Transcription start site analysis reveals widespread divergent transcription in D. melanogaster and core promoter-encoded enhancer activities

Rennie, Sarah; Dalby, Maria; Lloret-Llinares, Marta; Bakoulis, Stylianos; Dalager Vaagensø, Christian; Jensen, Torben Heick; Andersson, Robin

Published in:
Nucleic Acids Research

DOI:
10.1093/nar/gky244

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Published online 6 April 2018

Nucleic Acids Research, 2018, Vol. 46, No. 11 ***5455–5469***
doi: 10.1093/nar/gky244

Transcription start site analysis reveals widespread divergent transcription in *D. melanogaster* and core promoter-encoded enhancer activities

Sarah Rennie[1,†], Maria Dalby[1,†], Marta Lloret-Llinares[2], Stylianos Bakoulis[1], Christian Dalager Vaagensø[1], Torben Heick Jensen[2] and Robin Andersson[1,*]

[1]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N, Denmark and [2]Department of Molecular Biology and Genetics, Aarhus University, C.F. Møllers Allé 3, Building 1130, 8000 Aarhus C, Denmark

Received January 22, 2018; Revised March 19, 2018; Editorial Decision March 21, 2018; Accepted March 22, 2018

ABSTRACT

Mammalian gene promoters and enhancers share many properties. They are composed of a unified promoter architecture of divergent transcripton initiation and gene promoters may exhibit enhancer function. However, it is currently unclear how expression strength of a regulatory element relates to its enhancer strength and if the unifying architecture is conserved across Metazoa. Here we investigate the transcription initiation landscape and its associated RNA decay in *Drosophila melanogaster*. We find that the majority of active gene-distal enhancers and a considerable fraction of gene promoters are divergently transcribed. We observe quantitative relationships between enhancer potential, expression level and core promoter strength, providing an explanation for indirectly related histone modifications that are reflecting expression levels. Lowly abundant unstable RNAs initiated from weak core promoters are key characteristics of gene-distal developmental enhancers, while the housekeeping enhancer strengths of gene promoters reflect their expression strengths. The seemingly separable layer of regulation by gene promoters with housekeeping enhancer potential is also indicated by chromatin interaction data. Our results suggest a unified promoter architecture of many *D. melanogaster* regulatory elements, that is universal across Metazoa, whose regulatory functions seem to be related to their core promoter elements.

INTRODUCTION

Spatio-temporal control of metazoan gene expression is mediated in part by factors acting at gene promoters and at gene-distal transcriptional enhancers. Although major efforts have been made to identify the locations of transcriptional regulatory elements (TREs, here denoting enhancers and promoters) and their cell type-restricted activities, the regulatory mechanisms of these genomic regions are not well understood. Careful characterization of the properties of TREs and the determinants of their regulatory activity is crucial to better understand the means by which cells control gene expression. Despite the often adopted view on enhancers and gene promoters as distinct entities with discernible functions and local chromatin characteristics (1–3), e.g. different levels of H3K4me1, H3K4me3 and H3K27ac at nucleosomes flanking TREs, recent observations suggest large similarities between mammalian enhancers and gene promoters (4–7). In particular, mammalian TREs are characterized by a high prevalence of divergent transcription initiation (4,8–15). In addition, enhancers frequently contain core promoter elements (4,13), bind general transcription factors (16–18), and may act as alternative gene promoters (19). Gene promoters themselves form stable chromatin interactions with other promoters, often resulting in the co-expression of genes in a tissue-specific manner (20–22). Several examples of mammalian gene promoters exhibiting enhancer function have also been identified (20,23–25). Taken together, these observations raise the question whether the repertoire of TREs may be treated as a unified class (4–7). However, it is currently unclear how promoter (expression) strength relates to enhancer strength, whether a TRE with strong enhancer function also possesses strong promoter function or vice versa, or if enhancer function is inversely related to promoter function. In addition, the inherent state of divergent transcription at TREs in Mammalia has not been well supported across Metazoa. Observations in *Drosophila*

*To whom correspondence should be addressed. Tel: +45 35330245; Email: robin@binf.ku.dk
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

melanogaster have suggested a distinct reduction in divergent events at gene promoters (26–28) and less widespread occurrences of enhancer transcription (27). These observations raise the question of whether divergent transcription and, hence, the unifying promoter architecture across TREs is not a conserved property across Metazoa.

Transcription of mammalian protein-coding genes into mRNA is coupled with upstream transcription in the reverse orientation (8–12). The latter results in relatively short, non-coding transcripts, commonly referred to as promoter upstream transcripts (PROMPTs) or upstream antisense RNAs (uaRNAs) (8,9). While core promoters in general possess unidirectional transcription initiation (29), divergent transcription is accomplished by a pair of divergent core promoters contained within the same nucleosome deficient region (NDR) (11,30). Divergent transcription initiation is also widespread at regulatory active mammalian transcriptional enhancers (13,14). The resulting transcripts, known as enhancer RNAs (eRNAs), are, similar to PROMPTs, short, low abundant and non-coding. Both PROMPTs and eRNAs are unstable and targets of the ribonucleolytic RNA exosome complex (9,10,13,31,32). The path to RNA decay is, at least in part, linked to the presence of early polyadenylation sites and a depletion of U1 small nuclear ribonucleoprotein (snRNP) binding via a lack of 5′ splice sites downstream of PROMPT and eRNA transcription start sites (TSSs), leading to early transcriptional termination (31,33). Like PROMPTS, eRNAs are of low abundance and seldom transcribed from evolutionary constrained DNA (10,34), indicating high similarities between RNA species and that both PROMPTs and eRNAs likely possess little functional relevance. Nevertheless, promoter activity of mammalian enhancers, as observed from local transcription initiation events, is an accurate indicator of active enhancer regulatory function (13), suggesting a link between distal regulatory enhancer function and local promoter activity. A notable difference between gene promoters and gene-distal enhancers is that while the transcriptional activity at mRNA promoters is strongly favoring the production of stable, exosome-insensitive RNAs on the sense strand, enhancers are generally associated with more balanced production of low-abundant unstable eRNAs on both strands (4,10,13). The apparent commonalities and differences in transcription initiation patterns, frequencies, and RNA decay have therefore been utilized for classification of TRE function (4,10,13).

Efforts to catalogue genome-wide the enhancer potential of genomic sequences for activating transcription at a given core promoter have provided insights into the regulatory potential of *D. melanogaster* genomic sequences. Assays based on self-transcribing regulatory regions (STARR-seq) (35), have revealed apparent differences between housekeeping and developmental enhancer activities, as measured by STARR-seq constructs containing core promoters associated with broad, housekeeping activities (hkCPs) and cell type-restricted or developmental core promoters (dCPs), respectively. Sequences activating the former appear to be gene promoter-proximal while sequences activating the latter are generally gene promoter-distal (36). However, the DNA sequence itself is not the only determinant of enhancer activity. A recent study utilising massively parallel reporter assays emphasizes the importance of the positioning of TREs within chromatin contexts (37). In addition, the boundaries of *D. melanogaster* topologically associating domains (TADs) (38,39) may constrain sequence-compatible enhancer-promoter regulation. Concomitantly, enhancer classes, as characterized by STARR-seq, follow distinct chromatin architectures with respect to TADs, with housekeeping enhancers enriched at domain borders and developmental enhancers enriched at the anchors of loops (39). These results are supported by a preferential enrichment of housekeeping gene promoters at TAD boundaries (40).

In this study, we set out to investigate the link between promoter activity and enhancer function and how invertebrate and mammalian genomes compare in their RNA decay and transcription initiation frequencies at TREs. To this end we measured TSS usage, RNA abundance and exosome sensitivity in *D. melanogaster* to assess whether properties (abundance, stability, directionality, divergent transcription) of TREs are conserved across Metazoa and what determines their transcriptional activities. We find that divergent transcription is a common state of *D. melanogaster* gene promoters and gene-distal enhancers, which is supported by three recent studies using alternative assays (28,41,42). Characterization of open chromatin loci into major classes, unbiased to gene annotation, solely by their transcriptional properties recapitulates mammalian archetypical groupings, which reflect gene annotations and enhancer potentials. We show that fly TREs carry remarkable similarities in terms of promoter functionality, regardless of type, pointing to a unified architecture of TREs that is similar across Metazoa. We identify quantitative relationships between TRE expression level and enhancer function, which seem to be encoded by core promoter element strengths, pointing at a regulatory trade-off between developmental enhancer function and promoter functionality and a joint encoding of promoter and enhancer functionality for housekeeping TREs. Our results further suggest at least two layers of transcription control, which are also supported by chromatin interaction data. One, in which housekeeping gene promoters act as enhancers to other gene promoters alike and one in which gene-distal developmental enhancers control the transcription of developmental genes.

MATERIALS AND METHODS

S2 cell culturing and RNA interference

Drosophila melanogaster S2 cells were cultured at 27°C in Schneider's medium (Sigma, S0146) supplemented with 10% FBS (Sigma, F7524) and 1% penicilin/streptomycin (Sigma, P0781). Double-stranded RNAs (dsRNA) to deplete Rrp6 and Dis3 were prepared by *in vitro* transcription from a PCR template with T7 promoters on both ends using the Megascript RNAi kit (Ambion, AM1626) according to the manufacturer's instructions (Supplementary Table S3). DsRNA against GFP was used as a control. For each condition, 3×10^6 cells were seeded in a well of a six-well plate. The following day, cells were washed twice with Schneider's media with no FBS and no antibiotica, a mixture of 40 μg of dsRNA (20 μg Rrp6 dsRNA and 20 μg Dis3 dsRNA or 40 μg GFP dsRNA) in 500 μl media was

added dropwise to the cells, the plates were agitated for 30 s and incubated for 6 h at 27°C. Finally, 2.5 ml of media with FBS and penicilin/streptomycin were added. The treatment was repeated 2 days later and the cells were harvested 4 days after the first dsRNA treatment.

CAGE library preparation, sequencing and mapping

CAGE libraries were prepared as described elsewhere (10,43) from total RNA purified from S2 cells with TRIzol (Ambion, 15596018) according to the manufacturer's protocol. Sequenced reads were trimmed to remove linker sequences and subsequently filtered for a minimum sequencing quality of Q30 in 50% of the bases. Mapping to *D. melanogaster* (dm3, r5.33) was performed using Bowtie (44) (version 1.1.1), allowing max two mismatches per read and keeping only uniquely mapped reads.

Processing of DNase-seq data and identification of DNase I hypersensitive regions

Sequencing reads from DNase-seq and input data (35) were trimmed using Trimmomatic (45) (version 0.32) using a sliding window approach, trimming off the ends of reads in 4 nucleotide windows that did not fulfil a quality ≥Q22. Reads trimmed to a length shorter than 25 nucleotides were discarded. Kept reads were mapped to the dm3 (r5.33) reference genome using Bowtie (44) (version 1.1.1), allowing max three mismatches per read and keeping only uniquely mapped reads. DNase I hyperensitive sites (DHSs) were called using hotspot (46) at a FDR threshold of 0.01, on input data, pooled DNase-seq data and in each of two DNase-seq replicates. DHS hotspot peaks called from pooled replicates that overlapped peaks from individual replicates and not peaks from input DNase-seq data were used for further analyses. This resulted in a final set of 11 947 DHSs.

CAGE tag clustering and expression quantification

Tag clustering was performed on pooled CAGE data, including all four replicates from each condition, using a summit-fraction strategy to remove tails from wide TCs and split multi-modal peaks (see Supplementary Methods). This resulted in a set of 670 681 TCs.

The expression of each TC in each individual CAGE replicate was quantified by counting of CAGE 5′ ends falling into their defined genomic regions. In addition, CAGE genomic background noise levels were estimated (see Supplementary Methods). Only TCs whose expression was above the CAGE genomic background noise threshold in at least two out of four replicates in each condition were considered. Noise level filtering resulted in 121 809 TCs in control CAGE libraries and 147 379 TCs in exosome knockdown CAGE libraries. Expression levels were converted to tags per million (TPM), by counting the CAGE tags per TC and normalising to library size scaled by 10^6. TCs were annotated to FlyBase gene TSSs based on a max distance between TC summit positions and gene TSSs of 250 bp (upstream or downstream).

To measure the effect of the knockdowns, the mean expression and standard deviation over the four replicates for the knockdown and control experiments were calculated for TCs annotated to the primary FlyBase TSSs of the genes *Rrp6* and *Dis3*.

RT-qPCR validations of CAGE expression levels and exosome sensitivities

Expression levels and exosome sensitivities measured by CAGE were validated using RT-qPCR (Supplementary Table S4). The RNA was treated with TURBO DNA-free kit (Ambion, AM1907) and cDNA was prepared with the SuperScript II kit (Invitrogen, 18064014), using 1 μM oligo dT18 and 5 ng/μl/ μl random primers. qPCRs were performed with Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen, 11744500) in a MX3000P (Agilent technologies) machine. RNA levels were normalized to that of *Act5C*.

DHSs as focus points for transcription initiation

DHSs were used as focus points for characterising patterns of transcription initiation events at TREs as described elsewhere (10), with minor modifications (see Supplementary Methods). DHSs were annotated to FlyBase (release 5.12) genes by intersection of DHSs with gene TSSs extended 200 bp upstream and 100 bp downstream with respect to the annotated gene strand. DHSs were subsequently categorized into those associated with gene TSSs on plus strands or minus strands only (unidirectional genes), those associated with both gene TSSs on plus strands and minus strands (divergent gene pairs), or those that were not annotated on any strand (gene TSS-distal).

Unsupervised clustering of DHSs was performed on the basis of transcriptional properties derived from each replicate CAGE library in a two-step approach (see Supplementary Methods), to guarantee high agreement between individual replicates.

Evaluation of enhancer potential by STARR-seq data

Wiggle track STARR-seq data (36) were used to evaluate the enhancer potential of transcribed DHSs. For each DHS, the summit signal within a 401 bp region centered on the DHS mid point was identified from STARR-seq data generated using *RpS12* and *even skipped* core promoters, for housekeeping (hkCP) and developmental (dCP) enhancer potential, respectively. At the summit position, the $\log_2$ fold change of STARR-seq signal over STARR-seq input signal was calculated. STARR-seq active regions were defined as those DHSs having a $\log_2$ fold change of at least 1.5.

Core promoter element scans and clustering

Core promoter element occurrences were scanned around each transcribed DHS using MEME FIMO (v4.11.2) (47). A genome sequence database of ±50 bp around major and minor strand CAGE summit positions within each DHS was considered. Position weight matrices (PWMs) of MTE and TATA core promoter elements were retrieved from JASPAR POLII database (48), species *D. melanogaster*. DRE, Inr, Trl, E-box motifs were retrieved

from DMMPMM *D. melanogaster* motif collection (49). Finally, DPE, Ohler6 (Motif6) and Ohler 1 (Motif1) motifs were collected from (50). Motif PWMs and consensus sequences were converted into the Minimal MEME Motif Format using the tools chen2meme and iupac2meme, respectively. FIMO scans were performed with a statistical threshold (*P*-value) equal to 1 and a maximum number of motif occurrences retained in memory at 100×10^6. The FIMO output was filtered to only contain the motif hits with maximum score for each motif occurrence for each DHS strand window. Motif hits were subsequently considered significant if they passed a *P*-value threshold <0.001. Core promoter element clustering was performed using the R function pheatmap (Ward.D agglomeration) on scaled and centered data for each core promoter element. Core promoter clusters were determined by the cutree R function with $k = 10$ desired groups.

Analyses of histone modifications and transcription factor binding at DHSs

Binarised (51) modENCODE ChIP-chip data in 50 bp regions for histone modifications, histone variants and transcription factor binding were investigated around transcribed DHSs. In order to generate the heatmap against classes, a 50 bp window surrounding the center of each DHS was overlapped with the 50 bp ChIP-chip regions and the mark was recorded as present at a DHS if it overlapped with at least one element. For ChIP-chip footprint plots, the binary signal was averaged across sites in 50 bp bin intervals from the CAGE summits of the considered DHSs, up to a maximum of 5000 bp away from the summits. Cases where a given interval for a DHS overlapped another DHS were filtered from the analysis. The background level was generated based on randomising the ChIP-chip locations, 10 times for each combination of set and mark.

ChIP-seq data (52) for histone modifications H3K4me3, H3K4me1 and H3K27ac were processed by the AQUAS ChIP-seq pipeline (http://github.com/kundajelab/). Mapped ChIP-seq signal was then quantified within each 401 bp DHS region. ChIP-seq signal within DHS windows was plotted as a function of housekeeping (hkCP) or developmental (dCP) enhancer potential, using the binned 1−99th percentile STARR-seq signal. ChIP-seq signal within DHS windows was also plotted against the binned 1−99th percentile CAGE TPM expression.

Assessment of the relationship between chromatin architecture and type of regulatory element

TADs and significantly interacting regions for *D. melanogaster* Kc167 cells based on 1kb resolution HIC data (39) were considered. For all HiC and TAD-associated analyses, the coordinates for ±200 bp around the CAGE summits of DHSs were lifted over to dm6, keeping all DHSs whose width were preserved in the liftover coordinates (9454, corresponding to 99.8% of DHSs defined in dm3). DHSs were allocated a TAD number if they overlapped the TAD by at least 200 bp. All pairs of TREs within a maximum distance of 1 Mb between the DHS center points were identified and annotated according to DHS class membership and whether the pair overlapped coordinates of significantly interacting regions.

For each DHS class, the proportion of elements annotated as falling inside of a TAD region, or between (not overlapping a TAD) was calculated. For each TAD, the number of DHSs of each class was aggregated. To calculate the enrichment of elements according to TAD size, TADs were split according to the total number of DHSs that were within them, grouping all TADs with more than six elements, and the proportion of each of the classes calculated per total size. The $\log_2$ scaled data containing the number of TREs per class in each TAD, for a minimum TAD size of three elements were further clustered using the kmeans++ algorithm, generating seven clusters of TADs (as determined based on inspection of a scree plot for 2–20 possible clusters). To calculate enrichments of TAD boundary vicinities of DHS classes, a cut-off of 1 kb from the nearest TAD boundary was applied to determine inclusion or exclusion of a class element from a boundary region.

Statistics of interactions and co-occurrences within TADs were analysed using generalized linear models (see Supplementary Methods).

RESULTS

Fly regulatory elements are associated with divergent transcription and RNA species-specific decay

To characterize transcription initiation events in *D. melanogaster*, we performed deep Cap Analysis of Gene Expression (CAGE (43)) sequencing (33.5–46.4 million mapped reads per library) in Schneider line 2 (S2) cells. To measure exosome sensitivity, cells were subjected to a double knockdown of the catalytic subunits Dis3 and Rrp6 of the ribonucleolytic exosome complex (by RNA interference, Materials and Methods). This resulted in a marked reduction in the abundances of tags aggregating at the annotated TSSs of *Dis3* and *Rrp6* genes when compared to control (dsRNA against GFP) libraries (Supplementary Figure S1A). The great majority of CAGE tags were proximal to open chromatin regions as measured by DNase I hypersensitivity (82–86% within 500 base pairs (bp), Supplementary Figure S1B), indicating a high signal-to-noise ratio of the mapped CAGE data. We observed several instances of divergent transcription initiation, with exosome-sensitive PROMPTs originating upstream of FlyBase (53) gene TSSs in a divergent manner and in a good agreement between replicates (exemplified by Figure 1A). In addition, many enhancers were associated with replicate-consistent exosome-sensitive divergent eRNAs (exemplified by Figure 1B). These observations suggest that fly biogenesis and decay of eRNAs and PROMPTs may match those of human.

To quantify the extent and characteristics of divergent transcription in fly cells, we clustered proximally mapped CAGE tags into genomic regions representing CAGE-inferred TSSs (referred to as tag clusters (TCs)). Wide TCs were trimmed and those representing multi-modal peaks were split into narrow single-peak TCs (Supplementary Methods). Although TC expression levels were largely concordant between biological replicates (Supplementary Figures S2 and S3), we filtered out TCs with low-level expres-

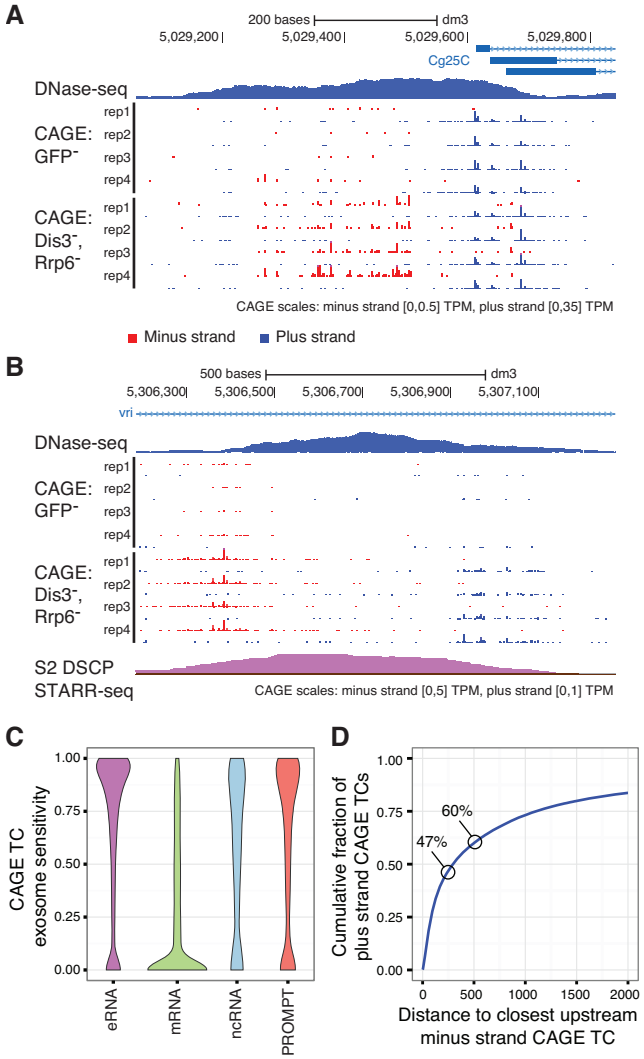

Figure 1. Fly regulatory elements are associated with divergent transcription initiation. (A, B) Genome browser views around FlyBase annotated TSSs of the *Cg25C* (also known as *Col4a1*) gene (**A**) and a *vri* intragenic dCP STARR-seq enhancer (**B**). DNase-seq, control CAGE, and exosome KD CAGE data are shown. All four replicates per CAGE condition are displayed (red: minus strand, blue: plus strand). For visibility reasons, the scales of CAGE signal differ between strands and are provided below each panel. See Supplementary Figures S8 and S9 for RT-qPCR validations. (**C**) Distributions of exosome sensitivity, ranging between 0 (insensitive) to 1 (CAGE expression not observed without exosome KD), for eRNAs (associated with dCP STARR-seq enhancers), FlyBase mRNAs, FlyBase ncRNAs and PROMPTs (upstream of and antisense to annotated FlyBase gene TSSs). (**D**) Cumulative fraction (vertical axis) of plus strand CAGE TCs that are within a certain distance (horizontal axis) of minus strand CAGE TCs. The percentages of divergent events are highlighted for distances of 250 and 500 bp.

sion (not statistically distinguishable from genomic background noise estimated from TSS-unlikely loci, see Supplementary Methods), allowing us to accurately assess the transcriptional patterns of TREs naturally associated with low abundant RNAs (like eRNAs and PROMPTs). For the remaining TCs, we measured the fraction of expression in knockdown conditions to that observed in control libraries, providing a quantitative measure of exosome sensitivity ranging between expression levels fully captured by control CAGE data (exosome sensitivity 0) to expression levels only observed upon exosome knockdown (exosome sensitivity 1). Overall, the majority of TCs associated with annotated mRNA TSSs (~62%) displayed low (<0.25) exosome sensitivity (Figure 1C). In contrast, a large fraction of PROMPTs (TCs <500 bp upstream of and antisense to annotated FlyBase gene TSSs), ncRNAs (TCs associated with annotated FlyBase ncRNA TSSs) and eRNAs (TCs associated with gene TSS-distal dCP STARR-seq enhancers (36)) were mainly highly (>0.75) exosome sensitive (~51%, ~42% and ~60%, respectively).

We next calculated the distance from each plus strand TC to the nearest upstream (non-overlapping) minus strand TC (Figure 1D). ~47% and ~60% of CAGE TCs had, regardless of annotation, a nearest upstream minus strand TC within 250 and 500 bp, respectively. The relative increase in the divergent fraction was reduced at larger distances, suggesting that many fly divergent events are contained within the same NDR (that are most often <500 bp in size). Among specific TREs, we observed that transcribed gene-distal dCP enhancers were frequently (~81%) divergently transcribed. Proximal bidirectional (head-to-head) gene pairs showed the highest degree of divergent transcription (~90%). Stand-alone mRNA promoters, on the other hand, exhibited the least degree of divergent transcription (~46%). These fractions are likely underestimates, since rare transcripts can fall below imposed noise thresholds. Nevertheless, these results demonstrate that a considerable proportion of transcription initiation events in *D. melanogaster* are divergent, and that a fraction of these events are associated with exosomal RNA decay, in accordance with human cells.

Expression and exosome sensitivity patterns characterize distinct regulatory elements

With an aim to systematically characterize transcription initiation events and associated RNA turnover at TREs unbiased to existing annotations, we focussed the remaining analyses on DNase I hypersensitive sites (DHSs) (Materials and Methods). Overall, DHSs showed preferences for minus strand expression upstream (relative to the genome reference) and plus strand expression downstream of DHS center points in exosome depleted libraries (Supplementary Figure S5, $P < 2.2 \times 10^{-16}$, odds ratio 13.3, Fisher's exact test). Only 139 (1.5%) DHSs were associated with convergent transcription. Based on these observed trends, we quantified DHS-associated expression through aggregation of CAGE tags in strand-specific divergently oriented windows of 200 bp immediately flanking DHS center points that maximized CAGE tag coverage (Supplementary Figure S4, Materials and Methods). 9471 out of 11 947 (~79%) called DHSs were significantly expressed (above estimated background noise levels) on any strand in at least two control or exosome depleted libraries. Below, we refer to the most highly expressed strand from a DHS as the 'major' strand and the other strand as the 'minor'. For each DHS, we quantified expression-associated properties on a per-replicate basis according to the knockdown-ascertained major and minor strand expression levels, ma-

jor and minor strand exosome sensitivity scores, and exosome knockdown-derived transcriptional directionality (Materials and Methods). We then performed unsupervised clustering of the transcriptional properties from each of the four replicates across transcribed DHSs, based on a two step clustering procedure (Supplementary Methods). First we clustered all DHSs into six groups, regardless of replicate. We then compared the resulting group allocation across the replicates for each DHS and clustered a second time, thus generating a final set of clusters, which strongly agreed per DHS across replicates (Supplementary Figure S6). This resulted in six major groups (Figure 2A, Supplementary Figure S7), each of which disagreed on average by at most one replicate within-group (Supplementary Table S1). Only 15 DHSs were removed from further analyses due to lack of replicate agreement, demonstrating that inferred DHS groupings are robust against biological replicate variance. Results from CAGE data were further validated across randomly selected loci by RT-qPCR (Supplementary Figures S8 and S9, Supplementary Table S2), demonstrating the accuracy in determining RNA abundance and turnover even at lowly expressed TSSs.

The clustering of DHSs revealed several interesting relationships between expression levels, transcriptional directionality and exosome sensitivity, and displayed widely different enrichments of annotated gene TSS proximities (Figure 2B, Supplementary File 1). Three identified classes of DHSs were associated with stable (exosome insensitive) RNAs on their major strand. The proximal regions of *unidirectional stable* and *unidirectional stable w/ PROMPT* DHSs were highly enriched ($P < 2.2 \times 10^{-16}$, Chi-squared test) with annotated unidirectional gene TSSs, consistent with a strong directional expression bias resulting from high expression levels and low exosome senstivity from their major strands. In contrast to *unidirectional stable* DHSs, DHSs in the *unidirectional stable w/ PROMPT* category were in addition associated with lowly expressed and highly unstable RNAs from their minor strands, properties reminiscent of human mRNA gene promoters associated with PROMPT transcription (9,10,13,31,32). We also identified a smaller class with more balanced, stable, high expression on both strands (*bidirectional stable*). DHSs in this class were enriched ($P < 2.2 \times 10^{-16}$, Fisher's exact test) in annotated head-to-head gene TSSs. We collectively refer to these three classes as *stable* TREs, due to the low exosome sensitivity of RNAs transcribed from their major strands.

The remaining DHSs were grouped into three classes associated with exosome-sensitive RNAs emitted from their major strands (*unstable* TREs, Figure 2A). One class (*weak bidirectional unstable*) gathered DHSs associated with balanced low output of unstable RNAs. *Intermediate bidirectional stable* DHSs exhibited moderately higher expression on the major strand resulting in a more biased directional transcription. DHSs having close to unidirectional output of unstable RNAs were grouped in the third class of *unstable* DHSs (*weak unidirectional unstable*). All three unstable clusters were highly enriched ($P < 2.2 \times 10^{-16}$, Chi-squared test) in gene TSS-distal regions (Figure 2B).

We next compared DHS classes by their association with genome-wide signals from STARR-seq data using constructs based on housekeeping (hkCP) and developmental (dCP) core promoter types (*RpS12* and *even skipped* core promoters, respectively) (36). Only 51% (2767 out of 5408) of previously called dCP enhancers (36) overlapped the set of 11 947 DHSs identified in this study, indicating that many STARR-seq positive regions are not regulatory active *in vivo* but rather should be considered to possess enhancer potential. In contrast, 82% of DHS-overlapping dCP enhancers were transcribed. To investigate the enhancer potential of transcribed DHSs with respect to DHS classes, we considered the $\log_2$ fold change between STARR-seq signal and input (>1.5) around DHSs (Figure 2C). With the exception of *weak unidirectional unstable* DHSs, *unstable* DHSs were enriched ($P < 2.2 \times 10^{-16}$, Chi-squared test) in dCP enhancers, with the largest overlap (∼55%) found among *weak bidirectional unstable* DHSs. Importantly, these results provide external evidence that balanced bidirectional output of exosome-sensitive RNAs is a marker of gene promoter-distal open chromatin TREs with enhancer potential in *D. melanogaster*, which has previously been established in human cells (4,10,13). Interestingly, in agreement with previous reports (36), a large fraction of *stable* DHSs overlapped with hkCP positive enhancers. The largest overlaps were observed for *bidirectional stable* (∼80%) and *unidirectional stable* (∼71%) DHSs. *Weak unidirectional unstable* DHSs had modest overlap with STARR-seq positive enhancers and displayed no real preference to either housekeeping or developmental core promoters. We observed similar STARR-seq enrichments when restricting our analyses to intergenic regions (Supplementary Figure S10), indicating that the hkCP and dCP enrichments are not counfounded by proximity to annotated gene promoters.

In conclusion, clustering of DHSs by their transcription initiation frequencies and associated exosome sensitivity reveals overall similarities between derived clusters of DHSs in *D. melanogaster* S2 cells and those identified in human cells (10). In addition, a large proportion of *D. melanogaster* TREs show archetypal mammalian-derived properties of PROMPTs and eRNAs. This suggests that mammalian and invertebrate genomes share similar classes and promoter architectures of TREs.

DNA sequence elements reflect transcriptional directionality and RNA instability

The differences in annotation preferences and transcriptional directionalities between *stable* and *unstable* DHSs prompted us to investigate the relationships between transcriptional output (directionality and RNA exosome sensitivity) and DNA sequence elements at the core promoters and in regions downstream of TSSs of transcribed DHSs. First, we assessed the frequencies of predicted 5′ splice sites and termination signals (polyadenylation sites) at the locations of minor and major strand CAGE summits and up to 1000 bp downstream (Figure 3A, B, Supplementary Methods). Similar to what has been previously observed in human (31,33), we observed an enrichment in downstream 5′ splice sites on the major strands of *stable* DHSs while site frequencies were close to or under the genomic background level for their minor strands and for both strands of *unstable* DHSs (Figure 3A), indicating that the instability of RNA is inversely related to downstream flanking 5′ splice site se-

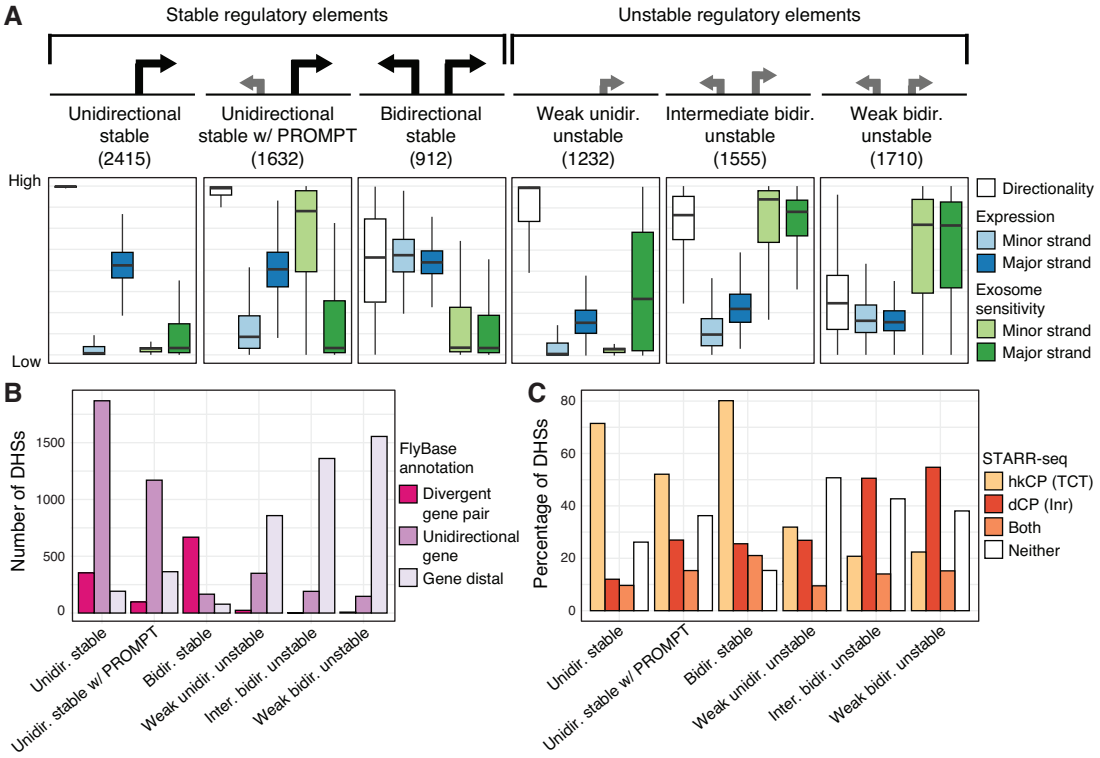

Figure 2. Transcriptional directionality, expression level and exosome sensitivity reveal major groupings of *D. melanogaster* regulatory elements. (**A**) Manual labeling and properties of the six identified clusters of transcribed DHSs with similar transcriptional directionality, expression levels, and exosome sensitivities (displayed in box-and-whisker plots). DHS clusters associated with stable or unstable RNAs on their major strands are indicated above DHS cluster labels. See Supplementary Figure S4 for a schematic of the measures used for clustering and the strategy behind expression quantification of DHSs. The lower and upper hinges of boxes correspond to the first and third quartiles of data, respectively, and the whiskers extends to the largest and smallest data points no further away than 1.5 times the interquartile range. For improved visibility, outlier data points are not visualized. (**B**) The number of DHSs in each cluster that are in close proximity with divergent (head-to-head) FlyBase gene TSS pairs (divergent gene pair), stand-alone FlyBase gene TSSs (unidirectional gene), or distal from FlyBase gene TSSs. (**C**) The percentage of DHSs in each cluster that are overlapping with or are distal to called STARR-seq enhancers, broken up by those overlapping with hkCP enhancers, dCP enhancers or both classes.

quences. Enrichments of 5′ splice sites were supported by a higher prevalence of multi-exonic transcripts, inferred from RNA-seq data (54), arising from the major strands of *stable* DHSs (Figure 3C). In contrast, unstable RNAs were mostly unspliced. In further agreement with the human system, we noted that polyadenylation sites (AWTAAA consensus hexamers) were in general depleted downstream of stable RNA TSSs, but above or at genomic background levels for unstable RNAs (Figure 3B). However, in contrast to human (10), we found an enrichment of polyadenylation sites in the immediate region (within 100 bp) downstream of stable RNA TSSs before falling below the genomic background further (>200 bp) downstream. High frequencies of polyadenylation sites in the 5′ untranslated region (UTR) of many fly mRNAs have previously been characterized (55). These enrichment differences between fly and human might reflect differences in sequence preferences between fly and mammalian gene promoters (56,57), such as a depletion of CpG islands in the *D. melanogaster* genome.

Next, we investigated the prevalence of core promoter elements on minor and major strands of transcribed DHSs (Supplementary File 2). We focused on eight functional core promoter elements (50,56–58) and the Trl element (GAGA motif of *Trithorax-like*) that, based on motif finding, either had clear preferences for expected positions (TATA, Inr, DPE, MTE (Ohler10) and E-box (Ohler5), Supplementary Figure S11) or an enrichment in individual DHS classes compared to random genomic background regions distal to DHSs but with weaker positional bias (Ohler1, DRE, Ohler6 and Trl, one-sided Mann–Whitney *U* test $P < 1 \times 10^{-20}$). Overall, TATA, Inr, DPE and MTE elements displayed the strongest positional preferences on major strands among investigated motifs (Figure 3D). Among the nine motifs, DRE, Ohler6, TATA and Trl elements showed the highest presence on minor strands (Supplementary Figures S11 and S12A), although at lower frequencies than on major strands, suggesting that TREs of *D. melanogaster* may be composed of two divergent core promoters. Indeed, a considerable fraction of DHSs were associated with at least one significant core promoter element motif match on both strands (ranging between ∼20% for *weak bidirectional unstable* DHSs to ∼52% for *bidirectional stable* DHSs, Figure 3E). In comparison, ∼12% of random genomic background regions had significant motif matches on both strands. While these results reflect a potential of having two divergent core promoters across TREs, calling of motif instances in genomic sequences can be inexact and does not

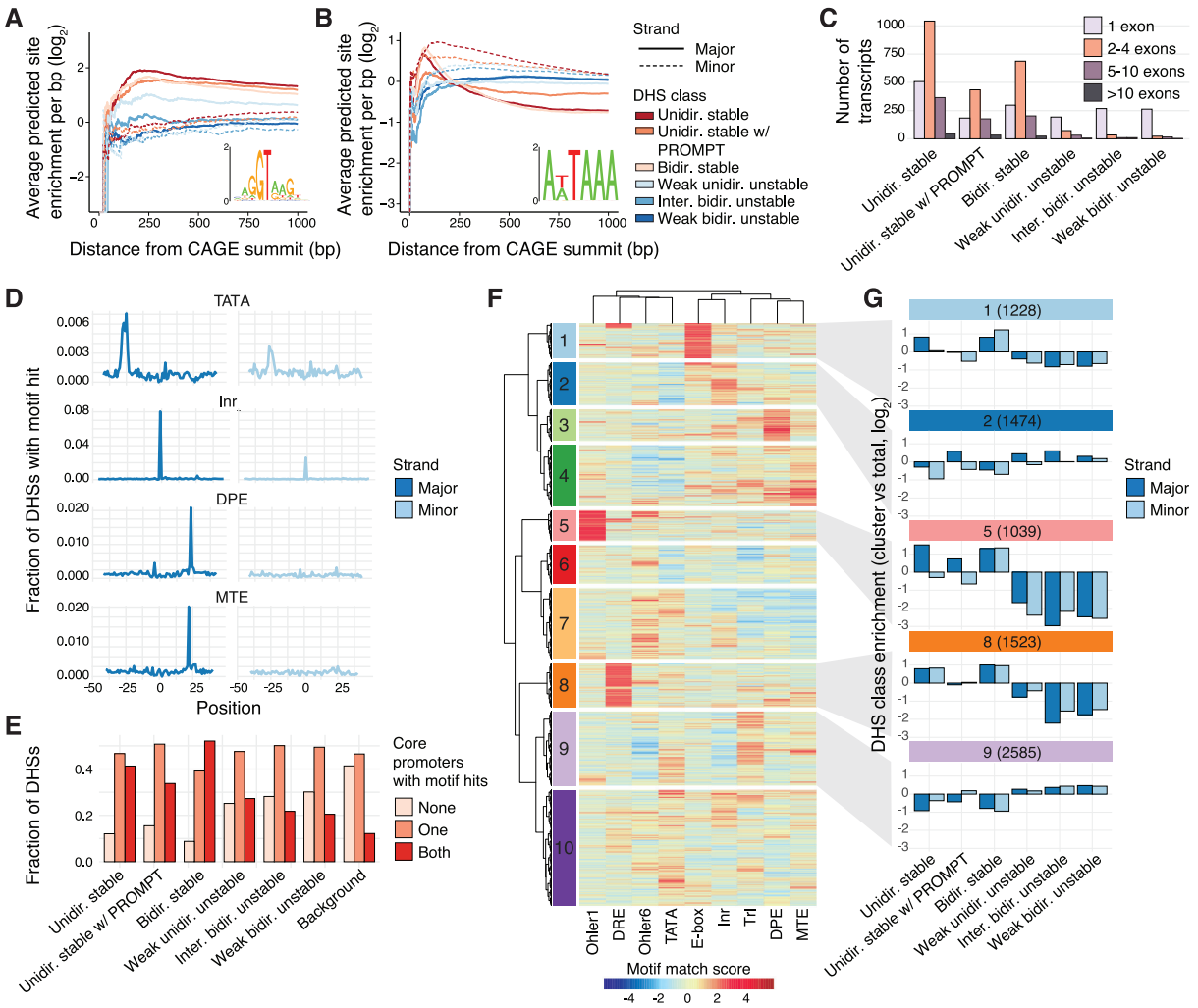

Figure 3. DNA sequence elements impact the stability, directionality and expression strengths of regulatory elements. (**A**, **B**) Frequencies of RNA processing motifs (A: 5′ splice site, B: polyadenylation site) downstream of CAGE summits broken up by DHS class and strand. Vertical axes show the average number of predicted sites per bp within an increasing window size from the TSS (horizontal axis) in which the motif search was done. 0 indicates the expected hit frequency from random genomic background. (**C**) Histogram of de novo-assembled transcript counts (vertical axis), broken up by number of exons and associated DHS class. (**D**) Fraction of transcribed DHSs (vertical axis) with an identified core promoter element (TATA, Inr, DPE, or MTE) at a given position relative to the major (left panels) and minor (right panels) strand CAGE summits. (**E**) Fraction of transcribed DHSs within each DHS class (vertical axis) associated with at least one out of nine core promoter elements identified on one or both strands. In addition, the fraction of DHSs with no core promoter elements are shown (none). (**F**) Hierarchical Ward agglomerative clustering of motif match scores for the nine considered core promoter elements on major and minor strands of transcribed DHSs. Ten clusters of core promoter element compositions are shown. (**G**) DHS class enrichments, calculated as the fraction of DHSs in each DHS class associated with each core promoter element cluster versus the fraction of total transcribed DHSs, displayed in $\log_2$ scale enrichment, broken up by major and minor strand. See Supplementary Figure S13 for DHS class enrichments for all core promoter clusters.

directly reflect the strengths of considered core promoter elements.

To investigate potential differences between DHS classes as well as between minor and major strands, taking into account the strengths of motif matches, we clustered the maximum match scores for each considered core promoter element motif in regions surrounding the CAGE summits of minor and major strands (Methods). This revealed a complex combination of core promoter elements across DHSs, for which many core promoter elements were restricted to only a subset of DHS regions (Figure 3F), reflecting the wide diversity of core promoter compositions in *D. melanogaster* (50,57). Individual core promoter clusters displayed strong match scores for individual core promoter elements, such as E-box (cluster 1), Inr (cluster 2), DPE (cluster 3), MTE (cluster 4), Ohler1 (cluster 5), DRE (cluster 8), and Trl (cluster 9). We observed several preferences between core promoter elements and the stability of associated RNAs (Figure 3G, Supplementary Figures S12B and S13A). For instance, cluster 5 (associated mainly with Ohler1) had a strong enrichment of major strands of *unidirectional stable* and *unidirectional stable w/ PROMPTs* DHSs (Fisher's exact test, $P < 1 \times 10^{-16}$, $P < 1 \times 10^{-10}$, respectively), as well as both strands of *bidirectional stable* DHSs (Fisher's exact test, $P < 1 \times 10^{-16}$, both), but a strong depletion of core promoters of *unstable* DHSs (Fisher's ex-

act test, $P < 1 \times 10^{-10}$). In contrast, Trl elements (cluster 9) were mostly found in core promoters of *unstable* DHSs. Other core promoter clusters were not mainly associated with the stability of produced RNAs. In particular, Inr elements (mostly identified in cluster 2) showed no clear differences between *stable* and *unstable* DHSs, but rather a preference for major over minor strands. Interestingly, cluster 8 (associated mainly with DRE) was enriched with *unidirectional stable* and *bidirectional stable* DHSs (Fisher's exact test, $P < 1 \times 10^{-15}$, all) but not *unidirectional stable w/ PROMPTs* DHSs. In general, we found that *bidirectional stable* and *unidirectional stable* major strands showed similar enrichment within core promoter clusters, while *unidirectional stable w/ PROMPTs* major and minor strands showed higher similarities with *unstable* DHS groups (Supplementary Figure S13B).

Taken together, we conclude that *D. melanogaster* TREs are frequently associated with core promoter elements regardless of DHS class, but possess strong diversity in core promoter composition that are reflecting RNA stability and transcriptional directionality.

Genomic positioning and core promoter elements may impede divergent transcription

We next wanted to investigate the nature of absent PROMPT transcription from *unidirectional stable* DHSs. Invertebrate genomes have an unexpectedly high fraction of head-to-head gene pairs, not immediately explained by their more compact genomes compared to mammalian ones (59). Given that the distance between head-to-head gene pair TSSs has an observable impact on human PROMPT transcription (60), we compared the distances between upstream antisense gene TSSs and major strand CAGE summits of *unidirectional stable* and *unidirectional stable w/ PROMPTs* DHSs. We noted a clear difference in positional preference between DHS classes (Figure 4A). Major strand TSSs of *unidirectional stable* DHS were more frequently positioned in close proximity (within 1000 bp) of upstream annotated head-to-head gene TSSs than the major strand TSSs of *unidirectional stable w/ PROMPTs* DHSs (Fisher's exact test, $P < 2.2 \times 10^{-16}$). In support, we found that plus strand TCs (regardless of DHS class) positioned within 500–1000 bp from minus strand gene TSSs were more frequently associated with unidirectional than divergent transcription (evaluated by the frequency of divergent events within 500 bp), while the divergent fraction increased at distances >1000 bp (Figure 4B). In contrast, at distances below 500 bp most transcription events were divergent. Hence, PROMPT transcription seems to be impeded when the promoter is placed in close proximity (within 1000 bp) with other gene TSSs in a head-to-head orientation. However, a considerable fraction of *unidirectional stable* DHS could not be explained solely by distance constraints (~40% of such DHSs are >2,000 bp from upstream head-to-head gene TSSs).

At distances >2000 bp to upstream head-to-head gene TSSs, we observed notable sequence differences between *unidirectional stable* DHSs compared to those with PROMPTs (Figure 4C). In particular, Ohler1 and DRE core promoter elements on the major strand were more frequently associated with unidirectional DHSs than those

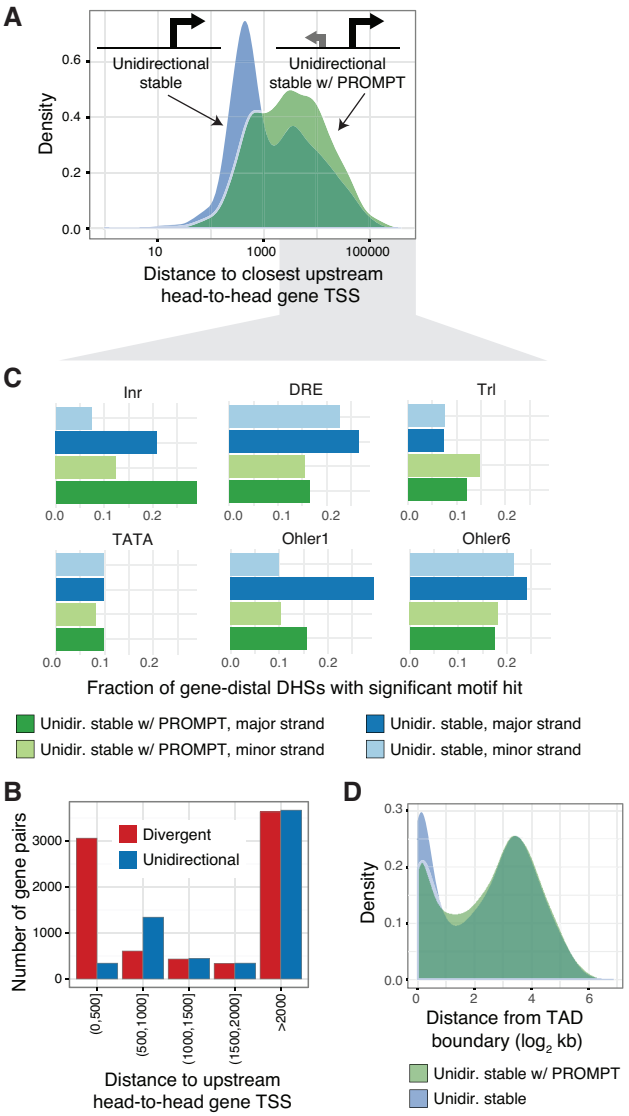

Figure 4. PROMPT transcription is impeded by positional and core promoter element constraints. (**A**) Densities of the distances between DHS major strand CAGE summits and the closest upstream antisense FlyBase gene TSS (head-to-head composition) for *unidirectional stable* and *unidirectional stable w/ PROMPT* DHSs. (**B**) The number of divergent and unidirectional events (vertical axis) for CAGE TCs at various distances from the closest upstream antisense FlyBase gene TSS (head-to-head composition). Divergent events were defined as divergent TC summits within 500 bp. (**C**) Fraction of *unidirectional stable* and *unidirectional stable w/ PROMPT* DHSs positioned >2,000 bp from the closest upstream antisense FlyBase gene TSS having core promoter elements Inr, DRE, Trl, TATA, Ohler1, and Ohler6 on major and minor strands. (**D**) Densities (vertical axis) of distances between transcribed DHSs to TAD boundaries (horizontal axis) for *unidirectional stable* and *unidirectional stable w/ PROMPT* DHSs.

with PROMPTs (Fisher's exact test, Ohler1: $P < 3.5 \times 10^{-16}$, DRE: $P < 4.3 \times 10^{-8}$). Overall, core promoter clusters (Figure 3F) defined by these elements (clusters 5 (Ohler1) and 8 (DRE)) were associated with a higher transcriptional directionality score (Supplementary Figure S13C). In contrast, Trl element occurrence was associated with PROMPT transcription (Figure 4C) and the core pro-

moter cluster (cluster 9) most strongly associated with this element displayed the weakest transcriptional directionality scores (Supplementary Figure S13C). Interestingly, both Ohler1 and DRE elements are associated with broader, ubiquitous expression (56), while Trl elements do associate with regulated, cell type-constrained gene expression (36). In addition, it is known that DRE and Trl elements have different positional preferences in chromatin architectures, with DRE elements frequently co-occurring with housekeeping TREs at TAD boundaries (39). In agreement, *unidirectional stable* DHSs were more frequently positioned in the vicinity of TAD boundaries (39) than *unidirectional stable w/ PROMPTs* DHSs (Figure 4D) (Fisher's exact test, $P = 1.297 \times 10^{-9}$).

In summary, the prevalence of divergent transcription in *D. melanogaster* may be impeded by constraints on core promoter element composition, genomic positioning, and proximal chromatin architectures. Since many of these features differ in frequencies from mammalian genomes, we suggest that these characteristics explain the lower tendency of divergent transcription at *D. melanogaster* gene promoters.

Enhancer potential is related to endogenous expression level

Next, we investigated the association between chromatin state and DHS class. We overlaid transcribed DHSs with locations of histone modifications, histone variants, and TF binding sites (binarised modENCODE (61) ChIP-chip data (51)), and calculated the per-class binding proportions within 100 bp of the major strand CAGE summit (Figure 5A). Several chromatin marks clearly distinguished *unstable* from *stable* DHSs, including H3K4me1, H3K18ac, H4K8ac, H4K5ac, and H3K27ac at *unstable* DHSs, and H2AV and H3K4me3 at *stable* DHSs. Other histone modifications did not specifically follow the inferred stability classes. H3K9ac was mostly found at *bidirectional stable* DHSs, while H2BK5ac displayed a promiscuous association with transcribed DHSs. We observed preferential GAF binding to *unstable* DHSs and to some extent also *unidirectional stable w/ PROMPT* DHSs, confirming the observed Trl element enrichment in these DHS classes (Figure 3F, G). Interestingly, architectural proteins frequently residing at chromatin domain boundaries (38,62), such as CTCF, BEAF32, CP190, Chriz (Chromator) and its associated kinase JIL1, frequently overlapped *stable* DHSs.

Although H3K4me1 and H3K4me3 showed preferential overlaps with *unstable* and *stable* DHSs in agreement with preferential gene TSS-distal enhancer associations (Figure 2C), we wanted to investigate their association with STARR-seq enhancers regardless of DHS class. First, we split *bidirectional weak unstable* and *unidirectional stable w/ PROMPT* DHSs according to whether they overlapped with a hkCP enhancer or a dCP enhancer and plotted enrichments of H3K4me3 and H3K4me1 along with two other chromatin marks distinguishing *stable* from *unstable* DHSs, namely H3K18ac and H2AV (Figure 5B, Supplementary Figure S14). For all four marks, we observed a binding profile which appeared remarkably similar according to the overlapped enhancer class, with binding reflecting the hkCP or dCP enhancer potential as opposed to the DHS class itself. H3K4me1 and H3K18ac were both depleted at the center of DHSs for both classes when overlapping with hkCP enhancers, whilst enriched for both classes when overlapping dCP enhancers. In contrast, H2AV and H3K4me3 overlaps were frequent in cases where either DHS class overlapped hkCP enhancers and showed a much reduced frequency around DHSs overlapping dCP enhancers. Interestingly, based on ChIP-seq data (52), we observed quantitative relationships (Spearman rank correlation test, $P < 2.2 \times 10^{-16}$) between H3K4me3 and H3K4me1 levels and enhancer strengths (Figure 5C, D). In line with preferential enrichments in DHS classes and enhancer classes, H3K4me3 displayed a positive correlation with hkCP enhancer potential (Spearman's rho = 0.49) and a weaker negative correlation with dCP enhancer strength (Spearman's rho = –0.26), while H3K4me1 showed the opposite trends. Enhancer strength associations were also strong for the ratio between H3K4me1 and H3K4me3, but interestingly not for H3K27ac (Supplementary Figure S15). Thus, it appears that H3K4me1 and H3K4me3 levels, as well as the ratio between these, reflect the underlying DNA sequence and in particular the ability of the sequence to act as an enhancer.

However, H3K4me3 and H3K4me1 are both associated with transcriptional levels (4), and likely reflect transcriptional memory and consistency between cells (63). Congruently, H3K4me1:H3K4me3 ratio and H3K4me1 levels (Supplementary Files 3 and 4) were negatively associated with endogenous CAGE expression levels (Supplementary Figure S15, Spearman rank correlation test, $P < 2.2 \times 10^{-16}$, rho = –0.46). In line with this observation, dCP enhancers were less expressed than DHSs not associated with dCP enhancer potential (*t*-test, $P < 1 \times 10^{-16}$), while hkCP enhancer-associated transcribed DHSs were associated with the highest expression levels (*t*-test, $P < 1 \times 10^{-16}$). In general, we observed a striking positive correlation between hkCP signal and the major strand CAGE expression level of DHSs regardless of attributed class, indicating that the stronger the housekeeping enhancer potential, the more transcription is observed from the DHS (Figure 6A, Spearman rank correlation test, $P < 2.2 \times 10^{-16}$, rho = 0.43). These results argue that the observed chromatin mark enrichment over DHS clusters and enhancer classes might reflect local core promoter strength. Indeed, the strength of DRE elements (as determined by motif match score) was positively correlated with hkCP enhancer potential (Supplementary Figure S16), which has been reported earlier (36). Hence, the enhancer potential to activate hkCP core promoters is related to core promoter strength (as observed for DRE), which is itself biased toward *stable* DHSs (Figure 3F, G). In contrast, endogenous expression levels were lowest for DHSs with the strongest dCP enhancer potential (Figure 6B, Spearman rank correlation test, $P < 2.2 \times 10^{-16}$, rho = –0.20), suggesting that dCP enhancer function is incompatible with strong promoter function. Importantly, the overall trends observed for H3K4me1, H3K4me3 and CAGE expression levels versus housekeeping enhancer potential were consistent both for DHSs that were proximal and those that were distal to FlyBase gene TSSs (Supplementary Figure S17). In addition, gene TSS-proximal DHSs with strong dCP enhancer poten-

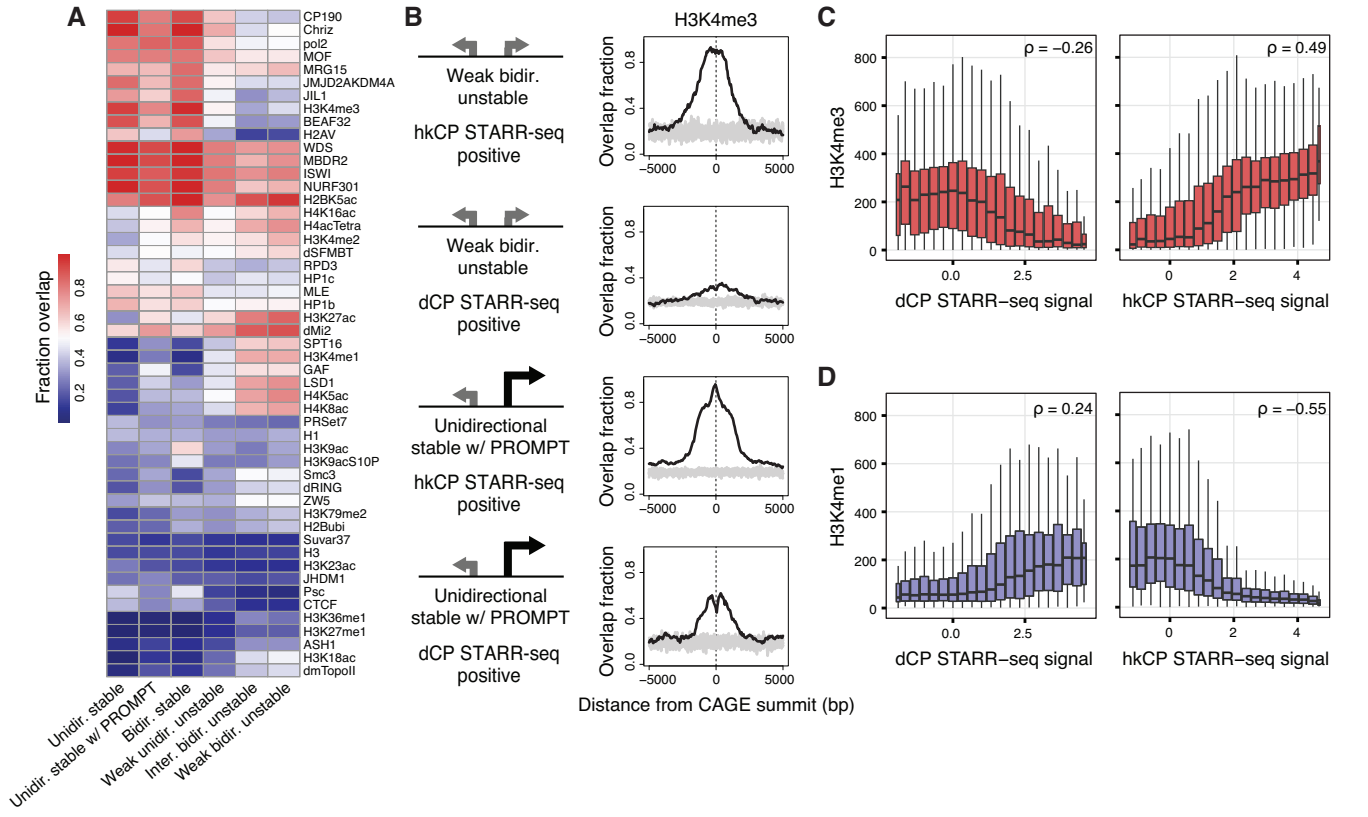

Figure 5. Histone modifications and architectural protein binding reflect enhancer potential. (**A**) Heat map representing DHS class proportions of binarised ChIP-chip data. Rows and columns are hierarchically clustered and binding is defined as at least one binding event observed within ±100 bp of the major strand CAGE summit. (**B**) Detailed binding enrichments for H3K4me3 at *weak bidirectional unstable* and *unidirectional stable w/ PROMPT* DHSs, broken up according to STARR-seq enhancer potential (overlapping either a hkCP or dCP enhancer), based on binding proportions within 5,000 bp from the CAGE summit. Grey represents background distribution based on randomized locations, generated 10 times per plot. See also Supplementary Figure S14 for the profiles of H3K4me1, H3K18ac, H2AV. (**C**) Normalized H3K4me3 ChIP-seq data (vertical axis) versus binned dCP (left) and hkCP (right) STARR-seq signal (horizontal axes). Spearman's rho statistics calculated on non-binned data are displayed in the top right corners of panels. (**D**) Like C but for normalized H3K4me1 ChIP-seq data. Box-and-whisker plots (C, D) displayed as in Figure 2A.

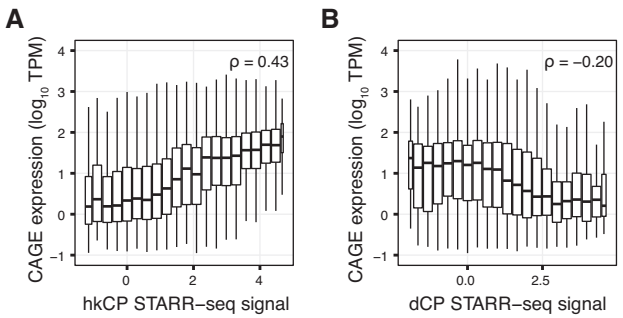

Figure 6. Enhancer potential is related to local endogenous expression levels. (**A**) DHS major strand CAGE expression levels ($\log_{10}$ TPM, vertical axis) versus binned hkCP STARR-seq signal (horizontal axis). (**B**) Like (A), but for binned dCP STARR-seq signal. Spearman's rho statistics calculated on non-binned data are displayed in the top right corners of panels. See also Supplementary Figure S15 for quantitative relationships between expression levels and H3K4me1, H3K4me3, and H3K27ac. Box-and-whisker plots (A, B) displayed as in Figure 2A.

tial tended to be more lowly expressed than those with weak dCP enhancer potential.

In conclusion, our results suggest that TRE housekeeping enhancer strength is reflecting local core promoter strength and thus, endogenous expression levels. The opposite trend for dCP enhancer potential and local expression level further suggest a regulatory trade-off between promoter strength and developmental enhancer strength.

Three dimensional architectures reveal multiple layers of transcriptional regulation

Next, we utilized TAD information based on high resolution Kc167 HiC data (39) to investigate how defined DHS classes in *D. melanogaster*, with respect to RNA exosome-sensitivity, directionality and expression strength, behave within their three-dimensional contexts. The number of transcribed DHSs within TADs ranged from 1 to 31, with a clear skew toward fewer sites and single-element TADs having the greatest frequencies (Supplementary Figure S18A). Interestingly, the number of DHSs within a TAD only very weakly correlated with the size of the TAD in which they belonged (Supplementary Figure S18B), suggesting that the TREs within multi-element TADs are more densely situated than in TADs with fewer TREs. In general, *stable* DHSs were frequently positioned closer to TAD borders, whereas

unstable DHSs were often positioned away from boundaries (Supplementary Figure S19). *Unidirectional stable* DHSs were also more likely to be positioned between annotated TADs as opposed to within ($P < 2.2 \times 10^{-16}$, Fisher's exact test), while on the contrary, *unstable* DHSs were strongly enriched within the TADs themselves ($P < 2.2 \times 10^{-16}$, Fisher's exact test, Figure 7A).

We asked whether TREs co-localized preferentially with other types of TREs within TADs. Considering TADs containing at least three transcribed DHS, we applied generalized linear models (GLMs) to calculate the odds of a partnering TRE of each DHS class appearing within the same TAD (correcting for the distance between TREs within and between TADs and against a background of encountering a random TRE, see Materials and Methods). Overall, the *unidirectional stable* and *bidirectional stable* DHSs showed a preference for co-localization with DHSs of their own classes, and a reduced preference for *unstable* DHSs (Figure 7B). Similarly, *unstable* DHSs showed a clear preference for co-localization with other DHSs of the same category. Interestingly, *unidirectional stable w/ PROMPT* DHSs showed a preference toward grouping with the *unstable* DHSs, thus showing a very different trend to its counterpart DHSs without detected PROMPTs.

To investigate if these results reflect a general property or a consequence of multiple layers of regulatory architectures occurring within the genome of *D. melanogaster*, we clustered the TADs according to their memberships of transcribed DHSs (see Materials and Methods), generating a total of 7 clusters (Supplementary Figure S20). In agreement with the GLM results, TADs that were highly enriched in *unstable* DHSs also contained *unidirectional stable w/ PROMPT* DHSs but very few DHSs classified as *unidirectional stable* or *bidirectional stable* (e.g. clusters 1, 3, 4). In addition, some TADs predominantly contained *stable* TREs (e.g. clusters 6 and 7). The clustering of TAD memberships also revealed other combinations of DHS classes, not apparent from the GLM analysis, e.g. cluster 5, which had a tendency to contain a mixture of classes, in particular combinations of *unidirectional stable* elements with *unstable* DHSs. This cluster might reflect cases in which enhancer-core promoter preferences do not follow our generalized DHS class attributions.

In order to investigate the architectural relationship of DHS classes with respect to enhancer potential, we generated models investigating the context of individual chromatin interactions (39) according to the hkCP or dCP enhancer strengths of target DHSs (utilizing sequences with no STARR-seq enhancer potential as background). Notably, DHSs associated with *stable* classes interacted preferentially with DHSs possessing hkCP enhancer potential, but *unstable* classes did not show such a preference (Supplementary Figure S21), reflecting the strong association between *stable* DHS and local hkCP enhancer potential. Further supporting the separate architectures of dCP and hkCP enhancers and their links with DHS classes, the interaction targets of DHSs possessing dCP enhancer potential were depleted of *stable* DHSs.

Taken together, our results suggest multiple and distinct regulatory architectures for hkCP and dCP enhancers, that can be categorized into two general regulatory programmes. One in which housekeeping gene promoters may act as enhancers of other gene promoters and another in which gene promoters are regulated by gene-distal developmental enhancers, likely affecting cell-type restricted expression levels.

DISCUSSION

In this study we provide an extensive annotation-unbiased characterization of transcriptional regulatory elements in *D. melanogaster*, based on the biogenesis and properties of produced transcripts, including expression levels, transcriptional directionality and RNA exosome-sensitivity. TSS data of capped RNAs in exosome-depleted S2 cells provide clear evidence that a large fraction of TREs in general are divergently transcribed. In particular, the vast majority of active gene-distal transcriptional enhancers are characterized by divergent transcription of exosome-sensitive eRNAs. In addition, a considerable proportion of gene promoters are associated with divergent transcription of stable mRNAs and unstable PROMPTs. These archetypical properties, distinguishing gene-distal enhancers from gene promoters, allow for accurate classification of regulatory function from expression data alone. We further show that similar principles of downstream RNA processing seen for mammalian PROMPTs and eRNAs are linked to exosome sensitivity also in *D. melanogaster*. Our observations support a unified divergent promoter architecture for many TREs (4–6), which is similar across Metazoa. Importantly, the prevalence of divergent transcription at fly enhancers is supported by three recent studies using alternative assays (28,41,42).

Despite the prevalence of divergent transcription at TREs, a fraction of gene promoters in *D. melanogaster* are unidirectionally transcribed. We find that genomic positioning and localization within chromatin architectures might explain some of these exceptions. When a gene TSS is positioned proximal to an upstream gene TSS in a head-to-head conformation, PROMPT transcription is impeded and the distance between divergent events are determined by the distance between paired gene TSSs, lending support to observations in human cells (60). Interestingly, divergent gene pair occurrences in invertebrates are much more frequent than what can be explained by their constrained genome size alone (59). Positioning with respect to TAD structures also seems to have an effect, with unidirectionally transcribed TREs more frequently positioned in close proximity with TAD boundaries than divergent ones. Binding of upstream architectural proteins also seems to have an effect on divergent transcription in human cells (64), potentially acting as barriers to elongating RNAPIIs. These features may, at least partially, explain the previously claimed lack of divergent transcription at *D. melanogaster* gene promoters (26–28).

Systematic characterization of core promoter elements at TREs revealed a complex landscape of core promoter compositions. We observed distinct associations of certain core promoter elements to subsets of TREs, which were strongly associated with their expression strength, directionality and RNA stability. Presence of Trl elements is associated with TREs characterized by balanced bidirectional

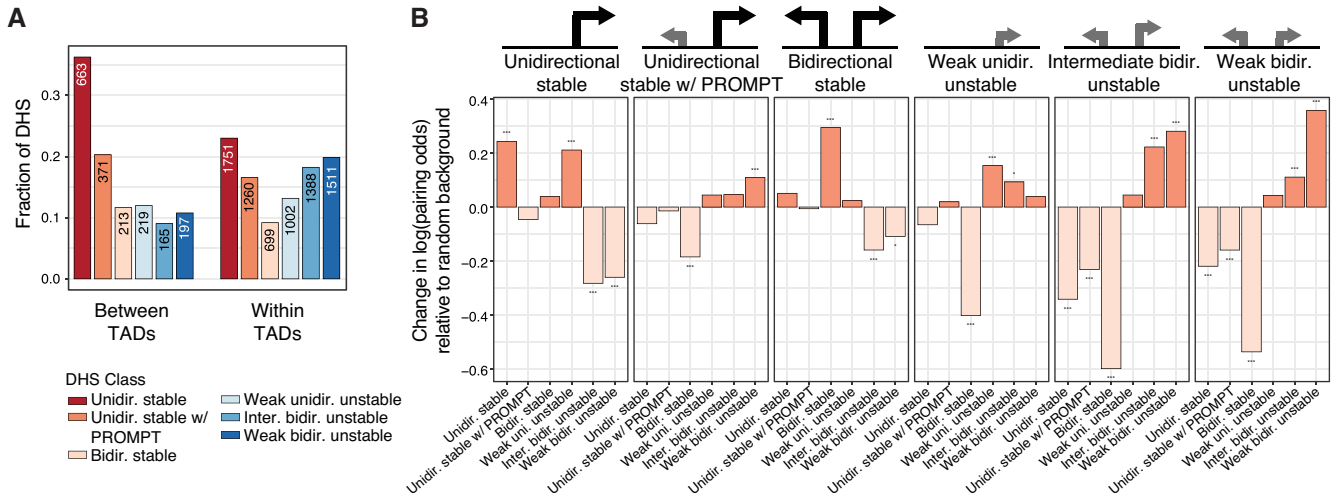

Figure 7. Chromatin architectures suggest multiple layers of transcriptional regulation. (**A**) Fractions of DHSs per class, out of either the total within or between annotated TADs. The number of DHSs in each class is denoted on top of bars. (**B**) The change in log(odds) of grouping within the same TAD for DHS classes, split according to DHS class. Significance stars interpreted as: $*P < 0.1$, $**P < 0.01$ or $***P < 0.001$.

transcription of unstable RNAs, thereby providing a signature of many gene-distal enhancers. Other elements, including Ohler1 and DRE, are associated with directionality of transcription. Since Ohler1 and DRE are specific to invertebrates (56), their enrichments provide an additional explanation of the reduced prevalence of PROMPTs at *D. melanogaster* gene promoters compared to mammals.

Integration of enhancer potential data, as measured using STARR-seq assays, provided clear insights into the relationship between transcriptional properties of TREs and the link between promoter and enhancer function. Gene promoter-like loci had a tendency to overlap housekeeping enhancers, as reported previously (36). Interestingly, we found that housekeeping enhancer strength is related with endogenous CAGE expression levels. This is potentially driven by core promoter element strength, in particular for DRE elements, which itself is correlated with housekeeping enhancer strength. Thus, the stronger the promoter, the greater potential it has to act as a housekeeping enhancer. In contrast, strong developmental enhancers are associated with weak promoter expression levels. Developmental enhancer strength was associated with low endogenous expression levels, suggesting a regulatory trade-off between promoter function and developmental enhancer function. Our identified link between enhancer function, core promoter strength and promoter expression level provides insights into frequently used histone modifications to discern enhancers from gene promoters, including H3K4me1 and H3K4me3, which we find to be chiefly related to expression levels, as observed elsewhere (4,42). Such histone modifications are therefore likely indirect markers of enhancer function, reflecting the generally weaker promoter strengths and thereby expression levels of gene-distal enhancers. However, although H3K4me1 tended to be prevalent at lowly expressed developmental enhancers, it is important to note that H3K4me1 on its own cannot distinguish between active and inactive enhancers (65) and that the relationship between enhancer strength and expression level (and thereby H3K4me1 levels) may differ between groups of TREs (42).

In line with observed differences between developmental and housekeeping enhancers, chromatin conformation data suggest a model involving separate architectures of transcriptional regulation, in which TREs are strongly biased to interact with those with the same transcriptional properties and regulatory potentials. Gene-distal developmental enhancers are enriched within TADs and gene promoters with housekeeping enhancer potential are enriched near TAD borders, reflecting the constraints chromatin architecture can have on transcriptional activity and regulation. Our results suggest at least two distinct regulatory programmes for housekeeping and developmental enhancers. For developmental or cell type-restricted regulation, gene-distal enhancers seem to regulate gene promoters with cell-type restricted expression levels constrained within the same TAD. Housekeeping gene promoters, on the other hand, are frequently located close to TAD borders, and may act as enhancers to other gene promoters alike.

Importantly, our study implies that many enhancers (defined as DHSs with STARR-seq activity) are RNAPII promoters. Our findings also agree with recent observations (41,42) that regulatory function might not be discernible on a per-element basis. Rather, a fraction of metazoan TREs possesses both strong enhancer and strong promoter function, while others are characterized by strong enhancer function and weak promoter function or vice versa. Based on these observations we favor the most parsimonious model, in which TREs (classically labeled as enhancers or promoters) should be referred to as promoters, whose regulatory activities, effects (local promoter and/or distal enhancer) and strengths are determined by local core promoter strength and the genomic landscape and chromatin architecture in which they are placed.

DATA AVAILABILITY

Data generated in this study have been deposited in GEO under accession number GSE109588.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

We thank Neus Visa for advice and sharing of reagents for the exosome knockdowns.
Author contributions: R.A. and T.H.J. conceived the project. M.L.-L. and C.D.V. conducted the experiments. M.L.-L. performed the experimental validations. S.R., M.D., S.B. and R.A. analyzed the data. S.R., M.D. and R.A. wrote the manuscript with input from all authors. All authors reviewed the final manuscript.

FUNDING

R.A. laboratory was supported by the Danish Council for Independent Research [6108-00038B]; European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [638173]; T.H.J. laboratory was supported by the ERC [339953]; Danish National Research Council; Danish Council for Independent Research [1333-00059B to M.L.L.]. Funding for open access charge: European Research Council [638173].
Conflict of interest statement. None declared.

REFERENCES

1. Heintzman,N.D., Stuart,R.K., Hon,G., Fu,Y., Ching,C.W., Hawkins,R.D., Barrera,L.O., Van Calcar,S., Qu,C., Ching,K.A. *et al.* (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.*, **39**, 311–318.
2. Heintzman,N.D., Hon,G.C., Hawkins,R.D., Kheradpour,P., Stark,A., Harp,L.F., Ye,Z., Lee,L.K., Stuart,R.K., Ching,C.W. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, **459**, 108–112.
3. Creyghton,M.P., Cheng,A.W., Welstead,G.G., Kooistra,T., Carey,B.W., Steine,E.J., Hanna,J., Lodato,M.A., Frampton,G.M., Sharp,P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21931–21936.
4. Core,L.J., Martins,A.L., Danko,C.G., Waters,C.T., Siepel,A. and Lis,J.T. (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.*, **46**, 1311–1320.
5. Andersson,R., Sandelin,A. and Danko,C.G. (2015) A unified architecture of transcriptional regulatory elements. *Trends Genet.*, **31**, 426–433.
6. Kim,T.-K. and Shiekhattar,R. (2015) Architectural and functional commonalities between enhancers and promoters. *Cell*, **162**, 948–959.
7. Andersson,R. (2015) Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays*, **37**, 314–323.
8. Seila,A.C., Calabrese,J.M., Levine,S.S., Yeo,G.W., Rahl,P.B., Flynn,R.A., Young,R.A. and Sharp,P.A. (2008) Divergent transcription from active promoters. *Science (New York, N.Y.)*, **322**, 1849–1851.
9. Preker,P., Nielsen,J., Kammler,S., Lykke-Andersen,S., Christensen,M.S., Mapendano,C.K., Schierup,M.H. and Jensen,T.H. (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science (New York, N.Y.)*, **322**, 1851–1854.
10. Andersson,R., Refsing Andersen,P., Valen,E., Core,L.J., Bornholdt,J., Boyd,M., Heick Jensen,T. and Sandelin,A. (2014) Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat. Commun.*, **5**, 5336.
11. Andersson,R., Chen,Y., Core,L., Lis,J.T., Sandelin,A. and Jensen,T.H. (2015) Human gene promoters are intrinsically bidirectional. *Mol. Cell*, **60**, 346–347.
12. Sigova,A.A., Mullen,A.C., Molinie,B., Gupta,S., Orlando,D.A., Guenther,M.G., Almada,A.E., Lin,C., Sharp,P.A., Giallourakis,C.C. *et al.* (2013) Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 2876–2881.
13. Andersson,R., Gebhard,C., Miguel-Escalada,I., Hoof,I., Bornholdt,J., Boyd,M., Chen,Y., Zhao,X., Schmidl,C., Suzuki,T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455–461.
14. Kim,T.-K., Hemberg,M., Gray,J.M., Costa,A.M., Bear,D.M., Wu,J., Harmin,D.A., Laptewicz,M., Barbara-Haley,K., Kuersten,S. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**, 182–187.
15. Danko,C.G., Hyland,S.L., Core,L.J., Martins,A.L., Waters,C.T., Lee,H.W., Cheung,V.G., Kraus,W.L., Lis,J.T. and Siepel,A. (2015) Identification of active transcriptional regulatory elements from GRO-seq data. *Nat. Methods*, **12**, 433–438.
16. Koch,F., Fenouil,R., Gut,M., Cauchy,P., Albert,T.K., Zacarias-Cabeza,J., Spicuglia,S., de la Chapelle,A.L., Heidemann,M., Hintermair,C. *et al.* (2011) Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.*, **18**, 956–963.
17. Liu,Z., Scannell,D.R., Eisen,M.B. and Tjian,R. (2011) Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell*, **146**, 720–731.
18. Zhou,H., Kaplan,T., Li,Y., Grubisic,I., Zhang,Z., Wang,P.J., Eisen,M.B. and Tjian,R. (2013) Dual functions of TAF7L in adipocyte differentiation. *eLife*, **2**, 190.
19. Kowalczyk,M.S., Hughes,J.R., Garrick,D., Lynch,M.D., Sharpe,J.A., Sloane-Stanley,J.A., McGowan,S.J., De Gobbi,M., Hosseini,M., Vernimmen,D. *et al.* (2012) Intragenic enhancers act as alternative promoters. *Mol. Cell*, **45**, 447–458.
20. Li,G., Ruan,X., Auerbach,R.K., Sandhu,K.S., Zheng,M., Wang,P., Poh,H.M., Goh,Y., Lim,J., Zhang,J. *et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, **148**, 84–98.
21. Chepelev,I., Wei,G., Wangsa,D., Tang,Q. and Zhao,K. (2012) Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.*, **22**, 490–503.
22. Ghavi-Helm,Y., Klein,F.A., Pakozdi,T., Ciglar,L., Noordermeer,D., Huber,W. and Furlong,E.E.M. (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*, **512**, 96–100.
23. Dao,L. T.M., Galindo-Albarrán,A.O., Castro-Mondragon,J.A., Andrieu-Soler,C., Medina-Rivera,A., Souaid,C., Charbonnier,G., Griffon,A., Vanhille,L., Stephen,T. *et al.* (2017) Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.*, **49**, 1073–1081.
24. Leung,D., Jung,I., Rajagopal,N., Schmitt,A., Selvaraj,S., Lee,A.Y., Yen,C.-A., Lin,S., Lin,Y., Qiu,Y. *et al.* (2015) Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature*, **518**, 350–354.
25. Engreitz,J.M., Haines,J.E., Perez,E.M., Munson,G., Chen,J., Kane,M., McDonel,P.E., Guttman,M. and Lander,E.S. (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, **539**, 452–455.
26. Nechaev,S., Fargo,D.C., dos Santos,G., Liu,L., Gao,Y. and Adelman,K. (2010) Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science (New York, N.Y.)*, **327**, 335–338.
27. Core,L.J., Waterfall,J.J., Gilchrist,D.A., Fargo,D.C., Kwak,H., Adelman,K. and Lis,J.T. (2012) Defining the status of RNA polymerase at promoters. *Cell Rep.*, **2**, 1025–1035.
28. Meers,M.P., Adelman,K., Duronio,R.J., Strahl,B.D., McKay,D.J. and Matera,A.G. (2018) Transcription start site profiling uncovers

divergent transcription and enhancer-associated RNAs in Drosophila melanogaster. *BMC Genomics*, **19**, 157.
29. Duttke,S. H.C., Lacadie,S.A., Ibrahim,M.M., Glass,C.K., Corcoran,D.L., Benner,C., Heinz,S., Kadonaga,J.T. and Ohler,U. (2015) Human promoters are intrinsically directional. *Mol. Cell*, **57**, 674–684.
30. Scruggs,B.S., Gilchrist,D.A., Nechaev,S., Muse,G.W., Burkholder,A., Fargo,D.C. and Adelman,K. (2015) Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol. Cell*, **58**, 1101–1112.
31. Ntini,E., Järvelin,A.I., Bornholdt,J., Chen,Y., Boyd,M., Jørgensen,M., Andersson,R., Hoof,I., Schein,A., Andersen,P.R. *et al.* (2013) Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.*, **20**, 923–928.
32. Andersen,P.R., Domanski,M., Kristiansen,M.S., Storvall,H., Ntini,E., Verheggen,C., Schein,A., Bunkenborg,J., Poser,I., Hallais,M. *et al.* (2013) The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Struct. Mol. Biol.*, **20**, 1367–1376.
33. Almada,A.E., Wu,X., Kriz,A.J., Burge,C.B. and Sharp,P.A. (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*, **499**, 360–363.
34. Marques,A.C., Hughes,J., Graham,B., Kowalczyk,M.S., Higgs,D.R. and Ponting,C.P. (2013) Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biol.*, **14**, R131.
35. Arnold,C.D., Gerlach,D., Stelzer,C., Boryn,L.M., Rath,M. and Stark,A. (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science (New York, N.Y.)*, **339**, 1074–1077.
36. Zabidi,M.A., Arnold,C.D., Schernhuber,K., Pagani,M., Rath,M., Frank,O. and Stark,A. (2015) Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, **518**, 556–559.
37. Corrales,M., Rosado,A., Cortini,R., van Arensbergen,J., van Steensel,B. and Filion,G.J. (2017) Clustering of Drosophila housekeeping promoters facilitates their expression. *Genome Res.*, **27**, 1153–1161.
38. Sexton,T., Yaffe,E., Kenigsberg,E., Bantignies,F., Leblanc,B., Hoichman,M., Parrinello,H., Tanay,A. and Cavalli,G. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, **148**, 458–472.
39. Cubeñas-Potts,C., Rowley,M.J., Lyu,X., Li,G., Lei,E.P. and Corces,V.G. (2017) Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.*, **45**, 1714–1730.
40. Hug,C.B., Grimaldi,A.G., Kruse,K. and Vaquerizas,J.M. (2017) Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell*, **169**, 216–228.
41. Mikhaylichenko,O., Bondarenko,V., Harnett,D., Schor,I.E., Males,M., Viales,R.R. and Furlong,E.E.M. (2018) The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev.*, **32**, 42–57.
42. Henriques,T., Scruggs,B.S., Inouye,M.O., Muse,G.W., Williams,L.H., Burkholder,A.B., Lavender,C.A., Fargo,D.C. and Adelman,K. (2018) Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.*, **32**, 26–41.
43. Takahashi,H., Lassmann,T., Murata,M. and Carninci,P. (2012) 5' end-centered expression profiling using Cap-analysis gene expression (CAGE) and next-generation sequencing. *Nat. Protoc.*, **7**, 542–561.
44. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
45. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, **30**, 2114–2120.
46. John,S., Sabo,P.J., Thurman,R.E., Sung,M.-H., Biddie,S.C., Johnson,T.A., Hager,G.L. and Stamatoyannopoulos,J.A. (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.*, **43**, 264–268.
47. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
48. Mathelier,A., Fornes,O., Arenillas,D.J., Chen,C.-Y., Denay,G., Lee,J., Shi,W., Shyr,C., Tan,G., Worsley-Hunt,R. *et al.* (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **44**, D110–D115.
49. Kulakovskiy,I.V., Favorov,A.V. and Makeev,V.J. (2009) Motif discovery and motif finding from genome-mapped DNase footprint data. *Bioinformatics*, **25**, 2318–2325.
50. Ohler,U., Liao,G.-c., Niemann,H. and Rubin,G.M. (2002) Computational analysis of core promoters in the Drosophila genome. *Genome Biol.*, **3**, RESEARCH0087.
51. Zhou,J. and Troyanskaya,O.G. (2016) Probabilistic modelling of chromatin code landscape reveals functional diversity of enhancer-like chromatin states. *Nat. Commun.*, **7**, 10528.
52. Herz,H.-M., Mohan,M., Garruss,A.S., Liang,K., Takahashi,Y.-h., Mickey,K., Voets,O., Verrijzer,C.P. and Shilatifard,A. (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes Dev.*, **26**, 2604–2620.
53. Gramates,L.S., Marygold,S.J., Santos,G.d., Urbano,J.-M., Antonazzo,G., Matthews,B.B., Rey,A.J., Tabone,C.J., Crosby,M.A., Emmert,D.B. *et al.* (2017) FlyBase at 25: looking to the future. *Nucleic Acids Res.*, **45**, D663–D671.
54. Lim,S.J., Boyle,P.J., Chinen,M., Dale,R.K. and Lei,E.P. (2013) Genome-wide localization of exosome components to active promoters and chromatin insulators in Drosophila. *Nucleic Acids Res.*, **41**, 2963–2980.
55. Guo,J., Garrett,M., Micklem,G. and Brogna,S. (2011) Poly(A) signals located near the 5′ end of genes are silenced by a general mechanism that prevents premature 3′-end processing. *Mol. Cell. Biol.*, **31**, 639–651.
56. Lenhard,B., Sandelin,A. and Carninci,P. (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.*, **13**, 233–245.
57. Kadonaga,J.T. (2011) Perspectives on the RNA polymerase II core promoter. *Wiley interdiscipl. Rev. Dev. Biol.*, **1**, 40–51.
58. Rach,E.A., Yuan,H.-Y., Majoros,W.H., Tomancak,P. and Ohler,U. (2009) Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. *Genome biology*, **10**, R73.
59. Yang,L. and Yu,J. (2009) A comparative analysis of divergently-paired genes (DPGs) among Drosophila and vertebrate genomes. *BMC Evol. Biol.*, **9**, 55.
60. Chen,Y., Pai,A.A., Herudek,J., Lubas,M., Meola,N., Järvelin,A.I., Andersson,R., Pelechano,V., Steinmetz,L.M., Jensen,T.H. *et al.* (2016) Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. *Nat. Genet.*, **48**, 984–994.
61. modENCODE Consortium, Roy,S., Ernst,J., Kharchenko,P.V., Kheradpour,P., Nègre,N., Eaton,M.L., Landolin,J.M., Bristow,C.A., Ma,L. *et al.* (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science (New York, N.Y.)*, **330**, 1787–1797.
62. Wang,Q., Sun,Q., Czajkowsky,D.M. and Shao,Z. (2018) Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells. *Nat.Commun.*, **9**, 188.
63. Howe,F.S., Fischl,H., Murray,S.C. and Mellor,J. (2017) Is H3K4me3 instructive for transcription activation? *Bioessays*, **39**, 1–12.
64. Bornelöv,S., Komorowski,J. and Wadelius,C. (2015) Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. *BMC Genomics*, **16**, 300.
65. Bonn,S., Zinzen,R.P., Girardot,C., Gustafson,E.H., Perez-Gonzalez,A., Delhomme,N., Ghavi-Helm,Y., Wilczyński,B., Riddell,A. and Furlong,E. E.M. (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.*, **44**, 148–156.