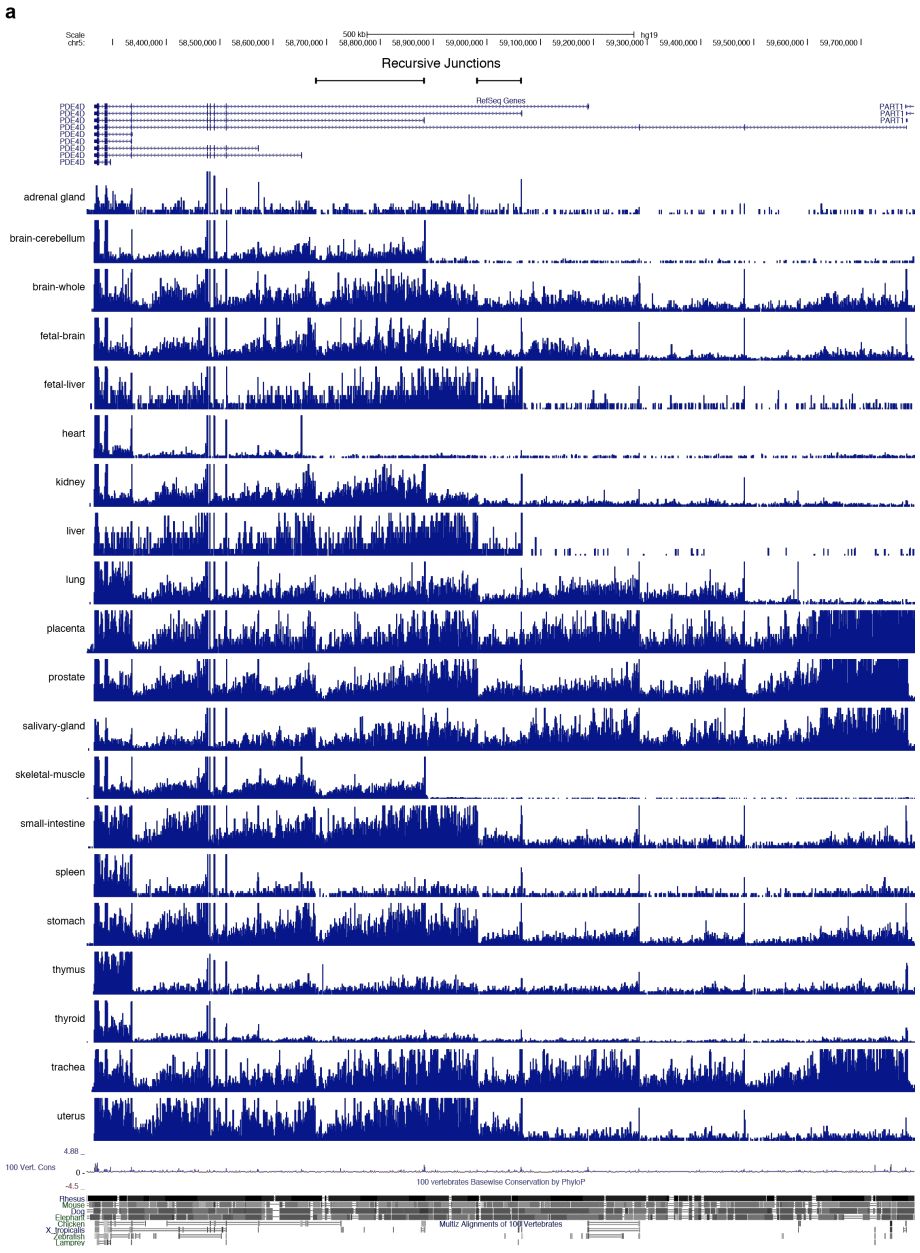
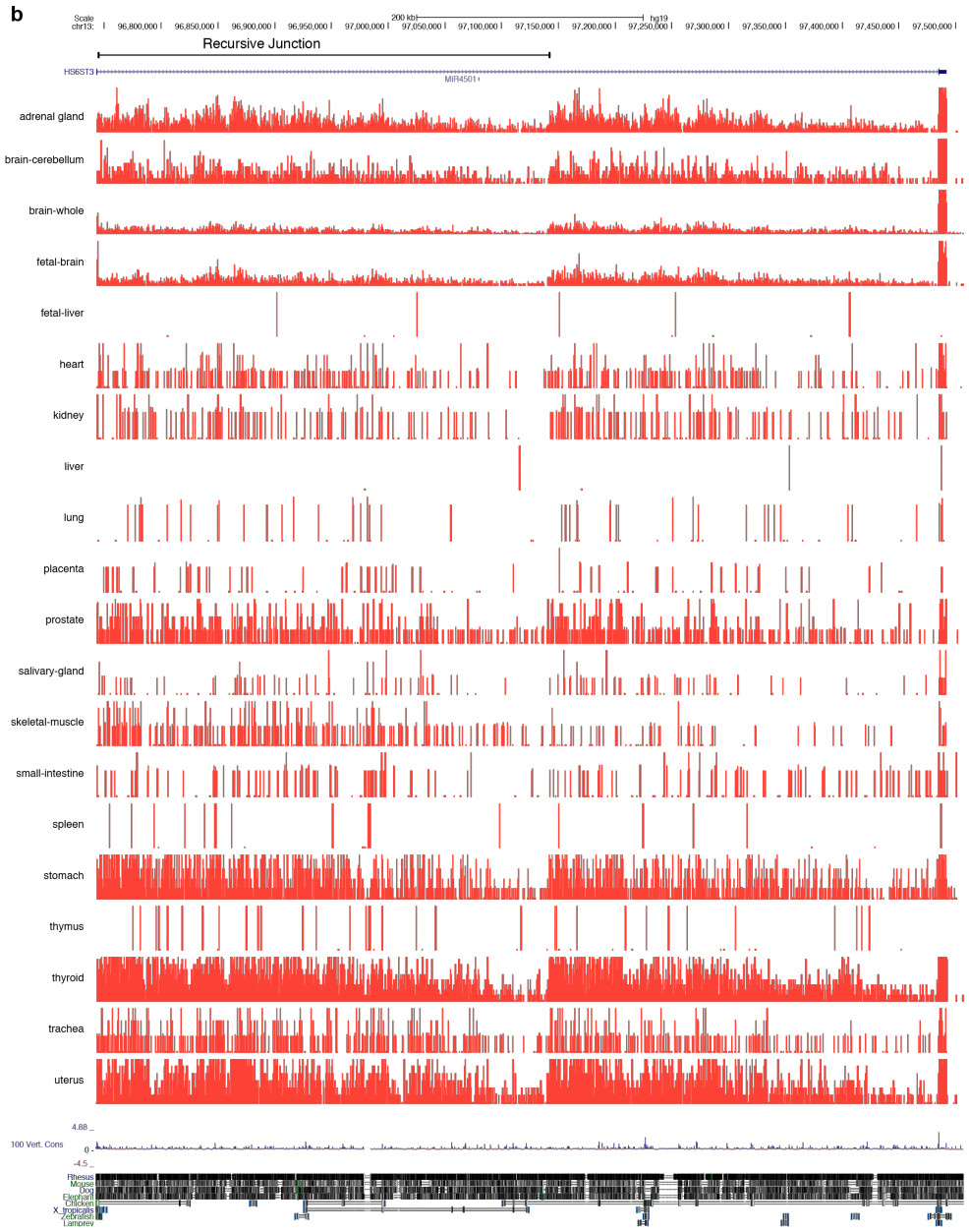
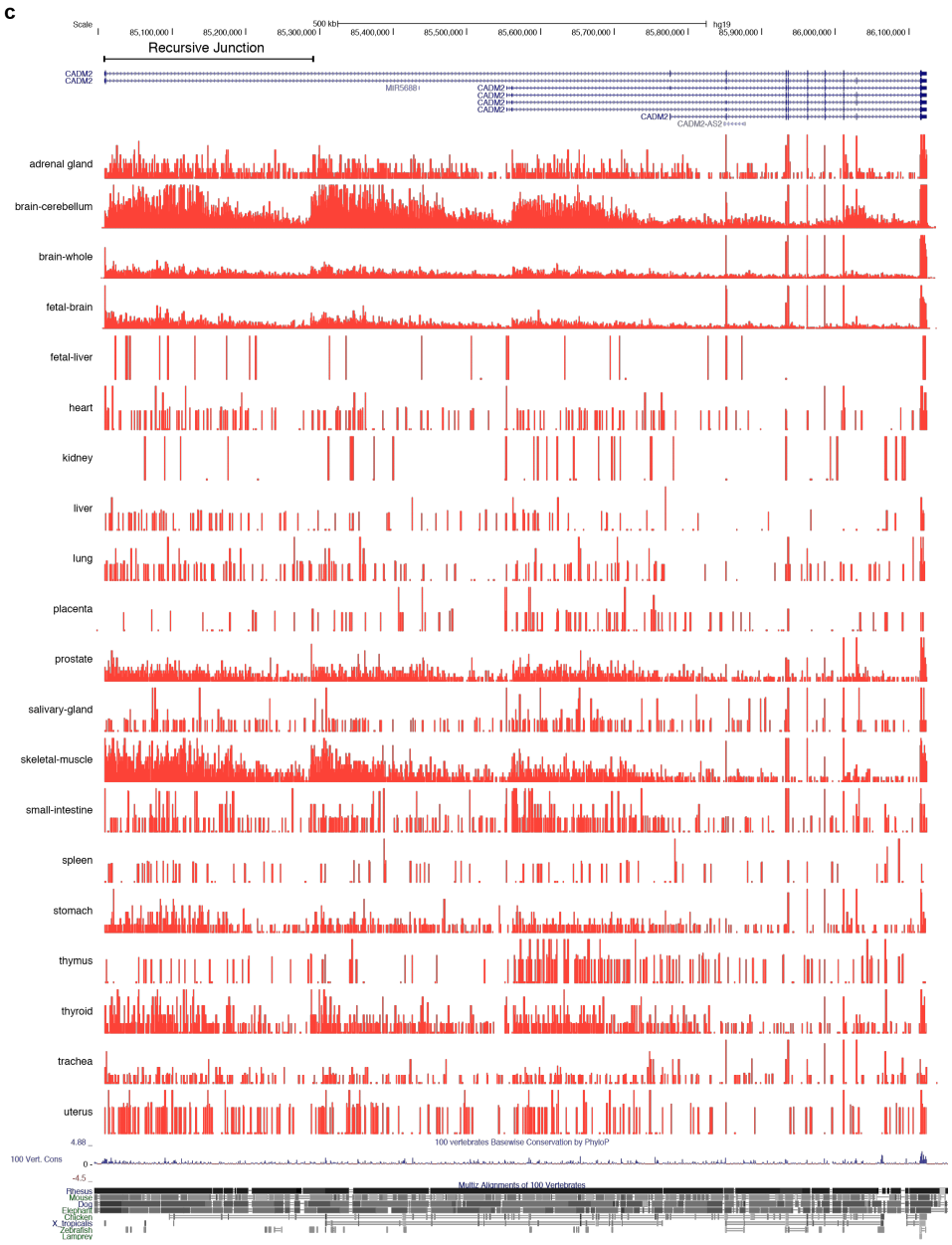
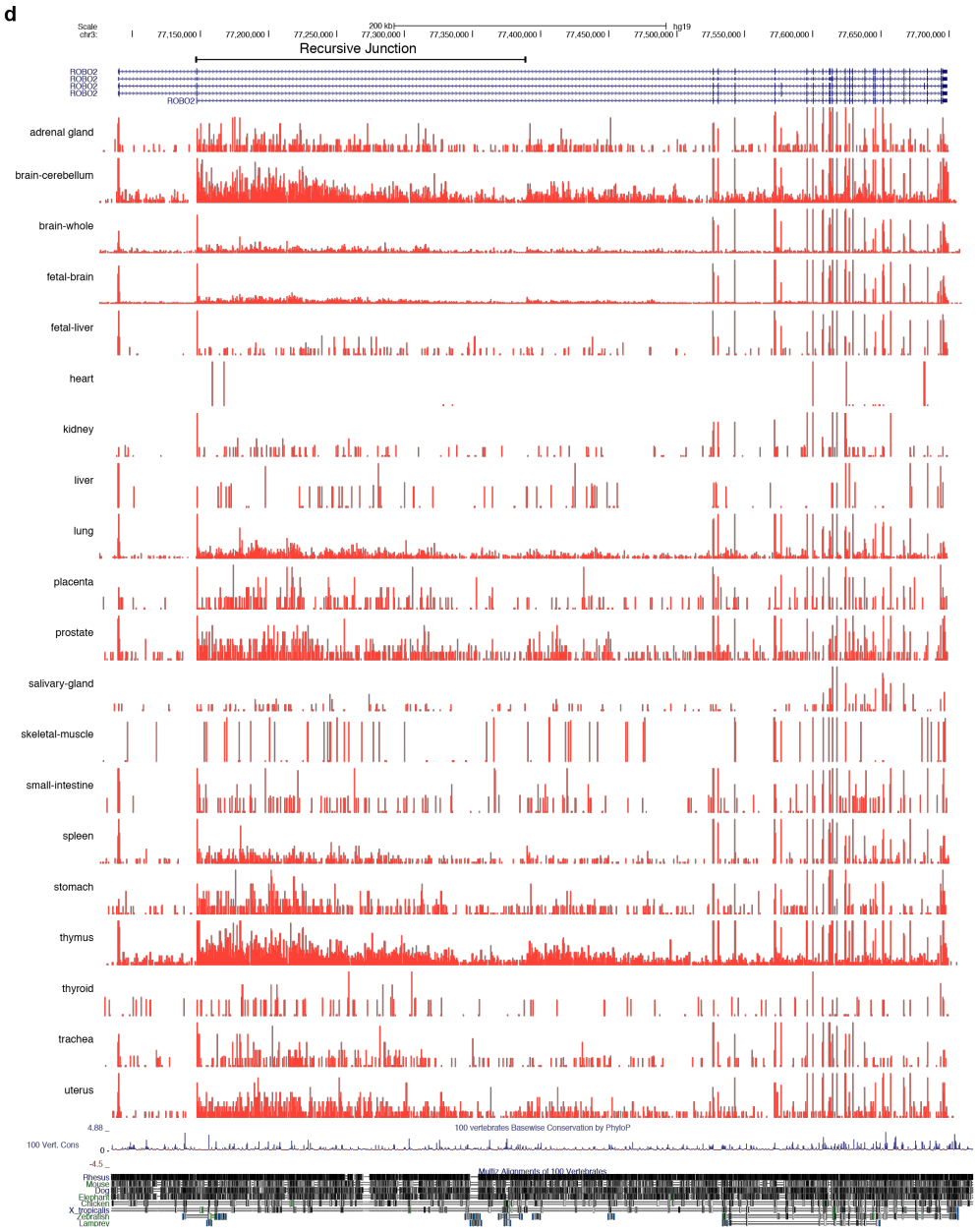
a


Scale
chr5:
58,400,000
58,500,000
58,600,000
58,700,000
58,800,000
58,900,000
59,000,000
59,100,000
59,200,000
59,300,000
59,400,000
59,500,000
59,600,000
59,700,000
500 kb
hg19
Recursive Junctions
RefSeq Genes
PDE4D
PART1
adrenal gland
brain-cerebellum
brain-whole
fetal-brain
fetal-liver
heart
kidney
liver
lung
placenta
prostate
salivary-gland
skeletal-muscle
small-intestine
spleen
stomach
thymus
thyroid
trachea
uterus
4.88
100 Vert. Cons
0
-4.5
100 vertebrates Basewise Conservation by PhyloP
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Multiz Alignments of 100 Vertebrates


b
Scale
chr13:
200 kb
hg19
96,800,000
96,850,000
96,900,000
96,950,000
97,000,000
97,050,000
97,100,000
97,150,000
97,200,000
97,250,000
97,300,000
97,350,000
97,400,000
97,450,000
97,500,000
Recursive Junction
HS6ST3
MIR4501
adrenal gland
brain-cerebellum
brain-whole
fetal-brain
fetal-liver
heart
kidney
liver
lung
placenta
prostate
salivary-gland
skeletal-muscle
small-intestine
spleen
stomach
thymus
thyroid
trachea
uterus
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey


c
Scale
500 kb
hg19
85,100,000
85,200,000
85,300,000
85,400,000
85,500,000
85,600,000
85,700,000
85,800,000
85,900,000
86,000,000
86,100,000
Recursive Junction
CADM2
CADM2
MIR5688
CADM2
CADM2
CADM2
CADM2
CADM2
CADM2-AS2
adrenal gland
brain-cerebellum
brain-whole
fetal-brain
fetal-liver
heart
kidney
liver
lung
placenta
prostate
salivary-gland
skeletal-muscle
small-intestine
spleen
stomach
thymus
thyroid
trachea
uterus
4.88
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
0
-4.5
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey



Supplementary Figure 1 | Identification of human recursive splicing events from total RNA-Seq data. Genome browser shots of total RNA-Seq data for 20 human tissues for the *PDE4D* (**a**), *HS6ST3* (**b**), *CADM2* (**c**), and *ROBO2* (**d**) genes. The splice junctions we identified that support recursive splicing are indicated near the top of each panel.

Supplementary Tables

Supplementary Table 1. Summary of *D. melanogaster* RNA Sequencing Samples Generated for these studies. For each library, the following information is provided: RNA ID, Biological ID, Biological Sample, Strain (or cell line if appropriate), Data Type (CELL LINE, DEVELOPMENT, ECDYSONE, POPULATION, TISSUE), Total Reads, Uniquely Aligned Reads (from TopHat alignments),% Uniquely Aligned, SRA Accession.

Supplementary Table 2. Summary information for 197 *D. melanogaster* ratchet points. For each ratchet point identified the following information is provided: Ratchet Point ID (chromosome_strand_location), Gene Name, Alias, TopHat aligned reads (total across all samples), de novo aligned reads (total across all samples), de novo mismatches (average frequency of mismatches for all reads across all samples from the de novo alignments), de novo entropy (entropy score for the ratchet points across all samples from the de novo alignments), intron chr (chromosome of the recursive intron), intron start (start position of the recursive intron), intron stop (stop position of the recursive intron), intron strand (strand of the recursive intron), intron chr (chromosome of the recursive intron if more than one exon can be recursively spliced), intron start (start position of the recursive intron if more than one exon can be recursively spliced), intron stop (stop position of the recursive intron if more than one exon can be recursively spliced), intron strand (strand of the recursive intron if more than one exon can be recursively spliced).

Supplementary Table 3. Comparison of ratchet points identified in this study and in Burnette *et al*. All of the predicted ratchet points listed in supplementary Table S1 of Burnette *et al*.[2] were compared to our experimentally identified ratchet points. To do this, we used BLAT to identify the location of each ratchet point from Burnette *et al*. and tabulated whether or not the ratchet point was identified in our RNA-Seq data. If the ratchet point was not identified, we listed the reason. For each ratchet point the following information is provided: Gene Name, Intron Number, *D. melanogaster* RP Sequence, RP Score (as calculated by Burnette *et al*.), Identified in this study (1 = yes, 0 = no), Reason (exon = ratchet point is an exon, Lack of sawtooth = no compelling sawtooth pattern of RNA-Seq read density observed, missed = not identified in our computational pipeline, no matches = the ratchet point sequence listed from Burnette *et al*. did not match the *D. melanogaster* genome by BLAT).

Supplementary Table 4. Summary of recursive intron lariats identified from total RNA-Seq data. For each read aligned to a recursive lariat intron junction, the following information is provided: read ID, strand (of read alignment to the lariat junction), lariat junction ID (build_chr_intronstart_intronstop_strand_RPchr_strand_ratchetpoint:gene:recursive segment_location of 5' end of the 3' segment of the intron included_number of nucleotides included from the 5' end of the intron: e.g., dm3_chr3L_10512711_10527930_+_chr3L_+_10527931:CG32062:segment2_-127_94), alignment start (within the junction), window start (location of 5' end of the 3' segment of the intron included), branchpoint (location of 3' end of the 3' segment of the intron included), Sequence (of aligned read), mismatches, gene, segment (segment of the recursive intron), location (first segment, internal segment, last segment).

Supplementary Table 5. Details of ratchet points experimentally validated in other *Drosophila* species. For each ratchet point in *D. melanogaster, the number of reads mapping to the orthologous ratchet point in D. simulans, D. sechellia, D. yakuba, D. pseudoobscura*, or *D. virilis* is indicated. In addition, the total number of reads identified from the other species is indicated in the last column.

Supplementary Table 6. Summary information for 5 human ratchet points. For each ratchet point identified the following information is provided: x_RP(chr_str_loc), Gene_Name, #samples the ratchet point was observed in, the number of recursive junction reads observed, the number of reads observed in each tissue (adreanal-gland, brain-cerebellum, brain-whole, fetal-brain, fetal-liver, heart, kidney, liver, lung, placenta, prostate, salivary-gland, skeletal-muscle, small-intestine, spleen, stomach, thymus, thyroid, trachea, and uterus), the coordinate of the ratchet point junction (m_chromo_strand_start1_end1_start2_end2), intron coordinates, intron length, gene name, n_align_strictn_align, avg mismatch per read, and offset entropy.

Supplementary Table 7. GO analysis of recursively spliced *Drosophila* genes. Results of GO analysis of recursively spliced genes using Funcassociate 2.0[ref. 17].