

Cancer as a Complex Phenotype: Pattern of Cancer Distribution within and beyond the Nuclear Family

Laufey T. Amundadottir^{1*}, Sverrir Thorvaldsson¹, Daniel F. Gudbjartsson¹, Patrick Sulem¹, Kristleifur Kristjansson¹, Sigurdur Arnason^{1□}, Jeffrey R. Gulcher¹, Johannes Bjornsson², Augustine Kong¹, Unnur Thorsteinsdottir¹, Kari Stefansson^{1*}

1 deCODE Genetics, Reykjavik, Iceland, **2** National University Hospital, Reykjavik, Iceland

Competing Interests: JB and SA have declared that no competing interests exist. LTA, ST, DFG, PS, KK, JRG, AK, UT, KS have stock in deCODE Genetics as well as equity interests.

Author Contributions: LTA, ST, DFG, PS, KK, SA, JRG, JB, AK, UT, and KS designed the study. LTA, ST, DFG, PS, AK, and UT analyzed the data. LTA, ST, DFG, PS, KK, SA, JRG, JB, AK, UT, and KS contributed to writing the paper.

Academic Editor: Cathryn Lewis, Guy's King's and St. Thomas' School of Medicine, United Kingdom

Citation: Amundadottir LT, Thorvaldsson S, Gudbjartsson DF, Sulem P, Kristjansson K, et al. (2004) Cancer as a complex phenotype: Pattern of cancer distribution within and beyond the nuclear family. *PLoS Med* 1(3): e65.

Received: August 27, 2004
Accepted: October 27, 2004
Published: December 28, 2004

DOI:
10.1371/journal.pmed.0010065

Copyright: © 2004 Amundadottir et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviations: ICR, Icelandic Cancer Registry; KC, kinship coefficient; RR, risk ratio

*To whom correspondence should be addressed. E-mail: laufey.amundadottir@decode.is (LTA), kari.stefansson@decode.is (KS)

□¹ Current address: Keflavik County Hospital, Keflavik, Iceland

ABSTRACT

Background

The contribution of low-penetrant susceptibility variants to cancer is not clear. With the aim of searching for genetic factors that contribute to cancer at one or more sites in the body, we have analyzed familial aggregation of cancer in extended families based on all cancer cases diagnosed in Iceland over almost half a century.

Methods and Findings

We have estimated risk ratios (RRs) of cancer for first- and up to fifth-degree relatives both within and between all types of cancers diagnosed in Iceland from 1955 to 2002 by linking patient information from the Icelandic Cancer Registry to an extensive genealogical database, containing all living Icelanders and most of their ancestors since the settlement of Iceland.

We evaluated the significance of the familial clustering for each relationship separately, all relationships combined (first- to fifth-degree relatives) and for close (first- and second-degree) and distant (third- to fifth-degree) relatives. Most cancer sites demonstrate a significantly increased RR for the same cancer, beyond the nuclear family. Significantly increased familial clustering between different cancer sites is also documented in both close and distant relatives. Some of these associations have been suggested previously but others not.

Conclusion

We conclude that genetic factors are involved in the etiology of many cancers and that these factors are in some cases shared by different cancer sites. However, a significantly increased RR conferred upon mates of patients with cancer at some sites indicates that shared environment or nonrandom mating for certain risk factors also play a role in the familial clustering of cancer. Our results indicate that cancer is a complex, often non-site-specific disease for which increased risk extends beyond the nuclear family.

Introduction

Highly penetrant susceptibility variants explain only a small fraction of the genetics of all cancer cases. As an example, mutations in the *BRCA1* and *BRCA2* genes account for around 2%–3% of all breast cancer cases [1,2], although more prevalent founder mutations in these genes can explain up to about 10% of the disease in some populations [3,4,5,6,7]. However, the role of genetics in the remaining breast cancer cases and the majority of other cancers is not clear.

Family studies have given insight into the contribution of genetic and environmental factors to the etiology of cancer. Case-control, registry- and population-based studies have evaluated familial clustering using either risk ratio (RR) estimations for relatives of cancer patients, or kinship coefficient (KC) estimations for cancer patients. The largest of these studies, utilizing either the Utah Population and Cancer Registry Database or the Swedish Family-Cancer Database, have demonstrated excess familial clustering at practically all cancer sites in the body [8,9,10,11,12]. Most of these studies have been able to evaluate familial clustering only within the nuclear family, thus making it more difficult to separate the roles of shared environmental and genetic factors in the familial aggregation of cancers. However, in one of these studies [12], in which familial clustering was evaluated for more distant relatives, significant clustering outside the nuclear family was demonstrated for a number of cancer sites. Extended familial clustering has also been reported in studies of individual cancers [13,14,15,16,17,18,19,20,21,22].

Twin studies have also evaluated the role of genes versus environment in cancer susceptibility. The largest study involved close to 45,000 twins from Denmark, Sweden, and Finland where the RR of same type of cancer was calculated for individuals with affected twins and compared to those without an affected twin [23]. The authors concluded that for the majority of cancer sites only a limited part of the risk could be explained by heritable factors. Exceptions to this were cancers of the prostate, colon and breast.

In addition to well documented familial clustering for the majority of individual cancers, aggregation of different types of cancers in families has also been observed. Reports have been published on the results of systematic analysis of the aggregation of different cancers using the Utah Population and Cancer Registry Database [24,25]. In addition to demonstrating excess familial clustering for most cancer sites, these studies also indicate that an excess is also shared by different cancer sites. In these studies, cancer clustering was evaluated either by calculating the RR for first-degree relatives or KC between different cancer sites. While distant relationships contributed to the overall calculation of KC, their contributions were not evaluated separately in the studies between cancer sites, hence making it more difficult to separate the effects of genetic and environmental factors in these studies.

We have studied a registry of all cancer cases diagnosed in Iceland from 1 January 1955 to 31 December 2002, with the aim of searching for evidence of genetic factors both at individual cancer sites and those shared by different sites. By cross-referencing cancer prevalence in relatives of cases with the aid of a comprehensive nationwide genealogy database, we have estimated RR separately for first- to fifth-degree relatives of all cancer patients diagnosed in Iceland over 48 y.

We demonstrate here an increased cancer risk in relatives outside the nuclear family (third- to fifth-degree relatives) for many cancer sites. These relatives share significant genetic makeup but are less likely to share environmental factors beyond those shared by the general population, indicating that genetic factors may be involved. By applying the analysis across different cancer sites we also demonstrate shared familiarity between certain cancer sites both in close and distant relatives. These results suggest that cancer can be considered a broad phenotype with shared genetic factors crossing different cancer sites. That is, the difference between cancers at various sites may in part be the consequence of variable expressivity of the same cancer-predisposing genes.

Methods

This study was approved by the National Bioethics Committee of Iceland, the Data Protection Authority of Iceland, and the Icelandic Cancer Society. All names of patients listed in the Icelandic Cancer Registry (ICR) and the genealogic database were encrypted through a process approved by the National Bioethics Committee and the Data Protection Authority before being analyzed [26].

Cancer Registry

The ICR of the Icelandic Cancer Society is a carefully constructed database containing practically complete records of all cancer cases diagnosed in Iceland after 1 January 1955 [27]. Records are received at the ICR from all hospitals in the country that treat cancer patients, and the very few not listed are individuals who are diagnosed while living abroad. Furthermore, the records are verified by a continuous interaction between the ICR and Icelandic hospitals and clinicians. Approximately 95% of cases are histologically verified [28]. In the present study we used International Classification of Disease version 10 codes as the basis for defining phenotypes. A total of 81 unique phenotypes (sites) were analyzed. In this paper we present data from 27 sites with more than 200 cases each (Table 1). For the 48 years (1 January 1955 to 31 December 2002) a total of 32,534 individuals were found in our genealogy database. Cancer incidence in Iceland is comparable to the Nordic countries of Europe and is detailed in [27].

Genealogic Database

deCODE Genetics has built a computerized genealogy database of more than 687,500 individuals [29,30]. The names of all 288,000 Icelanders currently alive and a large proportion of all Icelanders who have ever lived in the country are in the database. The genealogy of the entered individuals is recorded from multiple sources including church records and censuses from previous centuries and, more recently, from published genealogy books. The genealogy database is quite complete from the 18th century on, thus allowing quite distant relationships to be traced accurately.

Mates are defined as individuals of the opposite sex who have one or more children in common, regardless of marital status.

Calculations of RRs

The RR for relatives is a measure of the risk of disease for a relative of an affected person compared to the risk in the population as a whole. More precisely, for a given relation-

Table 1. RR Estimates of Cancer at the Same Site for Relatives and Mates for Cancer Sites with 200 or More Cases

Cancer Site	ICD10	Number Affected	RR [90% Confidence Interval]										Combined <i>p</i> Value for Relatives		
			1° Relatives	2° Relatives	3° Relatives	4° Relatives	5° Relatives	6° Relatives	Mates	1°-5° Relatives	1° and 2° Relatives	3°-5° Relatives			
Breast	C50	3,812	2.02 [1.88,2.15]	1.36 [1.27,1.43]	1.21 [1.15,1.25]	1.13 [1.08,1.16]	1.05 [1.01,1.06]	2.02 [0.83,5.68]	< 0.0001 ^a	< 0.0001	< 0.0001				
Prostate	C61	3,380	1.89 [1.75,2.01]	1.36 [1.26,1.45]	1.19 [1.13,1.24]	1.10 [1.07,1.12]	1.04 [1.01,1.08]	na ^b	< 0.0001 ^a	< 0.0001	< 0.0001				
Lung	C34	2,904	2.00 [1.83,2.16]	1.39 [1.26,1.50]	1.10 [1.03,1.16]	1.02 [0.97,1.07]	1.04 [1.01,1.08]	1.68 [1.35,2.06]	< 0.0001 ^a	< 0.0001	< 0.00135				
Stomach	C16	2,890	1.90 [1.74,2.05]	1.38 [1.25,1.48]	1.23 [1.16,1.29]	1.15 [1.09,1.19]	1.09 [1.04,1.11]	1.72 [1.33,2.18]	< 0.0001 ^a	< 0.0001	< 0.0001				
Colon	C18	2,224	1.92 [1.71,2.14]	1.26 [1.12,1.40]	1.16 [1.07,1.24]	1.05 [0.98,1.10]	1.06 [1.01,1.09]	1.46 [1.03,2.08]	< 0.0001 ^a	< 0.0001	0.00029				
Bladder	C67	1,384	1.68 [1.39,2.05]	1.19 [0.98,1.43]	1.26 [1.13,1.41]	1.08 [0.99,1.19]	1.03 [0.97,1.10]	0.41 [0.09,1.23]	< 0.0001 ^a	0.00002	0.00033				
Kidney	C64	1,227	2.30 [1.89,2.80]	1.31 [1.06,1.57]	1.32 [1.15,1.48]	1.15 [1.03,1.26]	1.03 [0.95,1.09]	1.20 [0.52,2.30]	< 0.0001 ^a	< 0.0001	0.00034				
Thyroid	C73	957	3.02 [2.33,3.85]	1.64 [1.29,2.02]	1.30 [1.07,1.51]	1.08 [0.93,1.21]	1.11 [1.00,1.19]	1.12 [0.23,3.93]	< 0.0001 ^a	< 0.0001	0.00506				
Pancreas	C25	930	2.33 [1.83,2.96]	1.28 [0.97,1.66]	1.09 [0.90,1.29]	1.06 [0.93,1.21]	0.99 [0.91,1.08]	1.29 [0.53,4.08]	< 0.0001 ^a	< 0.0001	0.20677				
Ovary	C56	906	2.01 [1.48,2.70]	1.62 [1.27,2.05]	1.24 [1.03,1.47]	1.18 [1.07,1.29]	0.92 [0.83,1.01]	na ^b	< 0.0001 ^a	< 0.0001	0.04804				
Non-melanoma skin	C44	781	1.46 [0.97,2.07]	0.96 [0.64,1.35]	1.44 [1.18,1.71]	1.04 [0.87,1.20]	1.12 [1.00,1.23]	2.16 [0.85,5.73]	0.00727	0.17809	0.00233				
Rectum	C20	767	1.68 [1.17,2.42]	1.63 [1.22,2.15]	1.31 [1.08,1.61]	0.90 [0.76,1.06]	1.00 [0.90,1.12]	1.63 [0.52,4.83]	0.00051 ^a	0.00034	0.15784				
Endometrium	C54	753	1.86 [1.31,2.62]	1.26 [0.90,1.74]	1.57 [1.30,1.89]	1.05 [0.89,1.24]	1.12 [1.00,1.24]	na ^b	< 0.0001 ^a	0.00298	0.00018				
Cervix uteri	C53	724	1.74 [1.12,2.73]	1.71 [1.24,2.26]	1.10 [0.86,1.40]	1.16 [0.98,1.36]	0.99 [0.88,1.12]	na ^b	0.00053 ^a	0.00079	0.11505				
Brain	C71	663	1.41 [0.74,2.40]	1.10 [0.69,1.62]	0.83 [0.59,1.10]	1.31 [1.09,1.54]	0.99 [0.85,1.11]	1.22 [0.20,4.45]	0.16029	0.18892	0.30278				
Melanoma skin	C43	618	1.86 [1.06,3.35]	1.61 [1.08,2.31]	1.23 [0.90,1.58]	1.00 [0.78,1.19]	0.92 [0.77,1.04]	1.52 [0.19,4.90]	0.02082	0.06066	0.52545				
Esophagus	C15	535	2.09 [1.30,3.31]	1.62 [1.07,2.38]	1.04 [0.74,1.41]	1.14 [0.91,1.40]	1.07 [0.92,1.23]	2.40 [0.66,5.94]	0.0015 ^a	0.00134	0.16738				
Diffuse NHL ^c	C83	422	1.46 [0.65,4.20]	1.19 [0.61,2.11]	1.42 [0.98,2.02]	0.96 [0.70,1.30]	0.92 [0.74,1.13]	2.07 [0.13,5.70]	0.12743	0.17125	0.24982				
Multiple myeloma	C90	391	2.66 [1.54,6.10]	1.32 [0.69,2.62]	0.57 [0.29,0.99]	1.15 [0.85,1.55]	1.10 [0.89,1.34]	3.65 [0.80,7.84]	0.03144	0.00743	0.58679				
Lymphoid leukemia	C91	368	4.56 [2.52,8.86]	1.53 [0.74,2.80]	1.57 [1.01,2.31]	1.25 [0.89,1.67]	1.06 [0.82,1.29]	0/281 ^d	0.00004 ^a	0.00024	0.03474				
Myeloid leukemia	C92	342	0.99 [0.23,3.37]	0.98 [0.38,2.01]	1.02 [0.57,1.70]	0.94 [0.62,1.35]	0.99 [0.76,1.26]	0/317 ^d	0.54272	0.48934	0.56875				
Meninges	C70	291	3.15 [1.50,6.84]	1.36 [0.54,4.03]	1.60 [0.94,2.70]	1.47 [1.00,2.06]	1.20 [0.89,1.53]	0/289 ^d	0.00141 ^a	0.01853	0.56875				
Liver	C22	257	2.00 [0.71,5.08]	1.28 [0.41,3.89]	0.64 [0.24,1.35]	1.14 [0.70,1.74]	1.02 [0.73,1.36]	0/263 ^d	0.00141 ^a	0.13249	0.60908				
Lip	C00	244	5.04 [2.75,9.52]	0.49 [0.03,2.37]	0.91 [0.40,1.73]	0.63 [0.30,1.12]	1.41 [1.06,1.80]	0/220 ^d	0.02346	0.0091	0.48947				
Hodgkin's disease	C81	239	3.27 [1.19,7.01]	1.85 [0.61,4.84]	0.79 [0.26,1.74]	0.94 [0.50,1.57]	0.79 [0.50,1.16]	0/191 ^d	0.03064	0.03064	0.81722				
Testis	C62	222	3.52 [1.18,7.37]	0.99 [0.09,3.37]	1.81 [0.89,4.78]	1.86 [1.16,3.04]	1.21 [0.83,1.73]	na ^b	0.11431	0.09278	0.00647				
Larynx	C32	208	3.02 [1.06,6.65]	0.65 [0.05,2.71]	0.27 [0.02,0.90]	1.30 [0.75,2.33]	0.83 [0.53,1.28]	0/244 ^d	0.01076	0.14907	0.76038				

Shown are the estimated RRs with 90% confidence interval for first- to fifth-degree (1°-5°) relatives and mates of the 27 cancer sites with ≥200 cases, in bold when the 90% CI does not include 1.00, which corresponds to one-sided *p* < 0.05. Also shown are combined *p* values to evaluate the significance of the increased RR for all relatives (first- to fifth-degree) and for close (first- and second-degree) and distant relatives (third- to fifth-degree). *p* values nominally significant at the 0.05 level are shown in bold.

^a Nominal *p* values that remained significant after Bonferroni correction for the 27 individual tests (*p* < 0.00185).
^b na, not applicable (sex-specific cancers).
^c NHL, non-Hodgkin's lymphoma.
^d Number of mates with cancer/total number of mates for each cancer site.
 ICD10, International Classification of Disease version 10.
 DOI: 10.1371/journal.pmed.0010065.t001

ship the RR for disease B in the relatives of probands with disease A is defined as

$$RR = \frac{P(R_B|P_A)}{P(R_B)}, \tag{1}$$

where P_A denotes the event that the proband is affected with disease A, and R_B denotes the event that the relative is affected with disease B. Note that disease A and disease B can be the same in this definition which applies when estimating RR at individual cancer sites. Using Bayes' rule it can be shown that for symmetric relationships, RR is the same if the roles of A and B are switched, i.e., the RR for disease A in the relatives of probands with disease B is the same as the described above. In this study we always chose the less common phenotype as the proband when estimating RR.

A basic underlying assumption in our estimation of RR is that of conditional independence of ascertainment, or censoring, (O_{RB} and O_{PA} are the events that the relative and proband are observed with diseases A and B, respectively):

$$P(O_{RB}, O_{PA}|P_A, R_B) = P(O_{RB}|R_B)P(O_{PA}|P_A). \tag{2}$$

Some form of this assumption is used by most methods estimating RR [31].

Obtaining valid estimates of the RR is not always straightforward, since the method of ascertainment of affected cases critically affects the estimation, and inappropriate estimators can lead to bias or inflated estimates [32]. The use of a nationwide registry of patients covering close to five decades decreases much of the potential sampling bias. However, the ascertainment of the ICR depends on the year of birth of individuals. This dependence needs to be addressed when estimating the RR.

The approach chosen here is to estimate the RR for a number of subpopulations, where prevalence is reasonably constant, and combine them into a single estimate of RR for the full population. Let r be the number of relatives of probands, counting multiple times individuals who are relatives of multiple probands [33], let a be the number of relatives of probands that are affected (again possibly counting the same individual more than once), let n be the size of the population, and finally let x be the number of affected individuals in the population. If $P(R_B)$ and $P(R_B | P_A)$ can reasonably be assumed to be constant in the population, then x/n and a/r , respectively, are estimates of these probabilities. Given these estimates, RR is consistently estimated by

$$\frac{a/r}{x/n}. \tag{3}$$

Assuming the population can be split into N subpopulations, such that within each subpopulation $P(R_B)$ and $P(R_B | P_A)$ can be assumed to be constant, although they may vary between subpopulations, and assuming furthermore that RR is the same in all the subpopulations, then the RR is consistently estimated by a convex combination of the estimates for the subpopulations. We selected weights for the combination such that the efficiency of the estimator was at maximum for RR equal to one. Making the simplifying assumption that the relatives are independent (while this assumption is not entirely correct, it affects only efficiency, not validity), the optimal weight for group j is

$$w_j = \frac{x_j r_j}{n_j - x_j} \tag{4}$$

(this is the inverse of the variance of the estimate for RR in subpopulation j), where a , r , x , and n are defined as above, restricted to the subpopulation j . Note that probands are not restricted to the subpopulation. Given these weights, our estimate of RR is

$$\frac{\sum_{j=1}^N w_j \frac{a_j/r_j}{x_j/n_j}}{\sum_{j=1}^N w_j} = \frac{\sum_{j=1}^N \frac{a_j n_j}{n_j - x_j}}{\sum_{j=1}^N \frac{r_j x_j}{n_j - x_j}}. \tag{5}$$

In this study, the most relevant variations in $P(R_B)$ and $P(R_B | P_A)$ stem from time-dependent censoring of affected status and sex-specific differences. Hence, we have stratified the population so that j runs over groups of people of the same sex and born in the same 5-y periods. For a fixed year-of-birth stratum, there is censoring of affected status (missing data) based on year of onset because of the fact that records cover only the period 1955–2002. Our approach is designed to address this type of missing data. As an example of the stratification, the breast cancer patients in our analysis were born in the years 1865 to 1970 (5-y strata), yielding 35 subpopulations, 22 for female patients, but only 13 for male, as this cancer is rare for males.

To assess the significance of the RR obtained for a given group of patients, we compared their observed values with the RR computed for up to 100,000 independently drawn and matched groups of control individuals. Each patient was matched to a single control individual in each control group. The control individuals were drawn at random from the genealogic database with the conditions that they had the same year of birth, the same sex, and the same number of ancestors recorded in the database at five generations back as the matched patients. Empirical p values can be calculated using the control groups; thus, a p value of 0.05 for the RR would indicate that 5% of the matched control groups had values as large as or larger than that for the patient's relatives or mates. The number of control groups required to obtain a fixed accuracy of the empirical p values is inversely proportional to the p value. We therefore selected the number of control groups generated adaptively up to a maximum of 100,000. When none of the values computed for the maximum number of control groups were larger than the observed value for the patient's relatives and mates, we report the p value as being less than 0.00001. Using a variance-stabilizing square-root transform, an approximate confidence interval may be constructed based on the distribution of RR for control groups [33].

As another test for significance of RR between cancer sites, we used combined estimators for risk in relatives of degree 1 and 2 together, degrees 3, 4, and 5 together, and degrees 1 through 5 together. If RR_d is the RR for relatives of degree d , then $RR_d - 1$ is known to decrease proportional to 2^{-d} as d increases for a monogenetic single variant or additive disease models, and faster for more complex disease models [34]. With the estimate of RR_d denoted by \hat{RR}_d , we then chose a test statistic of the form

$$\sum_d w_d \frac{(\hat{RR}_d - 1)}{2^d}, \tag{6}$$

with d summed over the relevant degrees. For RR_d close to one, the variance of the estimate \hat{RR}_d is inversely proportional to the number of relatives of degree d for the proband. Based on the Icelandic genealogy for the cancers being studied here, the number of relatives is proportional to γ^d , where the value of γ quantifies how the number of relatives grows with each degree of relatedness to the proband. This factor γ varies only slightly between cancers and is on average 2.46. Minimizing the variance of the test statistic in equation 6 with respect to the weights yields the statistic

$$\sum_d 1.23^d (\hat{RR}_d - 1). \quad (7)$$

As above, the choice of weights and the form of the statistic affects only power, not validity. To assess significance, the observed value of the statistic was compared to its value for multiple matched control groups as described above.

Although our evaluations of familial clustering, for both close and distant relatives, are based on RR, an alternative approach based on comparing KCs among patients and among controls exists [12,24,25]. The two approaches are closely related, and our choice was made in part because relative risk is a less technical concept and its application to genetic counseling more direct. Also, the relationship between relative risk and the power to map disease genes by linkage analysis has been thoroughly investigated [34,35].

Results

We have studied the familial clustering of cancer by estimating RR for first- and up to fifth-degree relatives both within and between all cancer sites. Here we present results for 27 sites that contain 200 or more cancer cases each, based on International Classification of Disease version 10 codes. These 27 sites represent 89% of all cancer cases in the ICR.

Risk Estimations for Cancer at Same Site

A significantly increased RR to first-degree relatives of patients with cancer was seen for 22 of the 27 cancer sites (Table 1). Among the statistically significant RRs, the highest estimates were for lymphoid leukemia, Hodgkin's disease, and cancer of the thyroid, meninges, lip, testis, and larynx (RR above three). These cancers, except for thyroid cancer, were among the least prevalent sites (200–400 cases), as reflected in the large standard deviation of the RR estimates (Table 1). First-degree relatives of individuals with breast, lung, kidney, pancreatic, ovarian, and esophageal cancer and multiple myeloma, had between 2- and 3-fold increased risk of developing the same cancer.

The medians of the estimated RR values for the 27 sites in first- to fifth-degree relatives were 2.00, 1.32, 1.21, 1.10, and 1.04, respectively.

Combined p -values incorporating the increased risk for first- to fifth-degree relatives identified 21 sites being significant at a nominal level of 0.05. Sixteen of those sites remained significant after Bonferroni adjustment for the 27 individual tests (p value < 0.00185) (Table 1). To discriminate between familial clustering in close and distant relatives, combined p values were also calculated for first- and second-degree relatives on one hand and for third- to fifth-degree relatives on the other hand (Table 1). Fourteen sites were nominally significant for the distant relationships (third- to fifth-degree relatives) of which eight were significant after

Bonferroni adjustment. These eight sites were all within the group of 16 sites demonstrating significant familial clustering in all relationships.

The RR for developing cancer at the same site was also estimated for mates of cancer patients at 22 out of the 27 individual sites. The remaining five sites are sex-specific and calculations thus not applicable. For seven rare cancer sites, affected mates were not observed, corresponding to a RR of zero. Only lung, stomach, and colon cancer were characterized by significantly increased RR values in mates (Table 1).

Risk Estimations between Cancer Sites

We calculated RR between all cancer sites for first- and up to fifth-degree relatives and mates (results for the 27 largest sites are shown in Table S1). As done for the individual cancer sites, p values were calculated for all (first- to fifth-degree), close (first- and second-degree), and distant (third- to fifth-degree) relationships. Figure 1 shows a diagram representing 20 pairs of cancer sites that associate with a combined p value, significant at a level of 1×10^{-4} , for first- to fifth-degree relationships. This level was significant at the 0.05 level after Bonferroni adjustment for the 351 tests (number of unique pairs of cancers). The strength of the distant familiarity (i.e., the p value for third- to fifth-degree relatives) between these pairs of cancers is represented by the thickness of the lines joining sites in Figure 1.

In total, 17 cancer sites were involved in 20 significant pairs of sites (Figure 1). Stomach and prostate cancer were involved in most pairs, seven and six pairs, respectively, followed by colon, ovarian, and cervical cancer, each involved in three pairs. The estimated RRs for the 20 pairs are between 1.1 and 1.7 for first-degree relatives and between 1.1 and 1.5 for second-degree relatives (Figure 1; Table S1). The highest RRs in first-degree relatives between cancer sites were seen for esophagus–cervix, with a RR of 1.74, pancreas–ovary, with a RR of 1.66, and colon–rectum, with a RR of 1.64.

All of the 20 pairs shown in Figure 1 were nominally significant (p value < 0.05) for distant relationships, of which nine were significant at the 0.001 level. In the latter group, prostate, rectum, stomach, and cervical cancers each appeared in two pairs, and colon cancer in three.

Discussion

In this study we have comprehensively analyzed familial aggregation of cancer cases in a whole nation, both within and between pairs of cancer sites. The completeness of our genealogy database allows us to accurately trace distant relationships, which we believe is unique to this study. Linking the ICR to our nationwide genealogy database thus has made it possible to uncover distant familial connections between cancer cases, and reach beyond shared environmental factors to identify individual and combined cancer sites with the strongest genetic influences. Furthermore, even though the genetic effect decreases with more distant relationships, the sample sizes used to estimate familiarity are dramatically larger for the distant relationships than for the closer ones. This compensates to some extent for the lower effect and adds considerable statistical power to the study.

In this paper we restrict the presentation and discussion to the most significant findings. However, we provide results for all pairs of 27 cancer sites in Table S1, as a resource for other researchers interested in the familiarity of specific cancers.

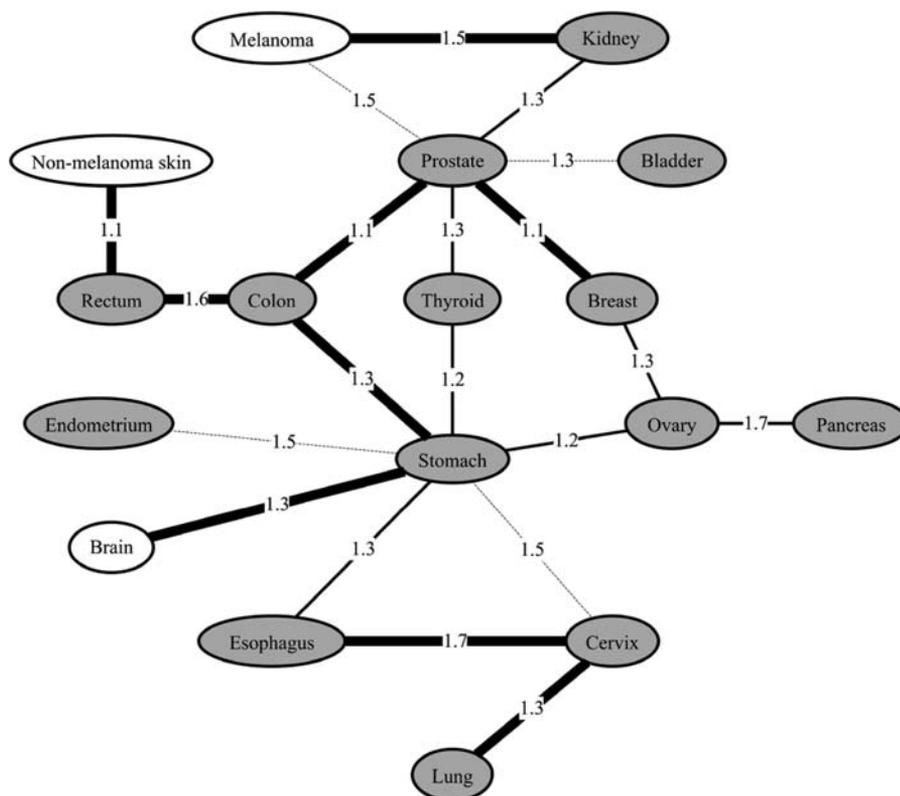


Figure 1. A Schematic Representation of Cancer Pairs Demonstrating Significant Familial Aggregation

Cancer pairs that demonstrate significant familial co-clustering (first- to fifth-degree relatives) at the 0.05 level after adjustment for multiple testing (nominal p value $< 1 \times 10^{-4}$) are joined by lines. The thickness of the lines joining the pairs are based on nominal p values corresponding to the significance of the familiarity in distant relatives (third to fifth degree): bold, $p \leq 0.001$; solid, $p \leq 0.01$; and dashed, $p \leq 0.05$. The number on the lines joining each pair indicates the cross-cancer RR in first-degree relatives. Shaded ovals correspond to individual cancer sites that were significant for the combined group of first- to fifth-degree relatives at the 0.05 level after Bonferroni adjustment (see Table 1). DOI: 10.1371/journal.pmed.0010065.g001

The largest population-based studies reported to date, evaluating familial clustering within the same cancer site, are from Utah and Sweden [8,9,10]. These studies report RR values for first-degree relatives [36] that are comparable to those presented here for first-degree relatives. For example, the median RRs for the occurrence of the same cancer in first-degree relatives were 2.15, 1.86, and 2.00 for the Utah, the Sweden, and our study, respectively. Also, RR values in first-degree relatives ranged between 1.5 and 3.0 for the majority of sites, i.e., 69%, 82%, and 60%, in Utah, Sweden, and this study, respectively.

As seen in Utah and Sweden, high RR values were found in this study for multiple myeloma, lymphoid leukemia, and thyroid, testicular, and laryngeal cancer. The RR for thyroid cancer in first-degree relatives was much higher in Utah and Sweden (8.48 and 9.51) than in Iceland (3.02). One possible explanation of the lower RR may be the high incidence of thyroid cancer in Iceland, due to an excess of the papillary subtype [18,37], which is not a part of the multiple endocrine neoplasia syndromes.

The cancer sites showing the highest RR for first-degree relatives tend to be among the rarer sites. There are two potential reasons why rare tumors tend to show higher RRs than common cancers. Being common, the baseline frequency is not low and that creates a bound on how large the RR can be. Also, common cancers are expected to be genetically complex, whereas it is more likely for a rare tumor to be closer to a Mendelian trait, caused by rare alleles with high penetrances.

Most individual cancer sites, or 16 out of the 27 studied here, showed familiarity as evidenced by significant p values (after adjustment for multiple testing) for the combined group of first- to fifth-degree relatives. Furthermore, eight of these 16 sites remained significant even after exclusion of the

first- and second-degree relatives (after adjustment for multiple testing). The majority of the 16 significant cancer sites are among the sites of the most prevalent cancers, indicating that we may lack power to detect extended familiarity for the less prevalent cancer sites. Indeed the median number of cases per cancer site was 943 for the 16 significant sites compared to 342 for the non-significant sites. Nevertheless, significant familial clustering (first- to fifth-degree relatives) is seen for some of the less prevalent sites, i.e., lymphoid leukemia and esophagus and meningeal cancer.

The largest cancer twin study reported to date [23] documented significant heritability of prostate (42%), colorectal (35%), and breast cancer (27%) and provided suggestive evidence for limited heritability of leukemia and stomach, lung, pancreas, ovarian, and bladder cancer. All of these cancer sites showed significant familial clustering in our study. However, when the analysis was restricted to distant relatives, lymphoid leukemia, pancreatic, and ovarian cancer were no longer significant. Although close to 45,000 pairs of twins were included in the study (of which 10,803 had been diagnosed with cancer), the study clearly lacked statistical power to detect the effects of heritable factors for the less prevalent cancer sites.

A significantly increased risk of the same cancer was seen in mates only for individuals diagnosed with stomach, lung, or colon cancer. These results are in accordance with previous reports, including Swedish population-based studies, except for colon cancer [38,39,40,41]. Environmental factors in adult life (including lifestyle and infections) or nonrandom mating could explain the higher risk of these cancer types in mates. The RR was not significant or not observed in mates for other sites.

We also assessed the significance of familial clustering

between cancer sites by calculating combined p values corresponding to the increased risk for first- to fifth-degree relationships. With this method, we detected 17 cancers that linked into 20 pairs of sites that were significant after adjustment for multiple testing. Stomach and prostate cancer appeared more frequently in the pairs than other cancer types, followed by colon, ovarian, and cervical cancer. We emphasize again, as with the same-cancer calculations, that we might lack power to connect rare cancers to other cancer sites. This possibility is highlighted by the fact that the 17 cancers in the significant pairs are the most prevalent cancer sites in Iceland.

Some connections seen here between cancer sites may be partly explained by known high-risk genes involved in heritable syndromes. Thus, mutations in genes associated with hereditary nonpolyposis colorectal cancers could explain a part of the risk shared between stomach, colon, rectal, and endometrial cancer, and possibly brain and ovarian cancer [42,43]. In a similar manner, mutations in *BRCA1* and *BRCA2* may explain in part the cluster seen between prostate, breast, ovarian, and possibly pancreatic cancer [20,44,45,46]. Other known but even rarer cancer syndromes are likely to explain only a handful of cases.

Undiscovered genetic factors could contribute to some connections seen here to a much greater extent than the known susceptibility factors. Although these could include unknown high-risk susceptibility genes, they are more likely multiple genetic variants, each conferring small to moderate risk.

Familial clusters were identified between cancer sites, both in close and distant relatives, that do not correspond to known cancer syndromes. These include lung, esophageal, cervical, and stomach cancer, which, interestingly, have been associated with environmental rather than genetic factors. One explanation for this excess familiarity between these cancer sites is an interaction of genetic susceptibility factors with environmental carcinogens (e.g., tobacco and diet) or infectious agents. Thus, the same environmental factor could interact with the same genetic susceptibility factor or factors to induce different cancers (i.e., smoking in lung and cervical cancer). Alternatively, different environmental factors could interact with the same genetic susceptibility factor or factors to increase the risk for different cancers (i.e., smoking in lung cancer and human papilloma virus in cervical cancer).

Hormone-related cancers form another risk cluster. Thus, shared genetic susceptibility factors could directly influence the hormonal metabolism to induce breast, prostate, thyroid, or ovarian cancer in carriers. Alternatively, shared genetic factors could interact with dietary factors to induce aggregation of cancers at these sites in related individuals. A significantly increased risk of breast, prostate, cervical, and non-melanoma skin cancer was recently reported in first-degree relatives of early-onset breast cancer patients from Sweden that tested negative for *BRCA1* and *BRCA2* mutations [47]. Our data support the notion that unknown susceptibility variants that increase the risk of breast and prostate cancer and melanoma remain to be characterized.

Two more groups of cancers with shared risk were identified that each include sites that share the same developmental progenitors: the prostate, kidney, and bladder are sites derived from the nephrogenic ridge while colon, rectum, and stomach are derived from the primitive gut tube. Therefore, the sites in each group may share risk alleles that regulate embryonic development, which can later play a role in oncogenesis.

Interestingly, three cancer sites/types, non-melanoma skin, brain, and melanoma, that do not have significant same-cancer familial clustering demonstrate significant cross-cancer familial clustering with more prevalent cancer sites, i.e., rectum, stomach, and kidney cancers, respectively.

Previous reports systematically evaluating the significance of co-clustering of cancer pairs in families have utilized the Utah Population Database. In these studies lip and prostate cancers appear to associate most frequently with other cancer sites. The same is true for prostate cancer in our study, whereas lip cancer does not significantly associate with any other cancer sites. This can at least in part be explained by the difference in age-standardized incidence rates for lip cancer in Iceland and Utah (Iceland 1.1 and Utah 2.4) [48]. In contrast, stomach cancer associates with seven other cancer sites out of the 20 significant pairs in our study, but only three other sites in the Utah study. Of the 20 cancer pairs that significantly associate in our study, eight concur with the Utah studies.

Because the increased cross-site RR extends beyond the nuclear family, shared genetic factors may contribute to the risk of more than one cancer type. This suggests that cancer could be considered a broad phenotype with shared genetic factors across cancer sites. Therefore cancer should in certain cases be studied in a broader context than previously done. Combining multiple cancers that show increased cross-site RR may serve to increase the power of linkage and case-control studies. Our results also have implications for genetic counseling and imply that the focus of attention should broaden to the history of multiple cancer types in relatives within and outside the nuclear family. These results also suggest the utility of comparing expression profiles and in vitro biological processes across the cancers that we have identified as sharing genetic risk. The isolation of cancer predisposition genes with broad effects may define new rate-limiting pathways that can be used to search for drug targets for a more focused treatment with fewer side effects but with utility across multiple cancers.

Supporting Information

Table S1. Cross-Site RR Estimates for Relatives and Mates of Patients Diagnosed with Cancers at 27 Sites with 200 Cases or More Found at DOI: 10.1371/journal.pmed.0010065.st001 (1.9 MB DOC).

Accession Numbers

The LocusLink (<http://www.ncbi.nlm.nih.gov/projects/LocusLink/>) accession numbers for the genes discussed in this paper are *BRCA1* (LocusLink ID 672) and *BRCA2* (LocusLink ID 675).

Acknowledgments

We thank the ICR for providing lists of cancer patients. We also thank Hersir Sigurgeirsson and Snæbjörn Gunnsteinsson for assistance with data processing, and Jon Thor Bergthorsson, Simon N. Stacey, and Julius Gudmundsson for their critical reading of the manuscript and helpful comments. The study was funded by deCODE Genetics. ■

References

1. Anglian Breast Cancer Study Group (2000) Prevalence and penetrance of *BRCA1* and *BRCA2* mutations in a population-based series of breast cancer cases. *Anglian Breast Cancer Study Group. Br J Cancer* 83: 1301–1308.
2. Syrjakoski K, Vahteristo P, Eerola H, Tamminen A, Kivinummi K, et al. (2000) Population-based study of *BRCA1* and *BRCA2* mutations in 1035 unselected Finnish breast cancer patients. *J Natl Cancer Inst* 92: 1529–1531.
3. King MC, Marks JH, Mandell JB (2003) Breast and ovarian cancer risks due to inherited mutations in *BRCA1* and *BRCA2*. *Science* 302: 643–646.
4. Johannesdottir G, Gudmundsson J, Bergthorsson JT, Arason A, Agnarsson

- BA, et al. (1996) High prevalence of the 999del5 mutation in Icelandic breast and ovarian cancer patients. *Cancer Res* 56: 3663–3665.
5. Thorlacius S, Sigurdsson S, Bjarnadottir H, Olafsdottir G, Jonasson JG, et al. (1997) Study of a single BRCA2 mutation with high carrier frequency in a small population. *Am J Hum Genet* 60: 1079–1084.
 6. Hartge P, Struewing JP, Wacholder S, Brody LC, Tucker MA (1999) The prevalence of common BRCA1 and BRCA2 mutations among Ashkenazi Jews. *Am J Hum Genet* 64: 963–970.
 7. Thorlacius S, Struewing JP, Hartge P, Olafsdottir GH, Sigvaldason H, et al. (1998) Population-based study of risk of breast cancer in carriers of BRCA2 mutation. *Lancet* 352: 1337–1339.
 8. Dong C, Hemminki K (2001) Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families. *Int J Cancer* 92: 144–150.
 9. Vaitinen P, Hemminki K (1999) Familial cancer risks in offspring from discordant parental cancers. *Int J Cancer* 81: 12–19.
 10. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH (1994) Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 86: 1600–1608.
 11. Czene K, Lichtenstein P, Hemminki K (2002) Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 99: 260–266.
 12. Cannon-Albright LA, Thomas A, Goldgar DE, Gholami K, Rowe K, et al. (1994) Familiality of cancer in Utah. *Cancer Res* 54: 2378–2385.
 13. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA (1997) Family history and the risk of breast cancer: A systematic review and meta-analysis. *Int J Cancer* 71: 800–809.
 14. Zogas A, Gildea M, Cohen P, Bringman D, Taylor TH, et al. (2000) Cancer risk estimates for family members of a population-based family registry for breast and ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 9: 103–111.
 15. Tulinius H, Egilsson V, Olafsdottir GH, Sigvaldason H (1992) Risk of prostate, ovarian, and endometrial cancer among relatives of women with breast cancer. *BMJ* 305: 855–857.
 16. Tulinius H, Olafsdottir GH, Sigvaldason H, Tryggvadottir L, Bjarnadottir K (1994) Neoplastic diseases in families of breast cancer patients. *J Med Genet* 31: 618–621.
 17. Tulinius H, Sigvaldason H, Olafsdottir G, Tryggvadottir L, Bjarnadottir K (1999) Breast cancer incidence and familiarity in Iceland during 75 years from 1921 to 1995. *J Med Genet* 36: 103–107.
 18. Hrafnkelsson J, Tulinius H, Jonasson JG, Sigvaldason H (2001) Familial non-medullary thyroid cancer in Iceland. *J Med Genet* 38: 189–191.
 19. Gudbjartsson T, Jonasdottir TJ, Thoroddsen A, Einarsson GV, Jonsdottir GM, et al. (2002) A population-based familial aggregation analysis indicates genetic contribution in a majority of renal cell carcinomas. *Int J Cancer* 100: 476–479.
 20. Tulinius H, Olafsdottir GH, Sigvaldason H, Arason A, Barkardottir RB, et al. (2002) The effect of a single BRCA2 mutation on cancer in Iceland. *J Med Genet* 39: 457–462.
 21. Imsland AK, Eldon BJ, Arinbjarnarson S, Egilsson V, Tulinius H, et al. (2002) Genetic epidemiological aspects of gastric cancer in Iceland. *J Am Coll Surg* 195: 181–186.
 22. Eldon BJ, Jonsson E, Tomasson J, Tryggvadottir L, Tulinius H (2003) Familial risk of prostate cancer in Iceland. *BJU Int* 92: 915–919.
 23. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, et al. (2000) 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.
 24. Thomas A, Cannon-Albright L, Bansal A, Skolnick MH (1999) Familial associations between cancer sites. *Comput Biomed Res* 32: 517–529.
 25. Gholami K, Thomas A (1994) A linear time algorithm for calculation of multiple pairwise kinship coefficients and the genetic index of familiarity. *Comput Biomed Res* 27: 342–350.
 26. Gulcher JR, Kristjansson K, Gudbjartsson H, Stefansson K (2000) Protection of privacy by third-party encryption in genetic research in Iceland. *Eur J Hum Genet* 8: 739–742.
 27. Moller B, Fekjaer H, Hakulinen T, Tryggvadottir L, Storm HH, et al. (2002) Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer Prev* 11: S1–S96.
 28. Jonasson JG, Tryggvadottir LT, Bjarnadottir K, Olafsdottir GH, Olafsdottir EJ, et al. (2002) Iceland. In: Parkin DM, SL Whelan SL, Ferlay J, Teppo L, Thomas DB, editors. *Cancer incidence in five continents, Volume VIII* (IARC Scientific Publications No. 143). Lyon: International Agency for Research on Cancer. pp. 354–355.
 29. Gulcher J, Stefansson K (1998) Population genomics: Laying the groundwork for genetic disease modeling and targeting. *Clin Chem Lab Med* 36: 523–527.
 30. Gulcher J, Kong A, Stefansson K (2001) The genealogic approach to human genetics of disease. *Cancer J* 7: 61–68.
 31. Wallace C, Clayton D (2003) Estimating the relative recurrence risk ratio using a global cross-ratio model. *Genet Epidemiol* 25: 293–302.
 32. Guo SW (1998) Inflation of sibling recurrence-risk ratio, due to ascertainment bias and/or overreporting. *Am J Hum Genet* 63: 252–258.
 33. Sveinbjornsdottir S, Hicks AA, Jonsson T, Petursson H, Guggmundsson G, et al. (2000) Familial aggregation of Parkinson's disease in Iceland. *N Engl J Med* 343: 1765–1770.
 34. Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 46: 222–228.
 35. Risch N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 46: 229–241.
 36. Risch N (2001) The genetic epidemiology of cancer: Interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol Biomarkers Prev* 10: 733–741.
 37. Jonasson JG, Hrafnkelsson J, Bjornsson J (1989) Tumours in Iceland. 11. Malignant tumours of the thyroid gland. A histological classification and epidemiological considerations. *Apmis* 97: 625–630.
 38. Hemminki K, Dong C, Vaitinen P (2001) Cancer risks to spouses and offspring in the Family-Cancer Database. *Genet Epidemiol* 20: 247–257.
 39. Hemminki K, Jiang Y (2002) Cancer risks among long-standing spouses. *Br J Cancer* 86: 1737–1740.
 40. Hemminki K, Chen B (2004) Familial risk for colorectal cancers are mainly due to heritable causes. *Cancer Epidemiol Biomarkers Prev* 13: 1253–1256.
 41. Mellemgaard A, Jensen OM, Lyng E (1989) Cancer incidence among spouses of patients with colorectal cancer. *Int J Cancer* 44: 225–228.
 42. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, et al. (2004) Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 96: 261–268.
 43. Oliveira Ferreira F, Napoli Ferreira CC, Rossi BM, Toshihiko Nakagawa W, Aguiar S Jr, et al. (2004) Frequency of extra-colonic tumors in hereditary nonpolyposis colorectal cancer (HNPCC) and familial colorectal cancer (FCC) Brazilian families: An analysis by a Brazilian hereditary colorectal cancer institutional registry. *Fam Cancer* 3: 41–47.
 44. Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE (1994) Risks of cancer in BRCA1-mutation carriers. *Breast Cancer Linkage Consortium. Lancet* 343: 692–695.
 45. Breast Cancer Linkage Consortium (1999) Cancer risks in BRCA2 mutation carriers. The Breast Cancer Linkage Consortium. *J Natl Cancer Inst* 91: 1310–1316.
 46. Baffoe-Bonnie AB, Kiemeny LA, Beaty TH, Bailey-Wilson JE, Schnell AH, et al. (2002) Segregation analysis of 389 Icelandic pedigrees with breast and prostate cancer. *Genet Epidemiol* 23: 349–363.
 47. Loman N, Bladstrom A, Johannsson O, Borg A, Olsson H (2003) Cancer incidence in relatives of a population-based set of cases of early-onset breast cancer with a known BRCA1 and BRCA2 mutation status. *Breast Cancer Res* 5: R175–R186.
 48. Parkin DM, Whelan SL, Ferlay J, Raymond L, Young J, editors (1997) *Cancer incidence in five continents, Volume VII* (IARC Scientific Publications No. 143). Lyon: International Agency for Research on Cancer. 1,240 p.

Patient Summary

Background Although a few cancers have a fairly simple genetic cause, most, especially the most common cancers, do not, and what makes one person rather than another develop cancer is not clear. One way of trying to work out how much genes rather than environment contribute to disease is to study large populations. One such population is the Icelandic nation: not only is detailed health information about individuals available, including information on cancer, but also very good genealogical information and a substantial amount of genetic data.

What Did the Study Find? Researchers examined all cancer records dating back to 1955 and then analyzed the chances of relatives and mates of these patients having cancer. They found that some cancers, especially rare ones, had a higher than baseline chance of occurring in relatives, but so did many common cancers, and for some cancers, the higher chances extended to quite distant relatives. In addition, the risk sometimes involved different cancer types.

What Does the Study Mean for Patients? Even for the highest risk cancers, the absolute increased risk for relatives remains very small. In addition, despite the large numbers of patients studied, the numbers of cancer cases are still not large enough to be completely certain of the results, apart from very common cancers, which had the lowest chance of occurring in relatives. So these results will not help doctors much at the present time in telling an individual patient what their risk is of getting cancer if a relative has it—but they will be useful for other researchers in knowing how to plan future studies to look at the underlying causes of cancer.

Where Can I Get More Information?

Icelandic Cancer Society: <http://www.krabb.is/cancer/>
 The United States National Cancer Institute's Cancer Information Service: <http://cis.nci.nih.gov/>
 CancerHelp UK, a free information service about cancer and cancer care: <http://www.cancerhelp.org.uk/>
 deCODE Genetics: <http://www.decode.com>