45. Hoyt, M. A., Totis, L. & Roberts, B. T. *S. cerevisiae* genes required for cell cycle arrest in response to loss of microtubule function. *Cell* **66**, 507–517 (1991).
46. Griffith, E. C., Licitra, E. J. & Liu, J. O. Yeast three-hybrid system for detecting ligand-receptor interactions. *Methods Enzymol.* **328**, 89–103 (2000).
47. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
48. Towbin, H. *et al.* Proteomics-based target identification: bengamides as a new class of methionine aminopeptidase inhibitors. *J. Biol. Chem.* **278**, 52964–52971 (2003).
49. Momparler, R. L. Cancer epigenetics. *Oncogene* **22**, 6479–6483 (2003).
50. Thiagalingam, S. *et al.* Histone deacetylases: unique players in shaping the epigenetic histone code. *Ann. NY Acad. Sci.* **983**, 84–100 (2003).
51. Luo, J. *et al.* Negative control of p53 by Sir2α promotes cell survival under stress. *Cell* **107**, 137–148 (2001).
52. Bereshchenko, O. R. *et al.* Acetylation inactivates the transcriptional repressor BCL6. *Nature Genet.* **32**, 606–613 (2002).
53. Frye, R. 'SIRT8' expressed in thyroid cancer is actually SIRT7. *Br. J. Cancer* **87**, 1479 (2002).
54. Panagopoulos, I. *et al.* Acute myeloid leukemia with inv(8)(p11q13). *Leuk. Lymphoma* **39**, 651–656 (2000).
55. Emerling, B. M. *et al. MLL5*, a homolog of *Drosophila* trithorax located within a segment of chromosome band 7q22 implicated in myeloid leukemia. *Oncogene* **21**, 4849–4854 (2002).
56. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629 (2002).
57. Keats, J. J. *et al.* In multiple myeloma, t(4;14)(p16;q32) is an adverse prognostic factor irrespective of FGFR3 expression. *Blood* **101**, 1520–1529 (2003).
58. Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12**, 1052–1058 (2002).
59. van Leeuwen, F. *et al.* Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109**, 745–756 (2002).

Acknowledgements
The authors wish to thank the members of the Simon group for helpful comments and members of the National Cancer Institute's Developmental Therapeutics Program for support in the NCI yeast screen.

Competing interests statement
The authors declare that they have no competing financial interests.

Online links

DATABASES
The following terms in this article are linked online to:
Entrez Gene:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
aurora-related kinase 1 | BAK | BAX | BCL2 | BCL6 | *BRCA1* | BUB1 | CDK4 | INK4A | KIP1 | p38 | p53 | PARP1 | *PTEN* | RB | SIRT1
OMIM:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
hereditary non-polyposis colon cancer

FURTHER INFORMATION
NCI Yeast Anticancer Drug Screen web site:
http://dtp.nci.nih.gov/yacds/index.html
Access to this interactive links box is free online.

OPINION

The Icelandic Cancer Project – a population-wide approach to studying cancer

Thorunn Rafnar, Steinunn Thorlacius, Eirikur Steingrimsson, Mikkel H. Schierup, Jesper N. Madsen, Violeta Calian, Bjarki J. Eldon, Thorvaldur Jonsson, Jotun Hein and Snorri S. Thorgeirsson

Cancer initiation and progression require a complex interaction of genetic, environmental and clinical factors. Most research, however, has been focused on only a narrow aspect of the disease process. Data generated by the Human Genome Project, as well as large-scale molecular analysis of tumours, have indicated that a more systematic approach, in which the biological information is integrated with clinical features, is warranted. There are many aspects of the Icelandic population that make it well suited for such a broad-based approach. The Icelandic Cancer Project was therefore initiated to build a population-based clinical genomics database and biobank that can be used to study cancer — from genetic predisposition to clinical outcome.

Cancer research covers many fields, encompassing genetics, cell biology, epidemiology, clinical science, drug development, pharmacogenomics and studies on tumour–host interactions such as metastasis and angiogenesis. Accordingly, researchers draw on resources that range from large epidemiological surveys and experimental models in whole animals to *in vitro* studies on isolated signal-transduction pathways and genetic polymorphisms. It can be difficult to correlate findings obtained in different systems, and often results from one experimental system are not reproduced in the next. Examples include mouse models of human tumour-suppressor mutations that produce quite different phenotypes in mice and men[1].

To complicate matters even further, the very nature of cancer is to accumulate genetic alterations and this leads to many genetic variants within specific cancers. Response to treatment varies according to the genetic alterations of the tumours, which can be determined by the genetic makeup of the affected individual. So, there is an obvious need for the establishment of a comprehensive collection of samples and data through which all aspects of the disease, including patient genotype, tumour biology and clinical features, can be studied.

The Icelandic Cancer Project

In 2001, a group of public and private entities in Iceland launched a large, multidisciplinary study on cancer, termed the Icelandic Cancer Project (ICP). The main aim of the ICP is to create a population-based clinical genomics database and biobank that can be used to study the genetics of cancer, probe the relationship between genotype and phenotype, and create tools for functional analysis of tumour development and drug response. In addition to the research conducted by the partners of the ICP, it is envisioned that the ICP resources can be used by outside researchers, public or private, through collaborative efforts.

All Icelandic patients with cancer and their relatives are invited to participate in the ICP, along with a control population, randomly selected from the National Registry. The structure of the programme and the different types of biological samples and data that are collected are depicted in FIG. 1 and BOX 1. The ICP will eventually include all 8,000 current cancer survivors in Iceland, plus the approximately 1,100 individuals who are diagnosed each year. Furthermore, archived samples and cancer information from over 24,000 deceased cancer patients can be accessed and linked to the project. The ICP has been approved by The National Bioethics Committee and The Privacy and Data Protection Authority in Iceland, and all samples and data are collected with written, informed consent.

The Icelandic public has shown a remarkably positive response to the project. So far, over 19,000 individuals (patients, relatives and controls) have entered the study; 92% of those invited have agreed to participate. Based on this participation rate and statistics from the Icelandic Cancer Registry, it can be estimated that about 10,000 patients with cancer will have entered the study by the end of 2005. Of those, approximately 2,200 will have been diagnosed with breast cancer and 1,200 with prostate cancer, the two most common cancers in women and men, respectively.

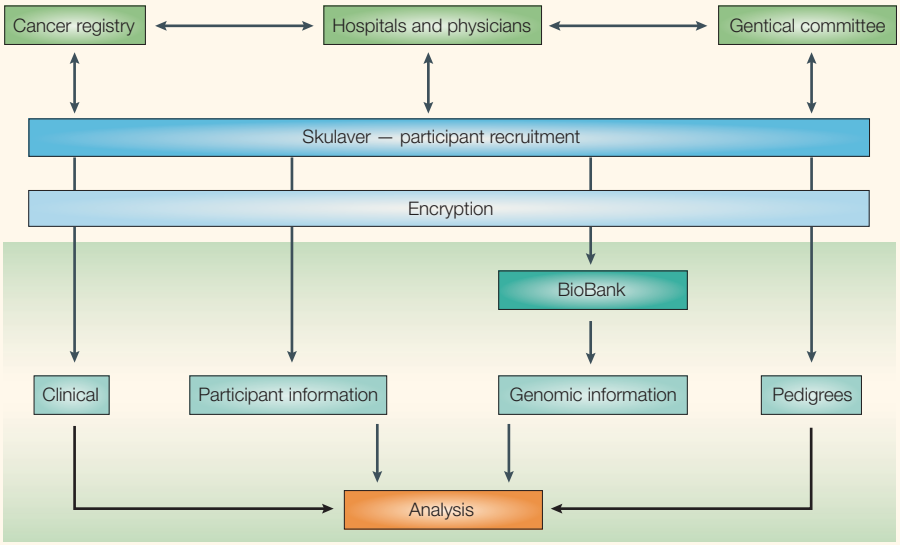

Figure 1 | **The structure of the Icelandic Cancer Project and the resulting biobank and database.** Participants in the ICP are recruited through the Skulaver patient-recruitment centre. Participants fill out a lifestyle questionnaire and a blood sample is collected. Skulaver also gathers clinical information about the patients' diseases from the Icelandic Cancer Registry, and from collaborating clinicians and hospitals. Genealogy information is collected from the Genetical Committee of the University of Iceland. All personal identifiers are encrypted. The patients' clinical characteristics, genomic (from the BioBank) and pedigree data are then analysed. The structure of the database and procedures of its operation are overseen by the Privacy and Data Protection Authority.

The working hypothesis of the ICP is that cancer is a group of related disorders that have common genetic causes and can be viewed as a single phenotype. Most known cancer-predisposing genes are associated with an increased risk of more than one type of cancer. This is exemplified in cancer syndromes such as the Li–Fraumeni syndrome[2], but is also evident for genes such as the hereditary non-polyposis colon cancer genes that, although identified by a strong association with colorectal cancer, predisposes mutation carriers to several other tumour types[3]. It is therefore likely that the same will be true for many of the cancer-predisposing genes still to be discovered — a promising prospect for the future of targeted molecular therapies and cancer-prevention strategies.

Why Iceland?

Several important features make the Icelandic population highly suitable for this broad-based approach to studying cancer. A national health-care system was established in 1915. Soon after, systematic collections of pathological samples and health-care records were established. Today, access to archived samples and medical records, dating back almost 100 years, can be granted to researchers who obtain the necessary permission. Lately, cancer care in Iceland has been largely centralized, which allows for consistency in diagnosis and tumour phenotyping, as well as in the types of treatment that patients receive. Information on diagnoses and pathology for all patients diagnosed with cancer since 1954 (over 32,000 cases total) is maintained by the Icelandic Cancer Registry. Some records extend even further back. For example, breast cancers are recorded back to the 1920s. Furthermore, details of births, marriages and deaths have been recorded since the seventeenth century. These records are now maintained in computerized form by the Genetical Committee, which provides the means to easily trace the genealogy of all Icelanders.

Another potential advantage of performing a population-based study on cancer in Iceland concerns the genetic composition of the population. Most Icelanders are descended from a limited number of settlers that came to the country about 1,100 years ago. These types of 'young' populations possess fewer genetic variants than older or larger populations. This results in less disease heterogeneity and increases the probability of finding the disease-causing variants(s). There has been considerable discussion on the values and limitations of isolated populations for genetic studies[4]. Regardless of the mixed views on these issues, the Icelandic population has been successfully used to find genes associated with single-gene disorders[5,6] as well as genes implicated in complex diseases[7,8].

Resources of the ICP database

Research activities in the ICP can be broadly divided into three categories: studies on genetic cancer predisposition; correlation of biological and environmental factors with disease history; and creation of tools for functional analysis of tumour development and drug response. Integrating data from all study categories using the same patient population offers valuable opportunities to probe the relationship between genotype, environment and disease.

Box 1 | **The Icelandic Cancer Project**

Collaborators

- Iceland Genomics Corporation.
- The Icelandic Cancer Clinicians Group.
- Landspitali University Hospital.
- Akureyri District Hospital.
- The Icelandic Cancer Society – Cancer Registry.
- Genetical Committee of the University of Iceland.

Samples collected

- Samples collected from all previously diagnosed patients, relatives and controls: blood (DNA, plasma).
- Samples collected from newly diagnosed patients: blood (DNA, plasma); tumour tissue; in selected cases, cells are cultured from normal tissue.

Data collected

- ICD codes, age at diagnosis and tumour morphology (patients).
- Clinical information on treatment and outcomes (patients).
- Participant-reported lifestyle questionnaire; for example, smoking history, reproductive history and hormone use (all participants).
- Genealogical information (all participants).

Box 2 | **Linkage mapping and association mapping**

Linkage mapping uses known pedigrees in a search for markers that are co-transmitted with a disease phenotype. The method is powerful when many multigenerational family pedigrees with several cases of disease are available, but resolution is low (typically 5–10 megabases), because few recombinations are observed in the pedigrees that cover only a few generations.

In considering low-penetrance mutations, most individuals that carry the mutations are unaffected, and few pedigrees exist with numerous cases. Association mapping can be used to identify disease genes in this situation, as it only requires a set of markers that are analyzed in a number of affected (cases) and unaffected (control) individuals. This approach uses the fact that the DNA of all individuals is related through an unknown, but very deep, genealogy (possibly over 10,000 generations) that differs along the genome because of recombination. Markers that are physically close to each other will be associated because of linkage disequilibrium (LD). LD is created around a new disease mutation, but rapidly disappears because of recombination, except in a closely linked region. The size of this region is inversely proportional to the age of the mutation, the effective population size and the recombination rate in the region. It follows that a marker allele will have different frequencies in the affected and unaffected individuals if it is sufficiently physically close to the variable position that is associated with the disease. The depth (number of meioses) of the gene genealogy that can be analysed with association mapping allows for a sufficient amount of recombination to have occurred such that a disease mutation can be located with much greater resolution than by linkage mapping. Another advantage to association mapping is that the power is increased by the fact that all available cases can be included, irrespective of whether their family history is known.

Finding new cancer-susceptibility genes in the ICP. Much debate has raged over the percentage of cancer risk attributable to genetic factors. The emerging picture is that rare, highly penetrant alleles such as *BRCA1* and *BRCA2* account for only a small fraction of all cancers. Most of the remaining genetic cancer susceptibility is probably conferred by low-penetrance mutations in many genes[9–11].

The resources of the ICP provide an opportunity to search for genes of both high and low penetrance using linkage and association mapping (BOX 2). Linkage mapping describes the physical mapping of a disease gene using polymorphic markers in pedigrees for which disease phenotypes are known. If a given marker shows co-inheritance with disease status, the recombination frequency between marker and disease gene allows estimation of the distance between the two. Association mapping involves the comparison of allele frequencies for a marker allele or combination of markers (a haplotype) between a disease population and a control population. When statistically significant differences in the frequencies of an allele/haplotype are found between a disease and control population, the marker allele/haplotype are said to be associated with the disease.

To identify mutations with relatively high penetrance, several large families with a high prevalence of cancer have been ascertained. These families are being studied by genome-wide linkage mapping to broadly identify regions of the genome involved, for subsequent fine-mapping by association. For the lower-penetrance genes, however, linkage mapping is expected to be less effective, except in a subset of large families in which cancer prevalence is high. To pinpoint the lower-penetrance genes (BOX 2), case–control association studies might be more likely to succeed[12]. The density of markers required for association studies is high, but this number is inversely related to the extent of linkage disequilibrium (LD) in the target population[13]. LD describes the non-random association of alleles at markers at different positions. When two polymorphic positions in the genome are in LD, knowledge of the allele at the first position is a predictor of the type of allele at the second position. This relationship can be exploited to map disease genes. From a practical standpoint, it is advantageous to perform association studies in an isolated population like Iceland, where LD might be expected to extend further than in a typical Western population and the allelic heterogeneity is lower than in larger cosmopolitan populations[14].

To examine the LD in the Icelandic population, we examined 195 microsatellite markers in thirteen 15 cM regions of the genome (including the regions around *BRCA1* and *BRCA2*) in 1,400 representative individuals from the Icelandic population. More than half of all pairs of markers within 2 cM of each other (on the average 2 megabases), and more than 15% of the marker pairs 2–5 cM apart, showed statistically significant LD (data not shown). This supports the concept that LD covers large regions in the Icelandic population and alleviates the need for very tightly spaced markers for association mapping. Furthermore, the data indicate that association studies can be performed in Iceland at a fraction of the cost required in a typical European population, for which LD is lower. Building on these results, we performed limited genetic analysis on our patient population. The same thirteen regions that were examined in the LD study were surveyed in approximately 2,400 cancer cases and compared with the results from the 1,400 controls. Promising results from several regions are being pursued at present.

A key concern for any statistical method used for LD mapping of cancer-susceptibility mutations is its power. Single-marker association tests require unrealistically large sample sizes (tens to hundreds of thousands) for low-penetrance mutations of low frequencies, and quite large sample sizes (order of thousands) for low-penetrance mutations of high frequencies. By contrast, multipoint LD mapping, based on explicit modelling of the genealogy — see online supplementary information S1 (box) — is not confronted with the same difficulties[15–17]. Multipoint LD mapping is a type of association mapping that uses all marker information in affected and unaffected individuals simultaneously to gain maximal statistical power. As the power of this method for the high-frequency, low-penetrance scenarios has not yet been rigorously investigated, we undertook a simulation study to test the method (BOX 3). The approach was to first simulate the evolution of data sets with markers and disease mutations for a region of the genome, and then estimate the most probable location of the disease mutation(s) from the markers. The preliminary conclusion is that this method can accurately predict the location of disease mutations even under conditions of allelic heterogeneity and low penetrance for sample sizes as small as 100 cases. Furthermore, by using family information, the disease mutation could be more accurately located (BOX 3).

Finally, it should be stressed that the availability of detailed medical information within the ICP allows the subdivision of cancer phenotypes. This subdivision is assumed to partly reflect the genetic heterogeneity of the disease. By subdividing cases based on cancer phenotype, the genetic heterogeneity might be reduced, thereby facilitating the gene-mapping process.

Although the number of independent susceptibility alleles and the number of loci that determine susceptibility to complex traits are expected to be smaller in founder populations than in an outbred population, genetic studies in founder populations also have disadvantages. The number of rare

Box 3 | Bayesian multipoint linkage-disequilibrium mapping

Bayesian multipoint linkage-disequilibrium (LD) mapping is superior to other methods for mapping disease genes in that it uses all the data within a biological model, and should therefore be a statistically powerful method. A simulation study was conducted to estimate the power of this approach in identifying a low-penetrance disease gene; for details, see online supplementary information S2 (box). Replicate simulations of data sets were done using a coalescent approach; for details, see online supplementary information S1 (box). Subsequently, Bayesian multipoint LD mapping was used to estimate the positions of disease-causing mutations. Data sets of 100 cases and 100 controls were investigated with 10 or 20 single nucleotide markers. The size of the genomic region that was simulated corresponds to 50–500 kb in an African population (where LD is short) and several megabases in the Icelandic population (where LD covers larger regions; BOX 2). The optimal situation — full penetrance of disease phenotype — was used as the baseline scenario to investigate the effect of adding more biological factors. A fraction of mutations were missed using this approach, but, importantly, in more than 75% of the simulations performed, the mutation was accurately localized; for details, see online supplementary information S3 (table).

In the basic scenario, a mutation with population frequency of 10% can be mapped with a precision of 4.4% of the total sequence length. This means that the method can detect the position with accuracy greater than the distance between two adjacent markers, as the ten markers are separated by 11% of the sequence lengths. Knowledge of genealogy increases the precision by 28%. Doubling the marker density does not double the precision, but decreases the size of the candidate region by one-third. Correcting for LD in controls slightly improves precision. Less idealized scenarios reduce precision. A lack of information on linkage, or the occurrence of two mutations that contribute to the disease, only cause a moderate reduction in precision. In the final scenario, disease was caused by a combination of a very low penetrance gene and factors unrelated to the surveyed region (such as other genes or the environment). These factors further decreased precision, but in more than half of the simulations the disease mutations were still concentrated within one-third of the mapped region. Extrapolation of these encouraging results to the situation assumed in the Icelandic Cancer Project demands great caution, as more allelic heterogeneity and a much larger proportion of cases caused by other genes are expected. However, the fact that much larger case data sets (greater than 1,000 cases) are available will improve the efficacy of this approach. Simulations of such a scenario are being investigated at present.

cancers is low, rendering a population the size of Iceland´s less suitable for genetic analysis of these diseases. Furthermore, the degree of mapping resolution is positively correlated with the number of generations since the founding of the population. Since the Icelandic population is younger than other European populations, fine mapping performed on the Icelandic population would be at a lower resolution than similar studies carried out in older populations. Despite these disadvantages, there is considerable evidence to support the opinion that cancer-associated genes with low penetrance, but relatively high frequency, that contribute to cancers with high prevalence can be found using the resources of ICP. This approach could also serve as a model for other genetically isolated populations, provided that similar levels of LD exist.

Correlating biological and environmental factors with disease history. Although the relative contribution of genetic and environmental factors to cancer development has been debated, a combination of both is clearly involved. One example of this is provided by various reports on the penetrance of the *BRCA* mutations. In this regard, it has been shown that the penetrance of the Icelandic founder mutation, *BRCA2*999del5, varies greatly between families[18]. Furthermore, some families with this mutation have high rates of prostate cancer, whereas male breast cancer (but not prostate cancer) is evident in other families. So genetic background and environmental factors are likely to be important for the penetrance of this single mutation. The collection of lifestyle information, genealogical information and clinical data relevant to cancer from all participants in the ICP provides an attractive opportunity to probe this relationship between genes and environment.

The cancer field has at last entered the stage where experimental results are being turned into therapies tailored to the causative molecular defects. One well-known example of this is the tyrosine kinase inhibitor imatinib mesylate (Glivec), a targeted therapy primarily used to treat patients with chronic myeloid leukaemia. During designer-drug development, it is important to be able to correlate detailed information on the molecular signature of the tumour with the patient's genotype and the clinical behaviour of the disease. This information can be used to answer vital questions such as who will respond to the therapy and what happens during resistance development. The resources collected by the collaborators of the ICP can be effectively used to answer these questions. Furthermore, analysis of archived samples and data can greatly add to the power of the study, as it is possible to retrospectively select patient groups with particular disease forms. Finally, by linking genealogical information to tumour profile, it might be possible to identify characteristic genetic alterations or pathways that can give clues to the underlying genetic predisposition.

The collection of plasma samples from all ICP participants provides an important resource for searching for specific blood-borne tumour markers. When possible, plasma is collected from newly-diagnosed patients before surgery. In several cases, it is also possible to collect plasma samples from patients before treatment as well as during subsequent stages of disease. These samples, combined with matching clinical information, will be valuable material for proteomic studies aimed at identifying prognostic factors.

Generation of tools for functional analysis of tumorigenesis and drug response. One aspect of the ICP is the collection of live normal epithelial cells from patients with cancer. The cells are collected for use in various functional studies, from cancer initiation to drug screening. Historically, transformed cell lines have been used for functional studies and, although these cell lines have proven to be workhorses for studies on pathways of cell signalling, the results do not always apply to primary human tumours. Using the resources of the ICP, it is feasable to produce a panel of cell lines from a genetically susceptible individual with a known genotype and clinical history. These cell lines could have many uses. For example, tumorigenesis could be studied in cell lines derived from individuals with known mutations in either of the *BRCA* genes. Through transfection of the cells with selected oncogenes, or by knocking out checkpoint/repair proteins, it might be possible to identify molecular events that cooperate with *BRCA* deficiency in malignant transformation. Matching frozen tumour samples are available for all of the breast cell lines produced in the ICP, allowing direct correlation of experimental findings to properties of the tumours.

Large databases are available that contain information of the human genome sequence as well as gene-expression data. The Icelandic Cancer Project is structured to take advantage of this genomic data to support many aspects of cancer research. As the project develops and more information and samples are added to this coordinated collection, it is our hope that it will serve as a valuable resource for the cancer research community for years to come.

Thorunn Rafnar, Steinunn Thorlacius, Eirikur Steingrimsson, Violeta Calian and Bjarki J. Eldon are at Iceland Genomics Corporation, Snorrabraut 60, 105 Reykjavik, Iceland.

Mikkel H. Schierup, Jesper N. Madsen and Jotun Hein are at the Bioinformatics ApS, Hoegh Guldbergsgade 10, 8000 Arhus C, Denmark.

Thorvaldur Jonsson is at Landspitali University Hospital, Reykjavik, Iceland.

Snorri S. Thorgeirsson is at the Laboratory of Experimental Carcinogenesis, Center for Cancer Research, National Cancer Institute, NIH, Building 37, Room 4146A1, 37 Convent Drive MSC4262, Bethesda, MD 20892-4262 USA.

Correspondence to S.S.T.
e-mail: snorri_thorgeirsson@nih.gov

doi:10.1038/nrc1371

1. Rangarajan, A. & Weinberg, R. A. Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nature Rev. Cancer* **3**, 952–959 (2003).
2. Malkin, D. *et al.* Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**, 1233–1238 (1990).
3. Vasen, H. F. *et al.* The tumour spectrum in hereditary non-polyposis colorectal cancer: a study of 24 kindreds in the Netherlands. *Int. J. Cancer* **46**, 31–34 (1990).
4. Abbott, A. Manhattan versus Reykjavik. *Nature* **406**, 340–342 (2000).
5. Palsdottir, A. *et al.* Mutation in cystatin C gene causes hereditary brain haemorrhage. *Lancet* **2**, 603–604 (1988).
6. Kristinsson, S. Y. *et al.* MODY in Iceland is associated with mutations in HNF-1α and a novel mutation in NeuroD1. *Diabetologia* **44**, 2098–2103 (2001).
7. Stefansson, H. *et al.* Neuregulin 1 and susceptibility to schizophrenia. *Am. J. Hum. Genet.* **71**, 877–892 (2002).
8. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nature Genet.* **35**, 131–138 (2003).
9. Lichtenstein, P. *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).
10. Peto, J. Breast cancer susceptibility — A new look at an old model. *Cancer Cell* **1**, 411–412 (2002).
11. Risch, N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol. Biomarkers Prev.* **10**, 733–741 (2001).
12. Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nature Rev. Genet.* **2**, 91–99 (2001).
13. Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genet.* **33**, 228–237 (2003).
14. Shifman, S. & Darvasi, A. The value of isolated populations. *Nature Genet.* **28**, 309–310 (2001).
15. Morris, A. P., Whittaker, J. C. & Balding, D. J. Fine-scale mapping of disease loci via shattered coalescent modeling of genealogies. *Am. J. Hum. Genet.* **70**, 686–707 (2002).
16. Liu, J. S., Sabatti, C., Teng, J., Keats, B. J. & Risch, N. Bayesian analysis of haplotypes for linkage disequilibrium mapping. *Genome Res.* **11**, 1716–1724 (2001).
17. Rannala, B. & Reeve, J. P. High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence. *Am. J. Hum. Genet.* **69**, 159–178 (2001).
18. Thorlacius, S. *et al.* Population-based study of risk of breast cancer in carriers of *BRCA2* mutation. *Lancet* **352**, 1337–1339 (1998).

Competing interests statement
The authors declare competing financial interests: see web version for details.

Online links

DATABASES
The following terms in this article are linked online to:
Cancer.gov: http://cancer.gov/
breast cancer | chronic myeloid leukaemia | male breast cancer | prostate cancer
Entrez Gene:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
BRCA1 | *BRCA2*
OMIM:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
hereditary non-polyposis colon cancer | Li–Fraumeni syndrome

FURTHER INFORMATION
The Cancer Genome Anatomy Project:
www.ncbi.nlm.nih.gov/ncicgap/
The Cancer Genome Project: www.sanger.ac.uk/CGP/

SUPPLEMENTARY INFORMATION
See online article: S1 (box) | S2 (box) | S3 (table)
Access to this links box is available online.