ENGAGE

Publishable Executive Summary January 2008 to December 2012

ENGAGE - European Network for Genetic and Genomic Epidemiology FP7-HEALTH-201413

ENGAGE CONCEPT AND OBJECTIVES:

This is a uniquely exciting time in human genetics as rapid advances in genomic technology permit deeper characterisation of the mechanisms underlying many human diseases. One feature of this recent scientific explosion has been the international collaboration fostered by realisation that the power of individual studies is limited. To identify the full range of genetic variation contributing to common disease and to uncover the effects of the complex interactions of genes, environment and lifestyle factors on disease risk, and thereby support translational advances, a more inclusive approach, based on epidemiological principles, is required.

Collectively, members of the ENGAGE consortium have access to an extensive range of well phenotyped and catalogued population cohorts representing >600,000 subjects. Genome wide association data (GWA) are available for >100,000 of these subjects and an early goal of the ENGAGE project has been to bring together these datasets to perform large scale integrated genetic association analyses. Adopting this approach allows the consortium to identify novel disease-susceptibility variants undetectable in individual studies. A key ENGAGE objective is to evaluate the clinical and public health relevance of the novel disease and trait-susceptibility genes identified and to demonstrate that these findings can be used as diagnostic indicators for common diseases helping us to better understand risk factors, disease progression and why people differ in responses to treatment.

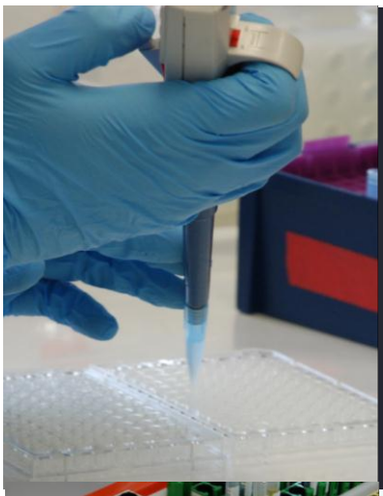
Project Coordinator:

Professor Mark McCarthy

Robert Turner Professor of Diabetes
Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM)
Churchill Hospital, Old Road
Headington, Oxford, OX3 7LJ, UK
Email: contact@euengage.org

ENGAGE was led by the world-leading human geneticist Prof. Leena Peltonen from FIMM, University of Helsinki for the first 26 months, until March 2010 when she sadly passed away after a long illness. The project co-coordinator, Prof. Mark McCarthy from University of Oxford has assumed the leadership as Scientific Coordinator since March 2010. University of Helsinki continues as EC contractual coordinator.

SEVENTH FRAMEWORK PROGRAMME



ENGAGE Structure:

ENGAGE activities are organized through ten work packages

- WP1 Genome Wide Data Integration
- WP2 Novel sources of Genome-wide Variation
- WP3 Novel Phenotypes
- WP4 Informatics and Bioinformatics
- WP5 Genetic Refinement of Identified Loci
- WP6 Epidemiology and Joint Effects
- WP7 Clinical Translation
- WP8 Societal Aspects
- WP9 Training and Dissemination
- WP10 Coordination

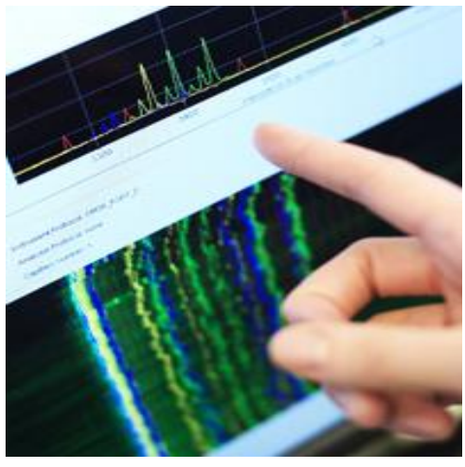
All work packages have been operational during the past 48 months of the project period with major progress around scientific activities supporting the sharing of data for large scale integration studies and the identification and characterisation of disease susceptibility variants and mechanisms through the meta-analysis of ENGAGE datasets.

ENGAGE OVERALL OBJECTIVES:

- To develop an enhanced supranational framework for research into genetic and genomic epidemiology that assembles the best researchers, the best sample and data sets in areas of primary focus (cardiovascular, metabolic, behavioural), the best ethical guidance and the best analytical and translational platforms;
- To accelerate discovery of disease-susceptibility genes through integrated analyses using multiple large-scale data sets and a range of experimental designs, thereby identifying novel aetiological pathways (with potential for pharmaceutical exploitation) and novel susceptibility variants and biomarkers (with potential as diagnostics as well as in guiding therapy development);
- To translate these findings into the clinical arena;
- To explore key methodological questions relevant to European research in this area;
- To develop novel technological and statistical approaches for the study of human disease;
- To disseminate research outputs to both the scientific and non-specialist audience
- To contribute to international efforts in large population cohorts as exemplified by our very close contacts with the P3G effort (Public Population Projects in Genomics).

ENGAGE is extending our integrated genetic analyses to encompass additional sources of genome variation as methods improve for the large-scale collection and analysis of these data types (copy number variation, rare variants etc), and to additional phenotypes as such datasets become available from ENGAGE partners. We are also exploring key methodological questions relevant to European research in genetic and genomic epidemiology and developing novel statistical approaches for data analysis.

Key to the success of the consortium in risk marker identification and clinical translation are the ENGAGE objectives for data sharing and harmonisation. We have developed new computational approaches supporting data sharing and the harmonization of cohort phenotypes whilst establishing protocols for managing the ethical aspects of sample and data sharing according to informed consent, local ethical approval and the governance structures of each ENGAGE partner.



ENGAGE ACTIVITIES AND MAIN RESULTS

DATA SHARING, HARMONISATION AND INTEGRATION: During the initial phase of the project, the WP1 and WP4 teams worked to identify the data submission and exchange requirements needed to support large scale integrated analyses within ENGAGE. The data submission system developed (SIMBioMS) enabled ENGAGE partners to share standardised data sets, in line with the data access policy and consent oversights established by WP8: currently, over 1200 datasets have been uploaded to support ENGAGE projects. SIMBioMS also facilitates data export to public data archives (e.g. EGA, ArrayExpress, PRIDE). Phenotype data for ENGAGE cohorts has been generated using a wide range of cohort-specific questionnaires, clinical protocols, and technology platforms. A strategic collaboration between ENGAGE and the P3G Consortium has supported data harmonization through mapping cohort-specific parameters onto controlled vocabularies: a web-based repository for these data (SAIL) developed by WP4 has been used in ENGAGE and related EU projects (e.g. SUMMIT). These data harmonization efforts are closely integrated with ongoing ESFRI activities (BBMRI, ELIXIR) to ensure the compatibility with European and global initiatives in this area. ENGAGE has further initiated efforts to map and document its data sharing experiences which can benefit other similar consortia at large-scale as well as funding agencies when considering data sharing principles. A manuscript based on consortium-wide

survey results is in preparation and will be submitted for publication.

DISCOVERY: ENGAGE has played a leading role (sometimes alone, often as part of wider consortia) in genome-wide association meta-analyses which have identified many hundreds of genetic loci influencing dozens of medically-significant traits, ranging from type 2 diabetes and obesity, to smoking behaviour and birthweight. These discoveries have often provided vital clues to the mechanisms influencing these phenotypes, catalyzing early steps towards novel therapeutic and preventative options. The maturity and low experimental cost of GWA arrays has meant that these datasets were the first to be widely available across ENGAGE cohorts. The consortium moved effectively to synthesize such data, and has been using similar approaches to mine additional sources of genomic variation (rare variants and copy number variants for example) as well as novel molecular 'omic' phenotypes (such as transcriptomic, epigenomic and metabolomic data). Several of these efforts have been highlighted as "flagship" projects which were pushed forward and finalized in period 4. The results emerging from these efforts have shown multiple novel hits for sequence variants (both common and rare) affecting the phenotypes studied. More than eight manuscripts have been submitted and/or in preparation for publication.

BIOLOGY: The first step in translation of these genetic discoveries is to define the molecular mechanisms through which they impact disease. ENGAGE WPs 5 and 6 have led consortium efforts to develop strategies for refining both the genetic and phenotypic basis of these associations. The first of these has involved deployment of fine-mapping, resequencing and imputation approaches, the objective being to track the specific causal alleles, a challenging task given the strong correlations that exist between nearby variants. The phenotypic efforts has focused on exploration of the wider biological consequences of associated variants, and on epidemiological studies to define the ways in which genetic variants interact with each other and with environmental exposures. These efforts have become, at least in part, focused around their respective resequencing and epidemiology "flagship" projects. For example, the epidemiology flagship has been exploring the wider phenotypic effects of a subset of SNP markers in data from ~200,000 ENGAGE samples. One of the studies in this flagship has demonstrated, through a robust Mendelian randomization analysis, a causal relationship between obesity and multiple cardio-metabolic including heart failure (Fall T et al, The Role of Adiposity in Cardiometabolic Traits: A Mendelian Randomization Analysis, PLoS Med, under review).

TRANSLATION: A key long-term objective has been to move ENGAGE findings towards clinical translation, and WP7 has led efforts to use ENGAGE findings to support stratified medicine. These efforts take several different forms including disease stratification, identification of non-genetic biomarkers (such as hs-CRP as a diagnostic marker for diabetes subtypes), improved prognostication (e.g. diabetes complication risk) and pharmacogenetics. In the final reporting period, further progress has also been made towards both biomarker discovery (e.g. mannose as a potential biomarker for abnormal glucose metabolism); and biomarker implementation in clinical diagnostics (such as a set of biomarkers to predict cardiovascular outcomes). We are also extending earlier ENGAGE work on the genetics of smoking and nicotine dependence by conducting a GWAS meta-analysis of cotinine levels. Cotinine is a biomarker of nicotine intake and metabolism, the genetics of which is only partly known). Since we expect that it will take some years for the full clinical impact of ENGAGE discoveries to reach fruition, we have increased our interactions with related EU-efforts (e.g. the SUMMIT, and DIRECT IMI projects) to support programs that will outlive ENGAGE.

TRAINING AND DISSEMINATION: WP9 organised workshops and training courses open to both ENGAGE and external participants including a series of joint P3G/ENGAGE/WT Summer Institutes, International Biobank Summit, EBI RNA-Seq courses and workshops focusing on ENGAGE Flagship projects. In total the ENGAGE exchange and mobility program supported a total number of 22 exchange visits and meeting/course attending. ENGAGE partners have published more than 250 manuscripts relating to project funded activities during the past 60 months, many of these in high profile journals that have attracted wider attention. The project website has also contributed to dissemination activities.

SELECTED PUBLICATIONS
(from a total of >250)

- Prokopenko *et al.*, Variants in *MTNR1B* influence fasting glucose levels. Nature Genetics (2009; epub 2008)
- Aulchenko, Ripatti *et al.*, Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nature Genetics (2009; epub 2008)
- Thorgeirsson *et al*., Sequence variants at *CHRNB3-CHRNA6* and *CYP2A6* affect smoking behavior and the risk of lung cancer. Nature Genetics (2010)
- Teslovich *et al*., Biological, clinical and population relevance of 95 loci for blood lipids. Nature (2010)
- Suhre *et al*., Human metabolic individuality in biomedical and pharmaceutical research. Nature (2011)
- Schumann *et al.*, Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (*AUTS2*) in the regulation of alcohol consumption. PNAS (2011)
- Surakka, Isaacs *et al*., A Genome-Wide Screen for Interactions Reveals a New Locus on 4p15 Modifying the Effect of Waist-to-Hip Ratio on Total Cholesterol. PLoS Genetics (2011)
- Budin-Ljosne I *et al*., Bridging consent: From toll bridges to lift bridges? BMC Med Genomics (2011)
- Thanabalasingham G et al, A large multi-centre European study validates high-sensitivity C-reactive protein (hsCRP) as a clinical biomarker for the diagnosis of diabetes subtypes. Diabetologia (2011)
- Deloukas P et al, Large-scale association analysis identifies new risk loci for coronary artery disease. Nature Genetics. (2012)
- de Moor MH et al, Meta-analysis of genome-wide association studies for personality. Mol Psychiatry (2012)
- Morris AP et al, Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature Genetics (2012)
- Yang J et al, FTO genotype is associated with phenotypic variability of body mass index. Nature (2012)
- Horikoshi et al, New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. Nature Genetics (2013)
- Codd V, Nelson C, Albrecht E et al. Identification of seven loci affecting mean telomere length and their association with disease. Nature Genetics (2013)

EXPECTED FINAL RESULTS AND POTENTIAL IMPACTS

DISCOVERY: ENGAGE has played a leading role in the integration of genetic data from diverse European data sets, and has catalyzed wider global efforts for several major traits. This leadership is already manifested in the publications arising from the project and in the high international profile of the consortium. This research has mostly focused on genome-wide association data because of its availability and relative ease of integration, and has been supported by considerable "behind-the-scenes" activity with respect to informatics, data access, trait harmonization, statistical methodologies and ethical compliance. ENGAGE is continuing to use this infrastructure to power further rounds of discovery beyond ENGAGE, that encompass a wider range of medical phenotypes (including a suite of behavioural and psychiatric traits), genomic traits (telomere length, metabolomics) and genetic variation (rarer variants, copy number variants). Many of these e fforts are focused around flagship projects which primary data have been generated during the final project years and further analyses will be continued under the collaboration framework built within ENGAGE. ENGAGE has worked with EGA to ensure the long-term availability of ENGAGE-generated data sets after 2012.

BIOLOGY: Ongoing efforts using next-generation sequence data (both directly, and indirectly through imputation) to extend the range of genetic variation examined will uncover novel loci but will also help to identify causal alleles (these may be driving the common variant associations discovered by GWAS or represent independent signals). These alleles (especially where coding) are catalyzing efforts to characterise biological mechanisms conferring risk and protection that will continue beyond the lifetime of ENGAGE.

TRANSLATION: Clinical translation represents the ultimate objective of human genetic discovery, but will require many years to play out. ENGAGE has contributed to some modest successes in this area (for example the demonstration that hsCRP is a useful diagnostic biomarker for HNF1A-MODY) and will continue to support the various efforts towards stratified medicine described earlier. To ensure the continuation of these efforts post-ENGAGE, we are interacting closely with related EU-efforts (e.g. SUMMIT, DIRECT) which have shared goals and some overlapping membership.

OVERALL: The primary aim of ENGAGE has been to generate fundamental research data related to human genetics, and then to take the first steps towards definition of the biological mechanisms through which they act. These discoveries will help to pave the way towards advances in clinical practice, including developments in "personalized medicine" and the identification of new pharmacological targets. During the past five years, ENGAGE has focused on dissemination of research results to the scientific community, for example, ENGAGE partners have published over 250 scientific manuscripts (many of them with open access), making the data and knowledge generated accessible to the scientific community. This has helped to maintain the competitiveness of European research excellence in the field of human genetic and genomic epidemiological research. ENGAGE has demonstrated a high level of integration and dynamic coordination and communication within a relatively large collaborative research consortium. The experience and knowledge obtained during the course of ENGAGE have been hugely beneficial for the field in general, and have also contributed to the training and development of a cadre of junior researchers with the skills and scientific temperament to support future projects of this kind. ENGAGE is making best efforts to document and disseminate such experience and provide recommendations for other consortia and funding decision bodies.

ENGAGE ON THE WEB:

Follow project news and learn about ENGAGE-sponsored training events and publications and progress in research activities at the project website:

www.euengage.org

ALSO ON YOUTUBE:

http://www.youtube.com/watch?v=YUlcSPkLNB8&feature=youtu.be

FUNDING AND PARTICIPANTS

The ENGAGE project is funded with €12 million through the EU 7th Framework Programme and is coordinated from the University of Helsinki. The ENGAGE consortium is comprised of 25 partners, from Europe, Canada and Australia, including 23 from universities and research institutes and 2 commercial partners. More information about the key scientists involved in the project at each partner site can be found on the project website.

ENGAGE PARTNERS

1. University of Helsinki, Institute for Molecular Medicine Finland (FIMM), Finland
2. University of Oxford, The United Kingdom
3. European Bioinformatics Institute (EMBL-EBI), The United Kingdom / European Molecular Biology Laboratory, Germany
4. Queen's University Belfast, The United Kingdom
5. King's College London, The United Kingdom
6. Illumina, Cambridge LTD, The United Kingdom
7. Leiden University Medical Centre Centre for Medical Systems Biology (CMSB) / Vrije Universiteit Amsterdam, The Netherlands
8. Erasmus Medical Center, The Netherlands
9. German Research Center for Environmental Health (former GSF), Germany
10. Karolinska Institutet, Sweden
11. Lund University, Sweden
12. Uppsala University, Sweden
13. Royal Institute of Technology, Sweden
14. Norwegian Institute of Public Health, Norway
15. University of Tartu Estonia Genome Project (EGP), Estonia
16. deCODE genetics, Iceland
17. Ontario Institute for Cancer Research, Canada
18. Université de Montréal, Canada (2008-July2009)
19. Centre for Genomic Regulation, Spain
20. University of Leicester, The United Kingdom
21. University Lübeck/University Medical Center Schleswig-Holstein, Germany
22. Imperial College London, The United Kingdom
23. Queensland Institute of Medical Research, Australia
24. The Wellcome Trust Sanger Institute, The United Kingdom
25. Institute of Mathematics and Computer Science, University of Latvia, Latvia
26. McGill University, Canada (since August 2009)