certainly require major funding of the order of an integrated project

to move any harmonization program forward significantly, but at least the groundwork will have been done and the costs, benefits and risks will be able to be properly weighed up before any major funding decision is taken.

As demonstrated by the Gantt chart, the ultimate aims of this CA demand a proper integration of the recommendations of the individual WPs. We do not underestimate the magnitude of that task, but integration will be assisted by cross representation on the WP expert groups and by the two joint conferences. If no obvious funding opportunity for a substantive major integrated project has arisen by the time of the concluding conference, one of the primary aims of that conference will be to determine future strategy to advance the harmonization of biobanks. It would obviously be both convenient and strategically sensible if an EU-based funding opportunity was to arise on the back of a successful CA, but we are in no sense dependent on this. The CA would put European population genomic researchers in a lead position to seek money from other sources both within and outside Europe.

WORKPACKAGE 1

Identification of potentially informative population biobanks in Europe

Leaders: George Davey-Smith, Max Baur

Aims and objectives

The aim of this workpackage is to identify and establish a network of completed, ongoing or planned biobanks in Europe that can contribute usefully to co-ordinated investigations of the genetic and environmental determinants of complex diseases. It is now recognised that most worthwhile studies in genetic epidemiology demand extremely large sample sizes. Furthermore, the recruitment of participants, their maintenance within a cohort, the collection of information and biomaterials, phenotyping, and genotyping are all extremely expensive (both in time and money). If Europe is to remain at the cutting edge of population-based biomedical research, it is realistically going to become necessary not only to conduct large expensive studies, but also to be able to properly share and combine information between these major studies. It is therefore clear that a reliable source of information listing completed, ongoing and planned population-based biobanks and large cohorts in Europe would be invaluable. Furthermore, if this could be linked to information about the accessibility and availability of material and data from each study, one would have a resource that could be of extreme value for any joint European endeavor in this highly competitive field of research.

Work Programme

As a first step it will be necessary to determine criteria for those biobanks and population cohorts that will be most informative. An expert group will be convened and will then meet as one component of the *Initial Conference for Population Biobanks for Health*. The expert group will determine which selection criteria are most appropriate. These will almost inevitably include:

(1) Sample size*:** The main effects of common genetic factors influencing complex diseases will be small (in general relative risks < 2 and often << 2), therefore studies require a large sample sizes to yield sufficient cases or to provide adequate statistical power for any useful genetic epidemiological analysis. ***(2) Availability of DNA: There is no question that the population based cohorts and biobanks which can contribute most significantly to modern genetic epidemiology are those in which DNA is available on the majority of individuals recruited. Although typing technologies are constantly improving, the number of genotypes to be produced by SNP-typing may be very large, particularly in a dense genomewide association screen. Therefore the amount of DNA that is available may become a limiting factor if the cohort is used for many projects. ***(3) Availability of information about family***

structure and recruited relatives: Although many of the studies conducted using population cohorts are case-control studies based on individuals, there may be a necessity to have diseased individuals and their relatives available for genotyping in some study designs (transmission oriented linkage analyses for complex disease models, TDT design, etc). ***(4) Population characteristics:*** The identification of genetic contributions to complex disease can be facilitated by use of particular population groups: isolated populations, admixed populations, and inbred populations. ***(5) Quality of phenotype information:*** There is a trade-off between quality of phenotype information and sample size with respect to statistical power. At a given sample size, studies with better characterisation of categorical outcomes (for example with pathological or genetic subclassification of tumour type for cancers) or of continuous traits (for example, through repeat measures of blood pressure or by utilising more precise markers of insulin resistance or atherosclerosis) have greater statistical power. Furthermore, some phenotype measures that are very precise are expensive and/or time consuming to collect, and are simply not available on very large sample size studies. ***(6) Availability of genotypes:*** Ideally, any genotype which has been produced for an individual recruited into a cohort should be returned to a central information store for that cohort and then made available to other users of the resource in a standardised format which provides full information regarding the marker and its actual genotypic expression. ***(7) Availability of biomaterial:*** Beyond the information of genotypes and phenotypes it will be of utmost importance to know about the availability of biomaterial beyond DNA (tissue, immortalized cells, organs, etc.) for functional studies subsequent or additional to the gene mapping studies. Non-availability under (6) or (7) may be consequent upon scientific limitations, ethical or legal restrictions, rules imposed by the agency funding the original cohort, or even the personal whim of an unreasonable investigator – it is important to know if any of these situations pertain.

Once the expert group has established the criteria to be used to select and classify population-based biobanks and cohort studies, a report will be published on the CA web site. The expert group will then check potential studies against the chosen criteria. Such studies will be identified by all available means such as: nomination by the attendees of the initial conference; response to the web posting; web searches; literature searches; personal contacts by way of any kind of correspondence *etc*. A particularly useful source of information will be the funding agencies for biomedical research across Europe with their comprehensive knowledge of ongoing and planned activities in the field. This work will be undertaken in collaboration with WP2 and WP3. As the first biobanks complying with the established selection criteria are identified and categorized, a content management system will have to be established in parallel in order to properly catalogue the studies electronically on the web. This tool must necessarily insure that information about studies that have already been entered can easily be updated and that new studies can quickly be added to the database.

In addition to meeting electronically to consider candidate studies for inclusion in the growing database, the expert group will formally reconvene at the *Concluding Conference for Population Biobanks for Health*. A final report summarising the current state of the database will have been prepared prior to the conference and will be presented and discussed. The conference will provide an important opportunity to discuss future strategy regarding population biobanks in the light of the state of the database and the main conclusions reflected in the final reports of other work packages.

Deliverables

(A) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(B) Report on study selection criteria. *Time: 12 months*

(C) Web-based content management system goes live. *Time: 18 months*

(D) Final report on status of studies in the database. *Time: 33 months*

(E) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

Milestones

(A) The sole "decision point" will occur when the selection criteria are selected as this will determine the studies that then become eligible to join the initiative. *Time: 12 months*

WORKPACKAGE 2

Opportunities for Future Biobanking in Europe

Leaders: Paulo Gasparini, Elio Riboli

The use of isolated populations to reduce disease heterogeneity of complex disorders has already proven very useful in identifying DNA polymorphisms associated with complex diseases and quantitative traits (Peeltonen et al. 2000, Angius et al. 2002, Gianfrancesco et al 2003, Revsnidottir et 2003, Thorgeirsson et al 2003, Krjstiansson et al 2002). The study of complex traits in geographically and culturally isolated populations is particularly useful because the entire population can be analyzed, the relative weight of environmental variation can be controlled and genetic factors can be more easily identified. In these genetically and culturally homogeneous populations, a large proportion of individuals presenting a given trait is likely to share the same trait-predisposing gene inherited from a common ancestor. Furthermore, inbreeding, typical of small communities, reduces genetic heterogeneity and increases homozygosity, providing greater power for detection of susceptibility genes.

Aims and objectives

a) to identify new opportunities for biobanking in Europe with a particular focus on genetic isolated populations.

A survey to identify completed, ongoing or planned European biobanking initiatives with a particular focus on studies of isolated populations will be carried out during the first 12 months of the CA. Currently, there are known initiatives in Italy, Croatia, Sweden, Iceland, Finland and Scotland. Additional ones are planned in Romania, The Netherlands, and Spain. This survey will be performed during the first year of the CA making use of literature review, web-searches, e-mail and telephone contact with known investigators as well as direct contact with representatives of each European country in Brussels. Final results, including a map showing the locations of key initiatives, will be made available as a report through the website for the CA. This activity will be carried out collaboratively with WP1 and WP3.

b) to establish common criteria for selection and collection of these populations.

We will aim to provide the international community with guidelines for the selection of genetically isolated populations for study. This work will sit nicely alongside the work programme of WP1 which is focusing on outbred populations. In addition, in collaboration with WP5 we will consider issues pertaining to harmonization of the assessment of phenotype and life-style exposure. Based on the collective experience within the CA consortium and of the larger expert group invited to contribute to the workpackage, we will explore standard criteria for selection of ideal populations for study which will include: availability of genealogical data; key parameters of population structure (*e.g.* number of founders, number of living inhabitants, endogamy rate, emigration/immigration rate); and the possibility of assessing and quantifying enviromental determinants. An expert group will be convened informally at the start of the CA with initial interaction taking place electronically or at face-to-face meetings that occur for other purposes. The group will meet formally at the *Initial*

Conference for Population Biobanks for Health and will agree upon a refined strategy to take things forward (including interaction with other workpackages) and will identify any additional experts who should be coopted onto the group. They will subsequently meet electronically or informally, with the intention of producing a final report at 33 months which will then be presented and discussed at the *Concluding Conference for Population Biobanks for Health*. It will then be revised, as required, and posted on the CA web site.

c) to stimulate development of analytic tools. Many of the biostatistical tools that are used to describe genetic data, construct pedigrees or draw even simple statistical inferences are poorly configured for the study of genetically isolated populations. These populations have an abnormally strong latent structure, many idiosyncratic features, and a complexity which demands the development of specific tools. Thus, we aim to establish a network of statisticians, bioinformatics and software constructors with experience and expertise in working with genetic isolates to explore issues pertaining to database construction, statistical analysis (including statistical power), the construction and drawing of pedigrees and data mining. We will build upon strong pre-existing interactions with industry (including IBM and HP) and invite them to join the collaboration. Their involvement should help catalyse later attempts to produce final products such as new software and/or new algorithms.

d) to solve common ethical issues. Due to their unusual structure and setting, there are many social, ethical and legal issues that are particular relevant to genetically isolated populations. In collaboration with WP6, we will explore the relevant issues. We will ensure that our expert group includes an ethicist with a special knowledge of genetic isolates and that there is cross-representation with the expert group on WP6.

Deliverables

(A) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(B) Report on European population-based biobanks in genetic isolates *Time: 12 months*

(C) Final report to address issues specific to the study of genetic isolates pertaining to: the selection of populations for study; statistical analysis, analytic tools and database construction; and ethical-legal issues. *Time: 33 months*

(D) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

(E) Final report to be posted on worldwide web. *Time: 36 months*

Milestones

(A) The sole ”decision point” will occur at the *Initial Conference on Population Biobanks for Health* when we identify the members of the expert group. *Time: 6 months*

WORKPACKAGE 3

Databases and Biobank Information Management Systems

Leaders: Jan-Eric Litton, Per Magnus, John Newton

Aims and objektives

The **long-term aim** of this workpackage is to aid research into the causes of common, complex diseases by arriving at a common strategy that will provide an efficient network of databases for population biobanks and cohorts in Europe and Canada. The challenge is to easily combine information from biobanks and cohorts so that very large sample sizes are generated, providing relevant, high quality and correctly documented sets of data to

researchers. Current confusion regarding gene-disease associations for a variety of disorders is to a large extent due to the severe lack of statistical power in the empirical studies that are presently being performed. The lack of a general information management system for biobanks is a bottleneck for processing the increasingly complex data structures incurred in modern clinical and epidemiological research.
The population cohorts that will participate in this initiative have developed or are in the process of developing diverse technologies and logistics to keep track of study subjects, environmental exposures, phenotypes, biological samples and analytic outcomes such as genotypes and gene expression profiles. The first objective in this work package will be to review existing and planned systems, and to arrive at a consensus on the requirements for a general information management system for biobanks.
The second objective is to agree on the unique identities for biobanks as well as for the secure identities for subjects from which samples are taken. However, the basic unit that requires an identity is the single specimen, which again may be further divided such that aliquots can be widely exchanged between laboratories and even be pooled for some analytic purposes. For each aliquot a number of analytic results will be reported, and this information, including meta-information on the validity and use of the laboratory results, must be easily traceable in the databases. Thus, samples and results must be localisable in a hierarchical system based on an agreed-on identity structure. Also, transactions and permissions for use must be built into the database.
The power of biobank-based research will increase enormously if multiple biobanks are interconnected to enable sharing of information and samples. The third objective is therefore to arrive at a complete strategy for biobank communication including a common nomenclature, and compatible software techniques.
The final requirement for biobank communication is a well-defined protocol for transmitting information between biobanks. This will necessarily incorporate a central biobank registry. Each biobank joining the communication network will use well-defined methods to register meta-information about the biobank contents in terms of samples and information. In a later stage, assuming a biobank wants to perform a specific search, it sends a request to the central biobank registry. Based on this request the registry will send a reply with information describing the biobanks containing such information. Finally, the requesting biobank sends the actual query to these biobanks in order to get results. In order to achieve these communication strategies, one may use Web Services. This is a highly standardised platform based on the SOAP protocol, which in turn is based on the widely-spread XML standard. Web services provide an optimal platform for communication between biobanks since it is entirely built on open standards and not connected to any particular vendor or programming language. Thus, it will enable biobanks with completely different IT infrastructures to communicate with a common protocol.

The **primary aims** of WP3 will be:
To convene an expert group as one part of the *Initial Conference on Population Biobanks for Health*. At the conference, this group will consider additional experts that should be invited to be involved in the work, and agree upon a strategy for on-going development of the scientific aims outlined above, within the necessarily limited framework of a Coordinated Action
The expert group will then produce a number of reports that will address the key issues underpinning the objectives detailed above: (a) A complete inventory of large European biobank projects and their information and storage strategies (18 months); (b) a format and variable standard for European biobank information and communication; (3 years) and (c) a biobank communication standard based on XML (3 years). In commencing the work in relation to (b) and (c), it is unclear whether the reports will conclude with a definitive scientific solution to each problem or with a list of further questions that need to be addressed

in order to arrive at these solutions. This is why it is entirely appropriate that this work be undertaken within the framework of a CA allowing relevant experts to pool their massed expertise.
The main conclusions of WP3 will be summarised in a final report [bringing together (a),(b) and (c)] which will be presented and discussed at the concluding Conference on Population Biobanks for Health and then, following appropriate revision, posted on the worldwide web.
Deliverables

(A) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*
(B) Interim report representing a complete inventory of large European biobank projects and their information and storage strategies. *Time: 18 months*
(C) Final report which will also address format and variable standards, communications standards and transmission policies. *Time: 33 months*
(D) Debriefing, presentation and discussion of final report, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*
(E) Final report to be posted on worldwide web. *Time: 36 months*

Milestones

(A) The sole "decision point" will occur at the *Initial Conference on Population Biobanks for Health* when we will determine the full composition of the expert group. *Time: 6 months*

WORKPACKAGE 4

Strategies for Genotyping in large Scale Biobanks

Leaders: Leena Peltonen, Tom Hudson

Aims and objectives

We plan to create an operational infrastructure for the evaluation of ongoing large-scale genotyping efforts in population cohorts, as well as provide a forum for expert opinions in regards to genotyping methods and quality assessment of these methods. We will collect information from pre-existing genotyping projects involving large population cohorts in Europe, and share it on the consortium web site. Based on the collective experience of consortium members, we will explore issues pertaining to the selection both of technologies and of sampling criteria for study design in future European research projects aiming to utilize these cohorts in genetic studies of complex traits and disease phenotypes. We will establish a network of statisticians to address the issues of statistical power of various study samples. Ultimately, our activities will provide the international community with advice and standards that will promote harmonization in regards to selection of markers, genotyping quality control and cost, data collection and storage, and genotype database structures.
Description of the work

The workpackage will consist of the following elements:

1) Exploration, integration and posting on-line (www) of existing information about strategies for genotyping in large cohorts (derived both from publically accessible databases/information sources and in the possession of individual experts in our research consortium); 2) Exploration, summary and posting of issues pertaining to standardization, quality control and cost of genotyping and related data storage and handling systems to assess pre-existing technologies and facilities; and 3) pooling of the collective expertise of statisticians to provide guidelines for genotyping strategy in large population cohorts.

For the purposes of 1), and in collaboration with other workpackages in this CA, we will establish a central web site for the European population genetic community that will provide information and links to genetic studies that involve large population cohorts in Europe. The web site will also contain links to the ever increasing number of databases relevant to population genetics (such as allele and haplotype freqencies in various populations). The web site will provide information on genotyping services that are available (location, technology, costs, etc.).

In relation to 2) we will establish a genotyping technology assessment centre. Given that genotyping technologies are evolving at a rapid pace, it is imperative that we not only be aware of developments, but that we be able to make objective comparisons between technologies, in regards to cost, throughput and accuracy. Every 6-12 months, we will establish a new set of test markers that will be shared with participating genotyping centres and new technology providers. We will undertake quality control exercises whereby particitating centres will analyse the test markers on a reference DNA set. We will compare results and report back the results to participating laboratories. Similar exercises used in the Human Genome Project and the HapMap project have been shown to help improve quality. Furthermore, we suggest this information to be essential when planning for ever larger and more expensive population genetic projects.

For the purposes of 3) we will establish a statistical network. It is widely acknowledged that the statistical expertise required for population genetic studies is lacking in most countries. We propose to create a network of such leaders in Europe, in order to promote the exchange of information, and explore and clarify pivotal issues (such as size of cohorts, advantages of specific populations, selection of markers based on allele frequencies, density, LD attributes, etc.). In collaboration with WP2 and WP3 we will establish a genotype database group. This group will establish database structures, definitions, language for data exchange, etc. that will allow comparisons of data among population genetic projects, including both outbred populations and genetic isolates.

We will organize bi-monthly meetings to exchange ideas, monitor progress, and prioritize specific issues involving high throughput genotyping and study design. The information generated by each group will be consolidated at workshops attended by leaders of large genotyping, phenotyping, and information technology centres. We will facilitate interactions with industry, and benefit from their expertise in regards to existing and future technologies.

Deliverables

These deliverables dovetail very naturally with the aims and deliverables of complementary workpackages in the GenomEUtwin project. See Leena Peltonen, form A2

(A) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(B) ”Genotyping cybercenter” providing on-line information about:

a) Major genotyping efforts ongoing in large European population biobanks and cohorts
b) Genotying quality reports – which will provide a comparison of genotyping accuracy among high throughput technologies and laboratories.
c) Genotyping cost reports. We will monitor costs offered to external clients by genotyping core facilities in Europe and worldwide.

Time: on-line 12 months, fully operational 33 months

(C) Annual report entitled ”Statistical issues in population genomics”. The report will contain a review of the literature, power tables, state-of-the-art commentaries on study design, etc. *Time: 12, 24, 36 months*

(D) Standard Operation Procedures (SOPs) for communicating genotyping data, as well as collection and storage. *Time: 33 months*

(E) Increased European cohesion and consensus on critical issues concerning the design and performance of research projects using large population cohorts. This will include involvement of junior investigators in the meetings of the experts of this WP. *Time: throughout CA*

(F) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

Milestones

(A) The sole "decision point" will occur at the *Initial Conference on Population Biobanks for Health* when we will determine the full composition of the expert groups. *Time: 6 months*

WORKPACKAGE 5

Harmonizing the assessment of phenotypes and life-style exposures

Leaders: Paul Burton, Dorret Boomsma, Jennifer Harris

Aims and objectives

The proposed Coordination Action will provide a substantive step towards the ultimate goal of optimizing the utility of future biobanking activity in Europe. A pivotal element of this goal is to develop a harmonized approach to the assessment of a wide range of complex phenotypes/traits (subsuming both disease phenotypes and intermediate endophenotypes) and environmental or life-style exposures.

Whilst it would be inappropriate to demand that all future European biobanks adopt *precisely the same protocol* for phenotypic assessment and exposure measurement, there is potentially great value in adopting a strategy with **long-term aims** to produce: (i) a concise "minimum data set" representing a common core assessment protocol for all phenotypes and exposures that *could* in principle be applied to all future European biobanks; (ii) agreed principles pertaining to the comprehensive assessment of key complex traits that go beyond the limitations imposed by the minimum data set itself; (iii) agreed principles pertaining to the comprehensive assessment of background data on environmental and life-style exposures and lifestyle. Armed with an agreed minimum data set and a consensus on the stated principles, major European biobanks and cohort studies would be better placed to collect data that: (a) meet current scientific "best practice" (recognizing the realistic bio-clinical limitations inherent to large epidemiological studies); (b) can be synthesized effectively across studies; (c) maximize the utility of the substantial phenotypic reservoirs that already exist within major cohort studies across Europe; and (d) are sensitive to potential gender/sex, age and cultural/ethnical differences in phenotypes and diagnostic techniques.

While the **long-term aims** outlined in the preceding paragraph may go beyond what is achievable within a Coordination Action they are undoubtedly important and it is crucial that we attempt to move towards them in manageable stages. Building on the relevant experience already obtained in major international initiatives such as COGENE and P3G, and in key individual studies such as GenomEUtwin and UK Biobank, the **primary aims** of the phenotyping workpackage represent logical steps in this process.

The phenotyping workpackage will bring together international experts (clinical, population and biological scientists) from Europe and elsewhere. Based on preliminary interactions via electronic media, and at the *Initial Conference on Population Biobanks for Health*, the preliminary **primary aims** of the phenotyping workpackage will be:

(1) To refine and agree upon the strategy that should be adopted in taking these next steps towards the long-term scientific aims (i-iii) stated above.

(2) To select the complex phenotypes and exposure classes that should be addressed with the highest priority. This prioritorisation will define a series of research subprogrammes each of which will focus upon a system-specific group of phenotypes (eg musculo-skeletal, metabolic etc) or a coherent class of exposures.

(3) To identify appropriate subgroups of experts to work on each subprogramme and to sit on an integrating committee.

On-going work within the workpackage over the funded period will then be based upon the activities of the research subprogrammes while the integrating committee will overview the activity within each subprogramme and will ultimately bring everything together to produce a final report at the end of the CA. The additional **primary aims** of the workpackage will be:

*(4) To undertake a complete workup of one major complex phenotype, that is **major depression:** a particularly complex multidimensional phenotype, which the World Health Organisation predicts will be ranked second as a global health burden (as measured by disability adjusted life years) by the year 2020* (The Global Burden of Disease: Summary. Eds: CJL Murray, AD Lopez, W.H.O.,1996). *This will act as a good exemplar of the types of procedure that will eventually have to be followed for **all** major phenotypes and exposure classes, if we are ultimately to attain the long-term aims stated above.*

(5) To explore the key issues (strengths, weaknesses, opportunities and threats) in relation to other common complex traits and life-style exposures that have been identified as priorities but are not designated exemplars, and in relation to potentially important rare traits that are at high prevalence in particular isolated populations.

(6) *To produce a final overall report that summarises the work undertaken across the whole workpackage (research subprogrammes and integrating committee). Final reports from the individual subprogrammes will appear as appendices to the overall report. The final report will be presented and discussed as one element of the* Concluding Conference on Population Biobanks for Health *to be held at the end of the Coordination Action and, following any final revisions, a definitive version will be posted on the worldwide web.*

Implementation of the work package

The workpackage will consist of the following elements:

(1) As a key component of the *Initial Conference on Population Biobanks for Health* to be held at the start of the CA, we will bring together experts who can make a substantive contribution to debate and discussion regarding the preliminary **primary aims (1-3)** stated above. After the conference, the designated integrating committee for the phenotyping workpackage will produce a preliminary report summarising the discussions, conclusions and recommendations.

(2) Each research subprogramme will be led by an expert chairperson and deputy chairperson. Each subprogramme will work towards **primary aims (4-5)** on the basis of appropriate literature review, electronic interaction, and (if appropriate) face-to-face meetings. It is to be anticipated that the integrating committee and the research subprogramme working on major depression will require more extensive interaction. Individual research subprogrammes will, wherever possible, include participants of previous initiatives that have aimed for consensus positions on specific components of a complex phenotype – *e.g.* the ISAAC study of asthma and allergies. Each subprogramme will consider harmonization pertaining to information from many sources including: (a) interviews and surveys; (b) blood samples; (c) other biological samples; (d) physiological

measurements; and (e) other sources (*e.g.* link to routine information systems or intensive ambulatory assessments).

(3) In working towards **primary aim (6)**, each subprogramme will provide updates on progress and will prepare a comprehensive final written report to be submitted to the integrating committee within 1 year of the *Initial Conference on Population Biobanks for Health* (the detailed report pertaining to major depression will be submitted within 2 years of the initial conference). At the end of the work package the integrating committee will prepare a comprehensive final report with the subprogramme reports as appendices. This will be presented and discussed at the *Concluding Conference on Population Biobanks for Health*. After the concluding conference, the final report will be subject to appropriate revision and made available on the worldwide web.

Deliverables

(A) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(B) Report following initial conference regarding **primary aims (1-3)**. *Time: 12 months*

(C) Final report (**primary aim 6**) addressing specific phenotyping recommendations for major depression and all key issues pertaining to other phenotypes and exposures (**primary aims 4,5)**. *Time: 33 months*

(D) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

(E) Final report to be posted on worldwide web. *Time: 36 months*

Milestones

(A) The sole ”decision point” will occur at the *Initial Conference on Population Biobanks for Health* when we will determine which phenotypes and exposures will be viewed as priorities, and which experts will sit on each subprogramme and on the integrating committee. *Time: 6 months*

Long-term strategy

On successful completion of the workpackage as described, European bioclinical and population science researchers will be in a position to roll out a definitive phenotype harmonization research program to address **all** three long-term aims across **all** prioritised phenotypic classes and exposure categories. This next stage of the initiative will be expensive and will require additional funding that goes beyond what may be applied for under a CA application. Depending upon the ultimate choice of topics for later calls for FP6, it is to be hoped that it will be possible for any work undertaken within the phenotyping workpackage under this CA to lead directly on to a definitive phenotype harmonization programme as a core component of a later application for a Network of Excellence or Integrated Project. From the strategic perspective of the EU, this would be the most coherent way to take things forward. However, if no appropriate topic arises under later calls for FP6, the phenotyping workpackage should be seen as placing European researchers at the cutting edge of international endeavours in the area, and that we would then be ideally placed to submit applications for funding for the definitive phenotypic harmonization programme from elsewhere.

WORKPACKAGE 6

Ethical, Legal, and Social Issues in Population Biobanks

Leaders: Jennifer Harris, Bartha-Maria Knoppers, Jaanus Pikani, Mary Dixon-Woods

Why a CA workpackage in ethical, legal and social issues of biobanking?
Population-based biobanks engender a host of scientific and ethical challenges. While the scientific challenges are met by rapid integration of new methodologies into labs worldwide, the picture is quite different regarding the ethical challenges. Socio-ethical and governance procedures for the establishment and use of biobanks have not kept pace with the rapid technological advances that have increased the demand for biobank studies. It is often the case that ethical protocols are not unified between projects within the same country or between countries. Rather, each country grapples with the same sets of questions (e.g. consent, data confidentiality, transfer of biological samples), and then adopts their own solutions. This situation poses clear hindrances to the growing interest in internationally coordinated biobank initiatives where ethical procedures have implications for nearly all key components of collaboration (e.g. data collection, analysis, sharing and storage of data). An increasingly common scenario is that researchers developing international projects are confronted with complex information regarding ethics regulations which they are ill-equipped to interpret and coordinate. Accordingly, the success of European multinational biobank initiatives will hinge on the coordination and harmonization of ethical-legal and governance issues.

Aims and Objectives
The goal of this workpackage is to increase the scientific and commercial value of pre-existing and planned population-based genetic databases and biobanks through coordination and harmonization of the ethical-legal and governance issues. We will establish a platform of ethical-legal and governance criteria that are consistent with international norms and will enable data and sample sharing for research purposes within a defined ethical framework.

The work subsumed under the WP can be broadly divided into two phases
Phase I consists of mapping extant database activities related to the major ethical questions raised by these projects. This includes examining questions of transparency, consent, recruitment, controlled access by 3rd parties, data sharing, role of oversight bodies, identifiability and privacy protection, commercialization, participant rights to withdraw, destruction of data, data confidentiality, informing participants of results. Defining and describing the current status of these issues will provide a comparative overview of current practices at major European and Canadian biobanks. A main objective under the **Phase I** mapping effort is to guide elaboration of collaborative biobanking projects such that they conform to international laws and norms derived from multiple sources including international: human rights law, ethical norms, professional norms, consumer positions and Canadian norms. This will require conducting an extensive overview and analyses of international legal and ethical norms governing the enumerated topics of Phase I. Such information is difficult to extract and its summary requires expertise in bioethics and law. Therefore, this aspect of the WP will be coordinated through our partnership with the Genetics and Society Project (GSP) at the Centre de recherché en droit public (CRDP), University of Montreal.

Phase II of this workpackage is a consensus building process. Information will be distilled from the phase I report in order to explore the possibility of creating a platform of ethical-legal and governance criteria that is consistent with international norms and will enable sharing data and samples for the research purposes within a defined ethical framework for

Europe. It is critical to coordinate this consensus finding effort with activities underway at the various projects associated with this CA. Specifically, P^3G and COGENE, together with other international bodies (e.g. UNESCO, WHO, CE, OECD, HUGO and HGVS) are developing a common taxonomy/nomenclature for biobanks with particular emphasis placed on an overarching ethical framework. This work strives towards maximum harmonization of ethical review methodologies used to evaluate population research protocols. For example, current approaches are built upon the ethics of protecting the individual and do not sufficiently address population research or even group benefits (promotion/prevention/protection) but only group "harms". Building such an understanding and consensus on norms and methodology requires networking and education.

Deliverables

(A) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(B) The main Phase I deliverable will be the interim report on ethical practices in Europe, Canada and other countries *Time: 18 months*

(C) The principal Phase II deliverable will be the final consensus report which will explore the possibility of an internationally consistent socio-ethical and governance platform in the light not only of the work undertaken in Phase I, but also consultation with international, regional and national partners. *Time: 33 months*

(D) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

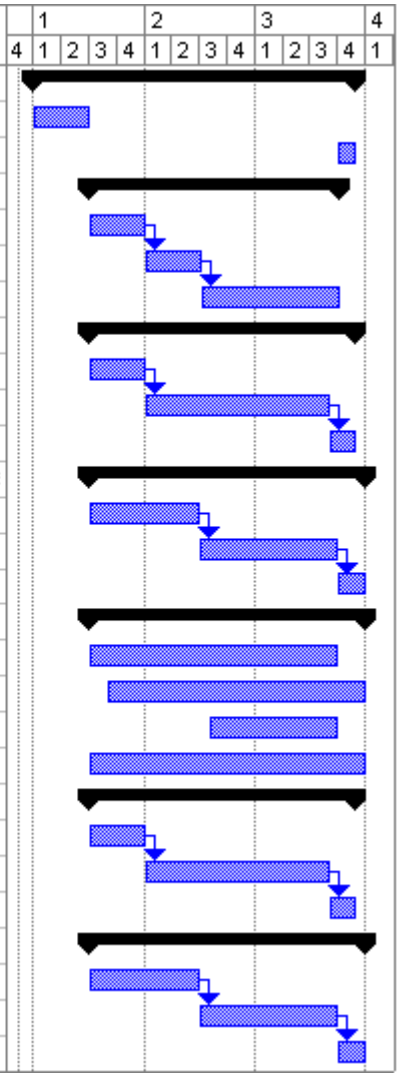
(E) Final report to be posted on worldwide web. *Time: 36 months*

Milestones

(A) The sole "decision point" will occur at the *Initial Conference on Population Biobanks for Health* when we will determine which additional experts should be invited to join the expert group. *Time: 6 months*

B.6.2 Workplanning

ID	Aktivitetsnavn
1	**Deliverables common to all WPs**
2	D1 Initial conference
3	D2 Concluding conference
4	**WP1 Identification of European biobanks**
5	D3 Report on study
6	D4 Web-based content management system
7	D5 Final report
8	**WP2 Opportunities for future biobanking**
9	D6 Report on biobanks and genetic isolates
10	D7 Final report
11	D8 World wide web report posting
12	**WP3 Databases and information management syste**
13	D9 Interim report on information storage
14	D10 Final report with standards/policies
15	D11 World wide web posting
16	**WP4 Strategies for genotyping**
17	D12 Genotyping cybercenter
18	D13 Annual reports
19	D14 SOP's for genotyping issues
20	D15 Statistical network
21	**WP5 Phenotype harmonisation**
22	D16 Preliminary report
23	D17 Final report
24	D18 World wide web report posting
25	**WP6 Ethical, legal and social issues**
26	D19 Interim report on ethical practises
27	D20 Final consensus report
28	D21 World wide web report posting



d) Detailed work description broken down into workpackages:

(Workpackage list, use Workpackage list form below)

Workpackage list (full duration of project)

Work-package No[1]	Workpackage title	Lead participant No[2]	Person-months[3]	Start month[4]	End month[5]	Deliv-erable No[6]
1	Identification of Potentially Informative Population Biobanks in Europe	2	2-8	0	36	1-5
2	Opportunities for Future Biobanking in Europe	11	3-22	0	36	1,2,6-8
3	Databases and Biobank Information Managment Systems	10	3-11	0	36	1,2,9-11
4	Strategies for Genotyping in Large Scale Biobanks	4	2-12	0	36	1,2,12-15
5	Harmonizing the assessment of phenotypes and life-style exposures	6	3-9	0	36	1,2,16-18
6	Ethical, Legal and social Issues in Population Biobanks and in Data Sharing	1	4-8	0	36	1,2,19-21
	TOTAL		17-70			

Deliverables list (full duration of project)

[1] Workpackage number: WP 1 – WP n.
[2] Number of the contractor leading the work in this workpackage.
[3] The total number of person-months allocated to each workpackage.
[4] Relative start date for the work in the specific workpackages, month 0 marking the start of the project, and all other start dates being relative to this start date.
[5] Relative end date, month 0 marking the start of the project, and all ends dates being relative to this start date.
[6] Deliverable number: Number for the deliverable(s)/result(s) mentioned in the workpackage: D1 - Dn.

Deliverable No[7]	Deliverable title	Delivery date [8]	Nature [9]	Dissemination level [10]
1	Initial Conference on Population Biobanks	6	O	RE
2	Concluding Conference on Population Biobanks for Health	35	O	RE
3	Report on study	12	R	PU
4	Web-based content management system	18	P	PU
5	Final report	33	R	PU
6	Report on biobanks and genetic isolates	12	R	PU
7	Final report	33	R	PU
8	World wide web report posting	35	R	PU
9	Interim report on information storage strategies	18	R	PU
10	Final report with standards/policies	33	R	PU
11	World wide web report posting	36	R	PU
12	Genotyping Cybercenter	33	P	PU
13	Annual report	12,24, 36	R	PU
14	SOP`s for genotyping issues	33	R	PU

(Deliverables list, use Deliverables list form below)

Deliverables list (full duration of project)

[7] Deliverable numbers in order of delivery dates: D1 – Dn.
[8] Month in which the deliverables will be available. Month 0 marking the start of the project, and all delivery dates being relative to this start date.
[9] Please indicate the nature of the deliverable using one of the following codes:
R = Report
P = Prototype
D = Demonstrator
O = Other
[10] Please indicate the dissemination level using one of the following codes:
PU = Public
PP = Restricted to other programme participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)

Deliverable No[11]	Deliverable title	Delivery date [12]	Nature [13]	Dissemination level [14]
15	Statistical network	36	O	PU
16	Preliminary report	12	R	PU
17	Final report	33	R	PU
18	World wide web report posting	35	R	PU
19	Interim report on ethical practises	18	R	PU
20	Final consensus report	33	R	PU
21	World wide web report posting	36	R	PU

[11] Deliverable numbers in order of delivery dates: D1 – Dn.
[12] Month in which the deliverables will be available. Month 0 marking the start of the project, and all delivery dates being relative to this start date.
[13] Please indicate the nature of the deliverable using one of the following codes:
R = Report
P = Prototype
D = Demonstrator
O = Other
[14] Please indicate the dissemination level using one of the following codes:
PU = Public
PP = Restricted to other programme participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)

Identification of and Information about Potentially Informative Population Biobanks in Europe

Workpackage number	WP1		**Start date or starting event:**	Start of CA
Participant id		2	3	
Person-months per participant		2	6	

Objectives

The aim of this workpackage is to identify and establish a network of completed, on-going or planned population-based biobanks or cohorts in Europe that can contribute usefully to co-ordinated investigations of the genetic and environmental determinants of complex diseases.

Description of work

A group of experts will meet at the *Initial Conference for Population Biobanks for Health.* They will define a list of criteria for the selection and categorization of population-based biobanks and population cohorts. These criteria will include the following: (i) Sample size; (ii) Availability of DNA; (iii) Availability of information about family structure and recruited relatives; (iv) Population characteristics; (v) Quality of phenotype information; (vi) Availability of genotypes; (vii) Availability of biomaterial.

Once appropriate criteria have been established, they will be published as a report on the worldwide web. Studies that might potentially be included in the initiative will then be identified and measured against the criteria. As the first biobanks meeting the stated selection criteria are identified and categorized, a content management system will be established in order to properly catalogue the studies electronically on the web.

In addition to meeting electronically to consider candidate studies for inclusion in the growing database, the expert group will formally reconvene at the *Concluding Conference for Population Biobanks for Health*. A final report summarising the current state of the database will be prepared prior to the conference and will be presented and discussed. The conference will provide an opportunity to discuss future strategy regarding population biobanks in the light of the state of the database and the main conclusions reflected in the final reports of other work packages.

Deliverables

(F) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(G) Report on study selection criteria. *Time: 12 months*

(H) Web-based content management system goes live. *Time: 18 months*

(I) Final report on status of studies in the database. *Time: 33 months*

(J) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

Milestones[15] and expected result

(B) The sole ”decision point” will occur when the selection criteria are selected as this will determine the studies that then become eligible to join the initiative. *Time: 12 months*

Opportunities for Future Biobanking in Europe

[15] Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Workpackage number	WP2	**Start date or starting event:**		Start of CA
Participant id	11	12	4	
Person-months per participant	16	3	3	

Objectives

The main objectives of this WP are: a) to identify new opportunities of biobanking in Europe with a focus on genetically isolated populations, b) to explore criteria for the selection and collection of these populations, c) to stimulate the development of standard analytic and other bioinformatic tools specifically designed for genetically isolated populations, d) to explore ethical/legal issues of relevance to isolates.

Description of work

An important aim of this work package is to identify completed, ongoing or planned European biobanks initiatives which have a particular focus on studies of isolated populations. Currently, there are known initiatives in Italy, Croatia, Sweden, Iceland, Finland and Scotland. Additional ones are planned in Romania, the Netherlands, and Spain. We will also set up a multi-disciplinary expert group to address a number of issues that offer particular problems for research based upon genetic isolates: (a) the selection of populations for study; (b) statistical analysis, analytic tools and database construction; and (c) ethical-legal issues. At the *Initial Conference for Population Biobanks for Health*, we will bring together a small group of relevant experts and, as one outcome of the conference, will determine other experts who should be asked to join the initiative. On the basis of a series of electronic and informal meetings and working with workpackages 1,3,5 and 6, we will produce two main reports. The first will detail current European population-based biobanks in genetic isolates. A final report will then explore issues (a)-(c) listed above. The final report will be presented and discussed at the *Concluding Conference on Biobanks*, and following revision, the final report will then be posted on the CA web site.

Deliverables

(K) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(L) Report on European population-based biobanks in genetic isolates *Time: 12 months*

(M) Final report to address issues specific to the study of genetic isolates pertaining to: the selection of populations for study; statistical analysis, analytic tools and database construction; and ethical-legal issues. *Time: 33 months*

(N) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

(O) Final report to be posted on worldwide web. *Time: 36 months*

Milestones[16] and expected result

The sole ”decision point” will occur at the *Initial Conference on Population Biobanks for Health* when we identify the members of the expert group. *Time: 6 months*

[16] Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Databases and Biobank Information Management Systems

Workpackage number	WP3	**Start date or starting event:**		Start of CA
Participant id	10	1	13	
Person-months per participant	5	5	1	

Objectives

The aim of this workpackage is to aid research into complex diseases by arriving at a common strategy that will provide an efficient network of databases for population biobanks in Europe and Canada. The challenge is to easily combine the cohorts so that very large sample sizes are generated, providing relevant, high quality and correctly documented sets of data to researchers.

Description of work

The first task in this work package will be to review existing and planned systems, and to arrive at a consensus on the requirements for a general information management system for biobanks. A second task will be to explore systems for maintaining unique and secure identities for specimens, subjects and biobanks, as well as for keeping track of the handling of permissions for use, analytical results and statistical output. Meta-information on quality of specimens and phenotypes will be integrated. The third task will be to explore a complete strategy for communication between population cohorts including a common nomenclature, compatible software techniques and appropriate information transmission policies. This all relates to information on specimens, laboratory results, phenotypes, exposures and genealogical data.

Deliverables

(P) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(Q) Interim report representing a complete inventory of large European biobank projects and their information and storage strategies. *Time: 18 months*

(R) Final report which will also address format and variable standards, communications standards and transmission policies. *Time: 33 months*

(S) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

(T) Final report to be posted on worldwide web. *Time: 36 months*

Milestones[17] and expected result

(C) The sole ”decision point” will occur at the *Initial Conference on Population Biobanks for Health* when we will determine the full composition of the expert group. *Time: 6 months*

Strategies for Genottyping in Large Scale Biobanks

[17] Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Workpackage number	4	**Start date or starting event:**	Start of CA
Participant id	4	5	
Person-months per participant	9	3	

Objectives

This WP aims to create an operational infrastructure for the evaluation of ongoing large-scale genotyping efforts in population cohorts and to provide a forum for expert opinions regarding genotyping methods and quality assessment of these methods. We aim to collect information from pre-existing genotyping projects involving large population cohorts in Europe, and share it on the consortium web site. Based on the collective experience of consortium members, we will explore issues pertaining to the selection both of technologies and of sampling criteria for study design in future European projects aiming to use large cohorts in genetic studies of complex traits and diseases. We will establish a network of statisticians to address issues of statistical power, study design and sampling protocol. This work will provide international researchers with advice and standards that will promote harmonization in regards to marker selection, genotyping quality control/cost and genotype database structures.

Description of work

1) Exploration, integration and posting on-line (www) of existing information about strategies for genotyping in large cohorts (derived both from publically accessible databases/information sources and in the possession of individual experts in our research consortium); 2) Exploration, summary and posting of issues pertaining to standardization, quality control and cost of genotyping and related data storage and handling systems to assess pre-existing technologies and facilities; and 3) pooling of the collective expertise of statisticians to provide guidelines for genotyping strategy in large population cohorts.

We will organize bi-monthly meetings to exchange ideas, monitor progress, and prioritize specific issues involving high throughput genotyping and study design. The information generated by each group will be consolidated at workshops attended by leaders of large genotyping, phenotyping, and information technology centres. We will facilitate interactions with industry, and benefit from their expertise in regards to existing and future technologies

Deliverables

These deliverables dovetail very naturally with the aims and deliverables of complementary workpackages in the GenomEUtwin project. See Leena Peltonen, form A2

(G) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(H) ”Genotyping cybercenter” providing on-line information about:
   a) Major genotyping efforts ongoing in large European population biobanks and cohorts
   b) Genotying quality reports – which will provide a comparison of genotyping accuracy among high throughput technologies and laboratories.
   c) Genotyping cost reports. We will monitor costs offered to external clients by genotyping core facilities in Europe and worldwide. *Time: on-line 12 months, fully operational 33 months*

(I) Annual report entitled ”Statistical issues in population genomics”. The report will contain a review of the literature, power tables, state-of-the-art commentaries on study design, etc. *Time: 12, 24, 36 months*

(J) Standard Operation Procedures (SOPs) for communicating genotyping data, as well as collection and storage. *Time: 33 months*

(K) Increased European cohesion and consensus on critical issues concerning the design and performance of research projects using large population cohorts. This will include involvement of junior investigators in the meetings of the experts of this WP. *Time: throughout CA*

(L) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

Milestones[18] and expected result

The sole ”decision point” will occur at the *Initial Biobanks Conference* when we will determine which phenotypes and exposures will be viewed as priorities, and which experts will sit on each subprogramme and on the integrating committee. *Time: 6 months*

Harmonizing the assessment of phenotypes and life-style exposures

Workpackage number	5	**Start date or starting event:**	Start of CA

[18] Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Participant id	6	7	8
Person-months per participant	3	3	3

Objectives

This workpackage is the next logical step in a strategy with **long-term aims** to produce: (i) a "minimum data set" representing a common core phenotyping protocol that could be applied to future European biobanking initiatives; (ii) agreed principles pertaining to the assessment of important complex traits that go beyond the limitations imposed by the minimum data set itself; (iii) agreed principles pertaining to the comprehensive assessment of background data on environmental and life-style exposures. The workpackage will bring together European and international experts and, based upon preliminary interactions via electronic media and at the *Initial Conference on Population Biobanks for Health*, its preliminary **primary aims** will be:

(4) To refine the strategy that should be adopted in taking the next steps towards the stated long-term scientific aims (i-iii).

(5) To select the complex phenotypes and exposure classes that should be addressed with the highest priority. This prioritorisation will define a series of research subprogrammes each of which will focus upon a system-specific group of phenotypes (eg musculo-skeletal, metabolic etc) or a coherent class of exposures. We will also set up an "integrating committee" to overview and synthesise the work within the individual research subprogrammes.

(6) To identify appropriate subgroups of experts to work on each subprogramme and to sit on the integrating committee.

Based upon a combination of face-to-face meetings, interactions via electronic media and at the *Concluding Biobanks Conference*, the research subprogrammes and the integrating committee will work towards the additional **primary aims**:

*(7) To undertake a complete workup of one major complex phenotype, that is **major depression.** This will act as a good exemplar of what must ultimately be done for all major phenotypes and exposures.*

(8) To explore the key issues (strengths, weaknesses, opportunities and threats) in relation to other common complex traits and life-style exposures that have been identified as priorities, and in relation to important rare traits in isolated populations.

(9) *To produce and discuss a final report that summarises the work undertaken across the whole workpackage (research subprogrammes and integrating committee) including final reports from the individual subprogrammes as appendices.*

Description of work

At the *Initial Conference on Population Biobanks for Health*, we will bring together experts who can make a substantive contribution to debate and discussion regarding **primary aims (1-3)**. After the conference, the designated integrating committee will produce a preliminary report summarising the discussions, conclusions and recommendations.

The integrating committee and each individual research subprogramme will be headed by an expert chairperson and deputy. Each subprogramme will work towards **primary aims (4-5)** using literature review, electronic interaction, and other meetings.

In working towards **primary aim (6)**, each subprogramme will provide updates on progress and a comprehensive final written report. The integrating committee will prepare a comprehensive overall final report. This will be presented and discussed at the *Concluding Biobanks Conference*. Following appropriate revision, it will then be posted on the worldwide web.

Deliverables

(U) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(V) Report following initial conference regarding **primary aims (1-3)**. *Time: 12 months*

(W) Final report (**primary aim 6**) addressing specific phenotyping recommendations for major depression and all key issues pertaining to other phenotypes and exposures **(primary aims 4,5)**. *Time: 33 months*

(X) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

(Y) Final report to be posted on worldwide web. *Time: 36 months*

Milestones[19] and expected result

The sole "decision point" will occur at the *Initial Biobanks Conference* when we will determine which phenotypes and exposures will be viewed as priorities, and which experts will sit on each subprogramme and on the integrating committee. *Time: 6 months*

(Description of <u>each</u> workpackage, use Workpackage description form below, one per workpackage)

Ethical, Legal and Social Issues in Population Biobanks and in Data Sharing

[19] Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Workpackage number	6	**Start date or starting event:**			Start of CA
Participant id	1	8	9	14	
Person-months per participant	2	2	2	2	

Objectives This workpackage concerns ethical, legal and social issues in population biobanks and data sharing. The objective is to increase the scientific and commercial value of pre-existing and planned population-based genetic databases and biobanks through coordination and harmonisation of the ethical-legal and governance issues. We will establish a platform of ethical-legal and governance criteria that are consistent with international norms and will enable data and sample sharing for research purposes within a defined ethical framework for Europe. In the long term this activity will help optimize public health through biobank research.

Description of work The work will be accomplished in two phases. Phase I consists of mapping extant database activities related to the major ethical questions raised by these projects. This includes examining questions of transparency, consent, recruitment, controlled access by 3rd parties, data sharing, role of oversight bodies, identifiability and privacy protection, commercialization, participant rights to withdraw, destruction of data, data confidentiality, informing participants of results. Major emphasis will be placed on guiding the elaboration of collaborative biobanking projects such that they conform to international laws and norms derived from multiple sources including international: human rights law, ethical norms, professional norms, consumer positions and Canadian norms. Phase II is a consensus building process in which information from phase I will be used to develop a platform of ethical-legal and governance criteria that are consistent with international norms and will enable sharing data and samples for the research purposes within a defined ethical framework for Europe. The activities of this WP will be highly coordinated with related activities in the ongoing projects P3G, COGENE and GenomEUtwin. Most of the work will be carried out through meetings, electronic communications, teleconferences and work by research assistants.

Deliverables

(Z) Preliminary planning, strategy setting and identification of full expert groups for all workpackages at *Initial Conference on Population Biobanks for Health. Time: 6 months*

(AA) Interim report on ethical practices in Europe, Canada and other countries *Time: 18 months*

(BB) Final consensus report exploring the possibility of an internationally consistent socio-ethical and governance platform in the light not only of the work undertaken in Phase I, but also following consultation with international, regional and national partners. *Time: 33 months*

(CC) Debriefing, presentation and discussion of final reports, and discussion of future strategy for all workpackages at the *Concluding Conference on Population Biobanks for Health. Time: 35 months*

(DD) Final report to be posted on worldwide web. *Time: 36 months*

Milestones[20] and expected result

(D) The sole "decision point" will occur at the *Initial Conference on Population Biobanks for Health* when we will determine which additional experts should be invited to join the expert group. *Time: 6 months*

B.7 Ethical, safety and other EC-policy related issues

Ethical, legal, social and safety issues

[20] Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Ethical, legal, social and safety issues are central to this CA and an entire work package (WP6) is dedicated to exploring this theme. That work package is entitled
'Ethical, legal and social issuses in population biobanks and in data sharing'. **There is no human subject data analysed in this CA. Rather, the activities concern assembling, reviewing, synthesizing and systematizing information. Therefore, this project does not raise ethical issues in the typical sense that there could be harm or mistreatment of personal rights, violations of confidentiality or privacy, or problems of informed consent.**
As described in the proposal, the purpose of the ethics-related work package is to undertake coordination activities that will lead to harmonization of the ethical-legal and governance issues. The outcome will be to establish a platform of ethical-legal and governance criteria that are consistent with international norms and will enable data and sample sharing for research purposes within a defined ethical European framework.

Other EC-policy related issues

Many of the ethics-related topics that will be studied under this work package are congruent with challenges and goals enumerated in EC policies. For example, the report 'Life sciences and biotechnology – A strategy for Europe' (COM, 2002, 27), stresses the importance of a unified, European strategy for the types of ethical governance issues and we strive towards this within the aims of this CA. Co-development of the scientific, ethical and legal issues is highly valued and reflected by CA plan that integrates key components into one harmonization initiative.

B.7.1 Ethical aspects

a) Specify if your project involves:

Does your proposed research involve:	YES	NO
• Human beings		X
Persons not able to give consent		
Children		
Adult healthy volunteers		
• Human biological samples		X
Human embryonic stem cells in culture		
Human foetal tissue/human foetuses		
• Personal data or genetic information		X
• Animals (any species)		X
Transgenic animals		
Non- human primates Dogs, pigs, cats,		

b) Confirm that the proposed research does not involve:

- research activity aiming at human cloning for reproductive purposes,
- research activity intended to modify the genetic heritage of human beings which could make such changes heritable[21],
- research activities intended to create human embryos solely for the purpose of research or for the purpose of stem cell procurement, including by means of somatic cell nuclear transfer,
- research involving the use of human embryos or embryonic stem cells with the exception of banked or isolated human embryonic stem cells in culture[22].

The proposed research does not involve any of the issues listed in point B.7.1.b)	CONFIRM No Involvement
	X

[21] Research relating to cancer treatment of the gonads can be financed.

[22] Proposers should note that the Council and the Commission have agreed that detailed implementing provisions concerning research activities involving the use of human embryos and human embryonic stem cells which may be funded under the 6th Framework Programme shall be established by 31 December 2003. The Commission has stated that, during that period and pending establishment of the detailed implementing provisions, it will not propose to fund such research, with the exception of the study of banked or isolated human embryonic stem cells in culture.

B.8 Gender issues

(for further explanation see Annex 4 (General approach across the programme) and Box 2 (Specific approach for the LifeSciHealth Priority) in the Guide for Proposers)

B.8.1 Participation of women

Answer the following questions:

- Are there women directly involved:
  - in the scientific management of the project? Yes Yes No
  - in the scientific partnership as scientific team leader in the project?Yes Yes No
- % of women scientists involved in the project[23]:
  - ⇒ Early researchers (less than 4 years after graduate)? ..……1..%
  - ⇒ Experienced researchers (minimum 4 years after graduate ……30…% or having a PhD)?
- Comment and justify if necessary
- Do you plan specific measure(s) regarding women role/participation in your project? Which? How? When?

[23] Definitions according to the FP6 mobility & Marie Curie activities.

B.8.2 Gender aspects in research

(If there are gender issues associated with the subject of the proposal, show how they have been adequately taken into account.)

(Recommended length - one page)

Answer the following questions:

	Yes	No
• Does the project involve human subjects?		
• Does the project use human cells / tissues / other specimens?		
• If human subjects are not involved or human materials not used, does the research involve animal subjects or animal tissues / cells / other specimens *(as models of human biology/physiology)* in such a way that it is expected that may have implications for humans?		
• Does the project use collection of data related to human subjects, human materials, animal subjects or animal materials		

A positive answer to any of these questions implies that gender/sex aspect should be taken into consideration in the research proposal.

	Yes	No
Are gender/sex differences with respect to the research documented in the literature?		

If yes please give details.

A negative answer to this question may imply some innovation in the proposal towards this issue that will be taken into account in the evaluation process.

If there are gender/sex aspects in your project:

- Detail the questions addressed in their proposal related to gender/sex aspects in research.
- Comment on the expected outcome.
- Describe how the gender/sex aspects will be taken into account in the research, methodology and interpretation of their results.

If you do not consider gender/sex differences, provide justification.

- *The evaluation panel will assess the relevance of the justifications provided.*
- *Neither additional costs, nor difficulties in obtaining female cells, female tissues, female specimens, or recruiting female subjects, would not normally be considered as a valid reason for excluding gender/sex aspects ("female" includes both animal and human subjects).*