

Bimodal sequence composition underlies functional duality of transcribed enhancers

Mario Flores

University of Texas

Ivan Ovcharenko (✉ ovcharen@nih.gov)

National Institutes of Health

Research article

Keywords: studies, advocate, gene regulation

Posted Date: August 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-55967/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Bimodal sequence composition underlies functional duality of transcribed enhancers

Mario Flores and Ivan Ovcharenko*

*Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD.

* correspondence should be addressed to ovcharen@nih.gov.

ABSTRACT

Background:

Recent studies have drawn attention to transcribed enhancers (trEs) as important regulatory elements of gene expression; however, their characteristics and mechanisms of action remain poorly understood.

Results:

We profiled the characteristics of trEs and obtained insights into their mechanisms of action. We found that trEs harbor functional duality related to bimodal sequence composition. TrEs are composed of nonoverlapping cores and flanking regions (flanks): cores function as regular enhancers, while flanks transcribe enhancer RNAs (eRNAs) that can potentially regulate the expression of their target genes in *trans*. Cores are evolutionarily conserved and compact, while flanks are significantly longer. We observed that approximately 25% of eRNAs transcribed from the flanks likely contribute to *trans* DNA:RNA triple helix formation, while another 10% likely employ classical mechanisms of RNA-based regulation. We found that the majority of human enhancers are not transcribed, and trEs are strikingly different from regular enhancers in their functional characteristics. In addition, we found evidence for trEs exhibiting functional duality in regulatory locus encapsulation (RLE), effectively providing localized control over the spread of gene expression upregulation by trE cores and other locus enhancers.

Conclusions:

In summary, our results advocate for enhancer transcription being an uncommon mechanism of gene regulation, and the duality of transcribed enhancer function being a product of additive, not overlapping, DNA sequence encryption.

Introduction

Transcribed enhancers (trEs) are DNA regulatory regions that couple enhancer function with bidirectional transcription of noncoding RNAs (ncRNAs) at their loci. TrEs were first reported several decades ago. Ashe et al. found nascent transcription across the human β -globin gene cluster, including the locus control region (LCR) (1). Plant et al. reported intergenic transcription in the human β -globin gene cluster (2). Masternak et al. observed intergenic transcription at the major histocompatibility complex (MHC) class II locus control region (3). In 2010, genome-wide evidence of transcription at enhancers was presented simultaneously in papers by Kim et al. (4) and de Santa et al. (4, 5). The former study reported that approximately 25% of enhancers are transcribed and named those enhancer transcripts “enhancer RNAs” (eRNAs). Synthesis of eRNAs is initiated near enhancer centers, where the CREB-binding protein (CBP) and RNA polymerase II (RNA Pol II), are bound and proceed bi-directionally extending to the flanks. Changes in eRNA expression levels at enhancers are strongly correlated with changes in mRNA expression at nearby genes (4). At the same time, de Santa and colleagues found that 70% of intergenic RNA Pol II peaks are associated with genomic regions with a canonical chromatin signature of enhancers and that enhancer transcription is regulated in response to endotoxin stimulation. Their data demonstrate that a large proportion of intergenic RNA Pol II transcription sites can be ascribed to cis-regulatory genomic regions (5). Furthermore, de Santa et al. pointed out the possibility that a part of the eRNA transcript may be nonfunctional and may only reflect noise generated by random collisions of the transcriptional machinery with the genome (5).

In 2014, the Functional Annotation of the Mammalian Genome (FANTOM) consortium generated a database of transcribed enhancers in the human genome (6) based on bidirectional transcription at enhancers. They identified 43,011 transcribed enhancers using cap analysis of gene expression (CAGE) (7) and showed that enhancer activity can be detected through the presence of bidirectional capped transcripts. They identified enhancers in 432 primary cells, 135 tissues, and 241 cell line samples from human and found that enhancers share properties with CpG-poor messenger RNA (mRNA) promoters but produce bidirectional and relatively short unspliced RNAs and that the production of eRNA transcripts is related to enhancer activity (6). In 2019, Hirabayashi et al. developed a robust method, called native elongating transcript-CAGE (NET-CAGE), that sensitively detects 5'-ends of nascent RNAs, including unstable transcripts such as enhancer-derived RNAs, and expanded the FANTOM database of transcribed enhancers to approximately 80,000 sequences (8).

Recently, an increasing number of studies have provided evidence of the functionality of trEs. Schaukowitch et al. found that eRNAs facilitate the transition of paused RNA Pol II into productive elongation by acting as decoys for the negative elongation factor (NELF) complex. They also found that the enhancer-promoter interaction was unaffected by eRNA knockdown (9). Allo et al. discovered that 80% of protein argonaute-1 (AGO1) clusters, required for RNA-mediated interference (RNAi), are associated with cell-type-specific trEs, with 73% overlapping active enhancers. They observed that the association is mainly mediated by long intronic eRNAs (10). A different perspective of the function of trEs was introduced by Sigova et al., who proposed that the transcription factor (TF) Ying-Yang 1 (YY1) binds to both gene regulatory elements and their associated RNA species across the genome where eRNAs produce a positive feedback loop that contributes to the stability of gene expression programs (11). In 2017, Cinghu et al. reported that intronic eRNAs interfere and attenuate host gene transcription during productive elongation (12). Later, Jiao et al. identified an eRNA of a super enhancer of the heparanase (HPSE) gene that is

able to bind to heterogeneous nuclear ribonucleoprotein U (hnRNP) to facilitate its interaction with E1A binding protein p300 (p300), which results in chromatin looping between the super enhancer and the HPSE promoter (13). There is also an increasing number of experimentally verified cases of *trans*-regulation by eRNAs. Tsai et al. showed that an enhancer region of the master regulator myogenic differentiation (MyoD) gene on chromosome 7 in mice gives rise to an eRNA that binds a component of a cohesion complex and is recruited to the myogenin locus on chromosome 1 to affect its expression in *trans* (14). They observed that the functional depletion of eRNAs reduces chromatin accessibility, prevents myogenin activation, and hinders muscle differentiation. Another case is presented by Alvares-Dominguez et al., who reported a long ncRNA (lncRNA), Bloodlinc, transcribed from a super enhancer of the erythroid membrane transporter SLC4A1/BAND3, that diffuses to *trans*-chromosomal loci to bind critical regulators of terminal erythropoiesis (15). They found that inhibition of Bloodlinc compromises the terminal erythropoiesis program.

The clinical importance of eRNAs has been under a spotlight in recent years. Studies show that eRNAs participate in several types of cancer, namely, bladder cancer (16, 17), castration-resistant prostate cancer (18), leukemia (19), head and neck cancer (20), basal cell carcinoma (20), colorectal cancer (21), and in cancer invasion and metastasis (13) (22) and p53 tumor suppressor regulation (23). There is also interest in the therapeutic prospects of trEs (24) (25) with some examples of successful attempts to regulate the expression of p53-dependent genes and to modulate the cell cycle with short interfering or silencing RNAs (siRNAs) against eRNA (26).

We developed a computational approach to dissect the characteristics and mechanisms of action of trEs. We found that trEs harbor a bimodal sequence composition defining their functional duality, in which their cores function as regular enhancers while their flanks produce eRNAs that

have a propensity to regulate the expression of their target genes in either *cis* or *trans*. Furthermore, eRNAs are bound by RNA binding proteins (RBPs) and contain DNA-binding domains (DBDs) that can potentially target genes via formation of triple helices with DNA sequences. We describe a case study of a trE within the locus of CD40 that highlights a multi-gene regulatory activity of this regulatory element that consists of an upregulating activity in *cis* and downregulating activity in *trans*.

Results

TrEs harbor enhancer-like epigenetic marks at cores and promoter-like marks at flanks

Using the set of 85,786 trEs, we divided each regulatory element into three segments, named reverse flank, core, and forward flank (Figure 1b; see Methods for a formal definitions and justifications of these regions). Cores of trEs have an average length of approximately 150 bp while flanks are enriched for transcription initiation clusters within 1 kb of the core. Although trEs are bidirectionally transcribed, 89% of trEs feature a predominant transcriptional directionality, which we named the leading flank (See Methods).

An analysis of histone modification profiles of trEs revealed that the cores harbor enhancer-characteristic epigenetic marks and are depleted in promoter-specific H3K4me3 (27). The leading flanks harbor enrichment of H3K4me3 and H3K27ac normally found at active promoters (27) (Figure 2a,b,c,f). The results suggest that trEs have a bimodal composition with cores that harbor enhancer-like epigenetic marks while flanks harbor promoter-like marks.

Two histone marks related to transcriptional activation, H3K9ac and H3K36me3 (28), are enriched at the flanks, but not the cores, and that verifies the transcriptional activity limited to the flanking regions (Figure 2g,k). In addition, we observed an enrichment of H3K9ac at the leading

flank accompanied by an RNA Pol II subunit A (POLR2A) enrichment, a component of the largest subunit of RNA Pol II (Supplementary Figure 2a,b).

We also observed elevated levels of H3K27me3 and H3K9me3 at the flanks (Figure 2i,j) at trEs, which are considered marks of bivalency related to low transcription levels (29). This is in agreement with trEs harboring low levels of transcription compared to promoters where the number of detected gene transcripts is tenfolds larger than the number of detected eRNAs (6).

Since the flanks share epigenetic modifications characteristic of promoters, we then verified that trEs are not promoters of unannotated genes or alternative promoters. To do this, we overlapped trEs with the histone modification H4K20me1 associated with transcriptional activation of genes (30). We found that this mark is depleted at flanks (Figure 2h), suggesting that transcription at intergenic trEs is not associated with transcription of promoters of unannotated genes. To reduce the possibility that trEs were unannotated TSSs of gene isoforms, we removed those in the proximity of the TSSs of genes (± 2 kb).

Overall, we observed that cores of trEs harbor enhancer-like epigenetic marks while flanks harbor promoter-like marks, suggesting distinct regulatory regions with potentially different functionality.

The duality of the regulatory code of the cores and flanks is also reflected in their evolutionary sequence conservation. While both cores and flanks are evolving under stronger negative selection than regular enhancers, the evolutionary conservation significantly spikes at the cores (p-value=1.65e-32, Student's t-test, Figure 3). This phenomenon could be explained if one assumes that the cores have two functions, i.e., enhancer-like regulation of their target genes as well as upregulation of the eRNA transcription from the flanks.

Cores of trEs are enriched in the binding of TFs commonly found at enhancers

We compared the enrichment of 132 TFs binding profiles in trEs and enhancers from the HepG2 cell-line using the ENCODE ChIP-Seq datasets (31, 32) (Figure 4a,b). We observed that trE cores and regular enhancers, but not trE flanks, are enriched in the binding sites of key liver TFs, including: the hepatocyte nuclear factor 4 alpha (HNF4A), a hub TF associated with hepatocellular carcinoma (33) (Figure 4d); the forkhead box A2 (FOXA2), a pioneer TF important for the differentiation of hepatoblasts (34) (Figure 4e); and, the CCAAT/enhancer binding protein (CEBPB), an important regulator of regeneration-associated hepatocyte transcription programs (35) (Figure 4g). However, the spatial binding profiles of these liver TFs are different between regulator enhancers and trE cores. Whereas the TF binding is spread across hundreds of bp of the regular enhancer sequence, it is compactly packed within the sequence of the trE cores (Figure 4a,b).

Several other liver TFs feature similar enhancer- and trE core-specific enrichment in binding lacking at the trE flanks. These include the general enhancer mark EP300 (Figure 4c), and TFs ETS1 (Figure 4f), TCF7 (Figure 4h), JUND (Figure 4i) FOSL2 (Figure 4j), and HNF1A (Figure 4k). Overall, we found that enrichment of the liver-specific TFs at the trE cores is similar to regular enhancers, thus providing additional evidence that the cores of trEs function as enhancers.

The analysis of computationally predicted TFBSs (see Methods) revealed a trend similar to the TF ChIP-seq profiles and also demonstrated that for some TFs binding is stronger at the trE cores as compared with the regular enhancers (Supplementary Figure 3).

TrEs and enhancers show similar interactions in 3D space

Chromatin looping brings enhancers in proximity to their target promoters, and the 3D structure of the chromatin is commonly characterized using Hi-C (36). We found no significant difference

in the number of chromatin contacts originating from trEs and regular enhancers compared to background that consisted of random regions of the human genome matched by length and GC content (Figure 5a), suggesting that trEs contact their target genes as often as regular enhancers and their function is not limited to eRNA production and eRNA-based regulation. The Hi-C results are supported by similar profiles of chromatin remodeling protein binding signals at regular enhancers and trEs (Figure 5b).

TrE flanks produce eRNAs bound by RBPs

In recent years experiments have shown that eRNAs mediate recruitment of RBPs to regulate transcription of their target genes (10, 14, 15, 37-39). We observed that RBPs binding is enriched at the flanks of trEs (Figure 6a,b). For example, we found that Protein Argonaute-2 (AGO2)(Figure 6c), Y Box binding protein 1 (YBX1)(Figure 6d) and Non-POU Domain Containing Octamer Binding (NONO) (Figure 6e) have enrichment of ChIP-Seq signals at flanks but not cores of trEs.

To gain additional evidence of the RBP binding to trE eRNAs, we computed the enrichment of RBP narrow peaks in HepG2 using 83 available enhanced crosslinking and immunoprecipitation (eCLIP) datasets from the ENCODE project (eCLIP is a recently developed technology that improves specificity in discovery of authentic binding sites (40)). We found that 43% of trEs feature transcription of eRNAs with significant binding of at least one in 83 RBPs (3 times more compared to background, $p\text{-value}=4.09\text{e-}78$, Fisher's exact test)(see Methods). We also observed that the binding is significantly localized to trE flanks but not to cores ($p\text{-value}=0.004$, Fisher's exact test). These results suggest widespread binding of RBPs to eRNAs.

***Trans-acting* trE eRNAs**

Previous research has shown experimental cases of gene regulation by eRNAs-RBPs, associations that can be categorized as *trans*-mode eRNAs (14, 15). The mechanism of action of

trans-mode eRNAs consist in the regulation of transcription of target genes located on different chromosomes relative to the chromosome harboring the locus for the transcribed eRNAs. To study *trans*-mode eRNAs, we first annotated eRNAs using several ncRNA databases.

Using Gencode v28 (41), we found that 1.4% (108/7613) of trEs have pseudogenes transcribed from their flanks in the HepG2 cell line (6.8 times more than expected by chance; p -value=4.7e-33, Fisher's exact test). In contrast, there were only 0.1% (39/30080) of HepG2 regular enhancers that feature transcription of pseudogenes. We also found that the preferential directionality of trEs agreed with the direction of transcription of the overlapping pseudogenes in 64% of cases compared to 35% in enhancers (p -value=0.002, Fisher's exact test). Although the enrichment of pseudogenes at the trE flanks was significant, the very small overall fraction of trEs containing pseudogenes advocates for largely non-genic transcription originating from the trE flanks.

Using RNA families (Rfam) (42) and microRNA database (miRbase) (43) Gencode data, we found that only 0.2% of trEs and 0.04% of regular HepG2 enhancers contain micro RNAs (miRNAs). By overlapping the trE flanks with known lncRNA loci, we found that ~3.9% (296/7613) of trEs and ~1.2% (376/30080) of regular enhancers harbor lncRNAs transcription. We also found that 0.3% of trEs harbor transcription of small nucleolar RNAs (snoRNAs), ribosomal RNAs (rRNAs), and small nuclear RNAs (snRNAs) (see Table 1).

Another potential mechanism of regulation by *trans*-mode eRNAs consists in the binding of antisense eRNAs (asRNAs) to their target mRNAs to block translation. We found that ~5% (380/7613) of trEs harbor eRNAs that likely function as asRNAs. Previous research reports an antisense ncRNA that represses transcription of the lymphoid enhancer-binding factor 1 (LEF1) gene by recruiting Polycomb repressive complex 2 (PRC2), which trimethylates H3K27 (44). We observed that ~75% (285/380) of these eRNAs are associated with mRNAs transcribed from loci

in a different chromosome (in *trans*) while ~25% associated with mRNAs transcribed from loci in the same chromosome (in *cis*) (p-value=1.8e-11), suggesting the potential of eRNAs to regulate transcription of their target genes in *trans*.

Next, we assessed the possibility that trans-mode eRNAs might regulate the transcription of target genes by direct binding to DNA to form triplex target DNA sites (TTSSs) (Supplementary Figure 4). We found that eRNAs are enriched for DNA binding domains (2.5-fold compared with background, p-value<2.34e-11, Fisher exact test) and up to 25% of eRNAs could be attributed to regulation through RNA:DNA triple helical structure formation.

In summary, we were able to identify some evidence of *trans* regulatory activity associated with trE flanks, but there is no single mode of *trans* action that could be attributed to the majority of trE eRNAs given currently available data.

Flanks of trEs are enriched in *trans*-eQTLs and the cores of trEs are enriched in *cis*-eQTLs

We hypothesized that if trans-mode eRNAs are functional and, indeed, act in *trans* to regulate their target genes, we should observe an enrichment of *trans*-eQTLs at the flanks of trEs. We found a significant 2.5-fold enrichment of *trans*-eQTLs at the trE flanks (p-value<2.77e-6, Fisher exact test; see Methods). We also observed a significantly higher enrichment of *trans*-eQTLs at the trE flanks as compared with the trE cores (Figure 7a). This difference between the eQTL density at the trE flanks and cores grows with the increase in the distance to the target gene (Figure 7b). For example, the first row corresponds to the density of eQTLs that are associated with genes located 200,000 bp or more (p-values: *<0.001 **<0.0001 Fisher's exact test). As the distance to the target gene rises to 800,000 bp, there are virtually no trE core eQTLs, and almost all eQTLs are at the trE flanks, which indicates that the true *trans* regulatory activity of trE is almost exclusively associated with the trE flanks. Finally, the directionality of the *trans* regulatory eQTL

effect indicates that the target genes of trE flank eRNAs are predominantly downregulated (Figure 7c).

The eQTL enrichment is different for *cis* eQTLs. At the <10 kb distance to the target gene, there are slightly more core than flank trE eQTLs, and the regulatory directionality corresponds to upregulation of gene expression. This indicates that trEs behave as classical enhancers at the shorter distance and that the trE flanks might be contributing to enhancer formation despite the lack of active TF binding to these outer parts of trEs.

TrE functional duality in regulatory locus encapsulation

Our results indicate that the core and flanking parts of a trE likely regulate different genes, i.e., a nearby gene is regulated by the trE core and a distant gene is regulated by an eRNA transcribed from one of the trE flanks, possibly in *trans*. As the eRNA regulatory effect is likely opposite in directionality to the upregulation by a trE core enhancer, we considered a trE regulatory mechanism, in which the trE core enhancer upregulates a small number of nearby genes while its eRNAs are used to limit the propagation of the upregulation to the surrounding genes. We named this mechanism regulatory locus encapsulation (RLE). Using thyroid eQTL data, we found evidence of an RLE in the locus of the Coiled-Coil Domain Containing 57 (CCDC57) gene (Figure 8). CCDC57 is surrounded by the Fatty Acid Synthase (FASN) gene downstream and the Solute Carrier Family 16 Member 3 (SLC16A3) gene upstream. While CCDC57 and SLC16A3 are highly expressed in thyroid and pancreas, the expression of FASN is almost absent in these two tissues (Figure 8a,b,c). FASN, which plays critical role in *de novo* synthesis of fatty acids, is critically important for proper liver function and is highly expressed in liver (45) (Figure 8a). However, FASN is linked to dysregulation of thyroid and is strongly associated with hypothyroidism (45) at the same time. In particular, FASN overexpression is a marker of papillary thyroid carcinoma (PTC) (46), and inhibition of FASN in PTC has been proposed as a clinical target for this cancer (46).

We found two eQTL SNPs in the core and a flank of the CCDC57 trE. The flanking eQTL SNP (rs7406326) is linked to the more distal FASN located 102 kb downstream. Its G>A mutation is associated with a significant 5.2-fold increase in the expression of FASN (Figure 8d), indicating that the native trE eRNA had been silencing FASN expression. The core eQTL SNP (rs8079572) is linked to SLC16A3 located 27 kb upstream, and its G>T mutation is associated with a decrease in expression of SCL16A3 (Figure 8e), which is consistent with an upregulatory enhancer function of this trE. In summary, these results indicate that the CCDC57 trE upregulates gene expression of two proximal genes, i.e., CCD57 and SLC16A3, in thyroid and pancreas, whereas the CCDC57 trE eRNA silences the expression of the flanking, but distal, FASN in these two tissues and that this RLE is critical for proper thyroid function.

Discussion

An increasing number of reports indicate a significant role for eRNAs in gene expression regulation in healthy and diseased states and suggest their clinical utility in eRNA-targeted therapy (16, 17, 20-22, 24, 47-49). However, trE characterization (50-52) and the mechanisms of action of trE eRNAs (53, 54) remain unclear. Here we developed an integrative approach to dissect the characteristics of trEs and investigated the duality of function of these regulatory elements.

We observed that trEs harbor a bimodal sequence composition, which is distinct from regular enhancers. TrEs consist of three distinct components: a central core and two flanks. The core shares functional characteristics and the TF binding capability of a regular enhancer, while the flanks are selectively used for eRNA transcription. Cores are much more compact than regular enhancers and evolve under stronger negative selection than regular enhancers.

Several groups have argued that eRNA production may be an incidental byproduct of a regulatory mechanism where the act of transcription, and not the transcripts, is the important regulatory function (55). To evaluate this hypothesis, we first tested if the direction of transcription of trEs agrees with the direction of transcription of their target genes. We observed that intronic trEs, but not intergenic trEs, follow the direction of transcription of their target genes in 60% of cases (Supplementary Figure 1a). This suggests that transcription of eRNAs at intronic trEs may be the result of gene transcription promoting inner eRNA transcription. Second, Cinghu et al. reported that transcription at intragenic enhancers interferes with, and attenuates, host gene transcription during productive elongation (12). We then tested this scenario by comparing the expression of target genes with intronic trEs in two groups. The first group consisted of intronic trEs with opposite directionality to that of their host genes, while the second group consisted of those trEs with the same directionality as that of their host genes (Supplementary Figure 1b,c). We found no significant difference in gene expression suggesting that the previously proposed regulatory mechanism may not have a genome-wide effect.

We observed that binding of eRNAs by RBPs is widespread and tissue-specific, which agrees with previous reports of extensive interaction of eRNAs with RBPs (9,68). Based on the spatial interaction of eRNAs with RBPs and their target genes, they can be categorized as either *cis*- or *trans*-mode eRNAs. *Cis*-mode eRNAs act in the proximity of their target promoters. For example, we observed that an eRNA can potentially regulate PRC2 occupancy at a promoter through the competitive binding of its core component RBP SUZ12 (Supplementary Figure 5). However, this regulation is only possible when gene transcription is comparable to trE transcription. In this scenario an increase of transcription at the trE can potentially prevent SUZ12 from forming the repressive complex and can allow an increase of gene transcription (Supplementary Figure 5c). If the model is correct, we should also observe a positive correlation between transcription at the trE and transcription at the gene. Furthermore, the correlation should

increase for higher values of SUZ12 at the trE as measured by ChIP-Seq (Supplementary Figure 5d). There are similar examples of this mode of action cited above that show that eRNAs facilitate the transition of paused RNA Pol II into productive elongation by acting as a decoy for the negative elongation factor (NELF) complex (9). Also, Sigova et al. reported that some DNA binding TFs also bind RNA generated at promoters and enhancers, which suggests a modest but important effect that contributes to the stability of gene expression programs (11). However, one limitation of this mode of action is the finding that co-transcription of eRNAs and mRNAs rarely occurs within closed enhancer–promoter loops (54), which also suggests, again, that this regulatory mechanism may not have a genome-wide effect.

On the other hand, the *trans*-mode eRNA mechanism commonly consists of the binding of RBPs to eRNAs in an effort to regulate their target genes in *trans*. We found that about 5% of flanks can transcribe eRNAs previously categorized as lncRNAs or pseudogenes and another 5% of flanks transcribe eRNAs that can be potentially bind asRNAs to target mRNAs. Previous reports have shown, for example, that an antisense ncRNA represses transcription of the LEF1 gene by recruiting PRC2. However, since asRNAs consist of repetitive elements and the genome is full of repetitive sequences, it is difficult to assess the true gene targets of eRNAs that function as asRNAs.

We also observed that eRNAs can potentially identify their target genes by RNA-DNA triplex interactions to regulate their transcription in *trans*. The implications of this *trans*-mode of action is that a subset of eRNAs harbors two types of sites. One is the site for binding RBPs, while the second consists of a DBD that potentially allows the eRNA to identify its target gene in *trans*. We found that, overall, approximately 25% of eRNAs can potentially interact with genes in a tissue through significant RNA-DNA triplexes, suggesting that triplex formation may be an important mechanism of eRNA regulation.

As additional evidence that eRNAs function to regulate target genes in *trans*, we found that flanks are enriched for *trans* eQTLs compared to cores and regular enhancers. We also observed that the primary effect of *trans* regulation by eRNAs is downregulation of target genes. However, we have to consider that the number of *trans* eQTLs that overlap flanks of trEs is very limited (<1% of trEs in B-cells), although this may just reflect the finite size of currently available data sets of *trans* eQTLs.

In addition to the classical *trans* regulatory activity of trE eRNA, we observed evidence of trEs contributing to RLE, where the role of a trE eRNA is to effectively shield the bystander genes surrounding the genes regulated by the trE core from the propagation of potentially undesirable upregulatory activity of this core.

Taken together, our results show that trEs play an important role in the precise orchestration of gene regulation in the human genome, accurately balancing the positive regulatory activity of trE cores with the negative regulatory activity of eRNAs transcribed from trE flanks.

Methods

Enhancers and trEs

To define enhancers, we used the histone modifications H3K27ac and H3K4me1 coupled with DNase I hypersensitivity (DHS) data downloaded from the Encyclopedia of DNA elements (ENCODE) (31) (Figure 1a). We defined regular enhancers in a tissue as 400 bp-regions centered on the intersection of the narrow peaks of H3K27ac and H3K4me1 that overlap by at least 20 bp with a DHS region of open chromatin. In total, we identified 30,080 HepG2 and 41,855 GM12878 enhancers.

For transcribed enhancers (trEs), we used 85,786 predicted trEs identified by bidirectional transcription using NET-CAGE across a panel of tissues (8), a robust method that sensitively detects 5'-ends of nascent RNAs, coupled with the FANTOM5 atlas of transcribed enhancers (6). We obtained 7,613 HepG2 trEs, 4,777 thyroid trEs, and 6,686 GM12878 trEs.

Segmentation, binning, and orientation of trEs

Based on transcription properties, we divided each trE into three segments: a reverse flank, a core, and a forward flank (Figure 1b). The segment locations are based on the positions of the leftmost/rightmost clusters of CAGE tags for reverse and forward flanks, respectively.

Next, since flanks are enriched for transcription initiation clusters within 1 kb of the core, we extended the flanks to 1,000 bp and binned them into 20 bins of 50 bp each (Figure 1c). Cores had an average length of 150 bp and were divided into 10 bins. Next, we oriented each trE so that the strongest transcription (the leading flank) was on the right side (Figure 1d). For this, we first assigned preferential directionality of a trE by comparing the number of CAGE tags (7) in the two flanks. Second, we reoriented all reverse directionality cases so that the strongest transcription pointed to the right. Using a 1.5x-fold difference in the number of CAGE tags between two flanks as a threshold of strong directionality, we observed that 89% of trEs have strong preferential directionality and 11% of trEs are bidirectional.

Chromatin immunoprecipitation sequencing (ChIP-seq) signals (histone marks, TFs) of trEs, enhancers, and promoters

To compare ChIP-seq and RNA-seq signals of trEs and enhancers, we calculated the average signal value S at each bin I (S_i) with respect to the background. The background consisted of regions of open chromatin in the tissue/cell line of study as marked by DHSs. The statistical

significances of differences in signal vs. background, trEs vs. enhancers, and cores vs. flanks signal were calculated using the Student's t-test (56).

Hierarchically-clustered heatmaps were generated using the Seaborn visualization library based on matplotlib (57).

Sequence conservation of trEs, enhancers and promoters

To evaluate sequence conservation of trEs, we used phylogenetic p-values for multiple alignments of 100 vertebrate genomes (58, 59). For this, we took advantage of the PhyloP scores available from the University of California, Santa Cruz (UCSC) Genome Browser (60) and computed the average value per bin with respect to the background. The background consisted of open chromatin regions marked by DHSs (length matched).

RBP enrichment in trEs and eRNAs

To calculate the enrichment of RBPs in flanks of transcribed enhancers, we calculated the average signal value S at each bin I (S_i), with respect to the background. RBP signals were obtained from the 32 ENCODE ChIP-Seq datasets (31, 32). Background consisted of open chromatin regions (length-matched) as marked by DHSs. Supplementary Table 1 shows the results for three RBPs: AGO2, non-POU domain-containing octamer-binding protein (NONO), and RNA binding motif protein 9 (RBFox2). To calculate enrichment of RBPs at trEs' flanks, we overlap trEs with sets of narrow peaks from ENCODE enhanced crosslinking and immunoprecipitation (eCLIP) datasets (31, 32) and compare to cores. eCLIP is a technique that allows robust transcriptome-wide discovery of RNA-binding protein binding sites and enable large scale and robust profiling, with amplification and sample requirements similar to those of ChIP-Seq (40).

TFBSs at trEs

To evaluate the enrichment of TFBSs at trEs and to compare them with enhancers, we used the Multiple Em for Motif Elicitation (MEME) library of TFBSs (61). We mapped TFBSs within trEs and enhancers and calculated the TFBS-enrichment p -values using the Poisson distribution with Bonferroni correction for multiple testing. The controls consisted of random regions matched for length and repeat content.

Gene transcription and eRNA directionality

To check if there is a trend of eRNA orientation “towards” the nearest transcription start site (TSS), we first calculated the directionality score (D) for each trE

$$D=(n_F-n_R)/(n_F+n_R)$$

where n_R is the number of tags overlapping the reverse flank and n_F is the number of tags overlapping the forward flank. Next, we associated each trE with its nearest gene and determined if the gene orientation was the same or opposite to D (Supplementary Figure 1a). We then compared the level of expression of the genes hosting intronic trEs that either match or oppose the transcriptional directionality of their host genes. A scatter plot of directionality vs. gene expression was generated using matplotlib (57).

eRNA-DNA triplex formation

To investigate the formation of eRNA-DNA triplexes, we first selected the GM12878 trEs that have a *trans* expression quantitative trait locus (eQTL) association. For this, we used a set of ~60,000 significant *trans*-eQTL associations (FDR<0.05) from the eQTLGen Consortium (62). Next, we used the Triplex Domain Finder (TDF) within the Regulatory Genomics Toolbox (RGT) (63). TDF

characterizes the triplex-forming potential between RNA and DNA regions (63). This method tests whether a particular eRNA region is likely to form a DBD using its RNA sequence homology to the potential target DNA regions (± 100 kb from center of gene body).

Analysis of *trans*-eQTLs at trEs

To evaluate the enrichment of *trans*-eQTLs at the flanks of trEs compared to the cores and to regular enhancers, we first downloaded the significant *trans*-eQTLs from the eQTLGen Consortium (62). After that, we calculated the density of *trans*-eQTLs at the cores and flanks with respect to the background, which consisted of random DNA regions matched by length and GC content.

References

1. Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.* 1997;11(19):2494-509.
2. Plant KE, Routledge SJ, Proudfoot NJ. Intergenic transcription in the human beta-globin gene cluster. *Mol Cell Biol.* 2001;21(19):6507-14.
3. Masternak K, Peyraud N, Krawczyk M, Barras E, Reith W. Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol.* 2003;4(2):132-7.
4. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature.* 2010;465(7295):182-U65.
5. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* 2010;8(5):e1000384.
6. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;507(7493):455-61.

7. Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A*. 2003;100(26):15776-81.
8. Hirabayashi S, Bhagat S, Matsuki Y, Takegami Y, Uehata T, Kanemaru A, et al. NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. *Nat Genet*. 2019;51(9):1369-79.
9. Schaukowitch K, Joo JY, Liu X, Watts JK, Martinez C, Kim TK. Enhancer RNA facilitates NELF release from immediate early genes. *Mol Cell*. 2014;56(1):29-42.
10. Allo M, Agirre E, Bessonov S, Bertucci P, Gomez Acuna L, Buggiano V, et al. Argonaute-1 binds transcriptional enhancers and controls constitutive and alternative splicing in human cells. *Proc Natl Acad Sci U S A*. 2014;111(44):15622-9.
11. Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, et al. Transcription factor trapping by RNA in gene regulatory elements. *Science*. 2015;350(6263):978-81.
12. Cinghu S, Yang P, Kosak JP, Conway AE, Kumar D, Oldfield AJ, et al. Intragenic Enhancers Attenuate Host Gene Expression. *Mol Cell*. 2017;68(1):104-17 e6.
13. Jiao W, Chen Y, Song H, Li D, Mei H, Yang F, et al. HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNPU/p300/EGR1/HPSE axis. *Oncogene*. 2018;37(20):2728-45.
14. Tsai PF, Dell'Orso S, Rodriguez J, Vivanco KO, Ko KD, Jiang K, et al. A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans. *Mol Cell*. 2018;71(1):129-41 e8.

15. Alvarez-Dominguez JR, Knoll M, Gromatzky AA, Lodish HF. The Super-Enhancer-Derived alncRNA-EC7/Bloodlinc Potentiates Red Blood Cell Development in trans. *Cell Rep*. 2017;19(12):2503-14.
16. Liu Y, Ding M, Liao X, Gao Q, He A, Liu B, et al. High expression of enhancer RNA MARC1 or its activation by DHT is associated with the malignant behavior in bladder cancer. *Exp Cell Res*. 2018;370(2):303-11.
17. Ding M, Zhan H, Liao X, Li A, Zhong Y, Gao Q, et al. Enhancer RNA - P2RY2e induced by estrogen promotes malignant behaviors of bladder cancer. *Int J Biol Sci*. 2018;14(10):1268-76.
18. Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*. 2011;474(7351):390-4.
19. Yu M, Al-Dallal S, Al-Haj L, Panjwani S, McCartney AS, Edwards SM, et al. Transcriptional regulation of the proto-oncogene Zfp521 by SPI1 (PU.1) and HOXC13. *Genesis*. 2016;54(10):519-33.
20. Irimie AI, Zimta AA, Ciocan C, Mehterov N, Dudea D, Braicu C, et al. The Unforeseen Non-Coding RNAs in Head and Neck Cancer. *Genes (Basel)*. 2018;9(3).
21. McClelland ML, Mesh K, Lorenzana E, Chopra VS, Segal E, Watanabe C, et al. CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. *J Clin Invest*. 2016;126(2):639-52.
22. Liu Y, Ding M, Gao Q, He A, Liu Y, Mei H. Current Advances on the Important Roles of Enhancer RNAs in Gene Regulation and Cancer. *Biomed Res Int*. 2018;2018:2405351.

23. Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Vrielink JAFO, et al. eRNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. *Mol Cell*. 2013;49(3):524-35.
24. Kashkin KN, Sverdlov ED. Properties, functions, and therapeutic prospects of enhancer RNAs. *Russ J Bioorg Chem*. 2016;42(5):473-8.
25. Leveille N, Melo CA, Agami R. Enhancer-associated RNAs as therapeutic targets. *Expert Opin Biol Ther*. 2015;15(5):723-34.
26. Leveille N, Melo CA, Rooijers K, Diaz-Lagares A, Melo SA, Korkmaz G, et al. Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. *Nat Commun*. 2015;6:6520.
27. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009;459(7243):108-12.
28. Xu YM, Du JY, Lau AT. Posttranslational modifications of human histone H3: an update. *Proteomics*. 2014;14(17-18):2047-60.
29. Matsumura Y, Nakaki R, Inagaki T, Yoshida A, Kano Y, Kimura H, et al. H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. *Mol Cell*. 2015;60(4):584-96.
30. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*. 2008;40(7):897-903.

31. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.
32. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46(D1):D794-D801.
33. Yan H, Wang Q, Shen Q, Li Z, Tian J, Jiang Q, et al. Identification of potential transcription factors, long noncoding RNAs, and microRNAs associated with hepatocellular carcinoma. *J Cancer Res Ther*. 2018;14(Supplement):S622-S7.
34. Yamamura N, Fugo K, Kishimoto T. Forkhead box protein A2, a pioneer factor for hepatogenesis, is involved in the expression of hepatic phenotype of alpha-fetoprotein-producing adenocarcinoma. *Pathol Res Pract*. 2017;213(9):1082-8.
35. Kuttippurathu L, Patra B, Cook D, Hoek JB, Vadigepalli R. Pattern analysis uncovers a chronic ethanol-induced disruption of the switch-like dynamics of C/EBP-beta and C/EBP-alpha genome-wide binding during liver regeneration. *Physiol Genomics*. 2017;49(1):11-26.
36. Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell*. 2016;62(5):668-80.
37. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol*. 2018;19(5):327-41.
38. Achsel T, Bagni C. Cooperativity in RNA-protein interactions: the complex is more than the sum of its partners. *Curr Opin Neurobiol*. 2016;39:146-51.
39. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. *Science*. 1969;165(3891):349-57.

40. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*. 2016;13(6):508-14.
41. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012;22(9):1760-74.
42. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res*. 2018;46(D1):D335-D42.
43. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 2011;39(Database issue):D152-7.
44. Beltran M, Garcia de Herreros A. Antisense non-coding RNAs and regulation of gene transcription. *Transcription*. 2016;7(2):39-43.
45. Sun J, Sun L, Chen W, Yin X, Lu Y, Jiang Q. A family with hypothyroidism caused by fatty acid synthase and apolipoprotein B receptor mutations. *Mol Med Rep*. 2018;18(6):4904-12.
46. Uddin S, Siraj AK, Al-Rasheed M, Ahmed M, Bu R, Myers JN, et al. Fatty acid synthase and AKT pathway signaling in a subset of papillary thyroid cancers. *J Clin Endocrinol Metab*. 2008;93(10):4088-97.
47. Ko JY, Oh S, Yoo KH. Functional Enhancers As Master Regulators of Tissue-Specific Gene Regulation and Cancer Development. *Mol Cells*. 2017;40(3):169-77.

48. Bal E, Park HS, Belaid-Choucair Z, Kayserili H, Naville M, Madrange M, et al. Mutations in ACTRT1 and its enhancer RNA elements lead to aberrant activation of Hedgehog signaling in inherited and sporadic basal cell carcinomas. *Nat Med.* 2017;23(10):1226-33.
49. Zhang Z, Lee JH, Ruan H, Ye Y, Krakowiak J, Hu Q, et al. Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat Commun.* 2019;10(1):4562.
50. Young RS, Kumar Y, Bickmore WA, Taylor MS. Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome Biol.* 2017;18(1):242.
51. Sharifi-Zarchi A, Gerovska D, Adachi K, Totonchi M, Pezeshk H, Taft RJ, et al. DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC Genomics.* 2017;18(1):964.
52. Espinosa JM. Revisiting lncRNAs: How Do You Know Yours Is Not an eRNA? *Mol Cell.* 2016;62(1):1-2.
53. Skalska L, Beltran-Nebot M, Ule J, Jenner RG. Regulatory feedback from nascent RNA to chromatin and transcription. *Nat Rev Mol Cell Biol.* 2017;18(5):331-7.
54. Rahman S, Zorca CE, Traboulsi T, Noutahi E, Krause MR, Mader S, et al. Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription. *Nucleic Acids Res.* 2017;45(6):3017-30.
55. Lam MT, Li W, Rosenfeld MG, Glass CK. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci.* 2014;39(4):170-82.
56. Gauvreau K, Pagano M. Student's t test. *Nutrition.* 1993;9(4):386.
57. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng.* 2007;9(3):90-5.

58. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 2010;20(1):110-21.
59. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 2005;15(8):1034-50.
60. Karolchik D, Hinrichs AS, Kent WJ. The UCSC Genome Browser. *Curr Protoc Bioinformatics.* 2012;Chapter 1:Unit1 4.
61. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics.* 2011;27(12):1696-7.
62. Võsa U, Claringbould, A. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv.* 2018.
63. Kuo CC, Hanzelmann S, Senturk Cetin N, Frank S, Zajzon B, Derks JP, et al. Detection of RNA-DNA binding sites in long noncoding RNAs. *Nucleic Acids Res.* 2019;47(6):e32.

Declarations

Not applicable

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Data supporting the results reported in the article can be found including, where applicable, hyperlinks to publicly archived datasets analyzed or generated during the study

Competing interest

None declared

Funding

This research was supported by the Intramural Research Program of the National Institutes of Health; National Library of Medicine

Authors' contributions

IO conceived and designed the study. MF performed data analyses. MF and IO wrote the manuscript. All authors read and approved the final manuscript.

Corresponding author

Ivan Ovcharenko

Acknowledgments

The authors are grateful to Dorothy L. Buchhagen for critical reading of the manuscript.

SUPPLEMENTARY DATA

Supplementary Data provided as a separate file “supplementary.pdf”.

TABLES

Table 1. Categories of known RNAs transcribed at flanks of trES

RNA class	% originated from trES
Pseudogenes	1.7%
miRNAs	0.2%
lncRNAs	4%

asRNAs	5%
Triplex eRNAs	25%
Other RNAs*	0.3%

* small nucleolar RNAs (snoRNAs), ribosomal RNAs (rRNAs), and small nuclear RNAs (snRNAs)

FIGURES

Figure 1. Definition of transcribed enhancers. (a) Procedures for the prediction of trEs and enhancers. (b) Segments of a trE based on transcription directionality: reverse flank, core, and forward flank. Blue indicates CAGE tags with reverse directionality, orange indicates CAGE tags with forward directionality. (c) Binning of segments of trEs. (d) Plot of the density of CAGE tags per bin on trEs after re-orienting each trE so that the strongest transcription (leading flank) is on the right side (downstream direction).

Figure 2. Composition of histone modifications for trEs and enhancers in HepG2. (a) Heat map of fold-change (respect to background) for nine histone modifications in trEs. (b) Heat map in enhancers. (c) Plot of fold-change for H3K27ac over the loci of trEs (red) and enhancers (blue). (d) Plot for H3K4me1, (e) H3K4me2, (f) H3K4me3, (g) H3K9ac, (h) H4K20me1, (i) H3K27me3, (j) H3K9me3, (k) H3K36me3.

Figure 3. Conservation of trEs. PhyloP fold-change per bin with respect to background.

Figure 4. TF enrichment at trEs and enhancers in HepG2. (a) Heatmap for 12 liver TFs at trEs. Scale indicates the fold-change with respect to background. (b) Heatmap for enhancers. On right, detailed binding profiles of individual TFs across trEs, regular enhancers, and promoters. (c) EP300. (d) HNF4A. (e) FOXA2. (f) ETS1. (g) CEBPB. (h) TCF7. (i) JUN D. (j) FOSL2. (k) HNF1A.

Figure 5. Chromatin loops originating from trEs and regular enhancers. (a) Number of Hi-C contacts at trEs and enhancers. (b) ChIP-seq signals of CTCF, RAD21, and SMC3 at trEs. Red plot: trEs; blue plot: enhancers.

Figure 6. RNA binding proteins (RBPs) binding at trEs and enhancers in HepG2. (a) Heatmap of ChIP-Seq signals (fold change) for RBPs at trEs, (b) ChIP-Seq signals of RBPs at enhancers. (c) AGO2. (d) YBX1. (e) NONO.

Figure 7. Enrichment of eQTLs at trEs. (a) Density of eQTLs (per 10 Kb) at the cores and flanks of thyroid trEs. (b) eQTL density as a function of distance D to the associated gene. (c) Percentage of eQTLs associated with negative ($S < 0$) or positive ($S > 0$) slope for distances D.

Figure 8. The regulatory landscape of the CCDC57 locus. Two eQTLs within the CCDC57 intronic trE are associated with the two genes, FASN downstream and SLC16A3 upstream, flanking CCDC57. Red: trEs is active in tissue; grey trEs is not active.

Figures

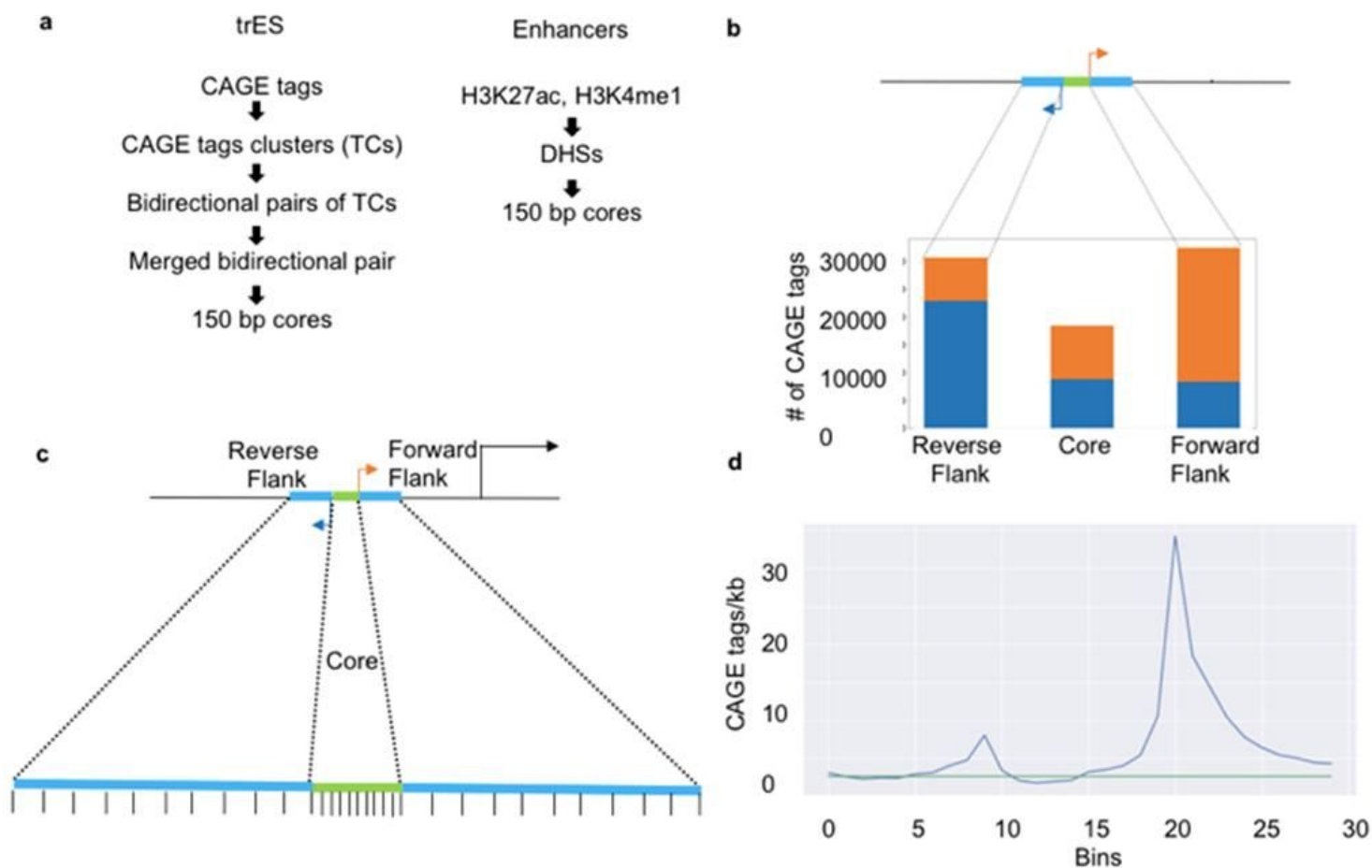


Figure 1

Definition of transcribed enhancers. (a) Procedures for the prediction of trEs and enhancers. (b) Segments of a trE based on transcription directionality: reverse flank, core, and forward flank. Blue indicates CAGE tags with reverse directionality, orange indicates CAGE tags with forward directionality. (c) Binning of segments of trEs. (d) Plot of the density of CAGE tags per bin on trEs after re-orienting each trE so that the strongest transcription (leading flank) is on the right side (downstream direction).

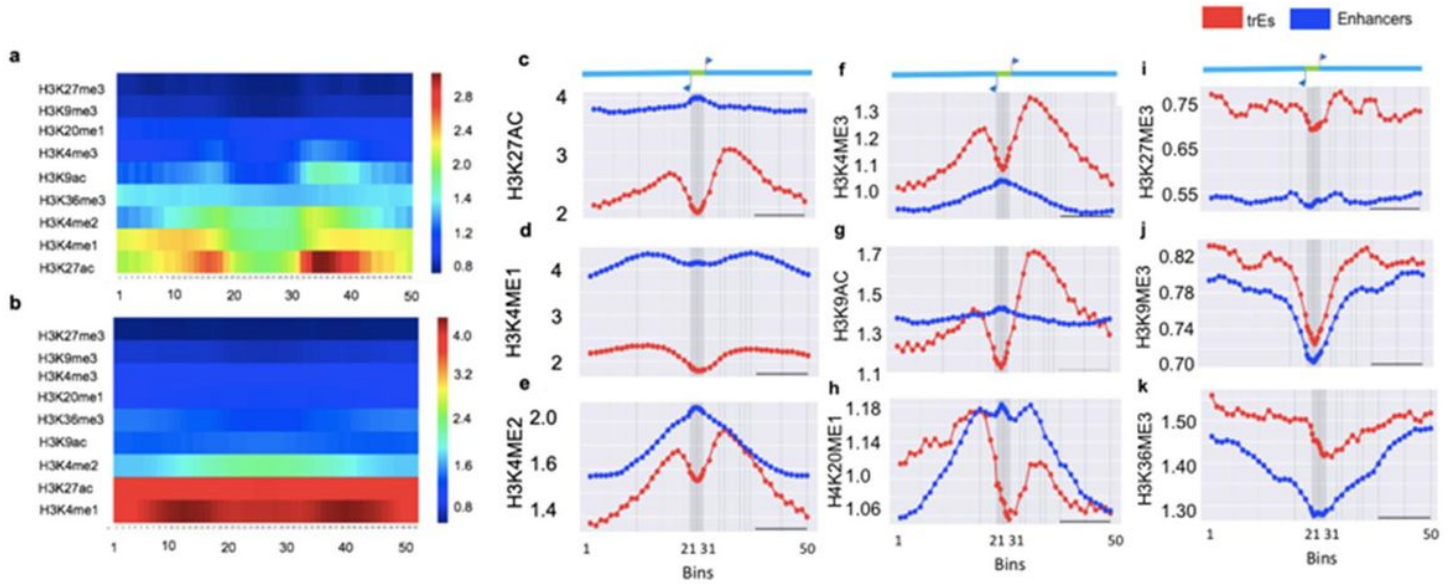


Figure 2

Composition of histone modifications for trEs and enhancers in HepG2. (a) Heat map of fold-change (respect to background) for nine histone modifications in trEs. (b) Heat map in enhancers. (c) Plot of fold-change for H3K27ac over the loci of trEs (red) and enhancers (blue). (d) Plot for H3K4me1, (e) H3K4me2, (f) H3K4me3, (g) H3K9ac, (h) H4K20me1, (i) H3K27me3, (j) H3K9me3, (k) H3K36me3.

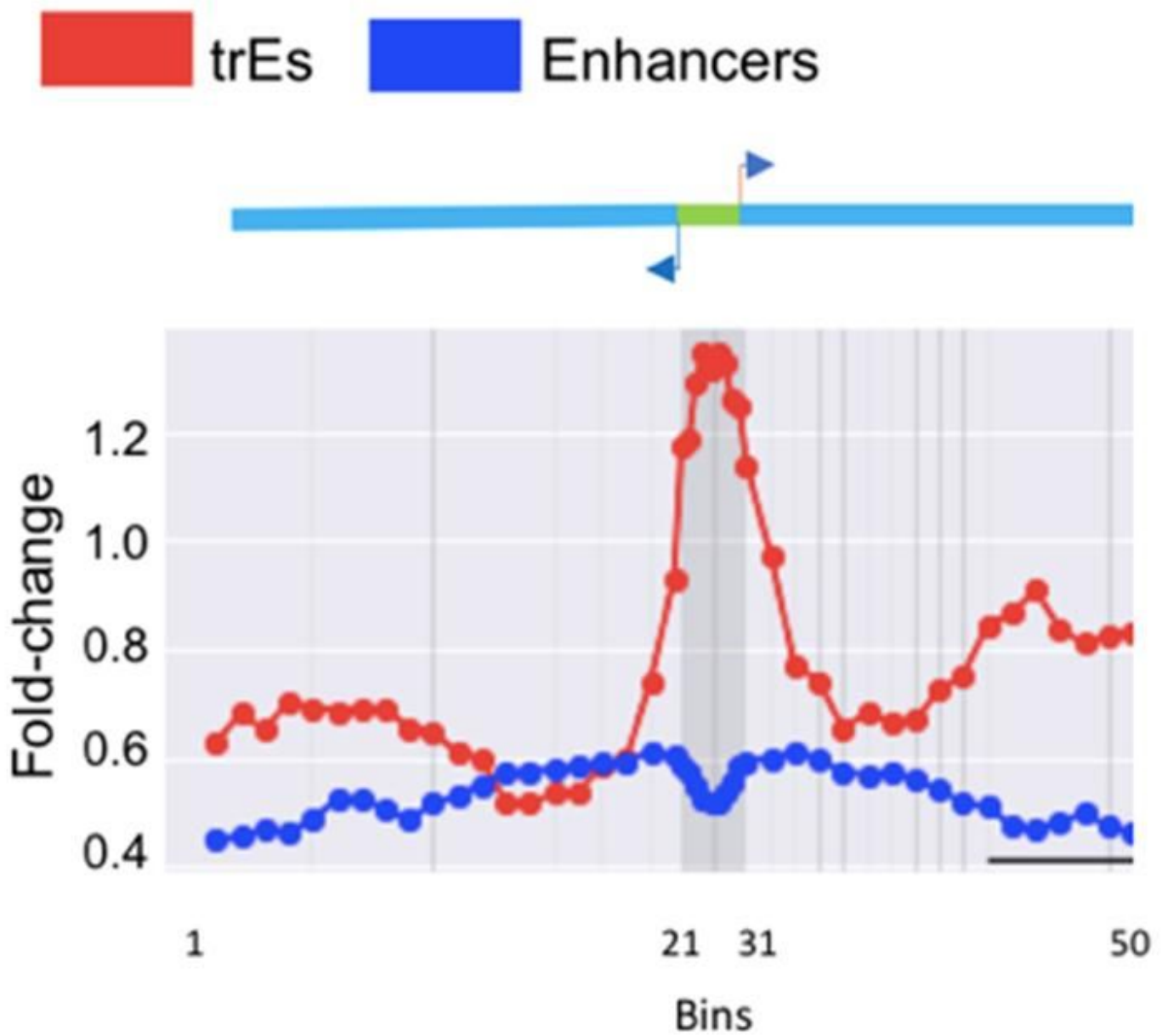


Figure 3

Conservation of trEs. PhyloP fold-change per bin with respect to background.

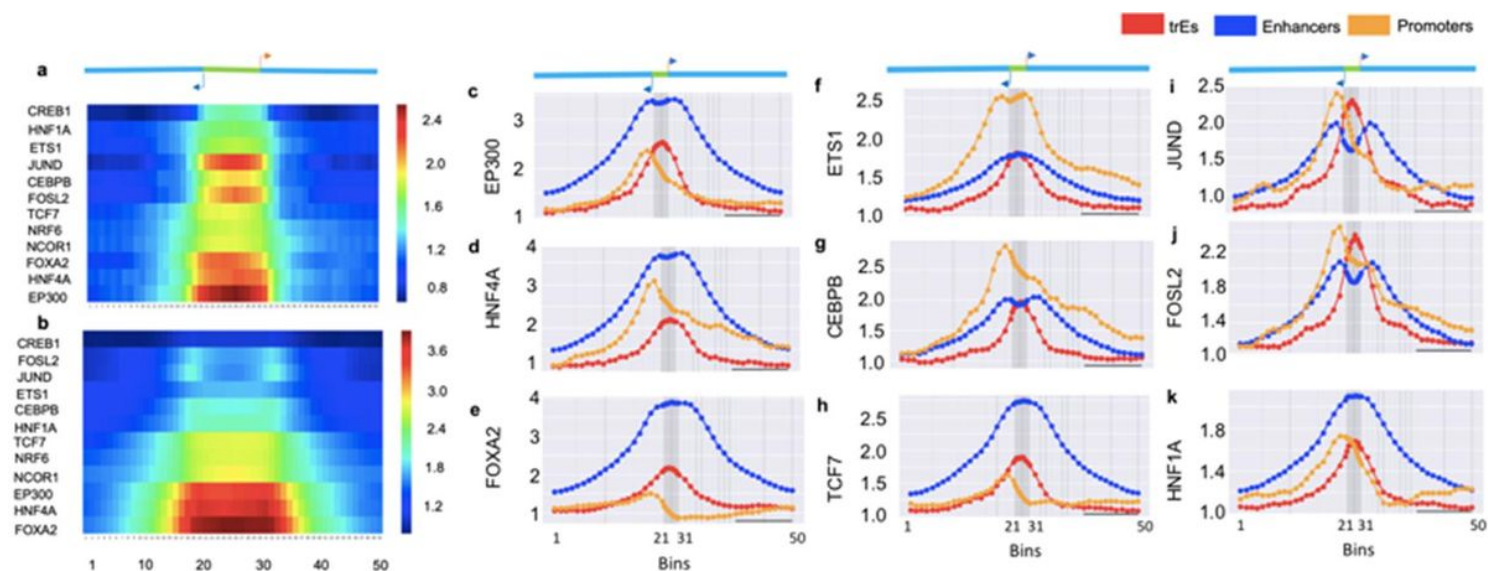


Figure 4

TF enrichment at trEs and enhancers in HepG2. (a) Heatmap for 12 liver TFs at trEs. Scale indicates the fold-change with respect to background. (b) Heatmap for enhancers. On right, detailed binding profiles of individual TFs across trEs, regular enhancers, and promoters. (c) EP300. (d) HNF4A. (e) FOXA2. (f) ETS1. (g) CEBPB. (h) TCF7. (i) JUN D. (j) FOSL2. (k) HNF1A.

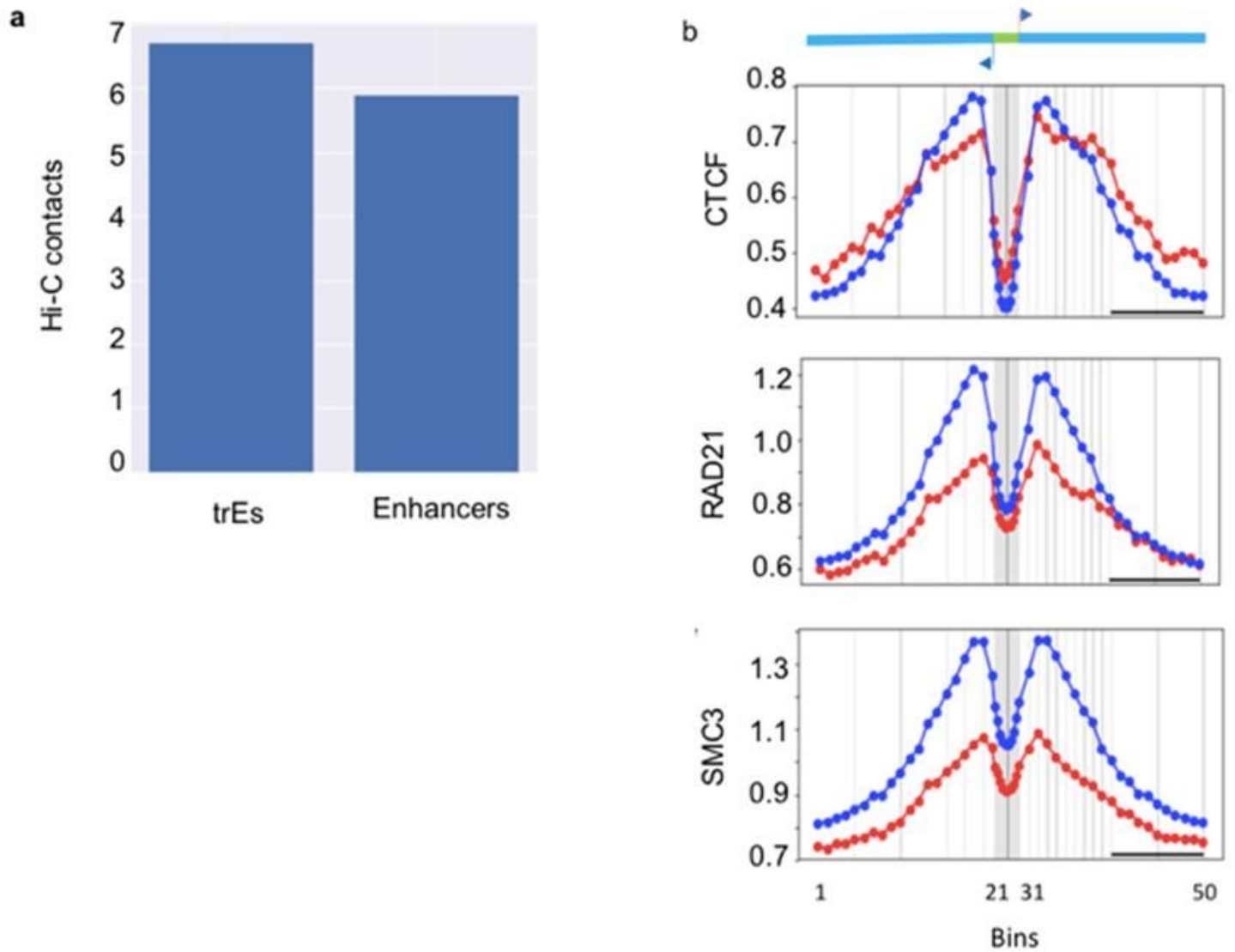


Figure 5

Chromatin loops originating from trEs and regular enhancers. (a) Number of Hi-C contacts at trEs and enhancers. (b) ChIP-seq signals of CTCF, RAD21, and SMC3 at trEs. Red plot: trEs; blue plot: enhancers.

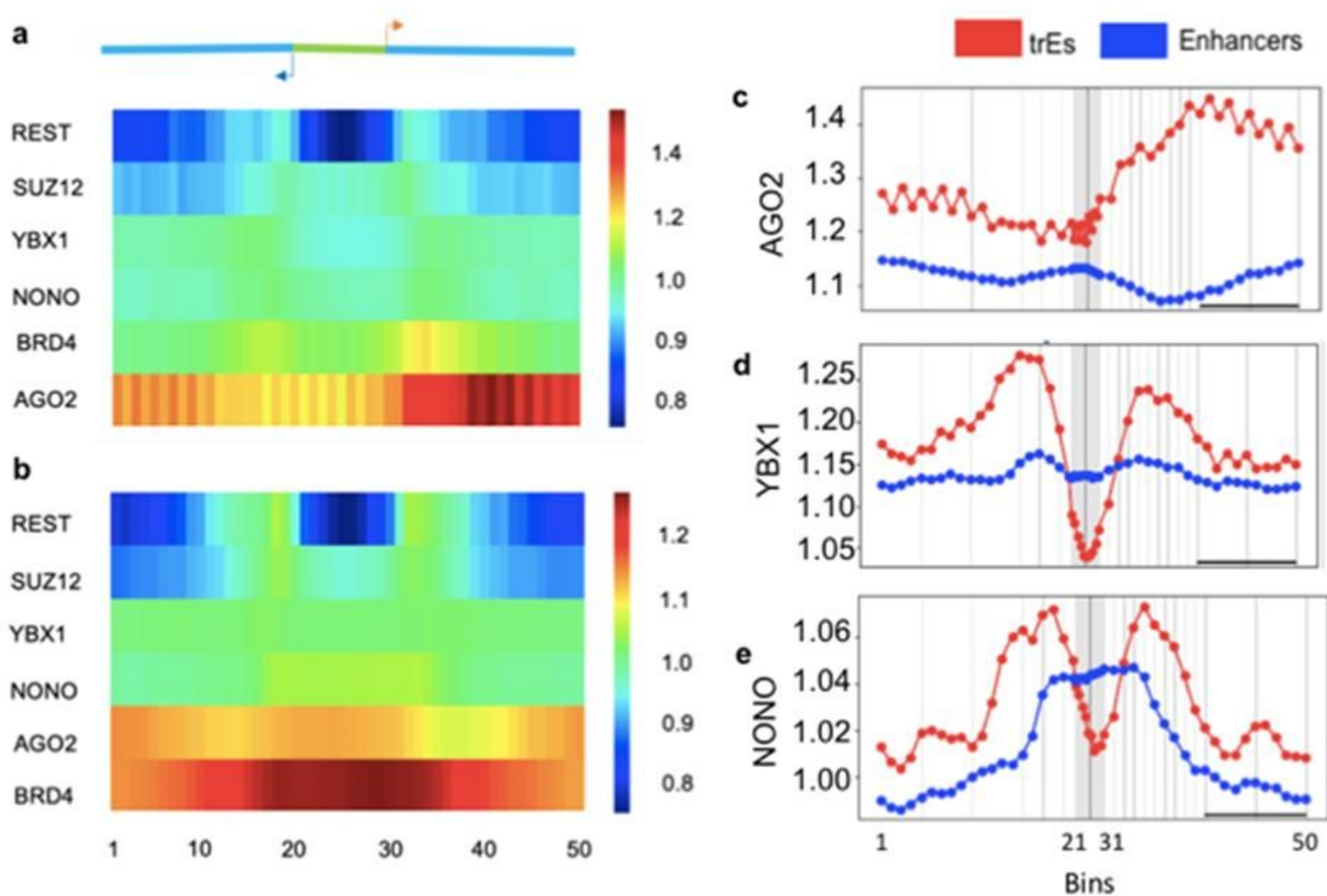


Figure 6

RNA binding proteins (RBPs) binding at trEs and enhancers in HepG2. (a) Heatmap of ChIP-Seq signals (fold change) for RBPs at trEs, (b) ChIP-Seq signals of RBPs at enhancers. (c) AGO2. (d) YBX1. (e) NONO.

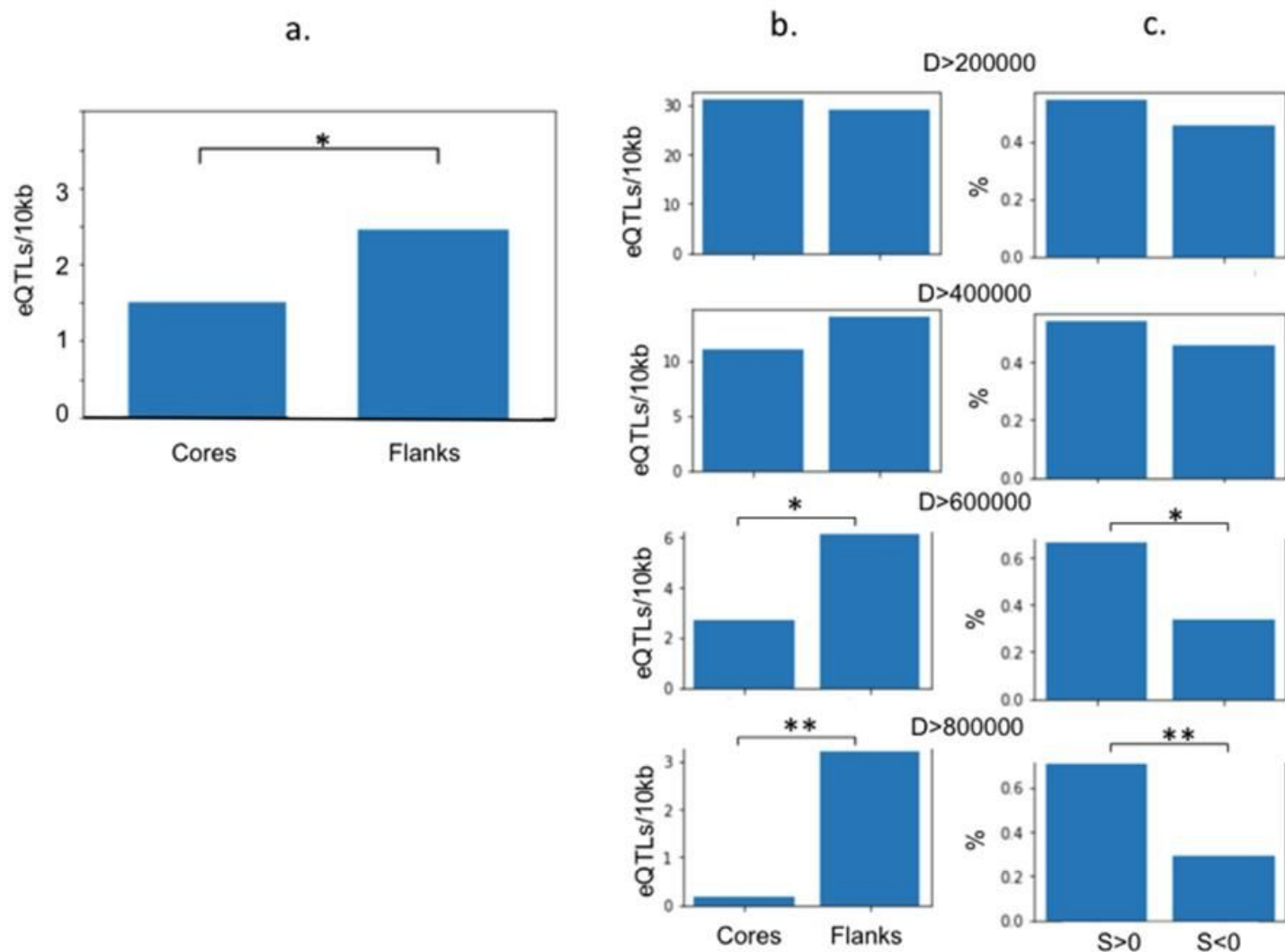


Figure 7

Enrichment of eQTLs at trEs. (a) Density of eQTLs (per 10 Kb) at the cores and flanks of thyroid trEs. (b) eQTL density as a function of distance D to the associated gene. (c) Percentage of eQTLs associated with negative ($S < 0$) or positive ($S > 0$) slope for distances D.

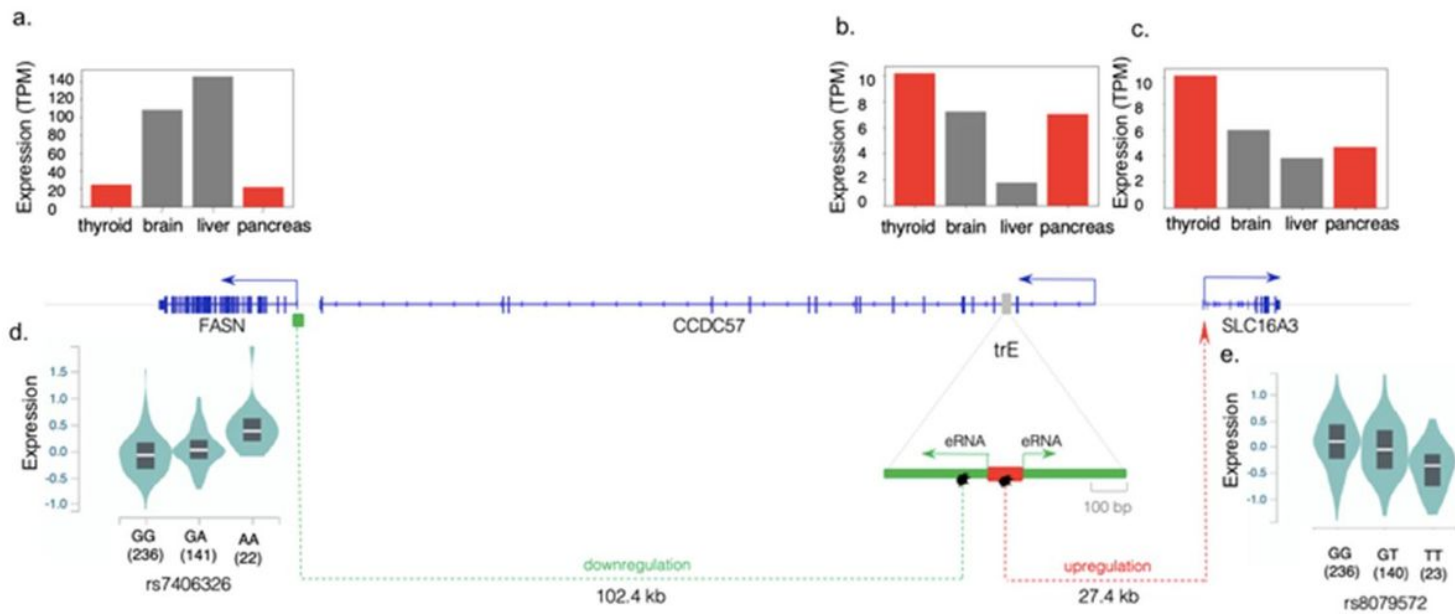


Figure 8

The regulatory landscape of the CCDC57 locus. Two eQTLs within the CCDC57 intronic trE are associated with the two genes, FASN downstream and SLC16A3 upstream, flanking CCDC57 Red: trEs is active in tissue; grey trEs is not active.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementary.pdf](#)