Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer

Huiling He^{a,1}, Wei Li^a, Sandya Liyanarachchi^a, Mukund Srinivas^a, Yanqiang Wang^a, Keiko Akagi^a, Yao Wang^b, Dayong Wu^a, Qianben Wang^a, Victor Jin^b, David E. Symer^{a,c,d}, Rulong Shen^e, John Phay^f, Rebecca Nagy^{a,d}, and Albert de la Chapelle^{a,1}

^aHuman Cancer Genetics Program and Department of Molecular Virology, Immunology, and Medical Genetics, ^cDepartment of Biomedical Informatics, ^dDepartment of Internal Medicine, ^eDepartment of Pathology, and ^fDepartment of Surgery, The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210; and ^bDepartment of Molecular Medicine, University of Texas Health Science Center, San Antonio, TX 78249

Contributed by Albert de la Chapelle, April 1, 2015 (sent for review February 4, 2015; reviewed by Federico Canzian and Yuri E. Nikiforov)

The [A] allele of SNP rs965513 in 9q22 has been consistently shown to be highly associated with increased papillary thyroid cancer (PTC) risk with an odds ratio of ~1.8 as determined by genome-wide association studies, yet the molecular mechanisms remain poorly understood. Previously, we noted that the expression of two genes in the region, forkhead box E1 (*FOXE1*) and PTC susceptibility candidate 2 (*PTCSC2*), is regulated by rs965513 in unaffected thyroid tissue, but the underlying mechanisms were not elucidated. Here, we fine-mapped the 9q22 region in PTC and controls and detected an ~33-kb linkage disequilibrium block (containing the lead SNP rs965513) that significantly associates with PTC risk. Chromatin characteristics and regulatory element signatures in this block disclosed at least three regulatory elements functioning as enhancers. These enhancers harbor at least four SNPs (rs7864322, rs12352658, rs7847449, and rs10759944) that serve as functional variants. The variant genotypes are associated with differential enhancer activities and/or transcription factor binding activities. Using the chromosome conformation capture methodology, long-range looping interactions of these elements with the promoter region shared by *FOXE1* and *PTCSC2* in a human papillary thyroid carcinoma cell line (KTC-1) and unaffected thyroid tissue were found. Our results suggest that multiple variants coinherit with the lead SNP and located in long-range enhancers are involved in the transcriptional regulation of *FOXE1* and *PTCSC2* expression. These results explain the mechanism by which the risk allele of rs965513 predisposes to thyroid cancer.

thyroid cancer | genetic susceptibility | long-range enhancer | functional variants | SNP rs965513

Papillary thyroid carcinoma (PTC) is the most common form of thyroid cancer, accounting for >80% of all thyroid malignancy. It is estimated that 62,450 individuals in the United States will be diagnosed with thyroid cancer in 2015 (www.cancer.org/Research/CancerFactsFigures/index). Although PTC is clearly influenced by both genetic and environmental factors, genetic predisposition plays a major role as evidenced by case-control studies (1–3).

Genome-wide association studies (GWASs) have linked SNP rs965513 in 9q22 to thyroid cancer, mainly PTC. The odds ratio (OR) for this SNP is as high as ~1.8 (4). The association between rs965513 and thyroid cancer risk has been independently confirmed in different populations (5–11). The association was also observed in familial PTC and in patients with radiation-induced PTC (12–15). SNP rs965513 resides ~60 kb upstream of forkhead box E1 (*FOXE1*) (also known as thyroid transcription factor 2), a critical transcription factor (TF) in thyroid development, differentiation, and function (16–19). It has been repeatedly proposed that *FOXE1* is involved in the tumorigenesis of PTC (11, 19, 20). A DNA variant (rs1867277) in the promoter region of *FOXE1* was reported as a functional variant involved in transcriptional

regulation of *FOXE1* (19). However, these studies cannot explain the genetic impact of the GWAS rs965513 signal. We recently described a thyroid-specific long intergenic noncoding RNA gene, PTC susceptibility candidate 2 (*PTCSC2*), in the 9q22 locus. *PTCSC2* is located on the opposite strand of *FOXE1* in the human genome; exon 1 and intron 1 of transcript isoform c of *PTCSC2* overlap with the promoter region of *FOXE1*, implying a shared promoter or transcriptional regulatory machinery (21). There are unspliced and spliced transcripts of *PTCSC2* in thyroid tissue; the spliced version of *PTCSC2* has 11 exons and several distinct splicing isoforms. SNP rs965513 resides in the unspliced transcript (or in an intron of the spliced transcript) of *PTCSC2* (21). Moreover, the risk [AA] genotype of rs965513 is significantly associated with decreased expression of *FOXE1*, *PTCSC2*, and *TSHR* in unaffected thyroid tissue of patients with PTC (21). Despite these findings suggesting an involvement of *FOXE1* and *PTCSC2* in PTC susceptibility and tumorigenesis, the molecular mechanism underlying the effect of SNP rs965513 remains poorly understood (22).

Cellular gene expression is critically determined by DNA regulatory elements and sequence-specific TFs, as well as chromatin modifications. We hypothesized that transcriptional regulatory

Significance

Papillary thyroid carcinoma (PTC) displays a strong hereditary component that is, in part, due to the additive effects of numerous low-penetrance genes or variants, but virtually no mechanistic information is available. Here, we studied a well-known low-penetrance variant (SNP rs965513) located in a region devoid of coding genes. We show that at least four variants located in the immediate vicinity of rs965513 reside in enhancer elements that bind to the promoter region shared by two adjacent thyroid-related genes, forkhead box E1 (*FOXE1*) and PTC susceptibility candidate 2 (*PTCSC2*), regulating their expression. The role of intergenic regulatory variants in cancer predisposition and carcinogenesis is growing. Further mechanistic understanding of how these variants work such as described here needs to be acquired.

Author contributions: H.H. and A.d.l.C. designed research; W.L., S.L., M.S., and Yao Wang performed research; Yanqiang Wang, D.W., R.S., J.P., and R.N. contributed new reagents/analytic tools; H.H., W.L., S.L., and K.A. analyzed data; H.H. and A.d.l.C. wrote the paper; D.W. and Q.W. supervised chromosome conformation capture assay; V.J. and D.E.S. supervised bioinformatics analysis; and R.N. provided genetic counseling.

Reviewers: F.C., German Cancer Research Center; and Y.E.N., University of Pittsburgh.

The authors declare no conflict of interest.

¹To whom correspondence may be addressed. Email: huiling.he@osumc.edu or albert.delachapelle@osumc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1506255112/-DCSupplemental.

elements may exist in the genomic region surrounding rs965513 in 9q22, which might regulate *FOXE1* and *PTCSC2* expression. We therefore sought to fine-map the 9q22 locus and search for functional variants that might contribute to the genetic predisposition in PTC. Our results allowed us to conclude tentatively that the 9q22 locus predisposes to PTC by multiple mechanisms involving enhancers, TF binding sites, and transcriptional alterations of at least two genes. These types of findings may explain why the mechanisms of action by GWAS-detected SNPs have been hard to elucidate so far (23, 24).

Results

Resequencing, SNP Genotyping and Imputation, and Haplotype Analyses.

To assess SNPs across the 9q22 locus and to identify functional SNPs, as well as additional associated variants that may contribute to PTC predisposition, we performed targeted next-generation sequencing in a gene-poor region of about 167 kb [genomic coordinates chromosome 9: 100455000–100622000 (GRCh37/hg19)] on DNA from 22 patients with PTC, including 10 [AA], 6 [AG], and 6 [GG] genotypes of the lead SNP rs965513. The region contains both coding genes (*XPA* and *FOXE1*) that flank the lead SNP. All of the variants with frequencies ≥ 0.001 observed in the patients can be found in the 1000 Genomes Project data in the European (EUR) population. Linkage disequilibrium (LD) analyses revealed an LD block of about 33 kb around the lead SNP in patients with PTC, as well as in EUR individuals in the 1000 Genomes Project (Fig. 1). We measured the strength of LD in this block. Twenty-eight SNPs showed $D' = 1.0$ and $r^2 > 0.97$, implying almost complete LD with rs965513. An additional 55 SNPs showed $D' > 0.99$, but the r^2 values were between 0.011 and 0.741, indicating that these SNPs are coinherited with rs965513 despite having different allele frequencies (Table S1).

To phase this smaller region further and find potential SNPs with independent association signals that could influence susceptibility to PTC, we genotyped four SNPs (including the lead SNP rs965513) in an Ohio cohort of 1,146 cases and 1,328 controls and performed association analyses (Table S2). Each of these SNPs showed significant association with PTC; however, the association became nonsignificant after conditioned analyses with the lead SNP rs965513. We imputed all cases and controls for variants with a minor allele frequency (MAF) of ≥ 0.001 in

the 1000 Genomes Project data and performed haplotype analyses. The estimated haplotypes with frequencies ≥ 0.005 in cases or controls are listed in Table S3. Notably, there were only three haplotypes containing the risk [A] allele of rs965513 in the list. The most common risk haplotype, Hap1, showed an allelic OR of 1.79. The risk haplotypes Hap2 and Hap3 showed ORs of 1.24 and 1.54, respectively. The relative risk of Hap1 was further assessed by diplotype analyses and by comparing the risk with individuals without Hap1. Heterozygous Hap1 carriers showed an OR of 1.695, whereas homozygous Hap1 carriers showed an OR of 3.423 (Table S4).

The above analyses suggested that we had narrowed the disease-associated region to an ~ 33 -kb interval that includes the lead SNP rs965513. The functional risk SNP(s) are likely to reside in this block and contribute to the risk haplotypes. These results corroborate our previous findings that demonstrated a clear-cut association between the risk genotype [AA] of the lead SNP and down-regulation of *FOXE1* and *PTCSC2* expression in unaffected thyroid tissue (21).

Enrichment of Enhancer Histone Markers in 9q22 in Thyroid Cancer Cell Line KTC-1 and in Unaffected Thyroid Tissue.

SNP rs965513 is located ~ 60 kb upstream of *FOXE1*. It is also in the unspliced transcript of *PTCSC2* and in an intron of the spliced *PTCSC2*. Previously, we have shown that the [AA] genotype of rs965513 is significantly associated with decreased expression levels of *FOXE1* and the unspliced *PTCSC2* transcript (21). We hypothesized that the functional SNP(s) might modulate critical *cis*-regulatory elements. For the lead SNP rs965513, no evidence supporting a functional role was found in publicly available databases. Therefore, we examined SNPs extensively within the 33-kb block. To establish whether these SNPs are functional variants affecting the binding sites of known TFs, we screened the variant nucleotides and adjacent DNA sequence using the transcription factor (TRANSFAC) database (25), Encyclopedia of DNA Elements (ENCODE) ChIP sequencing data, and HaploReg annotated information (SI Materials and Methods). Multiple SNPs showing potential functional effect were found in one or more of the tested cell lines, but none of these SNPs were derived from thyroid cancer. We noticed that several regions contain clusters of binding sites for multiple TFs (Fig. S1), a characteristic of regulatory modules. We selected three regions with candidate SNPs located in them and designated them as enhancer 1, enhancer 2, and enhancer 3 because of their putative roles as long-range enhancers. Notably, enhancer 1 (~ 1.2 kb) is highly conserved among mammals and is predicted to be an enhancer element by computer analysis of Vista HMR-Conserved Noncoding Human Enhancers from Lawrence Berkeley National Laboratory (enhancer element_1595). We focused on the above three putative enhancers for further analyses. Because histone markers are considered to be active enhancer markers, we performed site-directed ChIP assays for H3K4me1 and H3K27Ac in a human papillary thyroid carcinoma cell line (KTC-1) and an unaffected thyroid tissue sample. We tested four fragments corresponding to four SNPs in the regions of enhancers 1–3 within the risk block (Fig. 1). We observed significant enrichment of H3K4me1 and H3K27Ac, suggesting that these regions function as active enhancers in KTC-1 cells (Fig. 24) and thyroid tissue (Fig. 2B). These observations provided independent evidence for the likelihood that multiple *in vivo* enhancers exist in 9q22 in thyroid tissue.

Allele-Specific Enhancer Activities in Three Enhancer Elements. To test the above enhancer elements further, we performed luciferase assays. We hypothesized that the SNPs in the above three enhancers may be functional SNP(s) leading to a detectable allele-specific functional effect on transcription. To test this hypothesis, we amplified DNA fragments representing risk and WT

Fig. 1. Targeted next-generation sequencing and SNP LD in 9q22. A diagram of the sequenced region shows lead SNP rs965513, the flanking coding genes (*XPA* and *FOXE1*), and the long intergenic noncoding RNA gene *PTCSC2-c* (11 exons). Haplotype blocks of the intron 5 region of *PTCSC2-c* [chromosome 9: 100486812–100566532 (hg19)] were created using HaploView 4.1 and defined by confidence interval (41). The PTC risk haplotype (~ 33 kb), the three regions designated as enhancers 1–3, and the corresponding four SNPs (black bullets) in the haplotype are shown (not to scale).

Fig. 2. ChIP assays of enhancer histone markers H3K4me1 and H3K27Ac in the regions of enhancers 1–3 as shown in Fig. 1. KTC-1 cells (A) and thyroid tissue (B) are shown. The relative abundance of H3K4me1 and H3K27Ac was normalized to the input control. H3K4me1 and H3K27Ac showed significant binding in all of the tested regions, with $P < 0.005$ (Student t test).

alleles of each of the three enhancers. A 1.1-kb fragment surrounding rs7864322 (and rs7850258 due to proximity, $r^2 = 0.99$ with rs7864322), representing enhancer 1; a 160-bp fragment surrounding rs12352658, representing enhancer 2a; a 162-bp fragment surrounding rs7847449, representing enhancer 2b; and a 133-bp fragment surrounding rs10759944, representing enhancer 3 were tested. These DNA fragments carrying the risk [A] and WT [G] alleles were cloned into luciferase reporter constructs driven by a minimal promoter vector. Enhancer activities of the four fragments were determined by transient transfection and luciferase assays in the KTC-1 cell line (Fig. 3). We observed significant constitutive enhancer activities in enhancers 2a, 2b, and 3, but not in enhancer 1. Moreover, we observed that the [G] allele of rs12352658 and the [C] allele of rs7847449 in enhancer 2 and the [A] allele of rs10759944 in enhancer 3 showed significantly increased luciferase activity compared with the corresponding WT alleles (Fig. 3A); these alleles represent the risk allele in haplotype 1 (Table S3). We further examined enhancer 1 by performing a luciferase assay and cotransfecting with plasmids encoding CEBPa or CEBPb (Fig. 3B). Enhancer 1 showed detectable enhancer activities in the presence of CEBPa and CEBPb. Moreover, the risk [C] allele of rs7864322 was associated with a significant reduction of the luciferase activity in the presence of either CEBPa or CEBPb (Fig. 3B). We observed similar patterns of luciferase enhancer activities for enhancer 1 in the HeLa cell line (Fig. S2). These results pinpointed at least three enhancer elements in 9q22 where the risk nucleotides of the tested SNPs significantly altered the expression of firefly luciferase. Taken together, these data suggested that the region surrounding rs965513 harbors multiple enhancer elements functioning as allele-specific long-range enhancers.

Site-Directed ChIP Assay for TFs. To test the hypothesis further that multiple SNPs in the enhancer elements are functional, we performed ChIP assays to validate the binding of the predicted TFs to the enhancers and to test the differential occupancy between WT and risk alleles. The initial ChIP assays were performed in KTC-1 cells for a panel of TFs, including CEBPa, CEBPb, STAT1, YY1, cJUN, ER, and TFAP2A. We detected significant

binding of CEBPa, CEBPb, and TFAP2A; the remaining TFs did not show detectable binding activity. In thyroid tissue, we found that CEBPb showed significant binding to enhancer 1 but not to enhancers 2 (2a and 2b) and 3. TFAP2A showed significant binding to enhancers 2a and 3 (Fig. 4A). To evaluate whether these two TFs display differential occupancy between WT and risk alleles, we performed ChIP assays followed by a primer extension-based method for analyzing SNPs (SNaPshot assay) in three samples of unaffected thyroid tissue. These samples were heterozygous for SNPs rs7864322, rs12352658, and rs10759944. In enhancer 1, CEBPb showed stronger binding to the WT alleles than to the risk alleles; TFAP2A showed stronger binding to the WT allele than to the risk allele in enhancers 2 and 3 (Fig. 4B). These allelic differences of TF occupancy for SNP rs7864322 and rs10759944 were also observed in the KTC-1 cell line (Fig. S3).

Collectively, the results of the ChIP experiments and luciferase reporter assays are consistent with the hypothesis that the 9q22 locus harbors multiple intergenic long-range enhancers and multiple SNPs (but apparently not the lead SNP) appear to be functioning by affecting the intrinsic activity of the enhancers in thyroid cells.

Physical Interaction Between the Enhancers and the Promoter Region of *FOXE1* and *PTCSC2*. Previously, we have shown that SNP rs965513 is significantly associated with the expression of *FOXE1* and *PTCSC2* (21). We hypothesized that the above-observed 9q22 enhancer elements might regulate *FOXE1* and *PTCSC2* expression by enhancing their transcription through direct interaction. In an effort to demonstrate this physical interaction, we used the

Fig. 3. Multiple enhancer elements and differential allelic enhancer activities in 9q22. Luciferase assays show enhancer activity in the three regions of enhancers 1–3 in KTC-1 cells. (A) Constitutive enhancer activities in enhancers 2 and 3. (B) Enhancer activity of enhancer 1 in the presence of CEBPa and CEBPb. * $P < 0.05$; ** $P < 0.005$ (Student t test).

Fig. 4. Differential allelic TF binding activities in fresh unaffected thyroid tissue. (A) ChIP assays showing CEBPb and TFAP2A binding in the regions of

chromosome conformation capture (3C) assay and utilized the KTC-1 cell line for the initial analyses (26). We evaluated the interactions between fragments across the ~33-kb LD block and the upstream regions of *FOXE1* and *PTCSC2*. PCR primers were designed near the end of HindIII fragments in the same orientation with respect to successive restriction fragments. Enhancers 1–3 reside in 3C fragments 4–6, respectively (Fig. 5A and Table S5). When a constant primer and a probe in fragment 6 were used in the 3C assay, we observed a strong interaction with fragment 19, the genomic region immediately upstream of *FOXE1* and also comprising part of intron 1 of transcript isoform c of *PTCSC2*. The linear distance between fragment 6 and fragment 19 is ~51 kb, indicating that a long-range loop is formed (Fig. 5B). When a constant primer and a probe in fragment 19 were used in the 3C assay, we observed strong physical interactions with fragments 4–6, with fragment 6 showing the strongest interaction (Fig. 5C).

To explore the physical interactions further between 9q22 enhancer elements and the remote promoter region of *FOXE1/PTCSC2*, we performed 3C assays using an unaffected thyroid tissue sample. We noticed that the interaction between fragment 6 and fragment 19 was the strongest. However, interactions between fragment 4 or 5 and fragment 19 were quite low or undetectable (Fig. 5D). These results demonstrated that the 9q22 enhancers could form looping interactions with the promoter region of *FOXE1* and *PTCSC2*, and therefore cooperate in regulating the expression of these two genes.

Discussion

GWA studies have persistently linked the noncoding SNP rs965513 in 9q22 to thyroid cancer risk. It is worth noting that among GWAS-derived variants, rs965513 shows the strongest association with sporadic, familial, and radiation-induced PTC (4, 12, 15). However, advancing from the genetic association to an understanding of the functional biological process has been challenging (23). We are just beginning to understand the functional basis underlying the rs965513 risk in PTC. Similar scenarios are known from GWAS studies of several cancers and other phenotypes (27).

In this study, we fine-mapped the 9q22 genomic region surrounding rs965513 by performing targeted next-generation resequencing, genotyping/imputation, and haplotype analyses. After compiling these data, we identified a genomic block of ~33 kb containing rs965513. The most common and most significant risk haplotype, Hap1, had an allelic OR of ~1.79. As many as 83 SNPs in this block are co-inherited with rs965513; any of these SNPs could be functional.

A few genome-scale chromatin mapping studies have highlighted the enrichment of GWAS SNPs in regulatory DNA elements, suggesting that many functional variants may affect gene regulation (28–31). After performing computer analyses and studying the ENCODE data for the 9q22 locus, we selected three regions and four SNPs for experimental functional analyses in a thyroid cancer cell line and in unaffected thyroid tissue. We demonstrated three enhancer elements by ChIP assays for histone markers and TFs, as well as luciferase enhancer reporter assays. We showed that all three of these enhancer elements can form close looping interactions with the shared promoter region of *FOXE1* and *PTCSC2* in KTC-1 cells, whereas in unaffected thyroid tissue, only one of the enhancer elements (enhancer 3) showed detectable looping interaction. Our data suggest that long-range interactions underlying the mechanisms of transcriptional regulation of *FOXE1* and *PTCSC2* exist in the thyroid.

enhancers 1–3. (B) Relative allelic TF binding activities in enhancers 1–3 were determined using SNaPshot assay for SNPs rs7864322, rs12352658, and rs1075994, respectively. Differential allelic binding activities were calculated and normalized to “Input.” ** $P < 0.005$ (Student t test).

Fig. 5. Enhancers forming close contact with the shared promoter region of *FOXE1* and *PTCSC2* in a 3C assay. (A) 3C HindIII fragments in 9q22. (B) 3C assay using a constant primer in fragment 6 in KTC-1 cells. (C) 3C assay using a constant primer in fragment 19 in KTC-1 cells. (D) 3C assay using a constant primer in fragment 19 in a fresh unaffected thyroid tissue sample.

Our experiments suggest that at least four variants in the enhancer elements could be functional in thyroid but that the lead SNP rs965513 is seemingly not involved. The impact of functional variants on transcriptional regulation through multiple enhancers has been documented in other cancers (32). An example is the well-studied 8q24 locus in which GWAS SNPs are associated with risk of colon and prostate cancers (33–36). Multiple enhancer elements were identified within the cancer-associated regions in 8q24; these enhancers regulate Myc promoter activity through a long-range looping mechanism (37). Recently, it was reported that for six common autoimmune disorders (rheumatoid arthritis, Crohn’s disease, celiac disease, multiple sclerosis, lupus, and ulcerative colitis), the association detected by GWAS was due to multiple polymorphisms in LD that map to clusters of enhancer elements active in the same cell type (27). The authors observed that multiple enhancer variants within a given locus typically target the same gene and that multiple enhancer variants cooperatively contribute to altered expression of their target gene (27). Our data are consistent with this concept and provide further evidence for a “multiple enhancer variants” hypothesis by which noncoding variants can confer susceptibility to common traits. The effects of the 9q22 enhancer variants we describe in this study are likely combinatorial in dictating the gene expression of *FOXE1* and *PTCSC2*, and in conferring susceptibility to PTC.

The importance of long-range interactions in 9q22 may be especially relevant in the genetic predisposition to PTC, where we recently showed that the [AA] risk genotype of rs965513 is significantly associated with lower expression of *FOXE1*, *PTCSC2* unspliced transcript, and *TSHR* in unaffected thyroid tissue (21). The lower abundance of products from these genes is apparently a hallmark of dedifferentiation, allowing malignant transformation. The observation of differential allelic enhancer activities and direct looping interactions between these enhancers and the promoter region of *FOXE1* and *PTCSC2* strongly argues that multiple enhancer elements regulate *FOXE1* and *PTCSC2*, although we

cannot distinguish the differential effect on these two genes, if there is any, in this study.

In conclusion, our study provides further insight into the functional annotation of rs965513 in 9q22 as a risk factor contributing to PTC. We demonstrate that multiple functional variants displaying total or high LD and located in multiple long-range enhancer elements are the primary genomic basis for the predisposition displayed by the lead SNP rs965513. Only relatively recently has it begun to be realized that regulatory variants, mainly noncoding RNAs and enhancers, account for the major part of the genetic predisposition to complex traits, such as cancer. In one extensive study, regulatory variants accounted for much more heritability than coding variants (38). Consequently, the mechanisms need to be determined. The present study of the 9q22 locus in thyroid carcinoma serves as an example of such studies.

Materials and Methods

The study was approved by the Institutional Review Board at the Ohio State University (OSU), and all subjects gave written informed consent before participation.

Cell Line Culture and Samples from Patients. The KTC-1 cell line was incubated in antibiotic-free RPMI 1640 medium supplemented with 10% (vol/vol) FBS at 37 °C in humidified air with 5% (vol/vol) CO₂. Blood samples (1,146 PTC cases and 1,328 controls) included in this study were collected at OSU as part of ongoing studies. Fresh (*n* = 2) or frozen (*n* = 1) unaffected thyroid tissue samples were obtained from patients with PTC during surgery, and were either used immediately or snap-frozen in liquid nitrogen and kept at −80 °C. Clinical information on these samples will be made available on request.

Targeted Next-Generation Sequencing, SNP Genotyping, and Computer Analyses. To capture all genetic variations/mutations in the 9q22 locus, we resequenced blood DNA from 22 patients who had PTC. Customized sequence capture was conducted by means of an Agilent SureSelect Target Enrichment kit (Agilent Technologies) to cover the entire region [chromosome 19: 60,080,000–67,090,000 (hg19)]. Briefly, the nonrepetitive genomic DNA sequences were enriched using the SureSelect Target Enrichment system. Custom paired-end libraries were

prepared using the captured DNA. Deep sequencing was performed using an Illumina HiSeq2000 platform. The procedures of SNP genotyping, SNP imputation, LD, and haplotype analyses are described in *SI Materials and Methods*. The PCR primer sequences for SNP genotyping are provided in [Table S5](#). Computer analyses were performed using the UCSC Genome Browser (genome.ucsc.edu; GRCh37/hg19), TRANSFAC database data, and ENCODE project data.

ChIP Assay. ChIP was performed as previously described (39), with minor modifications. Briefly, protein/DNA cross-linking was performed by incubating cells with formaldehyde at a final concentration of 1% for 10 min at room temperature. After sonication, chromatin was immunoprecipitated with specific antibodies at 4 °C overnight. The antibodies against H3K4me1 (ab8895) and H4K27ace (ab4729) were from Abcam. Antibodies against CEBPa (sc-61), CEBPb (sc-150x), and TFAP2A (sc-8975) were from Santa Cruz Biotechnology. The immune complexes were then eluted from the beads, and the cross-links were reversed by incubating at 65 °C overnight. The DNA was purified with a QIAquick PCR purification kit (QIAGEN) and used as the template in quantitative PCR assays. Primers were designed to yield amplicons ranging in length from 70 to 110 bp spanning the enhancer regions. The primers used in quantitative PCR reactions are provided in [Table S5](#).

Constructs, Transient Transfections, and Luciferase Reporter Gene Assays. For enhancer-reporter gene assays, 133-bp to 1.2-kb DNA sequences containing the WT or risk allele in the selected regions in 9q22 were PCR-amplified and cloned into the PGL4.10-E4TATA vector. This vector contains a 50-bp minimal E4 TATA promoter sequence. The constructs were validated by Sanger

sequencing. Cotransfection in KTC-1 cells was performed with empty vector or the enhancer constructs and/or expression constructs of TFs and a *Renilla* luciferase reporter plasmid. The empty PGL4.10-E4TATA vector DNA was added if necessary to make sure that the same amount of total DNA was used for transfection in all groups. Cells were harvested after 24 h, and lysates were used for luciferase assays. At least three independent experiments were performed.

3C Assay. The 3C assay was performed according to the protocols previously described (40). We used HindIII as the restriction enzyme and obtained digestion efficiencies ranging between 65% and 82%. PCR primers were designed near the ends of the HindIII fragments in the same orientation with respect to successive restriction fragments. The BAC clone RP11-746L3 standard was used for data normalization to correct for the efficiency of the PCR amplification process between various experimental samples. PCR reactions were designed using Primer3 software and the coordinates of human genome GRCh39/hg19 assembly. The primer and probe sequences are provided in [Table S5](#).

ACKNOWLEDGMENTS. The Biomedical Genomics Core of the Research Institute at Nationwide Children's Hospital (Columbus, OH) performed targeted deep sequencing. SNP genotyping was kindly performed by Dr. Julius Gudmundsson at deCODE, Inc. Tissue samples were provided by the Cooperative Human Tissue Network at OSU, which is funded by the National Cancer Institute. This work was supported, in part, by an allocation of computing time from the Ohio Supercomputer Center. This work was also supported, in part, by National Cancer Institute Grants P30CA16058, P50CA168505, and R01 CA151979.

- Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH (1994) Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 86(21):1600–1608.
- Dong C, Hemminki K (2001) Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families. *Int J Cancer* 92(1):144–150.
- Teerlink CC, Albright FS, Lins L, Cannon-Albright LA (2012) A comprehensive survey of cancer risks in extended families. *Genet Med* 14(1):107–114.
- Gudmundsson J, et al. (2009) Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet* 41(4):460–464.
- Matsue M, et al. (2011) The FOXE1 and NKX2-1 loci are associated with susceptibility to papillary thyroid carcinoma in the Japanese population. *J Med Genet* 48(9):645–648.
- Jones AM, et al.; TCUKIN Consortium (2012) Thyroid cancer susceptibility polymorphisms: Confirmation of loci on chromosomes 9q22 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24. *J Med Genet* 49(3):158–163.
- Wang Y-L, et al. (2013) Confirmation of papillary thyroid cancer susceptibility loci identified by genome-wide association studies of chromosomes 14q13, 9q22, 2q35 and 8p12 in a Chinese population. *J Med Genet* 50(10):689–695.
- Köhler A, et al. (2013) Genome-wide association study on differentiated thyroid cancer. *J Clin Endocrinol Metab* 98(10):E1674–E1681.
- Liyanarachchi S, et al. (2013) Cumulative risk impact of five genetic variants associated with papillary thyroid carcinoma. *Thyroid* 23(12):1532–1540.
- Penna-Martinez M, et al. (2014) FOXE1 association with differentiated thyroid cancer and its progression. *Thyroid* 24(5):845–851.
- Zhu H, Xi Q, Liu L, Wang J, Gu M (2014) Quantitative assessment of common genetic variants on FOXE1 and differentiated thyroid cancer risk. *PLoS ONE* 9(1):e87332.
- Tomaz RA, et al. (2012) FOXE1 polymorphisms are associated with familial and sporadic nonmedullary thyroid cancer susceptibility. *Clin Endocrinol (Oxf)* 77(6):926–933.
- Bonora E, et al.; NMTC Consortium (2014) The FOXE1 locus is a major genetic determinant for familial nonmedullary thyroid carcinoma. *Int J Cancer* 134(9):2098–2107.
- Takahashi M, et al. (2010) The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Hum Mol Genet* 19(12):2516–2523.
- Damiola F, et al. (2014) Contribution of ATM and FOXE1 (TTF2) to risk of papillary thyroid carcinoma in Belarusian children exposed to radiation. *Int J Cancer* 134(7):1659–1668.
- Damante G, Tell G, Di Lauro R (2001) A unique combination of transcription factors controls differentiation of thyroid cells. *Prog Nucleic Acid Res Mol Biol* 66:307–356.
- Trueba SS, et al. (2005) PAX8, TTF1, and FOXE1 gene expression patterns during human development: New insights into human thyroid development and thyroid dysgenesis-associated malformations. *J Clin Endocrinol Metab* 90(1):455–462.
- Cuesta I, Zaret KS, Santisteban P (2007) The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure. *Mol Cell Biol* 27(20):7302–7314.
- Landa I, et al. (2009) The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. *PLoS Genet* 5(9):e1000637.
- Pereira JS, et al. (2014) Identification of a novel germline FOXE1 variant in patients with familial non-medullary thyroid carcinoma (FNMTC). *Endocrine*, 10.1007/s12020-014-0470-0.
- He H, et al. (2015) Genetic predisposition to papillary thyroid carcinoma: Involvement of FOXE1, TSHR, and a novel lincRNA gene, PTCSC2. *J Clin Endocrinol Metab* 100(1):E164–E172.
- Landa I, Robledo M (2011) Association studies in thyroid cancer susceptibility: Are we on the right track? *J Mol Endocrinol* 47(1):R43–R58.
- Fugger L, McVean G, Bell JI (2012) Genome-wide association studies and common disease—realizing clinical utility. *N Engl J Med* 367(25):2370–2371.
- Monteiro ANA, Freedman ML (2013) Lessons from postgenome-wide association studies: Functional analysis of cancer predisposition loci. *J Intern Med* 274(5):414–424.
- Wingender E, et al. (2000) TRANSFAC: An integrated system for gene expression regulation. *Nucleic Acids Res* 28(1):316–319.
- Wu D, et al. (2014) Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer. *Nucleic Acids Res* 42(6):3607–3622.
- Corradin O, et al. (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res* 24(1):1–13.
- Ernst J, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345):43–49.
- Paul DS, et al.; Cardiogenics Consortium; MuTHER Consortium (2011) Maps of open chromatin guide the functional follow-up of genome-wide association signals: Application to hematological traits. *PLoS Genet* 7(6):e1002139.
- Maurano MT, et al. (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337(6099):1190–1195.
- van der Harst P, et al. (2012) Seventy-five genetic loci influencing the human red blood cell. *Nature* 492(7429):369–375.
- Wang QF, et al. (2014) Regulation of MEIS1 by distal enhancer elements in acute leukemia. *Leukemia* 28(1):138–146.
- Zanke BW, et al. (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 39(8):989–994.
- Tomlinson I, et al.; CORGI Consortium (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39(8):984–988.
- Witte JS (2007) Multiple prostate cancer risk variants on 8q24. *Nat Genet* 39(5):579–580.
- Xiang J-F, et al. (2014) Human colorectal cancer-specific CCAT1-L lincRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 24(5):513–531.
- Sotelo J, et al. (2010) Long-range enhancers on 8q24 regulate c-Myc. *Proc Natl Acad Sci USA* 107(7):3001–3005.
- Gusev A, et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium (2014) Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet* 95(5):535–552.
- Im H, et al. (2004) Measurement of protein-DNA interactions in vivo by chromatin immunoprecipitation. *Methods Mol Biol* 284:129–146.
- Hagège H, et al. (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2(7):1722–1733.
- Gabriel SB, et al. (2002) The structure of haplotype blocks in the human genome. *Science* 296(5576):2225–2229.