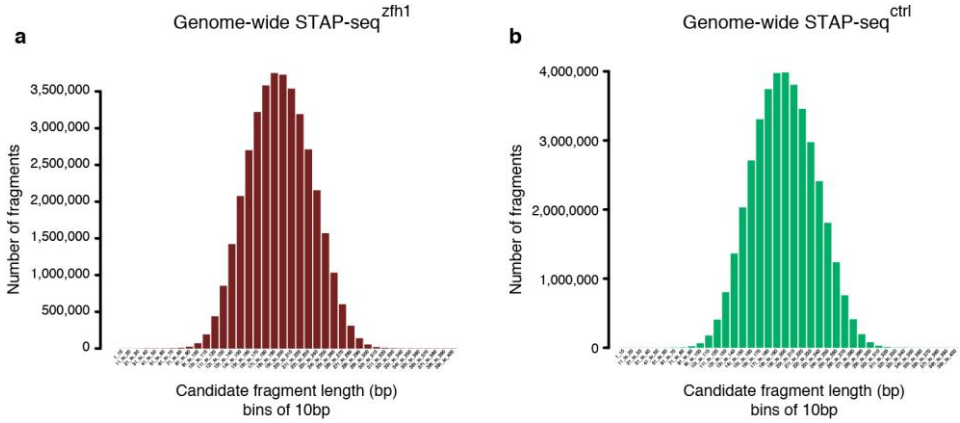
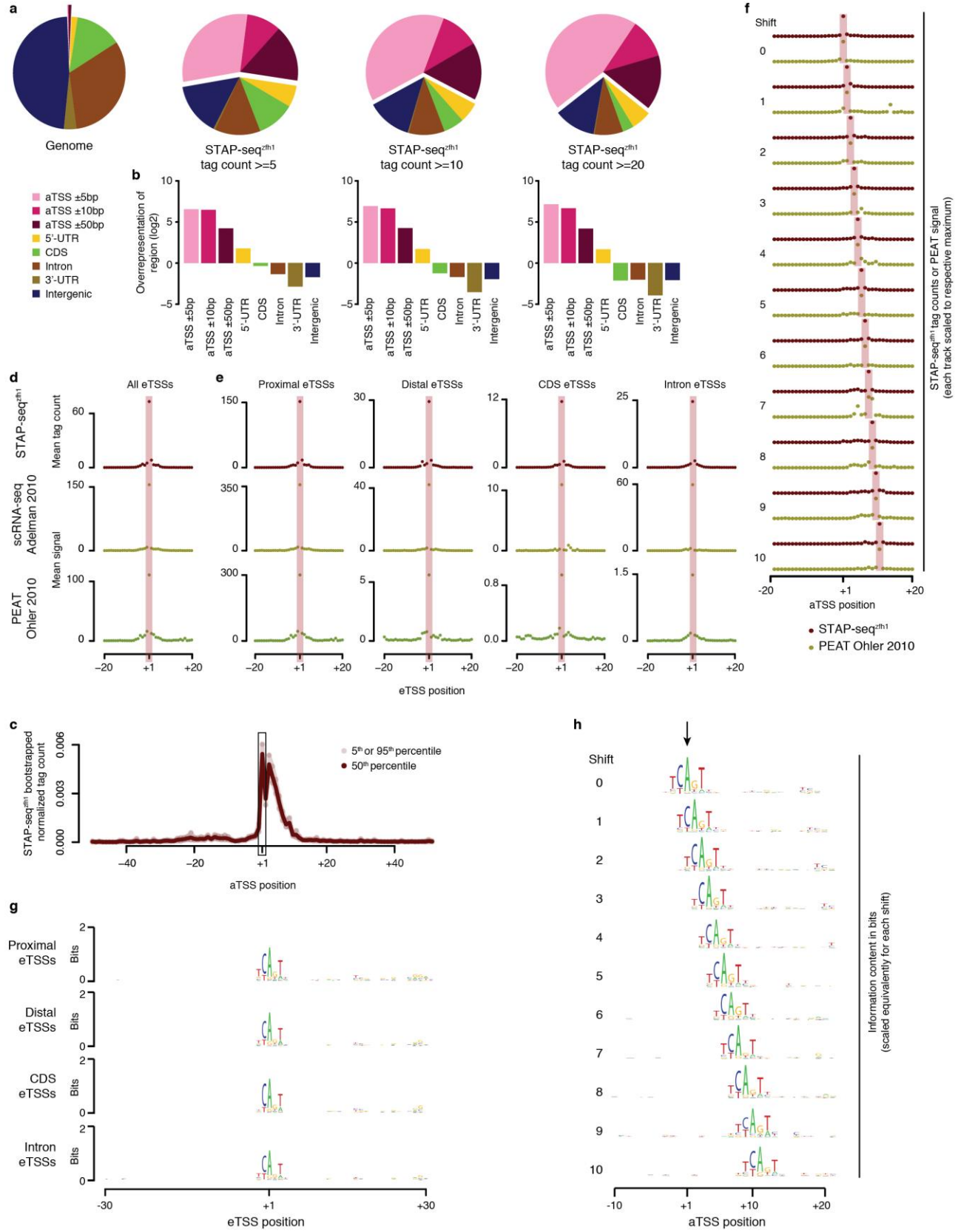


Supplementary Figure 1

Candidate fragment length distribution.

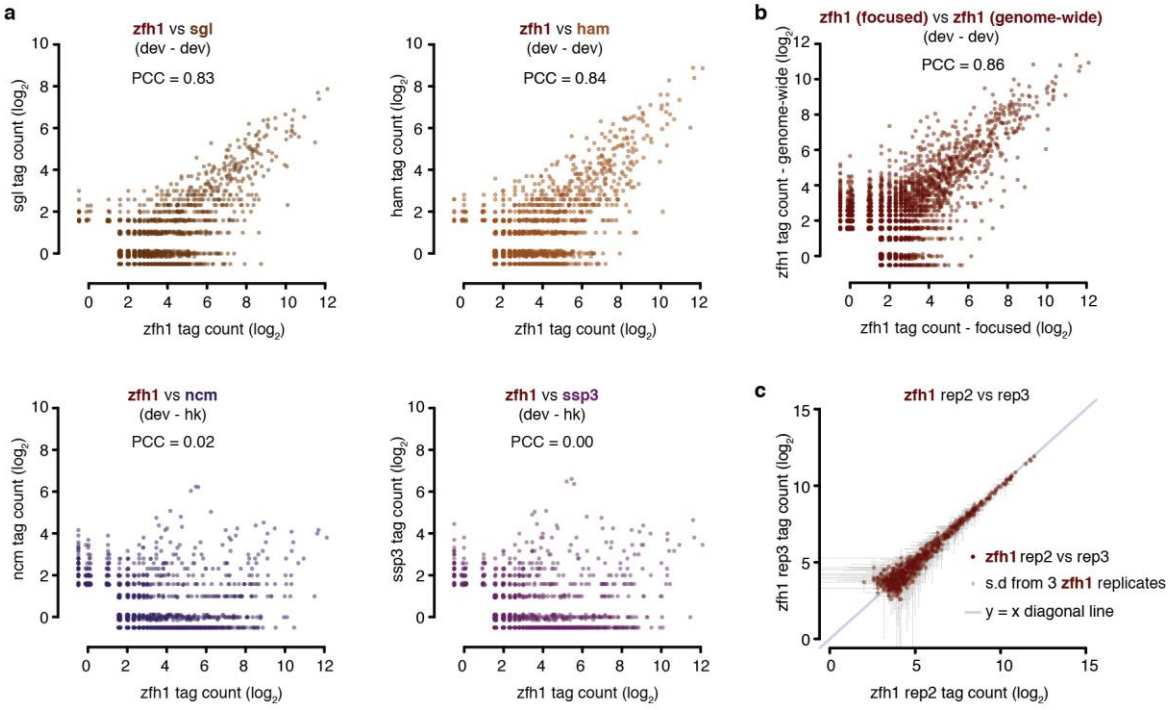
(**a**) Candidate fragment length distributions of STAP-seq^{zfh1} and (**b**) STAP-seq^{ctrl} input fragments in bins of 10bp (40 bins in total).



Supplementary Figure 2

Genomic distribution of eTSSs and agreement of eTSS positions with endogenous TSSs and established sequence motifs.

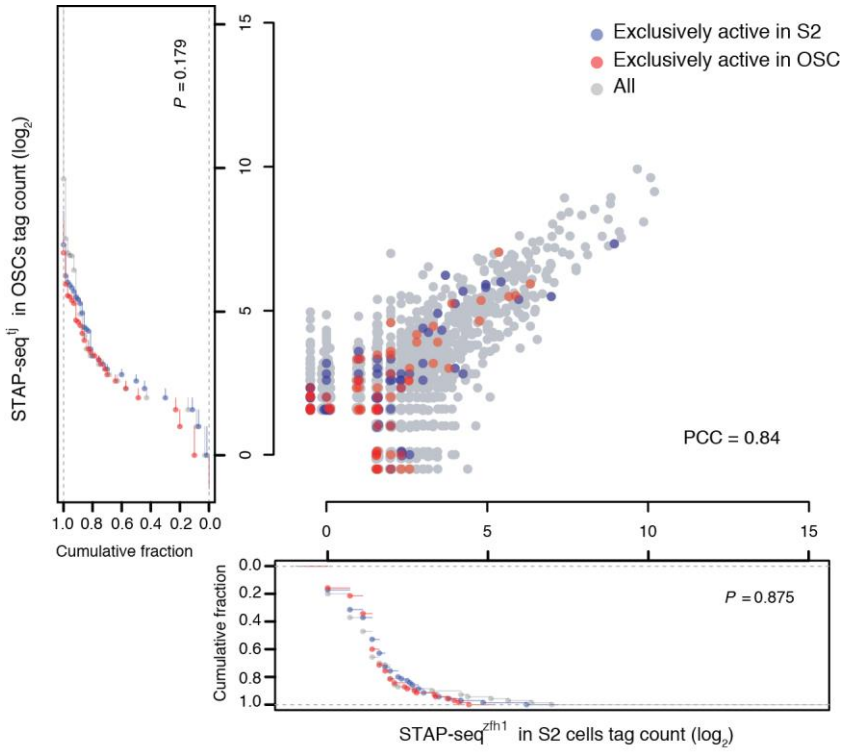
(**a**) Pie charts of the genomic distribution of all genomic sequences that initiate transcription with tag counts >=5, >=10, >=20, respectively, in response to the *zfh1* enhancer compared to the *D. melanogaster* genome (leftmost pie chart; the sector indicates aTSSs ±50bp that make up 0.32% of the genome)**.** (**b**) Bar plots visualizing the enrichment of the regions from **a,** over the genome. (**c**) Metagene profile of average normalized STAP-seq^{zfh1} tag counts around aTSSs, including the 5th and 95th percentiles determined by bootstrapping. (**d**) Metagene profiles of STAP-seq^{zfh1}, scRNA-seq[18] and PEAT[19] at all eTSSs. (**e**) As in **d**, but specifically for proximal, distal, within coding DNA sequence (CDS) and intronic eTSSs. (**f**) Agreement of STAP-seq eTSSs and embryo-derived TSSs by PEAT[19] that are shifted from aTSSs between 1 to 10 nucleotides (each row is scaled to the respective maximum; see **Figure 1e** for the equivalent comparison with scRNA-seq from S2 cells[18]). (**g**) Sequence-logos depicting position-specific nucleotide frequencies for eTSSs that are aTSS-proximal or -distal, within CDS, or intronic. (**h**) Sequence logos of eTSSs that coincide with aTSS (shift 0, top row, the +1 position of aTSSs is indicated by the arrow) or are proximally-misaligned by 1 to 10 base pairs (bps) from aTSSs (logos are moved closer together but not scaled differently relatively to each other).



Supplementary Figure 3

Reproducibility of STAP-seq^{zfh1} and comparisons between STAP-seq screens with different enhancers.

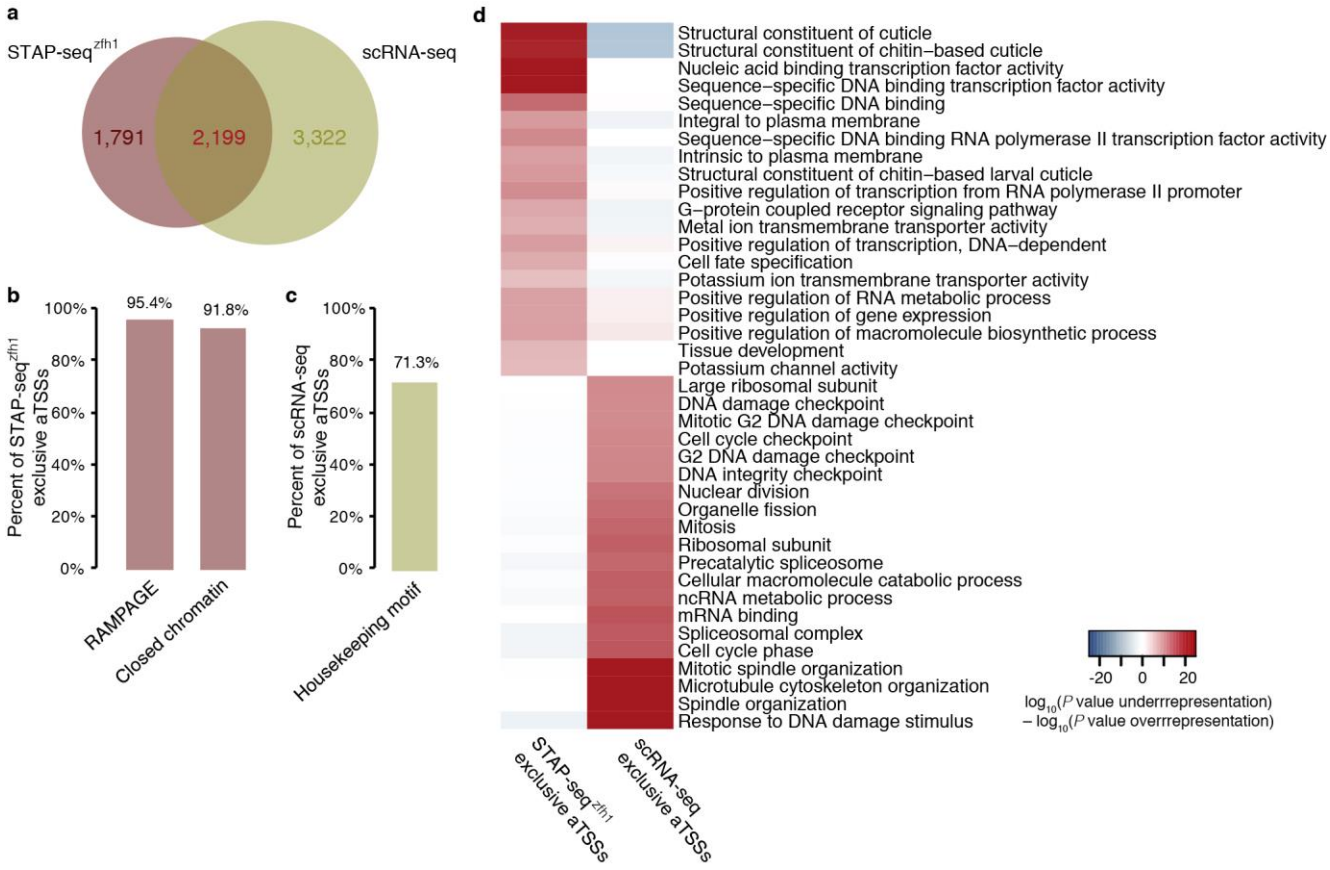
(**a**) Scatterplots comparing focused STAP-seq^{zfh1} screens with different screens with diverse developmental (dev; top) and housekeeping enhancers (hk; bottom; see **Figure 2** for further comparisons; PCC: Pearson correlation coefficient). (**b**) Scatterplot depicting the similarity between focused and genome-wide STAP-seq^{zfh1} screens. (**c**) Scatterplot for two independent biological replicates of focused STAP-seq^{zfh1} screens, including standard deviations (s.d) calculated across three independent biological replicates (error bars).



Supplementary Figure 4

Induced activities are consistent across cell types.

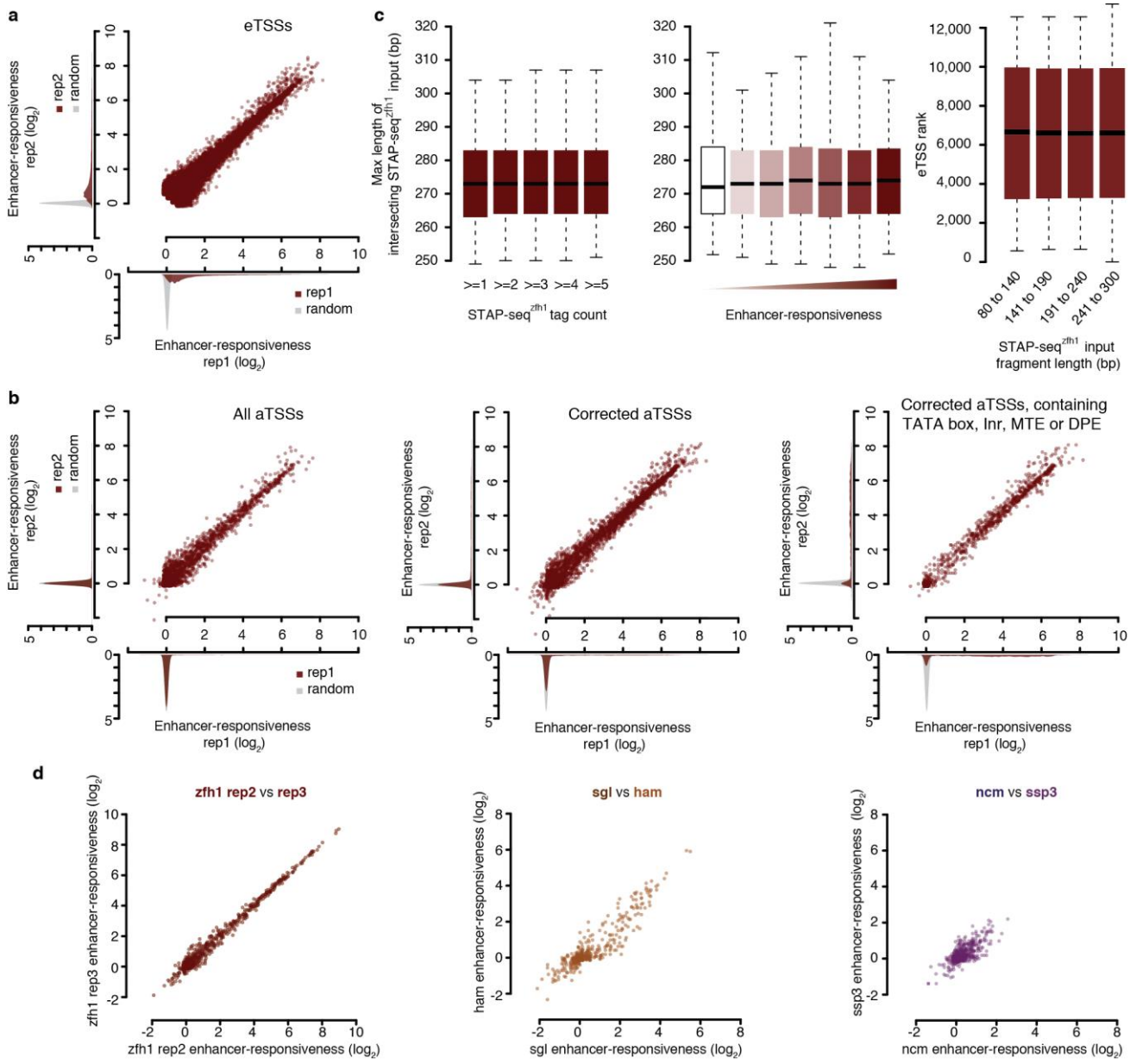
Scatterplot depicting STAP-seq tag counts for STAP-seq^{zfh1} in S2 cells (x-axis) versus STAP-seq^{tj} in OSCs (y-axis) and their similarity (expressed as PCC: Pearson correlation coefficient). TSSs that are endogenously – as measured by GRO-seq[22,23] – and exclusively active in S2 cells or OSCs are labeled blue or red, respectively. Aligned to each axis are the respective cumulative distributions used to assess the difference between TSSs that are endogenously exclusively active in either S2 cells or OSCs by Kolmogorov–Smirnov tests (*P* values indicated). The scatter plot corresponds to **Figure 3a**.



Supplementary Figure 5

STAP-seq is complementary to methods that assess endogenous transcription initiation.

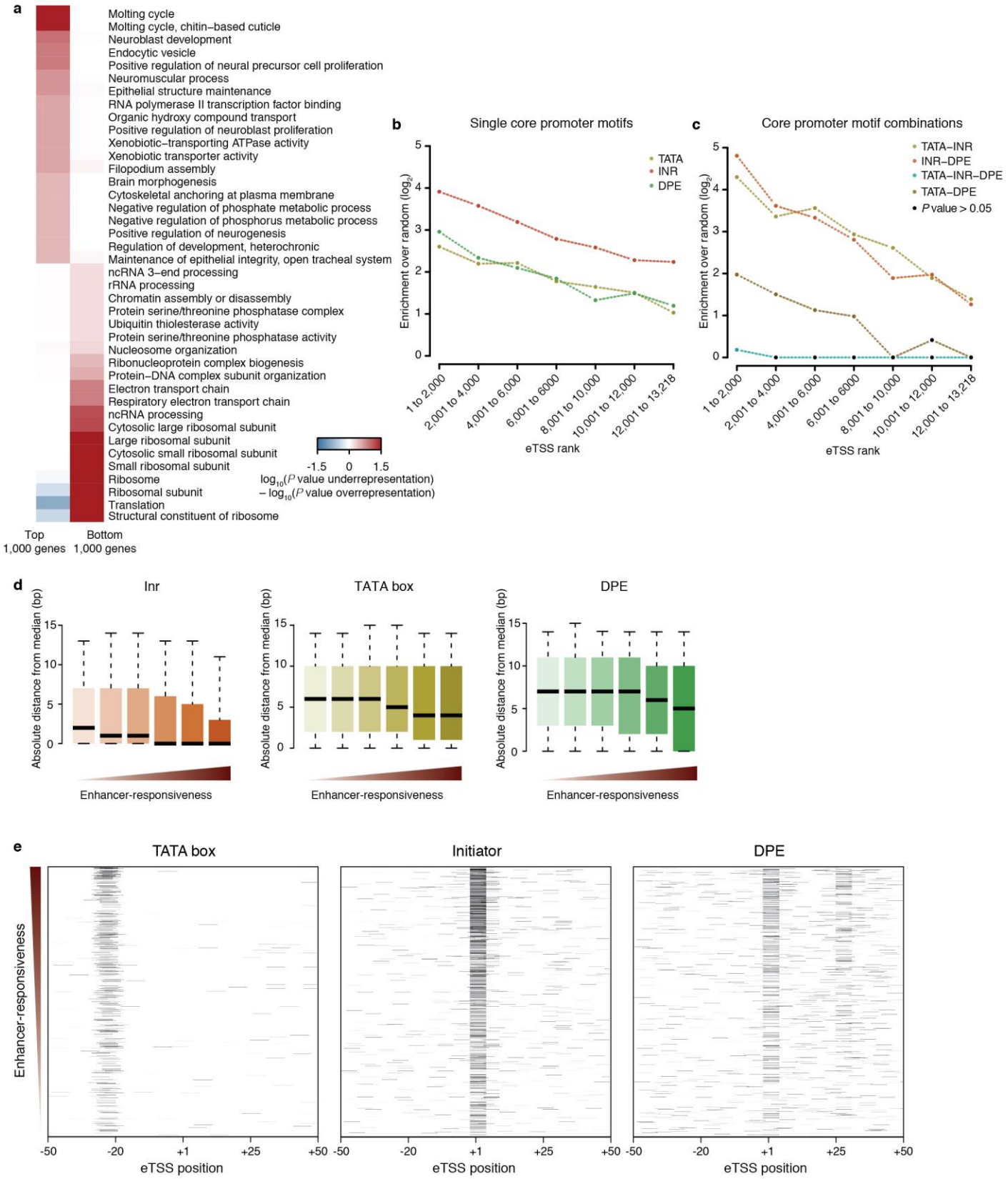
(**a**) Venn diagram depicting the overlap of aTSSs detected by STAP-seq^{zfh1} and scRNA-seq in S2 cells genome-wide (see **Figures 3d** and **e** for an equivalent analysis of the focused STAP-seq screens). (**b**) Proportion of aTSSs uniquely detected by STAP-seq^{zfh1} that are also detected during different developmental stages by RAMPAGE[44] (left bar) or are in closed chromatin in S2 cells (right bar). (**c**) aTSSs uniquely detected by scRNA-seq that contain housekeeping core promoter motifs. (**d**) Gene Ontology analysis of aTSSs uniquely detected by either STAP-seq^{zfh1} or scRNA-seq. The results from **b**-**d** suggest that aTSSs uniquely detected by scRNA-seq are housekeeping-type core promoters (see also **Figures 3e** and **f**), while aTSSs uniquely detected by STAP-seq^{zfh1} are endogenously not active in S2 cells (**b**, right bar) but in other cell types (**b**, left bar).



Supplementary Figure 6

A wide range of enhancer-responsiveness.

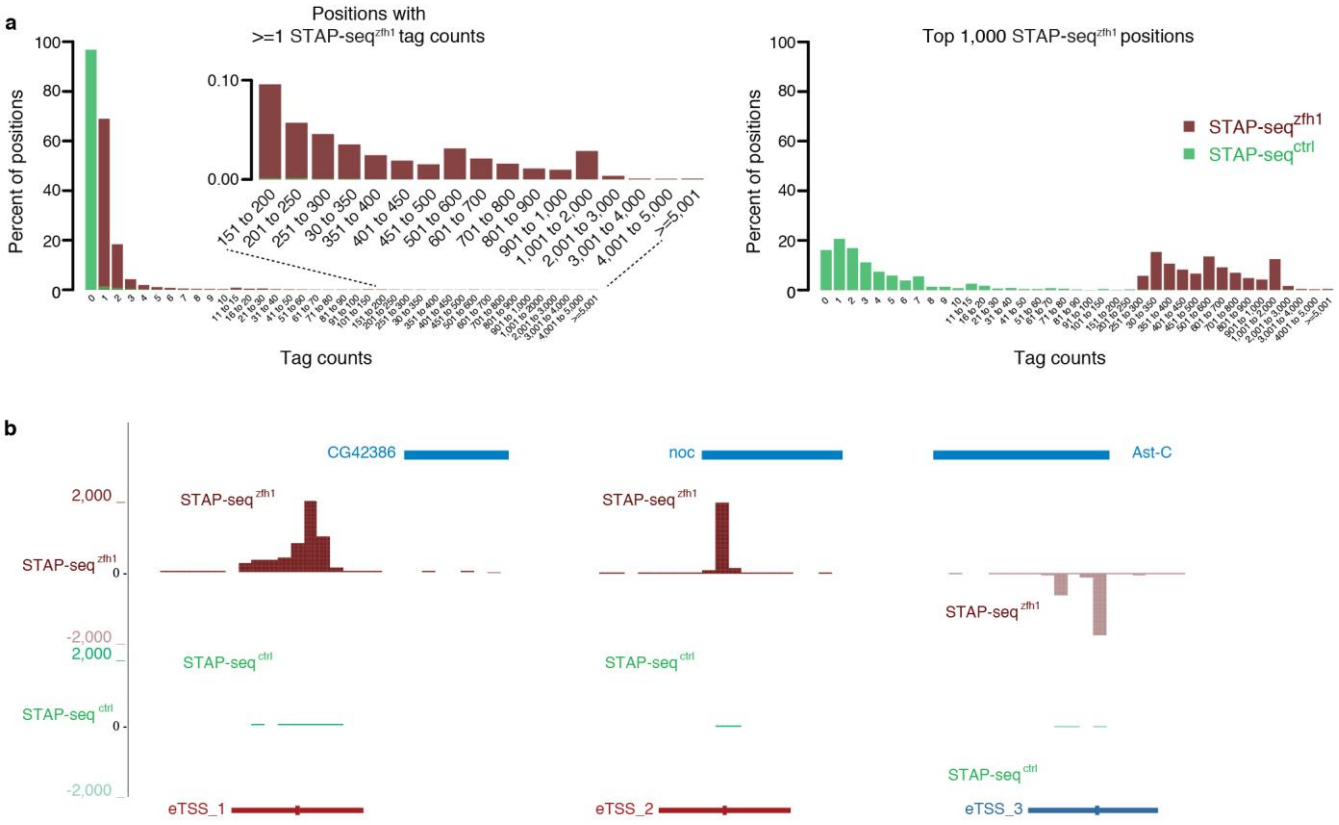
(**a**) Scatterplots depicting the range of enhancer-responsiveness as determined by STAP-seq^{zfh1} over STAP-seq^{ctrl} at eTSSs for replicate (rep) 1 versus 2. The distributions of enhancer-responsiveness at the respective TSSs (brown) and at random positions (grey) are shown by density plots along the axes. (**b**) As in **a,** but, at all, position-corrected, and TATA box, Initiator (Inr), MTE or DPE-containing and position-corrected aTSSs. (**c**) TSS strengths are independent of candidate lengths. Boxplots depicting maximal length of candidate STAP-seq^{zfh1} fragments intersecting positions with the corresponding STAP-seq^{zfh1} tag counts (left) or eTSSs of different enhancer-responsiveness (middle), and ranks of eTSSs which +1 positions intersect candidate STAP-seq^{zfh1} fragments of different lengths (right). Center line: median; limits: interquartile range; whiskers: 5th and 95th percentiles. (**d**) Housekeeping enhancers activate transcription from candidate sequences at a reduced dynamic range compared to developmental enhancers. Scatterplots depicting the range of enhancer-responsiveness of the indicated focused STAP-seq screens at aTSSs in genomic regions covered by the focused candidate libraries for developmental (left and middle) and housekeeping (right) enhancers. To account for the broad nature of initiation at housekeeping core promoters, enhancer-responsiveness in **d,** is calculated in a window of ±50bp around aTSSs.


a
Molting cycle
Molting cycle, chitin–based cuticle
Neuroblast development
Endocytic vesicle
Positive regulation of neural precursor cell proliferation
Neuromuscular process
Epithelial structure maintenance
RNA polymerase II transcription factor binding
Organic hydroxy compound transport
Positive regulation of neuroblast proliferation
Xenobiotic–transporting ATPase activity
Xenobiotic transporter activity
Filopodium assembly
Brain morphogenesis
Cytoskeletal anchoring at plasma membrane
Negative regulation of phosphate metabolic process
Negative regulation of phosphorus metabolic process
Positive regulation of neurogenesis
Regulation of development, heterochronic
Maintenance of epithelial integrity, open tracheal system
ncRNA 3–end processing
rRNA processing
Chromatin assembly or disassembly
Protein serine/threonine phosphatase complex
Ubiquitin thiolesterase activity
Protein serine/threonine phosphatase activity
Nucleosome organization
Ribonucleoprotein complex biogenesis
Protein–DNA complex subunit organization
Electron transport chain
Respiratory electron transport chain
ncRNA processing
Cytosolic large ribosomal subunit
Large ribosomal subunit
Cytosolic small ribosomal subunit
Small ribosomal subunit
Ribosome
Ribosomal subunit
Translation
Structural constituent of ribosome
Top 1,000 genes
Bottom 1,000 genes
-1.5 0 1.5
log10(P value underrepresentation)
– log10(P value overrepresentation)
b
Single core promoter motifs
TATA
INR
DPE
Enrichment over random (log2)
1 to 2,000
2,001 to 4,000
4,001 to 6,000
6,001 to 6000
8,001 to 10,000
10,001 to 12,000
12,001 to 13,218
eTSS rank
c
Core promoter motif combinations
TATA–INR
INR–DPE
TATA–INR–DPE
TATA–DPE
P value > 0.05
Enrichment over random (log2)
eTSS rank
d
Inr
TATA box
DPE
Absolute distance from median (bp)
Enhancer-responsiveness
e
TATA box
Initiator
DPE
Enhancer-responsiveness
-50
-20
+1
+25
+50
eTSS position

Supplementary Figure 7

Biological significance and sequence properties of eTSSs with different enhancer-responsiveness.

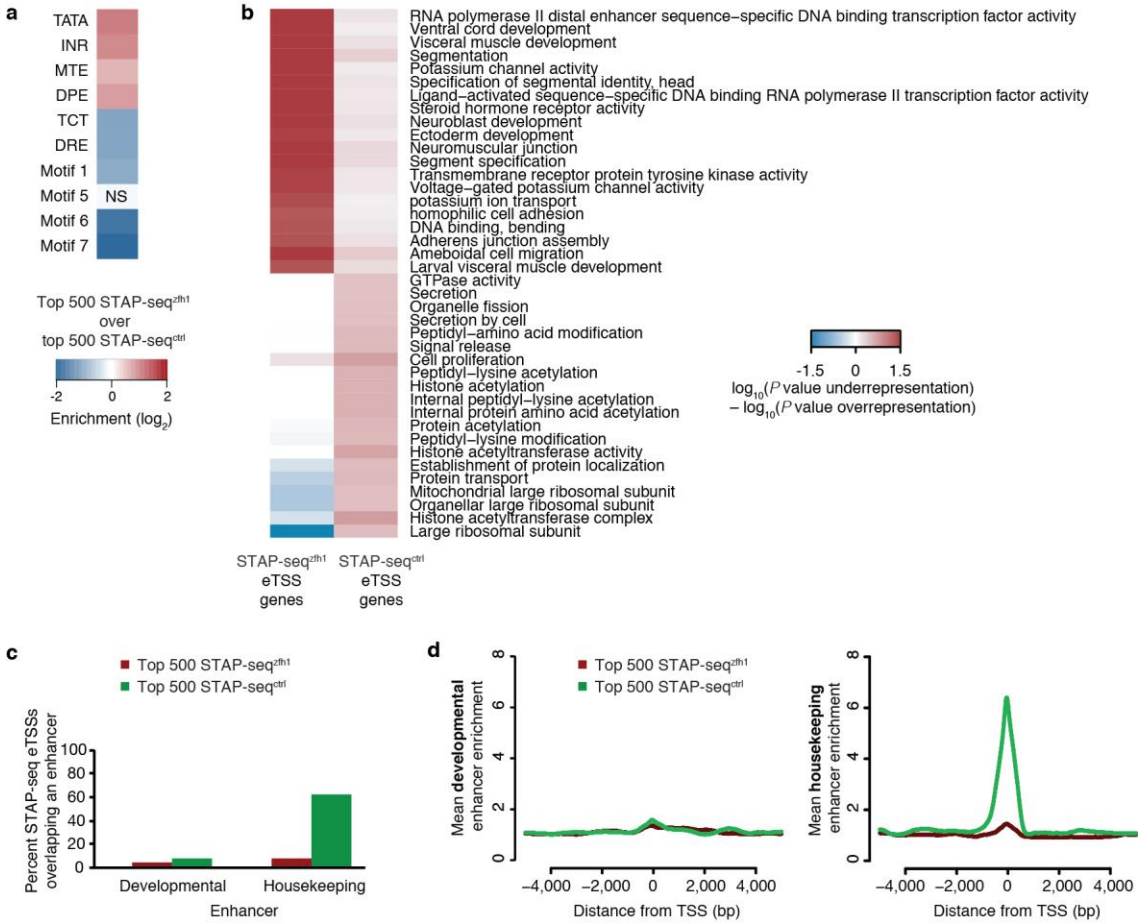
(**a**) Gene Ontology analysis of the top and bottom 1,000 genes associated with the strongest and weakest eTSSs, respectively (see **Figure 4f** for an equivalent analysis restricted to eTSSs containing exclusively TATA box, Inr, MTE or DPE). (**b-c**) Enrichment of single (**b**) or combinations (**c**) of core promoter motifs in eTSSs of different ranks compared to random genomic sequences. As has been observed before[20], the combination of TATA box and DPE motif is less strongly enriched compared to the TATA box and Inr or Inr and DPE motif pairs and the combination of all three motifs rarely occurs in the same eTSS. (**d**) Boxplots depicting motif deviation (distance in base pairs, bp) from their consensus positions (determined as the average position across all eTSSs) versus enhancer-responsiveness of eTSSs. Center line: median; limits: interquartile range; whiskers: 5^{th} and 95^{th} percentiles (**e**) Heatmaps depicting positional occurrences of core promoter motifs around eTSSs. Note that TATA boxes are somewhat less positionally constraint, as previously observed for mammalian core promoters[31].



Supplementary Figure 8

Induced and basal activities measured by STAP-seq^{zfh1} and STAP-seq^{ctrl}, respectively, reveal low basal activities for top eTSSs.

(**a**) Histograms depicting distributions of normalized STAP-seq^{zfh1} and STAP-seq^{ctrl} tag counts at positions that are covered by at least 1 tag (left) or the top 1,000 STAP-seq^{zfh1} positions (right). The inset on the left shows that STAP-seq^{zfh1} but not STAP-seq^{ctrl} reaches high normalized tag counts. (**b**) Representative screenshots of the top three eTSSs, depicting high STAP-seq^{zfh1} and low STAP-seq^{ctrl} tag counts.



Supplementary Figure 9

Sequences with the highest basal activities are at housekeeping genes and overlap housekeeping-type enhancers.

(**a**) Core promoter motif enrichment analysis of candidate sequences with the highest basal (STAP-seq^{ctrl}) versus the highest induced (STAP-seq^{zfh1}) activities (top 500 each). NS: not significant. (**b**) Gene Ontology (GO) analysis for genes associated with eTSSs that show the highest basal (STAP-seq^{ctrl}, right) versus the highest induced (STAP-seq^{zfh1}, left) activities. (**c**) Fraction of candidates as in **a,** that show STARR-seq enrichment values of at least 3-fold. Developmental and housekeeping STARR-seq data are from ref. 10. (**d**) As in **c**, but plotting average STARR-seq signals around the most prominent, or the center of dispersed TSSs.