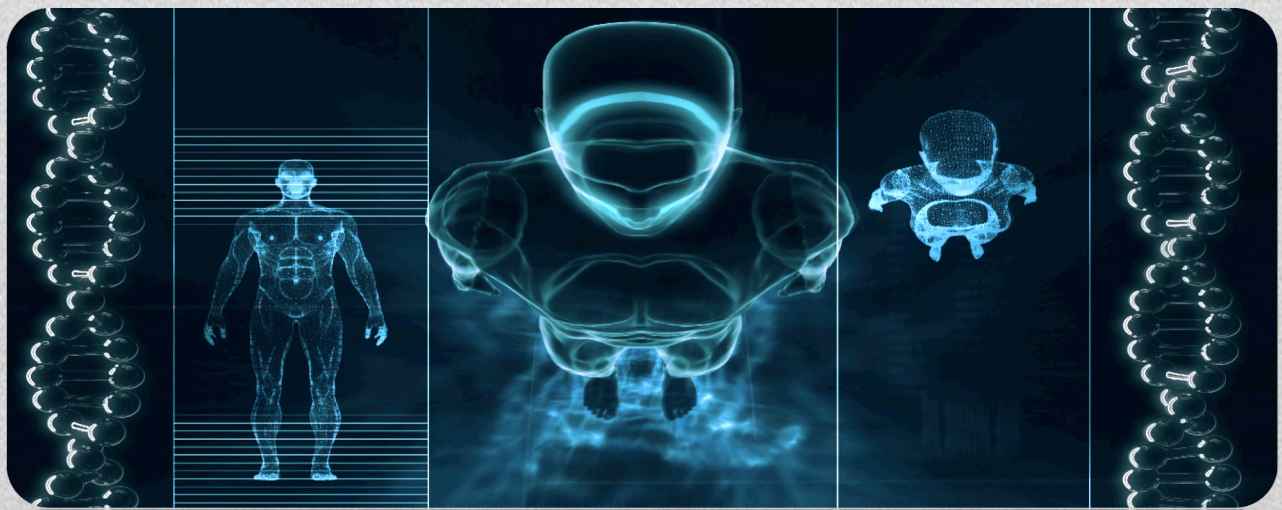




ENGAGE

European Network for Genetic and Genomic Epidemiology

FINAL PUBLISHABLE SUMMARY REPORT





The ENGAGE Consortium was supported through funds from the European Commission Seventh Framework Programme; FP7-Health-2007-A-2.1.1-2; Grant Agreement no.: 201413; Project duration: 1/1/2008-31/12/2012

TABLE OF CONTENTS

1.1 Executive Summary	3
1.2 Summary Description of Project Context and Objectives	4
1.2.1 Project Context	4
1.2.2 Overall Objectives:	5
1.2.3 ENGAGE Partners:	6
1.3 Main S&T Results/Foregrounds	8
1.3.1 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	8
1.3.2 Accelerate discovery of disease-susceptibility genes through integrated analyses using multiple large-scale data sets	14
1.3.3 Translate ENGAGE findings into the clinical arena	15
1.3.4 Explore key methodological questions relevant to European research in this area and develop novel technological and statistical approaches for the study of human disease	18
1.3.5 Contribute to international efforts in large population cohorts as exemplified by our very close contacts with the P3G effort (Public Population Projects in Genomics)	20
1.4 Potential Impact and Main Dissemination Activities and Exploitation of Results	21
1.4.1 Potential Impact	21
1.4.2 Main dissemination activities and exploitation of results	23
1.5 Address of Project Public Website and Contact Details	29

1.1 Executive Summary

The past decade was a remarkable era for human genetics and genomics. After the completion of the Human Genomic Project, tremendous efforts have been invested to translate the new information provided by human genome into knowledge that can benefit human health. With rapid development of technology, human genetic researchers are able to characterize the mechanisms underlying many diseases to a profound extent that was not possible previously. One feature of this recent scientific explosion has been the international collaboration fostered by realisation that the power of individual studies is limited.

The European Network for Genetic and Genomic Epidemiology (ENGAGE) aimed at translating the wealth of data emerging from large-scale research in genetic and genomic epidemiology from European (and other) population cohorts into information relevant to future clinical applications. The concept of ENGAGE was to enable European researchers to identify large numbers of novel susceptibility genes that influence metabolic, behavioural and cardiovascular traits, and to study the interactions between genes and life style factors.

Collectively, members of the ENGAGE consortium have access to an extensive range of well phenotyped and catalogued European population cohorts representing more than 600,000 subjects. In addition to clinical phenotypes, many of these subjects have provided biosamples that can be used for genetic and other genomic studies; over 100,000 of these subjects have had genome-wide association (GWA) analyses completed.

The collaboration fostered through the award established an international framework that has allowed us to:

- bring together expertise to develop research and analytical strategies;
- develop new computational tools to support data sharing and the harmonisation of cohort phenotypes;
- integrate and analyse large-scale datasets with robust statistical tools;
- uncover the effects of the complex interactions of genes, environment and lifestyle factors on disease risk;
- support and explore potential in translating early findings to clinical applications

Through inter-consortium initiatives and through collaboration with international consortia, ENGAGE partners have contributed to many of the largest genetic studies yet performed. These have used genome wide association approaches applied to many hundreds of thousands of individuals and have identified many hundreds of genetic loci influencing dozens of medically-significant traits, ranging from type 2 diabetes and obesity, to smoking behaviour and birth weight. These genetic discoveries have provided crucial information to accelerate definition of the biological mechanisms through which disease comes about; in turn this will further catalyse early steps towards better ways of treating, monitoring, predicting and preventing disease.

ENGAGE has made best effort to document and disseminate such experience and provide recommendations for other consortia and funding decision bodies. ENGAGE partners have published over 250 scientific manuscripts (many of them with open access and in high impact journals), 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. 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.

1.2 Summary Description of Project Context and Objectives

1.2.1 Project Context

The past decade was a uniquely exciting time in human genetics as rapid advances in genomic technology enabled deeper characterisation of the mechanisms underlying many human diseases. One feature of this rapid scientific development 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, was 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 (GWA) data are available for >100,000 of these subjects and an early goal of the ENGAGE project was to bring together these datasets to perform large scale integrated genetic association analyses. Adopting this approach has allowed the consortium to rapidly identify novel disease-susceptibility variants undetectable in individual studies. A key ENGAGE objective was 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.

ENGAGE activities were organized through ten work packages with WP1, 2 and 3 to encompass initial discovery studies using available GWA data, new genotypic and new phenotypic data, with WP5 and 6 to explore key biological questions through refinement and epidemiological approaches, with WP7 to attempt to translate key finding to clinical arena, and with WP4, 8, 9 and 10 to provide a comprehensive IT/bioinformatics, ethical, training and management framework to facilitate these activities.

List of ENGAGE work packages and leaders:

- WP1 Genome Wide Data Integration (Mark McCarthy)
- WP2 Novel sources of Genome-wide Variation (Xavier Estivill)
- WP3 Novel Phenotypes (Gert-Jan van Ommen)
- WP4 Informatics and Bioinformatics (Alvis Brazma)
- WP5 Genetic Refinement of Identified Loci (Kari Stefansson/Unnur Thorsteinsdottir)
- WP6 Epidemiology and Joint Effects (Nancy Pederson)
- WP7 Clinical Translation (Leif Groop)
- WP8 Societal Aspects (Jennifer Harris)
- WP9 Training and Dissemination (Jaakko Kaprio)
- WP10 Coordination (Leena Peltonen/Mark McCarthy)

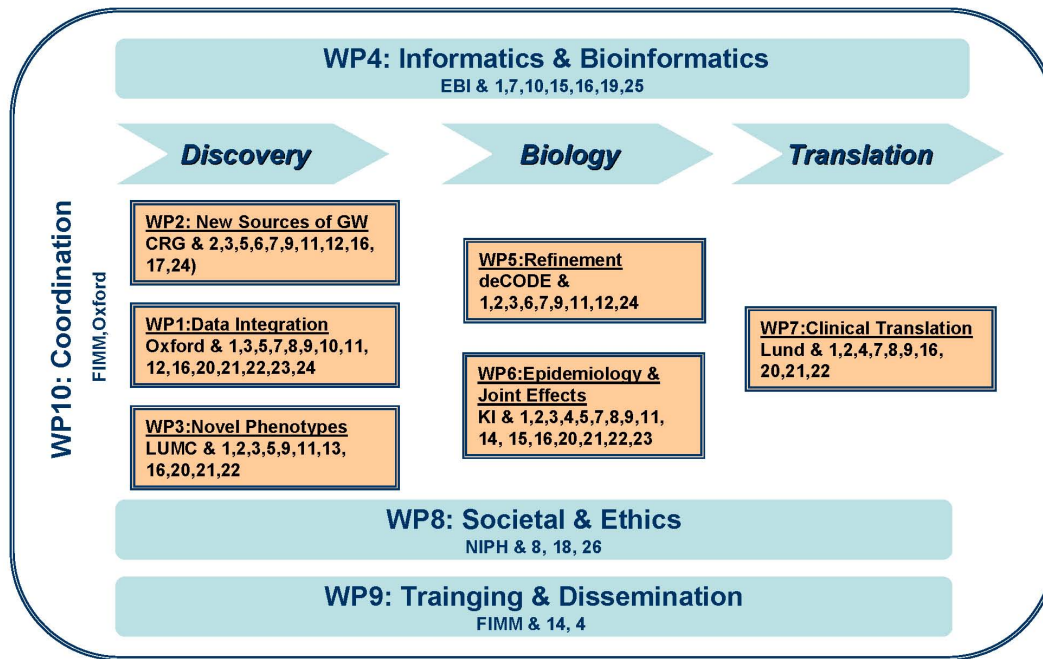


Figure 1: ENGAGE Structure. From discovery, biology to translation, and key participants in each work package

1.2.2 Overall Objectives

The overall objectives of ENGAGE are:

- 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 (including for example, the consequences of ethnic and environmental heterogeneity for gene discovery efforts and the allelic architecture of common disease);
- 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)

1.2.3 ENGAGE Partners

ENGAGE has brought together 24 leading research organizations and two biotechnology and pharmaceutical companies across Europe and in Canada and Australia. The project 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 assumed the leadership as Scientific Coordinator since March 2010.



ENGAGE Final Consortium Meeting, 15-16 November 2012, Rotterdam

Table 1: List of ENGAGE partners, respective countries and key personnel

No	ENGAGE Partners	Country	Key Personnel
1	University of Helsinki, Institute for Molecular Medicine Finland (FIMM)	Finland	Leena Peltonen (1952-2010), Coordinator Jaakko Kaprio, WP9 leader Samuli Ripatti Markus Perola
2	University of Oxford	United Kingdom	Mark McCarthy, Coordinator, WP1 leader
3	European Bioinformatics Institute (EBI)/ European Molecular Biology Laboratory (EMBL)	United Kingdom/ Germany	Alvis Brazma, WP4 leader
4	Queen's University Belfast (MORGAM studies)	United Kingdom	Alun Evans Frank Kee
5	King's College London	United Kingdom	Tim Spector
6	Illumina, Cambridge LTD	United Kingdom	David Bentley Mark Ross
7	Leiden University Medical Centre Centre for Medical Systems Biology (CMSB) / Vrije Universiteit Amsterdam	The Netherlands	Gert-Jan van Ommen, WP3 leader Peter 't Hoen Dorret Boomsma Brenda Pennix
8	Erasmus Medical Centre	The Netherlands	Albert Hoffman Cornelia van Duijn Cecile Janssens
9	German Research Centre for Environmental Health-Helmholtz Zentrum München	Germany	Erich Wichmann Konstantin Strauch Thomas Meitinger Thomas Illig Christian Geiger
10	Karolinska Institutet	Sweden	Nancy Pedersen, WP6 leader Jan-Eric Litton Erik Ingelsson
11	Lund University	Sweden	Leif Groop, WP7 leader
12	Uppsala University	Sweden	Ann-Christine Syvänen
13	Royal Institute of Technology	Sweden	Mathias Uhlen Cristina Al-Khalili Szigyarto
14	Norwegian Institute of Public Health	Norway	Jennifer Harris, WP8 leader
15	University of Tartu Estonia Genome Project (EGP)	Estonia	Andres Metspalu
16	deCODE genetics	Iceland	Kari Stefansson Unnur Thorsteindottir, WP5 leader
17	Ontario Institute for Cancer Research	Canada	Thomas Hudson
18	Université de Montréal Canada (2008-July2009)	Canada	Bartha Maria Knoppers
19	Centre for Genomic Regulation	Spain	Xavier Estivill, WP2 leader
20	University of Leicester	United Kingdom	Nilesh Samani
21	University Lübeck/University Medical Centre Schleswig-Holstein	Germany	Heribert Schunkert Jeanette Erdmann
22	Imperial College London	United Kingdom	Marjo-Riitta Jarvelin
23	Queensland Institute of Medical Research	Australia	Nick Martin; John Whitfield
24	The Wellcome Trust Sanger Institute	United Kingdom	Aarno Palotie Leena Palotie (1952-2010)
25	Institute of Mathematics and Computer Science, University of Latvia	Latvia	Juris Viskna
26	McGill University (August2009 >)	Canada	Bartha Maria Knoppers

1.3 Main S&T Results/Foregrounds

Through its high level of synergy and integration ENGAGE has successfully achieved its primary aim to translate the wealth of data emerging from large-scale genetic and epidemiological studies from well-characterized population cohorts into information that is relevant to future advances in clinical practice.

ENGAGE has extended its joint 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. ENGAGE has also explored key methodological questions relevant to European research in genetic and genomic epidemiology and developing novel statistical approaches for data analysis.

During its lifespan ENGAGE partners have actively engaged and performed research activities and further efficiently made their research outputs available to the scientific community, for example through publications in scientific journals (over 250 peer-reviewed scientific manuscripts published, many of them with open access and in high impact journals) and through presentations in conferences. This has contributed to promoting visibility and maintaining the competitiveness of European research excellence in the field of human genetic and genomic epidemiological research. 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 has made best efforts to document and disseminate such experience and provide recommendations for other consortia and funding decision bodies.

1.3.1 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

ENGAGE has established a dynamic framework across country borders that assembled world-renown expertise from the related fields and a constellation of genetic, genomic and phenotypic data and samples from an extensive range of populations. Key to the success of ENGAGE in risk marker identification and clinical translation are the ENGAGE objectives for data sharing and harmonisation. ENGAGE has 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. A set of flagship projects was launched to enhance synergy and collaboration across related core activities. ENGAGE also recognised the importance of next generation experts for this rapidly developing field of genetic research. Through a productive ENGAGE Training Policy and Exchange and Mobility Programme, ENGAGE junior scientists were offered excellent opportunities to visit partner institutions or outside ENGAGE to enhance their skills and knowledge using results and developments emanating from the ENGAGE project and beyond.

ENGAGE samples and data

ENGAGE partners have access to up to 600,000 subjects in cohorts with various formats (including healthy, disease-based, twins/family-based, population isolates, prospective/cross-sectional cohorts) and at least 100,000 of these subjects with genome wide association (GWA) data. To fully capitalize the resources available from these cohorts and foster epidemiological studies (e.g. GxG, GxE interactions), an extensive epidemiology survey was conducted within the scope of the 'Large-Scale Epidemiology Flagship Project'. In total, 38 ENGAGE cohorts representing over 250,000 subjects were included in this study and 150,000 samples were genotyped on a selected set of SNPs that were of ENGAGE main research interests.

Furthermore, also motivated largely through the flagship projects, new large-scale data sets have been generated in the interests of improving power and extending applicability to new research approaches. For example, metabolomic profiling covering 163 metabolic traits involving >7000 samples from 7 ENGAGE cohorts, produced through the 'Metabolomics Flagship Project' aims to identify new associations for common variants with weaker effect size and for rarer variants with moderate effective size. Telomere length (TL), a heritable trait associated with several diseases (e.g. CAD, cancer) and disease risk factors (e.g. oxidative stress, inflammation), has been measured for over 38,000 samples. The TL data together with GWA data were meta-analysed to identify genetic loci that affect telomere length, and to assess the relationship of TL with age-related diseases and other phenotypes. Genome-wide DNA methylation data has been generated for highly-informative data on trait-discordant MZ pairs.

Outreach to other international consortia

ENGAGE as a project and partners therein have deeply networked with several projects relevant to the area of study. Examples include the following:

Table 2: List of diseases/traits that ENGAGE partners have contributed to identification of new disease-susceptibility genes (most of the results were published in scientific peer-review journals)

Consortium	Relationship with ENGAGE
<i>A) Meta-analysis studies</i>	
HaemGen	<i>ENGAGE worked in collaboration with the HaemGen consortium in a successful meta-analysis of haematological phenotypes.</i>
CHARGE	<i>ENGAGE has worked with CHARGE on a number of meta-analysis studies including haematological phenotypes and Vitamin D analysis.</i>
TAG and OX-GSK	<i>These consortia shared data for the SNPs evaluated by ENGAGE in the smoking meta analysis study published in Thorgeirsson et al., Nat. Genet. 2010;42:448-453.</i>
Sunlight	<i>ENGAGE worked in collaboration with the Sunlight consortium on a successful meta analysis of vitamin D measurements.</i>
MAGIC	<i>ENGAGE has a longstanding relationship with the MAGIC consortium to jointly evaluate genetic traits in glucose metabolism.</i>
GIANT	<i>ENGAGE has a longstanding relationship with the GIANT consortium to jointly evaluate genetic traits in anthropomorphic phenotypes</i>
SpiroMeta	<i>Several ENGAGE partners have contributed data to this meta-analysis of lung function (Repapi et al, Nat Genet 2010), which is now continuing with further analyses with the CHARGE consortium (Soler Artigas et al, Nat Genet 2011) and others.</i>
CGASP	<i>CGASP is a NIDA coordinated consortium aiming to unravel the complex genetics of the nicotinic receptor gene alpha5/alpha3/beta4 complex associated with nicotine dependence, lung cancer and COPD. Again, several ENGAGE partners are involved</i>
CARDIoGRAM	<i>ENGAGE worked in collaboration with the CARDIoGRAM consortium in a number of successful meta-analyses studies for cardiovascular diseases</i>
Tobacco and Genetics Consortium	<i>ENGAGE worked in collaboration with the Tobacco and Genetics Consortium on a successful meta analysis of smoking behaviour measurements</i>
EGG	<i>ENGAGE has worked with the EGG consortium on a number of meta-analysis studies on phenotypes relevant to early growth.</i> <i>EGG has utilized ENGAGE database solution to facilitate data exchange for joint analysis</i>
EAGLE	<i>ENGAGE worked in collaboration with the EAGLE consortium in a successful meta-analysis of atopic dermatitis (Paternoster et al., Nat Genet 2011)</i>
DIAGRAM	<i>ENGAGE has worked with the DIAGRAM consortium on a number of meta-analysis studies for type 2 diabetes genetics</i>
Global Lipids Genetic Consortium	<i>ENGAGE has worked in collaboration with the Global Lipids Genetic Consortium on a number of meta-analysis studies relevant to lipid traits</i>

CREAM	<i>ENGAGE has worked with the CREAM consortium on a number of meta-analysis studies on refractive error and myopia</i>
EMADS	<i>Exome Chip Meta-Analysis of Drinking and Smoking has several ENGAGE partners involved in exomic analyses of alcohol and nicotine use. The consortium was initiated in 2012 and first analyses are expected in early 2013</i>
B) Data Handling and Harmonisation	
EGA	<i>Working closely with WP4 to ensure that ENGAGE data handling and harmonization approaches are compatible with the EGA efforts</i>
P3G	<i>ENGAGE is working with the Public Population Project in Genomics (P3G) around their DataShaper tool, unifying the efforts of ENGAGE and P3G to develop more streamlined and standardised procedures for harmonization efforts.</i>
BioSHaRE	<i>Working closely with WP8 to explore opportunity on establishing scientific (e.g. phenotype harmonization, biosample handling), legal, ethical and political interface between BioSHaRE-EU and ENGAGE and other leading biobank consortia.</i>
C) New Phenotypes	
MolPAGE	<i>MolPAGE has provided the WP3 team with early access to multi-omic datasets as training sets for new methods development.</i>
D) Clinical Translation	
SUMMIT	<i>The work on identification of novel variants for diabetic complications will be carried forward in close collaboration with the IMI project SUMMIT.</i>
CEED3	<i>Work on genes and biomarkers influencing beta-cell function in ENGAGE is further developed for diagnostic purposes within the EU project CEED3.</i>
DIRECT	<i>ENGAGE has collaborated with DIRECT for the work on exploring potential biomarkers to predict response to treatment for diabetes. The work will be continued post ENGAGE.</i>
BiomarCaRE	<i>The MORGAM biomarker project will be continued in this EU-funded project to assess the predictive value of existing and emerging biomarkers for cardiovascular risks and phenotypes</i>
E) Technologies	
GEUVADIS	<i>ENGAGE co-organized two workshops on RNA-seq courses with GEUVADIS and EBI.</i>

Database and computational infrastructure for data sharing

To support large scale integrated analyses within ENGAGE, during the initial phase of the project, the GWAS integration and database teams worked to identify the data submission and exchange requirements needed. 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. 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 the database team has been used in ENGAGE and partner

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 made 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 (manuscript submitted).

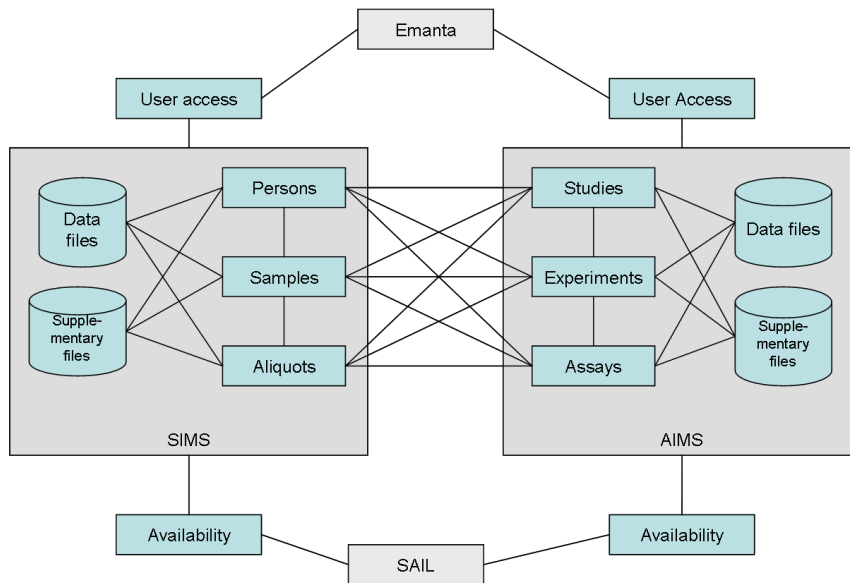


Figure 2: Illustration of relationship between the SIMBioMS software components

Ethical framework

The ENGAGE overarching aim has important social implications which have been well addressed within the context of the project. The ENGAGE ELSI (Ethical, Legal and Social Implications) team performed conceptual analyses on areas that were relevant to ENGAGE, for example, on convergence and divergence of ELSI when translating between population-based and clinical-based biobank studies; and on socio-ethical and legal approaches to policy issues in genome research.

ENGAGE has developed ELSI tools to promote inter-operability of large-scale projects in translational molecular epidemiology. More specifically, a policy paper was published to report ENGAGE experience on the mechanisms allowing the retrospective use of data and human biological samples. A confidentiality template specific to ENGAGE research as well as consent template to bridge consent between clinical and non-clinical research environment were formulated and published (i.e. through ‘ENGAGE Principles for Data Sharing Data release and intellectual Property’, and a research paper ‘Bridging Consent: From Toll Bridges to Lift Bridges?’ published in BMC Medical Genomics 2011, respectively). In collaboration with Public Population Project in

Genomic (P3G) and the Centre for Health, Law and Emerging Technologies (HeLEX), an international Data-Sharing Code of Conduct for Genomic Research was developed and published. Recent work has been focused on analysing current recommendations for genotype-driven recruitment research design as genotypic information becomes more commonly available in clinical care and for individuals in direct to consumer testing. The strategies to address the key challenges identified have been reported by the ENGAGE ELSI team.

Training and mobility programme

It is imperative that European research groups obtain 'state of the art' training in modern genetic epidemiology and related fields relevant to population genomics and emerging 'omics'. To ensure that future expertise in this rapidly developing field is well obtained, ENGAGE established a 'Training Policy' and launched an 'Exchange and Mobility Programme' to offer young pre- and postdoctoral researchers training opportunities in areas, such as interdisciplinary genetics and statistical/epidemiological methods and others. The programme also consisted of inter-institute exchanges, thematic workshops and hands-on courses at the participating partner centres. Through the training programme, ENGAGE has supported 22 exchange and mobility visits and organized or co-organized at least 15 training workshops/courses/meetings. These workshops have been very well-received, while the exchanges have been essential to promoting research collaborations between partners.



ENGAGE Young Investigators

(see detailed profiles at <http://www.euengage.org/yi.html>)

1.3.2 Accelerate discovery of disease-susceptibility genes through integrated analyses using multiple large-scale data sets

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 (see Table 3 for a list diseases/traits that ENGAGE partners have contributed). 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 synthesise 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 and new data are being generated.

Table 3: List of diseases/traits that ENGAGE partners have contributed to identification of new disease-susceptibility genes (most of the results were published in scientific peer-review journals)

A) Metabolic/cardiovascular traits (ENGAGE initial focus disease area)

Type 2 diabetes

Continuous glycaemic traits including HbA1c, proinsulin, random glucose, insulin secretion, fasting glucose, fasting insulin, and 2h glucose

Complications of type 1 and type 2 diabetes

Adiponectin

Protein biomarkers including transferrin glycosylation and C-reactive protein (CRP)

Anthropometric traits including BMI, obesity, fat distribution, height, waist, waist-hip ratio, brachial circumference

Lipids

Metabolic syndrome

Coronary artery disease

Blood Pressure

Serum urate level

Metabolomic phenotypes

B) Behavioural and psychiatric phenotypes (ENGAGE initial focus disease area)

Eating disorder-related symptoms

Use and dependence on alcohol, nicotine and caffeine

Temperament

Major depressive disorder

Sensation seeking

Handedness

Education attainment

Personality measures

C) Additional phenotypes beyond ENGAGE initial focus areas

Telomere length

Haematological traits, including blood cell formation and function, leukocyte telomere length, circulating plasminogen activator inhibitor-1, circulating haptoglobin levels, megakaryopoiesis

Liver enzyme

Liver function tests (AST, GGT, ALP, bilirubin)

Kidney function and chronic kidney disease

Lung Function

Smoking behaviour

Smoking and lung cancer

Alcohol intake and alcohol consumption

Early growth phenotypes including birth-weight and early growth, growth and adiposity in early life, childhood and adult height, tooth development during infancy

Pubertal traits including age at menarche, childhood and pubertal growth

Vitamin D levels

Sex hormones

Psoriatic arthritis

Psoriasis

Atopic dermatitis

Eye-related traits including intraocular pressure and glaucoma, myopia

Hippocampal and intracranial volumes

Menopause

Asthma

Migraine

Facial morphology

Self-rated health

Urate and gout

1.3.3 Translate ENGAGE findings into the clinical arena

A key long-term objective of ENGAGE has been to move ENGAGE findings towards clinical translation. These efforts take several different forms including disease stratification, identification of genetic and non-genetic biomarkers, improved prognostication (e.g. diabetes complication risk) and pharmacogenetics.

Through strategies combining discovery, validation and first attempts to translation, the success in identifying novel genetic variants that increase risk of disease have generated important molecular and biological insights, leading to better understanding of disease classifications and pathogenesis.

For example, ENGAGE partners reported a comprehensive description of mechanism by which type 2 diabetes (T2D) associated gene variants including *TCF7L2*, *MTRN1B*, *FTO*, *KCNQ1* cause altered glucose or lipid metabolism. In collaboration with other consortia through large-scale association studies, additional T2D susceptibility genetics variants have been identified. These findings have expanded the understanding of the genetic architecture and molecular basis of T2D, leading to potential risk stratification of T2D that can be used in future clinical practice. Similar efforts involving ENGAGE partners have also been made in other disease areas including cardiovascular diseases, lipids, metabolic syndromes and the complications of diabetes.

By means of a genome-wide scan with a robust statistical design, the ENGAGE lipids team (LIPIDACTION) evaluated the interactions between genes and epidemiological risk factors including lifestyle (e.g. smoking, alcohol consumption) and body composition (e.g. body mass index (BMI), waist-to-hip ratio (WHR)) for circulated serum lipid levels. The study identified a new genetic locus in chromosome 4p15 modifying the effect of WHR on total cholesterol (TC) (Surakka et al 2011). Although additional studies are required to further establish the interaction mechanism, this study might imply potential use for establishing targeted intervention strategies for people characterized by specific genomic profiles.

Within the ENGAGE Large-Scale Epidemiology Flagship Project, a Mendelian randomization approach was applied to the *FTO* locus variant, where the flagship team have confirmed the causal relationship between obesity and multiple cardio-metabolic traits using *FTO* variant as an instrumental variable. In particular, the study has provided novel insights in the causal effect of obesity on heart failure and increased liver enzymes levels. These findings are relevant to public health in relation to the global prevention efforts for obesity and hence diseases associated to obesity such as T2D and heart disease (Fall et al, PLoS Med in revision).

ENGAGE has made progress towards applications in clinical diagnostics. The Oxford team has successfully identified C-reactive protein (CRP) as a biomarker for MODY (maturity-onset diabetes of the young), a monogenic form of diabetes different from both type 1 and type 2 diabetes. Using samples from seven European populations, the Oxford team further validated high-sensitivity CRP (hsCRP) as a clinical biomarker for the diagnosis of diabetes subtypes HNF1A MODY. It is expected that hsCRP can be rapidly applied in clinical practise to improve diagnosis of this diabetes subtype, considering its relatively low cost and wide availability (Owen et al, 2010; Thanabalasingham et al, 2011). In more recent work the same team has identified plasma glycan profiles as an additional biomarker of HNF1A-MODY.

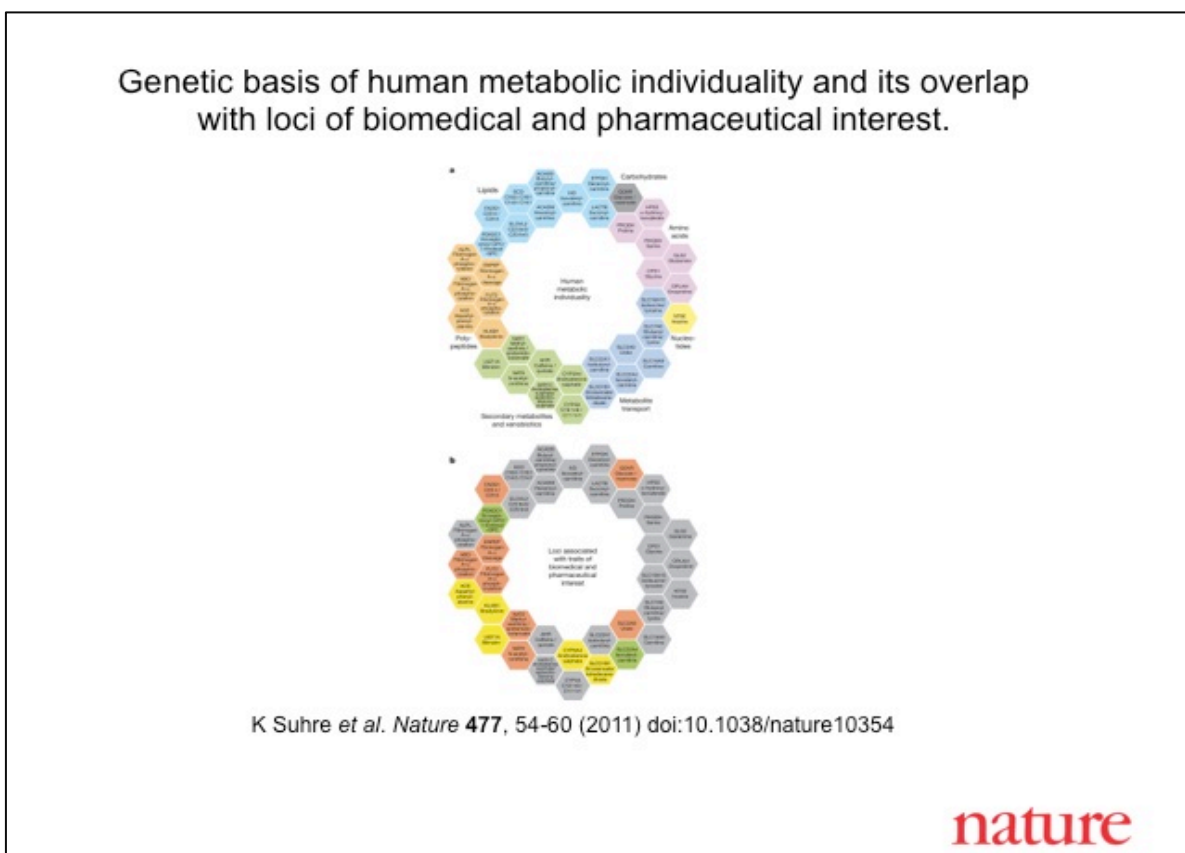
Several ENGAGE studies also contributed to gain knowledge in potential development of clinical diagnostics for prediction and prognostication. These efforts include studies such as identifying biomarkers to predict diabetic complications (through collaboration with IMI-SUMMIT); testing a number of 46 T2D related SNPs for their ability to predict future T2D (through a joint meta-analysis); and evaluating a set of biomarkers (including troponin 1, adrenomedullin and vitamin D) for their ability to predict cardiovascular outcomes (through the MORGAM biomarker project). Also by applying a two-stage population risk screening strategy using perspective population cohorts, the Helsinki team has evaluated genetic risk discrimination and reclassification for coronary heart disease (CHD). The findings showed that a genetic risk score based upon 28 previously reported risk variants for CHD improves risk prediction of CHD and thereby may be used to help identify individuals at high risk for the first CHD event (Tikkanen et al, ATVB accepted).

ENGAGE partners and collaborators have demonstrated the strength of combining genomics and metabolomics information in a global evaluation of genetic variance in human metabolism. The

study successfully identified 37 genetic loci associated with blood metabolites; with majority of the loci showing exceptionally large effect sizes (Suhre et al, 2011). Through association analyses of known disease-risk loci with these metabolites, the findings have provided new insights in biological pathways, offering potential possibilities for diagnosis and treatment. For example, in relevance to diabetes, one of the findings identified a strong association between a variant in the glucokinase regulatory protein (*GCKR*) gene and the metabolites mannose:glucose ratio; and that the association was found stronger between the *GCKR* and mannose, in comparison to *GCKR* and glucose. This may pave the way for further studies on the role of mannose as a potential biomarker for abnormal glucose metabolism.

This study also contributed new information to pharmacogenomics. Six out of the 37 genetic loci identified in this study were previously known to be associated with drug toxicity or adverse reactions. The associations with metabolites and these loci established in this study may provide new biochemical information that can be useful for drug design to avoid adverse reaction.

Although some ENGAGE findings have been initially translated into the clinical arena, it is expected that it will take some years for the full clinical impact of ENGAGE discoveries to reach fruition. We are interacting closely with related EU-efforts (for example SUMMIT, DIRECT, CEED3, BiomarCaRE) to support programmes that will outlive ENGAGE.



1.3.4 Explore key methodological questions relevant to European research in this area and develop novel technological and statistical approaches for the study of human disease

The exploration of key methodological questions and further development of new technological and statistical approaches relevant to genetic and genomic epidemiology has been an intrinsic component of many ENGAGE key initiatives, including the GWAS integration analyses, genetic as well phenotypic refinement activities and flagship projects.

Notably, many of our GWA meta-analyses have involved developing strategies for dealing with structure, quality control, cryptic relatedness, heterogeneity etc. These efforts include (but are not limited to):

- Imputation based on reference data sets (e.g. HapMap and 1000 Genomes) to support combination of data generated on different platforms and to extend analysis to other non-typed variants, using programs such as IMPUT and MACH;
- Data quality control including correction for stratification using genomic control and PCA (principal component analysis) based method;
- Strategies for harmonization of diverse phenotypic transformations and analytical protocols within individual genome-wide scans;
- Strategies for efficient replication on findings and for dealing with heterogeneity

Also through development of the SIMBioMS platform, ENGAGE also developed IT tools to facilitate efficient integration of summary GWAS data (with potential to extend to other data types) and associated meta-data.

To share the established methodological expertise and to ensure that the integration analyses as well as data organization were conducted in a standardised manner within the ENGAGE consortium, a two-day training course for pre- and post-doctoral researchers were organised in February 2010 to provide them hands-on training covering a variety of computational methods for GWA integration studies (including these key methodological topics listed above). In addition, a separate statistical workshop was held to tackle existing and emerging statistical modelling challenges of genome-wide studies with special focus on analysis of rare variants and interactions.

Other key methodological exploration work within ENGAGE is exemplified by the following list:

- Strategies to identify low frequency and rare variants, e.g. for finding missing heritability;

- Development of imputations and phasing tools allowing for analysis of millions of SNPs not directly genotyped for identification of new signals and refinement of signals;
- Strategies to evaluate biology, joint effects and pleiotropy of discovered variants in large datasets (e.g. through LIPIDACTION, Large-Scale Epidemiology Flagship project as well as through collaborations with other consortia);
- Exploration and development of the potential of NGS (next generation sequencing) technologies for transcriptome analyses in human cohort studies and eQTL analysis (through the ENGAGE DeepSAGE Transcriptomics Flagship project);
- Development of algorithms for CNV (copy number variation) detection on NGS data, at both whole-genome and whole-exome based dataset;
- Strategies specific for CNV analysis combining samples across different cohorts to search for CNV and metabolic associations;
- Strategies for analysis of MeCap sequence data as well as protocols for genome-wide methylation analysis (through the ENGAGE Epigenomics Flagship project);
- Development of new methods for data mining and for identification of candidate genes and disease relevance information including pathways, by exploring available internal and external genetic and genomic data as well as information from the literature
- Strategies for the analysis of longitudinal and multivariate phenotypes
- Exploration and development of statistical methods to combine genomics and other 'omics' (e.g. metabolomics) dataset for association studies (e.g. through the Metabolomics Flagship project)

1.3.5 Contribute to international efforts in large population cohorts as exemplified by our very close contacts with the P3G effort (Public Population Projects in Genomics)

Other than contributing data and analytical expertise to international trait-specific GWAS consortia through active collaboration, ENGAGE has been well connected with other major European and international projects relevant to large population cohorts. These include the EU-funded projects: 'Promoting Harmonisation of Epidemiological Biobanks in Europe' (PHOEBE), Harmonisation of phenotyping and Biosampling for Human Large-Scale Research Biobanks' (BioSHaRE-EU), 'Genotype to Phenotype Databases: a Holistic Approach' project (GEN2PHEN), the 'Biobanking and Biomolecular Resources Research Infrastructure' (BBMRI), the European Life Sciences Infrastructure for Biological Information (Elixir), 'BBMRI Large Perspective Cohorts' (BBMRI-LPC) and the Public Population Projects in Genomics (P3G) in Canada.

In particular, ENGAGE and P3G have several joint-force activities such as:

- Tool building to promote interoperability of large-scale international initiatives in the field, such as code of conducts for data sharing, bridging content for clinical and non-clinical environment, recommendation for recruitment by genotype...etc.
- Establishment of ENGAGE Study and Access Catalogue in P3G Observatory (<http://www.p3gobservatory.org/>);
- Organization of two Summer Institutes, the 1st International Biobank Summit;
- Organization of a joint workshop to explore the potential use of Datashaper, a harmonisation tool developed by P3G, in phenotypes relevant to metabolic syndrome



As a result from the ENGAGE –P3G joint workshop, a strategic collaboration towards a more streamlined and standardised procedure for harmonisation across cohorts was established. One of the key outputs from these harmonization efforts was the IT platform for phenotype harmonisation 'Sample avAILability system' (SAIL) developed by our integration and database teams. SAIL has been used in ENGAGE and other related EU projects (e.g. SUMMIT).

1.4 Potential Impact and Main Dissemination Activities and Exploitation of Results

1.4.1 Potential Impact

Europe has a strong track record as a pioneer in the field of human genetic and genomic epidemiology and remains the international leader in many respects. As the technologies rapidly advance and become more affordable, unified health care systems together with rich data and sample collections from populations as well as world-renowned molecular epidemiologists reflect unique opportunities for Europe to accelerate the translation of basic research findings into clinical applications. ENGAGE has used the special competitive niche of Europe to translate the wealth of data emerging from large-scale research efforts in genetic and genomic epidemiology conducted in well-characterised European (and other) samples into information of relevance to future clinical advance.

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 has continued using 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).

The first step in translation of these genetic discoveries is to define the molecular mechanisms through which they impact disease. ENGAGE made 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. The identification and characterisation of these alleles (especially where coding) will catalyze efforts to characterise biological mechanisms conferring risk and protection that will continue beyond the lifetime of ENGAGE. On-going efforts within the large-scale epidemiology flagship projects are expected to deliver further biological insights, including, the causal relationship between obesity and

other phenotype groups (such as cancer, neuropsychiatric outcomes), and complex pleiotropic relationships involving variants influencing the incretin axis.

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 and will continue to support the various efforts towards stratified medicine described earlier. To ensure the continuation of these efforts post-ENGAGE, ENGAGE is interacting closely with related EU-efforts (e.g. SUMMIT, DIRECT) which have shared goals and some overlapping membership.

The high profile research results generated by ENGAGE have helped to maintain the competitiveness of European research excellence in the field of human genetic and genomic epidemiological research and will have an impact of driving research advancement in the scientific community involved in the related fields.

Through various efforts from discovery and biology to translation, ENGAGE has provided the research community and industry (including research oriented SMEs) with advanced knowledge and tools for further preclinical/clinical exploitation that can ultimately lead to development of new biomarkers and companion diagnostics as well as identification of new targets for further drug discovery and development. The results (e.g. new biological mechanism) may also have potentials to provide new targets for interventions as prevention or treatment in clinical medicine.

ENGAGE results will also have a societal impact in multiple dimensions. These include updated ELSI guidelines and codes of conducts for moving genomic knowledge to the clinical research setting, as the molecular elucidation of pathways to disease is moving beyond a single focus on gene identification to other sources including diverse types of genetic variation, upstream phenotypes revealed through 'omic' profile, and gene-environment interactions that include behaviour traits. In relevance to health care system and welfare, the knowledge generated by ENGAGE, including genetic susceptibility and understanding the relationship between genetic and life-style/environmental risk factors, has good potential to lead to advances in clinical practice that will impact also at welfare level by influencing public health and policy decision.

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.

1.4.2 Main dissemination activities and exploitation of results

Dissemination activities

ENGAGE scientists have actively disseminated research results to both scientific and non-specialist audience. This has been exemplified by the extensive list of dissemination activities reported in the later section of this report (2.A2 List of Dissemination Activities), which is summarized here.

The ENGAGE project website (www.euengage.org) is served as central tool to widely disseminate ENGAGE main research outputs and activities to all audience.

Research community

ENGAGE scientists have made best efforts to disseminate research results to wider scientific audience through publishing papers on scientific journals, presenting information and project results at international conferences, meetings, workshops and courses (both external or co-organized by ENGAGE through the 'ENGAGE Training Policy').

So far, ENGAGE partners have published more than 250 scientific papers relating to its project activities; many of these are in high profile journals (such as Nature, Nature Genetics, Lancet) that have attracted wider dissemination activities. A list of publications is maintained under the ENGAGE public web site and is regularly updated.

An access catalogue has been developed to inform scientists outside of ENGAGE regarding research procedures for access to ENGAGE resources for genetic and genomic epidemiology research. The catalogue is available under the ENGAGE project web site (<http://www.euengage.org/data%20access%20catalogue.html>). Additionally, in collaboration with P3G, an ENGAGE Consortium Catalogue, served as a repository of standard information describing ENGAGE cohorts, is available in the P3G Observatory (<http://www.p3gobservatory.org/network/studies.htm?partner=engage>).

ENGAGE scientists have disseminated research data for access by wider scientific community. For example, summary level data of GWAS meta-analyses for major phenotypes are deposited and released through collaborative trait-specific consortia on the respective websites:

- MAGIC (the Meta-Analyses of Glucose and Insulin-related traits Consortium):
<http://www.magicinvestigators.org/>
- DIAGRAM (DIAbetes Genetics Replication And Meta-analysis Consortium)
<http://diagram-consortium.org/>
- GIANT (The Genetic Investigation of ANthropometric Traits Consortium)
http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium

- EGG (Early Growth Genetics Consortium)
<http://egg-consortium.org/>
- Global Lipids Genetics Consortium
<http://www.broadinstitute.org/mpg/pubs/lipids2010/>
- Human metabolic individuality (Suhre et al, Nature 2011)
<http://www.gwas.eu>

Furthermore, other type of data generated through ENGAGE activities have been (or are in the process of being) deposited in public data archives such as the European Genome-phenome Archive (EGA), ArrayExpress, PRIDE. These will be made accessible to bona-fide researchers outside of the consortium through a constituted Data Access Committee (DAC).

The ENGAGE Coordination Office has published an ENGAGE Newsletter distributed bi-monthly to all ENGAGE members to inform about activities in the project such as flagship projects, publications and training opportunities. Part of the Newsletter content that is of wider audience' interest, such as recently published high impact papers, ENGAGE's training opportunities (e.g. Summer Institute, open workshops...etc), has been published on the News bulletin of the project website by means of RSS feed.

Clinical and public health community

Clinicians are one of the key stakeholders for ENGAGE research results. Several ENGAGE scientists are involved in public health and publish scientific articles in this area. ENGAGE also (co)-organized workshops and seminars that were open to clinicians and the public health community. For example, the two ENGAGE Summer Institutes 'Genetics, Ethics and Clinical Translation' and 'Translational Genomics Pipeline: From Populations to Individuals' covered key clinical and public health components in the course programmes. In addition ENGAGE team on Ethics and Society worked on topics related to facilitating research projects involving data sharing clinical and non-clinical researchers.

General public

The ENGAGE project web site has been used as a channel to disseminate basic concept of ENGAGE that might be of interests to general public (<http://www.euengage.org/faq.html>).

Through local dissemination efforts in partner institutions, ENGAGE partners also disseminated their relevant research results to more layman audience, such as interviews by media, public lectures.

ENGAGE also explored the possibility to use audio-visual tools to introduce project concept and main results to the wider audience. An ENGAGE video clip was made in collaboration with a professional film producer FastFacts targeting to attract attention from general public. The video product has been widely distributed through web-based video sharing channels, such as Youtube, Vimeo as well as on the ENGAGE project website.

Link to ENGAGE video:

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



ENGAGE Video: From Genetic Discovery to Future Health (by ENGAGE and FastFacts.nl)

List of selected ENGAGE publications (from a total of over 250)

1. Aulchenko and Ripatti et al, 'Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts'. *Nature Genetics* 41, 47 – 55, 2009 (epub 2008)
2. Lyssenko V et al. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nature Genetics* 41, 82 – 88, 2009 (epud 2008)
3. Prokopenko et al, 'Variants in MTNR1B influence fasting glucose levels'. *Nature Genetics* 41, 77 – 81, 2009 (epub 2008).
4. Thorleifsson et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nature Genetics* 41, 18 – 24, 2009 (epub 2008)
5. Kolz et al. Meta-Analysis of 28,141 Individuals Identifies Common Variants within Five New Loci That Influence Uric Acid Concentrations. *PLoS Genet.* Jun;5(6):e1000504, 2009
6. Illig et al, A genome-wide perspective of genetic variation in human metabolism. *Nature Genetics*,42:137-141, 2010
7. Lango Allen H et al., Hundreds of variants clustered in genomic loci and biological pathways affect human height, *Nature*, Oct 14;467(7317):832-8, 2010
8. Teslovich et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466(7307):707-13, 2010
9. Thorgeirsson et al., Sequence variants at CHRNA3-CHRNA6 and CYP2A6 affect smoking behavior and the risk of lung cancer. *Nature Genetics* 42, 448-453, 2010.
10. Wang et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet*, 376(9736):180-8, 2010
11. Berglund E, Kiialainen A, Syvänen A-C. Next Generation Sequencing Technologies and Applications for Human Genetic History and Forensics. *Investigative Genetics* 2:23, 2011
12. Budin-Ljøsne L, Tassé AM, Knoppers BM, Harris JR, Bridging Consent: From Toll Bridges to Lift Bridges? *BMC Medical Genomics*. Oct 4;4:69, 2010
13. Chambers JC et al., Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature Genetics*. 43(11):1131-8, 2011
14. Ehret GB et al., Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. Sep 11;478(7367), 2011
15. Ferreira MA et al, Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet*. 78(9795): 1006-14, 2011
16. Gostev M, Fernandez-Banet J, Rung J, Dietrich J, Prokopenko I, Ripatti S, McCarthy MI, Brazma A, Krestyaninova M. SAIL – a software system for sample and phenotype availability across biobanks and cohorts. *Bioinformatics*;27:589-591, 2011

17. Knoppers BM, Harris JR, Tassé AM, Budin-Ljøsne I, Kaye J, Deschênes M and Zawati MH, Towards a data sharing Code of Conduct for international genomic research. *Genome Med.*, Jul 14;3(7):46. 2011
18. Schumann G et al., Genome-wide association and functional studies identify AUTS2 in the regulation of alcohol consumption. *PNAS*; 108(17):7119-24, 2011
19. Suhre K, Shin S-Y, Petersen A-K et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* ; 477: 54-60, 2011
20. 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* Oct;7(10):e1002333, 2011
21. 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*;54:2801-2810, 2011
22. Deloukas P et al, Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics*. Jan;45(1):25-33, 2013 (Epub 2012)
23. de Moor MH et al, Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry*. 7(3):337-49, 2012 (2010 epub)
24. Hughes et al. Genetic Markers Enhance Coronary Risk Prediction in Men: The MORGAM Prospective Cohorts. *PLoS ONE* 7(7): e40922. 2012
25. Ivliev AE, 't Hoen PA, van Roon WM, Peters DJ, Sergeeva MG. Exploring the Transcriptome of Ciliated Cells Using In Silico Dissection of Human Tissues. *PLoS One*. 2012;7(4):e35618.
26. Morris AP et al, Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics*. 44(9):981-90, 2012
27. Ripke S et al, A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry*, Apr 3 2012 (epub)
28. Scott RA et al, Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics*. Sep;44(9):991-1005, 2012
29. Yang J et al, FTO genotype is associated with phenotypic variability of body mass index. *Nature*. 490(7419):267-72, 2012
30. Horikoshi et al, New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nature Genetics*.;45(1):76-82, 2013
31. Codd V, Nelson C, Albrecht E et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nature Genetics*. Apr;45(4):422-7, 2013

Exploitation of results

As an integrated research project aiming to advance knowledge in human disease genetics and explore relevant translational opportunities in medicine, the research community is identified as one of the primary users to take up ENGAGE research results for further exploration before the outputs can be fully translated into the clinics. Such exploitable results include analytical data generated, methods and tools developed, laboratory protocols and technology platforms established as well as databases built. The main exploitation measures taken by ENGAGE for such results targeting scientific community have been through publication of research papers, presentations at international scientific conferences, training proposals through organisation of open courses/workshops, consultation through research collaboration and website resources. For more detailed activities, see 'Dissemination activities' in the previous section.

It is important to make clear that the majority of ENGAGE research takes place in the "precompetitive space" and adheres to the norms of the field with respect to not claiming IP positions on genetic discoveries. Rather we make them available as rapidly and widely as possible for others to develop further the scientific picture and to build the case for translational use.

ENGAGE research is built on the invaluable resources from large-scale cohort collections. Some of the results therefore are highly relevant to the biobanking community and can also be used by this community. Specifically, the various databases developed can be used for tracking meta-consortium activities (Emanta), for data storage, exchange and annotation (AIMS), and for inventories of available specimens as well as for harmonisation and indexing across several collections in biobanks (SAIL). Some of these database tools have been used by institutions involved in biobanking activities as well as international projects.

Also the ENGAGE ELSI tools built to improve interoperability can be exploited by the biobanking community. These include code of conduct for data sharing, recommendations for bridging consent in clinical and non-clinical environment, for retrospective access to data, and for recruitment by genotype.

Technology platforms improved for ENGAGE targeted or large-scale sample analyses, such as next generation sequencing (e.g. whole-genome, exome based), genotyping with custom arrays (e.g. metabochip), genome-wide DNA methylation analysis are results that can be further exploited for commercial utilisation (in research or clinical diagnostics). The exploitation of such results has been primarily through our industrial partners, Illumina and deCODE, who have been actively involved in several of these ENGAGE activities.

In addition to technology platforms, biomarkers as predictors of disease onset or progression would be potential targets for commercial exploitation. For example, the identification of hs-CRP as a non-genetic biomarker for diabetes subtypes can be one immediate target for clinical use. However, even though significant progress has been made to explore translational opportunities, ENGAGE

findings at this stage still represent early innovations and direct exploitation for diagnostic inventions is relatively modest. ENGAGE has established strategic collaboration with several EC funded projects (e.g. IMI-SUMMIT, DIRECT) to support programmes that will carry on our efforts towards translation.

1.5 Address of Project Public Website and Contact Details

Website: <http://www.euengage.org>

Email: contact@euengage.org

Contact details:

Professor Mark McCarthy, Robert Turner Professor of Diabetes
The Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM)
University of Oxford
Churchill Hospital
Headington
OX3 7LJ
UK

and

Professor Jaakko Kaprio, Professor of Genetic Epidemiology
Institute for Molecular Medicine Finland (FIMM)
Nordic EMBL Partnership for Molecular Medicine
University of Helsinki
Biomedicum Helsinki 2U
8, Tukholmankatu 8
PO Box 20
University of Helsinki
FI-00014
Finland

¹ Pictures on cover page and page 20 extracted from ENGAGE video 'From Genetic Discovery to Future Health' by ENGAGE and Fastfacts.nl