LETTER

doi:10.1038/nature13994

Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation

Muhammad A. Zabidi[1]*, Cosmas D. Arnold[1]*, Katharina Schernhuber[1], Michaela Pagani[1], Martina Rath[1], Olga Frank[1] & Alexander Stark[1]

Gene transcription in animals involves the assembly of RNA polymerase II at core promoters and its cell-type-specific activation by enhancers that can be located more distally[1]. However, how ubiquitous expression of housekeeping genes is achieved has been less clear. In particular, it is unknown whether ubiquitously active enhancers exist and how developmental and housekeeping gene regulation is separated. An attractive hypothesis is that different core promoters might exhibit an intrinsic specificity to certain enhancers[2–6]. This is conceivable, as various core promoter sequence elements are differentially distributed between genes of different functions[7], including elements that are predominantly found at either developmentally regulated or at housekeeping genes[8–10]. Here we show that thousands of enhancers in *Drosophila melanogaster* S2 and ovarian somatic cells (OSCs) exhibit a marked specificity to one of two core promoters—one derived from a ubiquitously expressed ribosomal protein gene and another from a developmentally regulated transcription factor—and confirm the existence of these two classes for five additional core promoters from genes with diverse functions. Housekeeping enhancers are active across the two cell types, while developmental enhancers exhibit strong cell-type specificity. Both enhancer classes differ in their genomic distribution, the functions of neighbouring genes, and the core promoter elements of these neighbouring genes. In addition, we identify two transcription factors—Dref and Trl—that bind and activate housekeeping versus developmental enhancers, respectively. Our results provide evidence for a sequence-encoded enhancer–core-promoter specificity that separates developmental and housekeeping gene regulatory programs for thousands of enhancers and their target genes across the entire genome.

We chose the core promoter of *Ribosomal protein gene 12* (*RpS12*) and a synthetic core promoter derived from the *even skipped* transcription factor[11] as representative 'housekeeping' and 'developmental' core promoters, respectively (hereafter termed hkCP and dCP; Fig. 1a and Extended Data Figs 1, 2) and tested the ability of all candidate enhancers genome wide to activate transcription from these core promoters using self-transcribing active regulatory region sequencing (STARR-seq)[12] in *D. melanogaster* S2 cells. This set-up allows the testing of all candidates in a defined sequence environment, which differs only in the core promoter sequences but is otherwise constant[12,13].

Two hkCP STARR-seq replicates were highly similar (genome-wide Pearson correlation coefficient (PCC) 0.98; Extended Data Fig. 1c) and yielded 5,956 enhancers, compared with 5,408 enhancers obtained when we reanalysed dCP STARR-seq data[12] (Supplementary Table 1). Interestingly, the hkCP and dCP enhancers were largely non-overlapping (Fig. 1b, c) and the genome-wide enhancer activity profiles differed (PCC 0.38), as did the individual enhancer strengths: of the 11,364 enhancers, 8,144 (72%) activated one core promoter at least twofold more strongly than the other, a difference rarely seen in the replicate experiments for each of the core promoters (Fig. 1d). Indeed, 21 out of 24 hkCP-specific enhancers activated luciferase expression (>1.5-fold and *t*-test $P < 0.05$) from the hkCP versus 1 out of 24 from the dCP (Fig. 1e and Extended Data Fig. 3). Consistently, 10 out of 12 dCP-specific enhancers were positive with the dCP but only 2 out of 12 with the hkCP, a highly significant difference ($P = 5.1 \times 10^{-6}$, Fischer's exact test) that confirms the enhancer–core-promoter specificity observed for thousands of enhancers across the entire genome.

Enhancers that were specific to either the hkCP or the dCP showed markedly different genomic distributions (Fig. 2a and Extended Data

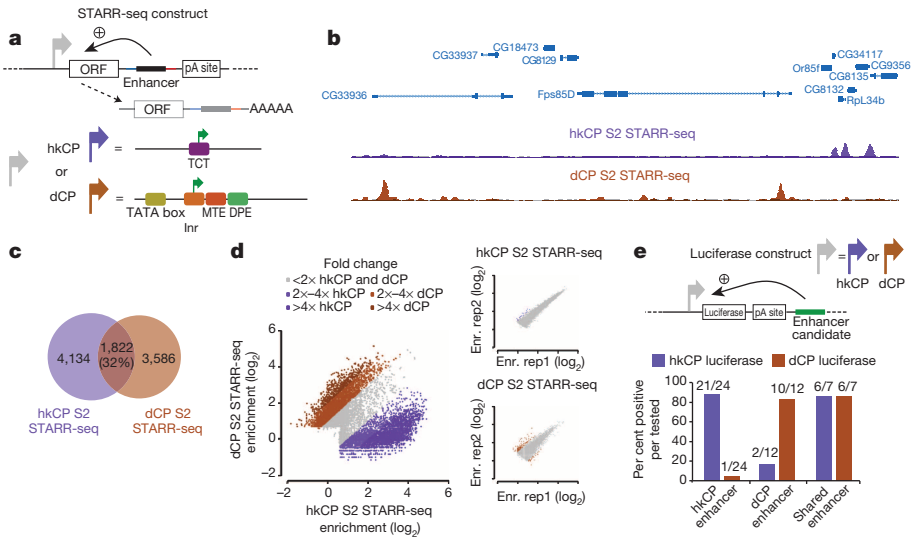


Figure 1 | Distinct sets of enhancers activate transcription from the hkCP and dCP in S2 cells. **a**, STARR-seq set-up using the hkCP housekeeping (*RpS12*; purple) and dCP developmental core promoters (*Drosophila* synthetic core promoter (DSCP)[11]; brown) **b**, Genome browser screenshot depicting STARR-seq tracks for both core promoters. **c**, Overlap of hkCP and dCP enhancers. **d**, hkCP versus dCP STARR-seq enrichments at enhancers (insets show enrichment for replicates (Enr. rep) 1 versus 2 for hkCP and dCP; dCP data reanalysed from ref. 12). **e**, hkCP, dCP or shared enhancers that activate luciferase (>1.5-fold and $P < 0.05$ (one-sided *t*-test); $n = 3$; Extended Data Figs 3 and 5) from hkCP (purple) or dCP (brown; numbers show positive/tested).

[1]Research Institute of Molecular Pathology IMP, Vienna Biocenter VBC, Dr Bohr-Gasse 7, 1030 Vienna, Austria.
*These authors contributed equally to this work.

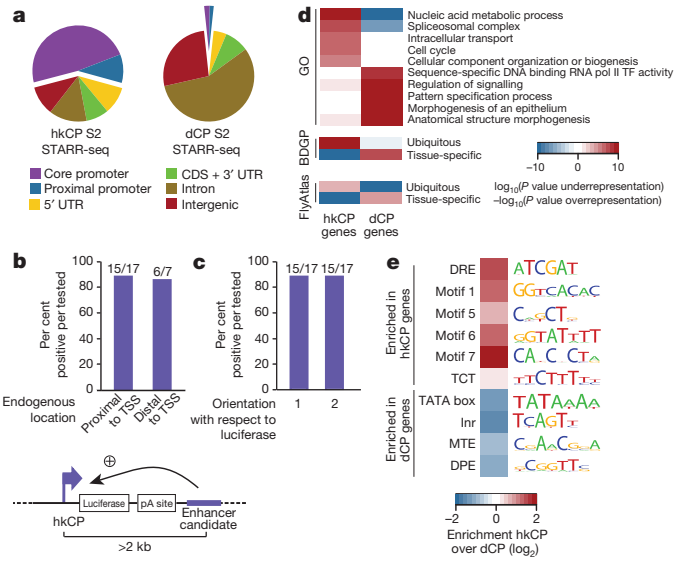


Figure 2 | hkCP and dCP enhancers differ in genomic distribution and flanking genes. **a**, Genomic distribution of hkCP and dCP enhancers. CDS, coding sequence; UTR, untranslated region. **b**, **c**, hkCP enhancers function distally in luciferase assays independent of their genomic positions (**b**) and orientation towards the luciferase TSS (**c**; orientation 1 from **b**; Extended Data Figs 3 and 5). **d**, **e**, GO (5 of the top 100 terms shown per column; Supplementary Table 11) and gene expression (terms curated from the Berkeley *Drosophila* Genome Project (BDGP) and FlyAtlas) analyses (**d**) and enrichment of core promoter elements at TSSs (**e**) for genes next to hkCP and dCP enhancers. TF, transcription factor.

Fig. 4): whereas the majority (58.4%) of hkCP-specific enhancers overlapped with a transcription start site (TSS) or were proximal to a TSS (≤200 bp upstream; Fig. 2a), dCP-specific enhancers located predominantly to introns (56.5%) and intergenic regions (26.9%; Fig. 2a)[12]. Importantly, despite the TSS-proximal location of most hkCP-specific enhancers, they activated transcription from a distal core promoter in STARR-seq (Fig. 1a and Extended Data Figs 1a, 2). Luciferase assays confirmed that they function from a distal position (>2 kb from the TSS) downstream of the luciferase gene and independently of their orientation towards the luciferase TSS (Fig. 2b, c and Extended Data Figs 3, 5). These results show that TSS-proximal sequences can act as bona fide enhancers[14] and that developmental and housekeeping genes are both regulated through core promoters and enhancers, yet with a substantially different fraction of TSS-proximal enhancers (3.4% versus 58.4%).

hkCP and dCP enhancers were also located next to functionally distinct classes of genes according to gene ontology (GO) analyses: genes next to hkCP enhancers were enriched in diverse housekeeping functions including metabolism, RNA processing and the cell cycle, whereas genes next to dCP enhancers were enriched for terms associated with developmental regulation and cell-type-specific functions (Fig. 2d, Extended Data Fig. 6a and Supplementary Tables 2–4). Consistently, hkCP enhancers were preferentially near ubiquitously expressed genes and dCP enhancers were near genes with tissue-specific expression (Fig. 2d and Supplementary Table 5).

The core promoters of the putative endogenous target genes of hkCP and dCP enhancers were also differentially enriched in known core promoter elements[15] (Fig. 2e and Extended Data Fig. 6b): TSSs next to hkCP enhancers were enriched in Ohler motifs[16] 1, 5, 6 and 7, consistent with the ubiquitous expression and housekeeping functions of these genes. In contrast, TSSs next to dCP enhancers were enriched in TATA box, initiator (Inr), motif ten element (MTE) and downstream promoter element (DPE) motifs, which are associated with cell-type-specific gene expression[7,15].

We next investigated whether the specificity that hkCP and dCP show to the two enhancer classes applies more generally. We selected three additional core promoters from housekeeping genes with different functions: from the *eukaryotic translation elongation factor 1δ* (*eEF1δ*), the putative splicing factor *x16*, and the cohesin loader *Nipped-B* (*NipB*). Importantly, all three contained combinations of core promoter elements that differed from that of hkCP, namely TCT[8] and DNA-replication-related element (DRE) motifs (*eEF1δ*), and Ohler motifs 1 and 6 (*x16* and *NipB*; Fig. 3a). In addition, we selected a DPE-containing core promoter of the transcription factor *pannier* (*pnr*) and the TATA-box core promoter of *Heat shock protein 70* (*Hsp70*), which can be activated by tissue-specific enhancers (for example, see ref. 17), thus covering the two most prominent core promoter types of regulated genes[9,16,18].

We performed STARR-seq for the five additional core promoters and grouped the genome-wide enhancer activity profiles of all seven core promoters by hierarchical clustering. This revealed two distinct clusters corresponding to the four housekeeping and the three developmental core promoters, respectively (Fig. 3b, Extended Data Fig. 7 and Supplementary Tables 6, 7), and the core promoters of both clusters indeed responded markedly differentially to individual genomic enhancers (Fig. 3c).

These results obtained for core promoters with diverse motif content and from genes with various functions suggest that the distinct enhancer preferences observed between hkCP and dCP apply more generally and that two broad classes of housekeeping and developmental (or regulated) core promoters exist. Differences within each class might correspond to differences in relative enhancer preferences of the core promoters[2–6], while similarities between both classes could reflect enhancers that are shared (Fig. 1c–e) or core promoters that can be activated to different extents by enhancers from both classes (for example,

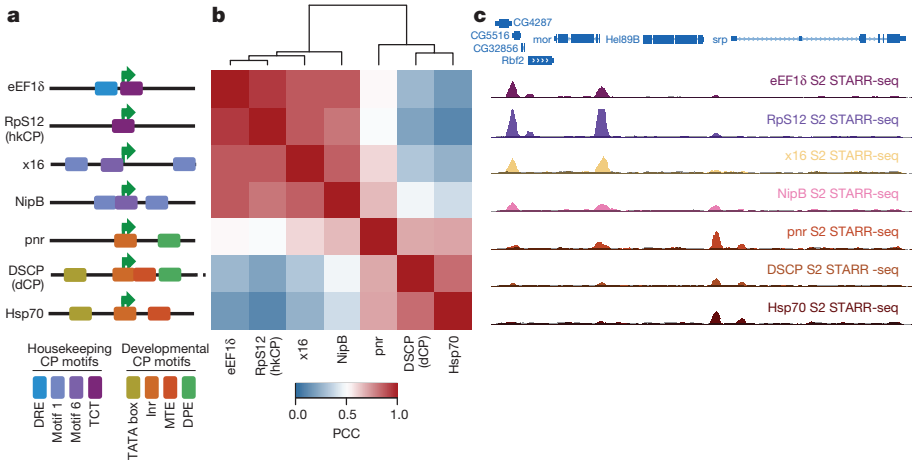


Figure 3 | Housekeeping and developmental core promoters differ characteristically in their enhancer preferences. **a**, Different housekeeping (top 4) and developmental-like (bottom 3) core promoters and their motif content (schematic). **b**, Bi-clustered heat map depicting pairwise similarities of STARR-seq signals (PCCs at peak summits). PCCs and dendrogram (top) show the separation between housekeeping and regulated core promoters. **c**, Genome browser screenshot depicting STARR-seq tracks for all seven core promoters.

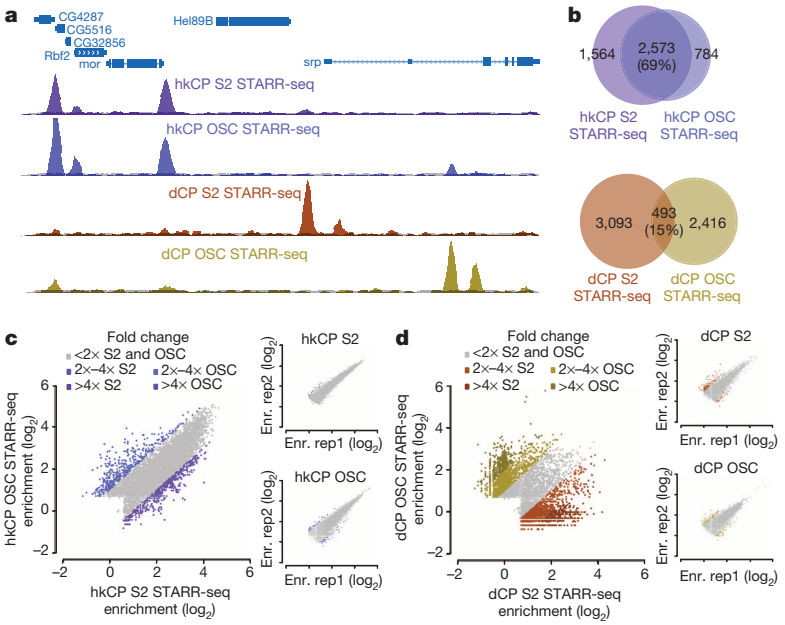


Figure 4 | hkCP enhancers are shared across cell types. **a**, Genome browser screenshot showing tracks for hkCP (top) and dCP STARR-seq (bottom) in S2 cells and OSCs. **b**, Overlap of hkCP (top) and dCP (bottom) enhancers between S2 cells and OSCs. **c**, **d**, hkCP (**c**) and dCP (**d**) STARR-seq enrichments in S2 cells versus OSCs at hkCP- or dCP-specific enhancers (insets show enrichments for replicates (Enr. rep) 1 versus 2; dCP data reanalysed from ref. 12).

NipB; Fig. 3b, c). The latter might be important if broadly expressed housekeeping genes need to be further activated in specific tissues.

To test whether hkCP enhancers function in different cell types, we performed STARR-seq using hkCP in OSCs, which differ strongly from S2 cells in gene expression and dCP enhancer activities[12]. Two hkCP STARR-seq replicates in OSCs were highly similar (PCC 0.97) and yielded 6,217 enhancers (Supplementary Table 1), compared with 5,774 enhancers obtained for dCP data from OSCs[12]. The OSC data confirmed the differences between hkCP and dCP enhancers observed in S2 cells (Extended Data Figs 8, 9 and Supplementary Tables 8–10). Strikingly, the hkCP-specific enhancers in OSCs and S2 cells (3,357 and 4,137, respectively) were almost indistinguishable, whereas dCP-specific enhancers (2,909 in OSCs and 3,586 in S2 cells) differed strongly between the two cell types[12] and from the hkCP enhancers (Fig. 4a). The observation that hkCP enhancers showed similar activities in both cell types while dCP enhancers were cell-type specific was true genome wide when comparing genomic locations (69% versus 15% overlap) or enhancer strengths as measured by STARR-seq (PCC at peak summits 0.83 versus 0.05; Fig. 4b–d and Extended Data Fig. 9c). Together, these results show that hkCP enhancers are shared between two different cell types,

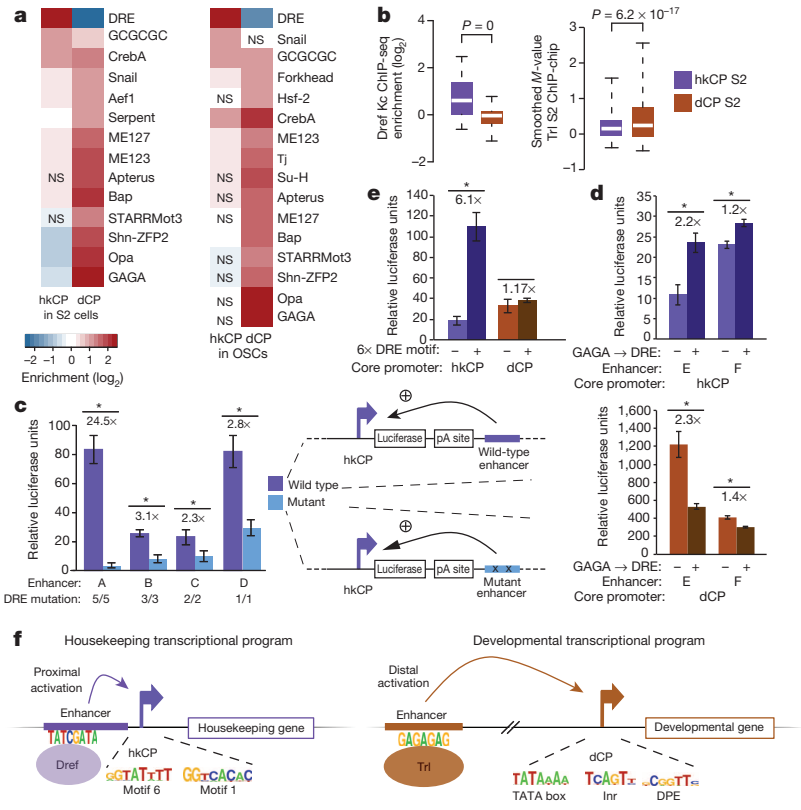


Figure 5 | hkCP and dCP enhancers depend on Dref and Trl, respectively. **a**, **b**, Motif enrichment (**a**) and ChIP signals for Dref and Trl (**b**) in hkCP and dCP enhancers. False discovery rate (FDR)-corrected hypergeometric $P > 0.01$; boxes: median and interquartile range; whiskers: 5th and 95th percentiles; two-sided Wilcoxon-rank-sum P values. NS, not significant. **c**, Luciferase assays for four wild-type and DRE-motif-mutant hkCP enhancers (numbers show mutated motifs). Error bars show standard deviation (s.d.) ($n = 3$, biological replicates). $^*P < 0.005$ (one-sided t-test). **d**, Luciferase assays for two dCP enhancers (−) and their GAGA → DRE-mutant variants (+) with hkCP (top) and dCP (bottom; details as in **c**). **e**, Luciferase assays for an array of DRE motifs with hkCP and dCP (details as in **c**). **f**, Model: housekeeping genes contain Ohler motifs 1, 5, 6, 7 and/or the TCT motif and are activated by TSS-proximal hkCP enhancers via Dref. Regulated genes contain TATA box, Inr, MTE and/or DPE and are activated by distal dCP enhancers via Trl.

whereas dCP enhancers are cell-type specific[12], presumably representing ubiquitous housekeeping versus developmental and cell-type-specific gene expression programs.

To assess whether the marked core promoter specificities of the hkCP and dCP enhancers are encoded in their sequences, we analysed the *cis*-regulatory motif content of both classes of enhancers[19]. This revealed a strong enrichment of the DRE motif in hkCP enhancers (Fig. 5a and Supplementary Tables 11, 12), whereas dCP enhancers were strongly enriched in the GAGA motif of *Trithorax-like* (*Trl*) and other motifs previously described to be important for dCP enhancers[20]. Published genome-wide chromatin immunoprecipitation (ChIP) data[21,22] confirmed that DRE-binding factor (Dref) bound significantly more strongly to hkCP enhancers than to dCP enhancers (Wilcoxon $P = 0$; Fig. 5b), while the opposite was true for Trl (Wilcoxon $P = 6.2 \times 10^{-17}$). Considering only distal enhancers (>500 bp from the closest TSS) yielded the same results (Extended Data Fig. 10a, b and Supplementary Tables 13, 14), suggesting that the differential occupancy is a property of both classes of enhancers rather than a consequence of the different extents to which they overlap with TSSs. Disrupting the DRE motifs in four different hkCP enhancers substantially reduced the activities of the enhancers as measured by luciferase assays in S2 cells (between 2.3- and 24.5-fold reduction; Fig. 5c), while dCP enhancers depend on GAGA motifs[20]. Adding DRE motifs to 11 different dCP enhancers significantly increased luciferase expression from the hkCP for 9 of them (82%; Extended Data Fig. 10c), and changing the GAGA motifs of two dCP enhancers to DRE motifs significantly increased the activities of both enhancers towards the hkCP but decreased their activities towards the dCP (Fig. 5d). Furthermore, an array of six DRE motifs was sufficient to activate luciferase expression from the hkCP but not the dCP (Fig. 5e). Together, these results show that hkCP and dCP enhancers depend on DRE and GAGA motifs, respectively, and demonstrate that DRE motifs are required and sufficient for hkCP enhancer function.

Our results show that developmental and housekeeping gene regulation is separated genome wide by sequence-encoded specificities of thousands of enhancers to one of two types of core promoter, supporting the longstanding 'enhancer–core-promoter specificity' hypothesis[2–6,23]. Our findings indicate that these specificities are probably mediated by defined biochemical compatibilities[24] between different *trans*-acting factors such as Dref versus Trl (at enhancers) and the different paralogues that exist for several components of the general transcription apparatus (at core promoters), presumably including the TATA-box-binding protein-related factor 2 (Trf2) at housekeeping core promoters[25,26]. As such paralogues can have tissue-specific expression and stage-specific or promoter-selective functions[27,28] (reviewed in refs 29, 30), sequence-encoded enhancer–core-promoter specificities could be used more widely to define and separate different transcriptional programs (Fig. 5f).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 May; accepted 20 October 2014.

Published online 15 December 2014; corrected online 25 February 2015 (see full-text HTML version for details).

1. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157,** 13–25 (2014).
2. Li, X. & Noll, M. Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the *Drosophila* embryo. *EMBO J.* **13,** 400–406 (1994).
3. Ohtsuki, S., Levine, M. & Cai, H. N. Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev.* **12,** 547–556 (1998).
4. Sharpe, J., Nonchev, S., Gould, A., Whiting, J. & Krumlauf, R. Selectivity, sharing and competitive interactions in the regulation of *Hoxb* genes. *EMBO J.* **17,** 1788–1798 (1998).
5. Merli, C., Bergstrom, D. E., Cygan, J. A. & Blackman, R. K. Promoter specificity mediates the independent regulation of neighboring genes. *Genes Dev.* **10,** 1260–1270 (1996).
6. Butler, J. E. & Kadonaga, J. T. Enhancer–promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev.* **15,** 2515–2519 (2001).
7. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* **1,** 40–51 (2012).
8. Parry, T. J. *et al.* The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. *Genes Dev.* **24,** 2013–2018 (2010).
9. Engström, P. G., Ho Sui, S. J., Drivenes, O., Becker, T. S. & Lenhard, B. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.* **17,** 1898–1908 (2007).
10. FitzGerald, P. C., Sturgill, D., Shyakhtenko, A., Oliver, B. & Vinson, C. Comparative genomics of *Drosophila* and human core promoters. *Genome Biol.* **7,** R53 (2006).
11. Pfeiffer, B. D. *et al.* Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc. Natl Acad. Sci. USA* **105,** 9715–9720 (2008).
12. Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339,** 1074–1077 (2013).
13. Shlyueva, D. *et al.* Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. *Mol. Cell* **54,** 180–192 (2014).
14. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27,** 299–308 (1981).
15. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Rev. Genet.* **13,** 233–245 (2012).
16. Ohler, U., Liao, G.-C., Niemann, H. & Rubin, G. M. Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol.* **3,** research0087.1–0087.12 (2002).
17. Smith, D., Wohlgemuth, J., Calvi, B. R., Franklin, I. & Gelbart, W. M. *hobo* enhancer trapping mutagenesis in *Drosophila* reveals an insertion specificity different from *P* elements. *Genetics* **135,** 1063–1076 (1993).
18. Kutach, A. K. & Kadonaga, J. T. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol.* **20,** 4754–4764 (2000).
19. Yáñez-Cuna, J. O., Dinh, H. Q., Kvon, E. Z., Shlyueva, D. & Stark, A. Uncovering *cis*-regulatory sequence requirements for context-specific transcription factor binding. *Genome Res.* **22,** 2018–2030 (2012).
20. Yáñez-Cuna, J. O. *et al.* Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. *Genome Res.* **24,** 1147–1156 (2014).
21. Gurudatta, B. V., Yang, J., Van Bortle, K., Donlin-Asp, P. G. & Corces, V. G. Dynamic changes in the genomic localization of DNA replication-related element binding factor during the cell cycle. *Cell Cycle* **12,** 1605–1615 (2013).
22. modENCODE Consortium Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330,** 1787–1797 (2010).
23. Ohler, U. & Wassarman, D. A. Promoting developmental transcription. *Development* **137,** 15–26 (2010).
24. van Arensbergen, J., van Steensel, B. & Bussemaker, H. J. In search of the determinants of enhancer–promoter interaction specificity. *Trends Cell Biol.* http://dx.doi.org/10.1016/j.tcb.2014.07.004 (2014).
25. Wang, Y.-L. *et al.* TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes Dev.* **28,** 1550–1555 (2014).
26. Isogai, Y., Keles, S., Prestel, M., Hochheimer, A. & Tjian, R. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* **21,** 2936–2949 (2007).
27. Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. C. & Tjian, R. TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* **420,** 439–445 (2002).
28. Deato, M. D. E. & Tjian, R. Switching of the core transcription machinery during myogenesis. *Genes Dev.* **21,** 2137–2149 (2007).
29. D'Alessio, J. A., Wright, K. J. & Tjian, R. Shifting players and paradigms in cell-specific transcription. *Mol. Cell* **36,** 924–931 (2009).
30. Müller, F., Zaucker, A. & Tora, L. Developmental regulation of transcription initiation: more than just changing the actors. *Curr. Opin. Genet. Dev.* **20,** 533–540 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Cochella and O. Bell for comments on the manuscript. Deep sequencing was performed at the CSF Next-Generation Sequencing Unit (http://csf.ac.at). M.A.Z. was supported by the Austrian Science Fund (FWF, F4303-B09) and C.D.A., K.S., M.R. and O.F. by a European Research Council Starting Grant (no. 242922) awarded to A.S. Basic research at the Research Institute of Molecular Pathology is supported by Boehringer Ingelheim GmbH.

Author Contributions M.A.Z., C.D.A. and A.S. conceived the project. C.D.A., K.S., M.P., M.R. and O.F. performed the experiments and M.A.Z. the computational analyses. M.A.Z., C.D.A. and A.S. wrote the manuscript.

Author Information All deep sequencing data are available at http://www.starklab.org and have been deposited in the Gene Expression Omnibus database under accession numbers GSE40739 and GSE57876. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.S. (stark@starklab.org).

METHODS

hkCP STARR-seq vector. We derived the hkCP STARR-seq vector from the original STARR-seq vector[12] by replacing the DSCP sequence with the sequence of the *RpS12* core promoter (−50 to +50 bp relative to the TSS; TTGTACCAATAGCTAAAAACTCACATCTCCAGCGCCATGCCGATTTTGTTCTCTTTCTTTCCGGTTGTCAAAAGGTACAGATGCTTGGATTTTATTTCTC). The STARR-seq vectors are available subject to a material transfer agreement (MTA). For both STARR-seq vectors, we confirmed that transcription initiates from within the respective core promoters' Inr (DSCP) and TCT (*RpS12*) motifs by 5′ rapid amplification of cDNA ends (RACE; Extended Data Fig. 2). All other STARR-seq vectors were derived from the hkCP STARR-seq vector by replacing the 100 bp sequence encompassing the *RpS12* core promoter by the sequences indicated in Supplementary Table 15 using the BglII and SbfI restriction sites.

hkCP and dCP luciferase vectors. For the dCP luciferase vector, the SV40 promoter of the pGL3-Promoter Vector (Promega) was replaced by the DSCP[11] and a Gateway cassette was inserted downstream of the luciferase gene and the SV40 polyA-signal into the AfeI restriction site, to allow Gateway LR cloning of candidate sequences[12]. For the hkCP luciferase vector, the SV40 promoter and the sequence until the translation start codon of the luciferase gene was replaced by the sequence encompassing the TSS of *RpS12* from −50 bp until its translation start codon: TTGTACCAATAGCTAAAAACTCACATCTCCAGCGCCATGCCGATTTTGTTCTCTTTCTTTCCGGTTGTCAAAAGGTACAGATGCTTGGATTTTATTTCTCCGAAATGAAGAGGTTTTCTTATCGAAAATGTAATAAATATGAACAATTAACTATCTTTTCCAGTGCAGTGCATCCTTAACCGCAGAACA. Constructs are available subject to an MTA.

Intrinsic activity of core promoters. All core promoters used in this study were cloned into the dCP luciferase vector (without the Gateway cassette), replacing the DSCP between the BglII and SbfI restriction site with the respective core promoter. For each core promoter, the intrinsic (or basal) activity was measured as firefly luciferase activity and is presented as relative luciferase units, normalized to *Renilla* luciferase signals.

Genome-wide STARR-seq screens. STARR-seq enhancer screens using the core promoters of *RpS12* (hkCP), *NipB*, *x16*, and *eEF1δ* (Supplementary Table 15) were performed in two biological replicates (independent transfections) as described previously[12] with the following exceptions. First, 1.6×10^9 S2 cells and OSCs[31] were transfected per biological replicate. Second, first-strand cDNA synthesis was performed in 30–60 reactions with the STARR-seq RT primer (CTCATCAATGTATCTTATCATGTCTG) as reverse transcription primer. Last, next-generation sequencing (NGS) was performed on an Illumina HiSeq 2000 machine using multiplexing according to the manufacturer's instructions. STARR-seq data using the DSCP (dCP STARR-seq) and *Hsp70* core promoters are from ref. 12, but were reanalysed using the same pipeline as for hkCP STARR-seq.

Focused STARR-seq BAC screens. The DSCP is a 137-nucleotide-long synthetic core promoter derived from the core promoter of *even skipped* (*eve*)[11]. To assess the functional similarity of the DSCP, its 137-nucleotide-long wild-type counterpart from the *eve* locus, and a version defined identically to all other core promoter used here (−50 to +50 nucleotides around the TSS), we performed STARR-seq screens with libraries derived from 29 different BACs containing a total of ~5 Mb of *D. melanogaster* genomic DNA (Supplementary Table 16). For comparison, we also screened all other core promoters with this library. For library cloning, all BACs were grown in individual bacterial cultures and were then mixed equally according to measurements of their optical density at 600 nm ($OD_{600\,nm}$) before BAC DNA isolation to achieve an equal distribution of all BACs. BAC DNA extraction, sonication and adaptor ligation was performed as described[12] and the same adaptor-ligated and PCR-amplified BAC DNA was used to clone all focused STARR-seq libraries. Per STARR-seq vector, four In-Fusion reactions were performed, which allowed five transformation reactions as described[12]. Each library was grown in 4 l liquid culture (LB medium) to an $OD_{600\,nm}$ of 2.0–2.5. Each BAC library was screened as described earlier for the genome-wide screens; however, only 1×10^8 S2 cells were used, accounting for the less complex library. Similarly, the number of reactions for all subsequent steps of the STARR-seq protocol was reduced fourfold.

Luciferase reporter assays. Luciferase assays were performed as described previously[12] with the exception that the candidate enhancers were cloned downstream of the luciferase gene and the polyA signal, more than 2 kb away from the respective core promoter (*RpS12* or DSCP). Candidate enhancers were selected manually based on different criteria to allow the systematic assessment of several aspects of this study, including enhancers that were (1) specific to one of the two different core promoters (24 hkCP and 12 dCP enhancers) or found in both screens (7 shared enhancers); (2) located proximally (17) or distally (7) to the hkCP; and (3) of different strengths according to STARR-seq (ranks 18 to 1,044). We cloned all candidates as described[12] (for their genomic coordinates and primer sequences see Supplementary Table 17), picking initially one orientation towards the luciferase TSS randomly. However, to test the influence of TSSs contained in the candidate sequences, we cloned and tested all TSS proximal candidates (hkCP_01 to hkCP_17) in both orientations using both core promoters. Candidate enhancers with DRE mutations were cloned from synthesized DNA fragments (GeneArt Strings; Supplementary Table 18). Candidates with DRE motifs that replace GAGA motifs were cloned similarly using synthesized DNA fragments (gBlocks) obtained from Integrated DNA Technologies (Supplementary Table 19). We also added an array of 6× DRE motifs into the AfeI restriction site of the dCP and hkCP luciferase vectors and cloned dCP_01 to dCP_11 into the middle of the DRE motif array (using AfeI) of the hkCP luciferase vector, such that these sequences were each flanked by three DRE motifs (Supplementary Table 19).

Luciferase assay data analysis. For all luciferase assays, we calculated standard deviations and one-sided Student's *t*-tests from three biological replicates (independent transfections). Core promoters have intrinsic (basal) activities that can differ between different core promoters. Therefore, when comparing enhancer activities for different core promoters, normalization to the core promoters' intrinsic activities is required, which we assessed with three different negative control fragments (nine biological replicates in total). For all measurements, we normalized firefly luciferase values first to *Renilla* luciferase values (controlling for transfection efficiency) and then to the normalized luciferase values of the three negative control sequences. Candidates with a significant ($P < 0.05$) enrichment greater than 1.5 fold over negative were considered positive.

5′ RACE of STARR-seq transcripts. To determine the exact TSSs of hkCP and dCP within the STARR-seq vectors we performed 5′ RACE of STARR-seq transcripts using one enhancer for each (an intergenic enhancer of *TpnC41C* for hkCP and an intronic enhancer of *zfh1* (shared_01 from ref. 12) for dCP) which we cloned with EcoRV at the position of the selection cassette used during library cloning (Supplementary Table 20). We transfected 3.2×10^7 cells with each of the constructs and isolated total RNA using the RNeasy mini prep kit (Qiagen; two columns per construct) followed by polyA+ RNA isolation using oligo-dT Dynabeads (Life Technologies) according to the manufacturer's instructions. We then performed 5′ RACE for both samples using the FirstChoice RLM-RACE Kit (Ambion; catalogue no. AM1700) according to the manufacturer's instructions. To reflect RNA processing of the STARR-seq pipeline, reverse transcription was, however, performed using SuperscriptIII (Invitrogen) according to the manufacturer's instructions and using the reverse transcription primer GFP-RT (Supplementary Table 20) as a gene-specific primer (using RNA amounts according to the FirstChoice manual). The first PCR was performed with the manufacturer-provided 5′ RACE Outer Primer and the transcript-specific primer RACE-01-rv, using 2× KAPA Hifi Hot Start Ready Mix (98 °C for 45 s; followed by 35 cycles of 98 °C for 15 s, 69 °C for 30 s, 72 °C for 30 s) with 1 µl of cDNA as template. The nested PCR was performed similarly (primer: 5′ RACE Inner Primer and RACE-02-rv; 98 °C for 45 s; followed by 30 cycles of 98 °C for 15 s, 67 °C for 30 s, 72 °C for 10 s). The PCR products were visualized on a 1% agarose gel. The PCR products for both samples were Sanger sequenced using the primer GFP-seq-rv (for all primer sequences see Supplementary Table 20).

STARR-seq NGS data processing. Paired-end STARR-seq and input read processing was performed as described[32]. The NGS data for dCP (DSCP) and *Hsp70* were obtained from ref. 12 and reanalysed. In the same cell line, a hkCP peak is considered to be 'specific' if the 501 bp window centred at the peak summit does not overlap with any such window for dCP peaks, and vice versa (note that this is only applied within each cell type, such that comparisons across cell types are not influenced). For screens with the BAC-derived libraries, we considered only fragments that originated from the BACs used and determined the relative abundance of each BAC from the NGS data of the respective inputs only. On the basis of this, we then adjusted both inputs and STARR-seq NGS data such that all BACs were equally represented and analysed the data as described earlier.

Venn diagrams and peak intersection. We used the same intersection method as described earlier, and plotted the Venn diagrams with areas proportional to the number of peaks.

Scatter plots. We calculated the STARR-seq enrichment over input at the summit positions of both data sets that were to be compared, using a pseudo count of 1, and computed the $\log_2$ of corrected ratio as described[12]. This plots one data point for each enhancer—even for closely spaced ones—exactly at the enhancer's summit position. For visualizing replicates, we called peaks on the merged data sets and plotted the values from both replicates at these peaks' summits.

Enhancer-to-gene assignment. We performed three different strategies of enhancer-to-gene assignments: (1) 'closest TSS', whereby an enhancer is assigned to the closest TSS of an annotated transcript; (2) '1 kb TSS', whereby an enhancer is assigned to all TSSs that are within 1 kb; and (3) 'gene loci', whereby an enhancer is assigned to a gene provided that it falls within 5 kb upstream from the TSS, within the gene body itself, or 2 kb downstream of the gene (multiple assigned genes are possible). In all cases we used annotation from *D. melanogaster* FlyBase release 5.50.

Genomic distribution. We assigned a unique annotation for each nucleotide in the genome by using the following priority order: coding sequence (CDS), core promoter (±50 bp around TSS), 5′ UTR, 3′ UTR, first intron, intron, proximal promoter (200 bp upstream of a TSS), intergenic region. We then assigned each peak to one of these categories by the annotation of the peak's summit.

GO analysis. We assessed whether genes assigned to hkCP or dCP enhancers were enriched for particular GO categories[33] by calculating hypergeometric *P* values for all categories, which we corrected for multiple comparisons (FDR-type correction in R). We then sorted all categories according to *P* values of overrepresentation, selected the top 100 of either hkCP or dCP, and removed redundant categories manually. For each category, we calculated $\log_{10}$(*P*-value underrepresentation) − $\log_{10}$(*P*-value overrepresentation), and sorted the terms in a descending order of difference between hkCP and dCP values. The colour intensity of the heat maps represents $\log_{10}$(*P*-value underrepresentation) − $\log_{10}$(*P*-value overrepresentation).

Gene expression analysis. We analysed enrichment in ubiquitous versus tissue-specific gene expression sets as described for the GO analysis above. To define the gene sets based on an *in situ* hybridization data set of fly embryos (BDGP[34]), we first removed maternal (stages 1 to 3) annotations, as well as genes with the annotation 'no staining' in all stages. We required each gene to have annotations for at least three stage groupings. We called a gene 'tissue specific' if at most one of these annotations contains the word 'ubiquitous', and called it 'ubiquitous' if at least 60% of them contain word 'ubiquitous'. We also defined gene sets based on microarray data sets from dissected fly tissues (FlyAtlas[35]). We defined genes as 'ubiquitous' if their expression does not change more than twofold compared with the whole fly for at least 15 out of 23 tissues. For this, we used the ratios and 'change_direction' calls from FlyAtlas directly and did not consider cell lines and carcasses. We similarly defined genes to be 'tissue specific' if they change more than twofold in at least three tissues. We do not consider genes with multiple conflicting entries as they can result from the use of multiple probes and removed genes that overlapped between the 'ubiquitous' and 'tissue-specific' gene sets from both sets.

Transcription factor motif and core promoter element enrichment analysis. We used previously employed position weight matrices (PWMs) for different transcription factors[13] with a cut-off of $4^{-6} = 2.4 \times 10^{-4}$. We selected random control regions by controlling for genomic and chromosome distribution, and required that they did not overlap with any peak. We scored each motif for its enrichment in 401 bp windows centred on the peak summits by multiple testing (FDR) corrected hypergeometric *P* values. We considered only motifs that showed $\log_2$(confidence ratio of motif counts in peak windows/motif counts in random control regions) > 1 and *P* value < 0.01 in hkCP or dCP enhancers (or both) and reduced motif redundancy by removing highly similar motifs as in ref. 13 and references therein. We sorted the motifs in a descending order by difference in $\log_2$(hkCP enrichment) − $\log_2$(dCP enrichment). When assessing whether the observed motif distribution persisted for distal enhancers (Extended Data Fig. 10a), we kept the motifs and their order as in Fig. 5a and only re-evaluated their enrichment in distal enhancers. The colour intensity of the heat maps represents $\log_2$(confidence ratio of motif counts in peak windows/motif counts in random control regions). We used previously published PWMs or created PWMs from published nucleotide counts for TATA box, Inr, MTE, DPE and Ohler motifs[16] 1, 5, 6, 7 and the TCT motif[8] restricted to 8 bp. We scanned for motif occurrences using MAST from the MEME suite[36] (version 4.9.0) and parameters that ensured specificity and sensitivity for each motif (Supplementary Table 21). For enhancer-to-gene assignment methods 1 and 2 described earlier, we determined the presence of each core promoter element in the core promoter region of all genes uniquely assigned to either hkCP or dCP enhancers, respectively. For assignment method 3, we took the core promoter elements of the TSSs of the longest messenger RNA isoform. We assessed the differential distribution of each core promoter element between the core promoters assigned to hkCP or dCP enhancers by confidence ratios and hypergeometric *P* values.

Transcription factor motif and core promoter element *de novo* discovery. We used MEME[36] (version 4.9.0) to discover *de novo* motifs with lengths between 5 and 8 nucleotides in the enhancer regions we identified using STARR-seq and in the core promoter regions around the nearest annotated transcription TSS. We provide all discovered motifs in Supplementary Table 22.

Core promoter similarity heat map. For all pairs of core promoters, we computed pair-wise PCCs between the respective STARR-seq fragment coverages at the summits of all peaks called in either of the two screens genome wide. We performed hierarchical clustering (complete linkage) in R, directly using the computed PCC values as similarities.

STARR-seq enrichment heat map. We computed the $\log_2$ of the corrected STARR-seq enrichment over input as described earlier, but for each nucleotide in a 20 kb window around all reference peak summit positions, and down-sampled the data points 50-fold by calculating one average data point per 50 nucleotides.

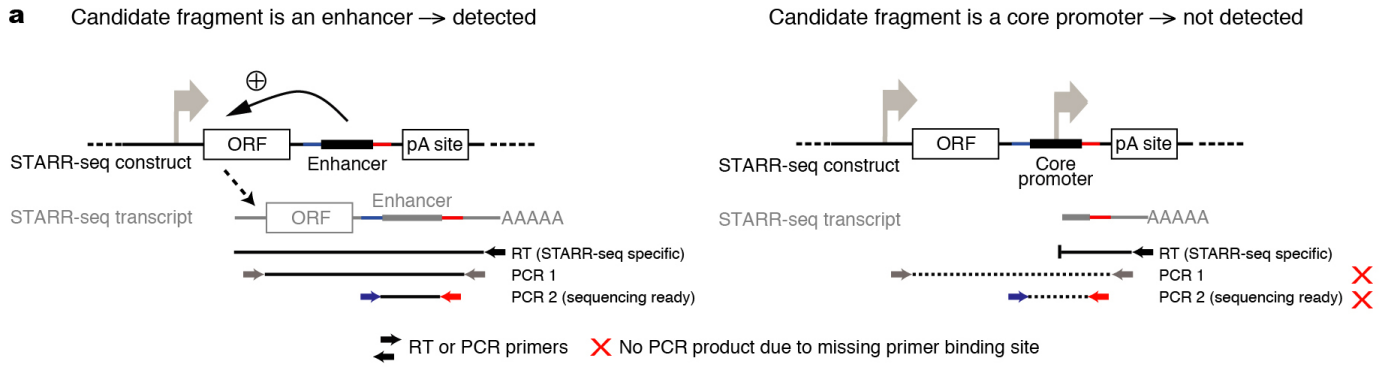
STARR-seq enrichment meta-profiles around TSSs. We calculated corrected STARR-seq enrichments ($\log_2$) as for the heat maps, but for 20 kb windows around TSSs, selected according to their core promoter motif content (see Extended Data Figs 4 and 8), corrected for the orientation of the TSSs within the genomic sequence. We then calculated the average for each position along the *x*-axis.

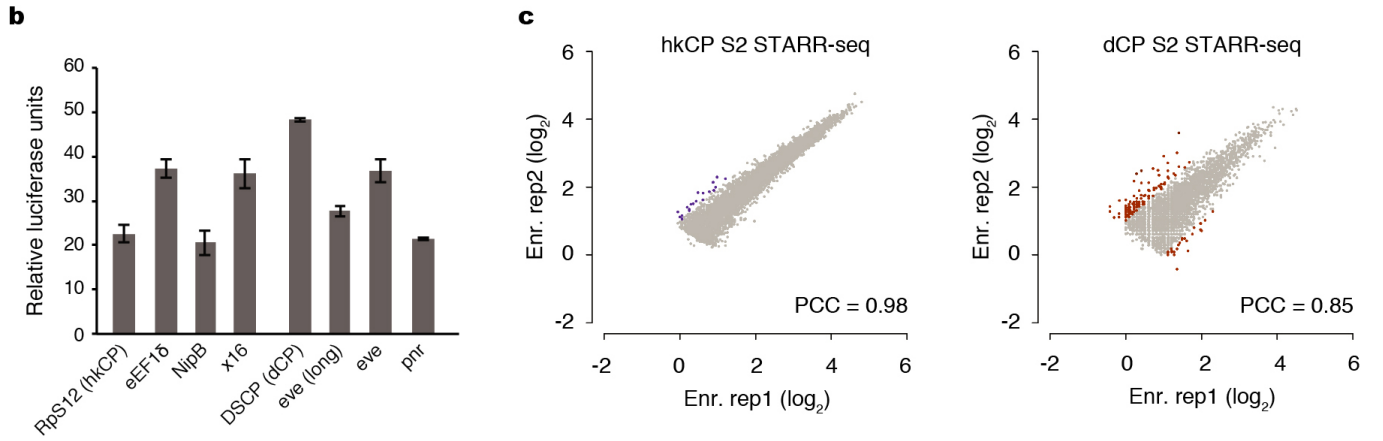
Boxplot. We obtained Dref ChIP-seq and input data (from Kc167 cells) from ref. 21 (Gene Expression Omnibus accession numbers GSM977024 and GSM762849) and mapped the 36-nucleotide reads using bowtie[37] (version 0.12.9) with the following parameters: -p 4 -q -v 3 -m 1 --best --strata --quiet. We extended the reads to 150 bp, calculated the coverage for ChIP-seq and input at the STARR-seq peak summit, normalized the value to the number of input fragments, added a pseudo count of 1, and computed the confidence ratio of ChIP-seq over input. For the Trl ChIP-chip data obtained from ref. 22, we used the signal of the chip-array probe at the peak summit if available or inferred the signal by linear extrapolation from the two nearest flanking probes (one on each side) provided that they were both within 10 nucleotides of the peak summit. We calculated statistical significance via Wilcoxon's paired rank tests.

Coordinate intersections. We performed genomic coordinate intersections using the BEDTools suite[38] (version 2.17.0).

Statistics. We performed all statistical calculations and created graphical displays with R[39].

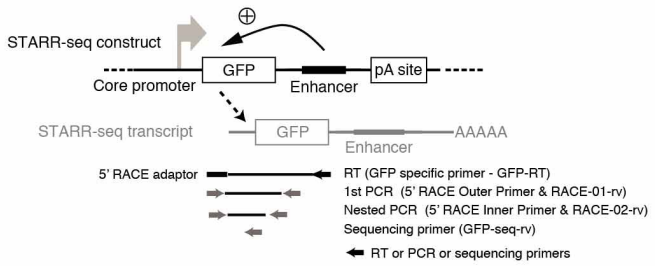
31. Saito, K. *et al.* A regulatory circuit for *piwi* by the large Maf gene *traffic jam* in *Drosophila*. *Nature* **461,** 1296–1299 (2009).
32. Arnold, C. D. *et al.* Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during *cis*-regulatory evolution. *Nature Genet.* **46,** 685–692 (2014).
33. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25,** 25–29 (2000).
34. Tomancak, P. *et al.* Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **8,** R145 (2007).
35. Chintapalli, V. R., Wang, J. & Dow, J. A. T. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nature Genet.* **39,** 715–720 (2007).
36. Bailey, T. L. & Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14,** 48–54 (1998).
37. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
38. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
39. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2010).
40. Zeitlinger, J. & Stark, A. Developmental gene regulation in the era of genomics. *Dev. Biol.* **339,** 230–239 (2010).
41. Kvon, E. Z. *et al.* Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature* **512,** 91–95 (2014).
42. Soler, E. *et al.* The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev.* **24,** 277–289 (2010).
43. Chen, K. *et al.* A global change in RNA polymerase II pausing during the *Drosophila* midblastula transition. *eLife* **2,** e00861 (2013).
44. Lagha, M. *et al.* Paused Pol II coordinates tissue morphogenesis in the *Drosophila* embryo. *Cell* **153,** 976–987 (2013).
45. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339,** 950–953 (2013).





Extended Data Figure 1 | Set-up of STARR-seq with different core promoters. **a**, STARR-seq detects enhancers but no promoters (reproduced with permission from ref. 12). Left, STARR-seq couples the enhancer activities of candidate fragments to the sequences of the candidates in *cis* by placing the candidates to a position within the reporter transcript. Enhancer activities can therefore be assessed by the presence of candidates among cellular messenger RNAs, which allows the parallel assessment of millions of candidates, enabling genome-wide screens. Sequences that activate transcription from the intended core promoter of the STARR-seq vector lead to a full-length reporter transcript and can be detected by STARR-seq. Shown are the reverse transcription (RT) and nested polymerase chain reaction (PCR) steps of the STARR-seq reporter RNA processing protocol that ensure this. Right, in contrast, STARR-seq does not detect truncated transcripts that result if a candidate fragment functions as a promoter to initiate transcription. Thus, core-promoter-containing (that is, TSS-overlapping) sequences that are detected by STARR-seq exhibit enhancer activity as they can activate transcription from a remote position, in addition to their ability to serve as core promoters endogenously[12]. **b**, Luciferase signals (firefly/*Renilla*) assessing the intrinsic (or basal) activity of the core promoters used in this study. The luciferase reporter constructs do not contain any enhancer and differ only in the respective core promoter sequences. The basal activities differ as expected, but do not differ consistently between housekeeping (*RpS12*, *eEF1δ*, *NipB*, *x16*) and developmental (DSCP, *eve* (long), *eve* and *pnr*) core promoters, nor between core promoters for which the STARR-seq screens appear most similar (for example, *RpS12* and *eEF1δ*; see Fig. 3). Note that all luciferase assays and STARR-seq screens are corrected for differences in intrinsic activity. **c**, Reproducibility of hkCP and dCP STARR-seq in *D. melanogaster* S2 cells. The reproducibility of hkCP and dCP STARR-seq as assessed by the STARR-seq enrichments (replicate 1 versus 2) at the summits of enhancer peaks called in the merged experiments (hkCP: 5,956; dCP: 5,408). Scatter plots are enlarged versions of the insets in Fig. 1d. "Enr. rep X", STARR-seq enrichment in replicate X. Note that the raw data for dCP have been re-analysed from ref. 12.

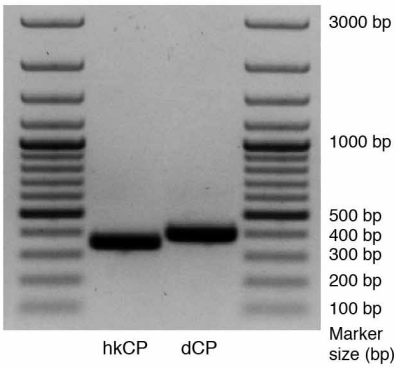
a Schematic overview of 5' RACE on the STARR-seq vectors



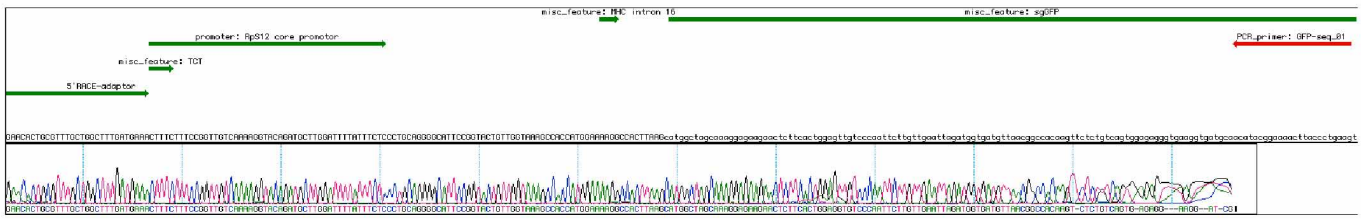
Core-promoter–enhancer pairs used:

hkCP (*RpS12* core promoter) and hkCP_19 (expected size of nested PCR product: ~350bp)
dCP (DSCP) and intronic enhancer of *zfh1* (expected size of nested PCR product: ~400bp)

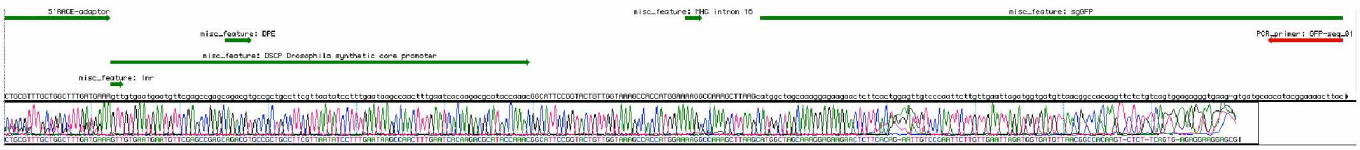
b Agarose gel electrophoresis of 5' RACE nested PCRs (PCR2)



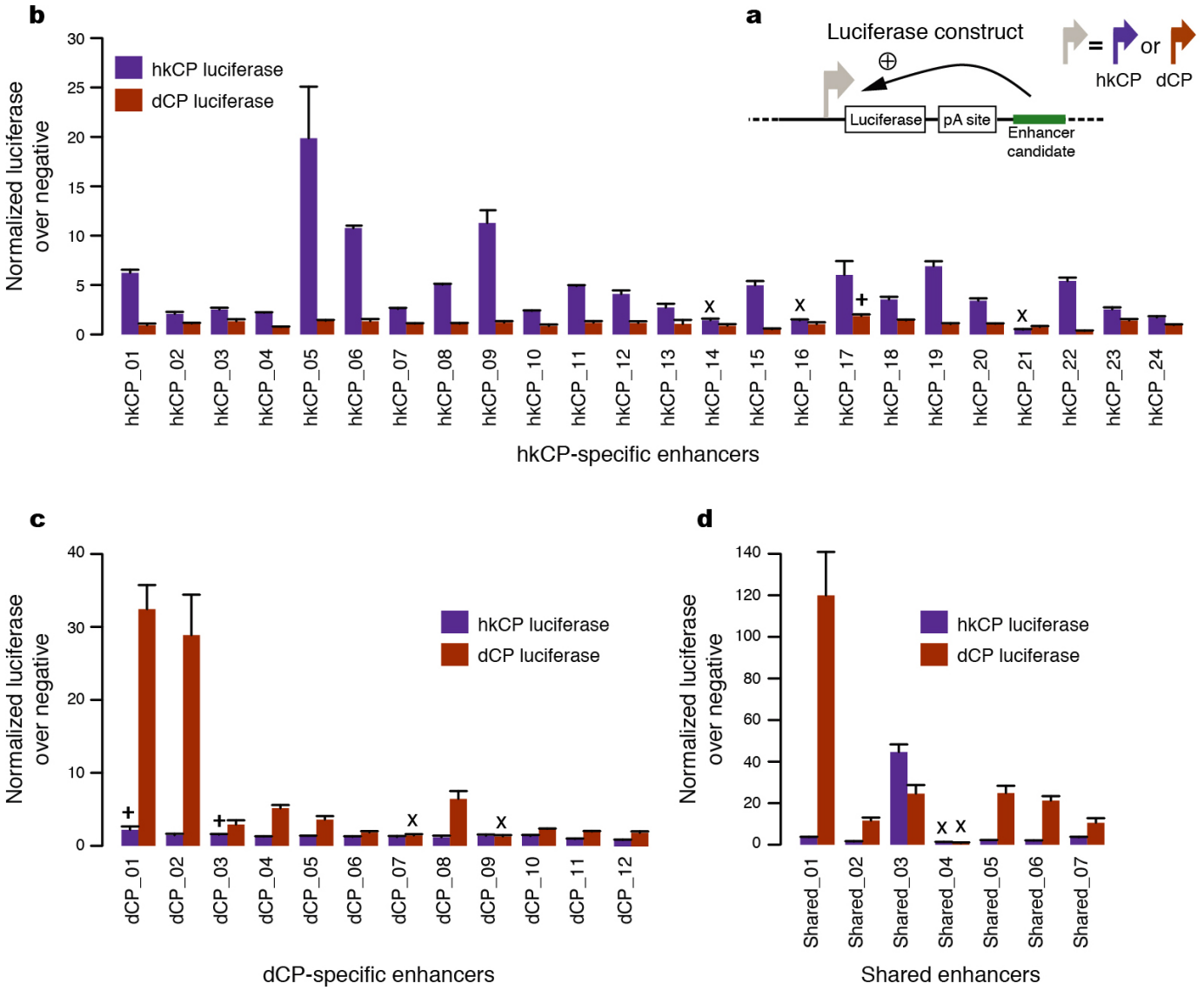
c Sequence alignment and chromatogram of Sanger sequencing of the PCR product (nested PCR) of 5' RACE of the hkCP STARR-seq vector



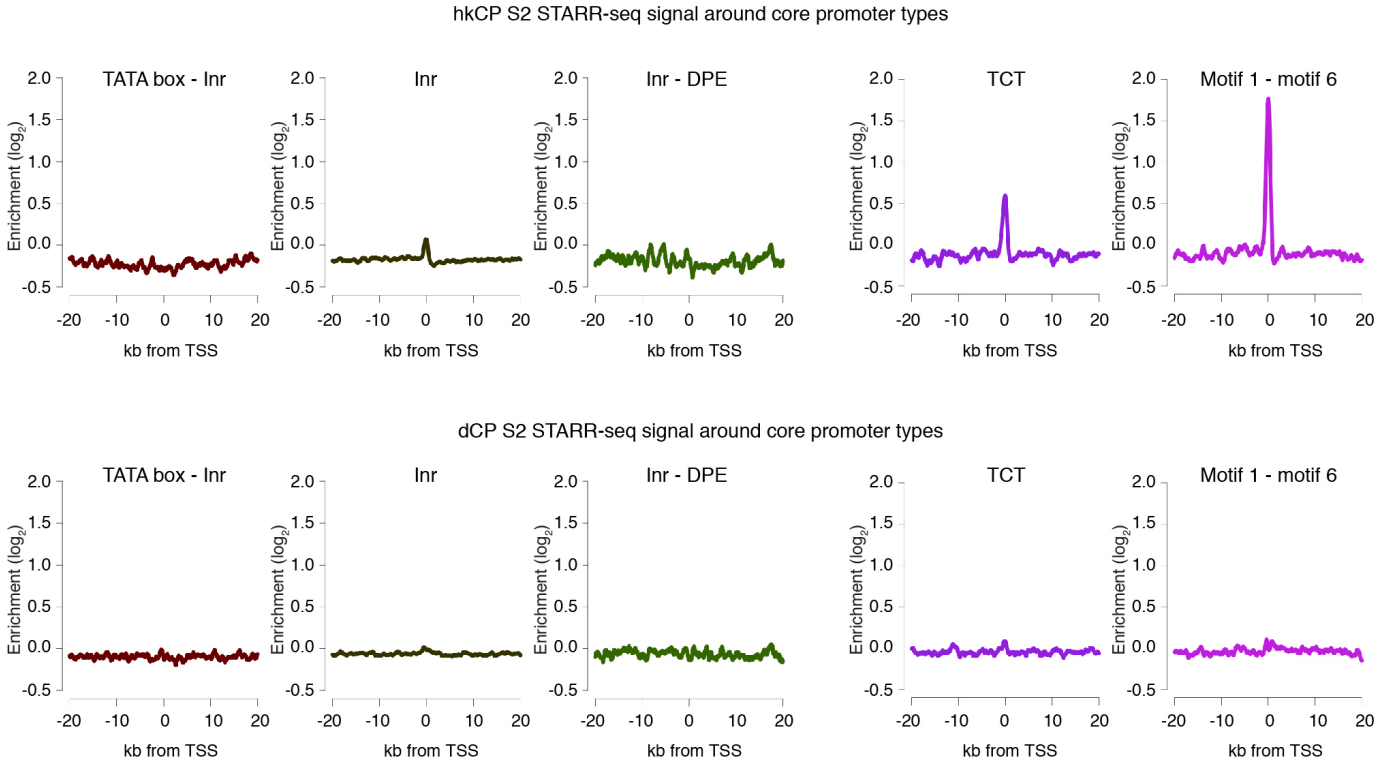
d Sequence alignment and chromatogram of Sanger sequencing of the PCR product (nested PCR) of 5' RACE of the dCP STARR-seq vector



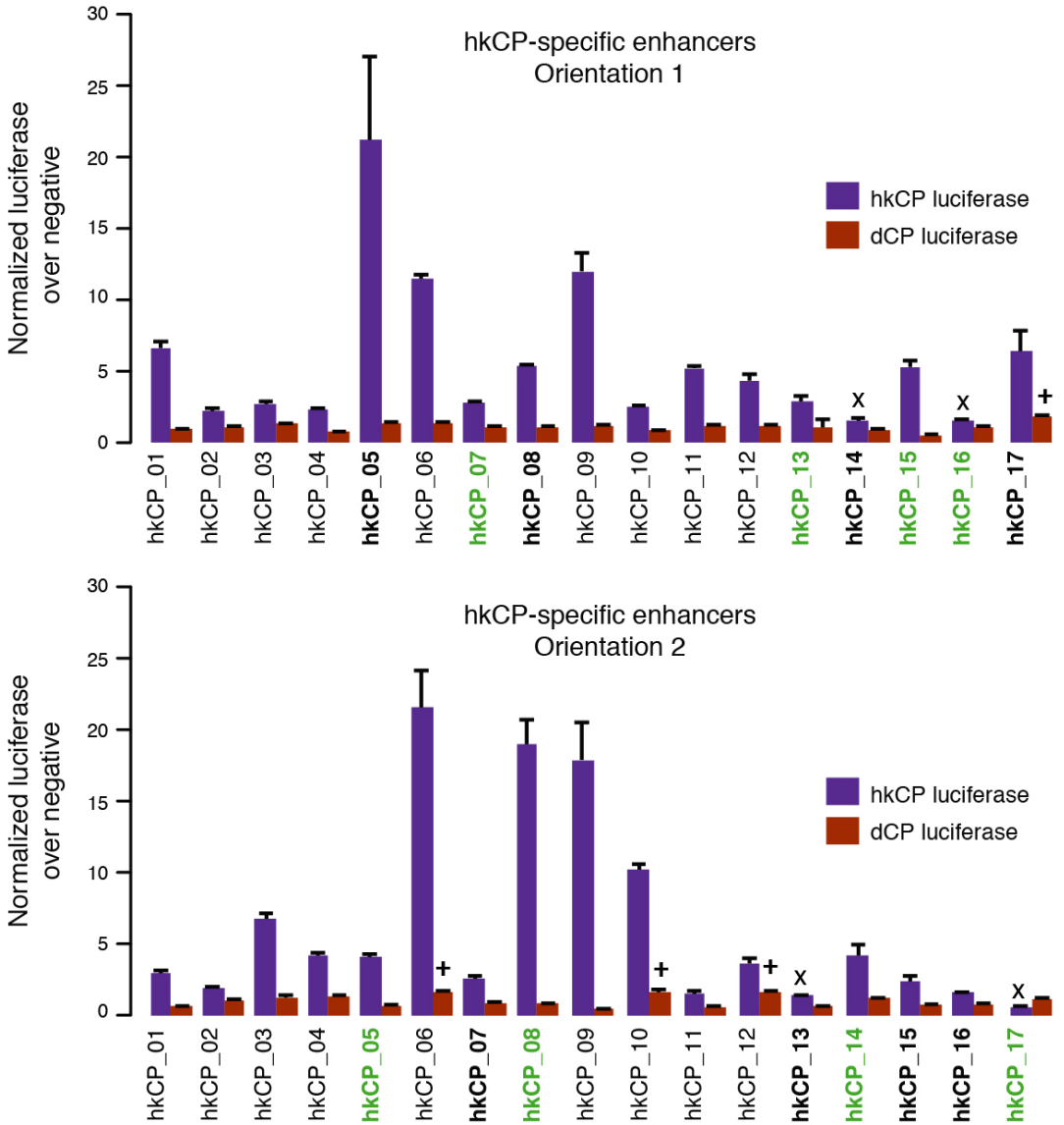
Extended Data Figure 2 | Transcription initiates within the core promoter of the STARR-seq construct. **a**–**d**, 5′ Rapid amplification of cDNA ends (5′ RACE) demonstrates that transcription initiates at the TCT and Inr motifs within the hkCP and dCP, respectively. **a**, Set-up of the 5′ RACE experiment, including the STARR-seq plasmid, used here with two defined enhancers, the STARR-seq transcript and the location of all primers used to specifically amplify 5′-capped STARR-seq transcripts. **b**, 5′ RACE nested PCR products separated on a 1% agarose gel. **c**, Screenshot of Sanger sequencing results (chromatogram and called bases) compared with the template sequence. Annotations are shown in green, in the following order: 5′ RACE adaptor, hkCP with TCT motif (only the part downstream of the TSS is annotated, as the 5′ part is not present in the sequenced complementary DNA), spliced intron, green fluorescent protein (GFP); the sequencing primer is shown in red (top). Also shown is a version that displays the template and Sanger sequencing results for the core promoter region only (zoom in). **d**, Same as in **c** but for the dCP for which transcription initiates within the Inr motif.



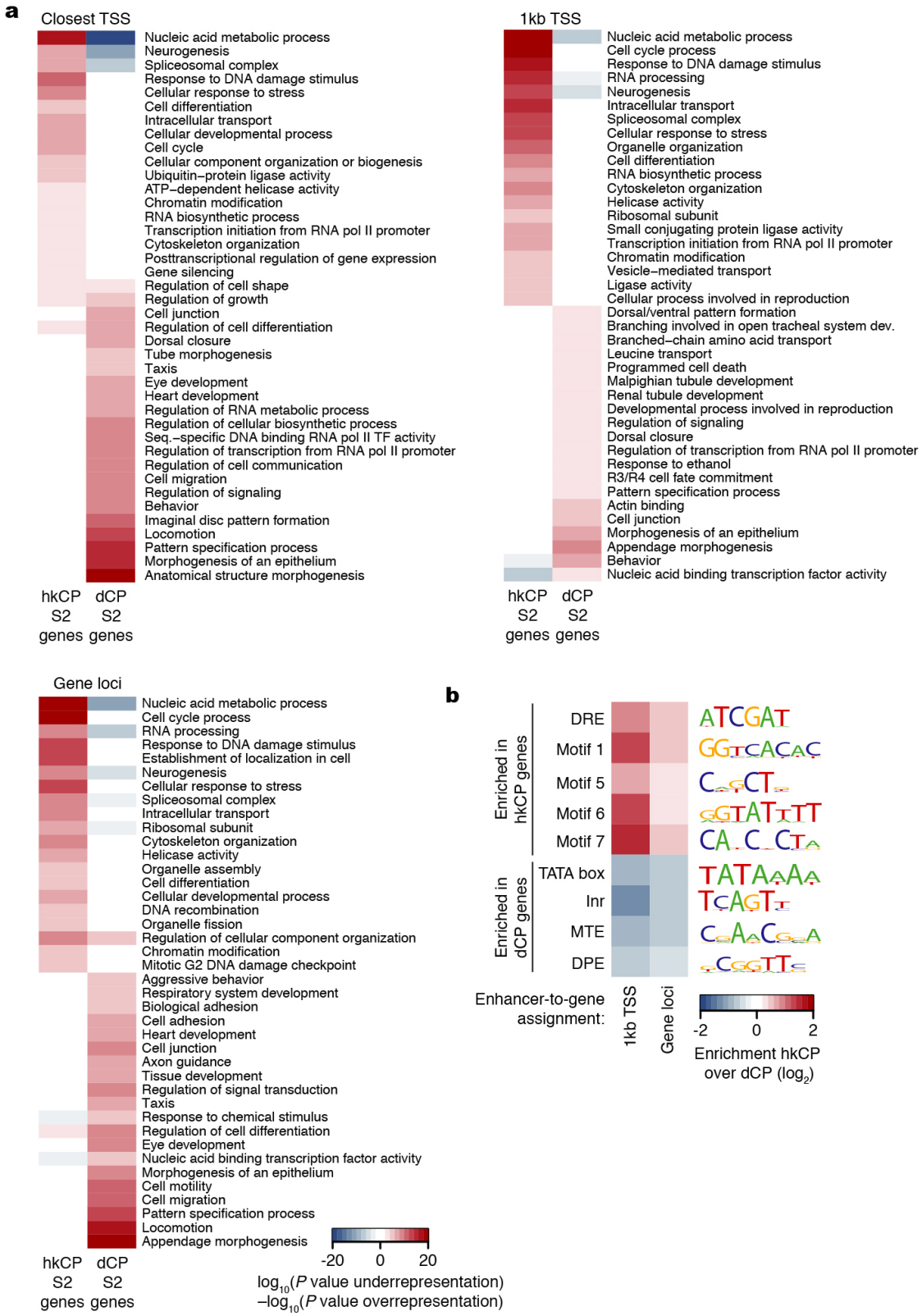
Extended Data Figure 3 | Specificity of hkCP and dCP enhancers to the hkCP and dCP assessed by luciferase assays. **a**, Luciferase reporter set-up with the hkCP or dCP (see also Fig. 1e). **b**, Luciferase signals of 24 hkCP-specific enhancers tested in a hkCP- (purple bars) as well as in a dCP-containing (brown bars) luciferase reporter. Twenty-one out of 24 hkCP enhancers showed luciferase activity (>1.5 fold over negative, $P < 0.05$ via one-sided unpaired Student's *t*-test, $n = 3$) with the hkCP, while only 1 out of 24 showed activity with the dCP (error bars are s.d. of three biological replicates, 'x' indicates candidates that are not active with the correct core promoter, and '+' indicates candidates for which the activity with the wrong core promoter is above the threshold (note that the activity with the correct core promoter is still higher in all three cases). **c**, As in **b** but testing dCP-specific enhancers. Ten out of 12 are positive with the dCP whereas only 2 out of 12 are positive with the hkCP. **d**, As in **b** and **c** but testing shared enhancers that were found by STARR-seq with hkCP and dCP; 6 out of 7 are active with both core promoters. See Supplementary Table 17 for the genomic coordinates of the enhancers and the primers used to amplify them.



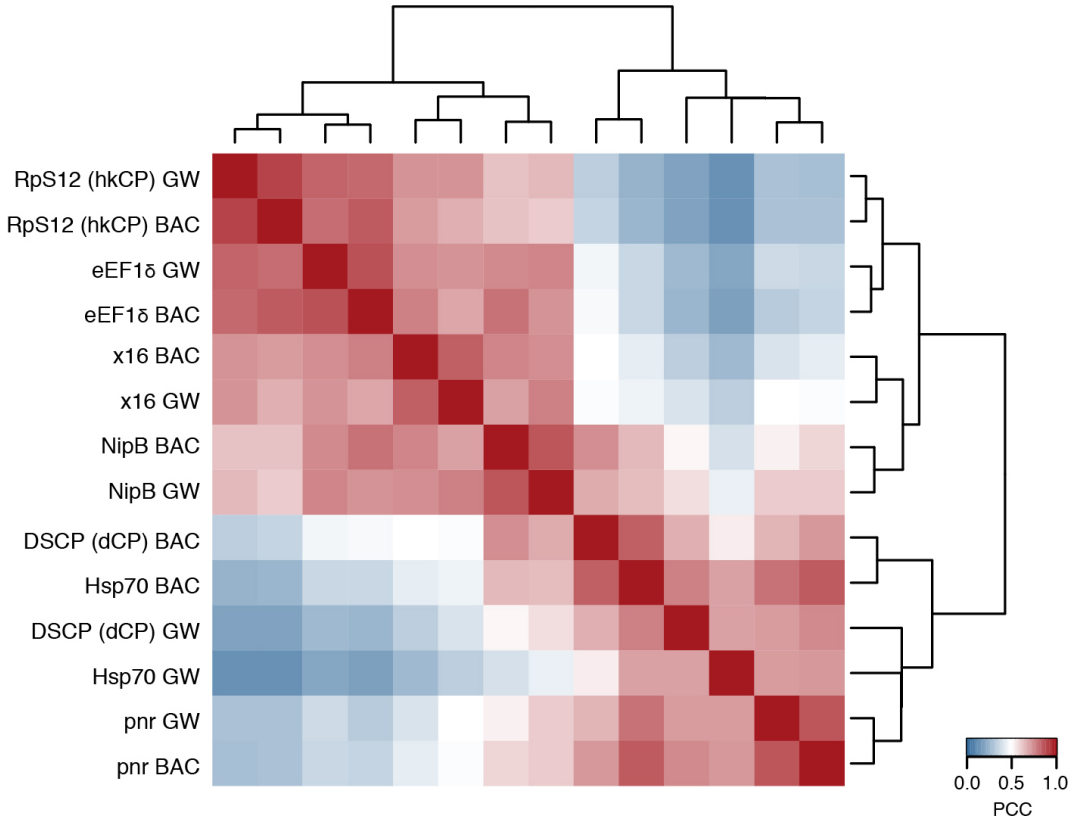
Extended Data Figure 4 | hkCP and dCP STARR-seq signal in S2 cells around different core promoter types. Average hkCP (top) and dCP (bottom) S2 STARR-seq enrichment in 40 kb intervals around TSSs that contain different combinations of known core promoter motifs. Shown are (left to right) TATA box–Inr (179 TSSs), Inr (that do not contain either TATA box or DPE; 1,901), Inr–DPE (100), TCT (303) and motif 1–motif 6 (266). According to their motif contents, the first three are developmental-type core promoters and the last two are housekeeping-type core promoters. Indeed, only the housekeeping-type core promoters show a strong enrichment of hkCP S2 STARR-seq signals at the TSS, which is not seen for the dCP STARR-seq signal (owing to enhancer–core-promoter specificity) nor for the developmental-type core promoters (owing to the dCP enhancers location at more distal sites).



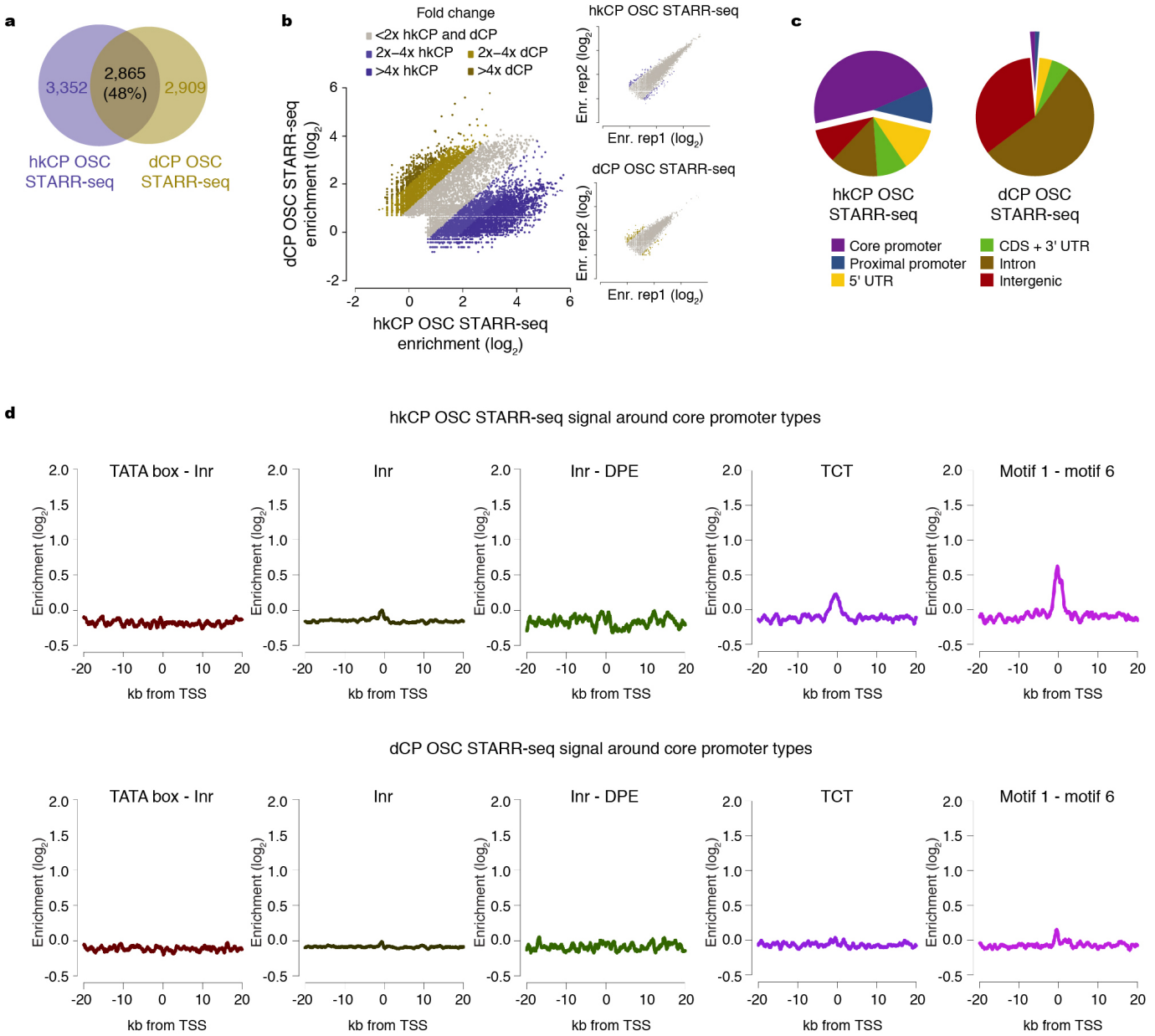
Extended Data Figure 5 | TSS-overlapping hkCP enhancers function independent of their orientation. Luciferase signals for all 17 TSS-overlapping hkCP enhancers (that is, containing one TSS or two divergent TSSs; see Supplementary Table 17) from Extended Data Fig. 3 cloned in the second orientation with respect to the TSS of the luciferase gene (bottom bar plot; the top bar plot corresponds to the initial orientation as in Extended Data Fig. 3 and is shown for comparison). In both orientations, 15 out of 17 enhancers showed activity towards the hkCP (details as in Extended Data Fig. 3). These results together with the findings in Extended Data Fig. 3 challenge the widespread notion that TSS-proximal sequences are promoters and even the concept of promoters more generally: sequences that autonomously activate gene expression—and are therefore often termed promoters—might in fact be the combination of a core promoter and a proximal enhancer. The TSS-proximal location of many housekeeping enhancers might be evolutionarily more ancient, consistent with regulatory mechanisms in simple eukaryotes such as yeast. In contrast, enhancers of genes with more complex regulation are typically located more distally, potentially simply because the several different cell-type-specific enhancers of these genes would not all fit to positions near TSSs. Consistently, such genes frequently have larger intergenic and intragenic regions[40] known to accommodate enhancers with diverse activity patterns[41].



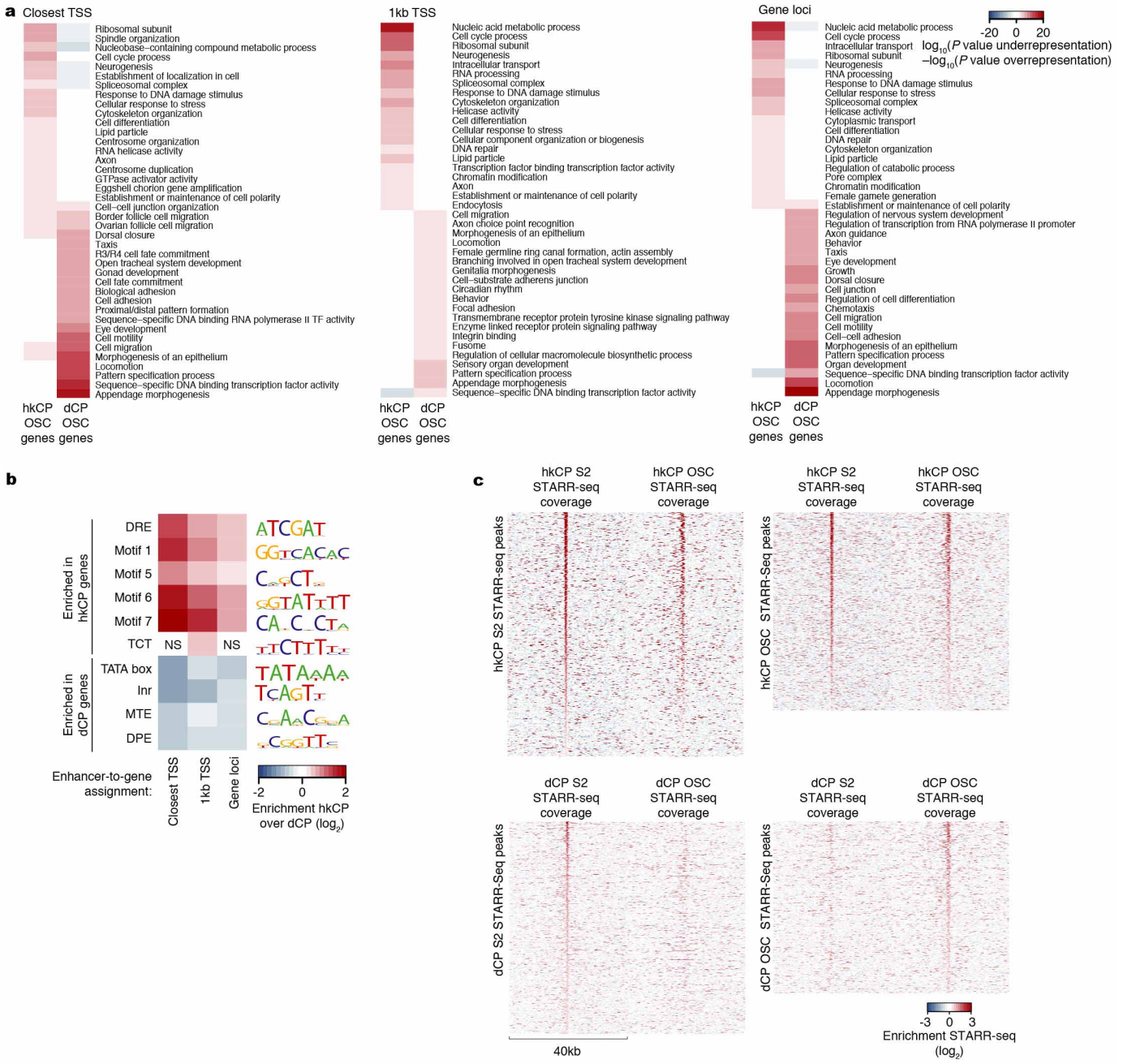
Extended Data Figure 6 | hkCP and dCP enhancers in S2 cells are associated with genes of different functions and core promoter elements. a, GO analysis of genes next to hkCP- and dCP-specific enhancers in S2 cells using different enhancer-to-gene assignment strategies (top left, 'closest TSS' as in Fig. 2; top right, '1 kb TSS'; bottom left, 'gene loci'; see Methods for details). Shown are 20 non-redundant GO categories selected from the 100 most significantly enriched categories associated with each enhancer class (see Supplementary Tables 2–4 for all categories). **b**, Enrichment of core promoter elements at genes next to hkCP- and dCP-specific enhancers in S2 cells. Similar analysis as in Fig. 2e, but using different enhancer-to-gene assignment strategies (see Methods for details). Consistent with Fig. 2e, core promoters of genes assigned to hkCP-specific enhancers are enriched in motifs 1, 5, 6, 7 and DRE, while core promoters of genes assigned to dCP-specific enhancers are enriched for TATA box, Inr, MTE and DPE motifs, irrespective of the assignment strategy.



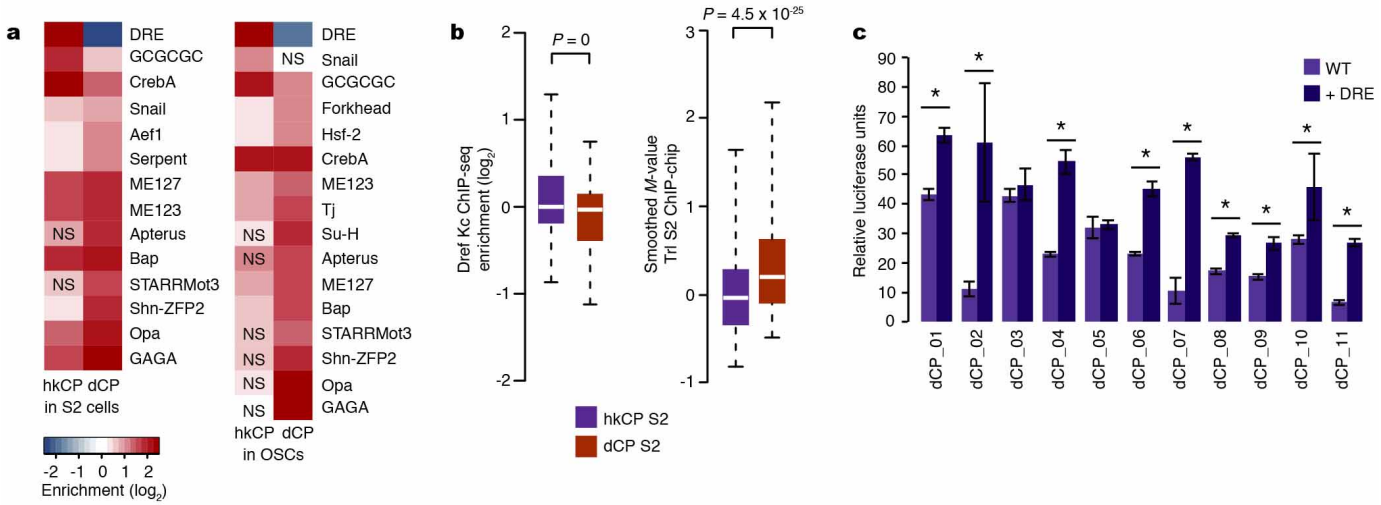
Extended Data Figure 7 | Housekeeping and developmental core promoters differ characteristically in their global enhancer preferences. As in Fig. 3b but including biological replicates with independently cloned focused bacterial artificial chromosome (BAC) libraries covering around 5 Mb of genomic sequence (BAC) and assessing the PCC at each position along these regions. GW, genome-wide screens as in Fig. 3b. The similarity observed for the TATA box- and DPE-containing core promoters (*Hsp70*, *pnr* and DSCP (dCP)) suggest that differences related to these core promoter elements might be more subtle or related to alternative mechanisms, including the potential preferences of more proximal or distal enhancers[42] or RNA polymerase II pausing and the dynamics versus stochasticity of initiation and elongation[43,44,45].



Extended Data Figure 8 | hkCP and dCP enhancers differ in OSCs. **a**, **b**, Different enhancers activate transcription from hkCP and dCP in OSCs. As Fig. 1c, d but for OSCs rather than S2 cells (data in bottom inset of **b** are re-analysed from ref. 12). **c**, Genomic distribution of hkCP and dCP enhancers in OSCs. As Fig. 2a but for OSCs rather than S2 cells. **d**, hkCP and dCP STARR-seq signal in OSCs around different core promoter types. As Extended Data Fig. 4 but for OSCs rather than S2 cells.



Extended Data Figure 9 | Differences between hkCP and dCP enhancers in OSCs. **a**, GO analysis of genes next to hkCP- and dCP-specific enhancers in OSCs. As Extended Data Fig. 6a but for OSCs rather than S2 cells (see Supplementary Tables 8–10 for all categories). **b**, Enrichment of core promoter elements at genes next to hkCP- and dCP-specific enhancers in OSCs. As Fig. 2e and Extended Data Fig. 6b but for OSCs rather than S2 cells. NS, not significant (hypergeometric $P > 0.05$). **c**, Heat maps of hkCP (top) and dCP (bottom) STARR-seq enrichments in S2 cells and OSCs. Heat maps on the left and right are centred on the summits of core-promoter-type-specific enhancers in S2 and OSCs, respectively.



Extended Data Figure 10 | The activities of hkCP and dCP enhancers are dependent on DRE and GAGA motifs, respectively. **a**, Differential motif enrichment in distally located hkCP- and dCP-specific enhancers (as in Fig. 5a but assessing enrichments of the same motif PWMs exclusively at distal enhancers >500 bp away from the closest TSSs). Key motifs including DRE and GAGA are also differentially enriched in distal hkCP- and dCP-specific enhancers. NS, not significant (FDR-corrected hypergeometric $P > 0.01$). S2 cells: hkCP $n = 790$, dCP $n = 3{,}013$; OSCs: hkCP $n = 556$, dCP $n = 2{,}555$. **b**, Distal hkCP- and dCP-specific enhancers are differentially bound by Dref and Trl, respectively. ChIP enrichments of Dref (left) and Trl (right) at S2 hkCP- and dCP-specific enhancers that are distal (>500 bp) from the closest TSSs. Equivalent to Fig. 5b, but considering exclusively TSS-distal enhancers to exclude potentially confounding effects for TSS-proximal enhancers for which it is not possible to discern whether binding occurs due to the enhancer sequence or core promoter function. The differential binding between Dref and Trl to hkCP- and dCP-specific enhancers, respectively, is also found in Kc167 cells, in which the Dref ChIP-seq experiment had been performed (data not shown). **c**, Addition of DRE motifs to dCP enhancers increases their activity towards hkCP. Relative luciferase activity values (firefly/*Renilla*) for 11 dCP enhancers without DRE motifs (wild type (WT), light purple) and with 3 DRE motifs flanking the enhancers on each side (+DRE, dark purple). *$P < 0.05$, one-sided unpaired Student's *t*-test; error bars denote the s.d. of three biological replicates.