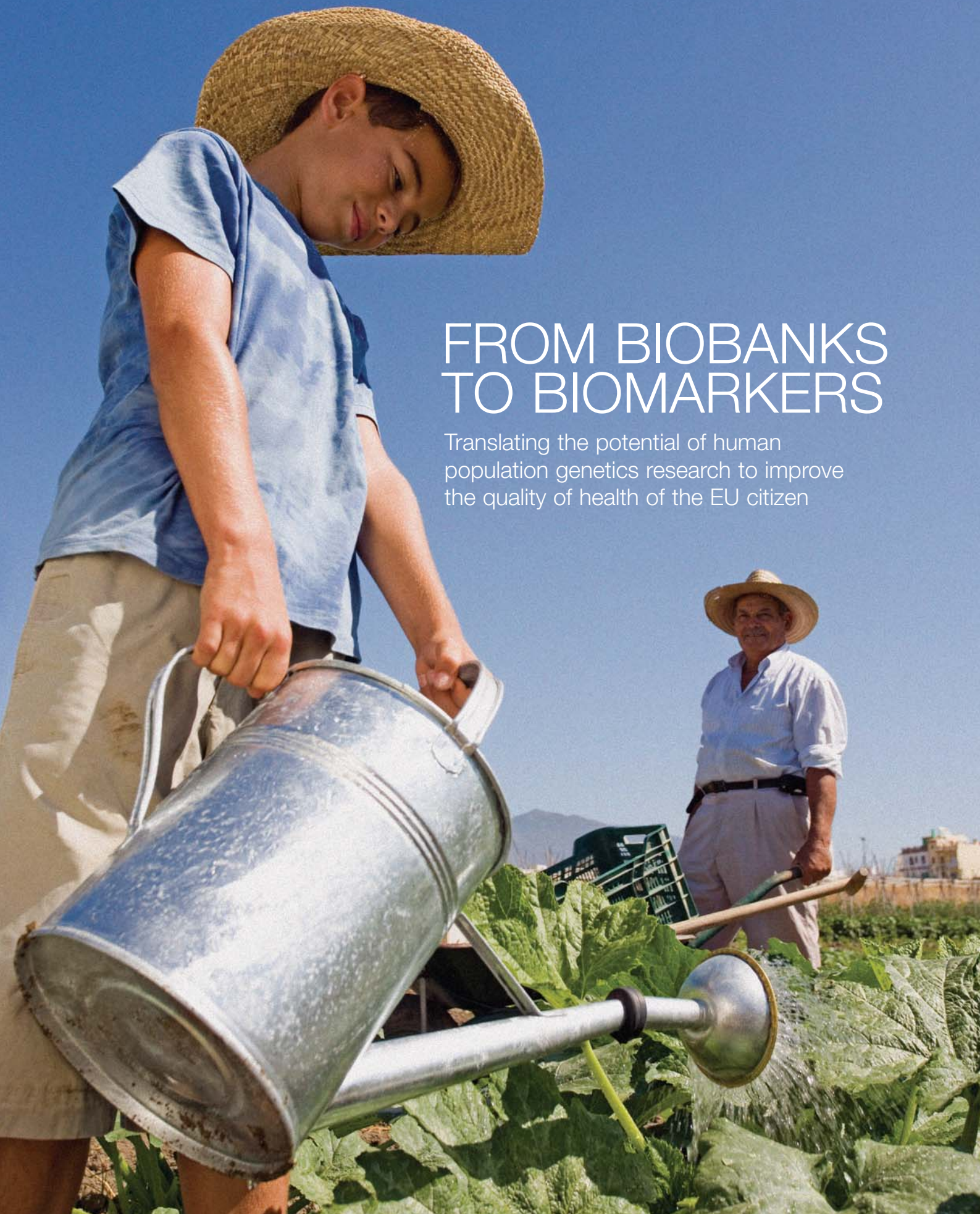# FROM BIOBANKS TO BIOMARKERS

Translating the potential of human population genetics research to improve the quality of health of the EU citizen

# FROM BIOBANKS
# TO BIOMARKERS

Translating the potential of human
population genetics research to improve
the quality of health of the EU citizen

Proceedings of a conference held at the
Wellcome Trust Conference Centre, Hinxton, Cambridge
20–22 September 2005

Jointly funded by the Health Research Directorate,
DG Research, European Commission, and the Wellcome Trust.

## Contacts

Dr S S Baig
Principal Scientist, Health Research
Research Directorate-General
European Commission, CDMA – 02/159
B-1049, Brussels
Belgium

**T** +32 2 296 34 37
**F** +32 2 296 05 88
**E** shahid.baig@cec.eu.int

Dr Alan Doyle
Science Programme Manager
Wellcome Trust
Gibbs Building
215 Euston Road
London NW1 2BE, UK

**T** +44(0)20 7611 8248
**F** +44(0)20 7611 7277
**E** a.doyle@wellcome.ac.uk

# Contents

# Foreword

At a time of rapid advances in the field of population genetics and plans for biobank ventures in many countries, the recent European Commission–Wellcome Trust conference on Population Genomics proved extremely timely. I am sure that all those who participated in the meeting will agree that it was a success. It brought together many of the key stakeholders to learn from experiences of current studies, establish cross-border networks and collaborations, to focus on opportunities and challenges for the future and how these might be achieved through the EU 7th Framework Programme (FP7). Europe has a long and successful history of research in this area; however large-scale population studies of this scale are of international importance. Partnership and collaboration across and beyond the borders of Europe will be key to ensuring the maximum utilisation of research outputs to ensure the most benefit to human health.

I hope you will find this report useful.

**Mark Walport**
Director, The Wellcome Trust

In the last decade, Member States across the EU have been investing in research on population genetics. This is due to the important knowledge gained through research in understanding the mechanisms which could contribute to specific diseases in humans. This is also due to the significant investments made in sequencing the human genome and the need now to harvest the results. However, the scope of these investments is limited by the resources available and the size of the population being served. Significant added value can be obtained if the objectives and protocols involved in human population genetics research at a national level can be harmonised to become representative of the entire EU population. The EU would thereby develop and maintain a leading global position in medicine, genetic epidemiology and population genetics. Advances in genetic medicine would be harnessed to improve the delivery of clinical medicine, public health and preventative healthcare for the EU citizen.

Common diseases of major public health importance are phenotypically complex with many having a heritable component. Population genetics can be used to characterise and stratify these complex diseases and associated pathophysiological states. Access to databases containing genotypic, clinical, environmental and lifestyle information on individuals, along with corresponding clinical specimens (biobanks), are an essential component for population genetics research.

The aim of this conference, as a first step in meeting the needs, was to mobilise European scientists and promote collaboration between population biobanks and other longitudinal cohort studies across Europe and elsewhere. Integrating the information held by these various organisations will be complex, as many scientific, technical, social, legal and ethical barriers need to be identified and addressed. A phased approach, therefore, will be needed to accomplish this and the current Framework Programme represents a first crucial step in the overall process.

We expect that the new Framework Programme (FP7) 2007–2013 will contribute to better integrating and strengthening the European Research Area in this fast developing field, by building upon the research initiatives taken during FP6. It will stimulate and maintain multidisciplinary research and will facilitate the translation of this research to build real, consistent and co-ordinated progress in medicine at a European level. It builds on existing resources for population genetics research across Europe, thus facilitating greater access to additional knowledge and expertise.

**S S Baig**
Principal Scientist, Health Research, European Commission

# Executive summary

EUHEALTHGEN is an initiative that aims to promote the translation of the outputs from research on population genetics into direct health benefits for European citizens. It is jointly funded by the Health Research Directorate, DG Research, European Commission, and the Wellcome Trust.

An international conference entitled 'From Biobanks to Biomarkers – Translating the potential of human population genetics research to improve the quality of health of the EU citizen' was held to promote the aims of this initiative. This document provides an account of the proceedings of this conference.

The conference programme was a mix of plenary lectures that reviewed the status of population genetics research and associated discussion sessions, designed to promote widespread debate on three topics identified as being of high priority. The objective was to engage as many of the conference participants as possible in an attempt to reach a consensus on these topics.

The first topic discussed concerned the expectations and goals for industry and healthcare. The overall conclusions were that Europe needs to invest in both the resources necessary to support population genetics research and the training and human capital to create research capacity in this important area. FP7 provides an ideal platform for making the required investments.

Other issues addressed during this discussion session included:

- The rationalisation of scientific, public and commercial interests in population genetics across Europe and the need for data collection to be harmonised, standardised and quality controlled to optimise the translation of research outputs into tangible health benefits.

- Barriers to the adoption of population genetics research, particularly those involving ethical considerations and issues of confidentiality and security of personal data.

- A common vocabulary for population genetics terms across Europe so there can be better dissemination of data and knowledge.

- Greater clarity on patenting policies for genetic discoveries across Europe to facilitate commercial development and translation into healthcare.

- The evidence base that needs to be developed before genetics-based health services can be delivered or accepted.

- The partnerships between industry, public bodies and the academic/clinical sectors needed to promote capacity building and the training of future researchers in population genetics.

- The size of epidemiological studies undertaken across Europe to ensure that these are sufficiently powered to give results that can be replicated and validated.

The second topic discussed concerned the future needs for resources and tools for statistical analysis. The conclusions made here reflected those made in the previous discussion and emphasised the need to increase training opportunities, particularly in statistical genetics so that capacity in this area across Europe could be improved. It was also concluded that improvements were needed in the legal and ethical frameworks for biobanks across Europe and that there should be better standardisation and quality control of data capture.

## Improvements were needed in the legal and ethical frameworks for biobanks across Europe.

Other issues raised during this discussion session included:

- The requirement for general access to biobanks and large clinical and sample databases to facilitate genetic association studies for common disease-related phenotypes – Europe contains some unique populations for these studies.

- The selection of appropriate analytical strategies and statistical methods, some of which may require further research and development in their own right.

- The lack of a comprehensive inventory of relevant biobanks, disease and other registries across Europe so that the resource available for population epidemiology is known.

- The lack of a similar catalogue of guidelines and regulations on the ethics, confidentiality and security requirements for sharing samples and clinical information on individuals across Europe.

- The infrastructural funding needed to support the collection of longitudinal clinical, life-style and disease end-point information databases across creating data warehouses for subsequent research use. This represents a long-term strategic commitment which national governments alone would find difficult to justify but which the European Commission, through FP7, could co-ordinate.

- The cultural changes that will be needed to achieve the required satisfactory collaborations between geneticists and epidemiologists and biologists and biostatisticians and the consequent need for the development of interoperable IT systems.

The final topic for discussion was about the research directions for the development of novel technologies and biomarkers, where the main conclusion was that better networking of existing and future biobanks across the European Research Area should be a high priority for the European Commission and FP7. This would facilitate the development of suitable EU standards and guidelines that would help to integrate population genetic analyses into clinical trials and new drug and diagnostic test approval procedures.

Other issues discussed during the following debate included:

- Access to large-scale genomic technologies within well-established high-level facilities across the European Research Area and the continuing funding requirement to support the development of novel technologies within them.

- The development of lower cost and higher sensitivity assays as well as low unit cost and highly portable devices to optimise the dissemination of appropriate technological capacity in a clinical settings. This should include the development of case-based molecular diagnostics, including pharmaco-diagnostic testing for advancing personalised medicine.

- The development of technologies targeting not only DNA but also RNA, protein and metabolite genomics, to integrate environmental, epigenetic and developmental variations.

- The special attention that needs to be given to the 'intermediate phenotypes' detected during disease progression.

- The small genetic effects that may influence the progression of complex diseases and how this knowledge might be used in the clinic because of the likely difficulties in interpretation. There is a gap between capacity, knowledge and application with little yet known on how genetic knowledge will exactly impact on diagnosis and treatment.

- The absolute need for method standardisation, validation and harmonisation with corresponding good clinical practice guidelines covering all aspects of informed consent.

The issues discussed at the conference and the conclusions made clearly demonstrated that the evidence base associated with human population genetics research needs to be further developed and consolidated as new and improved technologies are introduced. This cannot remain the responsibility of individual Member States and so greater emphasis needs to be placed on the creation of transnational and multi-disciplinary research collaborations between academia, health service providers, industry and governments in order to make the best use of the available resources.

Support for population genetics through the current proposed FP7 is therefore vital, particularly if the effective translation of research outputs into clinical applications was to be achieved. This support would provide the supporting infrastructure for research on fundamental genomics (to develop new techniques and technologies), translational research (to validate new protocols in a clinical environment and to make them economically viable, cost effective and readily available) and health services research (to develop methods for health promotion and prevention and so facilitate the introduction of sustainable and efficient healthcare systems).

Support for population genetics through the current proposed FP7 is therefore vital, particularly if the effective translation of research outputs into clinical applications was to be achieved.

The next steps being proposed for EUHEALTHGEN involve the organisation of further meetings between population geneticists to consider the use of the samples and information held by the various national biobanks with the specific objective of preparing high quality and relevant proposals across the major activity area discussed during this conference ready for submission following published calls for proposals during FP7. The major success criterion for EUHEALTHGEN will be the number of successful biobank-related proposals awarded in the acknowledged priority areas of mental health, metabolic disease, inflammatory disorders, ageing and child health and the subsequent adoption of the outputs of population genetics research by health service providers across Europe.

# 1 Background

**1.1** EUHEALTHGEN is a Specific Support Action jointly funded by the European Commission and the Wellcome Trust. Its main aim is to facilitate the development of a forward-looking strategy for translating the outputs of population genetics research into clinically useful and health enhancing initiatives, whilst improving EU industrial competitiveness in this area. It anticipates the creation of research collaborations that are able to respond creatively to the objectives and content of the genomics health theme in the coming 7th EU Research Framework Programme. By providing networking opportunities, it will also help to strengthen and contribute to the development of the European Research Area.

*An international conference that reviewed existing research in human population genetics was used to deliver the main EUHEALTHGEN objectives.*

**1.2** A consortium of senior researchers actively engaged in human population genetics research formed a Scientific Advisory Board that developed the programme for EUHEALTHGEN. The composition of the Scientific Advisory Board is given in Annex 1. An international conference that reviewed existing research in human population genetics was used to deliver the main EUHEALTHGEN objectives. This was held at the Wellcome Trust Conference Centre, Cambridge, UK on 20–22 September 2005. Some 200 scientists from over thirty countries, representing both academia and industry, attended. The list of participants is given in Annex 2.

**1.3** The conference consisted of keynote addresses and presentations on topics associated with population genetics research. These were supported by poster demonstrations and discussion sessions focusing on the themes of 'expectations and goals for industry and healthcare', 'future needs for resources and tools for statistical analysis' and 'research directions for the development of novel technologies and biomarkers'. The full meeting programme is given at Annex 3 and this report provides a summary of the conference proceedings.

## Abbreviations

| | |
|---|---|
| **CMSB** | Centre for Medical Systems Biology |
| **DNA** | Deoxyribonucleic Acid |
| **EGFR** | Epidermal Growth Factor Receptor |
| **EMEA** | European Agency for the Evaluation of Medicinal Products |
| **EPIC** | European Prospective Investigation into Cancer and Nutrition |
| **ERA** | European Research Area |
| **FDA** | Food and Drug Administration (USA) |
| **FP7** | The European Union's 7th R&D Framework Programme |
| **GRIP** | Genetic Research in Isolated Populations |
| **HLA** | Human Leucocyte Antigen (histocompatibility antigens) |
| **HSR** | Hypersensitivity Reaction |
| **IT** | Information Technology |
| **LD** | Linkage Disequilibrium |
| **NGI** | Netherlands Genomic Initiative |
| **NICE** | National Institute for Clinical Excellence |
| **P3G** | Public Population Project in Genomics |
| **RNA** | Ribonucleic Acid |
| **SME** | Small to Medium Sized (commercial) Enterprise |
| **SNP** | Single Nucleotide Polymorphism |
| **WTCCC** | Wellcome Trust Case Control Consortium |

# 2 From biobanks to biomarkers – the context

**2.1** Research groups across Europe have made significant contributions to progress made in human population genetics during the past two decades. This has demonstrated the importance and the significant added value of developing collaborative partnerships and consortia to address major research topics that cannot be accommodated by single countries alone because of either resource or research capacity constraints. A common challenge for all programmes on human population genetics is to capitalise on current research outputs and use these to improve public health, whilst enhancing industrial competitiveness and wealth creation.

**2.2** Common diseases of major public health importance are phenotypically complex with most having heritable, environmental and lifestyle components. Population genetics can be used to characterise and stratify these complex diseases and the associated pathophysiological states. However, this requires access to human genetic research databases that contain genotypic, clinical, environmental and lifestyle information on individuals along with corresponding clinical specimens. These databases have been variously referred to as population databases, cohort databases, gene banks and biobanks. For the purposes of this document the word biobank will be used as the generic term.

**2.3** Several biobanks already exist in, or are being planned by, a number of Member and Associated States and the Candidate Countries. These include Biobanks for Health in Norway, the UK Biobank, the Estonian Genome Project, GenomEUtwin, Generation Scotland, the Swedish national network of biobanks in functional genomics and LifeGene, deCode and the European Prospective Investigation into Cancer and Nutrition (EPIC). They are already extremely diverse with respect to the population included, the nature and size of the biological specimens held and the clinical and anthropomorphic data available. They are generally used for:

- Linkage studies to identify gene sequences associated with inherited diseases.

- Association studies to correlate a genetic change with a specific disease when there are no obvious patterns of inheritance.

- Investigating how genes, environmental factors and lifestyle act independently or in combination to influence susceptibility to disease and how, once a disease develops, it may progress.

- Investigating the diversity of genetic risk factors in different European populations.

- Pharmacogenetic analyses to determine the genetic basis of drug metabolism and induction of adverse drug reactions.

- Reclassifying disease phenotypes based on the genetic/biological background.

**2.4** At present, there is little collaboration between these biobanks, largely because of ethical, legal, practical and financial difficulties in sharing or exchanging material and/or information and because of the lack of standardised and quality controlled protocols for data collection, sample storage, analysis, and access etc. Furthermore, few of these databases provide for prospective studies of disease outcomes. These are considerable restrictions, particularly when investigating disease susceptibility where large cohort sizes are needed to give statistical accuracy to any of the correlations being made, for replication of original findings and for validation in standard healthcare settings. These restrictions also inhibit studying divergent as well as distinct isolated populations across the EU genetic cline.

## At present, there is little collaboration between these biobanks

**2.5** Significant advantage could therefore be gained if there was greater harmonisation of programmes, greater standardisation of protocols, greater harmonisation of regulatory standards and greater co-ordination of activities of biobanks across Europe and elsewhere. Not only would there be access to larger population groups but there would also be incentives for technological development or improvement and for industry involvement. There is therefore a need to develop a large multi-national programme to harness the potential of human population genetics research. It also provides the rationale for developing the EUHEALTHGEN project.

**2.6** EUHEALTHGEN has been established to:

- Promote communication between major biobank and longitudinal cohort initiatives across the ERA and restrict fragmentation.

- Serve as a powerhouse for the strategic planning of the large scale research and database infrastructure needed for major population genetic studies in Europe.

- Enhance effective sharing and dissemination of research information and materials.

- Promote common learning experiences and standardise methodologies.

- Promote population genetics research for the benefit of the EU citizen.

- Stimulate economic growth and enhance industrial competitiveness.

- Integrate ethical, legal and social issues and address barriers to collaboration and exchanges of research materials and data between studies.

- Involve accession and associated candidate countries.

- Promote positive communication of the benefits of educating the general public, healthcare professionals, policy makers and funders about human population genetics.

- Identify goals and prioritise approaches across the European network.

- Promote the uptake of best practice in human population genetics research and the adoption of a joined-up and fully integrated strategy from basic research, through clinical studies to the treatment of individual patients.

- Promote the anticipated paradigm shift in healthcare from disease diagnosis and treatment to the identification of personal disease risk and the development of appropriate personalised prevention strategies.

**2.7** EUHEALTHGEN builds on the pre-existing wisdom, knowledge and skills held within the consortium members, and it is anticipated that its outcomes will have a significant impact on the creation and working protocols for biobanks and related programmes at both national and international levels. It is anticipated that some degree of harmonisation of the standards adopted for data collection and management, genotyping and governance will be achieved.

**2.8** Developing a network of biobanks and longitudinal cohort studies across the ERA and elsewhere will create added value by pooling scarce resources and preventing their further fragmentation. Such a network has the potential to give researchers access to a greater cohort size and data set so that the effect of associations between genotype, environment and lifestyle at the individual and population level can be studied with greater statistical significance. Working at an integrated European level will enable a greater range of clinically important problems to be addressed and this, in turn, will ultimately lead to a faster development and uptake of targeted disease prevention strategies.

**2.9** The proposed outcomes of EUHEALTHGEN fit well with the objective of the health theme of FP7, which aims to improve the health of European citizens and increase the competitiveness of European health-related industries and businesses, while addressing global health issues including emerging epidemics. EUHEALTHGEN, like FP7, places great emphasis on the support of translational research, transferring new discoveries into clinical applications, developing and validating new therapies, methods for health promotion and prevention, diagnostic tools and technologies as well as promoting sustainable and efficient healthcare systems.

**2.10** EUHEALTHGEN recognises that clinical research on many diseases relies on international multi-centre trials and that epidemiological research requires access to a large diversity of populations to achieve significant results. A number of specific population cohorts are already available across Europe and these can be used to couple genotype with disease phenotype. However, better ways of making the data held within the associated databases available to all legitimate researchers are needed before they can be used for maximum effect. FP7 provides an ideal platform on which this activity can be promoted.

**2.11** FP7 will operate through three broad areas of work. These include:

- 'Biotechnology, generic tools and technologies for human health' that aims to develop and validate the necessary tools and technologies that will make possible the production of new knowledge and its translation into practical applications in the area of health and medicine.

- 'Translating research for human health' that aims to increase knowledge on processes and mechanisms involved in normal (healthy) functioning and specific disease situations, and to transpose this knowledge to clinical applications.

- 'Optimising the delivery of healthcare to European citizens' that aims to provide the necessary basis both for informed policy decisions on healthcare systems and for more effective strategies of prevention, diagnosis and therapy.

EUHEALTHGEN operates within, and will add value to, each of these areas of work.

# 3 From biobanks to biomarkers – the conference

**3.1** A conference was chosen as the main way of achieving the initial objectives of EUHEALTHGEN. This was because it could be used as a forum to engage all the major players in human population genetics across the Member and Associated States, the Candidate Countries and others in formal debate and informal discussion, particularly with respect to:

- High throughput research on human population genetics.

- Detection, diagnosis and monitoring of disease.

- Innovative therapeutic approaches and interventions for disease treatment.

- Predicting suitability, safety and efficacy of therapies at both population and individual levels.

**3.2** It was anticipated that the conference would facilitate the development of new research tools for modern biology that could significantly enhance data generation and improve data and specimens (biobanks) standardisation, acquisition and analysis. This, in turn, would aid future large-scale data gathering and the understanding of complex biological systems, thereby underpinning novel approaches to the detection, diagnosis and therapy of complex diseases. The conference would therefore facilitate the harmonisation of objectives and protocols in human population genetics research at a national level and make them more representative of the entire EU population.

**3.3** The conference programme was designed to review existing research on population genetics, analyse the associated strengths and weaknesses and take stock of the technologies needed to help build partnerships between existing and new national and international projects in the field. The stated performance indicator is the development of a forward-looking strategy for translating the outputs of human population genetics research into clinically useful and health enhancing initiatives, whilst improving EU industrial competitiveness.

*It was anticipated that the conference would facilitate the development of new research tools for modern biology that could significantly enhance data generation and improve data and specimens (biobanks) standardisation, acquisition and analysis.*

**3.4** The chosen format for the conference was a mix of plenary lectures on selected topics and associated discussion sessions involving both formal and informal presentations. Each discussion session was co-chaired by members of the EUHEALTHGEN consortium and rapporteurs were appointed to take a note of the proceedings.

**3.5** The discussions were designed to maximise participation from all of those attending the conference. There was an appropriate balance between a formal infrastructure and spontaneity so that contributions to the debate were given freely, honestly and without constraint. The objective was to identify the key issues associated with human population genetics research and so promote the cross-disciplinary dialogue necessary for the formation of networks that can respond effectively to future calls within the next European Research Framework Programme.

# 4 Overview and keynote presentations

## Session Chair: Dr Alan Doyle (Wellcome Trust)

**4.1** Dr Doyle and **Dr Bill Baig** (EU) welcomed everyone to the meeting.

**4.2** Dr Doyle began by outlining the scope of the meeting, which aimed to promote communication between the major biobanks and longitudinal cohort initiatives across Europe. It was hoped that the meeting would provide a discussion forum for the strategic planning of large-scale research and database infrastructure and help identify where the EU, national governments and the funding agencies could take best advantage of these studies.

**4.3** Dr Baig described the meeting as a first of its kind and went on to explain that the potential of developing this field would require a long-term strategy. Many biobanks were underway and it would be necessary to develop and share standard protocols and quality assurance measures. The European Union would provide the best basis for studying the susceptibility to disease, response to treatment, drug tolerance and other patient-related issues. In addition, standards and quality assurance could be best provided by collaborations within the EU. This was the start of visualising the 21st century mode of healthcare and Dr Baig asked the meeting to reflect on how major initiatives could be taken forward.

**4.4** **Dr Mark Walport** (Director, Wellcome Trust) emphasised the need for partnerships in this important and challenging area of research. When undertaking large-scale population studies, it is extremely important to fully engage the public. There is a substantial intersection between research of this nature and public policy on aspects of privacy, identity, data management and other issues that extend into many areas of daily life. The benefits of linking information on genetic variation to other information sources, such as diet, health, disease and environmental and social conditions would be enormous and enable research into the genetic pleiotropy of common disease.

**4.5** There are now many cohorts available for research purposes, including some very large studies that represent major investments. Key to the success of these studies is ensuring their accessibility and interest to other scientists, thereby allowing the use of these resources to be maximised. There are major challenges in linking large resources of phenotype and genotype data, but this will provide an important opportunity to develop new shared platforms.

It will be essential for scientists to be prepared to publicly discuss their projects, explain why they are important and what the realistic health benefits might be. These are global endeavours in which continued communication and partnerships are essential.

**4.6** **Dr Tom Hudson** (McGill University) discussed how the consideration of disease burden had moved rapidly from infectious to chronic diseases such as diabetes, cancer, cardiovascular disease, asthma and arthritis. The causes of chronic disease were multi-factorial with the 'web of causation' including genetics, environment, diet, lifestyle and social structure. Non-genetic risk factors were complex, numerous and varied over a lifetime. Because of this, it may be easier to pursue the genetic factors of disease, as these were more stable and inherited. The alleles discovered so far have been common alleles with modest effects on risk, for example the apoE4 allele associated with Alzheimer's disease.

**4.7** To identify further genetic risk factors, it would be necessary to obtain a comprehensive knowledge of genetic variation, develop efficient genotyping technologies and have available large cohorts that have been well phenotyped. International consortia are effective in developing these resources and to illustrate this, Dr Hudson presented data from the International HapMap Consortium, comparing the genetic sequences of different individuals in order to identify common, shared genetic variants.

**4.8** The HapMap project had discovered that some regions of the genome had very little diversity, such that 94% of the individuals studied had only one or two haplotypes. In contrast, other regions had very little correlation between Single Nucleotide Polymorphisms (SNPs) and it had been confirmed by others that these regions correspond to recombination hotspots. Some haplotypes identified in the HapMap project were themselves markers for disease and one such haplotype on chromosome 5, the IBD5 susceptibility haplotype, was shown to be a marker for Crohn's disease.

**4.9** When dissecting human disease using haplotype information, it may not be necessary to sample every individual SNP in order to identify functional associations. The example of the haplotype on chromosome 5 was true for the rest of the genome, and so understanding the architecture across the entire genome will be extremely valuable.

**4.10** Dr Hudson then discussed the lessons learned from the HapMap project, describing how the populations studied had not been consented for this scale of genotyping and unusually, how ethical and sociological experts were engaged before the project began. Each participating centre had used a different genotyping technology and the data produced was evaluated continuously to produce very low error rates. The resulting HapMap was therefore a robust tool for genetic studies and this was evidenced by the 20,000 'hits' per week on the HapMap database.

**4.11** The importance of collaboration and the benefits of size were exemplified by the Public Population Project in Genomics, or P3G, which was formed a little over a year ago. The P3G project aims to develop research tools to enable effective collaboration between biobanks and sharing of expertise and resources. Over the next three years, P3G will create a network of over 3 million participants for genetic epidemiological studies that will provide the statistical power to dissect both the genetic and gene-environment interactions in disease.

**4.12** The P3G consortium aims to harmonise methods in collecting and storing data and samples, and quality assurance, and to share an approach in governance. A key early goal was to invest in the development of a knowledge base that will track large-scale population genetic studies, including biobanks, and record their characteristics in order to benefit from existing knowledge and compare approaches. The knowledge base, called 'The Observatory', will contain a description of each study, a catalogue of questionnaires and key words for variables. This will allow queries for specific terms and allow users to identify specific features of each study. All of P3G's recommendations in terms of harmonised tools will be made available to the scientific community via The Observatory.

**4.13** In conclusion, Dr Hudson emphasised the need for large-scale studies to discover the complex genetic and gene-environment interactions associated with common disease and how international collaboration is required to realise these goals.

**4.14** **Professor Leena Peltonen** (University of Helsinki) presented her view on the emerging strengths in biomedical research in the post-genomic era.

**4.15** For the first time in human history, it is now possible to produce a high-resolution picture of our genes and the proteins they encode. The Human Genome Project led to rare successes in complex disease, but developments in understanding allelic variations across the entire human genome will ultimately make it possible to examine an individual's genetic variation and deduce the associated functional consequences. The interaction between genes and the environment is very important in the disease process, but it is very complex and difficult to study.

**4.16** The concept of addressing gene-environment interactions is not new. History has shown this to be a great strength within Europe. Our epidemiological research has long traditions. There is a European niche in biomedicine that flows from a reliable healthcare infrastructure, a well-educated population willing to participate in genetic studies and a high level of expertise in genetics, epidemiology, clinical medicine, mathematics and information technology. These factors combine to provide unique possibilities in healthcare and genomic research.

**4.17** In Europe, family studies, case-control studies and epidemiological studies have been underway for decades and it is important to recognise the complementary nature of the different approaches adopted. Once an allelic variant has been identified in exceptional families or case control studies, large epidemiological studies are required in order to show that the variant has significance at the population level and to determine the impact in different populations. A major challenge will be performing functional studies on DNA variants associated with common late-onset disease, as many of these will be represented by mutations in non-coding or regulatory regions, or by more complex rearrangements. We are still amazingly ignorant of the accurate structure of non-coding regions of the genome.

**4.18** As an example of this, Professor Peltonen described the basis of lactose intolerance in humans, which can be expressed early in newborns or as a late adult form. The early and late forms of lactose intolerance involve the same gene, but the variants differ. In newborns, a mutation in the coding region of the gene results in an inactive protein, whereas the late adult form is associated with a mutation that interrupts a regulatory element critical for controlling levels of expression of the gene in the gut. Genetic studies led to the observation that tolerance of milk actually represents a common global mutation that has conferred a selective advantage. This is a nice example how some 'common disease variants' are common since they have been favoured by selection.

**4.19** Finally, Professor Peltonen presented information on a number of collaborative biobank collections currently underway in Europe, including GenomEUtwin. Utilisation of already exisiting epidemiological cohorts and sample collections would be extremely cost-beneficial for European research. Maximising their impact will require the establishment of intellectual core facilities where expertise, technologies, databases and phenotypic information are shared. A major challenge within Europe will therefore be the integration of resources. She emphasised that Canada, Japan and the US must also be included in the exchange of information held within Europe.

# 5 Industry perspectives and translational research

**Session Chair**: Professor Nancy Pedersen (Karolinska Institute)

**5.1** **Dr Alun McCarthy** (GlaxoSmithKline) presented his perspective on why understanding populations and sampling is critical for the development of new medicines. He maintained that significant unmet needs of both patients and global populations still remained and that the safety of medicines must be addressed. The research and development expenditure of pharmaceutical companies continues to increase but, in general, the output of new chemical entities remains constant.

### Genomics and genetics impacts at every stage of the drug development process.

**5.2** When developing new medicines, candidate molecules are tested on increasing numbers of subjects at each phase of the clinical trial (from tens of subjects in phase 1 studies to several thousands in phase 3) and the risk of failure was exacerbated at each stage if an inappropriate population is selected for analysis. He emphasised that over 90% of new drugs fail during this period and that it was therefore essential to take into account sampling, ethnicity and the environment when selecting participants in a clinical trial. The importance of understanding disease heterogeneity was also highlighted, as it was possible to miss the potential positive effects of a new drug on a subset of patients. Dr McCarthy emphasised that genomics and genetics impacts at every stage of the drug development process and then went on to describe some specific examples.

**5.3** Genetic studies on the effect of the drug Tranilast showed that it was possible to use a large, standardised, ethnically matched population of cases and controls so that the power of the study could be increased. Increasing the number of random controls in the study also increased power, but the greatest impact on power was generated by geographically and ethnically matching the cases and controls.

**5.4** Studies on the action of the drug Ziagen, which is used to treat HIV/AIDS, have demonstrated the importance of including a range of ethnic groups in clinical trials. Ziagen produces a hypersensitivity reaction (HSR) in 4% to 5% of patients that is associated with a gene in the HLA region. When testing this marker as a possible predictor for hypersensitivity in subjects recruited from different ethnic groups, it was found that it was present in 46% of Caucasian hypersensitivity subjects tested but that it had no predictive power in a black population. He concluded that it is essential to ensure that sufficient numbers of relevant ethnic groups are included in clinical trials, as their under-representation could lead to the production of erroneous results.

**5.5** Dr McCarthy then described a method developed by GlaxoSmithKline to analyse populations using a large number of genetic markers. This statistical method provides a measure of the genetic similarity of an individual compared with a control group by analysing the average difference of variants across the entire genome. This methodology could be used to completely separate cases and controls into two distinct groups on the basis of genetic information alone. Some markers were influential in defining cases and controls, while others were important in defining the corresponding population heterogeneity. It was therefore possible to obtain information on what markers were important in disease along with information on population substructure.

**5.6** In an extreme example, it was possible to show genetic differences in controls that may relate to how they were ascertained for example, whether recruited from a local clinical centre or by advertising. Ascertainment was therefore vital in studies that would sample large numbers of variants across the genome and this would be especially important in very large genetic or epidemiological collections such as biobanks.

**5.7** Dr McCarthy ended by stating that academic and industry researchers should work in partnership to utilise population resources in recruiting cases and controls and in sharing the same datasets, subject of course to the consents available. Commercial input will be vital during the translation of the outputs of research on human population epidemiological into real health benefits.

**5.8** **Professor John Bell** (University of Oxford) presented an academic perspective on genomics, populations and clinical research with an emphasis on the use of new genetic technologies. He suggested that many in both Europe and the US now recognised that all the outputs generated from the basic science surrounding genetics and cell biology had not yet been translated into real benefits for healthcare and public health. However, this would change particularly as the interplay between genetics and the environment in disease was used to dissect pathogenic processes and define different disease sub-types at the population level. He emphasised that whilst the alignment of genetics and epidemiology as disciplines would facilitate the analysis of small effects on disease heterogeneity, this would require access to large databases containing genetic, environmental and clinical information.

*With other similar biobanks underway or being planned around the world, particularly in Europe, it was essential that a major effort should be made for their coordination.*

**5.9** Professor Bell then described some epidemiological studies on large prospective cohorts across the world and indicated that the UK has established many of these, some of which contain more than 10,000 subjects. In addition, the UK Biobank was in the process of being established. This would recruit 500,000 adults and would be used for a range of applications including the identification of disease genes, the validation of biomarkers of pre-morbid samples and the determination of environmental impacts on disease causation within different genetic backgrounds.

**5.10** Professor Bell presented a profile of the possible outcomes of the UK Biobank over the next 40 years to demonstrate the time it would take for the cohort to accumulate sufficient cases in a range of common diseases (estimated to be at least 1,000 cases) to allow cause and effect to be analysed. For many diseases this would take at least ten years to achieve but it would be less for the more common diseases such as diabetes and coronary heart disease. These considerations had determined the lower limit of 500,000 for the proposed cohort size of the UK Biobank.

**5.11** With other similar biobanks underway or being planned around the world, particularly in Europe, it was essential that a major effort should be made for their coordination. An immense resource for human population epidemiology would then become available to investigate primary genetic effects or gene-environment interactions in specific diseases.

**5.12** Professor Bell then suggested that the use of molecular phenotyping methodologies could be effectively scaled up and used for the epidemiological analysis of relatively large populations. As a consequence, the MolPAGE project was currently assessing which technologies could be used to undertake diagnosis of, and risk assessment for, type 1 diabetes and cardiovascular disease in large populations. MolPAGE was focusing on key areas such as proteomics and protein quantification and was utilising GenomeEUtwin as a sample population. He noted that a related activity was the antibody proteomic project led by Mathias Uhlen (KTH Biotechnology). This aimed to develop monoclonal antibodies to as many protein products of the genome as possible. Initial work has focused on proteins expressed by chromosome 21 where antibodies have been developed for use in tissue micro-arrays. The work done so far indicated that peptidomics (small proteins up to 15kDA) can be used to find markers of metabolic disease.

**5.13** Returning to genetics, Professor Bell asserted that there was a central flaw in many genetic studies, which assumed that diseases are caused by common variants. Evidence suggested that some diseases, and in particular those not subject to positive selection, may be caused by rare mutations. He believed that such rare variants could have a large effect on disease and that re-sequencing efforts will be the most powerful method in understanding these effects.

# 6 Expectations and goals for industry and healthcare

**Chairs:** Professor David Porteous (University of Edinburgh) and Professor Bartha Knoppers (University of Montreal)
**Rapporteur:** Dr Helena Andersson (Karolinska Institute)

**6.1** Initiating the discussion, Professor Porteous indicated that the EU populations are valid ones to study and that everyone could benefit from advances in human genome research. The life-science industry will be transformed, and personalised medicine, the use of marker-assisted diagnosis and targeted therapies derived from an individual's molecular profile will impact the way drugs are developed and prescribed. Gene and environment studies within and between EU-population isolates are a real strength. The research studies need to be well designed and powered to detect the anticipated risk (or protective) effects, which then need to be replicated and validated. Harmonisation, with common points of reference, will add value and facilitate the efficient translation of research findings.

**6.2** Identifying and understanding biomarkers will add to the information available on the molecular basis of disease. This work should not be limited to diseases such as cancer and cardiovascular disease but should be extended into other areas such as psychiatric disease, linking genetic studies to physiology and physical imaging. This will bring with it some challenging IT problems that must be understood and solved. This will require governmental, regulatory and ethical support if the corresponding research is not to be hindered.

**6.3** Critical questions that needed to be addressed in FP7 included: How can genetics be used to optimise healthcare at both the population level and individual level? Can biomarkers be identified that measure phenotype and predict disease or response to treatment? How can the biomarker industry be stimulated? How can collaborations be fostered and what will be the most productive areas for collaboration between academia and industry? Can research be disclosed and freely published? Are there conflicts between scientific, public and commercial interests? If so, how can these be addressed? To what extent can full access to data be negotiated by the research and health provider communities? What resources/facilities are available now or will be needed to support translational research in the future?

**6.4** Professor Knoppers noted that P3G was working in this area, attempting to understand governmental and ethical policies in order to identify the current barriers to the adoption of population-based research and how they

might be overcome. Some of the major issues in the translation of research findings into clinical settings will revolve around public trust and confidence in sharing genetic data. Political impediments to the support of longitudinal studies, which tend to be long-term and beyond the life-time of many governments, need to be overcome and this will necessitate the development of better research infrastructures and better knowledge creation pathways.

**6.5** Intercomparability and sharing of data in the future will require consistency of policies on confidentiality and consent. P3G has advocated a broad consent for this purpose and this will accommodate both unforeseen and unknown research uses of genetic data. However, this could only be accomplished if there were proper governance and an unassailable right for participants to be able to withdraw from longitudinal studies.

**6.6** A major barrier to translational research in this area was the lack of a common language. More than 30 different expressions are used around the world concerning the identifiability of samples. There is also a failure to achieve a cross-study level of data standards, security and protection. Whilst this is not unique to biobanks, population epidemiologists do need to become more transparent and robust regarding data handling procedures. Researchers must be able to maintain their own sample collections and document exactly what they are to be used for and how samples may be shared. The associated intellectual property rights must be rationalised and the granting of patents narrowed. Inappropriate ethical reviews must be avoided. From a legal point of view, aggregated datasets belong to human rights and data protection legislation rather than new genetic-specific legislation.

**6.7** The theme of intellectual property rights was continued in the open debate where it was suggested that having a suitable patenting policy with partnership rights for the findings would facilitate obtaining an appropriate return on any research investment. It was suggested that it may be possible to patent analytes where different sets of measurable markers could be predictive of the same disease. This would become more difficult if proximal markers of disease are found. It was felt that academia wants to publish and is eager to validate its technology

whereas companies want access to biological expertise and relevant biological samples. Whilst this may be possible using simple agreements, these, and the complex legal issues involved, need to be agreed before a project starts. All parties have to be realistic in their expectations.

6.8   It was subsequently suggested that industry today needs access to samples from large populations that are well phenotyped, and not just collected in comparably small clinical trials. In response, some members of the audience indicated that they had experienced problems with informed consent when they started collaborations between a public-based biobank and industry. The example given involved the collection of substantial cases and samples over time from 100,000 pregnant women. According to the informed consent obtained any release of samples or personal information to companies was prohibited. However, it was recognised that the example given did not represent a unique problem and, even so, subjects could be re-consented or the data can be anonymised. Industry often has the same problems as academia.

6.9   The audience then returned to the theme of interactions between academia and industry when it was stated that one of the biggest reasons for little collaboration is the lack of a clear statement of the potential for commercialisation. There are good ideas about which genes might influence health or susceptibility to disease but these have to be validated in a public health setting. It was agreed that it is in both academic and commercial interest to have access to well studied biomarkers. It was noted that currently there is strong support for translational research from the UK government.

6.10   The globalisation of clinical trials was then discussed and the increased dominance of China and India was mentioned. It was again recognised that companies may require exclusivity to biobank materials and that agreements over intellectual property rights will be needed. Companies involved in the development of diagnostic tests may have different requirements from those that are involved in drug development. It became clear from the discussion that the long-term stability and public endorsement for biobanks would depend heavily on public opinion. If the improvement of public health is neglected in the long term then public support for the creation of additional biobanks will not be forthcoming. It was subsequently suggested that the biobanks should assume responsibility (and liability) for the data they might contain.

6.11   A view was then expressed that, despite the existence of some very good health policies, reports and proposed supporting infrastructures, all health services around the world try to contain costs. It was therefore important to develop standards of evidence and evaluation before genetics-based health services could be adopted. The necessary platforms for these requirements are currently lacking across the EU and this deficit needs to be addressed. Clever science will not necessarily lead to healthcare providers across the EU Member States adopting new or novel biomarkers of disease

susceptibility into their funding portfolios. National governments are responsible for providing the majority of research funding but experiences across Europe can be very different. Research funding through the Framework Programmes provided by the European Commission was therefore crucial although some of the funding modalities used were sometimes not helpful to potential applicants in some Member States.

6.12   Several participants then emphasised the need to evaluate the outputs of genetics research with respect to the benefits for the patients and the consequences for different healthcare pathways. Attention was drawn to the use of biobanks for addressing psychiatric diseases. This was a poorly resourced area and much could be done to relate genetic testing to functional testing, as exemplified by neuroimaging studies. It was noted that the age groups sampled in each of the biobanks used for analysis would prove to be influential in determining what research questions could be answered. This again drew attention to the importance of issues surrounding the harmonisation and validation of any data collected.

6.13   Comments from other participants also indicated that capacity building and training of future researchers should not be neglected. It was reiterated that the industrial applications of research into human population genetics would be profoundly affected by large-scale population genetic studies. There is a need for partnerships between the industry, the public and the academic/clinical sectors to promote the ensuing opportunities and to share the cost and the expertise needed. The issues involved relate not only to technological advances but also to the many complex problems associated with the differences in culture, ethical views, legal requirements and sociological attitudes that exist across the Member States.

6.14   All agreed that it was essential for the epidemiological studies undertaken across Europe to be sufficiently powered; this can be achieved most easily by pooling data from all the biobank resources available. Common standards and quality control procedures are therefore necessary and these need to be developed. However, novel ways of obtaining consent, which would eliminate the need to return to study participants, were needed but this would necessitate the development of trust and confidence between researchers and the public.
In addition, good interoperable IT systems are required so that information contained in the different datasets available can be adequately mined. Europe needs to invest in the resources necessary to support population genetics and to invest in the training and human capital needed to create research capacity in this important area.

# 7 Resources and analytical tools for population genomics

**Session Chair:** Professor Luisa Bernardinelli (University of Pavia)

**7.1** **Professor Gert van Ommen** (University of Leiden) described the aims and philosophy behind the Netherlands Genomics Initiative (NGI). The initiative was a good example of a public-private partnership and there are four centres of excellence operating within it. Two of these are human health oriented, one focusing on multifactorial disease, the other on cancer. The other two centres of excellence are in the fields of plant genomics and microbial biotechnology. The NGI strategy includes seven other key activities, notably on proteomics, bioinformatics, nutrigenomics and infectious disease.

> Issues of informed consent, the rights and benefits of study populations and equitable access to screening were important issues.

**7.2** Professor van Ommen then described the work of the Centre of Medical Systems Biology (CMSB), one of the centres of excellence within the NGI initiative. This Centre aims to investigate primary disease processes in an integrative way and focuses on a broad range of diseases including Alzheimer's, migraine, metabolic and cardiovascular disease, cancer and arthritis. Specific platforms involving model systems, systems biology, epidemiology and technologies are used and the outputs are brought together through a central data integration platform. The epidemiology platform aims to establish collaborative biobanks that will build high-throughput technology and develop methods for the analysis of complex genetic traits.

**7.3** Professor van Ommen stated that the main aim of the CMSB studies was the integration of longitudinal, twin, case-control and isolate studies for discovery and validation. He cited the Rotterdam Study, the Genetic Research in Isolated Populations (GRIP) programme and the Netherlands Twin Registry as well as many clinical datasets as biobanks included in the initiative. The Rotterdam Study is a long-term study in an out-bred Caucasian population where the incidence of disease had been followed for 15 years. The GRIP study, in contrast, was a family-based study in an inbred population of 3,000 relatives in which any frequently occurring disease with a familial component could be studied within a known pedigree. The reduced genetic diversity of the GRIP population increases the statistical power to detect genes involved

in disease. A good example of this was the identification of an association between the *LIPIN2* gene and diabetes in the GRIP population. The *LIPIN2* gene was subsequently analysed in the Rotterdam Study where a variant of this gene was found to be associated with an increased deposition of body fat on the waist as compared to controls. This work highlighted the value of combining, and preferably embedding, family-based approaches in population studies.

**7.4** The strength of the downstream technological and animal modelling platforms was illustrated by a further study of familial migraine, using a gene that one of the CMSB groups had identified earlier. This work, using mouse knock-in experiments to affect the threshold for the cortisol spreading, showed that there was a relationship with the underlying neurological symptoms associated with migraine. Further analysis of the knock-in mice using a very high resolution MRI scanner identified subtle changes in the morphology of the brain, and underlined the power of such integrative approaches in analysing disease.

**7.5** Professor van Ommen finally identified some potential issues and barriers to future genetic research. These included the media misconceptions regarding genetic research exemplified by Dolly the sheep, headless frogs and green mice. He also mentioned the concern expressed by many that individuals will be profiled for health prediction purposes and that the information gained will then be used either to exclude them from health insurance schemes or to charge them higher premiums.

**7.6** Professor van Ommen argued that an open and sensible dialogue is emerging, and that this is an area where scientists should become actively involved. It is the main area where the battle for acceptance of the outputs of human population genetics research by a broader constituency will be won or lost. As an example, he indicated that while there is a need to reduce the time taken for drug development, this was difficult to reconcile with the need for the public to understand that new drugs may not be free of risk. He stated that issues of informed consent, the rights and benefits of study populations and equitable access to screening were important issues that must be taken into consideration when carrying out large population studies.

**7.7** **Professor Lon Cardon** (University of Oxford) addressed statistical issues associated with population genetics. He explained that the genetic data generated on large populations could be assessed using existing statistical procedures, although there was much work to do to develop new statistical methodologies. The success rate in identifying genes associated with complex traits was relatively low but major changes in the availability of technologies and the number of validated genetic markers across the genome from the HapMap, when taken together, led to great optimism in the field that this rate would improve and there would be many future successes. This renewed promise in genetics would lead to many new studies that would look at hundreds of thousands of markers in thousands of individuals.

**7.8** One such study was the Wellcome Trust Case Control Consortium (WTCCC) which aimed to detect common variants in a range of common diseases at a relative risk of around 1.5. The availability of very large numbers of SNPs provides greater coverage of the genome; however, this could make validation of genetic associations more difficult due to the statistical issues accompanying the many different tests conducted. This demonstrated the need for external validation of results.

**7.9** Replication would be the 'gold standard' in association studies but a lack of replication had unfortunately defined the field so far. While less than 1% of the 31,000 reported associations had been replicated, there was a measurable increase in the number of replications being reported. Replication can sometimes lead to a different gene or a different allele being found, or a different phenotype observed. However, an understanding of the underlying genomic architecture is essential before any possible associations are discounted.

**7.10** Many reasons can be presented in support of replication results that are not robust, and this raises the need for a definition of a true replication in an independent sample group. This will be a major issue when genetic studies are carried out using biobanks and large populations, where studies in the range of 20,000 well-phenotyped samples may be required to fulfil the potential promise of identifying genes in whole genome association studies. To facilitate these studies and the associated translational research, there will be a need to integrate national activities, preparing and storing samples today so that investigations on biomarkers can be conducted in the future.

# 8 Future needs for resources and tools for statistical analysis

**Chairs:** Professor Dorret Boomsma (Vrije University) and Professor Thomas Meitinger (GSF)
**Rapporteur:** Dr Marlies de Lange (Vrije University)

8.1 Professors Boomsma and Meitinger indicated that a major goal of the Human Genome Project is to determine a sufficient number of sequence variants and haplotypes in various populations for genome-wide linkage and genetic association studies for common disease-related phenotypes in global population samples. However, the polymorphisms involved often exert their effects through complex systems in which buffering, feedback, redundancy and robustness exist to dilute their effects. It is necessary to address and explore these issues if large-scale genome-wide studies are to be completed with any degree of certainty.

A key element of successful population genetic studies is the correct selection of the study strategies and statistical methods.

8.2 A key element of successful population genetic studies is the correct selection of the study strategies and statistical methods. This includes development of novel statistical strategies facilitating the maximal utilisation of all collected information in different study samples. Europe has unique study populations that provide additional advantages for special statistical analyses (isolated populations with wide linkage disequilbrium (LD) intervals, twin cohorts, populations with recent admixture). This diversity is beneficial but all these factors set high demands in the strategic design of studies as well as on the statistical and biocomputational expertise in various groups of Europe. Integration of the best biocomputational expertise at European level is essential if the crosstalk between the genes and environment and lifestyle is to be understood.

8.3 The session Chairs therefore suggested that the debate on future needs for resources and tools for statistical analysis should focus on three main issues. First, the challenges associated with collecting samples on an EU-wide scale and the ethical problems in sharing these resources. Second, the funding requirements, particularly for phenotype analyses. Last, the bottlenecks that exist for statistical analysis of the data collected.

8.4 During discussion on the challenges associated with collecting samples on an EU-wide scale it was acknowledged that, at present, there was a lack of appropriate samples on which suitable powered studies could be based. Many of the samples available were disease-based and it was essential for there to be further investment in longitudinal population-based prospective programmes so that premorbid samples could be collected for future association studies. However, this would require collaboration, harmonisation of approach, compatibility of databases and ethical consistency. In addition, there would have to be agreement on methods for sample collection, storage and subsequent access. Some of this has been achieved in the EPIC study where some 400,000 to 600,000 subjects have been recruited along with a full report of cause of death (if appropriate), health conditions and obesity assessments, biomarkers of diet and serum specimens.

8.5 It was then noted that other standardised collections also exist in a variety of projects. These include the MONICA/MORGAM collections which are used for investigating risk factors in coronary heart disease in 25 countries, the respiratory survey for asthma and allergy and the ISARC study investigating these chronic conditions in children. Whilst some of these studies were not originally designed for genetic studies, they could easily be used for this purpose. However, if this were to be the case then there could be problems in gaining informed consent from the respective study participants. There could also be additional problems with respect to subsequent attrition. Issues associated with how to keep people motivated and willing to participate in longitudinal studies have yet to be addressed. A further concern was that many of these collections are epidemiological samples that do not include biological relatives, which makes them less suitable for some genetic studies.

8.6 Other resources that could be used to provide datasets for analysis were then considered. These included disease registries and drug prescription registries. Some of these had now been in existence for over ten years. The meeting concluded that there was a real need to construct an inventory of all relevant resources across Europe and that further funding was required to collect clinical, lifestyle and end-point data in a longitudinal way. The meeting felt that investment was needed in what already existed so that

data warehouses could be constructed; however, this would necessitate the formation of networks of excellence by European researchers and consortia to develop integrated projects. This, in turn, will require national and international collaborations so that a joint infrastructure can be built. There are already good examples in other areas as to how this might be done, for example, the European Biological Resource Network. Biobank networks could learn from these examples. All agreed that samples and data from sufficient numbers of individuals should be available on an EU-wide basis to provide the analytical power on which validated genetic associations could be based.

8.7 It was felt that cultural changes may be needed to achieve the desired level of collaboration as geneticists and epidemiologists, and biologists and biostatisticians, often do not use the same vocabulary. It should be accepted by the scientific community that case-controlled studies, family studies and longitudinal studies are all important and that they could be used to address different questions. An important challenge would be the creation of interoperable databases and improved channels of communication, not only between the clinician scientists involved but also with the general public. The latter would be extremely important if/when the outputs of epidemiological research are used for public health purposes.

8.8 It was then questioned whether there really was a need to increase the number of participants in existing biobanks or whether more effort should be placed on obtaining additional information about existing participants. It was well recognised that obtaining more information per participant, whether this was in terms of collecting biological samples or basic clinical or anthropomorphic information, was difficult and costly. It required specialised people (trained research nurses) and more of these were needed. This highlighted the need for additional capacity building initiatives.

8.9 The ethical problems associated with sharing biobank resources were then discussed. It was clear that legislation across Europe needed to be integrated and harmonised. This legislation also needed to cover issues such as security and confidentiality of information and feedback to the participants of the results of relevant studies. A catalogue of some of the relevant European regulations and guidelines has been constructed but this needs to be continually updated as changes are introduced. The P3G initiative is working to gain European consensus on procedures for sharing data, transferring biological samples, security of data storage and publication, particularly where intellectual property might be involved. All of this work has to take place against the backdrop of an ever changing and developing science base.

8.10 The concept of harmonised ethical regulations was particularly welcomed as it has taken some translational studies in excess of two years to gain the necessary ethical approvals. It was noted that one of the simple ways of streamlining the acquisition of ethical approval was the sharing of best practice, whether this be consent forms, national guidelines or procedures. However, this required time and the issues involved were often regarded to be of low priority by researchers. It was also stated that some of the ethical committees approached for approval lacked understanding of the issues involved and even lacked professionalism. Two ways forward were proposed. First, to assume ethical approval for a study if objections were not raised within a specified time limit. Second, to adopt the principle of the need for only one ethical approval for both national and transnational studies. The European Commission could play an important role in this process.

8.11 The discussion then moved to consider how small genetic effects in complex disease could be identified and used. Such small effects are common but large population studies are needed to identify them. It was noted again that more detailed information on exposure, lifestyle, family background etc would be needed. This might lead to the introduction of different experimental designs for the analysis of different complex diseases. A systems approach was needed to find rare variants with strong effects and this might necessitate resequencing studies on large populations. It was noted that, whilst it was generally true that higher numbers of participants in any study would giver greater power and statistical significance, the outcome would still depend on the quality of the phenotypic data available.

8.12 The meeting was uncertain as to how knowledge of small genetic effects on complex diseases could be put into use by the clinic as it is not only the odds ratio that is needed but also the attributable risk. Clinicians would also need to be able to understand and interpret the

available data. It was felt that continued strategic funding of biobanks was essential if there was to be effective translation of epidemiological research findings into clinical benefit. This represented a long term commitment and one which national governments alone would find difficult to justify. EC Framework funding could play a crucial role in this process even though the construction and maintenance of the desired data warehouses was not regarded as primary research. Funding is needed for the construction of resources on which further research can be based. It was suggested that private funding from industry or charities should not be ignored but that this would need better dialogue between national governments and relevant industries, something again that the EC could promote.

8.13 The forum then discussed the resources available for statistical analysis. It was noted that the computational capacity needed in this area was growing exponentially. For several approaches, there is a need for empirically based significance levels. Over the last five years the field had moved from analysing just a few thousand SNPs over tens of individuals to tens of millions of SNPs over thousands of individuals. This is leading population epidemiology into an era where genetic screening is possible. Whilst costs may be contained as capacity increases, they will remain relatively substantial. The difficulties in obtaining clear and common definitions of phenotype were raised along with the associated problem of obtaining funding to support relevant research in this area. The ultimate goal should be a consensus set of standards.

8.14 It was considered likely that statistical analysis would become more complex as the size of the datasets became larger and the number of possible gene-environment and gene-gene interactions influencing disease susceptibility increases. This would promote greater demand for access to larger national and international electronic storage capacity and ultimately to the development of alternative statistical methodologies and strategies.

8.15 The Chairs then summarised and indicated that effective use of the funds to be made available during FP7 would be crucial. Better communication between geneticists and epidemiologists was needed and the two fields needed to be more integrated. It was essential to increase training opportunities, particularly in statistical genetics, so that capacity in this area across Europe could be improved. Last, improvements were also needed in the standardisation and quality of data collection and in the harmonisation of legal and ethical requirements across Europe.

# 9 The social impact of human population genomics

**Session Chair:** Professor Bartha Knoppers (University of Montreal)

**9.1** **Professor Hans Galjaard** (Erasmus University) addressed the issues surrounding social acceptance of gene technology and the potential contribution of population genomics to heathcare. He highlighted the large variation in the scale and quality of healthcare across Europe, which was greater still when global populations were considered. Over time, the burden of disease would move from illnesses causing mortality, such as cardiovascular disease, to other non-fatal but crippling conditions, such as Alzheimer's disease and depression. Medical genetics has developed from a clinical discipline to include DNA technologies that now allows individual DNA profiling and personalised diagnostics and treatment. There had been a corresponding increase in the level of genetic testing and counselling in the various healthcare systems across Europe, countering the view that genetic research had not led to real benefits for patients.

**9.2** Evidence from genetic testing in hypercholesterolemia and Huntington's disease had shown that an individual's response to a genetic test was directly related to what interventions were available for the disease. In hypercholesterolemia, the availability of effective medicines meant that 90% of potential carriers identified through family history attended the clinic for genetic testing. In contrast, only 10% to 20% of potential carriers of Huntington's disease are willing to be tested in the absence of any clinical intervention. In addition, invasive screening procedures, as in colorectal cancer, or severe clinical interventions such as mastectomies in breast cancer, also significantly reduce the numbers coming forward for genetic testing.

**9.3** Professor Galjaard then suggested that the main aim of genetic association studies is to identify causative genes and offer suitable interventions, if these are available. However, a major consideration has to be the behaviour of those identified as having a higher risk of disease and how they can be influenced to change their behaviour in response to this increased risk. A clear example is the risk-taking behaviour exhibited by smokers and how, despite the wealth of evidence confirming the dangerous effects, the incidence of smoking is increasing in some countries. Overall, it has taken 50 years to halve the number of smokers worldwide.

> Invasive screening procedures, as in colorectal cancer, or severe clinical interventions such as mastectomies in breast cancer, also significantly reduce the numbers coming forward for genetic testing.

**9.4** Access to genetic information may also prompt irrational behaviour by the public. An example of this is evidenced by the $13 billion per year spent in the US on food supplements that have not been shown to provide real health benefits. Whilst the barriers, such as time and development costs, in getting new medicines adopted are well known, continued investment in this area and the identification of new, effective medicines will be the key to health improvements in the future. The best example of this is the success of new drugs in lowering blood cholesterol levels.

**9.5** Professor Galjaard believed that the promise of individual gene expression profiling or pharmacogenetics was some way off and he had not seen any plans on how this would be paid for and delivered in a healthcare setting. Embryonic stem cell research was an essential activity that would be a real alternative to organ donor replacement. When considering embryonic research in a global perspective, the number of embryos used for research purposes was very small when compared to the number of unwanted pregnancies worldwide. Professor Galjaard asserted that most progress in this area will be made by those liberal countries that permit embryonic stem cell research.

# 10 Novel technologies and biomarkers

**Session Chair:** Professor Leena Peltonen (University of Helsinki)

**10.1** **Dr Panagiotis Deloukas** (Wellcome Trust Sanger Institute, Hinxton) discussed how DNA variation across the human genome could be used for future biomarker development and how this might impact on health.

> The HapMap Project had chosen to use a mix of platforms which enabled them to convert 99% of the available SNPs into working assays.

**10.2** Sequencing the human genome was now almost complete with relatively few gaps and the International HapMap Project had already identified around 10 million common variants in human populations. Despite this, the full spectrum of variation had not been captured. The Wellcome Trust Sanger Institute was addressing this issue in a project aimed at re-sequencing the exons and non-coding regions in Caucasian and African population samples. Other important sources of variation in the genome included large copy number variation, epigenetic variation, such as methylation, and transcription variation. The Sanger Institute had initiated specific projects to address each type of variation and was developing high-throughput approaches.

**10.3** Dr Deloukas explained the variety of genotyping platforms available and emphasised that it was necessary to have different platforms available for each phase of a large genotyping project. Each genotyping platform had its own advantages and disadvantages, but all required vigorous quality assurance and quality control measures. The HapMap Project had chosen to use a mix of platforms which enabled them to convert 99% of the available SNPs into working assays, and to reach a 99.5% overall accuracy across the project. Data from the HapMap Project has provided some understanding of the linkage disequilibrium architecture across the genome and has facilitated the selection of tagged markers of common variation.

**10.4** Dr Deloukas went on to describe the Wellcome Trust Case Control Consortium (WTCCC) which is a large collaborative network of geneticists, clinicians, statistical geneticists and technologists across the UK. The WTCCC aims to evaluate and optimise the design of large-scale genetic association studies and develop new methods for analysis. There will be two phases. In phase I, 675,000 SNP markers will be analysed in 1,000 cases in each of eight diseases. In phase II the best 5–10% of markers will be followed through in 1,000 new cases for each disease. In a separate experiment, 15,000 non-synonymous markers would be analysed in 1,000 cases for four other diseases.

**10.5** Both experiments would use common controls from the 1958 British Birth Cohort and the newly established National Blood Service control collection. The study was powered to detect effects with an odds ratio of around 1.5, which although less than the ideal of 1.2, was still reasonably well powered. Replication of any associations detected and further functional studies would be carried out by disease investigators. This type of genetic association study would be unlikely to identify rare variants. These will be best detected using re-sequencing technologies, of the type currently under development by companies such as 454 and Solexa.

**10.6** **Dr Klaus Lindpaintner** (Roche Genetics) addressed the potential benefits and realities of utilising biomarkers in medicine. Through a series of case studies, Dr Lindpaintner demonstrated the hurdles that need to be overcome before drugs are licensed, particularly emphasising those likely to be involved in the development of personalised medicines. For drugs, such as Herceptin, that are effective in only a percentage of patients their continued development by a pharmaceutical company will depend on both the science and the likely economic benefits. He suggested that personalised or pharmacogenomics-based medicines were never likely to be 'blockbusters' for a pharmaceutical company.

**10.7** Even when drugs have been licensed there are still hurdles to their introduction. Regulatory bodies, such as NICE in the UK, consider the incremental cost to efficacy ratio of every drug before giving approval for its general use. If the cost per adjusted life year is above a particular threshold then general approval will not be given.

However, this decision can be changed if the relevant patient groups have been already been segmented into those that respond and those that do not. Under these circumstances, the incremental cost per adjusted life year will be reduced for the responding group.

## Even when drugs have been licensed there are still hurdles to their introduction.

**10.8** Dr Lindpainter noted that the beneficial effects of a drug, even when published, may be less unambiguous than reported. His experience with a new drug targeted to the epithelial growth factor receptor implicated in 80% of lung cancers, highlighted the variance that can be obtained in different analyses. The Roche version of this drug, Tarceva, showed no efficacy with respect to EGFR inhibition, but instead increased longevity in patients. The determination of the efficacy of a drug is based on a series of experiments with the results analysed at each stage by analytical algorithms where the potential cut-off points are quite subjective. This leads to a range of potential outcomes at each stage of the analysis, and can result in a very complicated overall picture.

**10.9** The sensitivity and specificity of a drug or biomarker is also a major consideration, and for them to be useful, sensitivity and specificity levels of 90% or greater will be required. Significantly, the relative importance of sensitivity and specificity will differ when developing an intervention in a life-threatening disease compared to a trivial illness. When considering a serious illness, the drug must be very specific to avoid denying the medicine to a very ill person on the basis of an adverse reaction. In contrast, drugs developed for a trivial illness must be highly sensitive, to avoid adverse reactions in patients not greatly ill, whereas specificity is less important.

**10.10** Taking all these factors into consideration, Dr Lindpainter proposed that it is unlikely that clear-cut positive or negative biomarkers or drugs will be discovered. Instead, complex algorithms will need to be employed to determine the effectiveness of not one, but many biomarkers. This would involve many different assessments, including SNP analyses, expression profiles, clinical and demographic parameters, which together impact on the course of an illness. Dr Lindpaintner concluded that it is unlikely that the discovery of biomarkers will speed up the development of cheaper new drugs. However, discoveries in this area should produce safer, more specific and therefore more effective medicines.

# 11 The development of novel technologies and biomarkers

**Chairs:** Professor Achilleas Gravanis (University of Crete) and Professor Ségolène Aymé (INSERM)
**Rapporteurs:** Dr Cosetta Minelli (University of Leicester) and Dr Louise Leong (Wellcome Trust)

**11.1** Professors Gravanis and Aymé initiated the discussion by indicating that the Human Genome Project had not only resulted in the production of a consensus sequence for the human genome but had also facilitated the development of high-throughput methodologies for DNA sequencing, genotyping and the measurement of large-scale genomic re-arrangements. It is likely that these technologies will steadily increase in power and decrease in cost – essential if the potential of genomic technologies is to be harnessed for large-scale epidemiological studies. However the entry cost into large-scale genomic technologies is likely to be high and this argues for further investment in existing, well-established high-level facilities within the European Research Area. Access to these facilities should be made available to all EU Member States and Associated Countries. This was a role for the European Commission.

**11.2** A key challenge that must be met by existing (and prospective) high-level facilities or centres of excellence is the creation of standard operating procedures and associated quality control and quality assurance procedures to make sample and data exchange possible and to facilitate meta and mega analysis in a robust and statistically sound fashion. There is a concomitant need for a strong research and development component to secure lower cost and higher sensitivity assays and to develop low unit cost and highly portable devices to optimise the dissemination of appropriate technological capacity. This must go hand in hand with the development of IT infrastructures for data archiving, audit and statistical analysis; and will require web and grid-based interfaces for data acquisition, analysis and distribution. Data security and confidentiality will be of paramount concern for ethical, intellectual property development, and commercialisation reasons.

**11.3** Establishing and making available lymphoblastoid cell lines from representative members of different sample collections will be one important contribution to extending the scope for epidemiological research beyond DNA typing. Similarly it will be desirable to establish and maintain biopsy and tissue sample collection to agreed standards suitable for genomic analysis. In addition to the DNA-based technologies, there is an argument for developing complementary RNA-, protein- and metabolite-based genomics technologies. These technologies will introduce higher order challenges because the levels of RNA, protein and metabolites are more likely to be subject to environmental, epigenetic and developmental variation.

**11.4** During discussion, molecular diagnostics was identified as one of the fastest growing segments in the healthcare industry. The rapid growth of this segment is mainly due to novel applications becoming available to address unmet medical needs. This growth is promoting a shift from hardware development to the identification of biomarkers that can be used to answer a specific medical question. Different application categories can be distinguished: the establishment of diagnosis in a symptomatic person (Diagnostic Testing), the assessment of an individual's predisposition to get a certain disease (Risk Testing), the testing of an asymptomatic population in order to prevent and/or detect disease earlier (Screening), the prescription of personalised medicines guided by tests that assess the likelihood of an individual patient to respond to, or suffer adverse side-effects from, a particular drug (Pharmaco-Diagnostic Testing) and the assessment of recurrence or progression of disease (Monitoring). It was clear that sample sets with highly standardised and complete clinical records (biobanks) were needed to provide the biomarkers that form the technical basis of these tests.

**11.5** It was noted that non-coding regions of the genome harbour much DNA variation and that they have been given little attention to date. These regions are potentially important in determining disease susceptibility. However, determing each individual's complete genome might prove impossible. Emphasis should therefore be placed on population studies, which need to be broad, large-scale, pan-European and include major and minor ethnic groups. This brought into focus the need for quality standards, particularly where phenotyping and proteomics are considered important.

**11.6** The use of biomarkers was explored in greater detail. Longitudinal studies and access to population data will be essential in permitting identification of biomarkers for research and discovery of novel compounds. However, biomarkers might be poor predictors of disease susceptibility as some are known to show variation due to environmental conditions. Therefore, means of including environmental information in existing and new biobanks should be developed. This also highlighted the need to accurately

document not only clinical outcomes but also some of the 'intermediate phenotypes' detected during disease progression, as well as lifestyle and environmental factors. This would facilitate the development of earlier predictors of disease susceptibility.

11.7 A number of speakers considered some of the characteristics of the technologies that need to be developed for population genomic studies. Foremost amongst these was the need for better statistical tools to improve the analysis of an individual's susceptibility to disease. These tools would of course require validation and must be capable of showing the significance of small effects in the development of complex diseases.

11.8 There was a debate on the role of cytogenetics in detecting chromosomal duplications, inversions and translocations – cyotgenetics was acknowledged to be effective but its resolution needed to be improved. Some of this could be achieved by further developments in microarray and other whole genome scanning technologies. However, it was noted that the type of chromosomal alteration is not always important, but rather whether or not the relevant gene is active in expressing a functional product. This sometimes depends on distant positional effects of other genes or control elements. These considerations were then related back to the importance of biomarkers and the need to develop high-throughput proteomic technologies. The view was expressed that proteomics might prove useful for understanding the biological processes leading to diseases such as cancer but its use in the clinic was probably limited at present. The meeting was reminded that the context of clinical questions which needed to be answered should not be forgotten, at risk of the field simply being technology-driven.

11.9 The meeting returned to the need for method standardisation, validation and harmonisation – it was agreed that all biobanks should attempt to collect relevant information using agreed good clinical practice guidelines. However, it was recognised that these guidelines would need to accommodate ethical aspects so that concerns of confidentiality did not interfere with the quality of the research being done. There was some debate as to which organisations should develop such guidelines. It was felt that some established organisations, such as the EMEA

and the FDA, were inherently conservative and so should not be directly involved in this area. Nonetheless, they should be kept informed of developments in technologies and biomarkers and so be able to incorporate these in any changes to their approvals processes. This was considered important as the technologies used by biobanks were still developing rapidly, and annotation rather than standardisation might be more relevant. The discussion finished on a cautionary note with the comment that regulation operated at three distinct levels, with the appropriate level taken into consideration in clinical R&D: first, at the statutory level (for safety and efficacy); second, to satisfy reimbursement requirements (for cost effectiveness); and last, at the health provider level (for clinical governance).

11.10 The Chairs summarised the session by stating that a better networking of existing and future biobanks across the ERA should be a high priority for the European Commission. Suitable EU standards and guidelines should be developed to integrate population genetic analyses into clinical trials and new drug and diagnostic test approval procedures. There is currently a gap between capacity, knowledge and application with little yet known on how exactly genetic knowledge will impact on diagnosis and treatment. There is an urgent need for translational research, ideally carried out by multidisciplinary teams with clinicians, biochemists and statisticians working in consortia in order to make the best use of available resources for the development of high-throughput technologies, particularly those for biomarkers of disease. Research investment, particularly that supported by FP7, is needed in all these areas, and the involvement of all key stakeholders, in particular patient groups, was considered essential.

# 12 Closing keynote address and Commission overview

**Chair:** Dr Bill Baig (EU)

**12.1** **Dr Kari Stefansson** (deCode) discussed the approaches and experience of deCode in moving from population studies through a biobank to taking drugs into clinical trials. He believed that the central task in genetics of finding variants that impact on disease could be achieved in many ways. A genealogical approach is possible in Iceland because data on the entire population is contained in a unique nationwide database stretching back to the settlement of the country 1,100 years ago. The Icelandic population is relatively homogenous and has undergone little admixture, making it a powerful resource when searching for genetic traits. deCode places a strong emphasis on clinical phenotypes as defined by the consensus criteria used in common diseases. The coupling of a broad clinical phenotype with the genealogy allows the company to map particular defects and so circumvent phenotypic ambiguities.

**12.2** Experience with the Icelandic population has shown that the underlying population structure is extremely important when looking for relatively subtle effects in association studies. Using this approach, deCode has successfully identified many loci implicated in a range of diseases such as cardiovascular disease, peripheral vascular disease and stroke. Some of the gene mutations identified have been analysed in populations where the associations were replicated, but the phenotype associated with the disease was different.

**12.3** For example, a gene associated with the induction of atherosclerotic disease, leukotriene B4 hydrolase (LTB4), was found to be associated with a modest increased risk in 15% of Americans. In African-Americans, this variant is found in only 3% of the population, but interestingly, it is not found in Africans. This suggests that the variant associated with a very common disease has emerged since human populations left Africa. This finding underlines the importance of testing variants in different ethnic groups. Hence, the common assumption that the same variant will be identified in all populations is untrue as this would suggest that no new variants had emerged since man emerged from Africa.

**12.4** This is supported by an analysis of recombination rates in women which has shown that those with a greater level of recombination have more children. Selective pressure on diversity is therefore possible. Some of the variation generated in this way will be associated with common

disease and it should be assumed that the human genome is relatively dynamic. While linkage studies are driven by rare, highly penetrant variants, association studies will detect variants with lower penetrance and intermediate frequency and where the impact is great within the population as a whole. Association studies have shown that some common variants may confer a protective effect for disease induction. This is almost always associated with a difference in gene expression levels.

**12.5** Using the deCode experience, Dr Stefansson concluded that it appears unlikely that common diseases will be caused by a collection of rare alleles. Rare variants with a low impact can and will be detected, but this is made easier by the prior use of linkage information. deCode has taken advantage of the large families and population structure within Iceland and combined this with a huge amount of information on phenotype and genotype to take genes and discoveries into the clinic. The European Union can also make a significant impact in this area by similarly capitalising on the populations it has access to, in combination with the well-developed healthcare systems available.

**12.6** **Dr Octavi Quintana Trias** (European Commission) began by indicating that some pioneering work on population genetics had been initiated in FP6. These included the GenomEUtwin, MOLPAGE, GenOSept, DanuBiobank, EUROSPAN, Population Biobanks and EUROHEALTHGEN programmes. These projects have enabled researchers to better understand the ways in which interactions between genes and environmental factors are involved in the causes of common diseases of later life and to determine the influence of specific genetic variation on the development or severity of these diseases. A major outcome will be the ability to more accurately determine the numbers of people carrying a particular genetic variation and hence whether they are at a greater or lesser risk of developing a specific disease. FP7 will support additional research in this area to build on the excellent foundations already prepared.

**12.7** The EU Framework Programmes for research aim to promote fundamental research, the interdisciplinary networking of researchers and the creation of increased capacity by supporting training initiatives. It is intended that the outputs of different projects should converge and
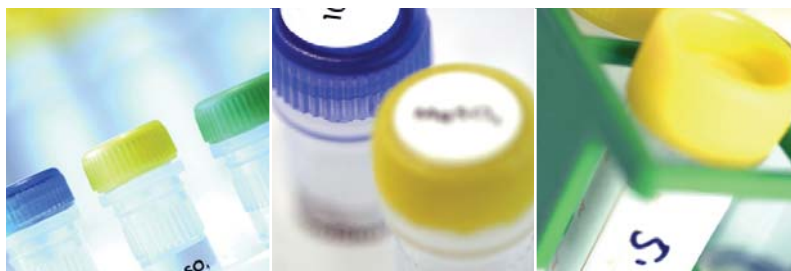
be used together to improve medicine and healthcare for the European citizen. However, this will only be possible if the needs of clinicians, researchers and healthcare providers are conveyed to politicians in Member States, to the European Parliament and to the European Commission so that adequate funding can be secured. It is important to recognise that science is important for society and not just for scientists.

12.8 The European Commission has drawn attention to the fact that European research, in general terms, is under-resourced and is therefore at a disadvantage in its ability to face world-wide competition. There are great strengths in biomedical research across Europe and great potential for transnational co-operation. Hence the transition from FP6 to FP7 will be marked by both a significant increase in the funding available and in the promotion of additional collaborative research that involves greater numbers of researchers in different countries. There will be a simplification in the application procedures for the funds available and there will be a transfer of the logistical and administrative tasks associated with research management to external infrastructures. FP7 will focus on investigator-driven translational research with a reduced political involvement.

12.9 Dr Trias indicated that FP7 will be flexible enough to address both emerging infections and unforeseen policy needs. The dissemination of knowledge and the transfer of results to clinical settings will be supported through collaborative research (Integrated Projects, Networks of Excellence, Co-ordination and Support Actions), joint technology initiatives, co-ordination of non-community research programmes and international co-operation, particularly with developing countries. The main health research theme for FP7 will have three objectives, namely, improving the health of EU citizens, increasing the competitiveness of EU health-related industries and businesses and addressing global health issues. The main rationale for supporting research at a European level was that individual Member States could not support all activities at a satisfactory level. Considerable added value could therefore be achieved by working together. There were good examples of where this had already been successful, as demonstrated by the programmes on human population genetics.

12.10 Research supported in FP7 will contribute to the development of new standards and an evidence base to establish the legislative framework needed to introduce new medical technologies such as stem cells and regenerative medicine into mainstream healthcare. This will provide additional support to European industrial development. However, it will have to be delivered within the context of different national healthcare systems and different models for health care delivery. This, in turn, will drive the need for more health policy research.

12.11 There will be three pillars to support activities within the health research theme of FP7. The first of these involves biotechnology, generic tools and technologies for human health. This will support high-throughput research, the detection, diagnosis and monitoring of disease, predicting the suitability, safety and efficacy of therapies and innovative therapeutic approaches and interventions. The second pillar will focus on the translation of research outputs to improve human health. This will promote large-scale data gathering initiatives, systems biology and research on the brain, human development and ageing. It will also facilitate translational research on new and emerging infectious diseases and on the important chronic diseases such as cancer and coronary heart disease. The third pillar aims to optimise healthcare delivery to all European citizens by supporting the translation of research outputs into mainstream clinical practice. It will underline the importance of ensuring universal access to high quality, safe and effective healthcare and healthcare technologies, irrespective of an individual's ability to pay.

12.12 Dr Trias concluded by stating that there would be a continuation of existing research activities during the transition from FP6 to FP7 but there would also be a greater emphasis placed on translational research. Whilst the focus on genomics might disappear, this did not imply that genomics research was no longer important but implied that the discipline had matured and was ready for integration into other areas of clinical research, adding to the information available to support the delivery of evidence-based medicine. He reiterated that research on human population genomics had great potential for improving the health of the European citizen.

# 13 Summary and conclusions

**Chair:** Dr Bill Baig (EU)

**13.1** Summarising the conference, **Professor David Porteous** first described its overall structure and indicated that the presentations and discussion had raised many issues and posed many questions for human population genetics research. There would certainly be many research outcomes and there would be many challenges in validating these and incorporating them into clinical settings to benefit human health. It was clear that the existing evidence base needs to be further developed as new and improved technologies are introduced. No one Member State would be able to undertake all aspects of the research that was needed and this meant greater emphasis had to be placed on the creation of more partnerships between academics, health services, industry and government organisations. The issues discussed at the conference had drawn attention to the great potential of this research and had clearly demonstrated how population genetics could influence and augment the EU FP7 research agenda.

*The existing evidence base needs to be further developed as new and improved technologies are introduced.*

**13.2** Professor Porteous indicated that the main outputs of the conference and the EUHEALTHGEN initiative should:

- further promote the development and use of both high-throughput DNA sequencing technology and proteomic analysis for clinical research, thereby enhancing data generation, standardisation, acquisition and analysis

- facilitate the detection, diagnosis, monitoring and prevention of disease, particularly the chronic diseases of later life

- generate a better understanding of human genetic variation in predicting disease and so achieve the better targeting of limited health resources

- provide innovative approaches to therapeutic interventions and technologies

- lead to the creation of additional population biobanks, generating and giving access to the large data sets needed to integrate biological data with clinical need, thus promoting translational research for improved human health

- promote the formation of partnerships between the public and private sectors and between the general public and the research community to create both health and wealth enhancing opportunities

- identify some of the sociological factors that may limit the uptake of, and demand for, genetics-based healthcare services

- facilitate a better understanding of both clinical decision-making and policy-making in this area and help improve the efficiency of translating the outcomes of clinical research into clinical practice

- create interdisciplinary and harmonised methodologies for the quality control and quality assurance of genetics-based healthcare services

- promote training and capacity building opportunities in the field of human population genetics across Europe

- shape the regulatory environment in which the advances in human genetics can be efficiently and effectively translated in health benefits for the European citizen.

# 14 Next steps

**14.1** A key objective for the health theme in future EU Research Framework Programmes will be to remain directed towards improving the health of European citizens and increasing the competitiveness of European health-related industries and businesses, while addressing global health issues including emerging epidemics. This will be achieved by supporting research on fundamental genomics (to develop new techniques and technologies), translational research (to validate new protocols in a clinical environment and to make them economically viable, cost effective and readily available) and health services research (to develop methods for health promotion and prevention and so facilitate the introduction of sustainable and efficient healthcare systems).

The consortium partners engaged in EUHEALTHGEN are currently considering how to capitalise on the successes achieved through the conference reported on in this document

**14.2** This represents a hierarchy of support that flows from basic research, through translation and into clinical application. Superimposed onto this supporting infrastructure will be the need to address some of the key health challenges that are currently facing European citizens. These are likely to include issues associated with mental health, metabolic disease, inflammatory disorders, ageing and child health.

**14.3** As demonstrated by this conference, human population genetics, but more specifically the material and data stored within national biobanks, will play an important role in analysing the complex interactions that occur in determining susceptibility and cause of the priority disease areas listed above. However for this to be realised it will be necessary to ensure that the biobanks operating across Europe are compatible so that validated reagents, samples and information can be exchanged in a safe and ethically acceptable way. This points to high priority areas that will be worthy of supporting during the coming 7th EU Framework for research and development. It also provides further justification for the main aim of EUHEALTHGEN, namely, to develop a forward-looking strategy for translating the outputs of population genetics research into clinically useful and health enhancing initiatives whilst improving EU industrial competitiveness in this area.

**14.4** Collaborative partnerships between those involved in human population genetics research across Europe are needed to achieve this aim and EUHEALTHGEN will facilitate the development and provision of essential networking opportunities. The obvious next step is to encourage smaller groups of population geneticists to meet and consider the further use of the samples and information held by the various national biobanks. It will be essential for these groups to be properly briefed by officials from the European Commission and for them to have, as a specific objective, the preparation of high quality proposals across the major activity areas ready for submission following relevant calls for proposals during FP7.

**14.5** The consortium partners engaged in EUHEALTHGEN are currently considering how to capitalise on the successes achieved through the conference reported on in this document and so initiate the necessary next steps to ensure that human population geneticists are able to make a significant and creative contribution to attaining the objectives of FP7. A suitable success criterion for the EUHEALTHGEN project will be the number of successful biobank-related proposals awarded in the acknowledged priority areas of mental health, metabolic disease, inflammatory disorders, ageing and child health. This success will then build on the considerable investment already made in this area by the two sponsors of this conference. The projects currently supported in the area of population genetics by the European Commission and the Wellcome Trust are listed in Annexes 4 and 5 respectively.

# Acknowledgements

A significant number of people have been involved in the production of these conference proceedings and their efforts and contributions are acknowledged accordingly and with grateful thanks.

# Annex 1
# The Scientific Advisory Board

**Professor Ségolène Aymé**
ORPHANET – INSERM SC11
Hopital Broussais
102 rue Didot
batiment Gaudart d'Allaines 5
F – 75014 Paris
**E** ayme@orpha.net

**Dr Bill Baig**
Principal Scientist
Health Research
Research Directorate-General
European Commission
CDMA – 02/159
B-1049 Brussels, Belgium
**E** shahid.baig@cec.eu.int

**Dr Kurt Berlin**
Epigenomics AG
Kastanienallee 24
DE – 10435 Berlin
**E** berlin@epigenomics.com

**Professor Luisa Bernardinelli**
University of Pavia
Dipartimento di Scienze Sanitarie
Applicate
Via Bassi 21
I – 27100 Pavia
**E** luisa.bernardinelli@mrc-bsu.cam.ac.uk
and lab.statgen@unipv.it

**Professor Di Boomsma**
Biological Psychology, VU
Van der Boechorstraat 1
NL – 1081 BT Amsterdam
**E** dorret@psy.vu.nl

**Dr Deborah Carley**
Wellcome Trust Conference Centre
Hinxton
UK – Cambridge CB10 1RQ
**E** d.carley@wtconference.org.uk

**Dr Alan Doyle**
The Wellcome Trust
215 Euston Road
UK – London NW1 2BE
**E** a.doyle@wellcome.ac.uk

**Professor Achilleas Gravanis**
Department of Pharmacology
University of Crete
Stavrakia
GR – 71110 Heraklion
**E** gravanis@med.uoc.gr

**Professor Thomas Meitinger**
Institute of Human Genetics
GSH National Research Centre
DE – Neuherberg
**E** meitinger@gsf.de

**Professor Nancy Pedersen**
Department of Medical Epidemiology
and Biostatistics
Box 281
Karolinska Institutet
SE – 17177 Stockholm
**E** nancy.pedersen@meb.ki.se

**Professor Leena Peltonen**
Department of Molecular Medicine
National Public Health Institute
Biomedicum Helsinki
Haartmaninkatu 8
FI – 00290 Helsinki
**E** leena.peltonen@ktl.fi

**Professor David Porteous**
Molecular Medicine Centre
Western General Hospital
Crewe Road
UK – Edinburgh EH4 2XU
**E** david.porteous@ed.ac.uk

**Dr Marco Tartaglia**
Istituto Superiore de Sanità
Viale Regina Elena 299
I – 00161 Roma
**E** mtartaglia@iss.it

**Project Coordinator**
**Dr Peter Greenaway**
Horus Research Management Ltd
St Brelades, Appledore Hill
Okehampton
UK – Devon EX20 2SR
**E** petergreenaway123@btinternet.com

# List of participants

| LAST NAME | FIRST NAME | INSTITUTION | COUNTRY |
| --- | --- | --- | --- |
| Amersdorfer | Peter | Oridis Biomed | Austria |
| Casari | Georg | Oridis Biomed | Austria |
| Endler | Georg | Medical University Vienna | Austria |
| Mannhalter | Christine | Medical University Vienna | Austria |
| Marsik | Claudia | Medical University Vienna | Austria |
| Schmidt | Konrad | Medizinische Universitat Innsbruck | Austria |
| Wrighton | Christopher | Oridis-Biomed | Austria |
| Miklos | Berta | Edutorial Consulting | Bahamas |
| Baig | Bill | European Commission | Belgium |
| Pisanelli | Federica | European Commission | Belgium |
| Trias | Octavi | European Commission | Belgium |
| Deschenes | Mylene | P3G Consortium | Canada |
| Fortier | Isabel | P3G Consortium | Canada |
| Gaudet | Daniel | University of Montreal | Canada |
| Goddard | Beatrice | Universite of Montreal | Canada |
| Hudson | Thomas | McGill University and Genome Quebec Innovation Centre | Canada |
| Knoppers | Bartha Maria | University of Montreal | Canada |
| Laberge | Claude | Laval University Medical Centre | Canada |
| Legault | Denis | P3G Consortium | Canada |
| Borovecki | Fran | Centre for Functional Genomics | Croatia |
| Jencova | Vera | Technology Centre of Academy of Sciences of the Czech Republic | Czech Republic |
| Macek | Milan | Charles University Prague | Czech Republic |
| Kyvik | Kirsten | Institute of Public Health | Denmark |
| Sorensen | Thorkild | Institute of Preventative Medicine | Denmark |
| Metspalu | Andres | University of Tartu | Estonia |
| Meultenbelt | Ingrid | Leiden University Medical Centre | Estonia |
| Pikani | Jaanus | Estonia Genome Foundation | Estonia |
| Kaprio | Jaakka | University of Helsinki | Finland |
| Palotie | Aarno | University of Helsinki | Finland |
| Peltonen | Leena | University of Helsinki | Finland |
| Perola | Markus | National Public Health Institute | Finland |
| Ayme | Segolene | INSERM | France |
| Cambien | Francois | INSERM | France |
| Cambon-Thomsen | Anne | INSERM | France |
| Clement | Bruno | INSERM Transfert | France |
| Riboli | Elio | International Agency for Research on Cancer (IARC-WHO) | France |
| Thomas | Gilles | Foundation Jean Dausset | France |
| Brand | Angela | German Centre for Public Health Genetics | Germany |
| Illig | Thomas | GSF Research Centre Epidemiology | Germany |
| Ising | Marcus | Max Planck Institute of Psychiatry | Germany |

| LAST NAME | FIRST NAME | INSTITUTION | COUNTRY |
|---|---|---|---|
| Jalilzadeh | Shapour | Institute of Human Genetics | Germany |
| Meitinger | Thomas | GSF National Research Centre for Environment & Health | Germany |
| Pfeufer | Arne | Institute of Human Genetics | Germany |
| Ropers | Hilger | Max Planck Institute for Molecular Genetics | Germany |
| Schambeck | Christian | University of Regensburg | Germany |
| Schmitz | Gerd | Institute for Clinical Chemistry & Laboratory Medicine | Germany |
| Wichmann | Erich | Institute of Epidemiology | Germany |
| Wiedmann | Silke | University of Regensburg | Germany |
| Castanas | Elias | University of Crete | Greece |
| Gravanis | Archilleas | University of Crete | Greece |
| Falus | Andras | Semmelweis University | Hungary |
| Fesus | Laszlo | Hungarian Genomic Consortium for Human Health | Hungary |
| Einarsson | Gudmundur | Decode Genetics | Iceland |
| Stefansson | Kari | Decode Genetics | Iceland |
| Cahill | Martha | Enterprise Ireland Biotechnology Directorate | Ireland |
| McKeigue | Paul | University College Dublin | Ireland |
| Bernardinelli | Luisa | University of Pavia | Italy |
| Coviello | Domenico | Ospedale Maggiore Policlinico | Italy |
| D'Adamo | Adamo Pio | Tigem | Italy |
| Foco | Luisa | University of Pavia | Italy |
| Gasparini | Paolo | Tigem | Italy |
| Klovins | Janis | University of Latvia | Italy |
| Lanzara | Carmela | Tigem | Italy |
| Stazi | Antonia | ISS | Italy |
| Bezzina Wettinger | Stephanie | University of Malta | Latvia |
| Kucinskas | Vaidutis | Vilnius University | Latvia |
| Shbaklo | Hadia | Lebanese American University / Chronic Care Centre | Lebanon |
| Dadoniene | Jolanta | Institute of Experimental & Clinical Medicine | Lithuania |
| Borg Buontempo | Mariella | Ministry of Health | Malta |
| Boomsma | Di | Biological Psychology, VU | Netherlands |
| de Lange | Marlies | Vrije Universiteit | Netherlands |
| Galijaard | Hans | Erasmus University | Netherlands |
| Heijmans | Bas | Leiden University Medical Centre | Netherlands |
| Janssens | Cecile | Erasmus University | Netherlands |
| Kluft | Cornelis | TNO-Quality of Life | Netherlands |
| Perbal | Louise | Ministry of Education, Culture & Science | Netherlands |
| Slagboom | Eline | Leiden University Medical Centre | Netherlands |
| van Duijn | Cornelia | Erasmus University | Netherlands |
| van Ommen | Gert-Jan | Centre of Human & Clinical Genetics | Netherlands |
| Verweij | Cornelis | VU Medical Centre | Netherlands |
| Willemsen | Gonneke | Vrije Universiteit | Netherlands |
| Bergsli | Heidi | Norwegian Institute of Public Health | Norway |
| Hagen | Eilsabeth | Norwegian Institute of Public Health | Norway |
| Harris | Jennifer | Norwegian Institute of Public Health | Norway |
| Stoltenberg | Camilla | Norwegian Institute of Public Health | Norway |
| Lubinski | Jan | Pomeranian Medical University | Poland |
| Lavinha | Joao | Instituto Nacional de Saude Ricardo Jorge | Portugal |
| Santos | Heloisa | Medicine Faculty of Lisbon | Portugal |
| Siska | Ioana | University of Medicine & Pharmacy | Romania |
| Ferak | Vladimir | Comenius University | Slovak Republic |
| Gasperikova | Daniella | Institute of Experimental Endocrinology | Slovak Republic |
| Klimes | Iwar | Institute of Experimental Endocrinology | Slovak Republic |
| Benitez | Javier | Spanish National Cancer Centre | Spain |

| LAST NAME | FIRST NAME | INSTITUTION | COUNTRY |
|-----------|-----------|-------------|---------|
| Bosca | Lisardo | Instituto de Bioquímica | Spain |
| Estivill | Xavier | Centre for Genomic Regulation | Spain |
| Gonzalez | Anna | Spanish National Cancer Centre | Spain |
| Andersson | Helena | Karolinska Institute | Sweden |
| Botling | Johan | National Biobanking Platform | Sweden |
| Hannelius | Ulf | Karolinska Institute | Sweden |
| Litton | Jan-Eric | Karolinska Institute | Sweden |
| Pedersen | Nancy | Karolinska Institute | Sweden |
| Tybring | Gunnel | Karolinska Institute | Sweden |
| Boggio | Andrea | University of Geneva | Switzerland |
| Ozdag | Hilal | Ankara University | Turkey |
| Ozguc | Meral | Hacettepe University | Turkey |
| Allen | Maxine | OCDEM Churchill Hospital | UK |
| Armstrong | Roma | Scottish Executive Health Department | UK |
| Balakrishnan | Sam | The Wellcome Trust | UK |
| Balding | David | Imperial College London | UK |
| Barnes | Michael | GlaxoSmithKline | UK |
| Bell | John | University of Oxford | UK |
| Berzuini | Carlo | MRC BSU Cambridge | UK |
| Bracegirdle | Pippa | Health Protection Agency | UK |
| Brazell | Celia | GlaxoSmithKline | UK |
| Brookes | Anthony | University of Leicester | UK |
| Burton | Paul | University of Leicester | UK |
| Caine | Noreen | Department of Health | UK |
| Camm | Tara | The Wellcome Trust | UK |
| Campbell | Harry | University of Edinburgh | UK |
| Cardon | Lon | University of Oxford | UK |
| Carley | Debbie | The Wellcome Trust | UK |
| Clarke | Robert | University of Oxford | UK |
| Craddock | Nicholas | Cardiff University, School of Medicine | UK |
| Danesh | John | University of Cambridge | UK |
| Davey Smith | George | University of Bristol | UK |
| Deloukas | Panos | Wellcome Trust Sanger Institute | UK |
| Dermitzakia | Emmanouil | Wellcome Trust Sanger Institute | UK |
| Dezateaux | Carol | Institute of Child Health | UK |
| Dhell | Jill | Department of Health | UK |
| Donnai | Dian | Academic Unit of Medical Genetics | UK |
| Doyle | Alan | The Wellcome Trust | UK |
| Dudbridge | Frank | MRC | UK |
| Duncanson | Audrey | The Wellcome Trust | UK |
| Elliott | Paul | Imperial College London | UK |
| Evseeva | Irina | University of Oxford | UK |
| Field | John | University of Liverpool | UK |
| Gallacher | John | Cardiff University | UK |
| Gaasenbeek | Michelle | BHF GCRC | UK |
| Golding | Jean | University of Bristol | UK |
| Goldsmith | Paul | Daniolabs Ltd | UK |
| Gooch | Daniel | Department of Health | UK |
| Goodwin | Pat | The Wellcome Trust | UK |
| Greenaway | Peter | Health Genetics Research Programme | UK |
| Haddow | Gill | University of Edinburgh | UK |
| Haites | Neva | University of Aberdeen | UK |
| Harris | Sarah | University of Edinburgh | UK |
| Henderson | Colin | IBM | UK |

| LAST NAME | FIRST NAME | INSTITUTION | COUNTRY |
|---|---|---|---|
| Holloway | Paul | University of Oxford | UK |
| Hunt | Adrienne | The Wellcome Trust | UK |
| Jarvelin | Marjo-Riitta | Imperial College London | UK |
| Jie | Zheng | Population Genetics Technologies Ltd | UK |
| Jones | Richard | University of Bristol | UK |
| Jones | Ian | The Wellcome Trust | UK |
| Keavney | Bernard | University of Newcastle | UK |
| Kennedy | Karen | The Wellcome Trust | UK |
| Kent | Alastair | Genetic Interest Group | UK |
| Kuh | Diana | MRC National Survey of Health & Development | UK |
| Kwiatkowski | Dominic | University of Oxford | UK |
| Leong | Louise | The Wellcome Trust | UK |
| Lewis | David | Health Protection Agency | UK |
| Lichtenstein | Conrad | Population Genetics Technologies Ltd | UK |
| Mastana | Sarabjit | Loughborough University | UK |
| McArdle | Wendy | University of Bristol | UK |
| McCarthy | Mark | University of Oxford | UK |
| McCarthy | Alun | GlaxoSmithKline | UK |
| Mikawa | Yoshikazu | Population Genetics Technologies Ltd | UK |
| Minelli | Cosetta | University of Leicester | UK |
| Mir | Kalim | Henry Wellcome Building for Genomic Medicine | UK |
| Ollier | Bill | University of Manchester | UK |
| Ouwehand | Willem | University Cambridge & National Blood Service | UK |
| Owen | Michael | Department of Psychological Medicine | UK |
| Peckham | Catherine | Epidemiology & Biostatistics Unit | UK |
| Pell | Jill | Department of Health | UK |
| Pembrey | Marcus | Institute of Child Health, UCL | UK |
| Porteous | David | University of Edinburgh | UK |
| Power | Chris | Institute of Child Health, UCL | UK |
| Pringle | Mike | UK Biobank | UK |
| Rastan | Sohaila | The Wellcome Trust | UK |
| Ring | Susan | University of Bristol | UK |
| Rohlff | Christian | Oxford Genome Sciences | UK |
| Spector | Tim | King's College London | UK |
| Sprosen | Tim | UK Biobank Ltd | UK |
| Strachan | David | St George's, University of London | UK |
| Sudlow | Cathie | University of Edinburgh | UK |
| Surdulescu | Gabriela | King's College London | UK |
| Taylor | Jenny | Oxford Genetics Knowledge Park | UK |
| Taylor | Marc | Department of Health | UK |
| Terry | Robert | The Wellcome Trust | UK |
| Trembath | Richard | University of Leicester | UK |
| Walport | Mark | The Wellcome Trust | UK |
| Webster | Andrew | University of York | UK |
| Whittaker | John | London School of Hygiene and Tropical Medicine | UK |
| Wilson | Jim | University of Edinburgh | UK |
| Worthington | Jane | University of Manchester | UK |
| Wright | Alan | MRC Human Genetics Unit | UK |
| Yuille | Martin | University of Manchester | UK |
| Zabaneh | Delilah | Imperial College London | UK |
| Zimmern | Ron | Strangeways Research Laboratory | UK |
| Brooks | Lisa | NHGRI | USA |
| Gottlieb | Michael | Foundation for the National Institutes of Health | USA |
| Weir | Bruce | North Carolina State University | USA |

# Conference programme

## 20 September

**13.00–15.30**
Registration and poster assembly

**Session Chair: Dr Alan Doyle**

**15.30–15.45**
Wellcome and Introduction
Dr Alan Doyle and Dr Bill Baig

**15.45–16.15**
Overview
Dr Mark Walport, Director, Wellcome Trust

**16.15–16.45**
Overview
Professor D Hiddemann,
University of Munich

**16.45–17.30**
Keynote Address
Dr Tom Hudson, McGill University

**17.30–18.15**
Keynote Address
Professor Leena Peltonen,
University of Helsinki

**18.30**
Buffet/reception and poster viewing

## 21 September

**Session Chair: Professor Nancy Pedersen**

**09.00–09.30**
Industry Perspectives of Population Genomics
Dr Alun McCarthy, GSK

**09.30–10.00**
Translational Research
Professor John Bell, University of Oxford

**10.00–10.30**
Coffee and poster viewing

**10.30–12.30**
Forum and debate – Expectations and goals
for industry and healthcare
Chairs: Professor David Porteous and
Professor Bartha Knoppers

**12.30–13.45**
Lunch and poster viewing

**Session Chair: Professor Luisa Bernardinelli**

**13.45–14.15**
Lecture on resources relevant to
population genomics
Professor G Van Ommen,
University of Leiden

**14.15–14.45**
Lecture on statistical analysis in
population genomics
Professor L Cardon, University of Oxford

**14.45–15.15**
Tea and poster viewing

**15.15–17.15**
Forum and debate – Future needs for
resources and tools for statistical analysis
Chairs: Professor Dorret Boomsma and
Professor Thomas Meitinger

**Session Chair: Professor Bartha
Knoppers, P3G**

**17.15–18.00**
Strategic lecture on the social impact
of human population genomics
Professor H Galjaard, Erasmus University

**19.30**
Conference dinner

## 22 September

**Session Chair: Professor Leena Peltonen**

**08.30–09.00**
Lecture on existing and emerging technologies
relevant to population genomics
Dr Panagiotis Deloukas,
Wellcome Trust Sanger Institute, Hinxton

**09.00–09.30**
Lecture on biomarkers
Professor Klaus Lindpaintner,
Roche Genetics

**09.30–10.00**
Coffee and poster viewing

**10.00–12.00**
Forum and debate – Research directions
for the development of novel technologies
and biomarkers
Chairs: Professor Achilleas Gravanis and
Professor Ségolène Aymé

**Session Chair: Bill Baig**

**12.00–12.30**
Keynote Address
Professor Kari Stefansson, deCode

**12.30–12.45**
Closing Remarks
Dr Octavi Quintana Trias,
European Commission

**12.45–13.00**
Summary and Conclusions
Professor David Porteous,
Edinburgh University

**13.00**
Lunch and end of meeting

# Projects supported
# by the European Commission

| TITLE | SUMMARY/TYPE OF COLLECTION | COORDINATOR'S DETAILS |
| --- | --- | --- |
| EuroBioBank (European network of DNA, cell and tissue banks for rare diseases) | EuroBioBank is a European network of biological banks established in 2001 by 16 founding partners and coordinated by Eurordis (European Organisation for Rare Diseases). The EuroBioBank network is fully dedicated to supporting research into rare diseases by facilitating access to quality human biological resources (DNA, cells and tissues) and their associated data from patients with rare diseases. From 2003 to 2006, the EuroBioBank network has received grants from the European Commission, under the 5th FP, 'Quality of Life and Management of Living Resources'. Its long-term objective is to expand by associating new members across Europe. | Dr Fabrizia Bignami<br>E fabrizia.bignami@eurordis.org<br>www.eurobiobank.org |
| European registry of severe cutaneous adverse reactions (SCAR) to drugs and collection of biological samples (RegiSCAR) | DNA, blood | Dr Maja Mockenhaupt<br>E dzh@haut.ukl.uni-freiburg.de |
| European Vascular Genomics Network (EVGN) | DNA, RNA, vascular cells and tissues, serum | Dr Alain Tedgui<br>E alain.tedgui@larib.inserm.fr |
| Rational treatment strategies combating mitochondrial oxidative phosphorylation (OXPHOS) disorders (EMUMITOCOMBAT) | Use of banked human cell lines with OXPHOS defects | Prof Jan Smeitink<br>E j.smeitink@cukz.umcn.nl |
| European Network on Functional Genomics of Type 2 Diabetes (EUGENE2) | DNA, RNA, cells. The aim of this project was to identify genetic causes of type 2 diabetes. Extensive phenotyping of healthy relatives to type 2 diabetic subjects was done with a focus on insulin resistance and secretion. Gene expression profiles in target tissues were also obtained to identify differentially expressed genes. The intention was to sequence potential candidate genes and to appropriately develop transgenic animal models to relate genotype with phenotype. | Prof Ulf Smith<br>E ulf.smith@medic.gu.se<br>www.eudg.invmed.gu.se/<br>groups.html |
| European genomics initiative on disorders of plasma membrane amino acid transporters (EUGINDAT) | DNA, blood, urine, cells | Prof Manuel Palacín<br>E mpalacin@pcb.ub.es |
| Wilson disease: creating a European clinical database and designing multicentre randomised controlled clinical trials (EuroWilson) | DNA, blood, urine | Prof M Stuart Tanner<br>E M.S.Tanner@sheffield.ac.uk |
| Autoimmune polyendocrine syndrome type I – a rare disorder of childhood as a model for autoimmunity (EURAPS) | DNA, serum, mRNA, immune cells | Prof Olle Kämpe<br>E olle.kampe@medsci.uu.se |
| Prader-Willi Syndrome: a model linking gene expression, obesity and mental health (PVS) | Use of already banked DNA and brain tissue | Prof Tony Holland<br>E ajh1008@cam.ac.uk |

| TITLE | SUMMARY/TYPE OF COLLECTION | COORDINATOR'S DETAILS |
|---|---|---|
| Rare genetic skin diseases: advancing diagnosis, management and awareness through a European network (GENESKIN) | Skin biopsies, cells, DNA and RNA | Dr Giovanna Zambruno E g.zambruno@idi.it |
| Network of European brain and tissue banks for clinical and basic neuroscience (BrainNet) | Human brain | Prof Hans Kretzschmar E hans.kretzschmar@med.uni-muenchen.de |
| Network of European Brain and Tissue Banks for Clinical and Basic Neuroscience (BrainNet Europe II) | Human brain.This project aims at establishing a network of European brain banks and will establish standardisation of brain banking throughout Europe, increase public awareness, include legal and especially ethical issues, and create a Europe-wide database. | Prof Hans Kretzschmar E hans.kretzschmar@med.uni-muenchen.de www.brainnet-europe.org/conference/index.htm |
| European integrated project on spinocerebellar ataxias: Pathogenesis, genetics, animal models and therapy (EUROSCA) | Human blood, CSF, epidemiological data, DNA. The project focuses on autosomal dominant spinocerebellar ataxias (SCA), ranging from basic science to clinical applications. In particular, it will establish a CAPIT-SCA (Core Assessment Programme for Interventional Therapies) clinical scale, maintain a SCA registry, perform research on genetic modifying factors in SCA, develop disease models and pursuit novel therapeutic strategies. | Prof Olaf Riess E Olaf.riess@med.uni-tuebingen.de www.eurosca.org |
| Genomics, mechanisms and treatment of addiction (GENADDICT) | Human blood, epidemiological data, DNA. This project is devoted to validation of animal models and to human studies on the role of genes in complex diseases. The core of the research effort will be the identification of genes associated with drug addiction using an unbiased genome-wide approach. | Prof Ian Kitchen E i.kitchen@surrey.ac.uk |
| From Immune Responses in Rare Autoimmune Diseases to Novel Therapeutic Intervention Strategies (AUTOROME) | DNA, blood | Prof Hans-Jürgen Thiesen E hans-juergen.thiesen@med.uni-rostock.de |
| Evaluation of the Role of Infections in Cancer using Biological Specimen Banks (ERICBSB) | This FP5 project joins seven biobanks together giving a total collection of 2 million blood samples from 1 million healthy individuals. About 50,000 participants have been diagnosed with cancer during follow-up (i.e. after the sampling). The biobank datafiles are linked with participating cancer registries and pathology and cervical smear registries to identify tissue and cell samples from the same subjects who have blood stored. During the contract that finished in 2005 almost 100,000 samples were actually retrieved and analysed in scientific studies on the role of infections in cancer. The consortium continues in FP6 as 'CCPRB' – Cancer Control using Population-based Registries and Biobanks. | Joakim Dillner E joakim.dillner@mikrobiol.mas.lu.se |
| Common molecular pathways in progression of kidney diseases (Chronic Kidney Diseases) | Tissue bank of renal biopsies | Mohamad Daha E m.r.daha@lumc.nl |
| Cancer Control using Population-based Registries and Biobanks (CCPRB) | Cancer Control using Population-based Registries and Biobanks (CCPRB) is a Network of Excellence on biobanking within FP6. CCPRB links 20 large biobank projects with blood samples from about 2 million subjects with up to 30 years follow-up and >100,000 prospectively occurring cancer cases, 7 cancer registries with >40 years of population-based registration and a number of technology platforms for advanced analysis of biobank samples in 9 European countries. Public CCPRB resources include biobank quality standards, sample and data handling services and formalised education programmes. For goals of CCPRB and further information see website. | Prof Joakim Dillner E joakim.dillner@mikrobiol.mas.lu.se E anna.olofsson_franzoia@med.lu.se www.cancerbiobank.org |
| European Randomized Study of Screening for Prostate Cancer (ERSPC) co-ordination, infrastructure and international pooled database | Frozen blood samples from >180,000 healthy men and men with prostate cancer | Prof Fritz Schroder E e.vandenberg@erasmusmc.nl |

| TITLE | SUMMARY/TYPE OF COLLECTION | COORDINATOR'S DETAILS |
|---|---|---|
| **Validation of recently developed diagnostic and prognostic markers and identification of novel markers for prostate cancer using European databases (P-MARK)** | The goals of P-MARK are to deliver novel diagnostic and prognostic serum and urine markers for all stages of prostate cancer (PCa) in order to define men in the general population at risk for life threatening PCa. For this purpose, a serum biorepository of about 2,350 samples has been established containing serum from healthy men and men with various stages of PCa. This biorepository includes subgroups with a clinical follow-up of at least 5 years. In addition, a prospective biorepository is being initiated that will contain consecutive blood (serum, plasma, DNA and RNA) samples from men suspected to suffer from PCa or treated for PCa. It is anticipated that matching urine and tissue samples will be added to this biorepository. | Prof Christaan Bangma<br>**E** h.j.vanalphen@erasmusmc.nl<br>**E** e.schenk-braat@erasmusmc.nl<br>www.p-mark.org |
| **Inherited risk of breast and prostate cancer (POLYGENE)** | Blood samples from breast and prostate cancer patients and controls. POLYGENE will study the genetic underpinnings of prostate and breast cancer using cohorts of the Icelandic and Dutch populations. It will probe the association between common polymorphisms in a large number of candidate cancer susceptibility genes and the risk of breast and prostate cancer and develop computational and statistical methods that focus on the analysis of genetic data from studies of complex diseases. | Thorunn Rafnar<br>**E** thorunn.rafnar@decode.is<br>www.uvs.is |
| **Outcome and Impact of Specific Treatment in European Research on Melanoma (OISTER)** | Human banked melanoma cell lines and clinical, patient bioregistry. Includes information on genotyping of oncogenes and expression of tumour antigens. | Prof Dirk Schadendorf<br>**E** d.schadendorf@dkfz-heidelberg.de |
| **European Searchable Tumour Cell Line Database (ESTDAB)** | ESTAB provides a service enabling investigators to search online for HLA-typed, immunologically-characterised melanoma cell lines available for distribution from a central bank. This enables investigators to identify cells possessing specific parameters important for studies of immunity. | Prof Graham Pawelec<br>**E** graham.pawelec@uni-tuebingen.de |
| **European Human Frozen Tumour Tissue Bank (Tubafrost) (virtual disease related collections)** | This European human tissue bank is composed of frozen tumour tissue sample collections in the major centres now part of the OECI. It is a virtual collection with samples stored locally under local conditions of governance and availability. Sample data is collected in a central database. There are standardised procedures for collecting, freezing and storing tissues to get samples of comparable high quality. There is a code of conduct for exchanging samples in Europe, accommodating the high variety of European and local laws on tissue samples. The rules for access and use allow local collectors to use their samples for their own purpose. The release of samples is determined by the local collector and this depends on co-operation, co-publication, importance of the proposed research and the possible reimbursement of costs. The central database handles requests for samples with difficult pathology and a virtual microscopic support is available. | OECI<br>Peter Riegman<br>**E** p.riegman@erasmusmc.nl<br>www.tubafrost.org |
| **The use of molecular biomarkers in early lung cancer detection (Early Lung Cancer)** | The overall goal of the project is to determine if specific genetic changes occurring in lung carcinogenesis are detectable in the respiratory epithelium of persons who have an increased risk of developing lung cancer. In order to pursue this objective the partners have decided to concentrate on individuals with a very high risk of developing lung cancer, i.e. Second Primary Lung Cancers (SPLC) in 12 centres throughout Europe. All of the recruited individuals will be followed up regularly and also be assessed for a range of molecular/pathological markers currently considered to be involved in carcinogenesis. A large lung cancer biobank including biopsies, serum and bronchial lavage will be created to establish biomarkers of early lung cancer detection. | Prof J K Field<br>**E** J.K.Field@liv.ac.uk<br>www.euelc.com |
| **Genetics of healthy ageing (GEHA)** | Blood; cells; PBMCs, granulocytes; plasma; serum. The object is to identify genes involved in healthy ageing and longevity from 90+ sibpairs from 11 European countries. The methods used include phenotype data collection of 2,650 sibpairs and 2,650 younger controls; 9 high throughput platforms for genome scanning; LD mapping of 3 candidate genomic regions; mitochondrial DNA analysis (aplogroups and subaplogroups identification and C150T mutation analysis). Phenotypic, genotypic and mtDNA data are stored in 3 databases, physically separated but interconnected in order to allow cross analysis and privacy protection. | Claudio Franceschi<br>**E** claudio.franceschi@unibo.it<br>www.geha.unibo.it |
| **The control of embryo implantation (EMBIC)** | DNA, tissues | Gerard Chaouat<br>**E** gerard_chaouat@wanadoo.fr |

| TITLE | SUMMARY/TYPE OF COLLECTION | COORDINATOR'S DETAILS |
|---|---|---|
| **Genetic Markers for Osteoporosis (GENOMOS)** | The aim of this project is to identify, by prospective meta-analysis of association studies, susceptibility alleles for parameters of osteoporosis i.e. BMD and fracture risk. It also looks for the influence of gene-gene and gene-environment interactions on the effect of these risk alleles. Over the course of three years the collection has grown from 18,000 samples to now over 40,000 DNA samples of which about 50% are kept and genotyped centrally (at ErasmusMC). Standardised procedures for genotyping (using a reference DNA plate) and phenotyping have been established. The samples are from well characterised epidemiological study cohorts across Europe, as well as Canada and the USA. The consortium is part of the HuGENET Network of Networks. | André G Uitterlinden **E** a.g.uitterlinden@erasmus.nl www.genomos.org (available May 2006) |
| **Viruses in Diabetes (Virdiab)** | Blood, plasma, stool. | Heikki Hyoty **E** heikki.hyoty@uta.fi |
| **Epidemiology of HPV infections in cutaneous squamous cell and basal cell carcinomas (EPI-HPV-UV-CA)** | Serum, DNA from eyebrow hairs, skin biopsies, epidemiological data. The object of this study was to clarify the role of HPV infections in the development of squamous cell and basal cell carcinoma in relation to sunlight exposure by combining methods of modern epidemiology and molecular biology. Specifically, the project studied: (i) the kinetics of HPV infection in organ-transplant recipients, (ii) the effect of immunosuppression on HPV load, by measuring the viral load around and at different time points after organ transplantation, (iii) the association of HPV infection with skin cancer in relation to exposure to sunlight by comparing northern Europe with southern Europe and Australia, (iv) the presence of high-risk HPV types by using newly developed HPV-DNA typing systems, and (v) the question whether all malignant cells contain at least one HPV genome, by determining the HPV load in histopathologically precisely defined skin cancers. | Jan Nico Bouwes- Bavinck **E** J.N.Bouwes_Bavinck@lumc.nl |
| **Viruses in Skin Cancer Risks (Viraskin)** | Skin biopsies | Ola Forslund **E** ola.forslund@mikrobiol.mas.lu.se |
| **Carotid Intima Media Thickness (IMT) and IMT-Progression as Predictors of Vascular Events in a High Risk European Population (IMPROVE)** | DNA, blood | Prof Rodolfo Paoletti **E** rodolfo.paoletti@unimi.it |
| **Genetic Regulation of the End-Stage Clotting Process that Leads to Thrombotic Stroke (EuroClot)** | EuroClot aims to identify and validate potentially therapeutically useful genes associated with thrombolytic stroke. This will be carried out using a novel approach, by means of uncovering the genes that control the end-stage of the coagulation process that leads directly to the production of the thrombus (clot) that causes vascular obstruction and tissue death. EuroClot will study stroke intermediate phenotypes in over 3,000 twins from northern and southern Europe. Genes will be validated in 1,000 stroke cases. Phenotyping will be standardised and harmonised and a European database established. | Prof Tim Spector **E** tim.spector@kcl.ac.uk www.twin-research.ac.uk/ euroclot.html |
| **Molecular mechanisms involved in organ-specific metastatic growth processes in breast cancer (MetaBre)** | Investigation of the molecular mechanisms of breast cancer metastasis, including the identification and functional analysis of new genes and proteins that characterise the organ-specific metastatic phenotype in breast cancer cells, and the interaction of cancer cells with the affected organ tissues. To support these studies MetaBre has used existing collections of metastases from breast cancer and primary breast tumours at Centre René Huguenin (Paris), supplemented by additional samples from pathologists and tumour banks in Italy, Spain and Belgium. Also, blood sera from patients has been collected for testing of biomarkers of organ-specific metastasis for analysis at Ghent University. All tissue samples are utilised after approval by the relevant ethical committee, and if applicable after approval by the patient. | Anna Teti **E** info@metabre.org www.metabre.org |
| **Genetic and environmental risk factors for melanoma: translation into behavioural change** | Development of a centralised bank of melanoma patient samples including DNA, tissues and lines | Julia A Newton Bishop www.genomel.org |

| TITLE | SUMMARY/TYPE OF COLLECTION | COORDINATOR'S DETAILS |
|---|---|---|
| Identification of low and moderate penetrance genes predisposing to colorectal cancer utilising established and novel biotechnology | Identification of low/moderate penetrance genes predisposing to colorectal cancer based on extensive sample series of patients representing different European populations. Collection of fresh as well as paraffin embedded normal and tumour tissue samples to identify tumour predisposition genes. The rationale is to understand mechanisms of tumorigenesis by identification of key human cancer susceptibility genes and modifier genes, as well as to create tools for diagnosing tumour predisposition to allow cancer prevention. | Lauri Aaltonen<br>E lauri.aaltonen@helsinki.fi |
| Defects in the tricarboxylic acid (Krebs) cycle genes in tumorigenesis | Improved prevention and management of TCAC associated hereditary cancers. The two key tasks to achieve this goal are: 1) characterising the natural history and prevalence of TCAC deficient cancers; 2) unravelling the molecular mechanisms driving TCAC associated tumorigenesis. Collection of fresh as well as paraffin embedded normal and tumour tissue samples to identify tumour predisposition genes. The rationale is to understand mechanisms of tumorigenesis by identification of key human cancer susceptibility genes and modifier genes, as well as to create tools for diagnosing tumour predisposition to allow cancer prevention. | Lauri Aaltonen<br>E lauri.aaltonen@helsinki.fi |
| Identification of risk genes for atherothrombosis in coronary artery disease by transcriptome and proteome analysis and high throughput exon resequencing (Bloodomics) | The main aim of this project is to discover genetic markers for the prediction of thrombus formation in coronary heart disease, and to design better anti-thrombics for improved prevention and treatment. | Dr Willem Ouwehand<br>E who1000@cam.ac.uk<br>www.bloodomics.org |
| From molecules to networks: understanding synaptic physiology and pathology of the brain through mouse models (EU-SYNAPSE) | The project aims at a comprehensive analysis of the molecular mechanisms mediating synaptic information processing and function, in the normal brain as well as of dysfunction in a group of neurological diseases defined as 'synaptopathies'. | Reinhard Jahn<br>E azimek@gwdg.de<br>www.eusynapse.mpg.de |
| X-linked adrenoleukodystrophy (X-ALD): pathogenesis, animal models and therapy | This project aims at understanding the pathogenesis of X-linked adrenoleukodystrophy and will include a wide range of activities: unravelling the function of the ALD protein using in vitro (liposomes) and in vivo (animal) models, and identification of genes and proteins differentially regulated in the target tissues of patients. | Johannes Berger<br>E johannes.berger@<br>meduniwien.ac.at |
| Paraneoplastic Neurological Syndromes (PNS) Strengthening the European Network (PNS-Euronet 2) | The aim of this project is to consolidate a European research partnership to collect data on patients with paraneoplastic neurological syndromes, develop clinical scenarios and identify relevant clinical issues to be addressed in prospective studies. A research inventory will also be established to exploit the database for clinical and scientific research on these rare diseases throughout the EU. A Sample Bank (serum, cerebrospinal fluid, lymphocytes) of patients with PNS has also been set up to provide standard guidelines for antibody detection and test the relevance of new antibodies initially detected in one or a few patients with PNS. The availability of centrally stored sera, fluids and biological material from PNS patients will give the entire scientific community access to samples for specific studies on adequate numbers of patients. Further it could enable any newly detected reactivities to be appropriately identified and characterised, given the availability of a higher number of sera with these new properties. | Bruno Giometto<br>E bruno.giometto@unipd.it<br>www.pnseuronet.org |
| New molecules in mood disorders: a genomic, neurobiological and systems approach in animal models and human disorder (NEWMOOD) | This is a study of new molecular mechanisms in the causation of depression and of effective drug-treatment in animal models and humans. Vulnerability, depression, and recovery will be assessed in humans with varying genetic and environmental risk factors. Endophenotypes will be defined and related to new candidate polymorphisms and genetic markers. Cohorts of about 200 persons each will be recruited from inner cities of Manchester, Würzburg, Budapest and Maastricht. | Bill Deakin<br>E newmood@manchester.ac.uk<br>www.medicine.manchester.ac.uk/<br>psychiatry/newmood |
| Migraine genes and neurobiological pathways (EUROHEAD) | Migraine is a common chronic pain syndrome with recurrent attacks of disabling headache, associated symptoms and, in 1/3 of patients, neurological aura symptoms. Genetic factors are involved in the mechanisms for the attack, pain and aura. EUROHEAD aims to identify and validate these genes, and unravel their functional involvement in migraine pathways by using clinical and (post-)genomics techniques in patients, cellular and transgenic mouse models. EUROHEAD collects blood and DNA from migraine patients and families and uses transgenic models of migraine genes. | Michel Ferrari<br>E M.D.Ferrari@LUMC.NL<br>www.eurohead.org/index.html |

| TITLE | SUMMARY/TYPE OF COLLECTION | COORDINATOR'S DETAILS |
|---|---|---|
| **Using European and international populations to identify autism susceptibility loci (AUTISM MOLGEN)** | Identification of susceptibility alleles for autism spectrum disorders in European genetically isolated populations. Data from 425 previously ascertained multiplex families will be pooled and a meta-analysis performed to identify the top 6 susceptibility regions for autism spectrum disorders. Over the course of the project the genotype of approximately 130 new multiplex and 280 new singleton families will be assessed. The top susceptibility regions will be tested in genetically isolated populations of Finnish and Friesland singleton trios to search for extended haplotypes that may further refine the critical regions. Subsequently brain expressed candidate genes in these narrowed six regions of linkage will be tested for mutations and association with autism. | Anthony Bailey<br>**E** anthony.bailey@psych.ox.ac.uk |
| **Dyslexia genes and neurobiological pathways (DYSLEXIA)** | Multicentre, multidisciplinary project to investigate the biological basis of dyslexia by collecting powerful samples of subjects consistently characterised across EU populations on three different levels: genetics, environment, and neuroscience. The aim is to understand the etiology of the disorder by integrating the results of the three levels. On the genetic level, a systematic two stage approach will be used to map and clone dyslexia susceptibility genes in samples of 800 families and 1,950 dyslexic cases and 1,950 controls. The identified risk-conferring genes will also be used to understand gene-gene and gene-environment interactions, as well as gene-specific contributions to a variety of neurobiological correlates of dyslexia. | Gerd Schulte-Körne<br>**E** schulte1@med.uni-marburg.de |
| **Disorders of sleep regulation: basic mechanisms and therapeutic perspectives (ENOUGH SLEEP)** | Consortium proposes to launch the first integrated effort to study the different aspects of sleep regulation at the genetic, molecular, cellular, and network levels. The proposal makes extensive use of the existing, representative DNA data bank recently collected in Finland and studies the correlations between polymorphisms in genes associated with sleep regulation and the behavioural phenotype assessed by extensive questionnaires and by a health examination. | Tarja Porkka-Heiskanen<br>**E** porkka@helsinki.fi |
| **Genome-wide analyses of European twin and population cohorts to identify genes predisposing to common disease (GenomEUtwin)** | The project has established a combined cohort of 850,000 twins and a cardiovascular diseases cohort (MORGAM) of 160,000 volunteers. There is a common approach to epidemiology and phenotyping, DNA extraction and genotyping, statistical analysis, database construction, ethics and education. The genotyping platform is operational and quality control procedures have been established. | Prof Leena Peltonen<br>**E** eena.peltonen@ktl.fi |
| **Molecular Phenotyping to Accelerate Genomic Epidemiology (MOLPAGE)** | MOLPAGE is a molecular phenotyping study that will lead to significant advances in the diagnosis, prognosis and therapy of diseases of high social relevance and impact on the European healthcare systems. It will develop tools for molecular phenotyping at epidemiologic scale including measuring small molecules (metabonomics), mRNA in white blood cells (transcription profiling), protein and peptide analysis using mass spectrometry methodology, affinity arrays and tissue arrays, DNA methylation patterns and genetics. Bioinformatic tools for data collection, data mining and statistical analysis will also be developed. | Prof John Bell<br>**E** regius@medsci.ox.ac.uk |
| **Genetics of Sepsis in Europe (GenOSept)** | This project uses a multidisciplinary fundamental genomics approach to examine genetic predisposition to sepsis. The GenOSept work packages aim to test the hypothesis that susceptibility to new treatments and fatal outcomes from severe sepsis are, in part, genetically determined. | The European Society of Intensive Care Medicine (ESICM) Brussels<br>www.esicm.org/PAGE_genosept |
| **The Danubian Bank Initiative – Towards Information-based Medicine (DanuBiobank)** | The DanuBiobank network will address the field of non-cancer ageing disorders focusing on diabetes-related endpoints, including vascular disease, metabolic disease and neurodegenerative disorders. Patient cohorts and quality controlled central banks for DNA, serum, plasma and cells/tissues/RNA/proteins will be established. | Gerd Schmitz<br>**E** gerd.schmitz@klinik.uni-regensburg.de |

| TITLE | SUMMARY/TYPE OF COLLECTION | COORDINATOR'S DETAILS |
|---|---|---|
| **European Special Populations Research Network: quantifying and harnessing genetic variation for gene discovery (EUROSPAN)** | The objectives of this study are to: 1) describe the variation in allele frequencies in genetic variants involved in health or disease states; 2) harmonise existing health and disease-related phenotype data and reach agreement on genotyping procedures; 3) adopt standard definitions, measurement methods and common data storage formats to form a common study database; and 4) undertake an extended pedigree based genome-wide linkage analysis to identify quantitative trait loci underlying a number of traits associated with many of the main diseases of public health importance in Europe. | Prof Harry Campbell<br>**E** Harry.Campbell@ed.ac.uk |
| **Harmonising population-based biobanks and cohorts studies to strengthen the foundation of European biomedical science in the post-genome era (POPULATION BIOBANKS)** | Population Biobanks is a coordination action which will be funded by the EU from 2006–2008. This coordination action aims to establish a collaborative research network that will identify and explore key issues that will help to ensure that Europe makes best use of its rich array of population-based biobanks and longitudinal cohort studies. These include major cohorts that already exist and new initiatives that are starting up. If we can ensure that our biobanks are able to work together to address pivotal research questions that fall outside the scope of projects funded by single nations, or even of large cohorts running across several nations, we can ensure that Europe remains at the cutting-edge of biomedical research internationally. | Camilla Stoltenberg<br>**E** camilla.stoltenberg@fhi.no |
| **Harnessing the potential of human population genetics research to improve the quality of health of the EU citizen (EUHEALTHGEN)** | EUHEALTHGEN is a Specific Support Action jointly funded by the European Commission and the Wellcome Trust. Its main aim is to facilitate the development of a forward-looking strategy for translating the outputs of population genetics research into clinically useful and health enhancing initiatives, whilst improving EU industrial competitiveness in this area. It anticipates the creation of research collaborations that are able to respond creatively to the objectives and content of the genomics health theme in the coming 7th EU Research Framework Programme. By providing networking opportunities, it will also help to strengthen and contribute to the development of the European Research Area. | Dr Alan Doyle<br>**E** a.doyle@wellcome.ac.uk<br>www.wellcome.ac.uk |
| **Co-ordination of genomes research across Europe (COGENE)** | The aims of COGENE (a strategic accompanying measure) were to co-ordinate national genomes initiatives, promote synergies between national genome research programmes in Europe and identify areas for collaboration. A survey of genomes research in Europe was conducted, and two workshops were organised on population databases and pharmacogenetics. Undoubtedly, COGENE increased information exchange between funding agencies and scientists. The workshops and meetings helped to identify topics that could benefit from community-wide co-ordination and this has influenced the formulation of science policy at national and EU level. Important areas for future development include harmonisation of data handling, biobanking and bioinformatics services for high-throughput 'omics' research and systems biology. | Eero Vuorio<br>**E** eero.vuorio@utu.fi<br>www.cogene.net |
| **Human Genomic Research Integration (HUMGERI)** | This project aims to establish an integrated human genomic program in Hungary based on studies and comparison of European human genome projects. It will form a stable bioinformatics background for genomic activities, develop a biobanking system, establish a network of Hungarian SMEs related to human genomics projects, clarify ethical and legal aspects of genomic research and organise a forum for genomic initiatives. | Dr Laszlo Fesus<br>**E** fesus@indi.biochem.dote.hu<br>www.humangenom.hu |
| **Advanced Molecular Tools for Array-based Analyses of Genomes (MOLTOOLS)** | The MOLTOOLS project aims to promote the development and implementation of molecular tools that will facilitate the identification of molecules encoded in the human genome, doing this at the level of the individual. The project brings together leading European research groups in the field of mostly array-based technologies such as genotyping, resequencing, protein arrays and cell arrays. The consortium partners provide a strong knowledge resource for what methods for genomic analyses are available, are being developed and what they can be used for. | Prof Ulf Landegren<br>**E** ulf.landegren@genpat.uu.se<br>www.moltools.org |

# Larger longitudinal studies supported by the Wellcome Trust in the UK and in developing countries

| TITLE | SUMMARY | COORDINATOR'S DETAILS |
|---|---|---|
| **CHILDHOOD COHORTS** | | |
| **The National Child Development Study – the 1958 birth cohort** | Demographic, environmental and social data collected on 17,000 births (random samples from UK population in 50:50 ratio of males and females) since March 1958. Biological measures have been added in the 1970s. Blood (DNA), serum and plasma samples collected. Cell lines available from 10,000 cohort members. | Prof Heather Joshi Prof Chris Power Centre for Longitudinal Studies and Institue of Child Health London Prof David Strachan Department of Epidemiology St George's Hospital Medical School, London |
| **Avon Longitudinal Study on Parents and Children (ALSPAC)** | Information collected on 14,000 children/28,000 parents representative of the UK population since 1991/1992. Clinics held approximately every two years with a 10% subset intensively studied. Blood, urine, teeth, nails and placenta collected. Cell lines available on 8,000 children; parents in progress. | Prof George Davey-Smith Department of Social Medicine University of Bristol |
| **The Twins UK Study** | 10,000 twins randomly sampled from the UK popluation. Initiated in 1999 collecting demographic, environmental and social health data (continuous measures). Blood-DNA and serum samples and lymphocytes for transformation. | Prof Tim Spector Kings College London, and Department of Epidemiology St George's Hospital Medical School |
| **Pelotas Birth Cohorts, Brazil – The health and nutrition transition in two Brazilian birth cohorts (1982 and 1993): impact of socioeconomic, behavioural, healthcare and biological factors on health** | The study aims to measure time trends in child and adolescent health indicators, relating them to changes in socioeconomic, cultural, environmental and healthcare factors. Including all hospital births in the city of Pelotas during the years 1982 and 1993. Clinical demographic, socioeconomic and environmental data obtained. Blood and DNA collected. | Prof Cesar Victoria Dr Barros Prof Santos Prof Barros Prof Menezes Department of Social Medicine Federal University of Pelotas Brazil |
| **Birth to Twenty Cohort, South Africa – Puberty and onset of sexual and lifestyle risk among urban South African youth** | The study aims to investigate the link between inter-generational, developmental and situational risk factors to: a) sexuality and reproduction and b) non-communicable diseases (obesity, hypertension and insulin resistance). Initiated in 1988, it involves 3,273 children with information on socioeconomic status, parental health, behaviour and fertility, maternal stress and social support, childcare, supervision andsocialisation, child health, development, adjustment and educational achievement, growth and blood pressure being collected. Blood and urine samples collected. | Prof Linda Richter Dr K Steyn Human Sciences Research Council South Africa |
| **UK Biobank** | The study aims to investigate the impact of genes and environment on health outcomes. 500,000 adults aged 40–69, representative of the UK population, will be recruited. There will be a questionnaire and biophysical data collected at recruitment with follow-up through NHS patients' records. Blood, urine and peripheral blood lymphocytes collected and stored. | Prof Rory Collins University of Manchester and six regional collaborating centres |

| TITLE | SUMMARY | COORDINATOR'S DETAILS |
|---|---|---|
| **ADULT COHORTS** | | |
| **Proyecto Coyoacan, Mexico – A prospective study of cause-specific mortality in 200,000 adults** | The study aims to analyse the main avoidable causes of chronic diseases in developing countries. The study started in 1998 and will involve 200,000 adult men and women aged >35 years who are a representative sample of adults in Coyoacan, Mexico. Information on height, blood pressure, smoking habits, alcohol consumption and diet will be obtained. Blood samples collected. Every 5 years a sub-sample is resurveyed with repeat interviews, measurement and blood collection. | Prof Richard Peto<br>Prof Rory Collins<br>Dr Tapia<br>Dr Kuri<br>Dr Alegre<br>Clinical Trials Unit<br>Nuffield Department of<br>Clinical Medicine<br>University of Oxford |
| **Urban Health Study, Beirut, Lebanon – Health status, well-being and population change in the Middle East: a research network approach** | The study will interview adolescents regarding health related attitudes and behaviour, family relations, peer groups, school and other influences. A subset may look into intentional and unintentional injuries and child labour. The study will also analyse the prevalence, causes and type of disability (physical and mental) and its impact on daily functioning and quality of life in the elderly. This started in 2005 in one community in Lebanon. There will be follow-up for 3 years at periodic intervals. | Dr Zurayk<br>Dr Khawaja<br>Dr Afifif<br>Dr Chaaya<br>Dr El-Kak<br>Centre for Research in<br>Population and Health<br>Faculty of Health Services<br>American University in Beirut<br>Lebanon |
| **Determinants of cardiovascular disease in Eastern Europe: a multi-centre cohort study** | The study will examine the causes of high rates of cardiovascular disease and other non-communicable diseases using 30,000 male and female adults aged 45–69 years randomly selected from Krakow (Poland), Novosibirsk (Russia) and four cities in the Czech Republic. Baseline information collected will include exposure to risk factors and a physical examination. Follow-up information will be obtained through an annual post-questionnaire for non-fatal enevnts and by monitoring national death registers for causes of mortality. Blood samples collected at baseline. | Prof Sir Michael Marmot<br>Department of Epidemiology<br>and Public Health<br>University College London<br>Dr Bobak<br>Dr Kubinova<br>Dr Malyutina<br>Dr Pajak |
| **Thai health-risk transition: A national cohort study** | The study will collect information on the progression of the Thai health transition. There will be a questionnaire on risk factors for disease given to 100,000 university students with follow-up after 4 years. Subset studies will be done on transition from agriculture setting to modern industrialised setting; risks with emerging affluence; fertility transitions, inter-generation contrasts, health service use, diseases beliefs, etc. | Dr Adrian Sleigh<br>Australian National University<br>Australia<br>Sukothai Thammathirat<br>Open University<br>Thailand |
| **Kanchanaburi DSS: Establishment of a population monitoring and evaluation studies centre in Thailand** | The aim is to study health consequences of population change and the effects of health interventions in households and communities in an area of migration in the Western Central region of Thailand. The main research areas include adolescent reproductive health in migrants, indoor air pollution and mental health. The study started in 2000 and involves 43,000 people in 20,000 households. Household census data, GIS information and satellite maps collected and updated annually. | Dr Yoddumnern-Attig<br>Dr Chamratrithirong<br>Dr Prasartkul<br>Dr Thongthai<br>Institute for Population<br>and Social Studies<br>Mahidol University<br>Thailand |

| TITLE | SUMMARY | COORDINATOR'S DETAILS |
|---|---|---|
| **FAMILY COHORTS** | | |
| **African Centre for Health and Population Studies (DSS and HIV surveillance)** | The study aims to describe the demographic, socio-economic and health impact of a rapidly spreading HIV epidemic in individuals and households in a rural area in South Africa and to assess the effect of intervention studies. It started in 2000 and involves 85,000 people. Demographic, health and socio-economic data will be collected on a bi-annual basis (births, deaths, migrations, family composition, economic circumstances and health). Since June 2003 all resident adults between 15–54 years and sub-sample of 12.5% non-residents are targeted annually for HIV testing. Blood samples. | Dr M-L Newell<br>Dr A Herbst<br>University of KwaZulu-Natal<br>South Africa |
| **The Vertical Transmission Study, South Africa: a cohort study to assess the impact of exclusive breastfeeding on post-natal transmission of HIV in a rural area of South Africa.** | The project will study the effects of infant feeding practices on HIV infection rates at 6 and 22 weeks of age and determine infant survival rate at 24 months according to feeding practices and HIV status. It started in 2001 and involves 3,508 mother-child pairs. The women were enrolled in the antenatal period and followed until all children reached 2 years of age. Blood from mother and child collected. | Prof H J Coovadia<br>Dr Coutsoudis<br>Dr Rollins<br>Dr Bland<br>Dr Newell<br>University of KwaZulu Natal<br>South Africa (linked with Africa Centre for Health & Population Studies DSS) |
| **Agincourt DSS: Investigating and responding to health, population and social transitions in rural South Africa** | The aim is to investigate and respond to changes in the health and population dynamics of rural South Africans. The main interests involve non-communicable diseases, health systems organisation and interventions, impact of health shocks on poor households, strategies to combat violence and injury. The study started in 1992 and involves 69,000 people (11,500 households in 21 villages in 390 km$^2$ northeast of Johannesburg. Data has been collected annually with some redesign in 1996/97 to focus on health and population dynamics. | Prof Tollman<br>Dr Kahn<br>Prof Pettiffor<br>Dr Garenne<br>University of Witwatersrand<br>South Africa |
| **KEMRI Wellcome Trust Programme for Health Research, Kilifi, Kenya** | Malaria and a range of other infectious diseases are being investigated through epidemiological questions in a semi-rural coastal area of Kenya with low HIV prevalence. The study started in 1990 and involves 220,000 people which also includes a logitudinal birth cohort. Blood and saliva collected. | Prof Marsh<br>Dr Peshu<br>Prof Snow<br>Dr Newton<br>University of Oxford/Kenya Medical Research Institute<br>Kenya |
| **Karonga DSS: Epidemiology of Mycobacterial, HIV and Helminth Infections in Northern Malawi** | The project aims to study the impact of HIV/AIDS in rural Africa. It was originally established to study leprosy. It is a multidisciplinary programme with a major focus on the immunology and genetics of tuberculosis and HIV. The project started in 1979 and involves the total population of Karonga district (approx 40,000 people). Two complete censuses were carried out in the 1980s. A new continuous registration system (CRS) has been established along with a database of all HIV infected individuals and a database of all tuberculosis cases reported in the district. Blood, stool, urine and DNA samples collected. | Prof Paul Fine<br>Dr Dockrell<br>Dr Zaba<br>Department of Infectious and Tropical Disease<br>London School of Hygiene and Tropical Medicine |

The Wellcome Trust is an independent biomedical research-funding charity, established under the will of Sir Henry Wellcome in 1936. It is funded from a private endowment, which is managed with long-term stability and growth in mind.

The Wellcome Trust's mission is to foster and promote research with the aim of improving human and animal health. During 2005–2010, our aims are:

**Advancing knowledge:** To support research to increase understanding of health and disease, and its societal context

**Using knowledge:** To support the development and use of knowledge to create health benefit

**Engaging society:** To engage with society to foster an informed climate within which biomedical research can flourish

**Developing people:** To foster a research community and individual researchers who can contribute to the advancement and use of knowledge

**Facilitating research:** To promote the best conditions for research and the use of knowledge

**Developing our organisation:** To use our resources efficiently and effectively.