A multi-enhancer *RET* regulatory code is disrupted in Hirschsprung disease

Sumantra Chatterjee,[1,+] Kameko M Karasaki,[2] Lauren E Fries[1], Ashish Kapoor[3] and Aravinda Chakravarti [1,+]

[1] Center for Human Genetics and Genomics, New York University Grossman School of Medicine, New York, NY 10016; [2] Department of Biology, Johns Hopkins University, Baltimore, MD 21205, USA; [3]Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

+*Send all correspondence to:*
Aravinda Chakravarti, Ph.D.
Center for Human Genetics & Genomics
New York University Grossman School of Medicine
435 East 30th Street, Room 802/3
New York, NY 10016
T: (212) 263-8023
E: aravinda.chakravarti@nyulangone.org

Sumantra Chatterjee, Ph.D.
Center for Human Genetics & Genomics
New York University Grossman School of Medicine
435 East 30th Street, Room 823R
New York, NY 10016
T: (212) 263-8029
E: sumantra.chatterjee@nyulangone.org

Abstract

The major genetic risk factors for Hirschsprung disease (HSCR) are three common polymorphisms within *cis* regulatory elements (CREs) of the *RET* receptor tyrosine kinase gene that reduce its gene expression during enteric nervous system (ENS) development. These variants have synergistic effects on *RET* gene expression and additionally dysregulate other ENS and HSCR genes in the *RET-EDNRB* gene regulatory network (GRN). Here, we use siRNA, ChIP and CRISPR/Cas9 deletion analyses in the SK-N-SH cell line to ask, how many HSCR-associated risk variants reside in CREs and affect *RET* gene expression? We demonstrate that 31 HSCR-associated variants reside in candidate *RET* CREs, ten with differential allele-specific *in vitro* enhancer activity and seven affecting *RET* gene expression; of these, five bind the transcription factors PAX3, RARB, GATA2 and SOX10. These and our prior results demonstrate that common sequence variants in at least 10 *RET* enhancers affect HSCR risk, extending the known *RET-EDNRB* GRN to reveal an extensive regulatory code modulating disease risk even at a single gene.

Introduction

It is now well established that most human complex traits and diseases arise from the additive genetic effects of hundreds to thousands of variants distributed across the genome[1]. At each locus, multiple statistically significant variants are detected but it is unknown how many of them make *functionally independent* contributions to the phenotype. The widespread existence of genetic association (linkage disequilibrium, LD) between local sequence variants makes this a difficult question to answer by statistical methods alone, and requires experimental perturbation and assessment of each candidate variant [2]. This is because genetic associations between variants depend on their recombination frequency, not their functional effects, thereby associating causal with innocent variants.

The majority of causal variants that contribute to trait variation reside within *cis* regulatory elements (CRE) and enhancers of a target gene thereby modulating its gene expression, usually in a cell-type specific manner [3,4]. Such gene expression control is assumed to occur within a topologically associating domain (TAD) [5,6], defining the physical locus within which CREs function. However, three major questions remain unanswered. First, since a TAD usually harbors multiple CREs and genes, which CREs affect which gene's expression? Second, do the different CREs of a specific gene have unique functions in space, time and cellular states or are they redundant (shadow enhancers)? Do they act independently, or are they synergistic and require clustering (super enhancers) for function [3,7,8]? Third, since most CRE effects are small how do such small gene expression effects modulate phenotypes? The existence of many experimental methods to identify CREs comprehensively now allows us to address these questions [9,10].

In this study, we use Hirschsprung disease (HSCR; congenital colonic aganglionosis) as an exemplar to identify the number and activity of all ENS enhancers with disease-associated variants at its major gene *RET*. HSCR is a complex neurodevelopmental disorder in which failure of differentiation of enteric neural crest cell (ENCC) precursors during ENS development leads to aganglionosis; >35 genes/loci explaining 67% of its population attributable risk have been identified [11]. Significantly, most of this risk arises from coding and enhancer variants of the *RET* receptor tyrosine kinase gene with smaller

contributions from other genes, all of whose functions in ENS development are united through a gene regulatory network (GRN) that co-regulates *RET* and *EDNRB* [2,11-13]. Thus, we also examine whether the CRE sequence variants we identify are individually sufficient to perturb *RET* gene expression as well as GRN activity.

Specifically, using the human neuroblastoma cell line SK-N-SH, we test 31 HSCR-associated common (minor allele frequency, MAF≥10%) polymorphisms at the *RET* locus [14] to demonstrate that 22 lie within CREs and of which 7 have sequence variants with differential enhancer and *RET* gene expression activity; in addition, 3 were identified before [2]. Two of the new CREs bind the transcription factor (TF) PAX3, known to affect neural crest cell migration and ENS differentiation [15,16]. Reducing *PAX3* expression decreases *RET* gene expression as well as that of *SOX10*, the major TF in the *RET-EDNRB* GRN [2,17,18]. Deleting individual *RET* CREs using CRISPR/Cas9 genome editing reduces *RET* expression but is insufficient to perturb genes expression for other GRN genes. We know that *RET-EDNRB* GRN effects are evident only when *RET* gene expression falls below 50% of its wildtype level [12]. Therefore, in HSCR, significant reductions of gene expression at *RET* and its GRN to affect ENS development only occurs in individuals with multiple CRE variants in combination with variants in other GRN genes. The corollary is that a diverse regulatory code affects complex disease genes.

Results

Enhancers at the RET locus

To create a complete catalog of common (MAF≥10%) *RET* regulatory variants associated with HSCR, we began with analysis of all 38 genome-wide significant non-coding single nucleotide polymorphisms (SNPs) discovered in a genome-wide association study (GWAS) of 220 HSCR trios comprising a proband and both of her/his parents [14]. These SNPs were distributed across 6 LD blocks in a 155kb TAD (chr10: 43434933-43590368; hg19) containing *RET* as the sole gene (**Figure 1A**). We previously analyzed 8 of these SNP-containing genomic elements because they each disrupted a predicted TF binding site (TFBS) determined form ENCODE ChIP-seq data [2].

We first asked: do the 30 remaining SNPs resided within CREs? We conducted *in vitro* functional tests of enhancer activity by cloning ~500 bp elements centered on each risk variant (**Table 1**) into a pGL4.23 luciferase vector, with a minimal TATA-box of the β-globin gene and transfecting them into the human neuroblastoma SK-N-SH cell line. SK-N-SH expresses all known members of the *RET-EDNRB* GRN and is an appropriate cell model system for studies of ENS transcriptional regulation [2,12]. Two pairs of SNPs (rs17158318/ rs17158320 and rs2506021/ rs2435342) were only 64bp and 108bp apart and were cloned into the same elements (E9 and E24, respectively) (**Table 1**). As positive controls, we re-analyzed 3 HSCR associated SNPs (rs2506030, rs7069590, rs2435357) previously shown to be *RET* enhancer variants [2] (**Table 1**). Our reporter assays showed that 22 new elements had significant enhancer activity (P <0.001 and >2X reporter activity over the promoter only control vector) of which 7 (E2, E4, E5, E14, E26, E27 and E28) also displayed differential reporter activity between the risk and non-risk alleles (**Table 1 and Figure 1B**). Among the latter, 71% (E2, E4, E26, E27 and E28) overlapped an open chromatin region or an enhancer-associated epigenetic mark in the human fetal gut and SK-N-SH [19,20] while 20% (E9, E16, E21) of the remaining 15 elements with reporter activity that had no allelic difference overlapped a potential enhancer mark (**Figure 1A**). Thus, at least 10 functionally distinct CREs within the *RET* TAD can potentially affect HSCR risk.

Haplotype specific effect of causal RET polymorphisms

The identification of 10 CRE-associated risk variants (rs788263, rs788261, rs788260, rs2506030, rs1547930, rs7069590, rs2435357, rs12247456, rs7393733 and rs2505541; **Table 1**) prompted us to ask which allelic combinations were associated with disease risk. First, we estimated haplotypes and their frequencies for all 10 SNPs in 220 unrelated HSCR cases [14] and 503 unrelated controls from the 1000 Genomes project [21], all of non-Finnish European-ancestry. Second, we estimated the odds ratio (OR) for all haplotypes with a frequency ≥1% in controls. We observed 10 distinct haplotypes of which CTGAACCAC**T** (risk allele in bold) was used as the reference because it had the smallest number (one) of risk alleles (**Table 2**) and we have previously shown that HSCR risk scales with increasing number of CRE variants [2,11-13]: significant risk was observed for two haplotypes, **GCAGGTTGGT** (OR 12.2, 95% CI: 5.97-24.93, $P=7.02\times10^{-12}$) and CTGA**GTTGGT** (OR 7.2, 95% CI: 3.26-15.91, $P=1.02\times10^{-6}$) (**Table 2**). Unsurprisingly, these two haplotypes contain

our previously identified risk increasing ATT and GTT haplotypes (for rs2506030, rs7069590, rs2435357) [2]. The 10-SNP risk haplotypes differ only for the first four SNPs (rs788263, rs788261, rs788260, rs2506030) which occur within the most 5' LD block. SNPs within this LD block do contribute to HSCR but we do not have a sufficient sample size to test our hypothesis that **GCAGGTTGGT** (OR 12.2) has a significantly higher risk than CTGA**GTTGGT** (OR 7.2)**.** Even discounting these variants, HSCR risk is clearly spread over at least three LD blocks suggesting multiple independent enhancer variants contributing to risk (**Figure 1**). This is *prima facie* evidence that these SNPs contribute to risk synergistically. In other words, risk of or protection from HSCR depends on the simultaneous binding or lack-of-binding of multiple independent TFs at *RET* CREs, implying a *RET* regulatory code.

Transcription factors regulating RET

To identify the TFs that underlie this code, we searched for TFBSs within the 7-novel risk-associated CREs and identified four candidate TFs: (1) PAX3 binding to E2 (AATAAACCC; $P=4.67x10^{-5}$) and E27 (TCGTCACTCTTAC; $P= 9.99x10^{-5}$), (2) ZBTB6 to E5 (TGGCTCCATCATG; $P=2.387x10^{-6}$), and, (3) ZNF263 binding to E14 (GCCTCACTGCTCCAG; $P=8.09x10^{-5}$). To determine their relevance for HSCR, we performed qPCR in SK-N-SH cells and observed no expression for *ZNF263* and *ZBTB6* (data not shown). Further, their expression was absent in the developing mouse gut [22], where *RET* expression is critical for ENS development [23], making it highly unlikely that these TFs control *RET* expression via specific enhancers. Moreover, *ZNF263* and *ZBTB6* are both zinc finger domain-containing proteins that are GC-rich, a feature which overestimates their statistical significance owing to their rarity.

In contrast, we detected robust expression of *PAX3* in both SK-N-SH and the developing mouse gut [22]. Thus, we performed ChIP-qPCR for *PAX3* in SK-N-SH cells and detected significant binding at both E2 (18-fold enrichment; $P= 10^{-3}$) and at E27 (26-fold enrichment, $P= 5x10^{-4}$) (**Figure 2**). We further demonstrated the specificity of this binding by performing ChIP-qPCR after siRNA-mediated knockdown of PAX3 in SK-N-SH cells to show a 1.3-fold ($P=8x10^{-3}$) reduced binding at E2 and 2-fold ($P=4x10^{-4}$) reduced binding at E27 (**Figure 2**). To further

prove that PAX3 does indeed control *RET*, we also measured *RET* gene expression after siRNA-mediated knockdown of *PAX3*. As positive controls, we measured *RET* levels after siRNA-mediated knockdown of the established *RET* TFs SOX10, GATA2 and RARB. These experiments showed that decreasing *PAX3* led to a 49% ($p=4x10^{-4}$) reduction in *RET* in comparison to 76% ($p=2.3x10^{-6}$), 50% ($p= 3.1x10^{-3}$) and 81% ($p= 4.1x10^{-5}$) decreases consequent to *SOX10*, *GATA2* and *RARB* knockdown, respectively; as a control, knockdown of *RET* by its specific siRNA reduced its expression by 96% ($p=4.4x10^{-6}$) (**Figure 3A**). We have previously demonstrated that there is considerable cross-talk between the established RET TFs [2]. Consequently, we measured gene expression of *SOX10*, *GATA2* and *RARB* after siRNA-mediated knockdown of *PAX3*: we observed only a significant drop in *SOX10* gene expression (32% decrease, $P= 3x10^{-3}$); *GATA2* and *RARB* levels were decreased but not significantly so (**Figure 3B)**.

in vivo evidence for RET enhancers:

The human genetic evidence for HSCR-associated polymorphisms within CREs identified from *in vitro* (reporter activity) and *ex vivo* (siRNA in SK-N-SH) experiments can be buttressed by deletion analysis of each enhancer rather than knockdown of its cognate TF. To do so, we designed a single guide RNA close to each HSCR-associated SNP for all 10 target enhancers to introduce non-homologous end joining-induced deletions in SK-N-SH cells. We screened 5 independently transfected wells for each guide and verified them by Sanger sequencing: none of the deletions were > 10bp. Except for enhancer E5, successful deletions were observed in the remaining CREs encompassing the polymorphic site. We used the *Inference of CRISPR Edits (ICE)* tool [24] to estimate that individual guides introduced deletions >3bp in 10-50% of the cells in all successfully targeted CREs (**Supplementary Table 1**).

We next measured *RET* gene expression in these enhancer-deleted cells. Our results demonstrate that except for enhancers E2 and E14, deletion of DNA sequences surrounding the HSCR-associated SNPs in all other CREs led to changes in *RET* expression. Thus, deletion of E4 (24%; $p=3.2x10^{-4}$) leads to higher expression while deletion of E26 (28%; $p= 3.7x10^{-4}$), E27 (19%; $p=1.2x10^{-3}$) and E28 (29%; $p=3.2x10^{-4}$) all lead to lower *RET* expression (**Figure 4A**). The positive controls, RET-7 (22%; $p=1.3x10^{-3}$), RET-5.5 (22%; $p=2x10^{-3}$), and

RET+3 (32%; $p=4.1x10^{-4}$) reduced *RET* gene expression as expected. It is interesting that all 4 intronic enhancers (RET+3, E26, E27 and E28) residing within a 8.5 kb region (chr10:43,581,812-43,590,347) in the first intron of *RET* independently control *RET* expression. Thus, these elements might be a part of a single *enhanceosome* critical for spatio-temporal expression of *RET*; note that the risk-associated variants at these sites are on a single haplotype (**Figure 1A**).

HSCR is associated with loss of *RET* function [25,26] and diminished *RET* expression leads to loss of ENS during gut development in mice, the hallmark of HSCR [22,27]. Thus, we predict that the cumulative effect of the disruption of all HSCR-associated CRE at E4, RET-7, RET-5.5, RET+3, E26, E27 and E28 should lead to reduced *RET* expression. Normalized *RET* expression averaged over individual enhancer deletion experiments indeed demonstrates that *RET* expression is diminished by 16% (p= 0.003) when compared to CAS9 only transfected (wildtype) cells (**Figure 4B**). Thus, irrespective of the direction of change in *RET* expression due to individual deletion, the cumulative disruption of these enhancers leads to *RET* deficiency.

We next asked whether these changes in *RET* expression had concomitant changes in the expression of the remaining members of the *RET* GRN by quantifying expression of *EDNRB*, *SOX10*, *GATA2*, *RARB, NKX2.5 and PAX3*, the members of the GRN which expresses in this cell line, in the individual enhancer-deleted cells. Our results demonstrate no significant changes in gene expression (**Figure 4C**). This result is not unexpected given our previous finding that GRN transcriptional dysregulation occurs only when *RET* gene expression falls below 50% of wildtype levels [2,12].

Since the individually strongest HSCR associations were within the previously discovered RET-7, RET-5.5 and RET+3 enhancers (OR of 1.8, 1.7 and 4.1 respectively) [2], we wanted to ascertain if DNA sequences beyond the polymorphic site impacted *RET* expression. Thus, we generated larger deletions in these enhancers using paired guide RNAs to direct espCAS9 to these specific DNA sequences in SK-N-SH (**Figure 4D**). We screened 48 independently transfected wells for each enhancer and detected only heterozygous deletions for each. For

RET-7, we identified 2 clones (4%) with 169bp and 183bp deletions, both of which deleted rs2506030. We detected 3 heterozygous clones for RET-5.5 deletion (6.25%): 2 clones with 147bp and 1 clone with a 139bp, all of which deleted rs7069590. Similarly, we also identified 3 clones (6.25%) with RET+3 deletion with 1 clone encompassing 147bp and 2 clones encompassing 138bp deletions, all of which deleted rs2435357(**Figure 4D**). Next, we measured gene expression of *RET*, *EDNRB*, *SOX10*, *GATA2*, *RARB, NKX2.5 and PAX3* by qPCR in these cells to detect the effect of heterozygous enhancer deletions on the expressed genes of the *RET* GRN. Deletion of RET+3 led to 32% ($P=3.1x10^{-3}$) loss of *RET* expression compared to cells transfected with only a Cas9 vector without the guide RNA; the corresponding decreases for RET-5.5 and RET-7 were 22% ($P=2.8X10^{-3}$) and 20% ($P=2.6X10^{-3}$), respectively, consistent with their individual risks (OR of 4.1, 1.7 and 1.8, respectively) (**Figure 4E**) (Chatterjee et al., 2016; Emison et al., 2010; Kapoor et al., 2015). Thus, deletion of sequences beyond the region immediately surrounding the polymorphisms do not seem to have any substantial effect on *RET* expression. Thus, for HSCR, these SNPs are the causal risk factors. Here as well, the remaining genes of the GRN are transcriptionally unchanged (**Figure 4E**) because *RET* expression never falls below 50%.

Discussion

It is evident that a multiplicity of enhancers control a gene's expression: this feature has many implications for complex disease genetics and its mechanisms. The data reported here, based on human genetics, siRNA, and ChIP analyses in the SK-N-SH cell line, together with our prior studies [2,11-13], have identified 38 distinct CREs of *RET* with sequence variants that are associated with HSCR. Of these, 10 contain common polymorphisms with differential allelic enhancer activity. CRISPR/Cas9 deletion analyses of these CREs demonstrate that at least 7 of these *RET* enhancers, binding the SOX10, RARB, GATA2, PAX3 and yet unknown TFs, have all of the hallmarks of harboring causal variants for HSCR (**Table 3**). Given that none of these experimental approaches are 100% efficient yet other enhancers are likely. Nevertheless, the many enhancers at the *RET* locus, identified through *in vitro* analyses, do not imply that all of them are involved in gut (as opposed to other tissue) development, and, if so, whether they have unique roles in the ENS across development, as we have shown earlier (Chatterjee et al., 2016) or are merely shadow enhancers [8]. *RET* is highly expressed in the

mid- and fore-brain, kidney and dorsal root ganglia: thus, these CREs may regulate *RET* in these tissues.

Definitive proof of our hypothesis that the 7 *RET* enhancers cause HSCR will require the creation of mouse models with multiple regulatory variants at the same locus, far more feasible using new methods of synthetic biology [28] as opposed to CRISPR/Cas9 genome editing. Nevertheless, this hypothesis is supported by the observation that RARB, GATA2, SOX10 and PAX3 binding to five of these enhancers have known roles in ENS development and HSCR [15,16,29]. In other words, if the trans factors lead to HSCR and the cis factors that bind them are HSCR-associated then these cis factors are direct risk factors of HSCR. Their distinct nature and LD relationships also suggests that they contribute independently to *RET* gene expression, and, therefore to HSCR.

The multiplicity of noncoding variants in CREs all controlling the same gene should give us pause in interpreting the effect size and functional effects of individual GWAS variants. As we have shown, the cumulative effect of all variant-bearing enhancers is to *significantly lower RET* gene expression, as expected from *RET* coding mutations [17,18]. As we have also demonstrated, the largest risks are associated with haplotypes with multiple risk alleles, even across LD blocks [2]. This accumulation of multiple CRE variants is expected to lead to a larger effect through reduced binding of multiple TFs to multiple enhancers. Given the role of multiple TFs on the promoter this is likely synergistic. Additionally, other interactions are important in ENS cells with these genotypes. Note that *RET* also controls the gene expression of its own TFs PAX3, RARB, GATA2 and SOX10. This feedback may be a secondary but an important cause of reduced *RET* gene expression further exacerbating the enhancers' effect. We do not yet know whether this diversity of genetic control with feedback is typical or not. *RET* is a highly dosage-sensitive gene with higher and lower than wildtype levels being associated with neuroendocrine tumors and aganglionosis, respectively. Although extensive regulatory control is common for developmental genes[30], the genetic lessons are likely to be universal.

The human genetic implications of these data, beyond understanding HSCR, are that genetically independent (unassociated) SNPs at a specific GWAS locus are not the only

candidate variants for understanding a phenotype. Additional SNPs may be involved, even those perfectly associated with one another provided they affect functionally independent enhancers, as has been demonstrated by the recent discovery of additional *RET* variants and CREs that control its expression [31,32]. Finally, such regulatory control by the variant allele can also be quite varied (E4, E27 versus others), decreasing versus increasing enhancer activity, depending on activator versus repressor TFs and their co-regulators, as well as decreasing as well as increasing target gene expression. This suggests that understanding the regulatory contributions to GWAS will require experimental data on enhancer effects beyond what statistical analysis can suggest. Furthermore, we need broader enhancer screens to define the full enhancer architecture of *RET* and do so *in vivo* at different developmental stages and by sex. We also need to elucidate the full repertoire of TFs that regulate *RET*. These pieces of information are crucial to understand the full extent and composition of the *RET-EDNRB* GRN, which in turn will identify new genes that become mutational targets of HSCR.

Materials and methods

Cell lines:

The human neuroblastoma cell SK-N-SH, purchased from ATCC, USA (# HTB-11), was grown under standard conditions (DMEM + 10% FBS and 1% Penicillin Streptomycin). It was maintained in 10 cm culture dishes and passaged every 48 hours when it reached ~80% confluency.

ChIP-seq peak calling:

Three epigenomic datasets generated from a 108-day human fetal large intestine, histone H3K27ac ChIP-seq (GSM1058765), histone H3K4me1 ChIP-seq (GSM1058775), and DNaseI-seq (GSM817188), were downloaded from the NIH Roadmap Epigenomics Project [19]. For the SK-N-SH cell line, DNaseI-seq data (GSM736559) were obtained from the ENCODE project [33] . For each dataset, MACS software v1.4 [34] with default settings was used to call ''peaks'' at genomic sites where sequence reads were significantly enriched over background. With the default peak-calling threshold ($P < 10^{-5}$), 51,771, 61,689, 66,930 and 52,534 genomic regions were identified in the GSM1058765, GSM1058775, GSM817188 and GSM736559 datasets, respectively.

Reporter assays:

400 ng of firefly luciferase vector (pGL4.23, Promega Corporation,USA) containing the DNA sequence of interest and 2 ng of Renilla luciferase vector (transfection control) were transiently transfected into the SK-N-SH cell line (5 – $6x10^4$ cells/well), using 6 ml of FuGENE HD transfection reagent (Roche Diagnostic, USA) in 100ml of OPTIMEM medium (Invitrogen, USA). Cells were grown for 48 hours and luminescence measured using a Dual Luciferase Reporter Assay System on a Tecan multi-detection system luminometer, per the manufacturer's instructions.

Chromatin Immunoprecipitation–qPCR (ChIP-qPCR) assays:

ChIP was performed thrice independently for each antibody using $1x10^6$ SK-N-SH cells for each transcription factor using the EZ-Magna ChIP kit (Millipore Sigma, USA), as per the manufacturer's instructions, with the following modifications: the chromatin was sonicated for

30s on and 30s off for 10 cycles; sheared chromatin was pre-blocked with unconjugated beads for 4 hours and specific antibodies separately conjugated to the beads for 4 hours before immunoprecipitation was performed with the pre-blocked chromatin. A polyclonal antibody was used against PAX3 (16HCLC; Invitrogen, USA) at 15μg concentration. ChIP assays were also performed on cells 48 hours after transfection with the PAX3 siRNAs (L-012399-00-0005; Dharmacon/Horizon Discovery, USA) at 25 μM to assess specificity of TF binding. qPCR assays were performed using SYBR green (Life Technologies, USA) and specific primers against enhancer E2 (E2_FWD 5' GCTGCAGATATGCAACTTCCAA 3' and E2_REV 5' AGATATGCTGGTGAGGGGCT 3') and enhancer E27 (E27_FWD 5' AGGAAGGTAGGCACCCTGTA and E27_REV 5' AGCCCTGTGTTAACTGTCCG 3').
The data were normalized to input DNA and enrichment was calculated by fold excess over ChIP performed with specific IgG as background signal. All assays were done in triplicate for each independent ChIP assay (n = 9).

siRNA assays:

PAX3 (L-012399-00-0005), *RET* (L-003170-00-0005), *SOX10* (L-017192-00), *GATA2* (L-009024-02 and *RARB* (L-003438-02) SMARTpool siRNAs (combination of 4 distinct siRNAs targeting each gene) along with ON-TARGET plus non-targeting siRNAs (D-001810-10, negative control) (Dharmacon/ Horizon Discovery, USA) were transfected at 20 μM in SK-N-SH cells at a density of 10^4 -10^5 cells using FuGene HD Transfection reagent (Promega Corporation, USA) per the manufacturer's instructions. Negative control siRNAs were always transfected at 25 μM concentration.

Gene expression assays:

Total RNA was extracted from SK-N-SH cells using TRIzol (Life Technologies, USA) and cleaned on RNeasy columns (QIAGEN, USA). 500μg of total RNA was converted to cDNA using SuperScriptIII reverse transcriptase (Life Technologies, USA) using Oligo-dT primers. The diluted (1/5) total cDNA was subjected to Taqman gene expression (ThermoFisher Scientific, USA) using the following transcript-specific probes and primers: *RET* (Hs01120032_m1), *EDNRB* (Hs00240747_m1), *PAX3*(Hs00240950_m1), *SOX10* (Hs00366918_m1), *GATA2* (Hs00231119_m1) and *RARB* (Hs00977140_m1)

Human β-actin was used as an internal loading control for normalization. For siRNA knockdown experiments, five independent wells of SK-N-SH cells were used for RNA extraction and each assay performed in triplicate (n = 15). Relative fold change was calculated based on the 2ΔΔCt (threshold cycle) method. For siRNA experiments, 2ΔΔCt for negative control non-targeting control siRNA was set to unity. P values were calculated from pairwise 2-tailed t tests and the data presented as means with their standard errors (SE).

CRIPSR/Cas9 induced deletions:

(i)*Single guide disruptions*

Each enhancer region centered on a polymorphic site was targeted using a single guide RNA (**Supplementary Table 2**) by transfecting a ribonucleoprotein complex containing 100pmol of specific gRNA, coupled with 5μg/μl of Truecut cas9 nuclease (ThermoFisher Scientific) in Lipofectamine Crisprmax solution (ThermoFisher Scientific). For each enhancer, 5 wells containing ~30,000 SK-N-SH cells were independently transfected. To increase efficiency of deletion, we re-transfected the cells with the same ribonucleoprotein mix a second time after 72 hours. The cells were further grown for 48 hours and then equally split into 2 tubes for DNA and RNA extraction. To confirm disruption of the enhancer regions, specific primers (**Supplementary Table 3**) were used to verify deletions by PCR followed by Sanger sequencing. We used the *Inference of CRISPR Edits (ICE)* tool [24] to estimate the percentage of cells carrying various insertion/deletions.

(ii) *paired guide disruptions*

The following genomic regions were targeted with pairs of guides for 3 enhancers: RET-7 (*RET-7_DGR- 5'GACCCCTCTGGACCCGTCGCGG3* and *RET-2_UGR-CCGAGTGTTTGCGCCCATGAGG*), RET-5.5 (*RET-5.5_DGR-AAGGATGGATACACCTTCCGGG* and *RET-5.5_UGR- AAGGATGGATACACCTTCCGGG*) and RET+3 (*RET+3_DGR- GCCTGGCTGGAGGTCCAAGAGG* and *RET+3_UGR-GCTCATGAGGAGCACACCGTGG*). The guide RNA and Cas9 were cloned into the vector *U6-gRNA/CMV-Cas9-GFP* (Millipore Sigma, USA) and each pair of guides were transfected at 10μg into 72 wells of low-density (~10,000 cells) plated SK-N-SH cells. We also transfected 72 wells with only the Cas9-GFP vector without the guide RNAs to serve as

controls against which to measure gene expression changes arising from our enhancer deletions.

After 24 hours cells were checked for GFP expression and wells with at least ~80% positive cells were re-plated, ensuring high transfection efficiency. For all 3 constructs and the control we obtained at least 48 wells with over 80% transfection efficiency. The cells were further grown for 24 hours and then split into 2 tubes of equal cells for DNA and RNA extractions.

To check for deletions, we used the following pairs of primers: RET-7 (*RET-7_FWD 5' GAGGAACCGCGCACATCAG 3'* and *RET-7_REV 5' CGCCTTGCCTGGCCCTCCA 3'*), RET-5.5 (*RET-5.5_FWD 5' GTCCTGGTTGCCCCACTCTG 3'* and *RET-5.5_REV 5' GGTCAGGTGTCACAAAGTCT 3'*) and RET+3 (*RET+3_FWD 5' AATGGGCAAGACCATCTCAG 3*' and RET+3_*REV 5' GTGGCCAGTAGCTGGAAGAGCAG 3*'). The primers amplify a wildtype sequence of 1.1kb for each enhancer region. Each cell line was PCR confirmed for enhancer deletion and Sanger sequenced to re-confirm the precise region of deletion. For quantifying gene expression changes, RNA was extracted from each cell line with deletions and qPCR assays performed in triplicate; n=6 for RET-7 and n=9 for RET-5.5 and RET+3. Quantitative gene expression changes were calculated as detailed above.

Estimating haplotype-specific HSCR risk:

Genotypes at 10 RET CRE variants in 220 S-HSCR cases and 503 European-ancestry controls were obtained from our published HSCR Genome Wide Association study [14] and the 1000 Genomes project [21], respectively. Haplotypes were generated from unphased genotypes using BEAGLE [35] and filtered to retain only those that had a frequency >1% in controls. Standard methods using χ^2 statistics were used to calculate haplotype-count based odds ratios (OR), their upper and lower confidence limits and significance of their deviation from the null hypothesis of no association (OR = 1) [13].

Identifying transcription factors for candidate enhancers:

We searched for TF binding sites (TFBS) within all putative CREs using FIMO [36,37] and 890 validated TF motifs in TRANSFAC [38]. We used the setting of "minimize false positives" and a stringent cut off of $P<10^{-4}$ to identify candidate cognate TFs.

Author Contributions:

S.C and A.C. conceived and designed the study. K.M.K. conducted all *in vitro* luciferase assays, A.K. all genotyping assays and S.C and L.E.F all *in vivo* CRISPR assays. S.C. and A.C. wrote and edited the manuscript.

Funding:

This work was supported by a National Institute of Health MERIT Award HD28088 to AC.

REFERENCES

1. Visscher, P.M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5-22 (2017).

2. Chatterjee, S. *et al.* Enhancer Variants Synergistically Drive Dysfunction of a Gene Regulatory Network In Hirschsprung Disease. *Cell* **167**, 355-368 e10 (2016).

3. Chakravarti, A. & Turner, T.N. Revealing rate-limiting steps in complex disease biology: The crucial importance of studying rare, extreme-phenotype families. *Bioessays* **38**, 578-86 (2016).

4. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).

5. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).

6. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).

7. Chatterjee, S. & Ahituv, N. Gene Regulatory Elements, Major Drivers of Human Disease. *Annu Rev Genomics Hum Genet* **18**, 45-63 (2017).

8. Kvon, E.Z., Waymack, R., Elabd, M.G. & Wunderlich, Z. Enhancer redundancy in development and disease. *Nat Rev Genet* (2021).

9. Kapoor, A. *et al.* Multiple SCN5A variant enhancers modulate its cardiac gene expression and the QT interval. *Proc Natl Acad Sci U S A* **116**, 10636-10645 (2019).

10. Inoue, F., Kreimer, A., Ashuach, T., Ahituv, N. & Yosef, N. Identification and Massively Parallel Characterization of Regulatory Elements Driving Neural Induction. *Cell Stem Cell* **25**, 713-727 e10 (2019).

11. Tilghman, J.M. *et al.* Molecular Genetic Anatomy and Risk Profile of Hirschsprung's Disease. *N Engl J Med* **380**, 1421-1432 (2019).

12. Chatterjee, S. & Chakravarti, A. A gene regulatory network explains RET-EDNRB epistasis in Hirschsprung disease. *Hum Mol Genet* **28**, 3137-3147 (2019).

13. Kapoor, A. *et al.* Population variation in total genetic risk of Hirschsprung disease from common RET, SEMA3 and NRG1 susceptibility polymorphisms. *Hum Mol Genet* **24**, 2997-3003 (2015).

14. Jiang, Q. *et al.* Functional loss of semaphorin 3C and/or semaphorin 3D and their epistatic interaction with ret are critical to Hirschsprung disease liability. *Am J Hum Genet* **96**, 581-96 (2015).

15. Bondurand, N. *et al.* Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome. *Hum Mol Genet* **9**, 1907-17 (2000).

16. Lang, D. *et al.* Pax3 is required for enteric ganglia formation and functions with Sox10 to modulate expression of c-ret. *J Clin Invest* **106**, 963-71 (2000).

17. Emison, E.S. *et al.* A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434**, 857-63 (2005).

18. Emison, E.S. *et al.* Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. *Am J Hum Genet* **87**, 60-74 (2010).

19. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-8 (2010).

20. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

21. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

22. Chatterjee, S., Nandakumar, P., Auer, D.R., Gabriel, S.B. & Chakravarti, A. Gene- and tissue-level interactions in normal gastrointestinal development and Hirschsprung disease. *Proc Natl Acad Sci U S A* (2019).

23. Natarajan, D., Marcos-Gutierrez, C., Pachnis, V. & de Graaff, E. Requirement of signalling by receptor tyrosine kinase RET for the directed migration of enteric nervous system progenitor cells during mammalian embryogenesis. *Development* **129**, 5151-60 (2002).

24. Hsiau, T. *et al.* Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv*, 251082 (2019).

25. Pasini, B. *et al.* Loss of function effect of RET mutations causing Hirschsprung disease. *Nat Genet* **10**, 35-40 (1995).

26. Angrist, M. *et al.* Mutation analysis of the RET receptor tyrosine kinase in Hirschsprung disease. *Hum Mol Genet* **4**, 821-30 (1995).

27. Uesaka, T., Nagashimada, M., Yonemura, S. & Enomoto, H. Diminished Ret expression compromises neuronal survival in the colon and causes intestinal aganglionosis in mice. *J Clin Invest* **118**, 1890-8 (2008).

28. Richardson, S.M. *et al.* Design of a synthetic yeast genome. *Science* **355**, 1040-1044 (2017).

29. Lang, D. & Epstein, J.A. Sox10 and Pax3 physically interact to mediate activation of a conserved c-RET enhancer. *Hum Mol Genet* **12**, 937-45 (2003).

30. Bolt, C.C. & Duboule, D. The regulatory landscapes of developmental genes. *Development* **147**(2020).

31. Fu, A.X. *et al.* Whole-genome analysis of noncoding genetic variations identifies multiscale regulatory element perturbations associated with Hirschsprung disease. *Genome Res* **30**, 1618-1632 (2020).

32. Kapoor, A. *et al.* Multiple, independent, common variants overlapping known and putative gut enhancers at RET, SEMA3 and NRG1 underlie Hirschsprung disease risk in European ancestry subjects. *bioRxiv*, 2020.06.07.138719 (2020).

33. Consortium, E.P. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).

34. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

35. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**, 1084-97 (2007).

36. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-8 (2009).

37. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-8 (2011).

38. Wingender, E., Dietze, P., Karas, H. & Knuppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* **24**, 238-41 (1996).

Figure Legends

Figure 1: **The *RET* regulatory landscape in the enteric nervous system.** (A) The 155kb *RET* locus (10:43434933-43590368; hg19) contains 38 putative cis regulatory elements (CRE) with HSCR-associated polymorphisms in 4 linkage disequilibrium (LD) blocks. Multiple enhancer-associated epigenetic marks (DHS: DNaseI hypersensitivity, H3K27ac, H3K4me1) in 108-day human fetal large intestine and the SK-N-SH neuroblastoma cell line, and transcription factor (TF) binding sites, from public sources, are noted. All common (≥10% allele frequency) variants associated with HSCR are shown, with those showing allelic difference in *in vitro* transcription assays marked in red (this paper). (B) Allele specific *in vitro* luciferase assays of the 38 CREs in (A) in SK-N-SH cells are shown: 22 CREs act as enhancers, as compared to a promoter-only control, of which, 7 also show allelic difference in luciferase activity (boxed) for its cognate HSCR-associated polymorphism in addition to RET-7, RET-5.5 and RET+3 positive controls. Error bars are standard errors (SE) of the mean (*P < 0.01, **P < 0.001) for 3 independent biological replicates for each assay.

Figure 2: Identification of cognate transcription factors bound to *RET* enhancers. Genome map of the *RET* locus with locations of the E2 and E27 CREs together with ChIP-qPCR results using a PAX3 antibody in SK-N-SH cells shows enrichment of binding as compared to the background. The specificity of binding is shown by siRNA knockdown of *PAX3* with concomitant reduction in ChIP-qPCR signals at both CREs. Error bars are standard errors (SE) of the mean (*P < 0.01, **P < 0.001) for 3 independent biological replicates for each ChIP assay.

Figure 3: Transcription factor-mediated *in vitro* and *in vivo* effects on gene expression.
(A) siRNA-mediated knockdown of *PAX3*, *SOX10*, *GATA2*, *RARB and RET* in SK-N-SH cells decreases *RET* gene expression. (B) siRNA-mediated knockdown of *PAX3* has significant transcriptional effects on *SOX10* but small yet statistically insignificant effects on

GATA2 and *RARB*. Error bars are standard errors (SE) of the mean (*P < 0.01, **P < 0.001) for 5 independent biological replicates in all experiments.

Figure 4: CRISPR/Cas9 induced deletions of *RET* enhancers with HSCR-associated variants reduce *RET* gene expression *in vitro*. (A) There is significant loss of *RET* gene expression from 7 of 9 CREs with small (≤10bp) deletions centered on the variant site; only the E14 enhancer shows increased gene expression. (B) There is significant loss of average *RET* expression from the cumulative effect of individual enhancer deletions as compared to wildtype cells. (C) The expression of other *RET* GRN genes are unaffected by these CRE deletions likely owing to *RET* expression never decreasing below 50%. (D) A genomic map of the *RET* locus showing 3 positive control CREs and their corresponding HSCR-associated polymorphisms: RET-7 (rs2506030), RET-5.5 (rs7069590) and RET+3 (rs2435357). The green boxed region is the deleted region for each enhancer as confirmed by Sanger sequencing and validated by agarose gel electrophoresis. (E) Heterozygous deletion of all 3 enhancers lead to significant loss of *RET* gene expression but not other *RET* GRN genes. Error bars are standard errors (SE) of the mean (*P < 0.01, **P < 0.001) for 5 independent biological replicates in all experiments for single guide deletions, across 6 replicates for RET-7 and 9 replicates for RET-5.5 and RET+3 in dual guide deletions.

TABLES

Genomic element	SNP ID	Genomic location (chr10)		LD Block	Non-Risk Allele	Risk Allele
		Start	End			
E1	rs788267	43434932	43435936	Block 1	C	T
E2	**rs788263**	43437007	43437506		C	G
E3	rs788262	43437440	43437943		A	G
E4	**rs788261**	43437726	43438225		T	C
E5	**rs788260**	43438228	43438727		G	A
E6	rs2995411	43440554	43441053		C	T
E7	rs1582227	43441446	43441952		T	C
E8	rs2488278	43446082	43446581		T	C
RET-7	**rs2506030**	43447620	43448074		A	G
E9	rs17158318/ rs17158320	43448581	43449090		G/C	A/A
E10	rs947696	43455083	43455582		G	T
E11	rs7908085	43460567	43461065		T	A
E12	rs10900290	43471322	43471823		C	T
E13	rs947690	43479479	43479979	Block 2	G	C
E14	**rs1547930**	43483056	43483559		G	A
E15	rs3004258	43483900	43484403		T	G
RET-5.5	**rs7069590**	43552669	43553121	Block 3	C	T
E16	rs3026703	43557546	43558048		T	C
E17	rs3026707	43558368	43559122		A	G
E18	rs2505989	43563195	43563700		G	C
RET+3	**rs2435357**	43581829	43582283	Block 4	C	T
E19	rs741763	43568087	43568586		G	C
E20	rs2505995	43569379	43569878		A	G
E21	rs2506010	43573167	43574025		C	T
E22	rs2506020	43578754	43579373		C	T
E23	rs2506021/ rs2435342	43583869	43584408	Block 5	C/T	T/C
E24	rs2506022	43584264	43584845	Block 6	C	T
E25	rs2435343	43585384	43585889		T	G
E26	**rs12247456**	43587982	43588420		G	A
E27	**rs7393733**	43588440	43588915		G	C
E28	**rs2505541**	43589862	43590368		T	C

Table 1: Genomic coordinates (hg19) of 31 elements containing 33 single nucleotide polymorphisms (SNPs), within 6 linkage disequilibrium (LD) blocks at the *RET* locus which are associated with Hirschsprung disease (HSCR). 25 elements in bold act as enhancers in *in vitro* transcriptional assays while 10 SNPs in bold show allelic difference between the risk and non-risk allele as well. Enhancer activities of elements RET-7 (rs2506030), RET-5.5 (rs7069590) and RET+3 (rs2435357) have been previously demonstrated to be affected by the HSCR associated risk alleles[2].

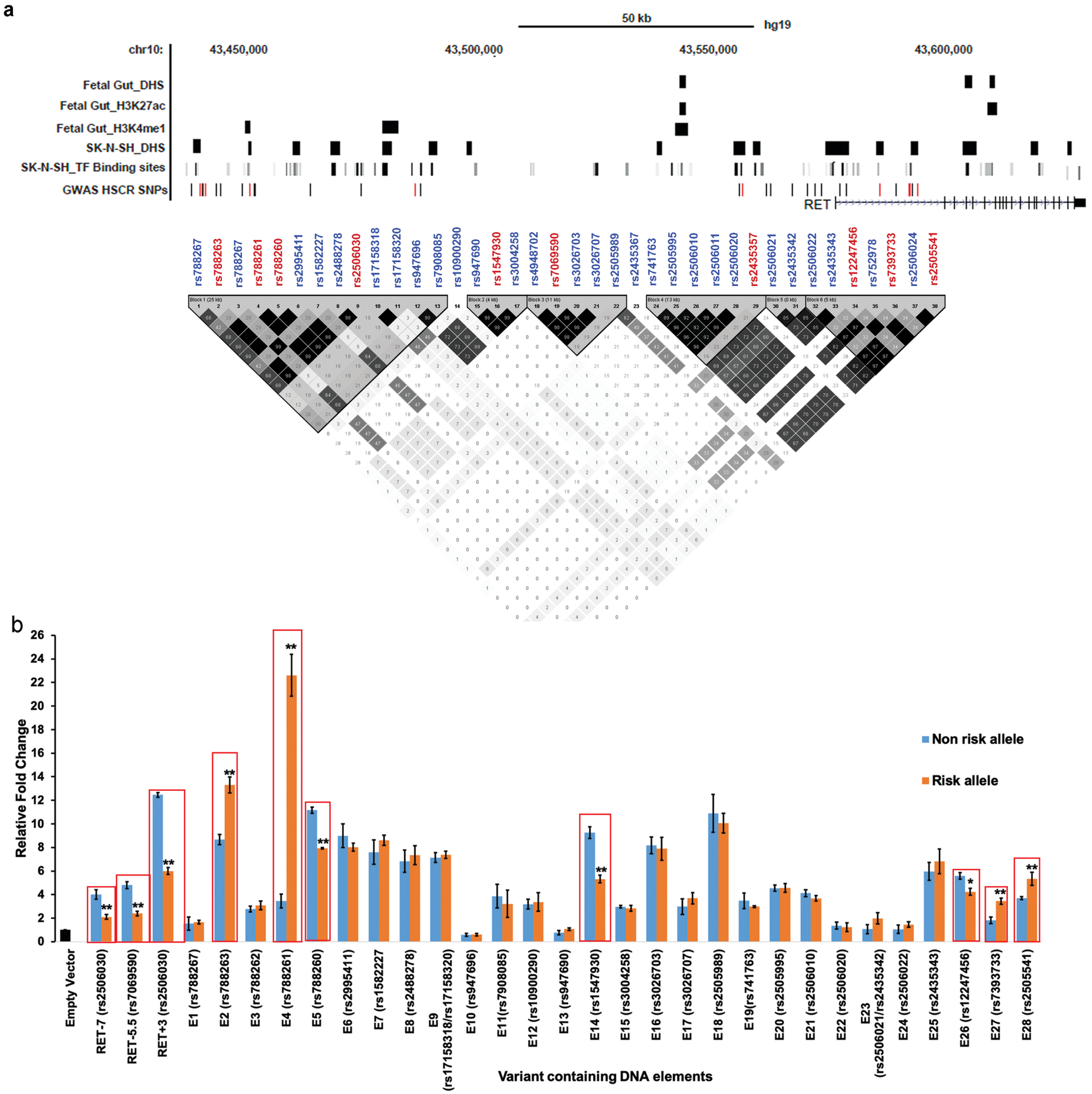
Haplotype	Case Frequency	Control Frequency	Odds Ratio (95% CI)	P
CTGAACCAC**T**	0.02	0.08	1	1
CTGAA**T**CAC**T**	0.04	0.07	2.24 (0.95-5.29)	0.07
CTGA**G**CCAC**T**	0.03	0.10	1.13 (0.45-2.82)	0.79
CTGAA**T**C**GG**C	0.04	0.06	2.84 (1.20-6.77)	0.02
CTGA**GT**C**GG**C	0.06	0.16	1.58 (0.71-3.51)	0.26
CTGAA**TTGGT**	0.02	0.01	4.35 (1.41-13.49)	0.01
CTGA**GTTGGT**	0.10	0.06	**7.2 (3.26-15.91)**	**1.02×10^{-6}**
GCAGGTCAC**T**	0.02	0.04	2.45 (0.92-6.54)	0.07
GCAGGTC**GG**C	0.06	0.15	1.62 (0.72-3.61)	0.24
GCAGGTTGGT	0.52	0.17	**12.20 (5.97-24.93)**	**7.02×10^{-12}**

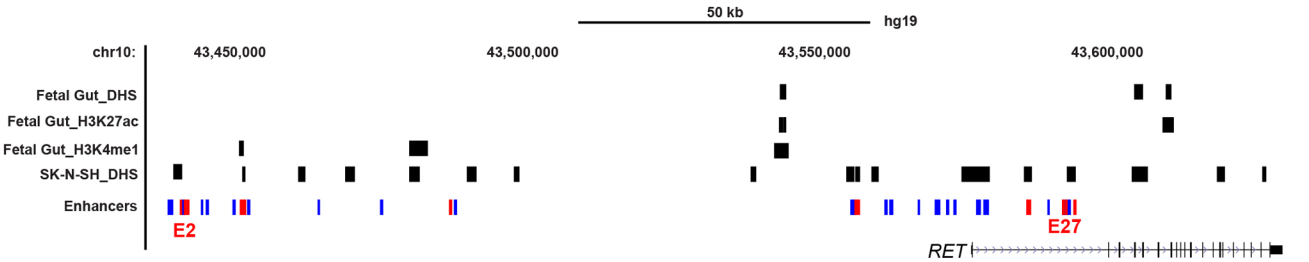
Table 2: Haplotypes for ten *RET* enhancer polymorphisms (rs788263, rs788261, rs788260, rs2506030, rs1547930, rs7069590, rs2435357, rs12247456, rs7393733 and rs2505541), with risk alleles denoted in bold. Observed frequencies of haplotypes among 220 European ancestry HSCR cases and 503 controls together with the odds ratio (with respect to the reference haplotype CTGAACCACT containing only one susceptibility allele; significant values bolded) and statistical significance (P) are shown.

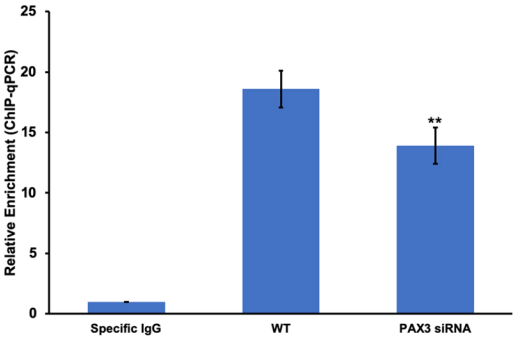
#	SNP ID	LD Block	DNA element	Coordinates (chr 10, hg19) Start	Stop	Epigenetic marks	Reporter assay	Risk allele effect	TF	*RET* expression
1	rs788267	Block 1	E1	43434932	43435936	+[a]	-	-	-	NT
2	rs788263		E2	43437007	43437506	+[a]	+	increase	PAX3	no change
3	rs788262		E3	43437440	43437943	+[a]	+	-	-	NT
4	rs788261		E4	43437726	43438225	+[a]	+	increase	-	increase
5	rs788260		E5	43438228	43438727	-	+	+	-	no change
6	rs2995411		E6	43440554	43441053	-	+	-	-	NT
7	rs1582227		E7	43441446	43441952	-	+	-	-	NT
8	rs2488278		E8	43446082	43446581	-	+	-	-	NT
9*	rs2506030		RET-7	43447346	43448347	+[a], +[b]	+	decrease	RARB	decreased
10	rs17158318		E9	43448581	43449090	+[a], +[b]	+	-	-	NT
11	rs17158320		E9	Included with SNP 10		+[a], +[b]	+	-	-	NT
12	rs947696		E10	43455083	43455582	-	-	-	-	NT
13	rs7908085		E11	43460567	43461065	-	+	-	-	NT
14	rs10900290	Block 2	E12	43471322	43471823	-	+	-	-	NT
15	rs947690		E13	43479479	43479979	-	-	-	-	NT
16	rs1547930		E14	43483056	43483559	+[a]	+	decrease	-	no change
17	rs3004258		E15	43483900	43484403	-	+	-	-	NT
18	rs4948702	Block 3	rs4948702	43551663	43552663	+[a]	+	-	-	NT
19*	rs7069590		RET-5.5	43552669	43553121	+[a]	+	decrease	GATA2	decreased
20	rs3026703		E16	43557546	43558048	-	+	-	-	NT
21	rs3026707		E17	43558368	43559122	-	+	-	-	NT
22	rs2505989		E18	43563195	43563700	-	+	-	-	NT
23	rs2435367		rs2435367	43566114	43567114	-	+	-	-	NT
24	rs741763	Block 4	E19	43568087	43568586	-	+	-	-	NT
25	rs2505995		E20	43569379	43569878	-	+	-	-	NT
26	rs2506010		E21	43573167	43574025	+[a]	+	-	-	NT
27	rs2506011		rs2506011	43574436	43575436	+[a]	+	-	-	NT
28	rs2506020		E22	43578754	43579373	-	-	-	-	NT
29*	rs2435357		RET+3	43581829	43582283	+[a], +[b]	+	decrease	SOX10	decreased
30	rs2506021	Block 5	E23	43583869	43584408	-	-	-	-	NT
31	rs2435342		E23	Included with SNP 30		-	-	-	-	NT
32	rs2506022	Block 6	E24	43584264	43584845	-	-	-	-	NT
33	rs2435343		E25	43585384	43585889	-	+	-	-	NT
34	rs12247456		E26	43587982	43588420	+[a]	+	decrease	-	decreased
35	rs752978		rs752978	43587929	43588929	-	+	-	-	NT
36	rs7393733		E27	43588440	43588915	+[a]	+	increase	PAX3	decreased
37	rs2506024		rs2506024	43588418	43589418	-	+	-	-	NT
38	rs2505541		E28	43589862	43590368	+[a]	+	increase	-	decreased

Table 3: 38 Hirschsprung disease (HSCR)-associated polymorphisms in 6 linkage disequilibrium (LD) blocks, contained within 36 DNA elements at the *RET* locus annotated with respect to epigenetic marks (DNaseI hypersensitivity (+[a]) in the SK-N-SH cell line, H3K4me1 (+[b]) marks in human fetal gut), luciferase reporter assays of alleles in the SK-N-SH cell line, allelic differences

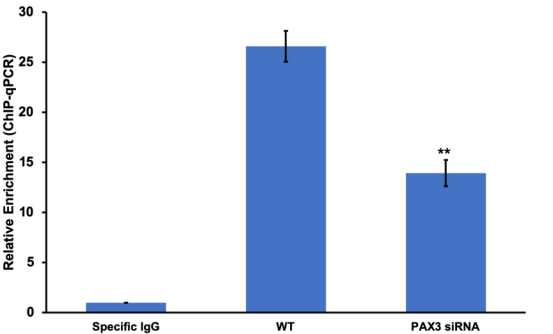
in luciferase assays in the SK-N-SH cell line, the transcription factor (TF) binding the indicated regulatory element, and whether deletion of the element affected *RET* gene expression. 31 elements have enhancer activity in *in vitro* luciferase assays, 10 of which demonstrated affect due to the HSCR associated risk polymorphisms. CRISPR based *in vivo* deletion of these 10 enhancers identified 7 which affect *RET* gene expression, 5 of which are bound by PAX3, RARB, GATA2 and SOX10 TFs. Elements #9, 18, 19, 23, 27, 29, 35 and 37 were reported in our previous study[2] and are included for completeness; of these 9,19 and 29, marked by *, are positive controls with *in vivo* evidence of enhancer activity in transgenic mice. NT: not tested.


a
50 kb
hg19
chr10:
43,450,000
43,500,000
43,550,000
43,600,000
Fetal Gut_DHS
Fetal Gut_H3K27ac
Fetal Gut_H3K4me1
SK-N-SH_DHS
SK-N-SH_TF Binding sites
GWAS HSCR SNPs
RET
rs788267
rs788263
rs788267
rs788261
rs788260
rs2995411
rs1582227
rs2488278
rs2506030
rs17158318
rs17158320
rs947696
rs7908085
rs10900290
rs947690
rs1547930
rs3004258
rs4948702
rs7069590
rs3026703
rs3026707
rs2505989
rs2435367
rs741763
rs2505995
rs2506010
rs2506011
rs2506020
rs2435357
rs2506021
rs2435342
rs2506022
rs2435343
rs12247456
rs752978
rs7393733
rs2506024
rs2505541
Block 1 (25 kb)
Block 2 (4 kb)
Block 3 (11 kb)
Block 4 (13 kb)
Block 5 (0 kb)
Block 6 (5 kb)
b
Relative Fold Change
Non risk allele
Risk allele
Empty Vector
RET-7 (rs2506030)
RET-5.5 (rs7069590)
RET+3 (rs2506030)
E1 (rs788267)
E2 (rs788263)
E3 (rs788262)
E4 (rs788261)
E5 (rs788260)
E6 (rs2995411)
E7 (rs1582227)
E8 (rs2488278)
E9 (rs17158318/rs17158320)
E10 (rs947696)
E11(rs7908085)
E12 (rs10900290)
E13 (rs947690)
E14 (rs1547930)
E15 (rs3004258)
E16 (rs3026703)
E17 (rs3026707)
E18 (rs2505989)
E19(rs741763)
E20 (rs2505995)
E21 (rs2506010)
E22 (rs2506020)
E23 (rs2506021/rs2435342)
E24 (rs2506022)
E25 (rs2435343)
E26 (rs12247456)
E27 (rs7393733)
E28 (rs2505541)
Variant containing DNA elements

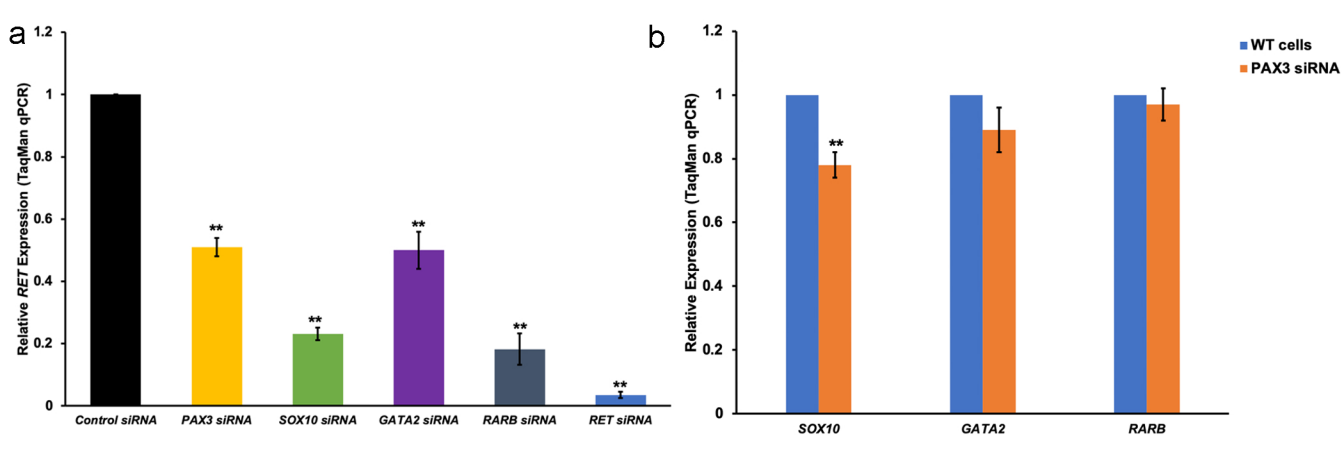


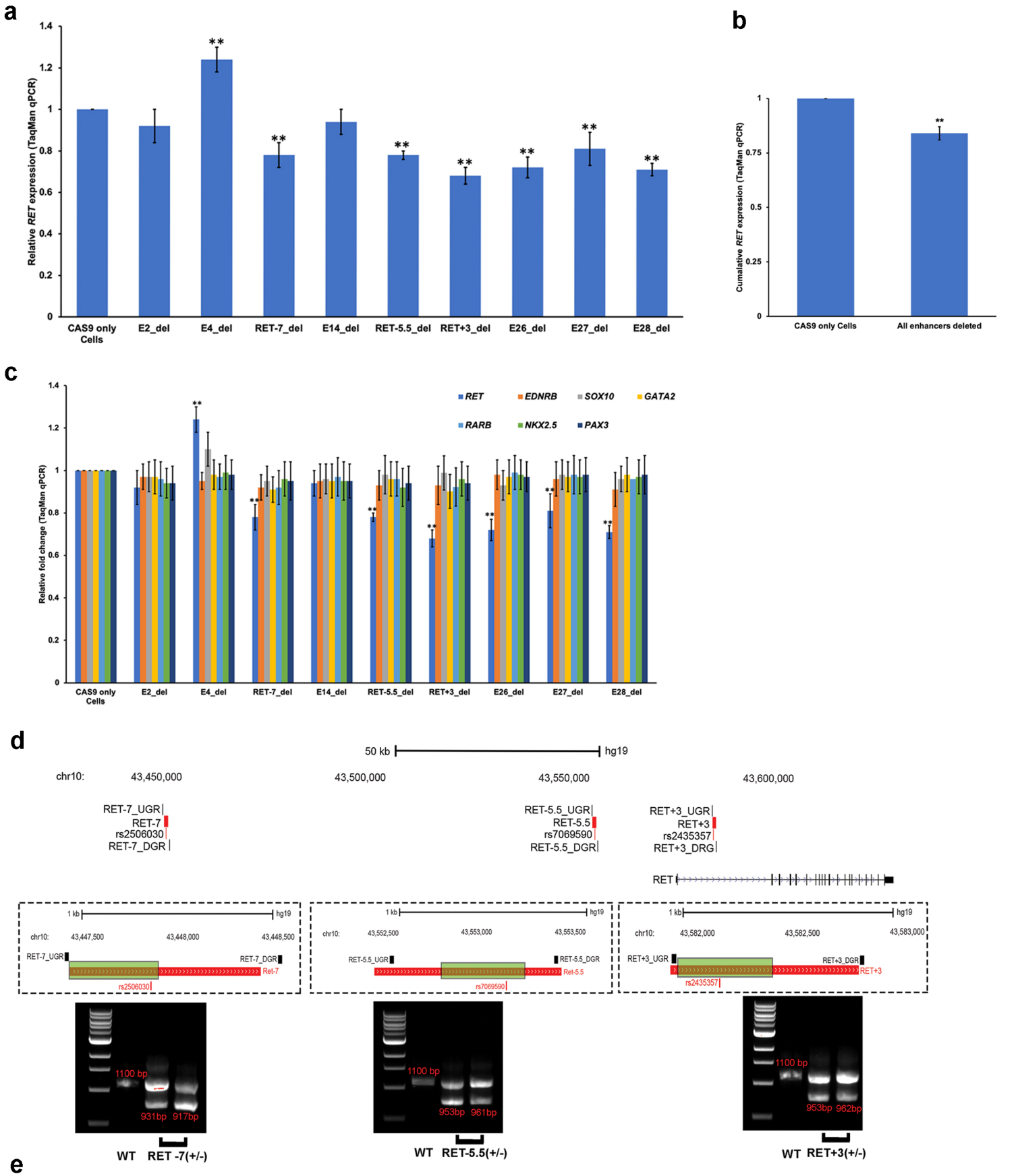


Enhancer E2



Enhancer E27


a
Relative RET Expression (TaqMan qPCR)
1.2
1
0.8
0.6
0.4
0.2
0
**
Control siRNA
PAX3 siRNA
SOX10 siRNA
GATA2 siRNA
RARB siRNA
RET siRNA
b
WT cells
PAX3 siRNA
Relative Expression (TaqMan qPCR)
SOX10
GATA2
RARB


a
Relative RET expression (TaqMan qPCR)
CAS9 only Cells
E2_del
E4_del
RET-7_del
E14_del
RET-5.5_del
RET+3_del
E26_del
E27_del
E28_del
b
Cumulative RET expression (TaqMan qPCR)
CAS9 only Cells
All enhancers deleted
c
RET
EDNRB
SOX10
GATA2
RARB
NKX2.5
PAX3
Relative fold change (TaqMan qPCR)
d
50 kb
hg19
chr10:
43,450,000
43,500,000
43,550,000
43,600,000
RET-7_UGR
RET-7
rs2506030
RET-7_DGR
RET-5.5_UGR
RET-5.5
rs7069590
RET-5.5_DGR
RET+3_UGR
RET+3
rs2435357
RET+3_DRG
RET
1 kb
43,447,500
43,448,000
43,448,500
43,552,500
43,553,000
43,553,500
43,582,000
43,582,500
43,583,000
1100 bp
931bp
917bp
953bp
961bp
953bp
962bp
WT
RET -7(+/-)
RET-5.5(+/-)
RET+3(+/-)
e
Relative Expression (TaqMan qPCR)
WT
RET+3(+/-)
RET-5.5(+/-)
RET-7(+/-)
RET
EDNRB
SOX10
GATA2
RARB
PAX3