

POLYGENE

Identification of common genetic variants that affect the risk of breast and prostate cancer



Dr. Thorunn Rafnar
deCode Genetics
Reykjavik (IS)



Prof. Dr. Bart Kiemeneij
Radboud University
Nijmegen Medical
Centre
Nijmegen (NL)

Summary

Linkage studies of families with many cancer patients have identified high-penetrance cancer genes such as BRCA1 and BRCA2. However, a large component of inherited cancer risk remains unaccounted for. It has been proposed that common low-penetrance variants in cancer susceptibility genes contribute significantly to genetic predisposition, either individually or through their interactions. In the POLYGENE project, genome-wide association studies of two European populations of different history and structure are being performed in order to identify common risk variants that contribute to the risk of prostate and breast cancer. At the same time new statistical methods will be developed that may identify the most important variants among the hundreds of thousands surveyed, direct the replication of positive findings and localise the causative genes. The major outcome of POLYGENE will hopefully be a catalogue of genetic variants that affect the risk and progression of prostate and breast cancer.

Problem

Cancer is caused by a complex interplay between genetic and environmental factors. The relative contribution of genetics to the risk of common cancers varies considerably between cancer types, with breast and prostate cancer considered to be the two cancers where genetics play the largest role. Using the combined twin registries of Sweden, Denmark and Finland, Lichtenstein *et al.* concluded that 42% of prostate cancer risk could be explained by genetic factors, which is higher than the genetic risk estimates for any other cancer type (1). The same study estimated that 32% of breast cancer risk could be contributed by genetics. Other researchers have pointed out that twin studies can yield only a lower limit of the proportion attributable to genetics, suggesting that genetic factors may contribute much

more than would be suggested by twin studies (2, 3). The analysis of cancer occurrence in relatives of cancer patients also lends strong evidence for genetic factors that increase the risk of cancer. A study of the familiarity of all cancers diagnosed in Iceland from 1955 to 2002 shows that relatives of breast and prostate cancer patients have a significantly increased risk of developing these cancers, as well as a subset of other cancers (4).

The identification of cancer susceptibility genes of large effect, such as BRCA1 and BRCA2, raised hopes that systematic studies of large cancer-prone families would lead to the identification of the major genetic determinants of cancer. However, the field of breast cancer genetics, where most of the effort has been focused, has not seen discoveries of highly penetrant breast cancer genes since BRCA1 and BRCA2 were found. This is despite the fact that population-based epidemiology studies have shown that only 15–20% of familial breast cancers occur in families carrying mutations in the BRCA1 or BRCA2 genes. In prostate cancer, mutations in three genes have been found in families with multiple cases of prostate cancer but the prevalence of these mutations is too low to be of clinical relevance at the population level. A collaborative effort for the mapping of additional and more relevant hereditary prostate cancer genes has been started (The International Consortium for Prostate Cancer Genetics, ICPG) in order to overcome two of the most important problems in linkage analyses: small sample sizes and genetic heterogeneity. This consortium, with linkage data available from more than 1200 HPC families, has come up with some new candidate genomic regions but candidate genes have not been identified yet (5).

“To increase our likelihood of success, we have chosen to study two European populations with different genetic backgrounds, the homogeneous population of Iceland and the more mixed population of the Netherlands.”

Accumulating evidence suggests that most of genetic cancer risk is due to multiple risk alleles, where the risk associated with each individual allele is small to moderate. Importantly, as the variants may interact in a multiplicative or supermultiplicative (epistatic) way, an individual with several susceptibility alleles may be at significant risk. If the polygenic model is correct, linkage analysis of families with multiple cases of breast or prostate cancer will not suffice because it lacks power to detect the low-penetrance risk variants. In addition, people with relatively high inherited risk of cancer may not have a significant family history and can thereby not be accounted for. The association approach (i.e. the comparison of genetic variants between cases and controls) is the current method of choice for identifying low-penetrance genes (risk ratio 1.1–2.0). Genotyping technologies are now available that allow 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.

Association studies, however, have several weaknesses which have been amply demonstrated by the large number of reports of genetic associations that have failed to be confirmed. The most prominent problems include biological and phenotypic complexity, population stratification and lack of power. Population stratification is a form of confounding which results from an unobserved different genetic background of the cases and controls. This may lead to many genetic variants being significantly different between cases and controls even though they have nothing to do with the pathogenesis of the disease. The huge amount of data generated by genome-wide association studies present a major statistical challenge with pressing need for improvements in analysis methods that can sort spurious associations from real ones and probe possible interactions between genetic variants. Last, but not least, genetic associations found in one population need to be replicated independently in one or more different populations in order to weed out false positive signals. These considerations were all taken into account in the structure of POLYGENE.

Aims

The major aim of POLYGENE is to identify common genetic variants that confer risk of prostate or breast cancer. This task will be approached at two levels; first, by large-scale genotyping of prostate and breast cancer patients and controls and simultaneously by improving statistical and computational methods for analysis of the resulting data.

To increase our likelihood of success, we have chosen to study two European populations with different genetic backgrounds, the homogeneous population of Iceland and the more mixed population of the Netherlands. The Icelandic population has proven very useful in genetic studies of complex diseases; notably, a genetic prostate cancer risk variant was recently identified in Iceland and subsequently replicated in several different populations (6). Good disease records are available which can help reduce phenotypic complexity and information on the substructure of the population can be taken into account to minimise stratification (7). In total, we will genotype over 2,000 prostate and 2,000 breast cancer patients from the two populations and compare these to the genotypes of 5,700 controls. By using a large group of controls, we increase our power for detecting disease variants with a modest effect. Thus, we will have more than 70% power to detect a risk variant of 10% allelic frequency or higher with a risk ratio of 1.4. For prostate cancer, we will have 63% power to detect a risk-variant of the same frequency and risk ratio. The collection of extensive clinical information allows the definition of different clinical phenotypes which is of crucial importance in genetic studies of complex diseases like breast and prostate cancer.

Analysis of association in a marker by marker fashion is often inefficient and the second aim of POLYGENE is to further develop existing methods that simultaneously use information from several markers (multipoint methods) (8, 9). Furthermore, methods that search for interactions between sets of unlinked markers which individually have only small effects will be developed. The importance of interactions is presently debatable, but the POLYGENE data set provides an excellent test case. Finally, since the statistical design consists of several stages, methods that can guide the selection of subsets of markers to be screened in a larger population in subsequent stages of the investigation will be developed and refined (10).

Potential results

If the goals of POLYGENE are achieved, it will lead to the identification of common genetic variants that have moderate but significant influence on the risk of developing prostate or breast cancer. Furthermore, it will be possible to explore the association between these variants and various clinical parameters, such as disease severity, progression and outcome. The project funds the establishment of a population-based biorepository of prostate and breast cancer samples in Nijmegen that can be used for additional studies of the genetics and epidemiology of these diseases. The computational and statistical methods developed in POLYGENE will be augmented by existing knowledge of genes that are known or suspected to play a role in cancer. However, the methods will also have general applicability in other large genome-wide association studies on complex human diseases.



Figure 2: Blood samples from over ten thousand cancer patients and controls will be analysed in the POLYGENE study.

Potential applications

The definition of all genetic risk factors for prostate and breast cancer is of high importance for several reasons. First, it will increase our knowledge of how minor differences in normal cellular physiology can



over time lead to cancer in these organs. Second, this information may unveil new cellular mechanisms that confer predisposition to prostate cancer and possibly other cancer types as well. Last, but not least, increased knowledge about the functional aspects of prostate and breast cancer predisposition may lead to the development of new methods for diagnosing and treating the disease.

Prostate and breast cancer are predominantly diseases of older age, their incidence is rising (especially considering the ageing population in Europe) and the social impact and healthcare costs are escalating. Early diagnosis and treatment are key factors in determining survival of breast and prostate cancer patients. Although there is no evidence yet for prostate cancer, regular screening of individuals at high risk for breast cancer greatly improves disease outcome. However, in both diseases, screening programmes have led to overdiagnosis (up to 50% in prostate cancer (11) and excessive treatment of localised lesions that might never progress to symptomatic cancer. Measures aimed at focusing the screening effort towards individuals at highest risk for developing prostate or breast cancer are needed. The definition of genetic risk variants may help focus screening towards individuals that have the highest risk of developing these cancers.

POLYGENE Partners:

1. deCODE Genetics, Reykjavik, Iceland. PI: Dr. Thorunn Rafnar
2. Dept. of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. PI: Prof. Bart Kiemeneij
3. Department of Statistics, Oxford University, Cambridge, UK. PI: Prof. Jotun Hein
4. Dept. of Genetics and Ecology, Inst. of Biology, University of Aarhus. PI: Assoc. Prof. Mikkel H. Schierup

References

1. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343(2):78-85.
2. Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. *Nat Genet* 2000;26(4):411-4.
3. Risch N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol Biomarkers Prev* 2001;10(7):733-41.
4. Amundadottir LT, Thorvaldsson S, Gudbjartsson DF, Sulem P, Kristjansson K, Arnason S, et al. Cancer as a Complex Phenotype: Pattern of Cancer Distribution within and beyond the Nuclear Family. *PLoS Med* 2004;1(3):e65 Epub 2004 Dec 28.
5. Xu J, Dimitrov L, Chang BL, Adams TS, Turner AR, Meyers DA, et al. A combined genomewide linkage scan of 1,233 families for prostate cancer-susceptibility genes conducted by the international consortium for prostate cancer genetics. *Am J Hum Genet* 2005;77(2):219-29.
6. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet* 2006;7:7.
7. Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K. An Icelandic example of the impact of population structure on association studies. *Nat Genet* 2005;37(1):90-5 Epub 2004 Dec 19.
8. Mailund T, Schierup MH, Pedersen CN, Madsen JN, Hein J, Schauer L. GeneRecon--a coalescent based tool for fine-scale association mapping. *Bioinformatics* 2006.
9. Mailund T, Schierup MH, Pedersen CN, Mechlenborg PJ, Madsen JN, Schauer L. CoaSim: a flexible environment for simulating genetic data under coalescent models. *BMC Bioinformatics* 2005;6:252.
10. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006;38(2):209-13 Epub 2006 Jan 15.
11. Draisma G, Boer R, Otto SJ, van der Cruisjes IW, Damhuis RA, Schroder FH, et al. Lead times and overdiagnosis due to prostate-specific antigen screening: estimates from the European Randomised Study of Screening for Prostate Cancer. *J Natl Cancer Inst* 2003;95(12):868-78.