Lessons from Functional Analysis of Genome-Wide Association Studies

Inderpreet Sur1,2, Sari Tuupanen3, Thomas Whitington1, Lauri A. Aaltonen3, and Jussi Taipale1,4

Abstract
Most cancer-associated single-nucleotide polymorphisms (SNP) identified using genome-wide association studies are located outside of protein-coding regions, and their significance and mode of action have been a source of continuing debate. One proposed mechanism of action of the SNPs is that they would affect the activity of enhancer elements regulating critical target genes. In this review, we summarize recent results that substantiate this model. These studies have identified a cancer-specific enhancer element at the 8q24 gene desert that controls the expression of the MYC oncogene. We further discuss implications of the observed difference between normal growth control and cancer for drug development, and the inherent features of genome-wide association studies that may specifically lead to identification of disease-specific regulatory elements. Cancer Res; 73(14); 4180–4. ©2013 AACR.

Introduction
Genome-wide association studies (GWAS) have identified more than 150 loci associated with increased susceptibility to cancer, according to the Catalog of Published Genome-wide Association Studies (http://www.genome.gov/gwastudies/ or ref. 1 for review). The susceptibility alleles identified by GWAS are generally common (minor allele frequency >10%) and confer small risks (OR < 1.5). Most of the susceptibility regions are located outside of known protein-coding sequences, and often it is hard to associate a single-nucleotide polymorphism (SNP) to a particular gene. In most cases, it is thought that the GWAS-identified SNPs do not themselves affect disease risk, but merely identify a region where a causative variant is located.

The genomic region on 8q24 is one of the most interesting regions pinpointed by GWAS because it harbors multiple risk loci for cancer (Fig. 1). For example, multiple independent risk loci for prostate cancer are located at 8q24 (2–6), and predisposition alleles to several other epithelial cancers are also found in distinct linkage disequilibrium blocks within this region. An intriguing aspect of the risk alleles is that they are located in a 1.2-Mb “gene desert” region. This region is part of a single chromatin topologic domain (Fig. 1; ref. 7) that contains the MYC gene. Because MYC is overexpressed in many tumors, it serves as the most likely candidate target gene.

Although the effects of most polymorphisms at 8q24 are small, they are among the strongest SNPs identified in cancer GWAS. Furthermore, some SNPs have a very high allele frequency, such that they account for a large number of cancer cases at the population level. For example, the polymorphism rs6983267 linked to colorectal cancer (CRC; ref. 8) and prostate cancer (6) contributes more to cancer morbidity and mortality than any other known inherited variant or mutation, including the classic high-penetration tumor suppressors such as RB, TP53, and APC.

Role of GWAS SNPs in Gene Regulation
How can the SNPs located far from any oncogene or tumor suppressor alter cancer susceptibility? A possible mechanism of action is thought to be via the influence of these SNPs on distal enhancer elements that regulate the expression of critical target genes (9–12). Evidence exists for regulatory activity of SNPs located in several regions, including 8q24, 8q21, and 17q24 (13, 14).

Several lines of evidence have suggested that the CRC predisposition SNP rs6983267 functions in such a way. We and others (9, 10) found earlier that rs6983267 resides in a binding site for T-cell factor 4 (TCF4; HUGO name TCF7L2), a crucial downstream effector of the Wnt signaling pathway. The SNP is located within a computationally predicted enhancer element MYC-335 (10). The MYC-335 element was subsequently shown to possess enhancer activity both in vitro and in vivo. It also carries histone mark H3K4me1 characteristic of enhancers, and the region containing rs6983267 physically interacts with the MYC promoter in CRC cell lines (9). Most importantly, the risk allele G creates a stronger binding site for TCF7L2 compared with the nonrisk allele T, thus providing a plausible mechanism by which this SNP affects cancer susceptibility (10). It is currently not known to what fraction of the GWAS SNPs function in this way. However, statistical enrichment analyses that indicate that GWAS-identified regions are
Research Reviews

enriched in expression quantitative trait loci (11, 15) and DNase I hypersensitive sites (12) suggest that considerable number of GWAS SNPs may act through a similar mechanism.

Reverse Genetic Analysis of 8q24 in the Mouse

The MYC-335 element is highly conserved in mammalian species, and so is the gene order in this region, with the exception of the POU5F1 pseudogene that is only present in humans. The binding of TCF7L2 to the site affected by rs6983267 is also conserved (16) at least in mice, which carry the risk allele. This made it possible to analyze the role of MYC-335 in cancer and MYC expression using a mouse model. We thus generated a knockout mouse in which the Myc-335 region was deleted. Consistent with the predicted role of MYC-335 in tissue-specific regulation of MYC, we observed a modest decrease in Myc expression in the colon of newborn mice. The mice were viable and fertile and developed normally under laboratory conditions. However, when challenged with the Apc^min^ mutation that confers strong, TCF7L2-dependent intestinal tumor predisposition, the mice showed a remarkable reduction in the frequency of intestinal tumor development (16).

These results have several important implications: (i) regulatory elements affected by SNPs can be much more important for tumorigenesis than suggested by the small effect of the SNP itself, (ii) normal growth and pathologic growth use different regulatory mechanisms, and (iii) GWASs are specifically biased to identify such disease-specific variants.

Even SNPs with Weak Effects Can Identify Central Mechanisms of Cancer

The weak effect of most SNPs identified in GWAS has elicited a lot of criticism of the methodology. The effects seen are
mostly so weak that their relevance to disease prediction and prevention is limited.

However, the SNPs can mark a region that is very important in cancer; they may for instance identify a region containing a causative mutation that is more rare and has higher penetrance. High-throughput whole-genome sequencing techniques have already identified one such SNP in 8q24, rs188140881, that confers relatively high risk (OR = 2.90) for prostate cancer (17). Such SNPs are more important for disease prediction and prevention, and more relevant for cancer screening. In addition, the SNPs may interact with each other epistatically to yield higher risk. Given that only approximately 5% of familial risk in most cancers can be attributed to common SNPs, such interactions between SNPs are a potential source of the "missing" familial risk (18).

In addition, even a weak SNP can pinpoint a regulatory element whose activity is critical for tumorigenesis. This is because it is possible that the weak SNP is weak simply because it only weakly affects the regulatory element in which it resides. Thus, such SNPs could act in an analogous way to a weak missense mutation in the coding region of an essential gene. The analysis of the effects of Myc-335 loss clearly shows that even though a SNP may lead to only a modest increase in risk, the element in which it resides can have a much stronger effect on disease.

Taken together, in the mechanistic sense, cancer-associated SNPs should be thought of as markers that have identified regions whose role in cancer can be very important or critical. As discussed below, many of the regions identified may also be cancer specific. Thus, further mechanistic studies of cancer susceptibility regions pinpointed by GWAS are clearly warranted.

**GWAS Specifically Identify Disease-Specific Mechanisms**

The importance of gene regulation in cancer is highlighted by the identification of several novel noncoding regions linked to cancer in GWAS. Although the number of novel findings coming from GWAS may appear large, the cancer-associated regions are expected to represent only a small fraction of the large number of regulatory elements in our genome identified by the ENCODE project (19).

Furthermore, our identification of a regulatory element that is not required for normal development, but is important for cancer, is not as surprising as one might initially think. In fact, GWAS may be specifically biased toward identification of such disease-specific functions. In a GWAS, SNPs that have a high allele frequency will more likely be identified with a low P value. However, high allele frequency indicates small effect on fitness, as any SNP that has a strong fitness effect would be relatively rapidly fixed in the population (20). Thus, identifying disease-linked SNPs using GWAS is biased toward finding regulatory elements that have a small effect on fitness and viability but a relatively large effect on the disease analyzed. Thus, investigating mechanisms of cancer-associated SNPs may yield promising new drug targets that are not required for viability, but could be targeted for therapy.

**Normal Functions of the Regulatory Elements at 8q24**

The high prevalence of the risk allele G of rs6983267, and the presence of additional SNPs predicted to increase MYC-335 activity (21) in the African-American population, suggests that increased activity of MYC-335 may provide a small fitness advantage at least in some populations. A biologic role for the gene desert at 8q24 is also suggested by the observation that several regions within it show high sequence conservation between species.

Given the conservation of elements within the 8q24 gene desert in mammals, it seems that this region does have a normal function. But what is it? Several lines of evidence suggest that elements other than MYC-335 also function in tissue-specific gene regulation. A comprehensive epigenetic mapping analysis has identified several enhancer elements within the risk loci that show tissue-specific activity both in vitro and in vivo (22, 23). Apart from the rs6983267, another regulatory SNP has been identified in the enhancers, namely rs11986220, which is strongly correlated with prostate cancer predisposition SNP rs10090154 (2). It affects a FoxA1 binding site within an androgen-responsive enhancer at PrCa-1 region (22). Furthermore, the enhancer elements at 8q24 are coupled to MYC via long-range interactions in a tissue-specific manner. Ahmadiyeh and colleagues (24) showed that the breast cancer risk locus interacts with MYC in the breast cancer line MCF7, but not in the prostate cancer cell line LNCaP. None of the risk loci interacted with MYC in a fibroblast cell line (24). Thus, although MYC is expressed in several tissues, distinct distal enhancers upstream of MYC could regulate its expression in specific tissues and different cancer types.

This still leaves a larger question: Why does all this regulation exist in the first place? As Myc is not required in the adult for most normal proliferation (25–29), it is likely that the biologic role of Myc and the 8q24 risk region has more to do with tissue repair and/or other responses that require temporary increase in proliferation rate.

**Difference between Growth Control in Normal Tissues and in Cancer**

The MYC gene is important for the development and proliferation of multiple types of cancers. Overexpression or deregulated expression of Myc in transgenic animal models causes unrestrained growth resulting in tumors in several tissues (30–33). Myc is also required for growth of normal cells in culture, and cells lacking Myc grow very slowly. The only identified gene that can rescue the slow growth phenotype of rat embryonic fibroblasts lacking Myc is Myc or its paralog N-Myc (34), and targeting Myc in cultured cells generally leads to growth arrest. However, at least in mice, Myc is dispensable for development until E11.5 with the exception of the hematopoietic lineage (27). Furthermore, although Myc is expressed in the proliferating compartment of several adult tissues, the postnatal deletion of Myc in mice does not result in prominent proliferation defects (25, 26, 28, 29). These results suggest that MYC is more important for growth of tumors and cultured cells than growth and homeostasis of normal tissues. Thus, growth
in cultured cells and cancer on one hand, and normal tissues on the other, seems to be controlled through different mechanisms. Our results have identified a molecular mechanism behind this difference. Myc-335 seems to be specifically required for tumorigenesis but not for normal growth control.

It is possible that cancer and adult tissue repair share mechanisms, whereas normal growth and proliferation might use a different mechanism. The connection between wound repair and cancer has long been suspected, and tumors are proposed to be wounds that do not heal (35). Cultured cells are usually exposed to serum proteins that are normally only present in healing wounds. If cancer cells in vivo and cultured cells in vitro share growth mechanisms that are not used by normal cells in vivo, this has profound implications for cancer drug development. Using cultured cells to assess growth effects or toxicity can result in rejection of lead compounds that specifically target tumors over normal tissues. Thus, cancer-specific compounds should be tested for their effects on normal cells in vivo, or using tissue culture conditions where the growth mechanisms of cells resemble normal tissues rather than healing wounds.

**Targeting MYC**

MYC is a very attractive cancer treatment target, given that it is upregulated in many cancer types and is critical for tumor growth. Compounds targeting MYC might thus be used in a much broader set of patients than the current mechanism-based cancer drugs. It might also be more difficult for a cancer to escape MYC loss, as no other genes except N-MYC and L-MYC are known that can compensate for its activity.

However, MYC is a hard-to-drug transcription factor that lacks small molecule binding pockets. Thus, it might be easier to target upstream or downstream effectors of MYC activation than c-MYC per se. Many current mechanism-based drugs act far upstream of MYC by targeting signaling pathways that drive expression of c-MYC or N-MYC (see, e.g., ref. 36). Genetic evidence in mouse models shows that targeting proteins that act downstream of Myc, for example via ribosomal haploinsufficiency or by targeting ornithine decarboxylase, is a successful strategy without significant adverse effects (37, 38). In addition, targeting a specific function of MYC itself, for instance, by expressing the dominant negative form of Myc (Omomyc) also results in an efficient and safe treatment in transgenic mouse cancer models (39). Our results imply that there might be an alternative way to generate the specificity in targeting MYC for cancer chemoprevention or therapy. It is feasible that in the future one could inhibit upstream mechanisms that control MYC-335 activity, for instance by targeting the mechanisms that regulate the activity of the transcription factors that control MYC-335 activity. If we can target a novel disease-specific enhancer without affecting normal function, it would provide a major advancement in our ability to target critical cancer genes without major side effects to the patients.

**Perspective**

Given that <6 million SNPs exist with allele frequency of >5%, the GWAS screens have not been saturating in the forward genetic sense. Thus, the fact that GWAS has not identified a risk SNP near a gene does not mean that gene has no role in cancer. However, the large number of novel cancer-associated loci identified by GWAS has provided a very rich resource for identification of novel mechanisms of cancer. The recent functional analysis of 8q24 variants has already uncovered a new complexity in regulation of a classic oncogene, MYC. Expansion of such functional analyses to genome-wide scale, using methods such as high-throughput ChIP-seq, chromatin conformation capture (3C), and genome editing using TALENs and/or CRISPRs will be required to make full use of the obtained genetic data.

Much work also remains before we understand the full implications of the mechanistic findings and can translate them to medical benefits. Open questions include the following: (i) What is the normal function of the gene desert at 8q24? (ii) Are all of the GWAS SNPs at 8q24 linked to SNPs or structural variants affecting MYC? (iii) Do SNPs at other loci act via similar mechanisms, (iv) Is MYC-335 activity still required in cancer cells? (v) Can we target it or upstream pathways that drive its activity? One thing is clear. Once again, several new and exciting lines of research have been opened by unbiased genetic analyses of cancer. And again, their value to patients will depend on the mechanistic studies to follow.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Authors’ Contributions**

Conception and design: I. Sur, T. Whitington, L.A. Aaltonen, J. Taipale

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. Tuupanen

Analysis and interpretation of data (e.g., statistical analysis, biosubjects, computational analysis): L.A. Aaltonen

Writing, review, and/or revision of the manuscript: I. Sur, S. Tuupanen, T. Whitington, L.A. Aaltonen, J. Taipale

Study supervision: J. Taipale

**Acknowledgments**

The authors thank Drs. B. Schmierer, K. Dave, and M. Taipale for critical review of the manuscript.

**Grant Support**

This work was supported by Center for Biosciences at Karolinska Institutet, Cancerfonden, the EU FP7 project SYSCOL, and the Academy of Finland Center of Excellence in Cancer Genetics.

Received March 15, 2013; revised April 19, 2013; accepted April 29, 2013; published OnlineFirst July 5, 2013.

**References**

Lessons from Functional Analysis of Genome-Wide Association Studies


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-13-0789

Cited articles
This article cites 39 articles, 13 of which you can access for free at:
http://cancerres.aacrjournals.org/content/73/14/4180.full.html#ref-list-1

Citing articles
This article has been cited by 13 HighWire-hosted articles. Access the articles at:
/content/73/14/4180.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.