Supplementary data

The role of *NRG1* in the predisposition to papillary thyroid carcinoma

Huiling He[1], Wei Li[1], Sandya Liyanarachchi[1], Yanqiang Wang[1], Lianbo Yu[2] Luke K. Genutis[1], Sophia Maharry[1], John E. Phay[3], Rulong Shen[4], Pamela Brock[5], Albert de la Chapelle[1]

[1] Department of Cancer Biology and Genetics; [2]Department of Biomedical Informatics, [3] Department of Surgery, [4]Department of Pathology, [5]Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, 43210, USA

Supplemental Table 1. Clinicopathologic characteristics of the Ohio PTC cohort

Characteristics		NRG1 expression analysis	Association and haplotype analysis	
		Case (n=163)	Case (n=1359)	Control (n=1605)
Gender				
	Female	120	1030	1184
	Male	43	329	421
Age_at_diagnosis				
	Age < 45 yrs	84	783	754
	Age >= 45 yrs	79	576	851
Histo_subtype				Not applied
	Classic PTC	133	792	
	MicroPTC	9	244	
	Follicular variant	9	260	
	Oncocytic variant	4	26	
	Tall cell	3	25	
	Columnar cell	1	5	
	Diffuse sclerosing	0	3	
	Other variants*	4	4	

*Other variant (one case for each): oxyphilic cell, Hurthle cell, Warthin-like, or solid squamous cell in NRG1 expression analysis; solid squamous cell, clear cell, macrofollicular, and PTC cystic tumor in association and haplotype analysis.

Supplemental Table 2. sequences of primers and probes	
Realtime RT- PCR	
NM_004495	
NM_004495-For	TCCCATTAGAATATCAGTATCCACAG
NM_004495-Rev	CATAAGCGACACACAGGATTTC
NM_004495-probe	AAGCACTCCCCTCCATTCACACAG
NM_001160008	
NM_001160008-For	CACTATACTTCCACAGCCCATC
NM_001160008-Rev	GCTTTGCCCTTCGGTTTTAC
NM_001160008-probe	AGAAACCCCTGATTCCTACCGAGACT
NM_013958	
013958-For	TCCCATTAGAATATCAGTATCCACAG
013958-Rev	TGGTAAGACACATCTAGCTCTAGG
013958-probe	ATACTTGTGCAAGTGCCCAAATGAGTTT
Luciferase assay	
Region1-F	aatttactcgagAACTGCGCGTGAAACCTACT
Region1-Rev	ttaattagatctACCCCAGAAAATGGAAAACC
Region2-F	aatttactcgagGTCCTACCGGGCGCCTAC
Region2-Rev	ttaattagatctCCTATAGCCCAGGTCCCACT
Region3-F	aatttactcgagTCAGTCCTGTGCTACGGAGA
Region3-Rev	ttaattagatctCAGCAAACAATTGGGTTCAA
Region4-F	aatttactcgagCAATGTTCTGTTAGCATTCTCCA
Region4-Rev	ttaattagatctTGGCATTTGGTTTGTTTTCA
Region5-F	aatttactcgagCGCAAATGAATGAACTTCTGG
Region5-Rev	ttaattagatctAAAAGAAATCCGTTTTCAAGC
Region6-F	aatttactcgagTGGGATTTGTGTCTGTGCAT
Region6-Rev	ttaattagatctTGTTGGGATTACAGGCATGA
Region7-F	aatttactcgagTTGGGTGAATCTGTCATCTCA
Region7-Rev	ttaattagatctAAGTCAGAAAATAAATATGTCGCAAA
Region8-F	aatttactcgagTGTGGGCCAGTTAAACCTCT
Region8-Rev	ttaattagatctGGAGAGTTAGGTGGCAAAGC
Region9-F	aatttactcgagTACTACCACCCACCCCCACT
Region9-Rev	ttaattagatctCCCAGGTTTCACAATTGACC

Continued Supplemental Table 2.	
ChIP assay	
rs3802160-F	CACGGAAGGATCCAGAATTG
rs3802160-Rev	ACCCCAGAAAATGGAAAACC
rs4733128-F	GTCCTACCGGGCGCCTAC
rs4733128-Rev	CCTATAGCCCAGGTCCCACT
rs4733130-F	TCAGTCCTGTGCTACGGAGA
rs4733130-Rev	GTTTGGGGCAGAGCTGATT
rs7835688-F	CAATGTTCTGTTAGCATTCTCCA
rs7835688-Rev	TGGCATTTGGTTTGTTTTCA
rs17646763-F	CGCAAATGAATGAACTTCTGG
rs17646763-Rev	AAAAGAAATCCGTTTTCAAGC
rs7825175-F	TGGGATTTGTGTCTGTGCAT
rs7825175-Rev	TGTTGGGATTACAGGCATGA
rs2439303&4-F	TTTGGTGTGTATAATTTTATTTGATGA
rs2439303&4-Rev	GTCCTCACCCAAGGGTTTCT
rs2439302-F	CTGGAGTTCCCTGAAAGCAG
rs2439302-Rev	TGCAAGAATGGCCTAACACA
rs2466076-F	CAATAAGCAGAGCCGATGAA
rs2466076-Rev	CCCAGGTTTCACAATTGACC
rs2466075-F	CGGCTCTGCTTATTGCCTAA
rs2466075-Rev	CCCTATCTCCAGCCCACTAC
SNaPshot	
rs2439302-For	tgtgggccagttaaacctct
rs2439302-Rev	ggagagttaggtggcaaagc
rs2439302-EXT	caatgtgtaatctttgtttcata

Supplemental Table 3. The features of 3 NRG1 isoforms

Reference gene ID	UCSC Isoform ID	Transcript variant	Number of exons	Expression in TCGA[a]	Expression in OSU[b]
NM_013958.3	uc003xis.2	HRG-beta3	7	71.58	2.86
NM_001160008.1	uc011lbf.1	HRG-beta2b	11	17.44	0.86
NM_004495.3	uc003xir.2	HRG-gamma	6	1.52	2.82

[a]Expression levels (FPKM, median) in unaffected thyroid tissues. RNA-Seq data obtained from TCGA (n=59).

[b]RNA-Seq data obtained from our previous study (n=12).

Supplemental Table 4. Candidate functional DNA variants in the ~32 kb DNA block

SNP ID	Position[a] (hg19)	Effective Allele	Opposite Allele	Case EAF[b]	Control EAF	Odds ratio[c]	P value[c]	P value GTEx	RegulomeDB score	Functinal assays[d]
rs3802160	32404629	G	A	0.535	0.469	1.305	2.06E-07	2.76984E-30	2b	Yes
rs3802158	32404896	T	C	0.535	0.469	1.305	2.06E-07	2.77203E-30	4	No
rs36213229	32405517	T	G	0.535	0.469	1.305	2.06E-07	2.37537E-30	4	No
rs7820838	32405979	C	T	0.828	0.790	1.278	0.000273	6.18786E-05	4	No
rs7834206	32406148	A	C	0.535	0.469	1.305	2.06E-07	2.70273E-30	4	No
rs73234136	32406382	C	T	0.535	0.469	1.305	2.06E-07	6.09176E-30	4	No
rs113350646	32406549	A	G	0.252	0.217	1.218	0.000778	6.1598E-23	4	No
rs4733128	32406602	T	C	0.535	0.469	1.307	1.74E-07	2.77246E-30	2b	Yes
rs4733129	32406768	C	T	0.535	0.469	1.307	1.74E-07	2.77203E-30	4	No
rs4733130	32406994	C	T	0.535	0.469	1.303	2.29E-07	2.77203E-30	2b	Yes
rs12548687	32408820	G	A	0.535	0.469	1.307	1.74E-07	3.01175E-30	4	No
rs7835688	32411499	C	G	0.536	0.469	1.308	1.67E-07	2.77203E-30	1f	Yes
rs17646763	32411656	C	T	0.536	0.469	1.306	1.89E-07	2.77203E-30	3a	Yes
rs2439312	32412359	G	A	0.828	0.789	1.283	0.000229	4.21578E-05	1f	No
rs9642727	32414032	C	A	0.535	0.470	1.298	3.95E-07	6.52708E-30	4	No
rs17646936	32414074	G	A	0.266	0.308	0.816	0.000527	1.40304E-08	4	No
rs17719705	32414332	T	A	0.535	0.470	1.298	3.95E-07	8.07485E-30	4	No
rs7825175	32416274	A	G	0.251	0.216	1.215	0.00107	6.21728E-21	1d	Yes
rs13258892	32423537	T	C	0.280	0.326	0.803	0.000129	1.36594E-08	4	No
rs35233333	32429734	C	T	0.269	0.307	0.832	0.001858	5.06777E-08	4	No
rs2439304	32430371	G	A	0.459	0.516	0.794	1.11E-05	1.33582E-23	3a	Yes
rs2439303	32430375	C	T	0.470	0.534	0.775	1.08E-06	1.50765E-24	3a	Yes
rs2439302	32432369	C	G	0.468	0.534	0.767	5.47E-07	6.46589E-25	1f	Yes
rs2466077	32432753	T	G	0.465	0.531	0.769	6.86E-07	8.78949E-25	4	No
rs2466076	32432796	T	G	0.465	0.531	0.769	6.86E-07	3.43194E-25	4	Yes
rs2466075	32432949	G	A	0.437	0.496	0.791	9.74E-06	1.43793E-24	3a	Yes
rs71512640	32432957	A	G	0.161	0.193	0.800	0.002864	5.85483E-07	3a	No

[a]Position is based on UCSC genome hg19, browser hg19.

[b]EAF, effective allele frequency

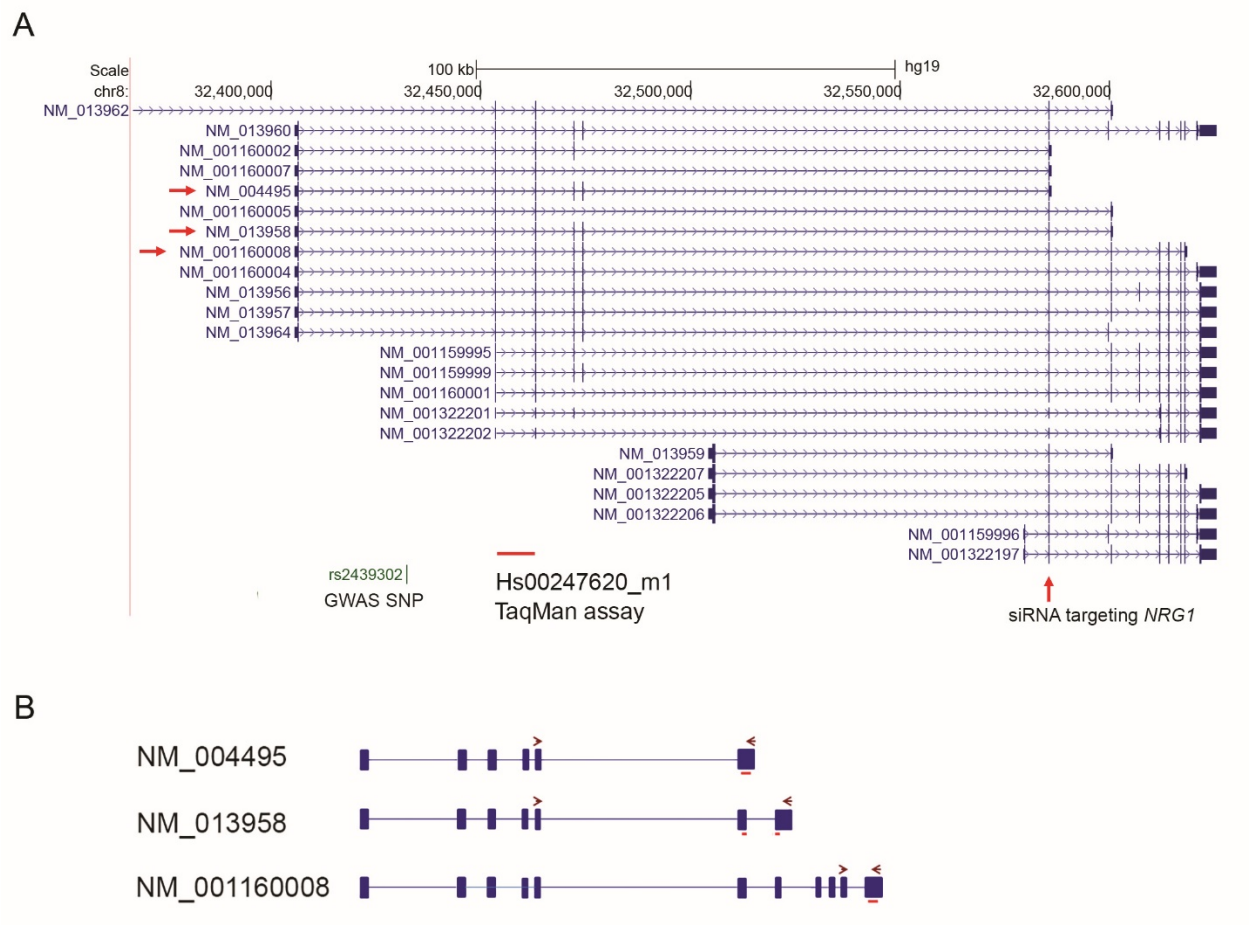
[c]Odds ratio and p values for association analysis between the SNP and PTC risk.

[d]Yes or No, ChIP and luciferase assays were performed or not.

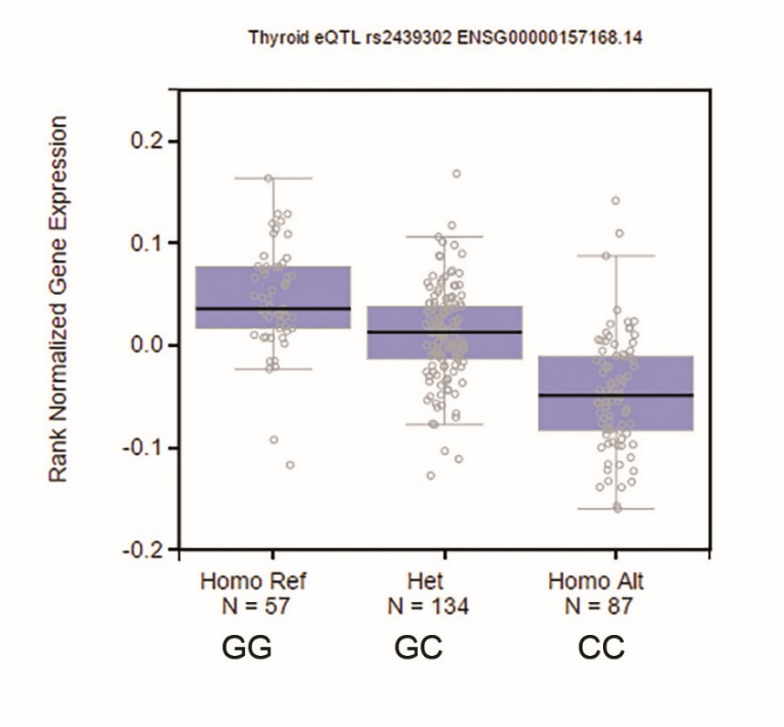
Supplemental Table 5. The top 10 diseases and biological functions	
Category	p-value
Neurological Disease	1.06E-06-8.94E-03
Psychological Disorders	1.74E-06-8.05E-03
Cancer	1.87E-06-8.94E-03
Organismal Injury and Abnormalities	1.87E-06-8.94E-03
Reproductive System Disease	1.87E-06-8.94E-03
Cellular Growth and Proliferation	2.63E-06-8.94E-03
Cell Death and Survival	2.97E-06-8.94E-03
Skeletal and Muscular Disorders	3.32E-06-8.94E-03
Cellular Compromise	8.13E-06-8.94E-03
Cellular Development	8.79E-06-8.94E-03
Ingenuity Pathway Analysis with the coding genes showing differential expression after siRNA knocking down NRG1	

Supplemental Table 6. The top molecular functions		
Name	p- value	#Molecules
Protein synthesis	8.79E-03 - 1.77E-07	39
Celluar growth and proliferation	8.84E-03 - 2.63E-06	83
Cell death and survival	8.84E-03 - 2.97E-06	74
Cellualr compromise	8.84E-03 - 8.13E-06	29
Cellular developlment	8.84E-03 - 8.79E-06	66
Ingenuity Pathway Analysis with the coding genes showing differential expression after siRNA knocking down NRG1		

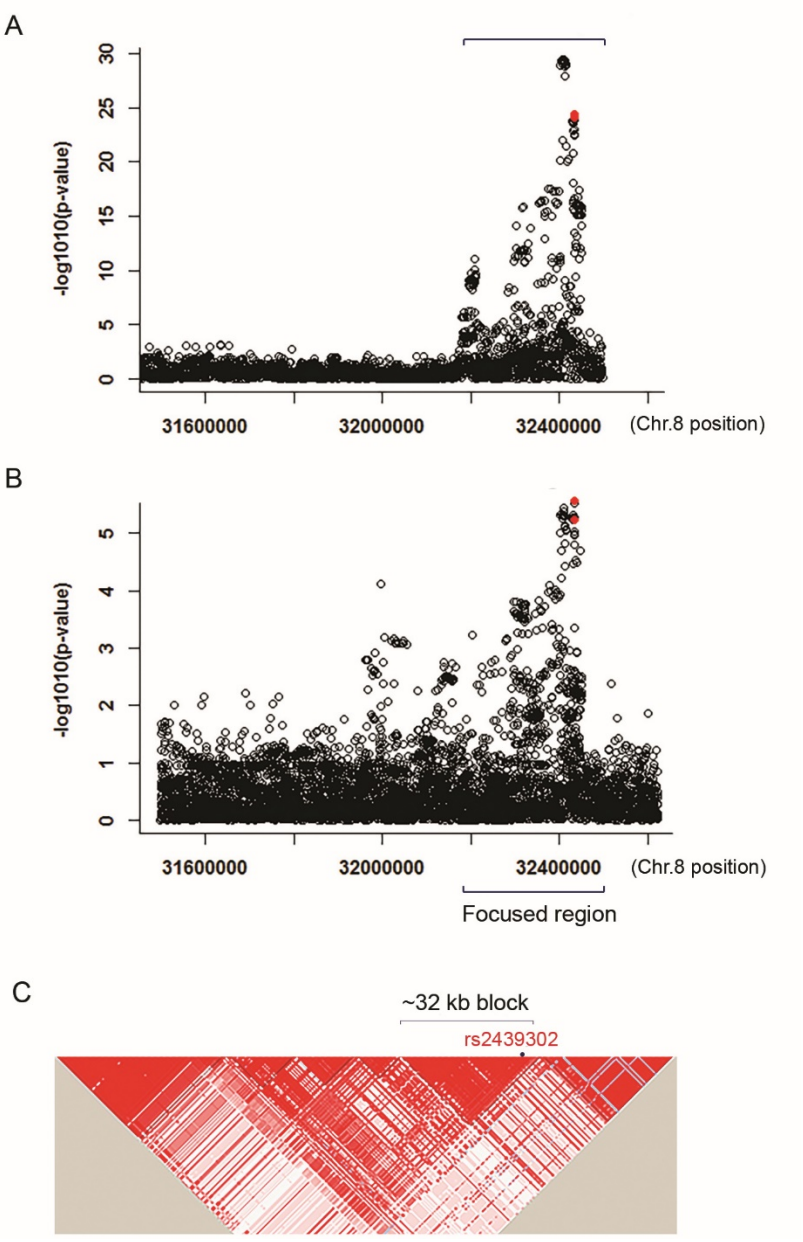
Supplemental Figures



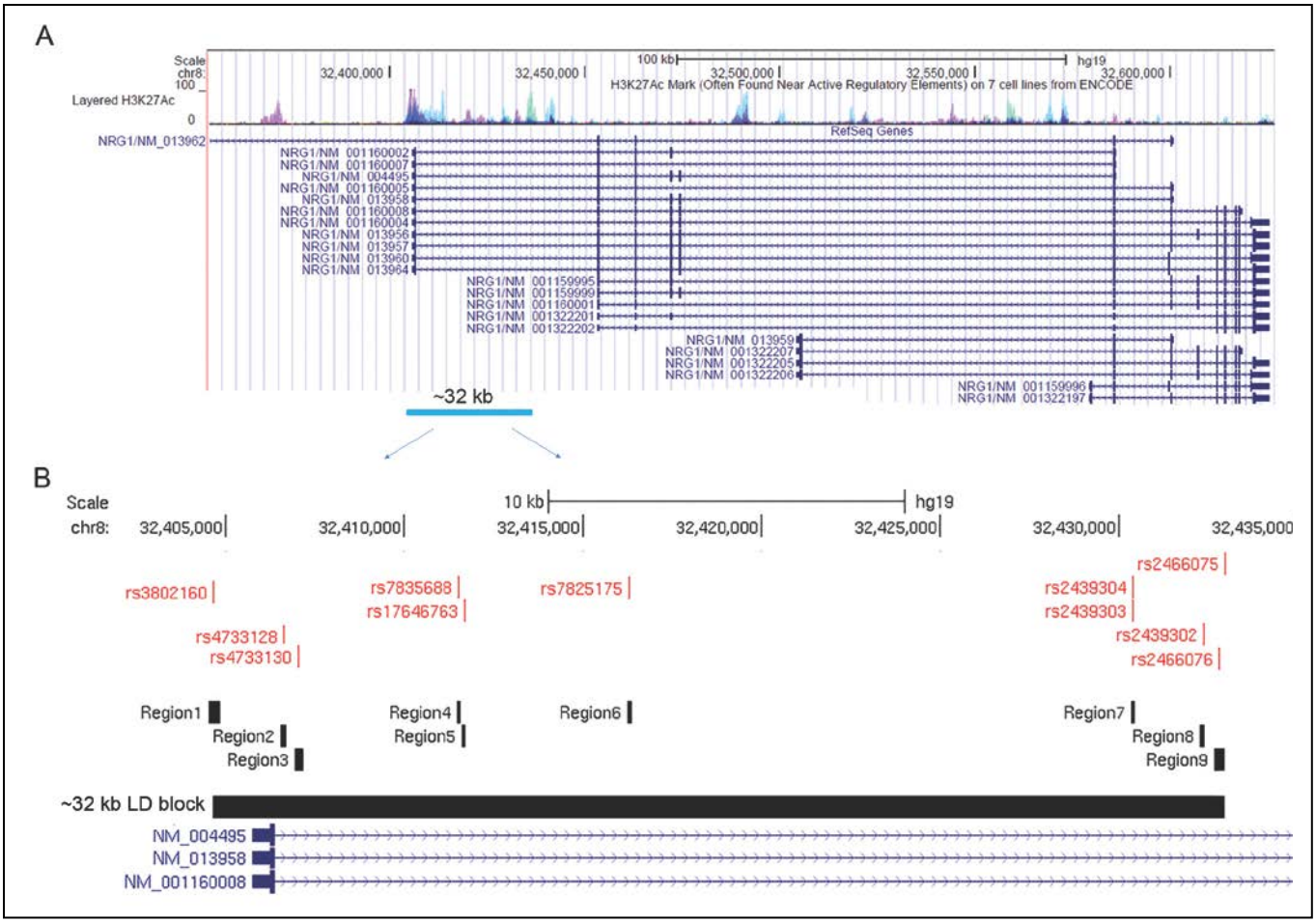
Supplemental Figure 1. The *NRG1* gene structure and multiple isoforms. (A) Three *NRG1* isoforms marked with red arrows were selected for gene expression analysis. The targeting locations of a predesigned TaqMan assay kit # Hs00247620_m1 (http://thermofisher.com) and predesigned siRNA oligos (http://dharmacon.gelifesciences.com/) are marked with a red line or an arrow. (B) Diagram of the three *NRG1* isoforms and the probes and primers used in quantitative RT-PCR assays. Vertical lines represent exons. The specific probes are indicated by red lines under one or two exons. The PCR primers are labeled by arrows.



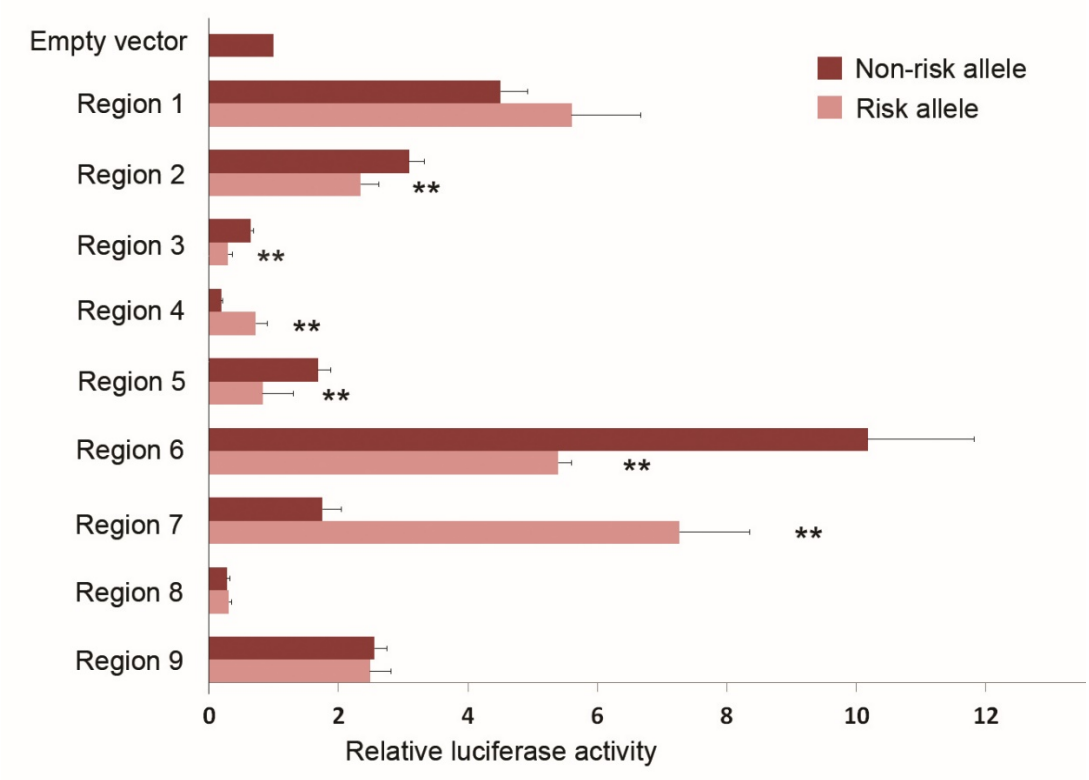
Supplemental Figure 2. Correlation between the *NRG1* expression and the genotypes of rs2439302 in thyroid tissue samples. The eQTL data were obtained from the GTEx consortium. (Date accessed July 6th, 2016; http://www.gtexportal.org/home/).



Supplementary Figure 3: Variants in the *NRG1* genomic region. (A) Plot of the association between the variants and the NRG1 expression; the p values data are obtained from GTEx; (B) Plot of the association between the variants and PTC risk; the p values data are obtained from an association in an Ohio cohort of 1359 cases/1605 controls. The open dots representing SNPs; the two red dots representing the two GWAS SNPs rs2439302 and rs2466076; (C) LD analysis in the 'Focused region" with HaploView (https://www.broadinstitute.org/haploview/haploview).



Supplemental Figure 4. The ~32 kb LD block in the *NRG1* locus. (A) Diagram of the layered H3K27Ac, the *NRG1* isoforms, and the ~32 kb LD block. All the information except the 32 kb block was obtained from UCSC genome browser (hg19) (https://genome.ucsc.edu/); (B) Selected functional candidate SNPs (n=11) and the cloned DNA fragments for the Luciferase assays are marked in the ~32 kb LD block.



Supplemental Figure 5. Luciferase assay in HeLa cells. DNA fragments containing the risk or wild type alleles in each region were cloned into a Luciferase enhancer reporter vector with a minimal promoter. The luciferase assay was performed in at least three experiments. The luciferase activities were normalized with the empty vector control. ** indicating the p values < 0.01.