

# The genetics of prostate cancer

Thorunn Rafnar, Patrick Sulem and Julius Gudmundsson  
*deCODE Genetics, Reykjavik, Iceland*

## **Abstract**

Prostate cancer is the most common incident cancer and the third most common cause of cancer related mortality in males in Europe, with over 89,000 deaths estimated in the year 2008 (9.3% of all cancer deaths in males). Apart from age, the strongest risk factor for prostate cancer is family history of the disease, highlighting the importance of genetics in disease development. Recently, considerable progress has been achieved in the identification of genetic factors that associate with prostate cancer risk, raising hopes that this knowledge may be used both to gain insights into the pathogenesis of the disease and to develop tools for risk assessment.

**Key words:** Prostate cancer, genetic factors, risk assessment

### Genetic factors play a major role in prostate cancer

Unlike most other cancer types, prostate cancer does not have any strong environmental risk factors. However, over 50 years ago, a report on the familial clustering of prostate cancer suggested that genetic factors play a role in the development of the disease (1). Subsequent studies on families with multiple cases of prostate cancer, as well as case control and cohort studies, all supported that the risk of prostate cancer was partly inherited. In a landmark study published in the year 2000, information from the combined twin registries of Sweden, Denmark and Finland were used to estimate that 42% of the variance in prostate cancer risk could be explained by genetic factors (2). This fraction was higher than the genetic risk estimates for any other cancer type except for thyroid cancer. Subsequently, other investigators pointed out that twin studies can yield only a lower limit of the proportion attributable to genetics, suggesting that genetic factors may contribute more than 60% of the variance in prostate cancer risk (3). Analysis of cancer in relatives of cancer patients also has consistently shown that first degree relatives of men with prostate cancer (sons, fathers and brothers) have approximately twofold risk of developing the disease relative to the general population (4, 5). This risk is even higher in family members of prostate cancer cases that are diagnosed at an early age (<60) or where there are multiple cases in the family (6).

The evidence that genetic factors play a major role in prostate cancer resulted in a large effort being put into studies of families with multiple prostate cancer cases in the hope of finding the causative genetic changes. However, strong susceptibility genes for prostate cancer have proven difficult to find. Notably, the International Consortium for Prostate Cancer Genetics (ICPCG) analyzed data from over 1200 prostate cancer families and although several genetic regions were highlighted as possible prostate cancer risk loci, it has not been possible to pinpoint the causative genes in any of these regions (7). This is mainly due to the fact that the candidate regions are large and, until now, it has been prohibitively expensive to re-sequence the DNA in order to find the pathogenic mutation. It should be noted that the breast cancer genes, *BRCA1* and *BRCA2*, have been shown to confer a moderate increase in risk of prostate cancer (8, 9).

Collectively, the results of family studies support the notion that no single susceptibility gene is likely to explain a large proportion of highly familial or early

onset prostate cancer. Rather, most of the inherited prostate cancer risk is due to multiple, moderate genetic risk variants. Each such variant would be expected to carry a small increase in risk but if man carried many such variants, his risk of prostate cancer would be high. This model is also consistent with the fact that in spite of the large genetic component of prostate cancer, the great majority of cases are sporadic, i.e. without notable family history.

### Genome-wide association studies for finding prostate cancer risk variants.

Statistical modeling soon demonstrated that family studies are not useful to find genetic variants that impart low to moderate risk on prostate cancer. Instead, genetic association studies, where the genomes of a large number of cases and controls are compared, are the method of choice. In brief, this methodology takes advantage of the fact that two unrelated people share about 99.9% of their DNA sequences, however, the remaining 0.1% can vary between individuals. This is the fraction that makes a person unique and determines attributes such as differences in risk of diseases. The most common variation in the human genome is the single nucleotide polymorphism (SNP). A SNP is a DNA sequence variation that occurs when a single nucleotide (A, T, C or G) in the genome sequence is altered. For example a SNP might change the DNA sequence AAGGC to ACGGC. SNPs occur every 100 to 300 bases along the 3-billion-base human genome and in most cases they do not affect the function of genes. However, because they are spread over the entire genome, they can serve as molecular markers for pinpointing an association between a particular region of the genome and disease risk.

Until recently, only a small fraction of SNPs in the genome was known and testing their association with disease risk was difficult. However, two major breakthroughs revolutionized genetic studies of common diseases. First, the international HapMap project ([www.hapmap.org](http://www.hapmap.org)) was launched with the aim of identifying and mapping all variants in the human genome. Secondly, major advances in genotyping technologies made it possible to genotype a single individual for hundreds of thousands of SNPs in a single experiment. By genotyping a large number of cases and controls, it became possible to pinpoint SNPs that differed significantly in frequencies between the two groups and therefore suggested the presence of a genetic risk

factor. For example, if a particular SNP variant is present in 22% of the chromosomes in cases and 15% of the chromosomes in controls, and this difference is statistically significant, we can assume that the genomic region where the SNP is located contains a genetic variation that increases risk of the disease.

The methodology of comparing the frequencies of SNPs all over the genome between cases and controls, termed genome-wide association studies (GWAS), has revolutionized genetic studies on common diseases such as prostate cancer. In the last 3 years GWAS of thousands of prostate cancer cases and controls have yielded close to 30 SNP variants that associate with risk of prostate cancer, some of which map to the same region (reviewed in (10)). Twenty five of these variants are listed in Table 1. As can be seen, almost all the risk variants are common and range in frequency between 3% and 85%. Also, each variant infers a small increase in risk as measured by the allelic odds ratio (OR). An OR is a measure of the risk of disease in an individual who has inherited the variant compared to the risk of a person that does not carry the variant. All individuals carry 2 copies of each chromosome, one inherited from the father and one from the mother. As an example, a man who has inherited one copy of the risk variant rs6983267(G) has a 1.27 fold risk of developing the disease relative to a person that does not carry the variant (non-carrier). The genetic model that best fits the observed inheritance is the multiplicative model which assumes that both copies carry the same increase in risk. Consequently, if the same individual has inherited two copies of the variant, his risk is  $1.27 \times 1.27 = 1.61$  times the risk of the non-carrier. No interaction has been observed between the currently known prostate cancer variants, therefore, their combined effect is the product of each effect.

### Calculation of genetic risk

The genetic risk measure is generally expressed as the relative risk (RR) of developing the disease compared to the general population. In Europe, the lifetime risk of prostate cancer is approximately 10%. If the 25 variants in Table 1 are combined in a multi-variant analysis using the multiplicative model, the cumulative risk of prostate cancer among men in the top 10% of the genetic risk distribution is more than 2-fold greater than the population average risk. For these individuals, this corresponds to an absolute risk of over 20% of being diagnosed with prostate cancer.

Conversely, males in the lowest 10% of the risk distribution have less than 0.45 fold risk compared to the population average. Thus, their risk of developing prostate cancer before the age of 70 is less than 4.5%. These risk estimates are largely independent of family history. Hence, the estimated risk for an individual will be increased if history of prostate cancer is known among close relatives (11). It should be emphasized that for the majority of the male population, the genetic risk estimates will not deviate greatly from the population average and the greatest benefit of genetic testing will be for the small percentage of males with the highest genetic risk.

As demonstrated above, the genetic risk calculations are based on population averages and care is needed when generalizing results from one population to another. In our example above, the average population incidence of prostate cancer in Europe was used to calculate RR based on genotype. However, prostate cancer incidence varies greatly between geographical regions within Europe, the difference mostly being caused by the prostate specific antigen (PSA) blood test which is widely used for prostate cancer screening in Western Europe. Within Europe, the incidence of prostate cancer varies almost 8 fold between regions, with highest estimated age-standardized incidence rates per 100,000 recorded in Ireland (183.1), France (178.7) and Norway (172.7) and the lowest incidence rates estimated in the Republic of Moldova (23.3), Ukraine (27.7) and Albania (30.7) (12). Importantly, the great majority of genetic studies of prostate cancer have been performed in PSA-screened populations (Western Europe and USA) and the resulting risk estimates have been calculated based on the resulting high lifetime risks. This fact demonstrates very clearly the importance of establishing risk estimates for genetic variants in each population. However, it should be noted that where variants found in one population have been tested in another population of the same ethnicity, little or no heterogeneity has been observed, i.e. the variants seem to confer similar risk in the two populations.

In Romania, the age-standardized incidence rate for prostate cancer was estimated at 32.0 per 100,000 in the year 2008, reflecting the relatively low level of PSA screening in the country (12). This translates into a lifetime risk of 2-4%. Currently, genetic analysis of Romanian prostate cancer cases and controls is being conducted through an EU-funded project, Promark ([www.promark-fp7.eu](http://www.promark-fp7.eu)). Within ProMark, researchers at

the University of Medicine and Pharmacy Carol Davila and Institute of Public Health in Bucharest, along with partner institutions in the Netherlands, UK and Iceland, are collecting samples and information from prostate cancer cases and controls for genetic studies. In addition to discovering new prostate cancer risk variants, it is hoped that analysis of this material will give a clear picture of the contribution of genetic factors to prostate cancer risk in the Romanian population.

### **Insights into the molecular mechanisms of prostate cancer development**

The genetic studies performed so far have yielded limited information about the possible molecular mechanisms of prostate cancer development. Notably, close to half of the variants in Table 1 are located in regions that have no known genes close-by, leaving little information as to the mechanism of increased disease risk. Even in the cases where the variants are located close to known genes, there are limited clues as to the actual function. In general, there is no evidence to suggest that the variants listed are the actual causative variants themselves, rather they serve as markers for the region and much work remains in pinpointing and verifying the actual underlying pathogenic variations.

The variants at chromosome 8q24 present a very intriguing case. This region has been shown to contain at least 6 independent risk variants for prostate cancer (13-17) and also separate risk variants for several other cancer types, including cancers of the colon and rectum, breast, urinary bladder and ovaries (18-21). The gene closest to the region containing the variants is *c-MYC*, the gene that is most often rearranged in malignant tumors. Thus, has been suggested that the cancer risk variants may facilitate rearrangement of *c-MYC* and that this mechanism is tissue specific, therefore, different variants are associated with different cancer types. However, this hypothesis needs to be tested by functional assays.

The prostate cancer risk variant at on chromosome 17q12 is also notable for several reasons. The variant is located in an intron of the *TCF2* gene, which encodes a transcription factor playing an important role in embryonic development of the kidney, pancreas and liver. Mutations in *TCF2* are associated with a particular form of diabetes (maturity onset diabetes of the young-MODY) and it has been shown that the same variant that increases risk of prostate cancer carries a slightly

reduced risk of developing adult-onset diabetes (22). This inverse association between prostate cancer and type 2 diabetes has been observed in several epidemiological studies, including a meta-analysis which estimated the relative risk of prostate cancer to be 0.84 among diabetes patients (23). Clearly, much work remains to be done in order to elucidate how variation in the *TCF2* gene affects these different diseases.

### **Future directions**

It has been estimated that the currently known genetic prostate cancer risk variants explain less than 30% of the familial risk; therefore, it is clear that more variants remain to be found. Furthermore, as mentioned above, there is little evidence to suggest that the SNPs identified are the actual causative variants, rather, it is more likely that they only tag the region where the causative variation may be found. The next step will be to scrutinize the genomic regions by direct sequencing in order to find all the variants in the region and test their association to the disease. Currently, large publicly funded projects are underway that aim to re-sequence a large number of individuals to provide a comprehensive resource on human genetic variation. The first of these projects, the 1000 Genomes Project, aims to find most genetic variants that have frequencies of at least 1% in the populations studied ([www.1000genomes.org](http://www.1000genomes.org)). It is hoped that this effort will facilitate the discovery of additional variants that influence risk of prostate cancer. As more genetic risk factors for prostate cancer are uncovered, it will be possible to refine the risk model described above, enabling the active monitoring of individuals with the highest genetic risk.

Another issue that needs to be resolved is the over-diagnosis of prostate cancer in PSA-screened populations. Most men with screen-detected prostate cancer have localized disease at diagnosis. Many of these men may harbor clinically insignificant disease that will not impact their quality of life or life expectancy while in other men prostate cancer will progress to an advanced or lethal disease if left alone. Because of these uncertainties, and the lack of reliable prognostic markers, most men with localized disease are subjected to radical prostatectomy or radiotherapy which can adversely impact their urinary and sexual health. The reasons why some cancers are more aggressive than others remain poorly understood and the need for diagnostic resources to help differentiate between the two is im-

SNP	Chromosome	Gene in region	Risk Allele	Non Risk Allele	Risk allele frequency in controls	Allelic OR
rs2710646	2p15	EHBP1	A	C	0.19	1.15
rs12621278	2q31.1	ITGA6	A	G	0.94	1.33
rs2660753	3p12	Intergenic	T	C	0.11	1.06
rs10934853	3q21.3	Intergenic	A	C	0.28	1.12
rs7679673	4q24	TET2	C	A	0.55	1.10
rs401681	5p15.33	TERT	C	T	0.56	1.07
rs9364554	6q25.3	SLC22A3	T	C	0.28	1.14
rs6465657	7q21.3	LMTK2	C	T	0.47	1.12
rs10486567	7p15	JAZF1	G	A	0.76	1.09
rs1512268	8p21.2	NKX3.1	A	G	0.45	1.18
rs1447295	8.q24.21	Intergenic	A	C	0.09	1.53
rs6983267	8q24.21	Intergenic	G	T	0.50	1.27
rs16901979	8q24.21	Intergenic	A	G	0.03	1.66
rs16902104	8q24.21	Intergenic	T	C	0.15	1.21
rs445114	8q24.21	Intergenic	C	T	0.64	1.14
rs10086908	8q24.21	Intergenic	T	C	0.70	1.16
rs10993994	10q11.23	MSMB	T	C	0.40	1.24
rs11228565	11q13	Intergenic	A	G	0.20	1.23
rs7127900	11p15	Intergenic	A	G	0.20	1.22
rs4430796	17q12	TCF2	A	G	0.49	1.22
rs1859962	17q24.3	Intergenic	G	T	0.46	1.20
rs27356839	19q13.33	KLK2/KLK3	G	A	0.85	1.12
rs8102476	19q13.2	Intergenic	C	T	0.54	1.12
rs5759167	22q13	Intergenic	G	T	0.53	1.16
rs5945572	Xp11.22	NUDT11	A	G	0.35	1.23

**Table 1. SNP variants associate with risk of prostate cancer.** Shown are the SNP name, chromosomal location, the reported gene, allele conferring risk and the alternate allele, the frequency of the risk allele in the population and the allelic odds ratio (OR).

mense. The possibility that inherited genetic variation can affect not only the risk of developing prostate cancer, but also be associated with a particular course of disease remains to be explored. Most of the prostate cancer variants discovered to date show a similar association with both forms of the disease; however, two of the variants in Table 1 (rs2710646 and rs1447295) show a slightly stronger association with a more severe disease (i.e. stage T3 or T4 and/or Gleason score >6 and/or metastatic disease) than with the more indolent form (Stage T1 or T2 and Gleason score <6). If genetic variants that associate strongly with aggressive forms of prostate cancer can be found, they could be of use in directing therapy of early stage prostate cancer. In

this regard, genetic studies in populations where PSA screening has not become common are extremely useful since a larger fraction of prostate cancer cases are diagnosed with clinically significant disease.

As new prostate cancer variants are found, they will need to be subjected to functional studies aimed at elucidating the actual mechanism by which the variation affects prostate cancer development. Identifying the causative genetic factors is pivotal in order to understand the pathogenesis of the disease. Once the molecular mechanisms that affect genetic prostate cancer risk have been uncovered, we will have moved one step closer to effective prevention and treatment.

## References

1. Morganti G GL, Cresseri A, Arrigoni G, Lovati G. Clinico-statistical and genetic research on neoplasms of the prostate. *Acta Genet Stat Med*, 6: 304-305, 1956.
2. Lichtenstein P, Holm NV, Verkasalo PK, 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*, 343: 78-85., 2000.
3. Risch N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol Biomarkers Prev*, 10: 733-41., 2001.
4. Cannon-Albright LA, Thomas A, Goldgar DE, et al. Familiality of cancer in Utah. *Cancer Res*, 54: 2378-85, 1994.
5. Amundadottir LT, Thorvaldsson S, Gudbjartsson DF, et al. Cancer as a complex phenotype: pattern of cancer distribution within and beyond the nuclear family. *PLoS Med*, 1: e65, 2004.
6. Hemminki K, Granstrom C, Sundquist J, and Bermejo JL. The updated Swedish family-cancer database used to assess familial risks of prostate cancer during rapidly increasing incidence. *Hered Cancer Clin Pract*, 4: 186-92, 2006.
7. Christensen GB, Baffoe-Bonnie AB, George A, et al. Genome-wide linkage analysis of 1,233 prostate cancer pedigrees from the International Consortium for Prostate Cancer Genetics using novel sumLINK and sumLOD analyses. *Prostate*, 70: 735-44, 2010.
8. Thompson D, and Easton DF. Cancer Incidence in BRCA1 mutation carriers. *J Natl Cancer Inst*, 94: 1358-65, 2002.
9. Thompson D, and Easton D. Variation in cancer risks, by mutation position, in BRCA2 mutation carriers. *Am J Hum Genet*, 68: 410-9, 2001.
10. Ioannidis JP, Castaldi P, and Evangelou E. A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. *J Natl Cancer Inst*, 102: 846-58, 2010.
11. Hsu FC, Sun J, Zhu Y, et al. Comparison of two methods for estimating absolute risk of prostate cancer based on single nucleotide polymorphisms and family history. *Cancer Epidemiol Biomarkers Prev*, 19: 1083-8, 2010.
12. Ferlay J, Parkin DM, and Steliarova-Foucher E. Estimates of cancer incidence and mortality in Europe in 2008. *Eur J Cancer*, 46: 765-81, 2010.
13. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet*, 41: 1122-6, 2009.
14. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*, 39: 645-9, 2007.
15. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet*, 39: 638-44, 2007.
16. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet*, 39: 631-7, 2007.
17. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet*, 38: 652-8, 2006.
18. Kiemeny LA, Thorlacius S, Sulem P, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet*, 40: 1307-12, 2008.
19. Ghossaini M, Song H, Koessler T, et al. Multiple Loci With Different Cancer Specificities Within the 8q24 Gene Desert. *J Natl Cancer Inst*, 2008.
20. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet*, 39: 984-8, 2007.
21. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447: 1087-93, 2007.
22. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet*, 39: 977-83, 2007.
23. Kasper JS, and Giovannucci E. A meta-analysis of diabetes mellitus and the risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev*, 15: 2056-62, 2006.