Grant Agreement number: LSHC-CT-2005-018827

Project acronym: POLYGENE

Project title: Inherited risk of breast and prostate cancer

Funding Scheme: FP6 - LSH-2004-2.2.0-6: Prevention, detection and treatment of familial cancers, such as cancer of the prostate, ovary, breast, colon and skin

6.1 PUBLISHABLE FINAL ACTIVITY REPORT

Period covered from: December 1, 2005 – September 30, 2009

Date of preparation: December 10, 2009

Start date of project: December 1, 2005 **Duration:** 46 months

Project co-ordinator name, title and organisation:

Thorunn Rafnar, Ph.D.,

Head of Division of Cancer,

deCODE genetics, Reykjavik, Iceland

Tel: +354 570 1836

Fax: +354 570-1903

E-mail: thorunn.rafnar@decode.is

Project website address: **www.polygene.eu**

Table of Contents

1. PROJECT EXECUTION

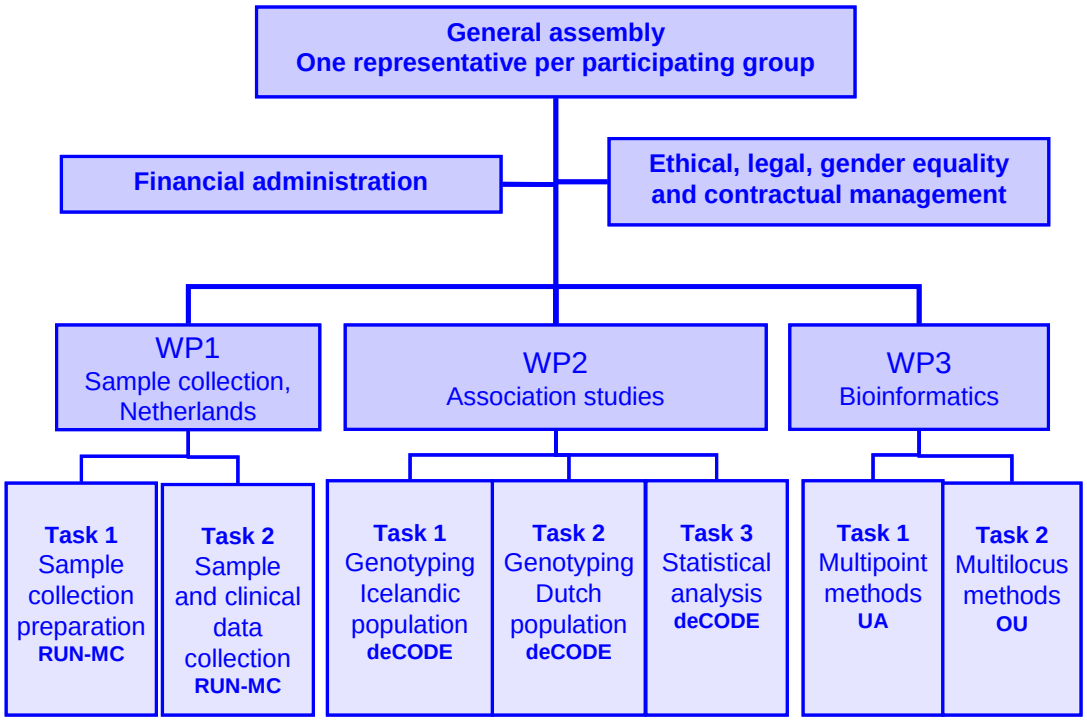
Project objectives

The overall objective of POLYGENE was to identify inherited risk factors of two important cancers that have been shown to have a substantial genetic component, prostate cancer in men and breast cancer in women. The specific objectives of the project were:

1. *Genome-wide association study of breast and prostate cancer.*
   Associations between SNP polymorphisms and the risk of breast and prostate cancer were to be examined in unselected samples from breast and prostate cancer patients in Iceland and the Netherlands, two populations with different structure and history.

2. *Development of efficient statistical and computational methods for the analysis of genetic and association data.*
   Two areas of research were proposed: 1) Multipoint methods and 2) multi-locus methods.

The long-term objective of POLYGENE is to gain an increased understanding of the genetic underpinnings of breast and prostate cancer which may, in turn, lead to more effective risk assessment, increase the efficiency of screening programs and lead to improved diagnosis and treatment. The improved analysis methods will also aid genomic research for other diseases. An overview of the project is depicted below.



Contractors involved

Part. Role*	Part. no.	Participant name	Participant short name	Country	Date enter project**	Date exit project**
CO	1	deCODE Genetics PI: Thorunn Rafnar	deCODE	Iceland	Month 3	Month 46
CR	2	University of Aarhus PI: Mikkel Schierup	UA	Denmark	Month 5	Month 46
CR	3	University of Oxford, Department of Statistics PI: Jotun Hein	OU	UK	Month 1	Month 46
CR	4	Radboud University Nijmegen Medical Center PI: Lambertus (Bart) Kiemeney	RUN-MC	Netherlands	Month 1	Month 46
CO	6[1]	Iceland Genomics/UVS PI: Eirikur Steingrimsson	IGC/UVS	Iceland	Month 1	Month 3
CR	5[2]	Bioinformatics ApS PI: Leif Schauser	Bioinformatics	Denmark	Month 1	Month 5

Table 1. POLYGENE partners

[1] The Icelandic SME, IGC/UVS, was the original coordinator of the project. In January 2006, ICG/UVS became a wholly owned subsidiary of deCODE genetics ehf. in Reykjavik, Iceland. Three months after the start date of the project, or on 28 February 2006, IGC/UVS withdrew from the consortium and its role was taken over by deCODE.
[2] In spring 2006, Bioinformatics Aps, an SME of Aarhus, Denmark, informed the consortium of its intention to withdraw, and was subsequently replaced by the University of Aarhus (Bioinformatics Centre) on 15 May 2006.

Work performed and summary of results (months 1-46)

WP1: Construction of a population-based repository of Dutch breast (700) and prostate (700) cancer cases with complete clinical and lifestyle information.

Background:

The major objective of WP1 was to acquire a truly population-based sample of Dutch breast and prostate cancer cases with complete clinical and lifestyle information. Such a collection had already been established in Iceland. Although several breast and prostate cancer cohorts were already in existence in the Netherlands, many of those were not appropriate for this study. A number of study populations focused on familial cases, in other instances there was lack of sufficient clinical information and/or enough high-quality DNA for large-scale genotyping, the informed consent did not cover the use of samples for new studies, or the sampling frame was not population-based which makes a valid case-control comparison almost impossible. Therefore, we believed it was crucial for the success of this study to start a new collection of cases that was specifically designed to include all the important variables required for a genetic study of a complex disease. The Icelandic study design was used as a prototype which ensured that all the information gathered on the Icelandic patients was also collected for the Dutch sample set.

Controls for the project were obtained from a survey conducted in 2002-2003 which was based on a random sample of the population registration (the Nijmegen Biomedical Study: NBS).

Results:

Preparation of control samples:
Life-style information, family history of cancer, reproductive and medical history as well as blood samples were already available from a group of 6,700 population controls. From this group, initially 700 female and 700 male controls were selected, frequency age-matched to the patient population. DNA was isolated from their blood samples, concentrations were measured and the samples were diluted to a standardized concentration 100 ng/μl. For each control 50 μl was shipped to Iceland for genotyping. In a second phase, another 300 men and 300 women were selected and DNA was shipped to Iceland in order to reach a total of 1000 male and 1000 female controls.

Preparation of data collection from cases:
Male- and female specific life-style questionnaires, which resembled the NBS questionnaires as closely as possible, were developed. The questionnaires were pilot tested, modified, transformed into Teleform scanning forms, and printed by a printing office. A brochure was produced explaining the goals and methods of the POLYGENE project, to be sent to the patients together with the invitation letter. All physicians (urologists, radiotherapists, medical oncologists, surgeons) in the hospitals in the area of the Comprehensive Cancer Center East (IKO) were contacted and asked for their consent to invite their patients for POLYGENE. All physicians consented. An invitation letter was written to be signed by the chairs of the Breast Cancer Working Group and Urological Tumor Working Group of the Comprehensive Cancer Center IKO, in the name of all members of these working groups. Furthermore, breast cancer and prostate cancer forms were developed to collect the baseline and follow-up clinical data of all the patients. The draft forms were discussed with several physicians in order to adapt them to Dutch clinical practice. The definitive forms were transformed into Teleform scanning forms and printed by a printing office.

Preparation of patient recruitment and study approval:
Five regional Thrombosis Service Centers, covering the region of the Comprehensive Cancer Center IKO were contacted and asked to draw blood samples of all the consenting patients. Practical arrangements were made and agreed upon by the IKO (coordinative center), the Thrombosis Service Centers and the Department of Clinical Chemistry (sample administration and storage). All procedures and written materials were submitted to the ethical committee of the Radboud University Nijmegen Medical Center (CMO Regio Arnhem-Nijmegen). Approval for the study was granted. Personnel responsible for patient contact, sample and data collection and DNA isolation were trained and the database infrastructure to handle the data was set up.

Collection of samples and data from cancer cases:
In total, 1,266 breast cancer patients and 1,400 prostate cancer patients with an age at diagnosis under 75 were selected from the IKO regional cancer registry. First, data on vital status was updated with data from the Municipality Population Register (GBA). All patients still alive and with a known address were contacted by mail with the invitation letter and the brochure. If they consented to participation, they were sent the lifestyle questionnaire and stickers for their blood samples. The original goal was to collect samples and data from 700 prostate cancer patients and 700 breast cancer patients. However, the number of prostate cancer patients with blood samples was further increased to a total of close to 1,200 patients and breast cancer to 800 cases. One of the 2 blood tubes of each patient was shipped to Iceland for DNA isolation and genotyping.

Collection of clinical data:
Cancer registry personnel visited all 7 hospitals in the IKO area regularly to extract clinical information on diagnosis, treatment and outcome, from both paper and electronic medical files. Clinical data was also collected from prostate cancer patients who could not be recruited for the project in order to obtain information on potential selection bias for disease severity. Through this, a clinical database of an unselected cohort of patients was assembled for prognostic studies. At the end of POLYGENE, primary data has been collected for all approximately 800 breast cancer cases and 1,500 prostate cancer cases and follow-up data has been collected for a large fraction of the cases.

In summary:

All the objectives of WP1 have been reached or exceeded. At the end of the POLYGENE a sample collection and clinical database of 800 breast cancer cases and 1,200 prostate cancer cases has been established and has been extensively used in the genetic studies of POLYGENE and a number of other projects (see end of this section and a complete list of publications in report 6.2; Final plan for using and disseminating knowledge).

WP2: Genetic association studies of Icelandic and Dutch breast and prostate cancer patients and controls

Background:
The major objective of WP2 was to perform genome-wide association (GWA) studies to identify genetic variants that affect susceptibility to breast and prostate cancer. At the outset of POLYGENE, the Icelandic cancer patient samples had already been collected, i.e. samples from 1,600 breast cancer patients and 1,400 prostate cancer patients. All patients had a clinically verified disease and clinical information such as tumor stage, laterality, grade, histology, treatment and disease recurrence were available for each patient as well as information on a number of potential risk factors, including height, weight, smoking history and family history of cancer. Patients with breast cancer and female controls also had submitted thorough information on menstrual history, reproductive history and hormone use. In addition to the clinical data and lifestyle data, the complete genealogy of all study subjects is registered in the genealogy database of deCODE, the "Book of Icelanders".

Results:

In order to find variants that affect the risk of prostate or breast cancer, we genotyped Icelandic cases and controls on the Illumina Hap300 bead chip. The control group (>35,000 subjects) consisted of individuals from other ongoing genome-wide association studies at deCODE and represent over 15% of the adult population of Iceland. No individual disease group is represented by more than 10% of the total control group. After removing SNPs failing quality control checks, 304,073 SNPs were tested for association with the two cancer types. The results were adjusted for relatedness between individuals and for potential population stratification using the method of genomic control.

Prostate cancer:
A total of 1,400 prostate cancer cases were genotyped on chip and the genotype frequencies compared to the frequencies in a large number of controls from the Icelandic population. From

the results of the GW scan, candidate markers were selected based on a combination of P-value rank and previous results from linkage studies, and assessed by single-SNP genotyping in the Dutch population, as well as a number of other external sample sets provided by collaborators in Spain, Sweden and the USA. To summarize, work in POLYGENE has resulted in the discovery of 6 common variants that associate with risk of prostate cancer. The variants are listed in Table 2, along with information on their location, risk allele, number of cases and controls genotyped, effect of the variant (odds ratio, OR), P-value and reference. The variants on 8q24 and 2p15 show a significantly stronger association with more aggressive forms of the disease. The variant on 5p15.33 was originally discovered in deCODE's research on basal cell carcinoma of the skin but subsequently confirmed to associate with risk of prostate cancer as well.

SNP	lLocus	Risk allele	# Cases	# Contr	OR	*P* value	Reference
rs16901979	8q24 - hap C	A	1,453	3,064	1.60	$6.4x10^{-18}$	Gudmundsson Nat Gen 2007
rs7501939	17q12	C	3,490	14,345	1.19	4.7×10^{-9}	Gudmundsson Nat Gen 2007
rs1859962	17q24.3	G	3,493	14,344	1.20	2.5×10^{-10}	Gudmundsson Nat Gen 2007
rs5945572	Xp11.22	A	10,054	28,897	1.23	3.9×10^{-13}	Gudmundsson Nat Gen 2008
rs721048	2q15	A	10,093	28,654	1.15	7.7×10^{-9}	Gudmundsson Nat Gen 2008
rs401681	5p15.33	C	9,473	37,901	1.07	3.6×10^{-4}	Rafnar Nat Gen 2009

Table 2. Prostate cancer risk variants identified in POLYGENE.

Breast cancer:
Close to 1,600 breast cancer cases were genotyped on chip and compared to the genotypes of a large number of controls. Signals were ranked by P-value and SNPs representing the best 10 loci were selected for scrutiny. The SNPs were genotyped in the breast cancer sample set from the Netherlands, as well as in an independent sample of Icelandic breast cancer patients and controls and in 2 independent case-control sets from Sweden, and Spain. SNPs that returned nominally significant signals in each replication set were then tested in a case-control set of European Americans from the U.S. Multiethnic Cohort. This analysis resulted in the discovery of two variants that associate with breast cancer risk, one on chromosome 2q35 and the other on 16q12. Subsequent scrutiny of the data in conjunction with the data from the Cancer Genetic Markers of Susceptibility (CGEMS) group (http://cgems.cancer.gov/data/) has resulted in the identification of two additional variants, one on 5p12 and another at the *ERS1* locus on chromosome 6. The variants are listed in Table 3 along with information on their location, risk allele, number of cases and controls genotyped, effect of the variant (odds ratio, OR), P-value and reference.

The first three variants were identified through a "traditional" GWAS approach. The fourth SNP was identified through a fine-mapping effort of a breast cancer locus originally reported in a Chinese population of breast cancer cases and controls. Here, we were able to show that the initially reported Chinese breast cancer risk variant rs2046210 cannot be used effectively as a risk marker in Europeans and Africans because it is not in strong LD with a causative variant in

all three ancestries. Thus, we identified another variant that confers risk of breast cancer in all three main ancestral populations. In summary, we have shown that ancestry shift refinement mapping can be useful in the identification of SNPs that associate with risk in populations of different ancestries. This has practical implications for genetic testing and highlights that a comprehensive approach is necessary when investigating whether a risk variant identified in one ancestral population is also present in another ancestry.

SNP	Locus	Risk allele	Sample set	#Cases	#Contr	OR	*P* value	Reference
rs13387042	2q35	A	All samples	4,533	17,513	1.20	1.3×10^{-13}	Stacey et al 2007
rs3803662	16q12	T	All samples	4,554	17,577	1.28	5.9×10^{-19}	Stacey et al 2007
rs4415084	5p12	T	All samples	5,028	32,090	1.16	6.4×10^{-10}	Stacey et al 2008
New variant	6q25.1	G	All Ancestries	10,176	13,286	1.19	3.9×10^{-7}	Stacey et al subm.
" "	" "	G	Asian	1,126	1,118	1.23	8.0×10^{-4}	Stacey et al subm.
" "	" "	G	European	7,899	11,234	1.15	1.2×10^{-3}	Stacey et al subm.
" "	" "	G	African & AA	1,151	934	1.35	0.014	Stacey et al subm.

Table 3. Breast cancer risk variants identified in POLYGENE.

In summary:

All the objectives of WP2 have been reached. In addition to identifying 10 variants that associate with risk of breast or prostate cancer, the variants have been incorporated into commercial tests for genetic cancer risk. Thus, deCODE ProstateCancer™ was launched in February 2008 and deCODE BreastCancer™ was launched in October 2008.

WP3: Development and application of bioinformatic solutions

Background:

Before the onset of POLYGENE, the partners at AU and OU had developed software for the fine-mapping of disease genes called *GeneRecon*. This software conducts multipoint analysis by modeling the unknown genealogy that relates diseased individuals, examined in a genetic study of the case – control type. The modeling of the *Coalescent with recombination* is a computationally complex problem which *GeneRecon* approaches by Markov-Chain Monte-Carlo technology in a Bayesian setting. Consequently, this and similar approaches are very slow for large data sets, allowing only 100s of markers to be analysed in reasonable time. Thus, when the focus of POLYGENE was shifted from the analysis of 100 candidate genes to GWA studies, it became necessary to develop further analysis tools that could handle GWA data.

Results:

Haplotype methods. In order to face whole-genome association data, a heuristic model, based on simpler aspects of the data generating process, namely marker compatibility, were developed. This approach was implemented in the software *Blossoc*, which is able to handle analysis of 1 million markers in 5,000 cases and 5,000 controls in a single day. The method

was also extended to be applied to quantitative traits (*QBlossoc*) and unphased data (*GBlossoc*). *Blossoc* has been released as a free software package (www.birc.au.dk/~mailund/Blossoc/index.html).

Aside from Blossoc two other methods were worked on:

- The *Haplotype Pattern Mining* (HPM) method, where we developed a faster algorithm and made an implementation efficient enough to scale to whole genome data. (Besenbacher et al.)
- The *HapCluster* method, a fine-scale mapping method, where we have developed a faster implementation of the basic method and extended it to deal with un-phased data. HapCluster is a Markov-chain Monte Carlo (MCMC) based method, so we use this approach in our work albeit in a very different way than originally planned in the grant proposal. HapCluster has been released as free software (www.birc.au.dk/~mailund/HapCluster/index.html).

Interaction methods. In this part, the focus was on two main issues: using biological networks to focus the search for interacting genes, and improving the statistical test for interaction. The idea behind using biological networks is to exploit the large amount of pre-existing knowledge about biological networks, and the fact that *a priori* we would expect interacting genes to be related in a network.

By only testing for interaction in genes closely related in known biological networks, it is possible to drastically reduce the number of tests while still testing all relevant pairs of genes. We have implemented this idea and tested it both on POLYGENE data and WTCCC data. The results were described in a paper in *EJHG* (Emily et al. 2009). As for improving tests, three new tests were developed: one based on testing for locus heterogeneity, a dual test which tests for locus multiplicity and a test that uses higher-order LD measures to improve tests based on contrasting LD in cases and controls. All three tests have been implemented and tested on the WTCCC data. Finally, in an attempt to combine haplotype methods with interaction, extensions of both *Blossoc* and *HapCluster* were developed that can search for pairwise interactions in unlinked genes. We also have an implementation of a logic regression method but have not tested this method fully yet.

Data Analysis. *Blossoc* was used to analyse prostate, breast and colon cancer data at deCODE, with and without the network-based interaction mapping method. The results confirmed the genetic locations identified by deCODE's analysis and suggested new loci that had not been found with previous methods. An attempt was made to replicate a total of 11 different haplotypes for the three cancers, using Dutch, Spanish and Swedish case control sample sets. None of the 11 loci showed an association in the replication samples sets. These results suggest that either the signals for these haplotypes were false positive or that the genetic risk factors are specific for the Icelandic population.

Finally, in order to cope with the data size and the various kinds of secondary data, we developed a new file format, *SNPFile*. This was not mentioned in the original workplan, but it is the main informatics framework underlying all our software.

Software developed and released:

- *SNPFile* (http://www.birc.au.dk/~mailund/SNPFile/index.html) a library and API for manipulating large SNP datasets with associated meta-data, such as marker names, marker locations, individuals' phenotypes, etc. in an I/O efficient binary file format.

- *SMA* (http://www.birc.au.dk/~mailund/sma/index.html) consists of a small collection of programs that perform different tests for association between genotypes at a single marker and a binary or continuous phenotype.
- *Blossoc* (http://www.birc.au.dk/~mailund/Blossoc/index.html) is a linkage disequilibrium association mapping tool that attempts to build (perfect) genealogies for each site in the input, score these according to non-random clustering of phenotypes and judge high-scoring areas as likely candidates for containing phenotype affecting variation. Building the local genealogy trees is based on a number of heuristics that are not guaranteed to build true trees, but have the advantage of more sophisticated methods of being extremely fast. *Blossoc* can therefore handle much larger datasets than more sophisticated tools, but at the cost of sacrificing some accuracy.
- *HapCluster* (http://www.birc.au.dk/~mailund/HapCluster/index.html) is a C++ implementation of the *HapCluster* MCMC association mapping method. Based on a simple model of relatedness, it searches the state space of haplotype clusters and scores for significant clustering of cases rather than controls in such clusters.
- *GeneRecon* (http://www.birc.au.dk/~mailund/GeneRecon/index.html) is a software package for linkage disequilibrium mapping using coalescent theory. It is based on a Bayesian MCMC method for fine-scale linkage-disequilibrium gene mapping using high density marker maps. *GeneRecon* explicitly models the genealogy of a sample of the case chromosomes in the vicinity of a disease locus. Given case and control data in the form of genotype or haplotype information, it estimates a number of parameters, most importantly, the disease locus position.
- *CoaSim* (http://www.birc.au.dk/~mailund/CoaSim/index.html) is a tool for simulating the coalescent process with recombination and gene conversion under various demographic models. It effectively constructs the ancestral recombination graph for a given number of individuals and uses this to simulate samples of SNP, microsatellite, and other haplotypes/genotypes. The generated sample can afterwards be separated into cases and controls, depending on states of selected individual markers. The tool can accordingly also be used to construct cases and control data sets for association studies.

The last two tools, *GeneRecon* and *CoaSim*, were not developed for POLYGENE but *GeneRecon* was included in the original workplan and *CoaSim* was used in most of our simulation experiments. In addition, we have developed tools for HPM and the various interaction tests, but these have not been released yet.

In summary:

All the objectives of WP3 have been reached or exceeded. The work has resulted in considerable improvements in several methods applied in the analysis of genetic data. The methods have been implemented in free software packages that have been made publically available.

Scientific publications from POLYGENE

Following is a list of publications that credit POLYGENE for support. The abbreviated reference is given here but full reference is presented in report *6.2 Final plan for using and disseminating knowledge.*

1. Rafnar T, Kiemeney LA. Polygene: Identification of common genetic variants that affect the risk of breast and prostate cancer. Eur Urol Today 2006; 18(4): 21.

2. Melchior M. Genetisch gewin. IJslands genenpakket wordt te gelde gemaakt. Medisch Contact 2006; 61(5): 192-4.

3. Kiemeney LALM, Aben KK. De eerste whole genome association studie bij prostaatkanker. Epistel 2007; 20: 9-10.

4. Gudmundsson J et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. 2007 Nat Genet, 39: 631-7.

5. Stacey SN et al. Common Variants on Chromosomes 2q35 and 16q12 Confer Susceptibility to Estrogen Receptor Positive Breast Cancer. 2007 Nature Genetics 39, 865-869

6. Kiemeney LALM et al. Nieuwe technologie onderzoek erfelijke risicofactoren. Tijdschrift Kanker 2007; 31: 28-31.

7. Gudmundsson J et al. Two Sequence Variants Conferring Risk of Prostate Cancer Identified on Chromosome 17 and One of Them, in TCF2, Shown to be Protective against Type 2 Diabetes. 2007 *Nature Genetics* 39, 977-983

8. Gudmundsson J et al. Common Sequence Variants on 2p15 and Xp11.22 Confer Susceptibility to Prostate Cancer. 2008 *Nature Genetics* 40, 281-283

9. Hellenthal G et al.Inferring human colonization history using a copying model. 2008 *PLoS Genet.* 4(5):e1000078.

10. Stacey SN et al. Common Sequence Variants in the MRPS30 Locus on Chromosome 5p12 Confer Susceptibility to Estrogen Receptor Positive Breast Cancer. 2008 *Nature Genetics* 40, 703-706

11. Kiemeney L. IJsland voedingsbodem genetisch onderzoek Bart Kiemeney. AOCN Magazine 2008; 4: 3-5.

12. Rafnar T et al. Sequence variants at the TERT- CLPTM1L locus associate with multiple cancer types. *Nature Genetics* 2009 Feb;41(2):221-7.

13. Emily M et al. Using Biological Networks to Search for Interacting Loci in Genomewide Association Studies, *Eur J Hum Genetics* 2009 Oct; 17(10): 1231-40. Epub 2009 Mar 11.

14. Stacey SN et al. Ancestry-Shift Refinement Mapping of the C6orf97-ESR1 Breast Cancer Susceptibility Locus. Manuscript to be submitted in December 2009.

15. Besenbacher S et al. Detecting interaction between unlinked loci as deviations from locus heterogeneity. Manuscript to be submitted in early 2010.

The impact of POLYGENE

The long-term goal of POLYGENE is to identify genetic factors that affect the risk of prostate and breast cancer. Both prostate and breast cancer are diseases with a considerable genetic factor that cannot be easily modified with environmental changes, high prevalence in the population and high impact on public health. It has been estimated that genetic factors contribute over 40% of the variation in prostate cancer risk and 27% of breast cancer risk. Thus they represent diseases with great potential for use of genetic information. Once all genetic risk factors for the two diseases have been characterized, it will be possible to produce genetic risk models that can be used to identify those individuals with the highest genetic risk. Both breast and prostate cancer have a high cure rate when detected early. Therefore, frequent screening of individuals with high genetic risk can facilitate early detection, improve prognosis and eventually result in lower mortality due to these diseases.

At the outset of the project, only a minor fraction of genetic risk for breast cancer had been characterized and most of this risk could be attributed to mutations in the breast cancer genes, *BRCA1* and *BRCA2*. The high genetic fraction of prostate cancer remained almost entirely unexplained. Accumulating evidence suggested that most of genetic cancer risk was due to multiple common risk alleles, where the risk associated with each individual allele was small to moderate. Importantly, as the variants may interact in a multiplicative or super-multiplicative (epistatic) way, an individual with several susceptibility alleles might be at significant risk. It had been realized that if the polygenic model of cancer risk was correct, linkage analysis of families with multiple cases of breast or prostate cancer would not suffice because it lacks power to detect the low-penetrance risk variants. The association approach was suggested as the method of choice for identifying low-penetrance genes (risk ratio 1.1-2.0). However, until about 2006, genome-wide genetic association studies had been prohibitively expensive.

In the beginning of 2006, genotyping technologies became available that allowed the scanning of the human genome with hundreds of thousands of genetic markers in order to detect small but possibly relevant differences in allelic frequencies of all genes. POLYGENE was perfectly poised to take advantage of this revolution, having already assembled the three most important ingredients; i.e. biological samples, clinical data and a dedicated team of researchers. To summarize the success of the project, POLYGENE has discovered a considerable fraction of the variants identified to affect risk of breast and prostate cancer. Specifically, 6 variants that affect prostate cancer risk were identified and 4 variants that affect risk of breast cancer.

These variants, along with additional variants identified by us and others, have been incorporated into genetic risk models for both cancer types. Although much of the genetic risk elements for breast and prostate cancer still remain to be found, our results indicate that a model including low risk variants can be useful in predicting risk for prostate cancer. A genetic diagnostic test incorporating 22 common low-risk variants for prostate cancer demonstrates that the average relative risk of prostate cancer in the top 10% of men having the highest risk is >2.3-fold the risk of the general population. Currently, cancer risk models which incorporate genetic and non-genetic risk factors are being developed for prostate and breast cancer by several groups in the UK and US.

The utility of the genetic tests for prostate and breast cancer will grow rapidly as more variants are uncovered and included. At the same time, we will gain knowledge about the biological pathways that play a role in prostate and breast cancer initiation and which may subsequently be used to develop more effective prevention and treatment strategies.

Further information

Information about POLYGENE can be found at www.polygene.eu. The web-site also links to publications resulting from the project.

2. DISSEMINATION AND USE

Description of exploitable results:

The major exploitable result from POLYGENE is the discovery of 6 genetic variants that associate with prostate cancer risk and 4 genetic variants that associate with risk of breast cancer. Knowledge about genetic factors that increase risk of cancer can be used to develop DNA-based test that evaluate a person's genetic risk of developing the disease. This information can in turn help to guide measures that aim to prevent the disease in those at high risk or increase their chance of early diagnosis through screening programs.

Market application and stage of development:

deCODE genetics, is a biopharmaceutical company whose major aim is to turn its discoveries in human genetics into the development of drugs and diagnostics for common diseases. deCODE currently offers two different kinds of DNA-based tests through its CLIA registered testing laboratory. First, the company produces a growing number of reference laboratory tests for estimating an individual's risk of developing diseases such as type 2 diabetes, atrial fibrillation, heart attack and certain cancers. The second kind of test offered by deCODE is a complete personal genomic scan offered directly to the customer, *deCODEme*™. Currently, the results report assessment of genetic risk for 46 diseases and traits as well as information about the person's ancestry. A reduced version of the complete scan, the *deCODEme Cancer*™ is a focused version of deCODEme which tests for genetic risk for 7 of the most common cancers, including breast cancer for women and prostate cancer for men.

The genetic variants discovered in POLYGENE will all be incorporated into the genetic tests for prostate cancer (*deCODE PrCa*™) and breast cancer (*deCODE BreastCancer*™) marketed by the company. Furthermore, all variants are included in the risk profile provided to the customers of *deCODEme*™ and *deCODEme Cancer*™. The variants may in the future be licensed to other producers of diagnostic tests.

Intellectual property rights granted or published:

Title of patent application	Pub date	Pub number
Prostate cancer		
Cancer Susceptibility Variants on CHR 8q24.21	2.5.2008	WO2008/050356
Genetic Variants Contributing to Risk of Prostate Cancer	14.8.2008	WO2008/096375
Genetic Variants Predicitve of Cancer Risk	Exp.feb 2010	
Breast cancer		
Genetic variants on CHR2 and CHR16 as Markers for Use in Breast Cancer Risk assessment, Diagnosis, Prognosis and Treatment	2.10.2008	WO2008/117314
Genetic variants on CHR5p2 and10q26 as Markers for Use in Breast Cancer Risk assessment, Diagnosis, Prognosis and Treatment	4.12.2008	WO2008/146309
New variant	Exp.2010	

Collaboration sought or offered:

Not applicable

Contact details:

deCODE genetics
Sturlugata 8
IS-101 Reykjavik
Iceland

Tel: +354 570 1900
Fax: +354 570 1903
E-mail: info@decode.is