Supporting Information

He et al. 10.1073/pnas.1506255112

SI Materials and Methods

Targeted Next-Generation Sequencing. To capture all genetic variations/mutations in the 9q22 locus, we resequenced blood DNA from 22 patients with PTC. Customized sequence capture was conducted by means of an Agilent SureSelect Target Enrichment kit to cover the entire region (chromosome 19: 60,080,000–67,090,000 (hg19)] as described (1, 2). Paired-end libraries were sequenced on an Illumina HiSeq2000 platform by the Biomedical Genomics Core of the Research Institute at Nationwide Children's Hospital (Columbus, OH). We sequenced 30–40 million reads (15–20 million pairs) for each sample, and 100-bp paired-end reads were aligned to the human genome (hg19) using Burrows-Wheeler alignment tool with default settings. Alignment data were stored in a binary alignment/map format and visualized using Integrative Genomics Viewer software (Broad Institute). Duplicate pairs were removed using SAMtools, and only reads with reliable alignment quality (mapping quality >30) were used to detect variants in this region. Variants were detected using the pileup function of Samtools/BCFtools.

In Silico Analysis. Computer analyses were performed using the UCSC Genome Browser (genome.ucsc.edu/cgi-bin/hgGateway; Assembly GRCh37/hg19, February 2009). Potential binding sites harboring the SNPs were searched for by TRANSFACT database scanning of the sequences in search of binding sites for known TFs (3). Score changes induced by the SNP were calculated. To explore the epigenetic profile of each SNP, the chromatin state segmentation profiles (ChromHMM) generated by the ENCODE project (genome.ucsc.edu) were studied. To assess the possible functional role of each SNP, we used the ENCODE-based tools HaploReg V2 and V3 (www.broadinstitute.org/mammals/haploreg).

SNP Genotyping, Imputation, and Haplotype Analyses. SNP genotyping was carried out using a standard protocol as described (4). The PCR primer sequences are provided in Table S5. The PCR assays were performed according to a standard protocol as follows: 2 min at 94 °C; followed by 30 cycles of 30 s at 94 °C, 30 s at 58 °C, and 30 s at 72 °C; followed by a final extension of 10 min at 72 °C. An ABI 3730 DNA Analyzer (Applied Biosystems) was used for the allele analysis. Genotyping of the four selected SNPs as reported in Table S2 was carried out using a single-track assay platform. All assays and data quality assessments were performed according to the manufacturer's instructions. Computation and visualization of LD in the intron 5 region was performed using Haploview V4.2 (5). To evaluate the SNPs in the locus more thoroughly, we studied the LD block (chromosome 9: 100532965–100566031) that contains the lead SNP rs965513 (Fig. 1). We imputed genotypes of the SNPs in this region for the Ohio cohort of 1,146 Caucasian PTC cases and 1,328 Caucasian controls, using a genotype imputation software package (MACH/MACH-Admix program) (5). Known genotypes of the Ohio cohort for the four SNPs in Table S2 and phased haplotypes from 503 EUR samples in 1000 Genomes Project phase3 data were used as references in the imputation. Preimputation quality filters of MAF > 0.01 and Hardy–Weinberg Equilibrium tests were applied for each SNP. Postimputation quality of estimated squared correlation between imputed and true genotypes (Rsq) > 0.5 was used to filter SNPs with any unreliable imputation calls, and we reran MACH/MACH-Admix to obtain phased haplotypes (Table S3).

Statistical Analyses. Association of each genotype or haplotype between cases and controls, adjusting for age and gender, was evaluated by applying logistic regression analysis. The comparison of luciferase activities between two groups was made by applying a *t* test (two tailed). A value of $P < 0.05$ was considered statistically significant. Data represent the mean ± SD.

1. He H, et al. (2011) Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* 332(6026):238–240.
2. He H, et al. (2013) Ultra-rare mutation in long-range enhancer predisposes to thyroid carcinoma with high penetrance. *PLoS ONE* 8(5):e61920.
3. Wingender E, et al. (2000) TRANSFAC: An integrated system for gene expression regulation. *Nucleic Acids Res* 28(1):316–319.
4. He H, et al. (2009) A susceptibility locus for papillary thyroid carcinoma on chromosome 8q24. *Cancer Res* 69(2):625–631.
5. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 44(8):955–959.

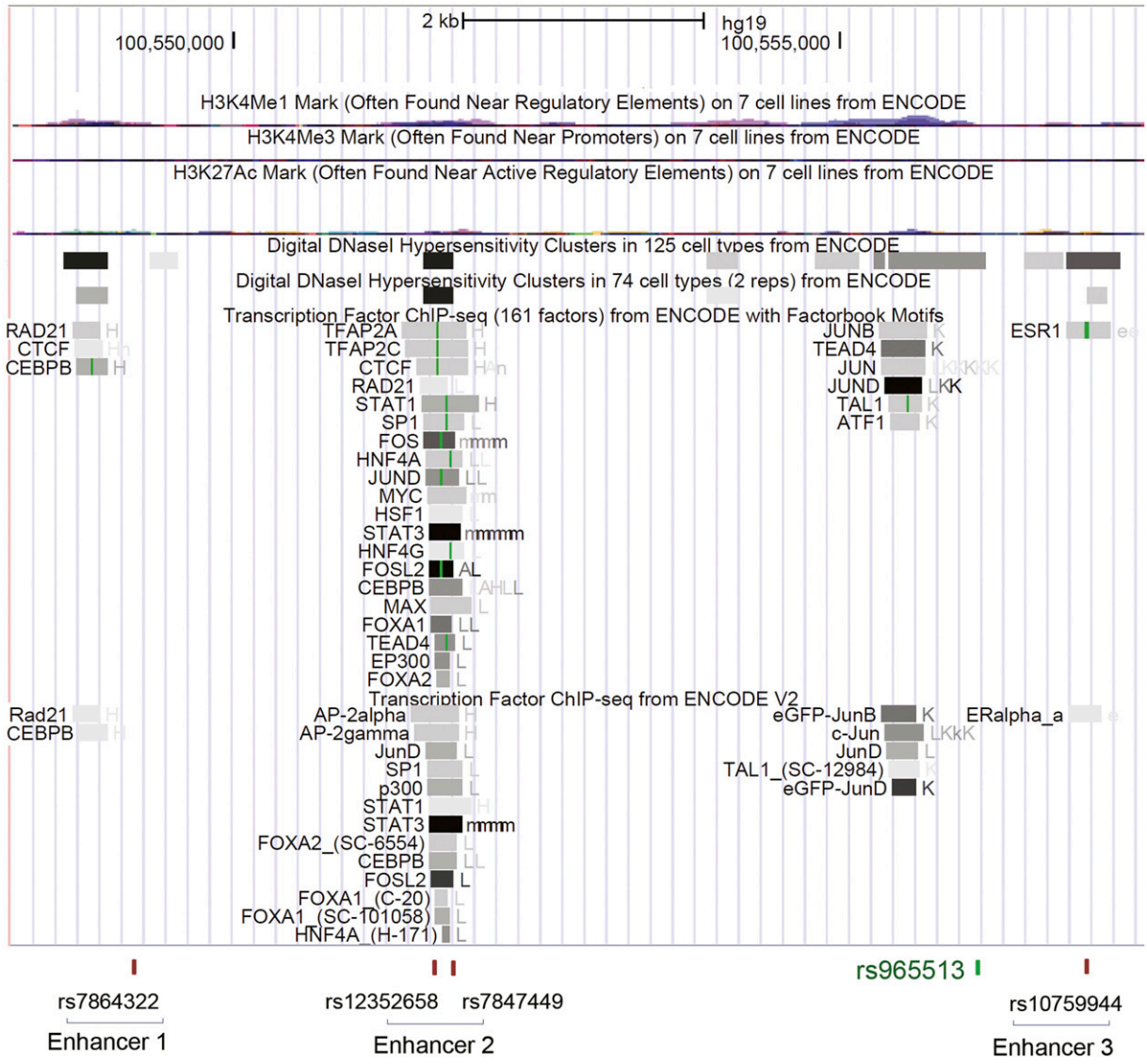

Fig. S1. ENCODE data of histone markers of regulatory elements and TF ChIP-sequencing in the rs965513 LD block in 9q22. The information was downloaded from the UCSC Genome Browser (genome.ucsc.edu; Assembly GRCh37/hg19). Three genomic regions (enhancers 1–3) and four SNPs tested with the KTC-1 cell line and thyroid tissues are indicated.

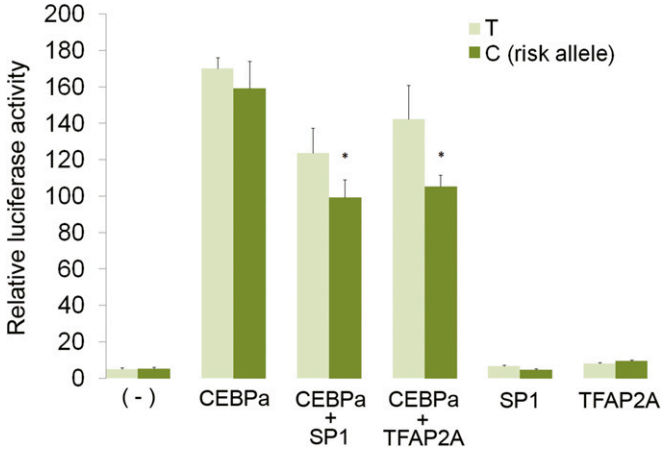

Fig. S2. Luciferase enhancer activities of enhancer 1 in HeLa cells. The enhancer activities of the two alleles of rs7864322 were tested and compared in the presence of TFs. *$P < 0.05$, Student t test.

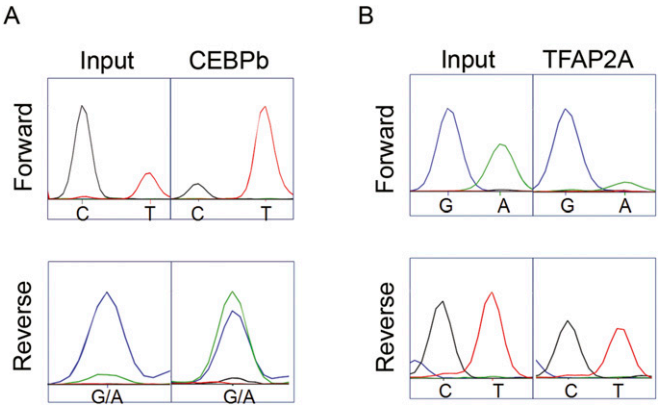

Fig. S3. Allelic TF occupancy in KTC-1 cells. Enhancer 1, rs7864322 (*A*) and enhancer 3, rs10759944 (*B*) are shown. ChIP assays were performed in KTC-1 cells, followed by a method for SNP analysis (SNaPshot assay). The terms "Forward" and "Reverse" indicate the genomic strand tested in the assay.

Table S1. SNPs in LD with rs965513 in an ~33-kb block in 9q22

SNP	Position in chromosome 9 (hg19)	D′	r^2	Major allele	Minor allele	MAF*	Comments	
rs1533180	100532965	1.000	0.368	G	T	0.421		
rs62573973	100533183	0.996	0.033	G	C	0.061		
rs5899324	100533317	0.995	0.722	A	ACT	0.411		
rs71501830	100533678	0.996	0.037	T	C	0.068		
rs12003414	100533996	0.996	0.037	T	C	0.068		
rs1877431	100534147	1.000	0.741	G	A	0.407		
rs7045465	100534823	1.000	0.741	T	A	0.407		
rs72751573	100535004	0.999	0.073	T	C	0.126		
rs7030256	100535203	1.000	0.991	G	C	0.339		
rs7030280	100535267	1.000	0.991	T	C	0.339		
rs16924115	100535317	0.998	0.039	C	T	0.072		
rs117964802	100535512	0.991	0.011	C	T	0.022		
rs62573974	100536542	1.000	0.290	C	T	0.364		
rs141224302	100536605	1.000	0.290	TGAAAAGTA	T	0.364		
rs10983700	100537455	1.000	0.991	C	T	0.339		
rs10983701	100537577	0.995	0.987	A	G	0.338		
rs71487344	100537665	0.998	0.057	AAAG	A	0.101		
rs1588636	100537763	0.999	0.075	C	T	0.128		
rs1588635	100537802	0.995	0.987	C	A	0.338		
rs7028661	100538470	0.995	0.987	G	A	0.338		
rs10818048	100538717	0.999	0.151	T	G	0.230		
rs10818049	100538922	0.998	0.057	C	T	0.101		
rs10818050	100538923	0.995	0.987	G	A	0.338		
rs140405800	100539449	0.995	0.046	C	T	0.023		
rs36052460	100539518	0.995	0.987	AT	A	0.338		
rs10983705	100539696	1.000	0.292	G	C	0.365		
rs12340862	100539830	0.998	0.057	T	C	0.101		
rs72751578	100540209	0.999	0.075	C	A	0.128		
rs7021576	100540541	1.000	0.991	T	C	0.339		
rs2401637	100541473	1.000	0.991	C	T	0.339		
rs12003512	100541776	0.996	0.037	G	A	0.068		
rs118174171	100541987	0.989	0.011	T	C	0.022		
rs10759927	100542176	1.000	0.991	G	A	0.339		
rs142727549	100542898	0.914	0.032	A	AAAAG	0.071		
rs1877432	100543880	1.000	0.290	G	A	0.364		
rs34388138	100544140	0.782	0.123	CT	C	0.283		
rs76230322	100544160	0.866	0.005	T	C	0.013		
rs7847663	100544868	1.000	0.991	T	C	0.339		
rs17429715	100544998	0.997	0.022	G	A	0.042		
rs4743130	100546040	1.000	0.991	T	C	0.339		
rs1561962	100546219	1.000	0.991	T	C	0.339		
rs925488	100546391	1.000	0.991	A	G	0.339		
rs35245612	100546486	1.000	0.290	TAA	T	0.364		
rs925489	100546600	1.000	0.991	T	C	0.339		
rs143350835	100547509	0.992	0.025	C	T	0.013		
rs146736169	100547600	0.991	0.011	A	G	0.022		
rs4273946	100547627	1.000	0.991	C	G	0.339		
rs72751583	100547884	0.996	0.038	G	A	0.019		
rs7020976	100547972	1.000	0.991	C	T	0.339		
rs7032019	100548144	1.000	0.991	A	G	0.339		
rs58187401	100548206	0.992	0.456	A	AGT	0.477		
rs7864322	100548934	1.000	0.995	T	C	0.338	Enhancer 1	3C fragment 4
rs7850258	100549013	1.000	0.995	G	A	0.338		
rs139121661	100550001	0.989	0.007	A	G	0.014		
rs1443438	100550028	1.000	0.995	C	T	0.338		
rs12347079	100550227	0.999	0.152	T	C	0.231		
rs10818071	100550253	0.998	0.058	G	A	0.102		
rs7030241	100550375	1.000	1.000	A	T	0.337		
rs7027030	100550455	1.000	1.000	C	A	0.337		
rs76293252	100550482	0.998	0.100	A	G	0.049		
rs118102144	100551260	0.991	0.011	A	T	0.022		
rs12352658	100551768	1.000	0.216	G	A	0.298	Enhancer 2a	3C fragment 5
rs7847449	100551908	1.000	0.242	C	A	0.323	Enhancer 2b	3C fragment 5

Table S1. Cont.

SNP	Position in chromosome 9 (hg19)	D′	r^2	Major allele	Minor allele	MAF*	Comments	
rs10739496	100552559	1.000	1.000	T	C	0.337		
rs10983761	100553957	1.000	1.000	C	A	0.337		
rs4743131	100554907	1.000	1.000	G	C	0.337		
rs34883424	100555363	0.996	0.037	G	A	0.068		
rs56202421	100555484	0.999	0.073	T	C	0.125		
rs965513	100556109	1.000	1.000	G	A	0.337	Lead SNP	
rs10759944	100556972	1.000	1.000	G	A	0.337	Enhancer 3	3C fragment 6
rs16924274	100557291	0.999	0.151	C	T	0.230		
rs141181809	100557521	0.986	0.006	T	C	0.012		
rs10983789	100558738	0.998	0.059	C	T	0.104		
rs147368517	100558780	0.998	0.059	C	T	0.104		
rs185843906	100558785	0.993	0.023	C	T	0.012		
rs568227990	100558818	0.996	0.021	G	T	0.041		
rs13295081	100559011	1.000	1.000	C	T	0.337		
rs76228779	100559045	1.000	0.387	G	A	0.432		
rs13290258	100559093	1.000	0.973	T	C	0.343		
rs13295254	100559114	1.000	0.991	C	T	0.339		
rs529832869	100559154	0.986	0.006	G	C	0.012		
rs139973818	100559360	0.996	0.037	G	C	0.068		
rs10983795	100559632	0.998	0.060	C	T	0.105		
rs572449192	100559923	0.024	0.000	TA	T	0.018		
rs10983796	100560368	1.000	0.292	T	C	0.365		
rs35342481	100561126	0.899	0.257	GA	G	0.385		
rs1867281	100561147	1.000	0.292	A	G	0.365		
rs73486628	100561352	0.994	0.022	G	A	0.042		
rs10818090	100561486	1.000	1.000	T	C	0.337		
rs187451397	100561600	0.987	0.005	T	C	0.011		
rs12341858	100562983	0.998	0.047	G	A	0.085		
rs79733232	100563136	0.992	0.025	T	C	0.013		
rs10818094	100563828	0.979	0.166	G	A	0.255		
rs112948091	100565441	0.481	0.089	A	AAAAG	0.164		
rs13289390	100565568	0.997	0.025	A	T	0.047		
rs77394667	100565813	0.479	0.088	G	A	0.163		
rs114804130	100566031	0.914	0.212	C	G	0.114		

*MAF was obtained from the 1000 Genomes Project (EUR = 503).

Table S2. Genotyping and association analyses of Four SNPs in 9q22 in an Ohio cohort

SNP	Major allele	Minor allele	MAF in Control	MAF in case	Risk allele	OR*	*P* value*	Risk allele frequency control	Risk allele frequency case	*P* value conditioned on rs965513
rs965513	G	A	0.341	0.487	A	1.830	1.40E-23	0.341	0.487	
rs16924274	C	T	0.224	0.170	C	1.410	2.97E-06	0.776	0.830	0.729
rs140405800	C	T	0.018	0.028	T	1.529	0.031	0.018	0.028	0.806
rs76293252	A	G	0.058	0.072	G	1.245	0.060	0.058	0.072	0.121

Ohio cohort: PTC (n = 1,146) and controls (n = 1,328).
*Age- and gender-adjusted.

Table S3. Haplotype analyses in cases/controls

Identification	Haplotype count		Frequency		OR*	*P* value*	Haplotypes†
	Case	Control	Case	Control			
Hap1	880	691	0.384	0.260	1.790	3.15E-20	GUAATCCCUTGCAATACAGCCTAGUCCCGUCGTGACATTGTAAGCCACTAACCCT-GCTCTACG
Hap2	165	154	0.072	0.058	1.243	0.061	GUAATCCCUTGCAATACAGCCTAGUCCCGUCGTGACATTGTAGGCCACTAACCCT-GCTCTACG
Hap3	62	46	0.027	0.017	1.541	0.029	GUAATCCCUTGCAATATAGCCTAGUCCCGUCGTGACATTGTAAGCCACTAACCCT-GCTCTACG
Hap4	751	1,068	0.328	0.402	0.701	2.09E-08	TAGTTGTTTCACCGTGCUCCTCGAUTTTATTCCAUTGCTGACAGATCGTGGCCCCA-TCCCGTG
Hap5	116	158	0.051	0.059	0.837	0.167	GAGTTGTCUCACCGGGCUGCTCGGCTTTAUTCCAUTGCCAACAACTCGTGGTTTC-GTCTTATA
Hap6	77	120	0.034	0.045	0.727	0.036	GAGTTGTCUCACCGGGCUGCTCGGCTTTAUTCCAUTGCCAACAACTCGTGGTTTC-GTCTTATA
Hap7	33	62	0.014	0.023	0.635	0.037	GAGTCGTCUCATCGGGCUGATCGGCTTTAUTCCAUTGCCGACAACTCGCGGTCCC-GTCCTATA
Hap8	17	40	0.007	0.015	0.493	0.016	GAGTTGTCUCACCGGGCUGCTCGGCTTTAUTCCAUTGCCGACAACTCGTGGTCCC-GTCCTATA
Hap9	17	35	0.007	0.013	0.578	0.067	GAGTTGTCUCACCGGGCUGCTCGGCTTTAUTCCAUTGCCGACAACTCGTGGTTTC-GTCTTATA
Hap10	16	25	0.007	0.009	0.725	0.310	GAGTTGTCUCACCGGGCUGCTCGGCTTTAUTCCAUTGCCGACAACTCGCGGTCCC-GTCCTATA
Hap11	11	30	0.005	0.011	0.420	0.015	TAGTTGTTTCACCGTGCUCCTCGAUTTTATTCCAUTGCTGACAGCTCGTGGCCCCA-TCCCGTG

U, small nucleotide insertion. The specific nucleotides are as follows: rs5899324, U = ACT; rs141224302, U = TGAAAAGTA; rs36052460, U = AT; rs34388138, U = CT; rs35245612, U = TAA; and rs58187401, U = AGT.

*Age- and gender-adjusted.

†SNPs included in haplotype analyses are as follows: (1) rs1533180, (2) rs5899324, (3) rs1877431, (4) rs7045465, (5) rs72751573, (6) rs7030256, (7) rs7030280, (8) rs62573974, (9) rs141224302, (10) rs10983700, (11) rs10983701, (12) rs1588636, (13) rs1588635, (14) rs7028661, (15) rs10818048, (16) rs10818050, (17) rs140405800, (18) rs36052460, (19) rs10983705, (20) rs72751578, (21) rs7021576, (22) rs2401637, (23) rs10759927, (24) rs1877432, (25) rs34388138, (26) rs7847663, (27) rs4743130, (28) rs1561962, (29) rs925488, (30) rs35245612, (31) rs925489, (32) rs4273946, (33) rs7020976, (34) rs7032019, (35) rs58187401, (36) rs7864322, (37) rs7850258, (38) rs1443438, (39) rs12347079, (40) rs10818071, (41) rs7030241, (42) rs7027030, (43) rs76293252, (44) rs12352658, (45) rs7847449, (46) rs10739496, (47) rs10983761, (48) rs4743131, (49) rs56202421, (50) rs965513, (51) rs10759944, (52) rs16924274, (53) rs10983789, (54) rs147368517, (55) rs13295081, (56) rs76228779, (57) rs13290258, (58) rs13295254, (59) rs10983795, (60) rs10983796, (61) rs1867281, (62) rs10818090, and (63) rs10818094.

Table S4. Relative risk of Hap1

Variable	Total samples	Diplotype		
		0/0	0/Hap1	Hap1/Hap1
Case	1,146	437 (38%)	538 (47%)	171 (15%)
Control	1,328	722 (54%)	521 (39%)	85 (6%)
OR*		1	1.695	3.423
P value*		—	1.18E-09	4.90E-17

Hap1 relative risk for heterozygous (0/Hap1) and homozygous (Hap1/Hap1) compared with risk for noncarriers (0/0) is shown.

*Logistic regression analyses, adjusting for age and gender.

Table S5. PCR primer and probe sequences

Primer name	Sequence
Primers for ChIP assay	
rs7864322-forward (enhancer 1)	CATACCAGAGCTGCGACTCA
rs7864322-reverse (enhancer 1)	GCATTCTAAGAGCCACATCCA
rs12352658-forward (enhancer 2a)	GGGAGGTGGTGTTTGAACTG
rs12352658-reverse (enhancer 2a)	GCATCCCACCCACATTTTAC
rs7847449-forward (enhancer 2b)	GAGTCCAAAGCAGCATGACA
rs7847449-reverse (enhancer 2b)	ATTGCAAGGAGGGAGGAGTT
rs10759944-forward (enhancer 3)	CCCACAGAGAAAAGGTTTGG
rs10759944-reverse (enhancer 3)	CAGGTCAAGGGGACCCTAAT
Primers for making luciferase constructs	
Enhancer 1-forward	AAAAATATTGGGGGCAGAGG
Enhancer 1-reverse	TTACTCCATTGCCCAGATCA
Enhancer 2a-foward	GGGAGGTGGTGTTTGAACTG
Enhancer 2a-reverse	GCATCCCACCCACATTTTAC
Enhancer 2b-foward	GTAAAATGTGGGTGGGATGC
Enhancer 2b-reverse	GTCATGGTGGTGGGAATAGG
Enhancer 3-forward	CAGCCATGAGTTCACAAACC
Enhancer 3-reverse	CAGAAACCCAAGCTCCTCAG
Primers for 3C assay	
Fragment 6_ConstantPrimer_Forward	CTCCCAAATCATTCCACACAAG
Fragment 6_Probe	6FAMACCACTGTAATTTCTCTGTTGGAGTCTGCMGBNFQ
Fragment15-forward	CCC TCT GGA GGA TGC AGT AA
Fragment16-forward	CTG ACA TAC GAA GCC GGA AT
Fragment18-forward	AGG ACA CCA GGG AGA AAG GT
Fragment19-forward	GTG CCT GGT TTA GTC CCA GT
Fragment20-forward	ACA CGT TCT TCC CTC ATT GC
Fragment21-forward	CCT CTG ACC CCT CAG CTC TA
Fragment 19_ConstantPrimer_Forward	ACCGTGCCTGGTTTAGTCC
Fragment 19_Probe	6FAMGGTTTTTAGACTTTGACTAGAAGCTTMGBNFQ
Fragment2-forward	TTA CTA TGG GCT GGG AGG TG
Fragment3-forward	GTC CCT TTC CCA GAT CCA TT
Fragment4-forward	TGC TAC AAA ACA CTC AGA AAT GC
Fragment5-forward	TTGCAGAAGCAAAGTCCTCT
Fragment6-forward	CTG CAG ACC CAT CCT CAA GT
Fragment7-forward	GAC CCA ATC CCA AGG TGT AAT
Fragment8-forward	TCG GGC AGA AAA AGA TAT CAG