Cancer genetic association studies in the genome-wide age

Sharon A Savage
Clinical Genetics Branch, Division of Cancer, Epidemiology and Genetics, National Cancer Institute,
6120 Executive Blvd, EPS/7018, Rockville, MD 20892, USA, Tel.: +1 301 496 5785, Fax: +1 301
496 1854, savagesh@mail.nih.gov

Abstract
Genome-wide association studies of hundreds of thousands of SNPs have led to a deluge of studies
of genetic variation in cancer and other common diseases. Large case–control and cohort studies
have identified novel SNPs as markers of cancer risk. Genome-wide association study SNP data have
also advanced understanding of population-specific genetic variation. While studies of risk profiles,
combinations of SNPs that may increase cancer risk, are not yet clinically applicable, future, large-
scale studies will make individualized cancer screening and prevention possible.

Keywords
cancer; cancer risk; genome-wide association study; genomics; population genetics; SNP

Genetic risk of cancer
Cancer is a complex, multifactorial disease with several well-described risk factors, such as
smoking and lung cancer [1], infection with human papilloma virus and cervical cancer [2],
and increasing age [3]. In addition to these factors, there is growing evidence that common and
rare genetic variants also contribute to cancer risk.

Studies of cancer-prone families have led to important advances in cancer biology. For
example, Knudson’s original description of the inheritance of retinoblastoma led to the
identification of mutations in the RB gene, a tumor suppressor and transcriptional regulator, in
these patients [4,5]. The study of families with Li–Fraumeni syndrome led to the identification
of mutations in the TP53 gene, which has since been shown to be a critical transcription factor
in normal cell growth, apoptosis and DNA repair [6-9]. Careful epidemiological study led to
the identification of familial breast cancer pedigrees and the discovery of the BRCA1 and
BRCA2 genes, which are important in DNA repair and highly penetrant breast and ovarian
cancer risk factors [10-12].

However, despite these advances, our understanding of cancer genetic risk factors is still
limited because the majority of people who develop cancer do not have a cancer predisposition
syndrome. Instead, it appears to be a polygenic disease that can cluster in families but does not
demonstrate a simple inheritance pattern. In this case, the combination of several common
genetic variants could contribute to increased cancer risk in certain populations. The common
disease/common variant hypothesis seeks to understand the effects of common genetic
variation (typically >1% minor allele frequency) that could contribute to polygenic disorders,
such as cancer. The basis of this hypothesis is that there may be several common variants that
together contribute to the risk of common disease.
Growing from candidate genes to genome-wide association studies

SNPs are the most common form of genetic variation in the genome. Of the roughly 3 billion nucleotide bases in the genome, it is estimated that 10 million of these sites (1 in 300) are SNPs with a minor allele frequencies of at least 1% [13]. Other types of genetic variation between individual human genomes include insertions and deletions (from a single base to thousands of bases) and copy-number variation. Advances in understanding these genetic variants are reviewed elsewhere [14].

Gene sequencing and genotyping initiatives, such as the International HapMap Project [13], the SeattleSNP Variation Discovery Resource [101] and the SNP500Cancer project [15], led to improved understanding of population-specific genetic variation and the haplotype structure of the human genome. The presence of linkage disequilibrium, or the nonrandom segregation of alleles, has made it possible to choose SNPs that can serve as surrogates for untested markers by using the basic principle that shared haplotypes contain variants that track together across generations. Through the use of haplotype-tagging methods, based on various statistical algorithms, fewer SNPs can be genotyped and still provide a thorough interrogation of the genetic variation across a gene, a specific locus or the entire genome [16,17].

Technological advances have, in part, driven the advancement from single SNP studies to genome-wide association (GWA) studies of hundreds of thousands of SNPs. In less than 10 years, SNP assays have expanded from the time-consuming restriction fragment length polymorphism (RFLP) methods that evaluated one SNP at a time, to high-throughput genotyping of GWA studies on tens of thousands of samples. The next step, complete genome sequencing, will generate a more detailed map of human genetic variation. The 1000 Genomes Project, an international research consortium, will build upon the HapMap project by completely sequencing the genomes of 1000 individuals [102]. This project seeks to identify genetic variants with frequencies of 0.5–1% in the population studied. This high-resolution map will be important in understanding the genetic contributions to health and illness. Pilot studies in a small number of individuals are underway to evaluate the best approaches for this large-scale project.

Replicating candidate gene associations

Until recently, the majority of genetic association studies were small, candidate gene or pathway studies of a relatively small number of SNPs. An a priori hypothesis regarding the role the gene of interest played in the pathogenesis of the disease was considered to be a crucial factor in study design. Using this approach, associations of increased gastric cancer risk and the -251 T/A promoter polymorphisms in the interleukin 8 gene, a proinflammatory chemokine, were replicated in several studies [18-21], but appear to be population-dependent [22,23].

Studies such as these were encouraging, but the majority of such studies have not been reproducible. In a comprehensive review of over 600 positive association studies, only 166 associations were studied more than three-times and only six were consistently replicated [24]. Similarly, a meta-analysis of 25 different associations in 201 genetic association studies of complex diseases showed evidence of replication in less than half of the studies [25].

A recent study of 161 meta-analyses and pooled analyses of genetic polymorphisms and cancer risk evaluated 344 gene-variant associations from the literature [26]. Of these, 98 gene-variant associations were reported as statistically significant in the published studies. Analyses of these studies using the False Positive Report Probability (FPRP) method [27], which incorporates biological plausibility into SNP data interpretation and is hypothesized to increase the probability of find a true association, the authors found that 13 and 4 gene-variant associations were reproducible at FPRP values of 0.001 and 0.000001, respectively. Genes encoding...
metabolizing enzymes were among the most consistently associated with cancer risk in that study.

Readily-accessible genotype datasets that are annotated with key clinical features from numerous studies are now critical for the continued advancement of genetic association research. The Genetic Association Database [103], which compiles data from published association studies, is increasingly useful in guiding study design, SNP selection, and genetic association study replication. Genetic association studies are also being compiled through the Human Genome Epidemiology Network (HuGeNet) [28]. HuGeNet forms the basis for meta-analyses of genetic association studies in cancer and other diseases. The systematic analysis of these studies will help to identify the true-positive associations and potentially generate leads for future studies.

There are numerous factors that may contribute to these reproducibility challenges. Publication bias, the tendency of scientists and journals to publish positive findings, plays a role. This is changing as more journals are reviewing and publishing null results in genetic-association studies [29]. Sample size and multiple testing comparisons may limit the statistical power of the study. Cancer is a heterogeneous disease and careful pathologic classification of cases is necessary. Matching of controls based on ethnic groups and other population genetic factors is crucial but also prone to error. There can also be some error in genotyping results due to factors including variable DNA quality, assay conditions and others. The candidate gene/pathway approach assumes that current knowledge of the cancer biology related to these genes is relevant, however, there are likely numerous presently unknown genes and pathways that contribute to cancer risk. Cautious interpretation of these and other such epidemiological studies has been encouraged [30,31].

**Genome-wide association studies in cancer**

While the candidate gene approach has not been abandoned, the availability of whole-genome technologies has led investigators to take an agnostic approach; assume we know nothing about the genetics of the disorder, compile a sufficiently large sample size, and test as many SNPs as possible, without any a priori biological hypothesis. This has long been the approach of family-based linkage studies, which evaluate differences in the inheritance patterns of genetic markers (microsatellites or SNPs) to find loci associated with disease. The strength of such family-based studies is the relatedness of the subjects; however, they require large families and/or large numbers of families with the same disorder and do not always identify causative loci.

The present approach with GWA studies continues to evaluate and expand upon the common disease/common variant hypothesis. It takes advantage of high-throughput genotyping methods to evaluate a very large number of common SNPs. The current genotyping platforms utilize common SNPs (>5% minor allele frequency) chosen through haplotype tagging algorithms and are based on data from the HapMap and other sequencing and genotyping initiatives.

The major strengths of a GWA study are that common SNPs are comprehensively interrogated across the entire genome, and due to this, it is hypothesis generating, rather than driven by a priori assumptions [32,33]. The challenges in genome-wide scans are largely due to the enormous size of the dataset and statistical methods [34]. GWA studies of common disease in which genetic variation may contribute relative risks of less than 1.5 require very large sample sizes and several replication strategies. Owing to the large number of statistical tests needed in these datasets, very high significance thresholds and replication are required. These efforts have resulted in the formation of large national and international consortia in which cases and controls are pooled.
A summary of the growing list of cancer GWA studies is presented in Table 1. Thus far, the majority of studies have evaluated common cancers, such as breast, prostate, and colon cancer. GWA studies identified SNPs in the 8q24.21 locus as associated with prostate and colon cancer [35-39]. Subsequent studies suggest that certain variants in this region are also associated with increased breast, and ovarian cancer risk [37,40]. This region at 8q24.21 is a ‘gene desert’ whose functional importance is yet to be discerned. Additional collaborative efforts between investigators in genomics, population genetics, epidemiology, molecular biology and other basic sciences will be critical to understanding the role of this region in cancer risk.

Two GWA studies of breast cancer identified SNPs in a biologically relevant gene, **FGFR2**, as associated with increased risk. This is particularly encouraging because the **FGFR2** gene encodes a receptor, tyrosine kinase, that is amplified or overexpressed in some breast cancers [41,42]. Other breast cancer GWA studies have found at least six loci associated with relatively modestly increased risk (Table 1).

To date, the majority of cancer genome-wide scans have been conducted in common cancers. Interestingly, a recent GWA study of neuroblastoma, a rare childhood tumor, found increased risk of neuroblastoma with certain SNPs in the 6p22.3 region [43]. Comprehensive genetic-association studies of childhood cancer have the potential for high-yield findings. It is possible that pediatric cancer risk has a stronger genetic component than cancers that affect the elderly, primarily because the latency period of environmental exposures that could contribute to pediatric cancer risk is significantly shorter. Several GWA studies of pediatric cancer are currently being planned to investigate this hypothesis.

**Genome-wide scans & population genetics**

Genome-wide scans are valuable beyond studies of disease risk. Population-specific genetic variation can be widely variable between different ethnic groups. The differences in genetic variation between ethnic groups due to a combination of evolutionary history, migration and population admixture are now incorporated into study design and analysis. Genome-wide analyses have further advanced this understanding.

For example, the genetic relationships between 938 unrelated individuals from 51 populations were studied using a panel of 650,000 SNPs [44]. This study showed that the relationship between haplotype heterozygosity and geography was consistent with serial founder effect originating in sub-Saharan Africa. An evaluation of 388,654 SNPs from 102 subjects in the SNP500Cancer panel showed that even with modest sample sizes, genes with high genetic distance between subpopulations could be identified [45]. In a larger subanalysis of two GWA studies in European Americans, genotyped in the Cancer Genetic Markers of Susceptibility (CGEMS) project, principal component analyses were used to identify sets of SNPs that can identify and be used to correct population stratification [46].

A challenge in the design of GWA studies is obtaining an appropriate number of cases and controls for adequate statistical power after correcting for multiple comparisons. One approach to increasing statistical power is through increasing the number of control subjects. Population genetic analyses of SNPs in GWA studies have suggested that sharing ‘genomic’ controls is a valid approach for increasing control sample size. This method evaluates the population-specific genetic variation in the control groups of several studies and tests for genetic similarities. Based on these analyses, subjects can be matched based on their SNP profiles. This approach was used in a large study of seven diseases in 14,000 cases and 3000 shared controls [47]. Previously known as well as novel risk loci were identified in diabetes, coronary artery disease, Crohn’s disease, and rheumatoid arthritis. The pooling of such controls could be used in studies of rare cancers in which case samples are most readily available through clinical trial networks and standard matched controls are not available. A major limitation of
shared genomic controls is that it is not possible to include environmental risk factors in the analyses.

Additional benefits of the genome-wide age are the formation of large international collaborations, the advancement of ‘team’ science, and data sharing. The NIH recently implemented a genome-wide association study data-sharing policy [104] for its grantees and intramural investigators that articulates the expectation that investigators will make such datasets available in the central NIH repository of genome-wide association study data, the Database of Genotypes and Phenotypes (dbGaP) [105]. Posting should occur as soon as the appropriate quality-control assessments are complete. The primary investigators will have a 12-month period of exclusive data access to permit them to submit genome-wide association studies dataset analyses for publication.

Clinical translation: are we ready?

The success of GWA studies in cancer and other diseases has helped to advance to the concept of genomic medicine. Goals of the clinical translation of genomic research include:

- Risk prediction and disease prevention based on these risks
- Improved mechanistic understanding of disease
- Development of preventive and/or therapeutic interventions
- Medication response and side effects (pharmaco-genomics)

On an individual patient level, this is challenging because the majority of individual SNP associations resulting from GWA studies of cancer risk have odds ratios less than 1.5 and relatively low population attributable risks (Table 1). For a large population, an odds ratio of 1.2 could have public health implications, but until an intervention has been proven to be effective for that population, translation of these risks to large or small populations will be limited.

Attempts are being made to evaluate the effects of the presence or absence of several alleles conferring a very modest increased risk with the overall goal to develop a test that is relevant to both the general population and the individual patient. The potential utility of a breast cancer risk profile based on seven SNPs generated in GWA studies was recently evaluated [48]. The authors created risk profiles based on these SNPs and evaluated the potential usefulness of the profiles in identifying women at increased risk of breast cancer who would benefit from cancer screening programs. They concluded that the risk profiles generated by these common moderate-risk alleles did not sufficiently identify individual women at increased risk who would benefit from individualized prevention. However, they suggest that if genotype data were available on all women in a large population the risk profiles could be useful at the population screening level.

The 8q24 locus and prostate cancer risk has been an active area of investigation at the individual level. Several SNPs in this region were identified in GWA studies and confirmed in several other studies that focused only on SNPs in that region [49]. The associations are statistically strong but most had small effect sizes (i.e., odds ratios were generally <1.5). The 8q24 SNP, rs1447295, had an odds ratio of prostate cancer of 1.93 for familial prostate cancer patients compared with control subjects [50]. In another study, men with a combination of two SNPs in 8q24 were 2.7-times more likely to develop prostate cancer than men with neither risk allele [51]. Subsequent studies of the same loci, suggested that men with a certain combination of five SNPs from 8q24, 17q12, and 17q23.3 had a ninefold increased risk of prostate cancer and that this was worth consideration as a clinically relevant test [52]. This led to much debate in
the literature, which made it clear that confirmatory studies and careful consideration of the global implications were warranted before such a test should move forward [53-60].

This debate continues to grow as several companies have now made available direct-to-consumer SNP panels that genotype the same SNPs as many of the GWA studies described here. The interested individual can, for a fee, have their SNP ‘profile’ determined and obtain a list of diseases and their level of risk [61]. The clinical utility and public health impact of such data is not proven. Testing women with a strong family history of breast and ovarian cancer for BRCA1/2 mutations has implications in cancer screening and even decisions for prophylactic surgery [62]. However, the effect of a SNP or SNPs that may increase an individual’s risk by 1.5-fold over the general population is not known.

Owing to these concerns, a framework for the translation of genomic research into healthcare and disease prevention was proposed which outlines suggested phases of this research [63]. Four translation research phases were proposed:

- Phase 1 translation – discovery
- Phase 2 translation – health application to evidence-based guidelines
- Phase 3 translation – practice guidelines to health practice
- Phase 4 translation – practice to population health impact

Within Phase 2 translation, the analytic and clinical validity, clinical utility, and ethical (ACCE), legal and social implications must also be considered.

Conclusion & future perspective

The rapid advancement of genomic findings in common diseases has led to expectations of clinical utility in the near future. However, there are several issues that need to be addressed before personalized prediction of cancer risk can be achieved. The dawn of the genome-wide age has led to a deluge of data that must be rigorously tested and replicated in studies with high statistical power. We must remember that SNPs are markers of the genome and are not likely to be causal on their own. To that end, thoughtful collaboration between basic scientists, genomicists, bioinformaticians and epidemiologists is required to advance understanding of the genetic basis of cancer risk and its contribution to cancer biology.

In order to follow-up on recent findings, comprehensive, prospective clinical studies of large cohorts will need to be conducted before the widespread application of genomic medicine occurs. In the next 5–10 years, as the body of GWA studies of cancer continues to grow, individual risk profiles based on a combination of a large number of SNP markers will be created and tested. It will be possible to adjust screening practices for common cancers, such as breast or prostate cancer, based on an individual’s genetic cancer risk profile. However, in order for genomics to have a positive impact on disease prevention and/or treatment careful study design, and consideration of the ethical, and legal consequences such studies must be included. When that occurs, the goal of patient-specific screening and/or intervention, informed by individual genotype data, may be achievable.

Executive summary

- Both common and rare genetic variants may contribute to cancer risk.
- Advances in genotyping technology have made large-scale genome-wide association (GWA) studies possible.
Prior to the advent of GWA studies, most studies of SNPs and cancer risk were not reproducible.

GWA studies comprehensively evaluate common (>5% minor allele frequency) genetic variation across the genome. They are hypothesis generating rather than driven by a priori assumptions.

Numerous novel risk loci have been identified in GWA studies of cancer.

The large number of SNPs evaluated in GWA studies has advanced our understanding of population-specific genetic variation and made the use of shared ‘genomic’ controls possible.

Collaboration between basic scientists, genomicsists, bioinformaticians and epidemiologists is required to advance understanding of the genetic basis of cancer risk and its contribution to cancer biology.

The creation of cancer risk profiles based on common genetic variants identified in GWA studies are not yet clinically applicable to individuals, but hold great promise for individualized cancer prevention.

Acknowledgments

Financial & competing interests disclosure

This research was supported by the Intramural Research Program of the National Institutes of Health and the National Cancer Institute. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Bibliography

Papers of special note have been highlighted as: • of interest


Per Med. Author manuscript; available in PMC 2009 September 1.


47. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447(7145):661–678, 2007; [PubMed: 17554300] • One of the largest studies thus far to use genomic controls.


Websites

102. 1000 Genomes www.1000genomes.org
## Table 1

Overview of cancer genome-wide association studies.

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Cases/controls in first stage (n)</th>
<th>Number of SNPs analyzed in first stage</th>
<th>Total number of cases/controls in study</th>
<th>Gene or locus</th>
<th>Estimated attributable risk (^%)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>1145/1142</td>
<td>528,173</td>
<td>2921/3214</td>
<td>FGFR2</td>
<td>16</td>
<td>[41]</td>
</tr>
<tr>
<td></td>
<td>390/364</td>
<td>205,586</td>
<td>26,258/26,894</td>
<td>FGFR2, TNRC9, MAP3K1, LSP1</td>
<td>5.5–10.5</td>
<td>[64]</td>
</tr>
<tr>
<td></td>
<td>249/299</td>
<td>150,800</td>
<td>1442/1465</td>
<td>6q22.33 (ECHDC1, RNF146)</td>
<td>7</td>
<td>[65]</td>
</tr>
<tr>
<td></td>
<td>1599/11,546</td>
<td>311,524</td>
<td>4533/17,513</td>
<td>2q35, 16q12 (TNRC9)</td>
<td>13–14#</td>
<td>[66]</td>
</tr>
<tr>
<td>Colorectal</td>
<td>930/960#</td>
<td>547,647</td>
<td>8264/6206</td>
<td>8q24.21</td>
<td>20</td>
<td>**[67]</td>
</tr>
<tr>
<td></td>
<td>1257/1336</td>
<td>99,632</td>
<td>7480/7779</td>
<td>8q24.21</td>
<td>NR</td>
<td>[35]</td>
</tr>
<tr>
<td></td>
<td>922/927#</td>
<td>547,647</td>
<td>18,794/18,453</td>
<td>8q23.3, 10p14</td>
<td>NR</td>
<td>**[68]</td>
</tr>
<tr>
<td></td>
<td>940/965#</td>
<td>547,647</td>
<td>8413/6949</td>
<td>18q21 (SMAD7)</td>
<td>15</td>
<td>**[69]</td>
</tr>
<tr>
<td></td>
<td>981/1002#</td>
<td>541,628</td>
<td>17,457/16,353</td>
<td>11q23, 8q24.21, 18q21 (SMAD7)</td>
<td>3.3–9.6</td>
<td>[70]</td>
</tr>
<tr>
<td>Gastric</td>
<td>188/752</td>
<td>85,576</td>
<td>932/1398</td>
<td>PSCA</td>
<td>NR</td>
<td>[71]</td>
</tr>
<tr>
<td>Lung</td>
<td>1926/2522</td>
<td>310,023</td>
<td>4493/7274</td>
<td>CHRNA3, CHRNA5, CHRNA6##</td>
<td>14</td>
<td>[72]</td>
</tr>
<tr>
<td></td>
<td>10,995##</td>
<td>306,207</td>
<td>1024/32,244</td>
<td>CHRNA3, CHRNA5, CHRNA6##</td>
<td>18</td>
<td>[73]</td>
</tr>
<tr>
<td></td>
<td>1154/1137</td>
<td>317,498</td>
<td>3878/4831</td>
<td>CHRNA##, CRP, IL1RAP</td>
<td>NR</td>
<td>[74]</td>
</tr>
<tr>
<td>Melanoma</td>
<td>864/864</td>
<td>535,150</td>
<td>2094/2215</td>
<td>20q11.22 (CDC91L1)</td>
<td>11</td>
<td>[75]</td>
</tr>
<tr>
<td>Neuroblastoma</td>
<td>1032/2043</td>
<td>464,934</td>
<td>1752/4171</td>
<td>6p22.3 (FLJ22536, FLJ44180)</td>
<td>NR</td>
<td>[43]</td>
</tr>
<tr>
<td>Prostate</td>
<td>1435/3064</td>
<td>316,515</td>
<td>3018/5881</td>
<td>8q24.21</td>
<td>10–24</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td>1501/11,290</td>
<td>310,520</td>
<td>3493/14,348</td>
<td>17q12 (HNF1B##), 17q24.3</td>
<td>36***</td>
<td>[76]</td>
</tr>
<tr>
<td></td>
<td>1854/21,372</td>
<td>310,520</td>
<td>10,093/28,962</td>
<td>2p15, Xp11.22</td>
<td>5 to 7</td>
<td>[77]</td>
</tr>
<tr>
<td></td>
<td>4044</td>
<td>11,555</td>
<td>1454/1636</td>
<td>1q25, 7p21</td>
<td>NR</td>
<td>[78]</td>
</tr>
<tr>
<td></td>
<td>1172/1157</td>
<td>538,548</td>
<td>4296/4299</td>
<td>8q24.21</td>
<td>21</td>
<td>[36]</td>
</tr>
<tr>
<td></td>
<td>1235/1599##</td>
<td>60,275</td>
<td>2477/2516</td>
<td>DAB2IP</td>
<td>NR</td>
<td>[79]</td>
</tr>
</tbody>
</table>

\* Total number of cases/controls in study includes cases and controls from multiple studies. \n\# Estimated attributable risks are based on meta-analysis. \n\‡‡‡ Gene or locus listed in the cancer genome-wide association studies. \n
---

**Note:** Per Med. Author manuscript available in PMC 2009 September 1.
<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Cases/controls in first stage (n)</th>
<th>Number of SNPs analyzed in first stage</th>
<th>Total number of cases/controls in study*</th>
<th>Gene or locus</th>
<th>Estimated attributable risk** (%)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate</td>
<td>1172/1157</td>
<td>527,869</td>
<td>5113/5121</td>
<td>8q24, 17q (<strong>HNF1B</strong>), chr10 (<strong>MSMB, CTBP2</strong>), chr7 <strong>JAZF1</strong></td>
<td>8–20</td>
<td>[80]</td>
</tr>
<tr>
<td></td>
<td>1854/1894</td>
<td>541,129</td>
<td>5122/5260</td>
<td>8q24, 17q, other loci on chr 3, 6, 7, 10, 11, 19, X</td>
<td>15</td>
<td>[39]</td>
</tr>
</tbody>
</table>

This is a general overview and is not intended to be a comprehensive description of the data.

* Includes number of cases and controls in the original scan plus those in the replication studies.

† This is the value, or range of values, reported in the manuscript for the population that was studied. Different methods to calculate the population attributable risk or fraction may have been used and as such, the values may not be directly comparable.

§ Included subjects with a family history of the cancer type studied.

¶ In the first stage cases were high-risk, BRCA1/2 mutation-negative Ashkenazi Jewish women with breast cancer who also had at least three family members with breast cancer. Controls were cancer-free Ashkenazi Jewish controls. In the replication studies, cases were Ashkenazi Jewish breast cancer cases.

# Combined population attributable risk for SNPs in both regions was 25%.

** Studies derived from familial colorectal cancer cases. The authors also reported data on subjects with adenoma only.

‡‡ This study included early-onset colon cancer cases.

§§ This study sought to identify genes that were associated with smoking, an important cause of lung cancer. The genes associated were nicotinic acetylcholine receptors and associated with both smoking and lung cancer.

## **HNF1B** was previously known as **TCF2**.

### **CHRNA5, CHRNA3 and CHRN** have also been associated with nicotine dependence.

*** This was a joint population attributable risk due to the combination of the two high-risk SNPs, rs4430796 and rs1859962.

‡‡‡ Combined SNPs from a scan of 498 cases and 494 controls with cases and controls from the CGEMS study. The SNPs reported were SNPs that overlapped between the two studies.

CGEMS: Cancer Genetic Markers of Susceptibility; chr: Chromosome; CI: Confidence interval; NR: Not reported.