Supplementary Table S1. Summary of *Dscam* isoform profile in the different samples

Sample	No. Quadruple Reads (million)	No. Detected Isoforms	% Reads from the most abundant 10 (100) isoforms
S2 cell	12.67	7,317	4.8 (25.6)
Embryo	12.89	16,862	1.1 (7.0)
L1	5.71	14,145	1.8 (10.3)
L2	14.35	15,118	2.0 (12.2)
L3	13.57	13,216	1.7 (10.8)
Pupa	11.90	16,876	0.67 (5.3)
Adult brain	15.22	16,886	1.1 (7.0)
Total	86.32	18,496	

Supplementary Table S2. *Dscam* isoform abundance (RPM: No. sequencing reads per million total reads) in different samples (see Supp_table2.xls)

Supplementary Table S3. The effect of different genetic deletions on the effective size of *Dscam* repertoire (excluding peusdo-exon 6.11)

Note: 1. delta 4.1-4.3, delta 4.4-4.12, aggregate 9, aggregate 6,9 strains were described in Hattori *et.al.* 2009; delta 4.2-4.6 and delta 4.4-4.8 strains were described in Chen *et.al.* 2006; 2. we assumed here that the biased usage of the remaining exons are unchanged after deletion. Since this assumption might not hold, as demonstrated by Chen et.al, the number should not be interpreted as the approximation of the real situation. Instead, we used the table to demonstrate the different effects of the deletions of the same number, but of different exons; and the unequal effect of the same deletion in different samples.

Strain	Uniform	Embryo	L1	L2	L3	Pupa	Brain	S2
WT	18,612	5,489	3,119	2,301	2,586	6,503	5,442	769
delta 4.1-4.3	13,959	5,303	3,206	2,079	1,966	5,235	5,376	726
delta 4.4-4.12	4,653	1,047	697	566	666	1,468	929	185
aggregate 9	564	429	409	385	422	463	432	300
aggregate 6,9	12	11	11	10	11	11	11	8
delta 4.2-4.6	10,857	4,165	2,462	2,122	2,006	4,191	3,337	411
delta 4.4-4.8	10,857	2,649	1,537	1,176	1,462	3,805	2,802	382

Supplementary Figure S1. Independent splicing choice between the different variable exon clusters. (**A**). Observed isoform frequency was depicted in X axis. The expected frequency was calculated by multiplying the frequencies of their respective variable exon 4, 6 and 9 and depicted in Y axis. To determine whether the splicing between the three clusters was independently controlled, the two frequencies were compared. The solid red line is the diagonal and the parallel dash red lines represent ranges of one-fold change. (**B**), (**C**), (**D**): In a similar way, we determined whether the splicing choice was independent between exon 4 and 6 (**B**), exon 6 and 9 (**C**), exon 4 and 9 (**D**), respectively. (**E**), (**F**): The whole dataset from each sample could be *in silico* decomposed into two (**E**) or three (**F**) groups. The splicing choice within each group was largely independent between exon 4 and 9. X axis depicted the observed isoform frequency from the whole dataset, whereas in Y axis, the expected frequency was the sum of the expected frequencies of the two (**E**) or three (**F**) groups.

Supplementary Figure S2. (**A**). The variable exon 4s and 9s were clustered based on their expression patterns, the exon 9s could be clearly divided into two groups, one containing only five exons and the other consisting of the remaining 27. (**B**). Given the differential usage of variable exon 4s within the two groups, we could *in silico* decompose the whole data into two sets with different usage of exon 4s and 9s.

Supplementary Figure S3. (**A**). The fraction of neuron pairs with the same *Dscam* identity (P_{kk}, Y axis; Methods) when each neuron expresses different number of *Dscam* mRNA molecules (X axis), if up to 0% (left), 10% (middle) or 20% (right) of isoform are allowed to share between any pair of neurons, calculated based on the distribution of *Dscam* isoform abundances in adult brain, S2 cells, a uniform distribution or in hypothetical samples with different effective size of *Dscam* repertoire (dash lines) (Methods). (**B**). The number of neurons that obtain unique

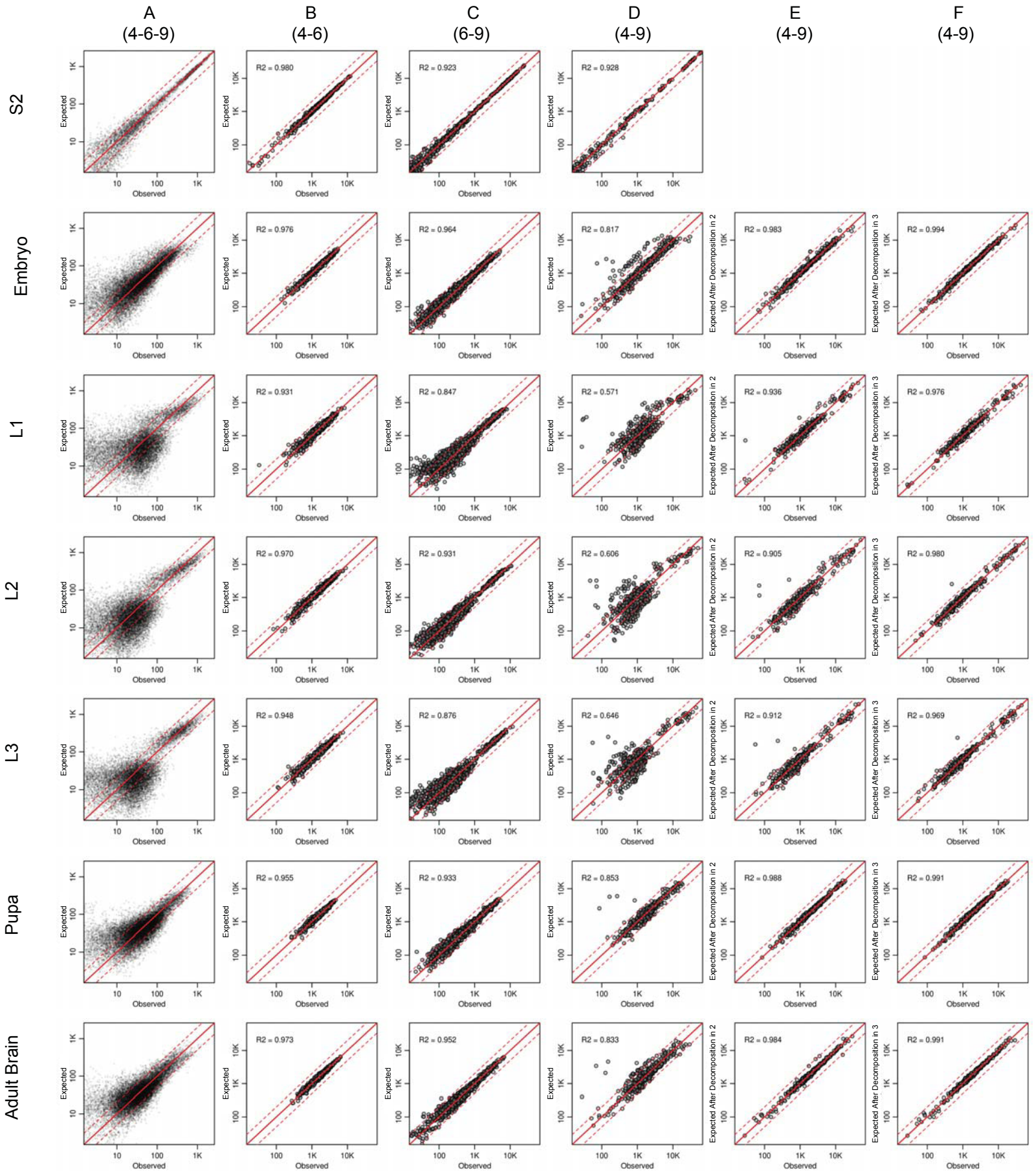
identities at more than 95% likelihood (Y axis) when each neuron expresses different numbers of *Dscam* mRNA molecules (X axis), if allowing up to 0% (left), 10% (middle) or 20% (right) of isoforms shared between any pair of neurons, calculated based on the distribution of *Dscam* isoform abundances in adult brain, S2 cells, a hypothetical uniform distribution or in hypothetical samples with different effective size of *Dscam* repertoire (dash lines) (Methods).

Reference:

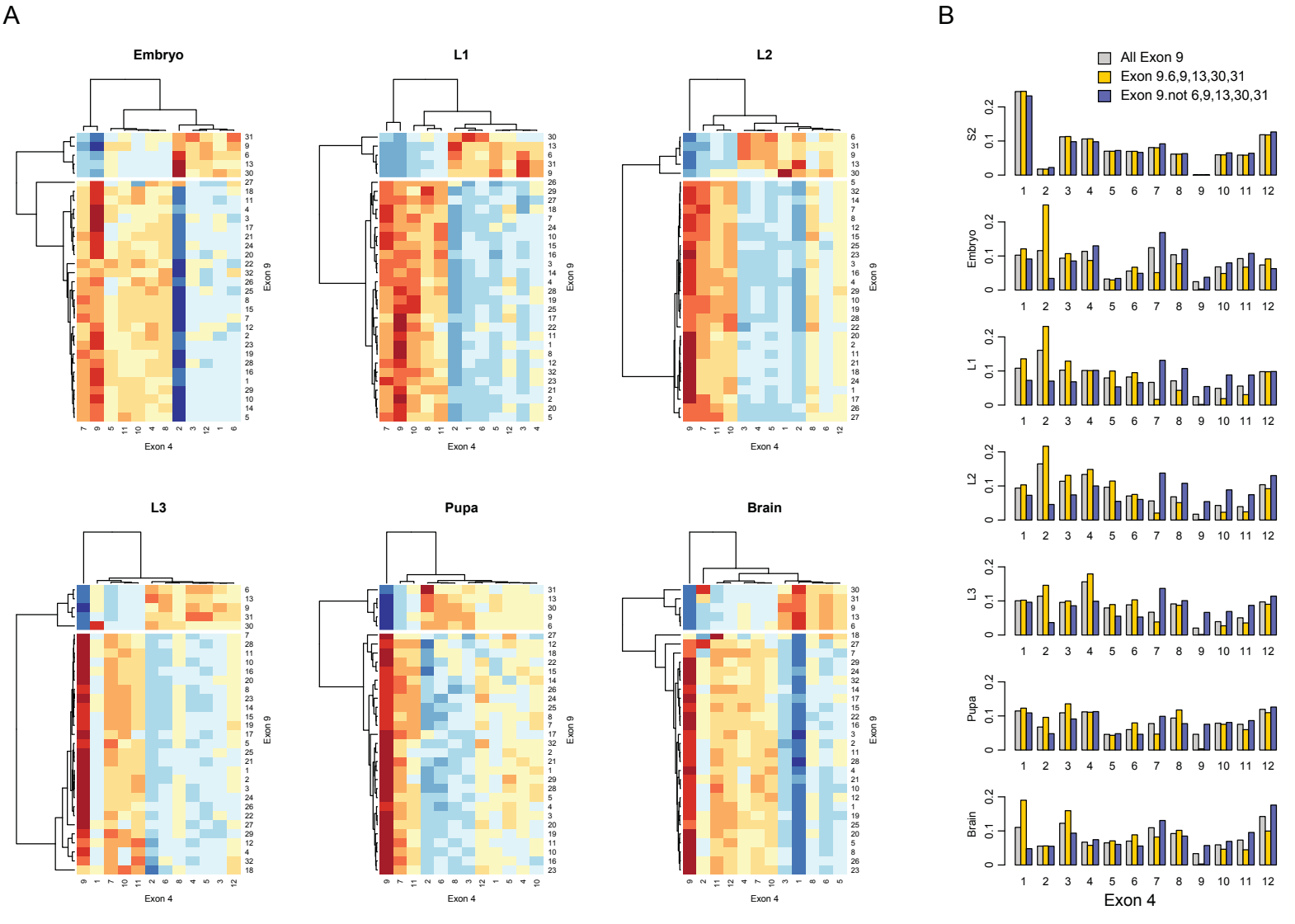
Chen, B.E., Kondo, M., Garnier, A., Watson, F.L., Püettmann-Holgado, R., Lamar, D.R., and Schmucker, D. (2006). The molecular diversity of Dscam is functionally required for neuronal wiring specificity in Drosophila. *Cell* **125:** 607–620.

Hattori, D., Chen, Y., Matthews, B.J., Salwinski, L., Sabatti, C., Grueber, W.B., and Zipursky, S.L. (2009). Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature* **461:** 644–648.

Supplementary Figure S1



Supplementary Figure S2



Supplementary Figure S3