Histol Histopathol (2008) 23: 507-514

**http://www.hh.um.es**

**Histology and Histopathology**

*Cellular and Molecular Biology*

## *Review*

# Current strategies in the search for low penetrance genes in cancer

**Roger L. Milne and Javier Benítez**
Human Cancer Genetics Programme, Spanish National Cancer Research Centre, Madrid, Spain

**Summary.** The genetic etiology of most cancers remains largely unclear and it has been hypothesised that common genetic variants with modest effects on disease susceptibility cause the bulk of this unexplained risk. Case-control association studies are considered the most effective strategy to identify these low-penetrance genes. While traditionally, such studies have focused on putative functional single nucleotide polymorphisms (SNPs) in candidate genes, a more comprehensive approach can now be taken, as a result of a number of recent developments: the mapping of the human genome, including the identification of almost ten million SNPs; and the development of high-throughput genotyping technologies that enable hundreds of thousands of SNPs to be genotyped in a single reaction, in multiple subjects and at an affordable cost. All common genomic variation can be captured by genotyping SNPs in gene-, pathway- or genome-wide-based strategies and these are now being applied to many diseases, including cancer. We present an outline of each of these approaches, including recent published examples, and discuss a number of challenges that remain to be addressed.

**Key words:** Cancer, Genes, SNPs, Study design

### Introduction

Many cancers are considered complex diseases, where multiple environmental and genetic factors interplay to cause the disease. While clear environmental causes have been identified for a number of cancers using classical epidemiological techniques, the ability to detect the genetic components of cancer etiology has until recently been very limited. Exceptions include genes that cause rare, largely monogenic family cancer syndromes such as Multiple Endocrine Neoplasia Type 2 (Mathew et al., 1987; Simpson et al., 1987; Mulligan et al., 1993), *CDKN2A* in familial melanoma (Cannon-Albright et al., 1992; Kamb et al., 1994; Holland et al., 1995), *BRCA1* in familial breast and ovarian cancer (Hall et al., 1990; Narod et al., 1991; Miki et al., 1994) *BRCA2* in familial breast cancer (Wooster et al., 1994, 1995), and MLH1 and MSH2 in hereditary nonpolyposis colorectal cancer (Lindblom et al., 1993; Peltomaki et al., 1993; Bronner et al., 1994; Fishel et al., 1993; Nystrom-Lahti et al., 1994). However, such exceptions tend to account for only a small portion of disease heritability. For example, the breast cancer susceptibility genes BRCA1 and BRCA2 explain only a minority (<30%) of familial breast cancers and a negligible proportion of sporadic breast cancers (Diez et al., 2003; Thompson and Easton, 2004). That is, the genetic etiology of most cancers remains largely unclear.

It seems most plausible that more common genetic variants with modest effects on disease susceptibility (low penetrance genes) cause the bulk of this unexplained risk (Risch, 2000; Pharoah et al., 2002, 2004; Botstein and Risch, 2003; Houlston and Peto, 2004; Zondervan and Cardon, 2004; Wang et al., 2005). Association studies comparing genotype distributions between individuals with cancer and disease-free subjects are considered the best strategy to detect these modest effects (Risch and Merikangas, 1996; Cardon and Bell, 2001). Several types of polymorphic variants exist in the human genome but the most frequent and useful for genotyping studies are single nucleotide polymorphisms (SNPs). SNPs are genomic loci where two alternative bases are present with appreciable frequency (greater than 1%). They are the most common type of variation in the human genome, occurring every several hundred base pairs, and nearly ten million SNPs

*Offprint requests to:* Javier Benítez, Human Cancer Genetics Programme, Spanish National Cancer Research Centre, C/ Melchor Fernández Almagro, 3, E-28029 Madrid, Spain. e-mail: jbenitez@cnio.es

have been identified to date. Until relatively recently, association studies have been undertaken by studying a small number of potentially functional SNPs (usually non-synonymous coding SNPs) in candidate genes, one gene at a time (Carlson et al., 2004; Neale and Sham, 2004). This has been referred to as the direct approach (see Fig. 1) (Carlson et al., 2004).

An alternative (indirect) approach is based on the argument that disease-associated variants with modest effects might be distributed proportionately between coding and non-coding sequences of the genome (Carlson et al., 2004; Hirschhorn and Daly, 2005). Indeed, the potential importance in gene function, and therefore disease susceptibility, of non-coding SNPs such as those located at transcription factor binding sites and the promoter in general, some intronic regions and, more recently, microRNA binding sites, is now well recognised (Beohar and Kawamoto, 1998; Dean and Clark, 1999; Kawada et al., 1999; Abelson et al., 2005; Clop et al., 2006). Furthermore, it has recently been suggested that conserved non-genic sequences are also likely to play a role in phenotypic variability and human disorders, although their function is largely unknown (Dermitzakis et al., 2005; Hirschhorn and Daly, 2005; Drake et al., 2006). Under the indirect approach, marker SNPs are ideally chosen to maximally capture the common variation across a candidate gene, or the entire genome (see Fig. 1). The idea behind this is that associations will be detected either directly with causal variants, if genotyped, or indirectly with markers in linkage disequilibrium (LD) with causal variants (Carlson et al., 2004; Neale and Sham, 2004).

The relatively recent mapping of the human genome and subsequent identification of almost ten million SNPs therein, over half of which have been validated in distinct populations and have genotype and allele frequency data available in the public domain (http://www.ncbi.nlm.nih.gov/projects/SNP/; http://www.hapmap.org/), has made this strategy viable. Other publicly available databases and search engines such as SeattleSNP, Ensemble, Tagger and PupaSuite (http://pga.gs.washington.edu/, http://www.ensembl.org/index.html, http://www.broad.mit.edu/mpg/tagger/, http://pupasuite.bioinfo.cipf.es/) have emerged to assist with SNP selection and studies are now beginning to emerge applying these to a range of complex diseases, including cancers, using case-controls designs. Furthermore, recent developments and the rapid advancement in high-throughput genotyping technology have made genotyping a large number of marker SNPs throughout the genome a possibility, both logistically and economically (Hirschhorn and Daly, 2005). Up to 1,000,000 predetermined marker SNPs, and up to 60,800 customised SNPs, can currently be genotyped in a single reaction in a time and at a cost that was previously unimaginable. A summary of the technologies currently available is provided in Table 1. Consequently, large-scale genome-wide association studies have also begun to emerge (Daimon et al., 2003; Hao et al., 2004; Kammerer et al., 2004; Li et al., 2004; Peters et al., 2004; Hu et al., 2005; Ozaki and Tanaka, 2005; Hampe et al., 2007). Here we outline different strategies to identify low penetrance genes in cancer.

## Gene-based strategies

To date, the most common approach taken to identify common polymorphisms has been to focus on genes likely to be involved in disease etiology on the basis of disease biology. For example, in female breast cancer, the well-established associations with hormone-related environmental factors such as age at menarche, parity, menopausal status and body mass index have highlighted the estrogen and other hormone metabolism genes as likely candidates. Other biological evidence, including gene expression studies and mouse studies may be also used to identify candidate genes. Once a candidate gene has been chosen, two main approaches may be adopted. Under the first, putative functional SNPs are selected across the gene and studied in a standard case-control study. The advantage of this direct approach is that it is hypothesis-driven and so the interpretation of positive findings may be more

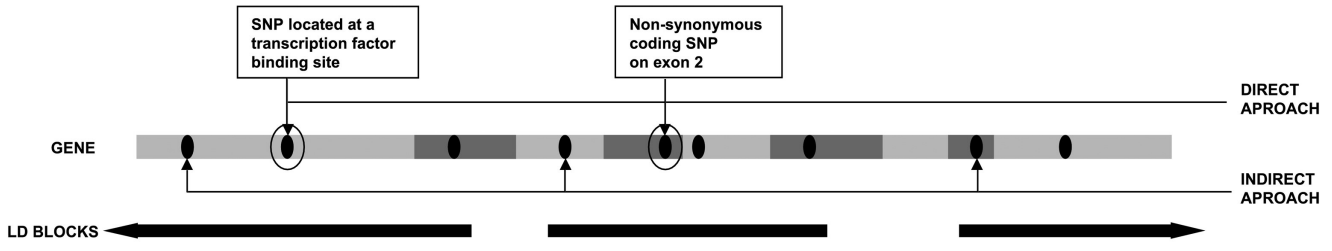


**Fig. 1.** Schematic representation of SNP selection according to a direct versus an indirect approach: gene pictured in grey with exons marked in darker grey; all SNPs with frequency greater than 5% depicted as black ovals in the gene; all putative functional SNPs are circled; regions of high linkage disequilibrium (LD) between SNPs shown as black bars below the gene. Arrows above the line indicate putative functional SNPs chosen under a direct approach, while arrows below the line indicate SNPs chosen under an indirect approach, according to LD patterns and regardless of function.

straightforward. What is not straightforward is the identification of putative functional SNPs. While a number of programs exist to guide such searches, it is not always clear what SNPs could be functional. Many studies have taken this approach by choosing only non-synonymous coding (missense) SNPs in the candidate gene in question, as proposed by Risch and Merikangas (1996). However, emerging evidence has directed attention away from just missense SNPs since, as mentioned above, it has been demonstrated that SNPs located in non-coding conserved regions outside of genes may also have functional consequences, as may synonymous coding SNPs (Chamary et al., 2006; Nackley et al., 2006; Kimchi-Sarfaty et al., 2007). The identification of functional SNPs is therefore not straightforward and so it is not clear how direct this direct approach can really be.

A perhaps more objective approach to SNP selection in candidate genes is to tag all common variation with tagging SNPs (tSNPs) and/or haplotypes formed by them and thereby test for association at any point in the gene and its flanking regions. This approach has been taken for a number of candidate susceptibility genes in breast cancer and ovarian cancer (Lesueur et al., 2005; Song et al., 2006b, 2007). There are two challenges inherent in this second approach. One is that the degree to which common variation in the gene is tagged depends on the density of the SNP map used. Some groups have chosen to re-sequence the gene rather then rely solely on HapMap data. The other challenge is the identification of the causal functional SNP if an association is found with a marker, since LD can extend across large regions containing many putative functional SNPs. Both these gene-based approaches depend on the identification of candidate disease-related genes and face the additional challenge of demonstrating that a putatively functional associated SNP is indeed functional and thereby potentially causal.

### Genetic-pathway-based strategies

An extension of the gene-based approach is to study a larger number of SNPs in genes in candidate genetic pathways, a strategy made possible by new technologies that permit the analysis of thousands of SNPs at the same time, in large series of samples, and in a short period of time (see Table 1). Some groups such as Wu et al. (2006) have done this by focussing on putative functional SNPs (in coding or promoter regions and/or those with previously reported associations with, in this case, bladder cancer) in genes from a given pathway or pathways (direct approach). As for gene-based studies, an alternative is to select tagging SNPs across genes as markers for any association, with no prior hypothesis regarding where causal loci might be located. This indirect pathway-based tagging approach has been applied to assess the potential role of genes involved in mismatch repair in ovarian cancer susceptibility (Song et al., 2006a).

We applied a combination of these direct and indirect strategies to the search for low penetrance genes involved in breast cancer, selecting SNPs as markers across the gene based on linkage disequilibrium patterns, but including putative functional SNPs where possible (Milne et al., 2006). The latter included coding and promoter SNPs, as well as those potentially causing alternative splicing, those located at putative transcription factor binding sites and/or those located in regions highly conserved across species. A total of 710 SNPs, representing 112 genes involved in cancer-related pathways such as DNA-repair, cell-cycle control, signalling and apoptosis, were genotyped in almost 2,000 cases and controls. The 10 SNPs with strongest evidence of association in our Spanish case-control series were then tested in an independent Finnish series and one of these, an intronic variant in ERCC4 from the nucleotide excision repair pathway, was found to be associated with protection from breast cancer.

The pros and cons of the different versions of this approach are identical to those outlined for gene-based strategies, with the added challenge of dealing with the issue of false-positive associations resulting from multiple testing. Two-stage designs such as that just described (Milne et al., 2006) are one way of avoiding false-positive associations (see Additional Considerations, below).

### Genome-wide scans

A number of genotyping platforms such as Illumina and Affymetrix now accommodate the genotyping of arrays of up to 1,000,000 SNPs in a single multiplex reaction (see Table 1). These SNPs capture most of the common variation in the human genome. There are also arrays of SNPs located on or near exons, transcripts and highly conserved regions across the entire genome. In addition, some platforms permit additional customised SNPs to be included. These options allow for direct, indirect and combined approaches to be adopted on a genome-wide scale and such studies are now beginning

**Table 1.** Summary of genotyping platforms commonly used at present.

| Platform | Assay design* | Number of SNPs† | Samples per plate†† |
|---|---|---|---|
| Taqman | Custom-built | 1 | 384 |
| Sequenom | Custom-built | 36 | 384 |
| SNPlex | Custom-built | 48 | 384 |
| Illumina – Golden Gate | Custom-built | 1,536 | 96 |
| Illumina – Infinium | Custom-built | 60,800 | 12 |
| Illumina – Infinium | Predetermined | 1,000,000 | 1 |
| Affymetrix | Custom-built | 25,000 | 48 |
| Affymetrix | Predetermined | 950,000 | 1 |

*Custom-built means that the SNPs to be genotyped can be selected by the investigator. †: Maximum number of SNPs genotyped in a single reaction. ††: Maximum number of samples genotyped in a single plate.

to be applied to the search for low penetrance genes in a number of complex diseases. The Wellcome Trust Case-Control Consortium (WTCCC) is perhaps the largest of these undertakings and includes both direct and indirect strategies (The Wellcome Trust Case Control Consortium, 2007). The WTCCC aims to test for "genetic signposts" among 675,000 SNPs for 8 diseases (tuberculosis, coronary heart disease, type 1 diabetes, type 2 diabetes, rheumatoid arthritis, Crohn's disease, bipolar disorder and hypertension). The study includes 2,000 cases for each of these diseases and 3,000 common controls, all from the United Kingdom. A sub-study will include 2,000 cases of each of tuberculosis and malaria along with 2,000 common controls, all from African populations. Finally 15,000 known non-synonymous coding SNPs will be assessed in case-control series for associations with each of four additional diseases (breast cancer, autoimmune thyroid disease, ankylosing spondylitis and multiple sclerosis). Other genome-wide association studies have been recently published, with convincing positive findings reported for breast and prostate cancer (Easton et al., 2007; Gudmundsson et al., 2007; Hunter et al., 2007; Stacey et al., 2007; Yeager et al., 2007).

The clear advantage of this approach is that it is not restricted by our currently relatively limited knowledge of disease biology, nor, in the case of the gene- or pathway-based approaches, of the genome. The challenges of genome-wide association studies include the cost (the WTCCC received almost £9 million from the Wellcome Trust to fund the project), the logistics of managing the volume of data generated, as well as minimising false-positive associations and maximising power. A further challenge that will arise is that of interpreting positive findings from indirect approaches as they emerge and are validated in sufficiently powered independent studies (see below). The identification of a causal SNP is no trivial task when a marker in or near a gene is found to be associated with disease. It is not at all clear how to tackle this issue if an associated marker SNP is nowhere near a known gene.

## Additional considerations

Association studies that test a large number of SNPs can lead to false-positive associations if multiple testing is not adequately accounted for. At the same time, correcting for a large number of tests requires large sample sizes to maintain adequate statistical power, and therefore avoid false-negative associations, and that can be prohibitively expensive. In addition, confirmation of associations identified in such studies by replication in independent and adequately sized samples is essential (Hirschhorn and Daly, 2005; Easton et al., 2007). These considerations present an economic and logistical challenge to investigators seeking to produce quality research in this field. Multi-stage study designs have been proposed as an efficient means of addressing these challenges (Hirschhorn and Daly, 2005; Shiffman et al., 2005). Under these designs all SNPs are genotyped in a set of cases and controls and only a reduced set of candidate SNPs, selected based on unadjusted p-values, are genotyped in a subsequent, independent set. If done appropriately, this staged design can substantially reduce the amount of genotyping, and therefore the costs incurred, without significant loss of statistical power (Song et al., 2007). Both the studies or subsets within the same study should be of sufficient size to avoid false-positive and false-negative findings (Cox et al., 2006). Selected cases (with early age at onset and/or a family history of the disease, for example) and controls (all disease free beyond a particular age) may be used to increase power for a given sample size (Antoniou and Easton, 2003; Houlston and Peto, 2003).

Often, additional replication studies in independent case-control series are required to conclude that a SNP is definitively associated with disease. This is the main goal of the Breast Cancer Association Consortium (BCAC), which was established in 2005 and currently includes more than 20 breast cancer case-controls studies from around the world (Breast Cancer Association Consortium, 2006; Cox et al., 2007; Easton et al., 2007). The extremely large combined sample of over 30,000 breast cancer cases and 30,000 controls allows for many SNPs to be definitively evaluated as susceptibility loci in breast cancer. To date, all but very small main effects have been ruled out for 11 SNPs previously found to be associated with breast cancer (Breast Cancer Association Consortium, 2006), while CASP8 has been confirmed to be a low penetrance breast cancer gene at genome-wide levels of statistical significance (Cox et al., 2007). Further replication studies are currently underway for other SNPs, including the one in ERCC4 identified in our study (Milne et al., 2006) as well as a number of strong candidates from other genome-wide scans. Large combined samples achieved through international collaborations are also likely to help identify true low penetrance genes in other complex diseases as more and more results from small- and large-scale association studies are published, many of which are likely to be false-positives (Wacholder et al., 2004).

Finally, demonstrating that a candidate SNP has functional consequences is often a very challenging final step in establishing its causal relationship to the disease. The most frequent approaches taken to evaluate the biological implications of a SNP include *in silico* studies, analysis of sequence conservation across different species, and *in vitro* and *in vivo* expression studies with constructs that include the variant. It is not clear how to establish the function of a putative causal SNP that is not located in or near a gene (Easton et al., 2007).

## Conclusions

The strategies applied to search for low penetrance genes in complex diseases such as cancer have changed

over the past few years as a combined result of the information that the Human Genome Project has supplied about our genome and advances in technology that allow for high-throughput genotyping. Large-scale studies can now be undertaken without any knowledge of the biology of the disease. Our experience to date has demonstrated the importance of using large sample sizes and validating positive results in adequately powered independent studies. It has also highlighted the challenges, once disease-associated loci are identified by indirect methods, of identifying potentially causal SNPs and establishing their biological role in the disease etiology. A large number of further genetic discoveries related to complex diseases in general, and cancer in particular, are expected to emerge in the near future.

## References

Abelson J.F., Kwan K.Y., O'Roak B.J., Baek D.Y., Stillman A.A., Morgan T.M., Mathews C.A., Pauls D.L., Rasin M.R., Gunel M., Davis N.R., Ercan-Sencicek A.G., Guez D.H., Spertus J.A., Leckman J.F., Dure L.St, Kurlan R., Singer H.S., Gilbert D.L., Farhi A., Louvi A., Lifton R.P., Sestan N. and State MW. (2005). Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science 310, 317-320.

Antoniou A.C. and Easton D.F. (2003). Polygenic inheritance of breast cancer: Implications for design of association studies. Genet Epidemiol. 25, 190-202.

Beohar N. and Kawamoto S. (1998). Transcriptional regulation of the human nonmuscle myosin II heavy chain-A gene. Identification of three clustered cis-elements in intron-1 which modulate transcription in a cell type- and differentiation state-dependent manner. J. Biol. Chem. 273, 9168-9178.

Botstein D. and Risch N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33 Suppl, 228-237.

Breast Cancer Association Consortium. (2006). Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. J. Natl. Cancer Inst. 98, 1382-1396.

Bronner C.E., Baker S.M., Morrison P.T., Warren G., Smith L.G., Lescoe M.K., Kane M., Earabino C., Lipford J., Lindblom A., Tannergard P., Bollag R.J., Godwin A.R., Ward D.C., Nordenskjold M., Fishel R., Kolodner R. and Liskay R.M. (1994). Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature. 368, 258-261.

Cannon-Albright L.A., Goldgar D.E., Meyer L.J., Lewis C.M., Anderson D.E., Fountain J.W., Hegi M.E., Wiseman R.W., Petty E.M., Bale A.E., Olopade O.I., Diaz M.O., Kwiatkowski D.J., Piepkorn M.W., Zone J.J. and Skolnick M.H. (1992). Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. Science 258, 1148-1152.

Cardon L.R. and Bell J.I. (2001). Association study designs for complex diseases. Nat. Rev. Genet. 2, 91-99.

Carlson C.S., Eberle M.A., Kruglyak L. and Nickerson D.A. (2004). Mapping complex disease loci in whole-genome association studies. Nature 429, 446-452.

Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X., Bibe B., Bouix J., Caiment F., Elsen J.M., Eychenne F., Larzul C., Laville E., Meish F., Milenkovic D., Tobin J., Charlier C. and Georges M. (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet. 38, 813-818.

Cox A., Dunning A.M., Garcia-Closas M., Balasubramanian S., Reed M.W., Pooley K.A., Scollen S., Baynes C., Ponder B.A., Chanock S., Lissowska J., Brinton L., Peplonska B., Southey M.C., Hopper J.L., McCredie M.R., Giles G.G., Fletcher O., Johnson N., Dos Santos Silva I., Gibson L., Bojesen S.E., Nordestgaard B.G., Axelsson C.K., Torres D., Hamann U., Justenhoven C., Brauch H., Chang-Claude J., Kropp S., Risch A., Wang-Gohrke S., Schurmann P., Bogdanova N., Dork T., Fagerholm R., Aaltonen K., Blomqvist C., Nevanlinna H., Seal S., Renwick A., Stratton M.R., Rahman N., Sangrajrang S., Hughes D., Odefrey F., Brennan P., Spurdle A.B., Chenevix-Trench G., Beesley J., Mannermaa A., Hartikainen J., Kataja V., Kosma V.M., Couch F.J., Olson J.E., Goode E.L., Broeks A., Schmidt M.K., Hogervorst F.B., Veer L.J., Kang D., Yoo K.Y., Noh D.Y., Ahn S.H., Wedren S., Hall P., Low Y.L., Liu J., Milne R.L., Ribas G., Gonzalez-Neira A., Benitez J., Sigurdson A.J., Stredrick D.L., Alexander B.H., Struewing J.P., Pharoah P.D. and Easton D.F. (2007). A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet. 39, 352-358.

Cox D.G., Hankinson S.E. and Hunter D.J. (2006). Polymorphisms in the ICAM gene locus are not associated with breast cancer risk. Cancer Epidemiol. Biomarkers Prev. 15, 178-179.

Chamary J.V., Parmley J.L. and Hurst L.D. (2006). Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat. Rev. Genet. 7, 98-108.

Daimon M., Ji G., Saitoh T., Oizumi T., Tominaga M., Nakamura T., Ishii K., Matsuura T., Inageda K., Matsumine H., Kido T., Htay L., Kamatani N., Muramatsu M. and Kato T. (2003). Large-scale search of SNPs for type 2 DM susceptibility genes in a Japanese population. Biochem. Biophys. Res. Commun. 302, 751-758.

Dean G. and Clark I.M. (1999). Transcriptional regulation of the human tissue inhibitor of metalloproteinases-1: mapping transcriptional control in intron-1. Ann. NY Acad. Sci. 878, 510-511.

Dermitzakis E.T., Reymond A. and Antonarakis S.E. (2005). Conserved non-genic sequences - an unexpected feature of mammalian genomes. Nat. Rev. Genet. 6, 151-157.

Diez O., Osorio A., Duran M., Martinez-Ferrandis J.I., de la Hoya M., Salazar R., Vega A., Campos B., Rodriguez-Lopez R., Velasco E., Chaves J., Diaz-Rubio E., Jesus Cruz J., Torres M., Esteban E., Cervantes A., Alonso C., San Roman J.M., Gonzalez-Sarmiento R., Miner C., Carracedo A., Armengod M.E., Caldes T., Benitez J. and Baiget M. (2003). Analysis of BRCA1 and BRCA2 genes in Spanish breast/ovarian cancer patients: a high proportion of mutations unique to Spain and evidence of founder effects. Hum. Mutat. 22, 301-312.

Drake J.A., Bird C., Nemesh J., Thomas D.J., Newton-Cheh C., Reymond A., Excoffier L., Attar H., Antonarakis S.E., Dermitzakis E.T. and Hirschhorn J.N. (2006). Conserved noncoding sequences are selectively constrained and not mutation cold spots. Nat. Genet. 38, 223-227.

Easton D.F., Pooley K.A., Dunning A.M., Pharoah P.D., Thompson D., Ballinger D.G., Struewing J.P., Morrison J., Field H., Luben R., Wareham N., Ahmed S., Healey C.S., Bowman R., the SEARCH collaborators, Meyer K.B., Haiman C.A., Kolonel L.K., Henderson B.E., Le Marchand L., Brennan P., Sangrajrang S., Gaborieau V., Odefrey F., Shen C.Y., Wu P.E., Wang H.C., Eccles D., Evans D.G.,

Peto J., Fletcher O., Johnson N., Seal S., Stratton M.R., Rahman N., Chenevix-Trench G., Bojesen S.E., Nordestgaard B.G., Axelsson C.K., Garcia-Closas M., Brinton L., Chanock S., Lissowska J., Peplonska B., Nevanlinna H., Fagerholm R., Eerola H., Kang D., Yoo K.Y., Noh D.Y., Ahn S.H., Hunter D.J., Hankinson S.E., Cox D.G., Hall P., Wedren S., Liu J., Low Y.L., Bogdanova N., Schurmann P., Dork T., Tollenaar R.A., Jacobi C.E., Devilee P., Klijn J.G., Sigurdson A.J., Doody M.M., Alexander B.H., Zhang J., Cox A., Brock I.W., Macpherson G., Reed M.W., Couch F.J., Goode E.L., Olson J.E., Meijers-Heijboer H., van den Ouweland A., Uitterlinden A., Rivadeneira F., Milne R.L., Ribas G., Gonzalez-Neira A., Benitez J., Hopper J.L., McCredie M., Southey M., Giles G.G., Schroen C., Justenhoven C., Brauch H., Hamann U., Ko Y.D., Spurdle A.B., Beesley J., Chen X., kConFab, AOCS Management Group, Mannermaa A., Kosma V.M., Kataja V., Hartikainen J., Day N.E., Cox D.R. and Ponder B.A. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447, 1087-1093.

Fishel R., Lescoe M.K., Rao M.R., Copeland N.G., Jenkins N.A., Garber J., Kane M. and Kolodner R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 75, 1027-1038.

Gudmundsson J., Sulem P., Manolescu A., Amundadottir L.T., Gudbjartsson D., Helgason A., Rafnar T., Bergthorsson J.T., Agnarsson B.A., Baker A., Sigurdsson A., Benediktsdottir K.R., Jakobsdottir M., Xu J., Blondal T., Kostic J., Sun J,, Ghosh S,, Stacey S.N,, Mouy M., Saemundsdottir J., Backman V.M., Kristjansson K., Tres A., Partin A.W., Albers-Akkers M.T., Godino-Ivan Marcos J., Walsh P.C., Swinkels D.W., Navarrete S., Isaacs S.D., Aben K.K., Graif T., Cashy J., Ruiz-Echarri M., Wiley K.E., Suarez B.K., Witjes J.A., Frigge M., Ober C., Jonsson E., Einarsson G.V., Mayordomo J.I., Kiemeney L.A., Isaacs W.B., Catalona W.J., Barkardottir R.B., Gulcher J.R., Thorsteinsdottir U., Kong A. and Stefansson K. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat. Genet. 39, 631-637.

Hall J.M., Lee M.K., Newman B., Morrow J.E., Anderson L.A., Huey B. and King M.C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250, 1684-1689.

Hampe J., Franke A., Rosenstiel P., Till A., Teuber M., Huse K., Albrecht M., Mayr G., De La Vega F.M., Briggs J., Gunther S., Prescott N.J., Onnie C.M., Hasler R., Sipos B., Folsch U.R., Lengauer T., Platzer M., Mathew C.G., Krawczak M. and Schreiber S. (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat. Genet. 39, 207-211.

Hao K., Wang X., Niu T., Xu X., Li A., Chang W., Wang L., Li G. and Laird N. (2004). A candidate gene association study on preterm delivery: application of high-throughput genotyping technology and advanced statistical methods. Hum. Mol. Genet. 13, 683-691.

Hirschhorn J.N. and Daly M.J. (2005). Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95-108.

Holland E.A., Beaton S.C., Becker T.M., Grulet O.M., Peters B.A., Rizos H., Kefford R.F. and Mann G.J. (1995). Analysis of the p16 gene, CDKN2, in 17 Australian melanoma kindreds. Oncogene 11, 2289-2294.

Houlston R.S. and Peto J. (2003). The future of association studies of common cancers. Hum. Genet. 112, 434-435.

Houlston R.S. and Peto J. (2004). The search for low-penetrance cancer susceptibility alleles. Oncogene 23, 6471-6476.

Hu N., Wang C., Hu Y., Yang H.H., Giffen C., Tang Z.Z., Han X.Y., Goldstein A.M., Emmert-Buck M.R., Buetow K.H., Taylor P.R. and Lee MP. (2005). Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. Cancer Res. 65, 2542-2546.

Hunter D.J., Kraft P., Jacobs K.B., Cox D.G., Yeager M., Hankinson S.E., Wacholder S., Wang Z., Welch R., Hutchinson A., Wang J., Yu K., Chatterjee N., Orr N., Willett W.C., Colditz G.A., Ziegler R.G., Berg C.D., Buys S.S., McCarty C.A., Feigelson H.S., Calle E.E., Thun M.J., Hayes R.B., Tucker M., Gerhard D.S., Fraumeni J.F. Jr, Hoover R.N., Thomas G. and Chanock S.J. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat. Genet. 39, 870-874.

Kamb A., Shattuck-Eidens D., Eeles R., Liu Q., Gruis N.A., Ding W., Hussey C., Tran T., Miki Y., Weaver-Feldhaus J., McClure M., Aitken J.F., Anderson D.E., Bergman M., Frants R., Goldgar D.E., Green A., MacLennan R., Martin N.G., Meyer L.J., Youl P., Zone J.J., Skolnick M.H. and Cannon-Albright L.A. (1994). Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. Nat. Genet. 8, 23-26.

Kammerer S., Roth R.B., Reneland R., Marnellos G., Hoyal C.R., Markward N.J., Ebner F,. Kiechle M., Schwarz-Boeger U., Griffiths L.R., Ulbrich C., Chrobok K., Forster G., Praetorius G.M., Meyer P., Rehbock J., Cantor C.R., Nelson M.R. and Braun A. (2004). Large-scale association study identifies ICAM gene region as breast and prostate cancer susceptibility locus. Cancer Res. 64, 8906-8910.

Kawada N., Moriyama T., Ando A., Koyama T., Hori M., Miwa T. and Imai E. (1999). Role of intron 1 in smooth muscle alpha-actin transcriptional regulation in activated mesangial cells in vivo. Kidney Int. 55, 2338-2348.

Kimchi-Sarfaty C., Oh J.M., Kim I.W., Sauna Z.E., Calcagno A.M., Ambudkar S.V. and Gottesman M.M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 315, 525-528.

Lesueur F., Pharoah P.D., Laing S., Ahmed S., Jordan C., Smith P.L., Luben R., Wareham N.J., Easton D.F., Dunning A.M. and Ponder B.A. (2005). Allelic association of the human homologue of the mouse modifier Ptprj with breast cancer. Hum. Mol. Genet. 14, 2349-2356.

Li Y., Nowotny P., Holmans P., Smemo S., Kauwe J.S., Hinrichs A.L., Tacey K., Doil L., van Luchene R., Garcia V., Rowland C., Schrodi S., Leong D., Gogic G., Chan J., Cravchik A., Ross D., Lau K., Kwok S., Chang S.Y., Catanese J., Sninsky J., White T.J., Hardy J., Powell J., Lovestone S., Morris J.C., Thal L., Owen M., Williams J., Goate A. and Grupe A. (2004). Association of late-onset Alzheimer's disease with genetic variation in multiple members of the GAPD gene family. Proc. Natl. Acad. Sci. USA 101, 15688-15693.

Lindblom A., Tannergard P., Werelius B. and Nordenskjold M. (1993). Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. Nat. Genet. 5, 279-282.

Mathew C.G., Chin K.S., Easton D.F., Thorpe K., Carter C., Liou G.I., Fong S.L., Bridges C.D., Haak H., Kruseman A.C., Schifter S., Hansen H.H., Telenius H., Telenius-Berg M. and Ponder B.A.J. (1987). A linked genetic marker for multiple endocrine neoplasia type 2A on chromosome 10. Nature 328, 527-528.

Miki Y., Swensen J., Shattuck-Eidens D., Futreal P.A., Harshman K., Tavtigian S., Liu Q., Cochran C., Bennett L.M., Ding W., Bell R.,

Rosenthal J., Hussey C., Tran T., McClure M., Frye C., Hattier T., Phelps R., Haugen-Strano A., Katcher H., Yakumo K., Gholami Z., Shaffer D., Stone S., Bayer S., Wray C., Bogden R., Dayananth P., Ward J., Tonin P., Narod S., Bristow P., Norris F., Helvering L., Morrison P., Rosteck P., Lai M., Barrett J.C., Lewis C., Neuhausen S., Cannon-Albright L., Goldgar D., Wiseman R., Kamb A. and Skolnick M.H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266, 66-71.

Milne R.L., Ribas G., Gonzalez-Neira A., Fagerholm R., Salas A., Gonzalez E., Dopazo J., Nevanlinna H., Robledo M. and Benitez J. (2006). ERCC4 associated with breast cancer risk: A two-stage case-control study using high-throughput genotyping. Cancer Res. 66, 9420-9427.

Mulligan L.M., Kwok J.B., Healey C.S., Elsdon M.J., Eng C., Gardner E., Love D.R., Mole S.E., Moore J.K., Papi L., Ponder M.A., Telenius H., Tunnacliffe A. and Ponder B.A.J. (1993). Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. Nature 363, 458-460.

Nackley A.G., Shabalina S.A., Tchivileva I.E., Satterfield K., Korchynskyi O., Makarov S.S., Maixner W. and Diatchenko L. (2006). Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science 314, 1930-1933.

Narod S.A., Feunteun J., Lynch H.T., Watson P., Conway T., Lynch J. and Lenoir G.M. (1991). Familial breast-ovarian cancer locus on chromosome 17q12-q23. Lancet 338, 82-83.

Neale B.M. and Sham P.C. (2004). The future of association studies: gene-based analysis and replication. Am. J. Hum. Genet. 75, 353-362.

Nystrom-Lahti M., Sistonen P., Mecklin J.P., Pylkkanen L., Aaltonen L.A., Jarvinen H., Weissenbach J., de la Chapelle A. and Peltomaki P. (1994). Close linkage to chromosome 3p and conservation of ancestral founding haplotype in hereditary nonpolyposis colorectal cancer families. Proc. Natl. Acad. Sci. USA. 91, 6054-6058.

Ozaki K. and Tanaka T. (2005). Genome-wide association study to identify SNPs conferring risk of myocardial infarction and their functional analyses. Cell Mol. Life. Sci. 62, 1804-1813.

Peltomaki P., Aaltonen L.A., Sistonen P., Pylkkanen L., Mecklin J.P., Jarvinen H., Green J.S., Jass J.R, Weber J.L., Leach F.S., Petersen G.M., Hamilton S.R., de la Chapelle A. and Vogelstein B. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. Science 260, 810-812.

Peters E.J., Slager S.L., McGrath P.J., Knowles J.A. and Hamilton S.P. (2004). Investigation of serotonin-related genes in antidepressant response. Mol. Psychiatry 9, 879-889.

Pharoah P.D., Antoniou A., Bobrow M., Zimmern R.L., Easton D.F. and Ponder B.A. (2002). Polygenic susceptibility to breast cancer and implications for prevention. Nat. Genet. 31, 33-36.

Pharoah P.D., Dunning A.M., Ponder B.A. and Easton D.F. (2004). Association studies for finding cancer-susceptibility genetic variants. Nat. Rev. Cancer 4, 850-860.

Risch N. and Merikangas K. (1996). The future of genetic studies of complex human diseases. Science 273, 1516-1517.

Risch N.J. (2000). Searching for genetic determinants in the new millennium. Nature 405, 847-856.

Shiffman D., Ellis S.G., Rowland C.M., Malloy M.J., Luke M.M., Iakoubova O.A., Pullinger C.R., Cassano J., Aouizerat B.E., Fenwick R.G., Reitz R.E., Catanese J.J., Leong D.U., Zellner C., Sninsky J.J., Topol E.J., Devlin J.J. and Kane J.P. (2005). Identification of four gene variants associated with myocardial infarction. Am. J. Hum. Genet. 77, 596-605.

Simpson N.E., Kidd K.K., Goodfellow P.J., McDermid H., Myers S., Kidd J.R., Jackson C.E., Duncan A.M., Farrer L.A., Brasch K., Castiglione C., Genel M., Gertner J., Greenberg C.R., Gusella J.F., Holden J.J.A. and White B.N. (1987). Assignment of multiple endocrine neoplasia type 2A to chromosome 10 by linkage. Nature 328, 528-530.

Song H., Ramus S.J., Kjaer S.K., Hogdall E., Dicioccio R.A., Whittemore A.S., McGuire V., Hogdall C., Jacobs I.J., Easton D.F., Ponder B.A., Dunning A.M., Gayther S.A. and Pharoah P.D. (2007). Tagging Single Nucleotide Polymorphisms in the BRIP1 Gene and Susceptibility to Breast and Ovarian Cancer. PLoS ONE. 2, e268.

Song H., Ramus S.J., Quaye L., DiCioccio R.A., Tyrer J., Lomas E., Shadforth D., Hogdall E., Hogdall C., McGuire V., Whittemore A.S., Easton D.F., Ponder B.A., Kjaer S.K., Pharoah P.D. and Gayther S.A. (2006a). Common variants in mismatch repair genes and risk of invasive ovarian cancer. Carcinogenesis 27, 2235-2242.

Song H., Ramus S.J., Shadforth D., Quaye L., Kjaer S.K., Dicioccio R.A., Dunning A.M., Hogdall E., Hogdall C., Whittemore A.S., McGuire V., Lesueur F., Easton D.F., Jacobs I.J., Ponder B.A., Gayther S.A. and Pharoah P.D. (2006b). Common variants in RB1 gene and risk of invasive ovarian cancer. Cancer Res. 66, 10220-10226.

Stacey S.N., Manolescu A., Sulem P., Rafnar T., Gudmundsson J., Gudjonsson S.A., Masson G., Jakobsdottir M., Thorlacius S., Helgason A., Aben K.K., Strobbe L.J., Albers-Akkers M.T., Swinkels D.W., Henderson B.E., Kolonel L.N., Le Marchand L., Millastre E., Andres R., Godino J., Garcia-Prats M.D., Polo E., Tres A., Mouy M., Saemundsdottir J., Backman V.M., Gudmundsson L., Kristjansson K., Bergthorsson J.T., Kostic J., Frigge M.L., Geller F., Gudbjartsson D., Sigurdsson H., Jonsdottir T., Hrafnkelsson J., Johannsson J., Sveinsson T., Myrdal G., Grimsson H.N., Jonsson T., von Holst S., Werelius B., Margolin S., Lindblom A., Mayordomo J.I., Haiman C.A., Kiemeney L.A., Johannsson O.T., Gulcher J.R., Thorsteinsdottir U., Kong A. and Stefansson K. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat. Genet. 39, 865-869.

The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661-678.

Thompson D. and Easton D. (2004). The genetic epidemiology of breast cancer genes. J. Mammary Gland Biol. Neoplasia 9, 221-236.

Wacholder S., Chanock S., Garcia-Closas M., El Ghormli L. and Rothman N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J. Natl. Cancer Inst. 96, 434-442.

Wang W.Y., Barratt B.J., Clayton D.G. and Todd J.A. (2005). Genome-wide association studies: theoretical and practical concerns. Nat. Rev. Genet. 6, 109-118.

Wooster R., Bignell G., Lancaster J., Swift S., Seal S., Mangion J., Collins N., Gregory S., Gumbs C. and Micklem G. (1995). Identification of the breast cancer susceptibility gene BRCA2. Nature 378, 789-792.

Wooster R., Neuhausen S.L., Mangion J., Quirk Y., Ford D., Collins N., Nguyen K., Seal S., Tran T., Averill D., Fields P., Marshall G., Narod S., Lenoir G.M., Lynch H., Feunteun J., Devilee P., Cornelisse C.J., Menko F.H., Daly P.A., Ormiston W., McManus R., Pye C., Lewis C.M., Cannon-Albright L.A., Peto J., Ponder B.A.J., Skolnick M.H.,

Easton D.F., Goldgar D.E. and Stratton M.R. (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 265, 2088-2090.

Wu X., Gu J., Grossman H.B., Amos C.I., Etzel C., Huang M., Zhang Q., Millikan R.E., Lerner S., Dinney C.P. and Spitz M.R. (2006). Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. Am. J. Hum. Genet. 78, 464-479.

Yeager M., Orr N., Hayes R.B., Jacobs K.B., Kraft P., Wacholder S., Minichiello M.J., Fearnhead P., Yu K., Chatterjee N., Wang Z., Welch R., Staats B.J., Calle E.E., Feigelson H.S., Thun M.J., Rodriguez C., Albanes D., Virtamo J., Weinstein S., Schumacher F.R., Giovannucci E., Willett W.C., Cancel-Tassin G., Cussenot O., Valeri A., Andriole G.L., Gelmann E.P., Tucker M., Gerhard D.S., Fraumeni J.F. Jr, Hoover R., Hunter D.J., Chanock S.J. and Thomas G. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat. Genet. 39, 645-649.

Zondervan K.T. and Cardon L.R. (2004). The complex interplay among factors that influence allelic association. Nat. Rev. Genet. 5, 89-100.

Accepted October 1, 2007