Supplementary Information for

Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during *cis*-regulatory evolution

Cosmas D. Arnold[1,5], Daniel Gerlach[1,4,5], Daniel Spies[1], Jessica A. Matts[2-4], Yuliya A. Sytnikova[2,3], Michaela Pagani[1], Nelson C. Lau[2,3], Alexander Stark[1]

[1] Research Institute of Molecular Pathology IMP, Vienna Biocenter VBC, Dr Bohr-Gasse 7, 1030 Vienna, Austria.
[2] Brandeis University, Department of Biology, Waltham MA, USA.
[3] Rosenstiel Basic Medical Science Research Center at Brandeis University, Waltham MA, USA.

[4] Present Adresses: Boehringer Ingelheim RCV GmbH & Co KG, Vienna, Austria (D.G.) and Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater OK, USA (J.A.M.)

[5] These authors contributed equally to this work.

Correspondence should be addressed to A.S. (stark@starklab.org).

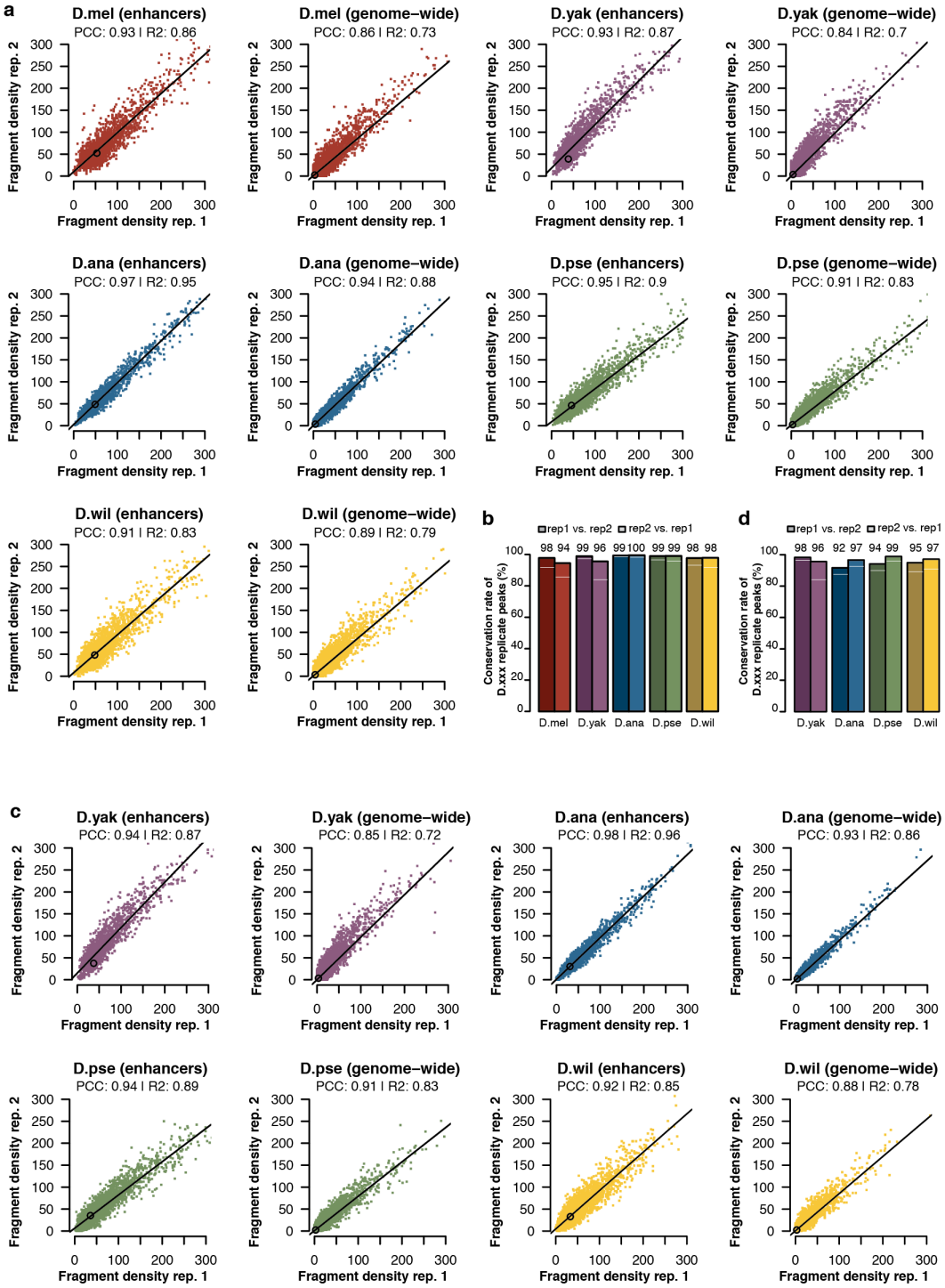
This file includes:

Supplementary Figs 1 to 15
Supplementary Tables 1, 3 and 5
Supplementary References

Nature Genetics: doi:10.1038/ng.3009

Table of Contents:

Nature Genetics: doi:10.1038/ng.3009

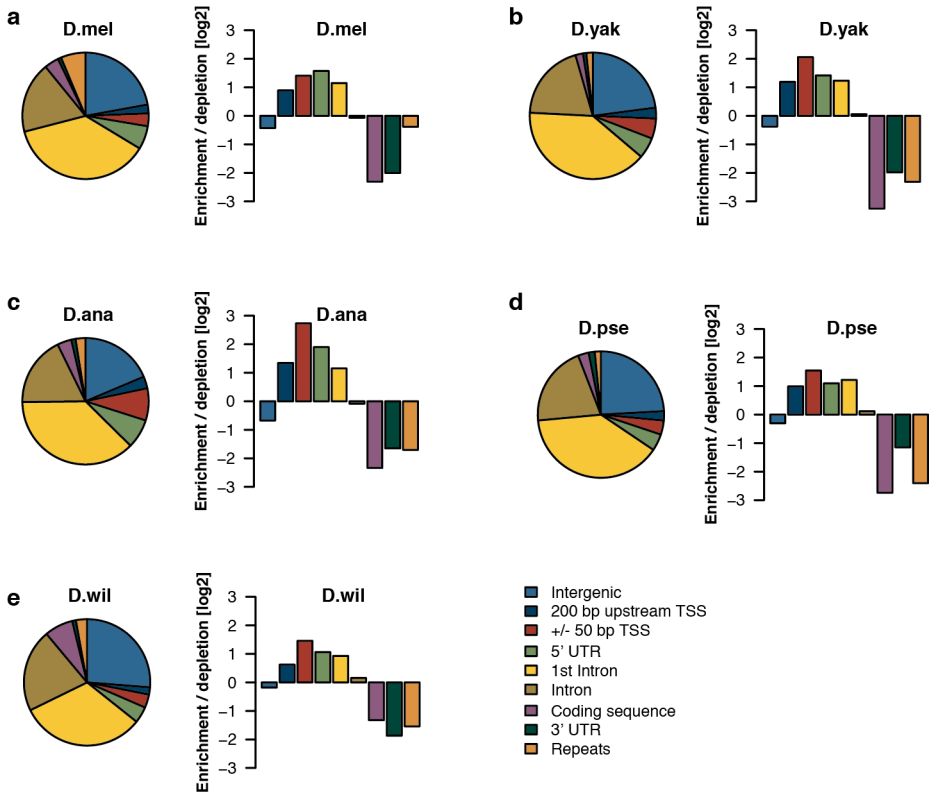


Nature Genetics: doi:10.1038/ng.3009

Supplementary Figure 1 | High quantitative and qualitative reproducibility of STARR-seq in *D. melanogaster* S2 cells.

(**a**) The quantitative reproducibility of STARR-seq in two independent biological replicates was assessed at enhancer peak summits (*D. melanogaster* (*D.mel*), 2,325 peaks; *D. yakuba* (*D.yak*), 2,293 peaks; *D. ananassae* (*D.ana*), 2,096 peaks; *D. pseudoobscura* (*D.pse*), 3,469 peaks; *D. willistoni* (*D.wil*), 2,860 peaks) and for 100,000 positions randomly sampled from the genome (common *D. melanogaster* coordinates for all species). Each data point represents the fragment density for both replicates normalized to 1 million mapped fragments (FPM). The Pearson correlation coefficient (PCC) and the coefficient of determination (R^2) for the linear fit (plus the regression line) are indicated in each subplot. The open black circle shows median values of coverage for replicate 1 versus replicate 2. (**b**) Qualitative reproducibility of STARR-seq measuring the consistency of enhancer calls between enhancers called in replicate 1 evaluated with enrichment data from replicate 2 equivalently to the assessment of conservation (bar height, relaxed settings with $P \leq 0.05$; white line, $P \leq 0.001$). The second bar for each species evaluates replicate 2 against replicate 1. (**c**,**d**) The same data are shown as for **a** and **b**, but using peak calls and fragment densities in the respective, original *Drosophila* genomes before coordinate translation.

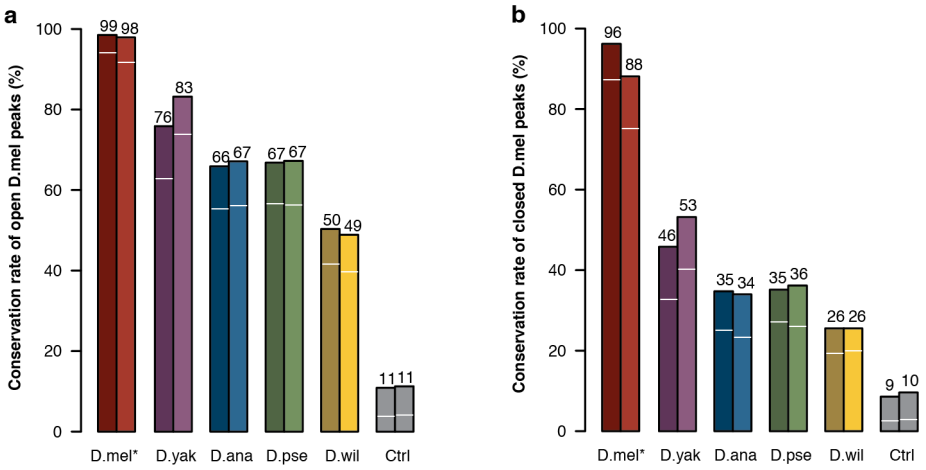
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 2 | Similar genomic distribution of STARR-seq enhancers for all five *Drosophila* species in *D. melanogaster* S2 cells.

The pie charts show the absolute genomic distribution of enhancers across different functional regions, and the bar charts show enrichment or depletion relative to overall region sizes in the genome (**a**, *D. melanogaster*; **b**, *D. yakuba*; **c**, *D. ananassae*; **d**, *D. pseudoobscura*; **e**, *D. willistoni*). Globally, the majority of identified enhancers were located within introns (53.2–59.4%) and in intergenic regions (18.7–26.3%), as described in ref. 1. Overall the genomic distribution of enhancers is comparable among the five *Drosophila* species. These data show that enhancer location with respect to different genomic regions is similar for all five *Drosophila* species.

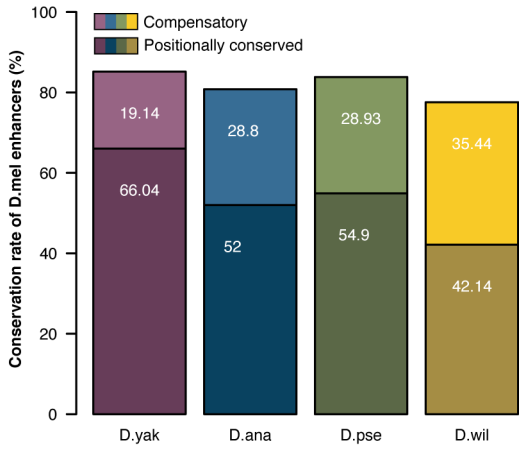
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 3 | Functional conservation of *open* and *closed* *D. melanogaster* S2 cell enhancers in *D. melanogaster* S2 cells.

D. melanogaster S2 cell enhancers were classified as open or closed depending on their accessibility in DNase I hypersensitivity (DHS) sequencing assays as described previously[1]. The functional conservation rates of (**a**) 1,554 open and (**b**) 771 closed *D. melanogaster* enhancers in the 4 other *Drosophila* species are shown (see **Figure 1c** for details of the conservation rate analysis). The conservation rate of open enhancers is roughly twice as high as for closed enhancers, whereas both show similar reproducibility in independent replicates (*D. melanogaster* bars, marked by an asterisk).

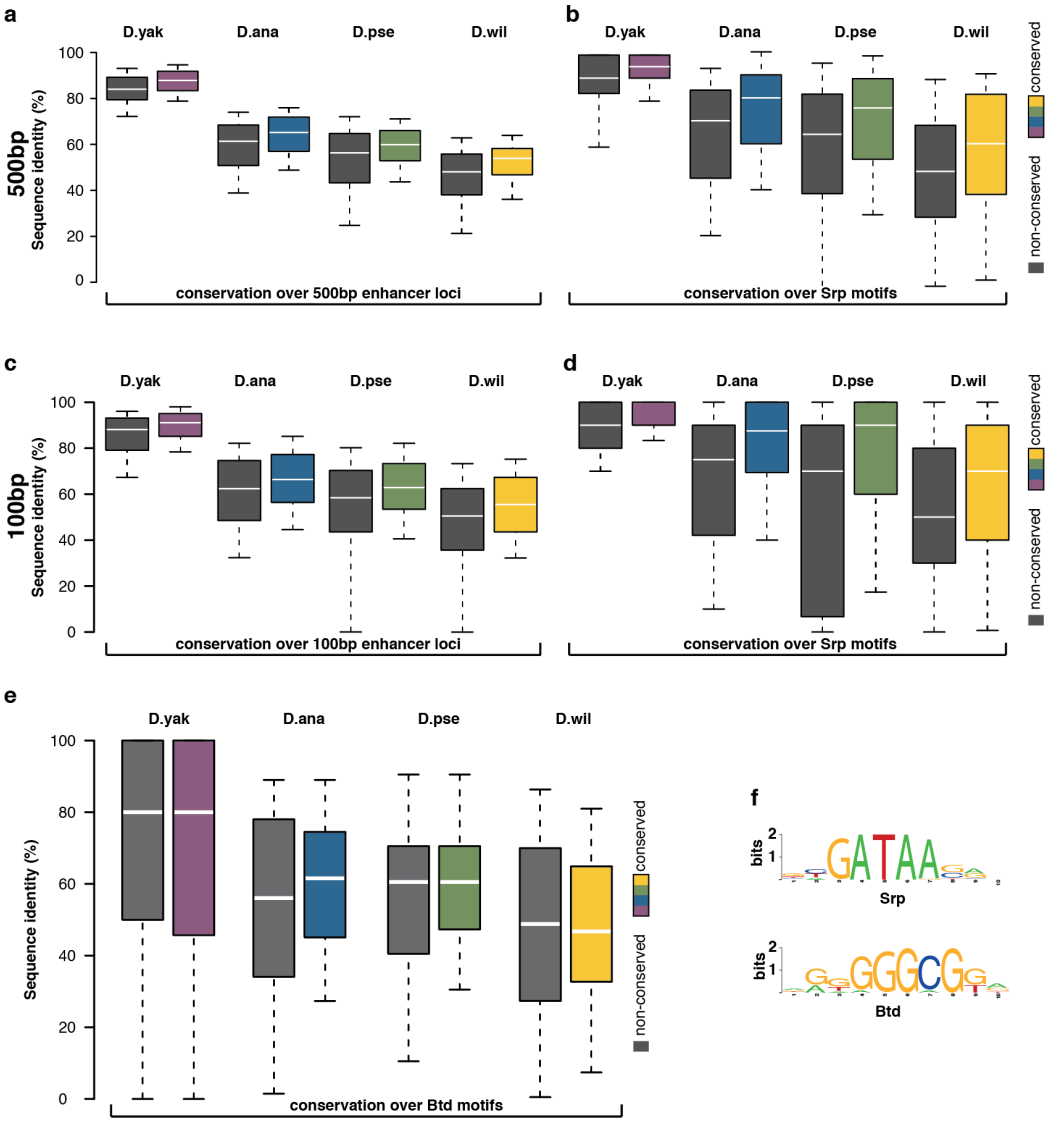
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 4 | Positional and compensatory conservation of enhancers.

The number of positionally conserved *D. melanogaster* enhancers declines with evolutionary distances. The drop in positionally conserved enhancers, however, is balanced by an increasing number of compensatory enhancers within the same respective gene loci, leading to a similar number of enhancers per gene locus. This might stabilize gene expression levels or confer regulatory robustness[2–4]. Note that this plot shows data from analysis of the two biological replicates combined. The white numbers inside the bars indicate the fraction of enhancers of each category as percentage of the total number of *D. melanogaster* enhancers (1,552) assigned to 1,201 gene loci.

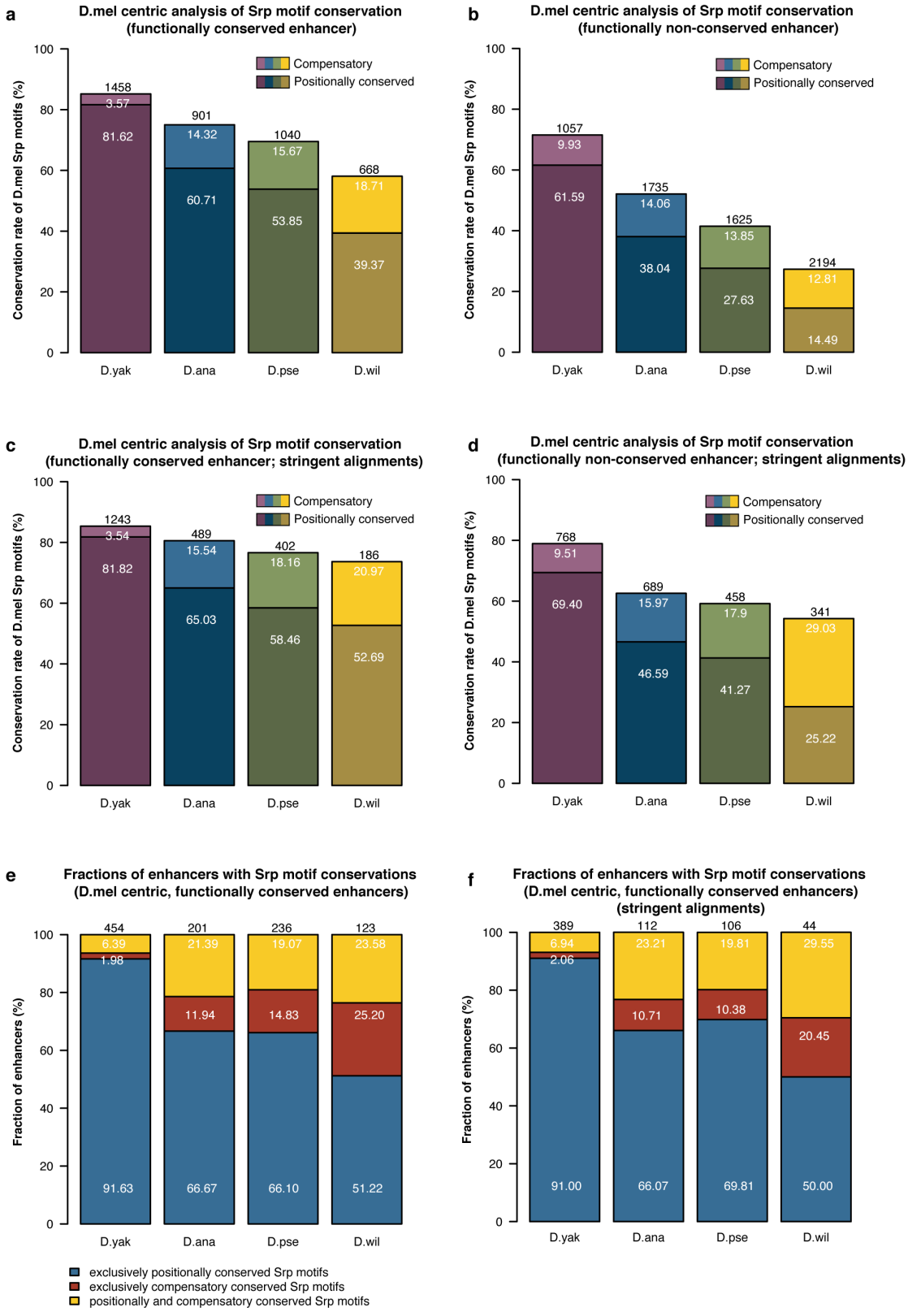
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 5 | Motif conservation by positional sequence constraints.

(**a**) Pairwise sequence identity for functionally conserved (colored) and non-conserved (gray) enhancers along the entire 500-bp enhancer sequence or (**c**) for a 100-bp core enhancer sequence (boxes depict the median and the interquartile range, and whiskers depict the 10th and 90th percentiles). (**b**,**d**) Sequence identity as in **a** and **c**, respectively, but restricted to positions that overlap with motifs of the transcription factor Serpent (Srp). Note: **a** and **b** show the same data as **Figure 3a**,**b** to allow for a comparison with **c** and **d**, demonstrating that the results are robust with respect to the lengths of the analyzed regions (for **b**–**d**, n = 214, 338, 413, 196, 361, 216, 366 and 174, respectively). (**e**) Sequence identity as in **b**, but for the motif of the transcription factor Buttonhead (Btd), which is not expressed in S2 cells. The largely overlapping sequence identities of the Btd motifs in conserved and non-conserved enhancers indicate that the motifs for Btd in conserved enhancers in S2 cells are not under constraint. (**f**) Position weight matrix logos for the Srp and Btd motifs.

Nature Genetics: doi:10.1038/ng.3009



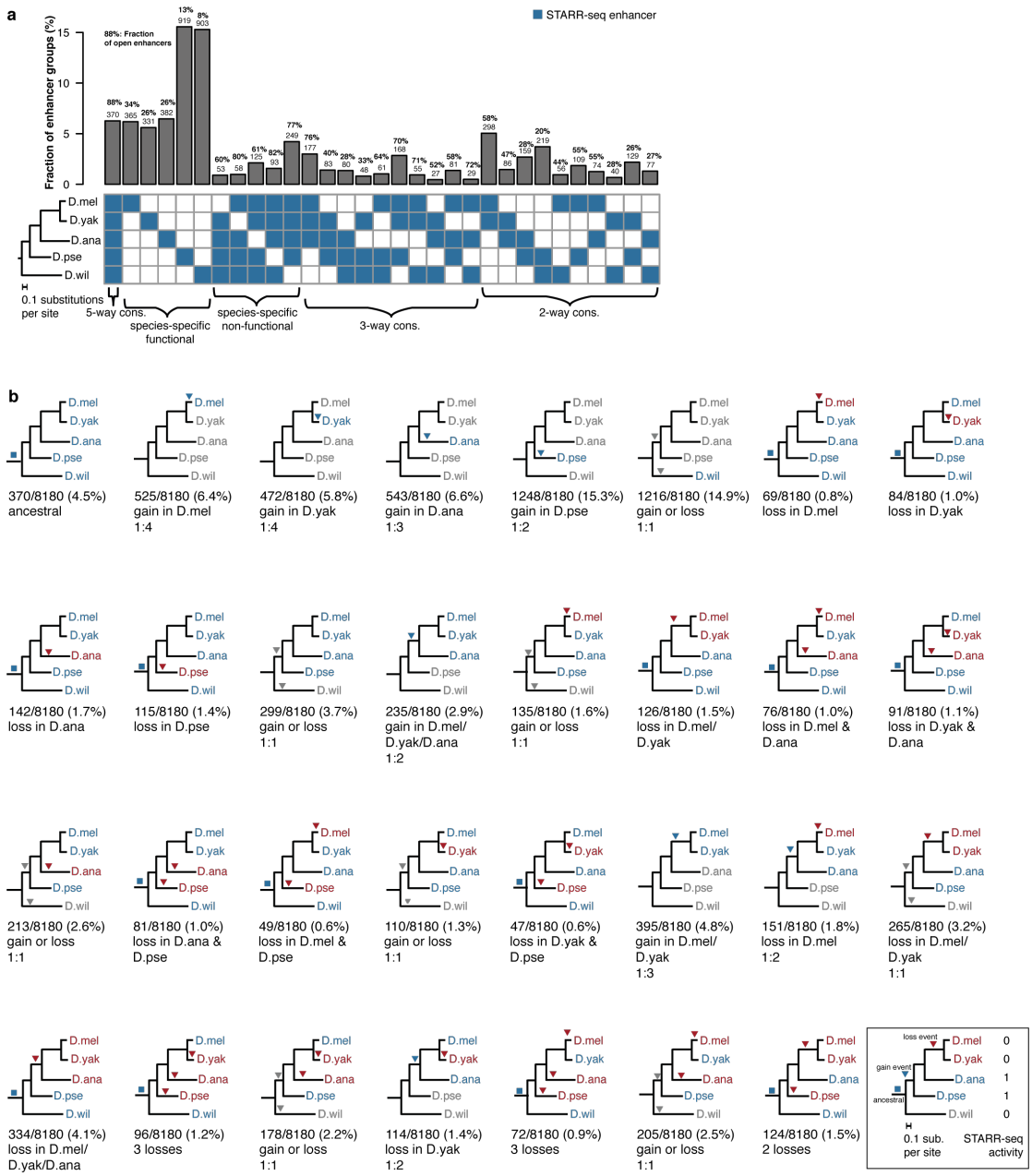
Nature Genetics: doi:10.1038/ng.3009

Supplementary Figure 6 | Positional and compensatory conservation of TF motifs in functionally conserved and non-conserved enhancers.

(**a**) Rate of positionally conserved and compensatory *D. melanogaster* Srp motifs in relation to all four other *Drosophila* species in functional conserved enhancer regions. The total number of *D. melanogaster* Srp motifs for each comparison is shown above the bars (black), and percent conservation values are shown within the bars (white). (**b**) Plot as in **a**, but in functionally non-conserved enhancer regions. (**c**,**d**) Plots as in **a** and **b**, but limiting the evaluation to a subset of well-aligned enhancer regions that have no undefined nucleotides (Ns) in the pairwise alignments and have non-gapped orthologous ends. Together this shows that motif turnover is common and that the loss of positionally conserved Srp motifs can be compensated by the gain of Srp motifs at different positions within the same enhancer. Further, Srp motifs are conserved at much higher levels in conserved enhancers compared to non-conserved enhancers, suggesting that they are important for S2 cell enhancer function. (**Supplementary Fig. 5**). When assessing well-aligned sequences (**c**), motif turnover maintains the number of serpent motifs at high levels of around 80%, even over large evolutionary distances.

(**e**,**f**) Relative contribution of compensatory motif turnover increases with evolutionary distance. (**e**) Fraction of functionally conserved enhancers with the same number of Srp motifs between species with positionally conserved motifs (blue), motifs conserved within an individual enhancer but not at the same position (compensatory; red) or a mix of positionally conserved and compensatory conserved Srp motifs (yellow). The total number of enhancers for each comparison is shown above each bar; percentages per category are plotted in white within the bars. (**f**) Plot as in **e**, but considering only a subset of enhancers that are well aligned (**c**,**d**). For the vast majority of conserved enhancers, the motifs are exclusively positionally conserved in closely related species such as *D. yakuba* in **e**. However, the fraction of enhancers with compensatory motifs increases significantly at larger evolutionary distances.

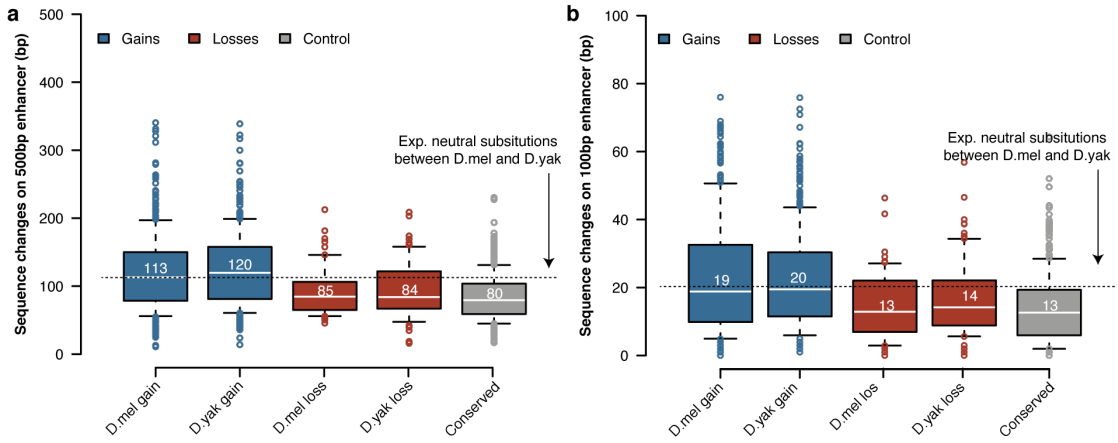
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 7 | Phylogeny of enhancer gain and loss events.

(**a**) Enhancer occurrences at non-redundant positions across species (binary representation in which blue boxes indicate enhancer presence/function). (**b**) Gain (blue triangles) and loss (red triangles) events assigned by parsimony to different branches of the phylogenetic tree. We assigned a gain event if two or more loss events would otherwise have to be assumed, but indicate the gain-versus-loss ratios below the trees (e.g., 1:4 = one gain or four losses; unclear events are shown in gray). Overall, the phylogeny of all 8,180 non-redundant enhancers identified in the genomes of the 5 species are shown on 31 (= 5^2 – 1) different trees.

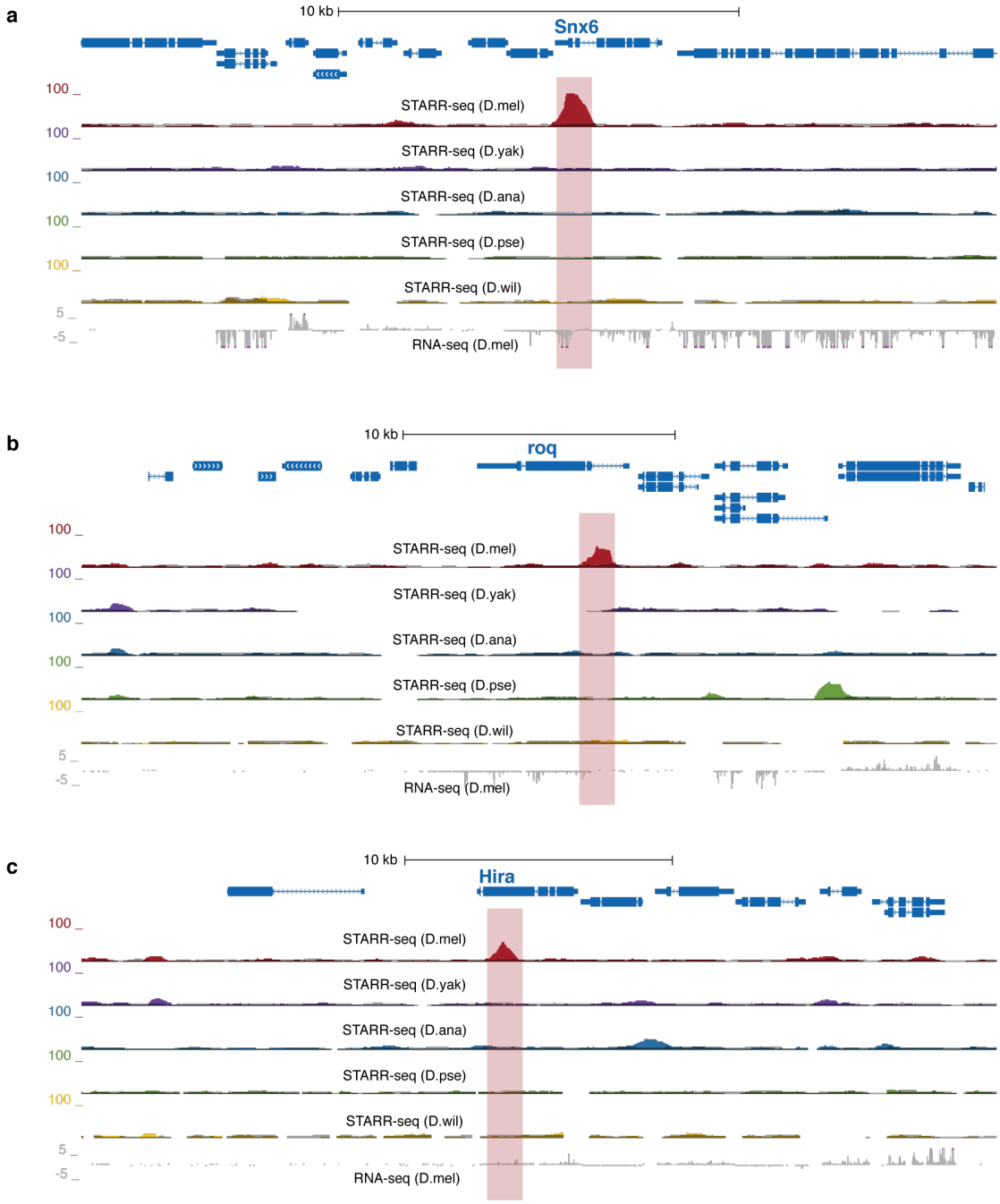
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 8 | Sequence changes of *D. melanogaster* and *D. yakuba* gained enhancers are similar to expected neutral substitutions between the two species.

(**a**) Same data as **Figure 4e** (boxes depict the median and the interquartile range, and whiskers depict the 10th and 90th percentiles; outliers are shown individually). (**b**) As in **a**, but for 100-bp core enhancer sequences. This shows that the patterns of sequence conservation in gained and lost enhancers are consistent between 500-bp enhancer sequences and shorter regions of 100 bp centered on the enhancer peak summit.

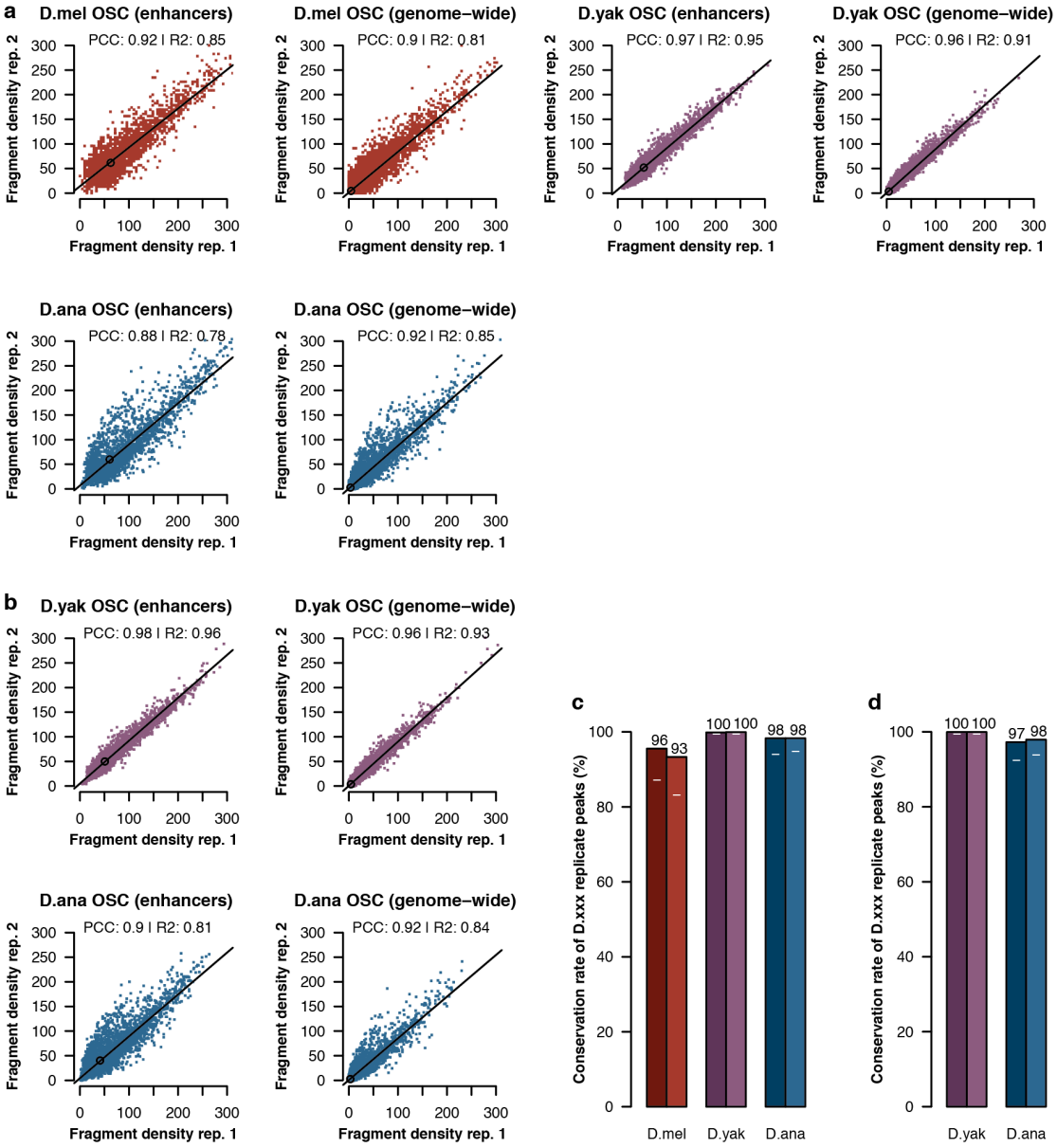
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 9 | Newly gained enhancers in *D. melanogaster* are associated with expressed genes.

UCSC Genome Browser screenshots of expressed genes in S2 cells that are exclusively associated with a newly gained (*D. melanogaster*–specific) enhancer (inputs in gray; *y*-axis labels depict normalized fragment counts). (**a**) *Sxn6* (RPKM 98.4). (**b**) *roq* (RPKM 25.9). (**c**) *Hira* (RPKM 12.3). RNA-seq data in *D. melanogaster* S2 cells are from ref. 1.

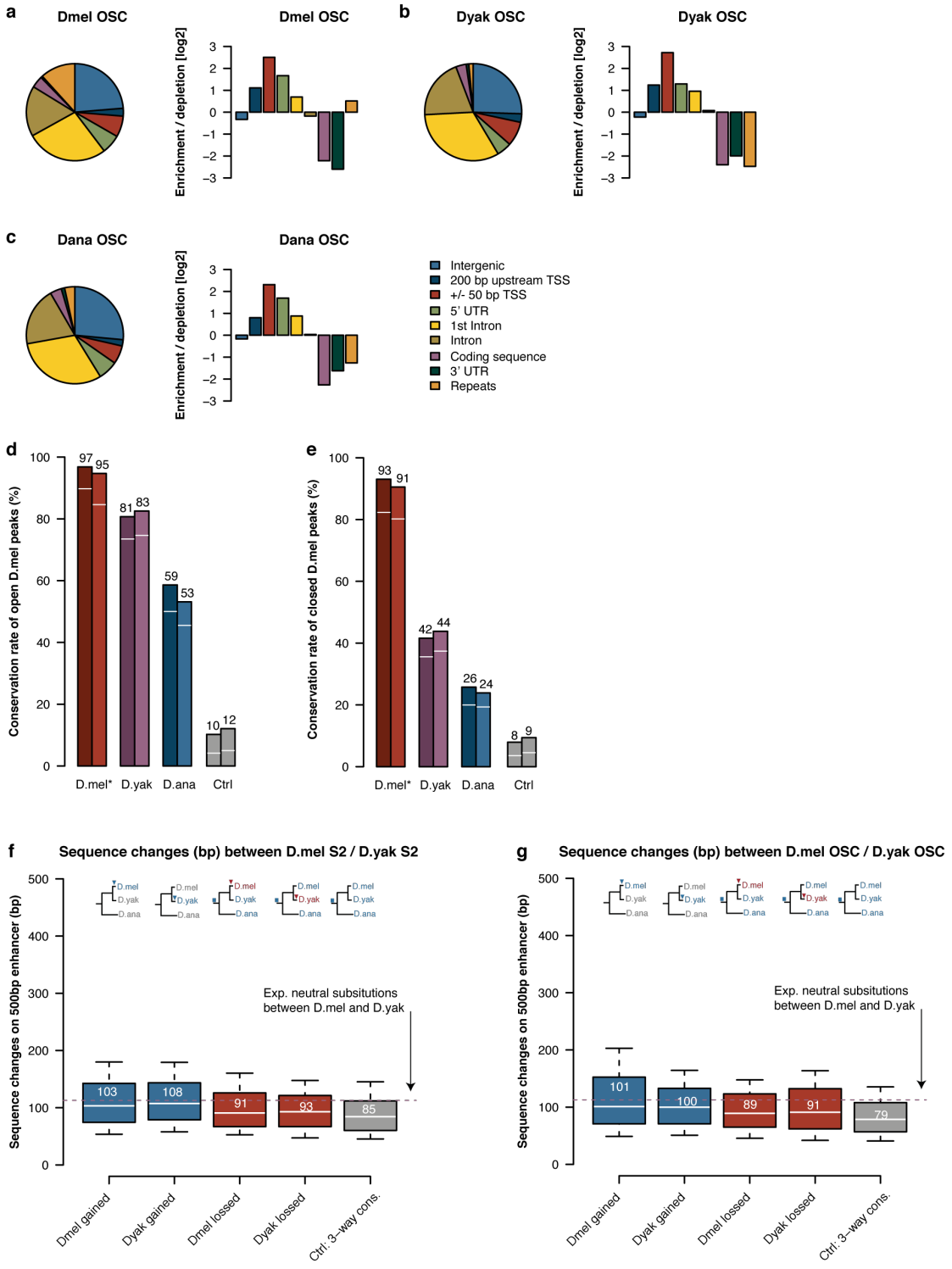
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 10 | High quantitative and qualitative reproducibility of STARR-seq in *D. melanogaster* ovarian somatic cells (OSCs).

(**a**) The quantitative reproducibility of STARR-seq in OSCs in two independent biological replicates was assessed at combined enhancer peak summits (*D. melanogaster*, 3,342 peaks; *D. yakuba*, 3,233 peaks; *D. ananassae*, 2,859 peaks) as in **Supplementary Figure 1a**. (**b**) Same data as in **a**, but using peak calls and fragment densities in the respective *Drosophila* genome coordinates before coordinate translation. (**c**) Qualitative reproducibility of STARR-seq in OSCs as in **Supplementary Figure 1b**. (**d**) Same as **c**, but using peak calls and fragment densities in the respective *Drosophila* genome coordinates before coordinate translation.

Nature Genetics: doi:10.1038/ng.3009



Nature Genetics: doi:10.1038/ng.3009

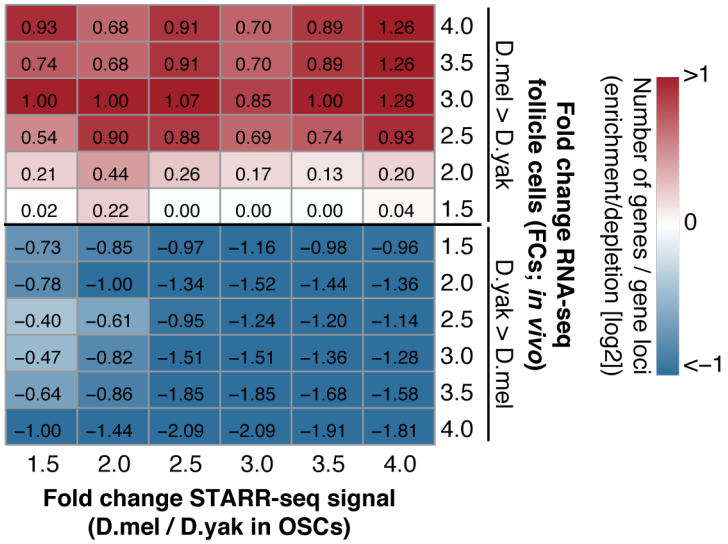
Supplementary Figure 11 | Genomic distribution, functional conservation and sequence changes of STARR-seq OSC enhancers.

(**a**–**c**) Similar genomic distribution of STARR-seq enhancers for three *Drosophila* species in *D. melanogaster* OSCs. Genomic distribution analysis for (**a**) *D. melanogaster*, (**b**) *D. yakuba* and (**c**) *D. ananassae* enhancers in OSCs as in **Supplementary Figure 2.** Globally, the majority of identified enhancers were located in introns (44.1–52.8%) and in intergenic regions (23.7–26.5%; see ref. 1).

(**d**,**e**) Functional conservation of open and closed *D. melanogaster* OSC enhancers in *D. melanogaster* OSCs. *D. melanogaster* OSC enhancers were classified as open and closed as described previously[1]. The conservation rates of (**d**) 2,269 open and (**e**) 1,073 closed *D. melanogaster* OSC enhancers in *D. yakuba* and *D. ananassae* (see **Figs. 1c** and **6a** for details).

(**f**,**g**) Number of sequence changes in *D. melanogaster* and *D. yakuba* gained, lost or deeply conserved S2 or OSC enhancers are similar. (**f**) Analysis as in **Figure 4e,** however, based on three species only (*D. melanogaster*, *D. yakuba*, *D. ananassae*) to allow the direct comparison between S2 cells and OSCs (boxes depict the median and the interquartile range, and whiskers depict the 10th and 90th percentiles). (**g**) As in **f**, but for OSC enhancers. The numbers of sequence changes for the different enhancer categories are highly similar between **f** and **g**, confirming the results shown in **Figure 4e** and suggesting that the reported numbers hold more generally, independent of the respective cell types**.**

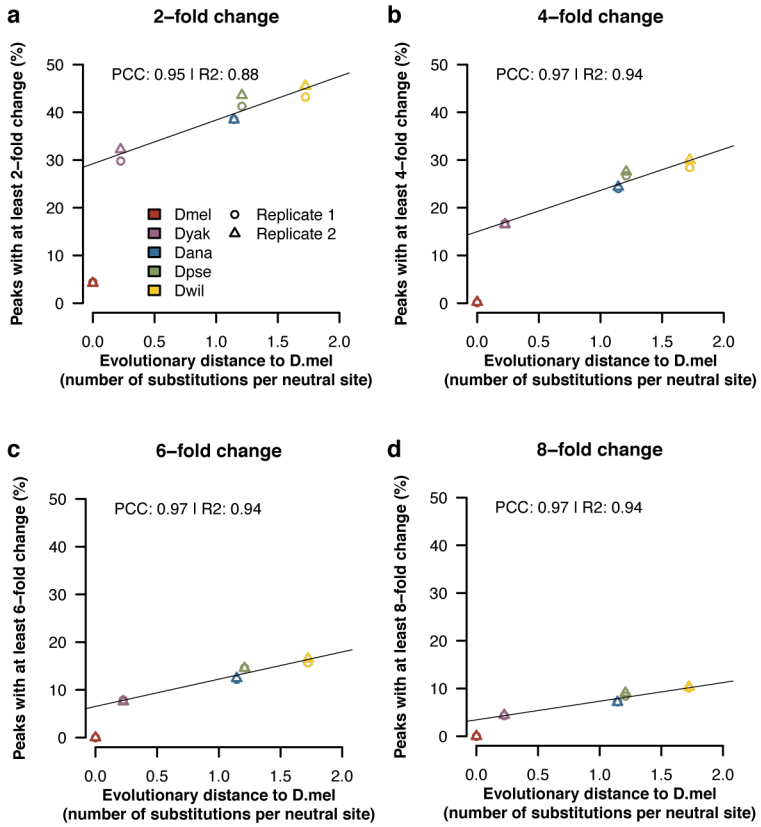
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 12 | Changes in OSC enhancer activities and follicle cell *in vivo* gene expression between *D. melanogaster* and *D. yakuba* correlate globally.

Same data and heat-map presentation as in **Figure 6d**, but with matrix cells colored according to enrichments irrespective of their significance.

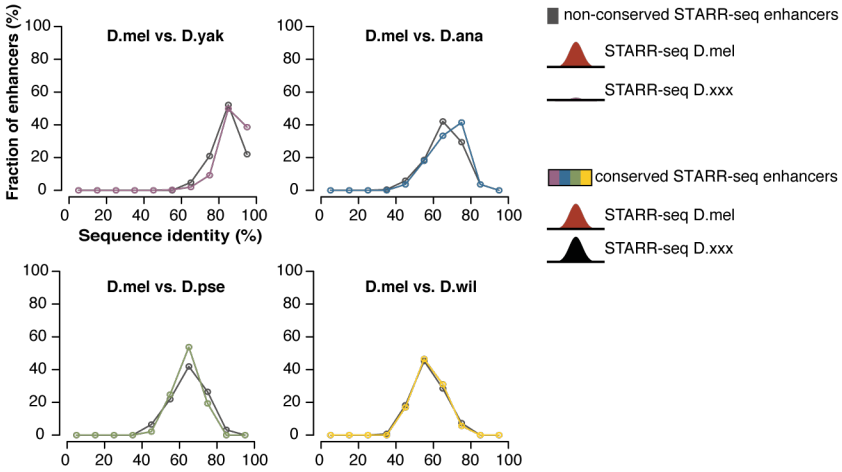
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 13 | Differences in quantitative enhancer strength follows a molecular clock.

Enhancer strength diverges with increasing evolutionary distance linearly, with the number of substitutions per neutral site (branch length) similar to qualitative enhancer conservation (**Fig. 1c**). The strong correlation of evolutionary distance and the fraction of enhancers with at (**a**) 2-fold, (**b**) 4-fold, (**c**) 6-fold and (**d**) 8-fold change in enhancer strength on non-redundant loci between *D. melanogaster* and other *Drosophila* species shows that enhancer strength is also conserved and follows a molecular clock. Note that the *D. melanogaster* replicate comparison considers replicate 1 against replicate 2 within the same species.

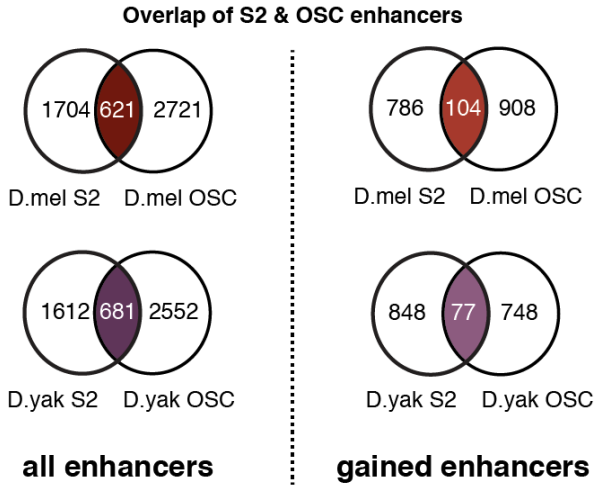
Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 14 | Global range of sequence identities for functionally conserved and non-conserved enhancers.

Sequence identity distributions for functionally conserved (colored lines) and non-conserved (dark gray lines) enhancers between *D. melanogaster* and other *Drosophila* species. The distributions are largely overlapping, suggesting that there is no selective pressure on the overall enhancer sequence. In addition, the two extreme boundaries of the distributions indicate that sequences can be up to 95% identical (between *D. melanogaster–D. yakuba*) yet without conserved function (only active in the *D. melanogaster* genome), whereas enhancer function can be conserved despite as little as 39% sequence identity between *D. melanogaster–D . willistoni*.

Nature Genetics: doi:10.1038/ng.3009



Supplementary Figure 15 | S2 cell and OSC enhancer gains are nearly additive.

Enhancers gained in S2 cells or OSCs show only limited overlap (right column), such that the number of gained enhancers is nearly additive for both cell types in *D. melanogaster* (top) and in *D. yakuba* (bottom; enhancer gains for both cell types are defined on the basis of three-way analyses considering only *D. melanogaster*, *D. yakuba* and *D. ananassae* as outgroup). The overlap of the gained enhancers is of the same magnitude as for the overlap of all enhancers in S2 cells and OSCs (left column), which suggests that different enhancers are gained in different cell types and the total number of enhancer gains in more complex tissues or organisms scales with the number of cell types and, presumably, the difference between cell types. (As the overlap of gained enhancers is even slightly lower than the overlap of all enhancers (1.8- to 3.6-fold), one could speculate that gained enhancers might have 'more unusual' sequence properties that are less likely to be shared by different cell types.)

Nature Genetics: doi:10.1038/ng.3009

Supplementary Table 1 | Number of reads and peak calls for all STARR-seq screens

Read pairs are the total number of paired reads mapped to the respective genome assembly. Unique fragments are read-pairs that were unique with regard to chromosome, start, end, and strand information and passed the non-heuristic redundancy filter (see methods). Lifted unique fragments are the number of mapped *D.xxx* fragments that were successfully lifted to dm3 coordinates. Peaks (called in *D.xxx*) are the number of enhancer peak calls within each respective species based on fragments mapped within the respective genomes. Peaks (dm3 lifted) are the number of peaks called in *D.xxx*, which could be lifted to dm3. Peaks Peaks (called in *D.mel*) are the number of peaks for each species based on peak calling with dm3 lifted fragments of each species. We restricted peaks to those in the euchromatic chromosomes (i.e. excluding heterochromatic chromosomes (Het) from the analysis).

Nature Genetics: doi:10.1038/ng.3009

Library	Read pairs	Unique fragments	Lifted unique fragments	Percentage liftable fragments	Peaks (called in Dxxx)	Peaks (dm3 lifted)	Peaks (called in Dmel)	Peaks (called in Dmel; no Het)
Input_Dmel	17,756,871	11,348,680	11,348,680	100.00	NA	NA	NA	NA
Input_Dyak	29,963,256	19,897,933	18,495,928	92.95	NA	NA	NA	NA
Input_Dana	45,839,408	31,748,014	24,727,167	77.89	NA	NA	NA	NA
Input_Dpse	33,532,553	27,456,215	23,023,734	83.86	NA	NA	NA	NA
Input_Dwil	28,272,857	22,520,804	16,666,846	74.01	NA	NA	NA	NA
cDNA_Dmel	38,041,487	688,076	688,076	100.00	2477	2477	2477	2325
cDNA_Dmel_rep1	11,787,165	338,786	338,786	100.00	2225	2225	2225	2139
cDNA_Dmel_rep2	26,254,322	380,024	380,024	100.00	2546	2546	2546	2361
cDNA_Dyak	52,617,454	1,205,818	1,112,892	92.29	2409	2314	2305	2293
cDNA_Dyak_rep1	43,416,059	562,597	515,032	91.55	1830	1751	1762	1752
cDNA_Dyak_rep2	9,201,395	679,209	632,725	93.16	2714	2614	2604	2591
cDNA_Dana	13,239,317	2,517,463	1,991,599	79.11	3426	2400	2142	2096
cDNA_Dana_rep1	7,716,324	1,355,477	1,080,379	79.70	3359	2369	2163	2107
cDNA_Dana_rep2	5,624,904	1,262,436	1,004,880	79.60	3085	2311	2146	2100
cDNA_Dpse	53,901,095	1,965,059	1,556,080	79.19	4471	3427	3493	3469
cDNA_Dpse_rep1	34,326,563	838,903	653,065	77.85	4487	3359	3391	3359
cDNA_Dpse_rep2	19,574,532	1,147,809	936,363	81.58	3946	3129	3246	3227
cDNA_Dwil	35,209,826	1,402,330	982,348	70.05	4573	3287	2922	2860
cDNA_Dwil_rep1	25,070,090	701,616	494,904	70.54	4271	3066	2775	2703
cDNA_Dwil_rep2	10,139,736	738,134	525,430	71.18	4199	3065	2792	2729
cDNA_Dmel_OSC	27,042,565	388,639	388,639	100.00	3915	3915	3915	3342
cDNA_Dmel_OSC_rep1	13,476,773	208,829	208,829	100.00	3625	3625	3625	3077
cDNA_Dmel_OSC_rep2	13,565,792	229,946	229,946	100.00	3879	3879	3879	3313
cDNA_Dyak_OSC	2,133,016	2,128,278	1,968,261	92.48	3497	3285	3280	3233
cDNA_Dyak_OSC_rep1	3,452,834	1,086,878	1,006,512	92.61	3408	3200	3204	3144
cDNA_Dyak_OSC_rep2	3,092,101	1,264,766	1,171,683	92.64	3533	3317	3335	3279
cDNA_Dana_OSC	2,015,227	2,008,536	1,605,149	79.92	4570	3474	2995	2859
cDNA_Dana_OSC_rep1	3,389,077	1,100,598	890,243	80.89	4741	3611	3308	3082
cDNA_Dana_OSC_rep2	2,950,703	1,109,725	900,845	81.18	4191	3219	2971	2777

Supplementary Table 2 | TF motif conservation in functionally conserved and *D. melanogaster*-specific S2 cell enhancers.

TF expression levels in S2 cells (RPKM values) and the TF motifs' preferential conservation in functionally conserved and *D.mel*-specific S2 enhancers. Table columns are: TF name, RPKM in S2 cells, TF motif (as in ref. 5), preferential conservation (fold increase), preferential conservation (binomial p-value).

→ This table is available for download at http://www.nature.com/ng and http://stark.imp.ac.at/data/arnold_gerlach_nature_genetics_2014

Supplementary Table 3 | RNA-seq in follicle cells – analysis statistics.

Gene expression data of follicle cells were obtained by RNA-seq from *D.mel* and *D.yak* adult females[6]. Number of reads, uniquely mapped reads, and uniquely mapped and lifted reads are shown.

Species	Reads	Uniquely mapped reads	Uniquely mapped reads (lifted)
D.mel	33,510,652	17,624,648	17,624,648
D.yak	24,308,648	17,370,054	14,017,986

Supplementary Table 4 | RNA-seq in follicle cells – gene expression (RPKM) values.

Gene expression levels in *D.mel* and *D.yak* follicle cell enriched samples in table format with the columns FlyBase gene ID, gene name, CG ID, RPKM *D.mel*, RPKM *D.yak*.

→ This table is available for download at http://www.nature.com/ng and http://stark.imp.ac.at/data/arnold_gerlach_nature_genetics_2014

Nature Genetics: doi:10.1038/ng.3009

Supplementary Table 5 | Oligonucleotide (primer) sequences.

Primer name	Sequence
STARR-seq RT	CTCATCAATGTATCTTATCATGTCTG
dyak_m1_fw	GCTGGCAATTGTTTTAATCGTTACAACGGCAAG
dyak_m1_rv	AAAGCCAAAGCTCCGTAATGATTATTCAGCGCTTCCTTTCGCTCGCCATC
dyak_y2_fw	GAAAGGAAGCGCTGAATAATCATTACGGAGCTTTGGCTTTGGCTTTGCCTATC
dyak_y2_rv	AAGGTTCCCTTTTGCCCAGCTTGGACGCAGTTC
dyak_y1_fw	GCGGGCAATTGTTTTAATCGCTACAACAGC
dyak_y1_rv	ACAGCCACCGCTTCGTAATGATTATTCAGCGCTTCCTTTCGCTCGATCCGCTG
dyak_m2_fw	GAAAGGAAGCGCTGAATAATCATTACGAAGCGGTGGCTGTCATATCGATCG
dyak_m2_rv	TCCCATTGCCATTCTTACCCACATTCGCATTGAC
dana_m1_fw	ACAAAAGTCTGCTGTTCGAAGGAACTTTAATCATAG
dana_m1_rv	TCCGAGGGGCTGTTAAATATCAGCATCTGTTGACCAGATGTAGTTTGTACAC
dana_a2_fw	CATCTGGTCAACAGATGCTGATATTTAACAGCCCCTCGGACGAGTGTGTGTG
dana_a2_rv	ACCTGACCTCTGCCATTGAAGGACCTCAAATC
dana_a1_fw	GCATAGTTTCCGTTGCACAGAGACCCTGATAAAG
dana_a1_rv	GCCTTATCAAGCCGCTTTATCACAGTCCCCGGAGCATTAACTTGTCCGTTTA
dana_m2_fw	TTAATGCTCCGGGGACTGTGATAAAGCGGCTTGATAAGGCCATCTCGGAAATC
dana_m2_rv	ACCGACCCTCCCGCCCAACCCCTTTTCACTTG
dpse_m1_fw	GAGATTTGCCTTCCGCAGCAAGCTGCCG
dpse_m1_rv	CGCCGGGGCGGGAGGATGACTCATCCGCTCAAGAAGCCGGGATTGATGCGTA
dpse_p2_fw	CCGGCTTCTTGAGCGGATGAGTCATCCTCCCGCCCCGGCGAGAACAGTCTCT
dpse_p2_rv	GAAGGATATCTTTTGATATGCTGATAAGGAGCGCC
dpse_p1_fw	ATGATATGCCTTGCTGCAGCGTCACCTGCTGC
dpse_p1_rv	CAGGATCGGTCTGACATGACTCATGCCGCCATTGATAAAACGGGATCACGGAG
dpse_m2_fw	TTTTATCAATGGCGGCATGAGTCATGTCAGACCGATCCTGAGAGTTCCGAGC
dpse_m2_rv	GAAGGATTTCTTTTGAGATACCCAAAAGGAGTGGC

Supplementary Data Set 1 | STARR-seq peak calls.

STARR-seq peak calls for each species, replicate, and cell type (ZIP file).

→ This table is available for download at http://www.nature.com/ng and http://stark.imp.ac.at/data/arnold_gerlach_nature_genetics_2014

Nature Genetics: doi:10.1038/ng.3009

Supplementary References

1. Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339,** 1074–1077 (2013).
2. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr Biol* **20,** 1562–1567 (2010).
3. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466,** 490–493 (2010).
4. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science* **321,** 1314 (2008).
5. 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 Research* **22,** 2018–2030 (2012).
6. Matts, J. A., Sytnikova, Y., Chirn, G.-W., Igloi, G. L. & Lau, N. C. Small RNA library construction from minute biological samples. *Methods Mol Biol* **1093,** 123–136 (2014).

Nature Genetics: doi:10.1038/ng.3009