PLOS | GENETICS

RESEARCH ARTICLE

Estimating Information Processing in a Memory System: The Utility of Meta-analytic Methods for Genetics

Tugce Yildizoglu[1,2], Jan-Marek Weislogel[1,2], Farhan Mohammad[1,2], Edwin S.-Y. Chan[3,4], Pryseley N. Assam[3,4], Adam Claridge-Chang[1,2,5]*

1 Program in Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School, Singapore, **2** Institute for Molecular and Cell Biology, Singapore, **3** Singapore Clinical Research Institute, Singapore, **4** Centre for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore, **5** Department of Physiology, National University of Singapore, Singapore

* claridge-chang.adam@duke-nus.edu.sg

Citation: Yildizoglu T, Weislogel J-M, Mohammad F, Chan ES-Y, Assam PN, Claridge-Chang A (2015) Estimating Information Processing in a Memory System: The Utility of Meta-analytic Methods for Genetics. PLoS Genet 11(12): e1005718. doi:10.1371/journal.pgen.1005718

Editor: Malcolm R Macleod, University of Edinburgh, UNITED KINGDOM

Received: May 26, 2015

Accepted: November 10, 2015

Published: December 8, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: TY, JMW and ACC were supported by a Biomedical Research Council block grant to the Neuroscience Research Partnership and the Institute of Molecular and Cell Biology. ACC received additional support from Duke-NUS Graduate Medical School, a Nuffield Department of Medicine Fellowship, a Wellcome Trust block grant to the University of Oxford and A*STAR Joint Council Office grant 1131A008. TY was supported in part by a Singapore Pre-Graduate Award from the A*STAR

Abstract

Genetic studies in *Drosophila* reveal that olfactory memory relies on a brain structure called the mushroom body. The mainstream view is that each of the three lobes of the mushroom body play specialized roles in short-term aversive olfactory memory, but a number of studies have made divergent conclusions based on their varying experimental findings. Like many fields, neurogenetics uses null hypothesis significance testing for data analysis. Critics of significance testing claim that this method promotes discrepancies by using arbitrary thresholds (α) to apply reject/accept dichotomies to continuous data, which is not reflective of the biological reality of quantitative phenotypes. We explored using estimation statistics, an alternative data analysis framework, to examine published fly short-term memory data. Systematic review was used to identify behavioral experiments examining the physiological basis of olfactory memory and meta-analytic approaches were applied to assess the role of lobular specialization. Multivariate meta-regression models revealed that short-term memory lobular specialization is not supported by the data; it identified the cellular extent of a transgenic driver as the major predictor of its effect on short-term memory. These findings demonstrate that effect sizes, meta-analysis, meta-regression, hierarchical models and estimation methods in general can be successfully harnessed to identify knowledge gaps, synthesize divergent results, accommodate heterogeneous experimental design and quantify genetic mechanisms.

Author Summary

Genetic analysis of learning in the black-bellied vinegar fly has revealed that a brain structure called the mushroom body is important to insect memory. The mushroom body contains three lobes with strikingly different shapes. A series of studies have concluded that the lobes have markedly different relevance to memory. For short-term memory, some studies have concluded that only a single lobe–the gamma lobe–is required. However,

Graduate Academy. PNA and ESYC are supported by a National Medical Research Council block grant to the Singapore Clinical Research Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

others have concluded that at least one of the other lobes is also involved. These studies used a data analysis method called 'null hypothesis significance testing' that may overemphasize differences between data. We examined whether estimation statistics, an alternative data analysis framework, could be used to verify or refute the lobular specialization hypothesis. Estimation statistics review methods were used to analyze published data on this topic. The estimation models indicate no evidence for lobular specialization, but instead show that neurons in all lobes contribute to short-term memory. These results verify a model in which learning is processed in a distributed manner across the mushroom body. These findings also demonstrate that estimation methods can be successfully harnessed for the analysis of complex experimental research data.

Introduction

Olfactory memory in *Drosophila* is measured using the classical T-maze olfactory conditioning assay, where groups of flies are conditioned by pairing an odor with an electric shock and subsequently assessed for their ability to avoid the conditioned odor when given a choice of two different odors presented at the end of the maze arms. Thirty years of T-maze experiments have elucidated many of the genetic, molecular and neural mechanisms of olfactory learning [1–5,9]. A landmark study showed that restoring the adenylyl cyclase gene *rutabaga (rut)* to a brain structure called the mushroom body is sufficient for short-term olfactory memory [6], connecting memory formation to cyclic adenosine monophosphate-mediated plasticity [10]. Experiments using inhibition of synaptic transmission by temperature-sensitive *shibire (shi)* [11–13] showed that neurotransmission from the mushroom body is essential [12,14]. Targeted expression of genes in specific neuronal circuits is possible with the use of transgenic 'driver' lines [15]. Manipulations based on *rut* restoration and *shi* inactivation form the foundation of a large number of studies aiming to further define the role of the mushroom body in olfactory learning. The mushroom body itself exists as three anatomically distinct lobes, αβ, α′β′, and γ [16]; studies on middle- and long-term memory (MTM and LTM) have revealed distinct lobe requirements in the different memory phases [13,17–20]. However, the three lobes' specializations remain unclear when it comes to short-term memory (STM). While the mainstream view is that *rut* activity in the γ lobes is sufficient to rescue STM [8], some studies have alternately concluded that *rut* restoration can only partially rescue [7], or is merely of importance to STM [6]. There is similar controversy on the role of *rut* activity in the αβ lobes, with *rut* restoration said to have either no effect [8], or to partially rescue STM for certain odors [7].

Contradictory research results are commonplace as they stem from sampling error and methodological differences, both unavoidable sources of variability. One concern is the widespread acceptance of weak significance testing power [21]. However, critics of significance testing itself claim that this statistical framework itself accentuates differences. The various conceptual and practical limitations of significance tests [22] include the inherent volatility of p-values, even with moderate statistical power [23,24]. Significance testing may also exacerbate discordance by using an arbitrary threshold to elicit a binary outcome (reject/accept) from continuous data [25]. To illustrate, a pair of alpha 0.05 tests on two replicated experiments with identical effect sizes could produce p-values of 0.049 and 0.051: the significance test results are starkly discordant even though the biological outcome is the same [25]. The reject/accept dichotomy might also lead to the impression that a substantial (but non-statistically

significant) effect is irrelevant. Conversely, a highly powered sample size could give the impression that a minuscule (but statistically significant) effect is of great importance [23].

In medical research, the complementary methods of systematic review and meta-analysis are routinely used to synthesize evidence from multiple studies and to reconcile divergent findings [26]. Meta-analysis forms part of estimation statistics, an alternative analysis framework to significance testing. Such approaches are increasingly applied to preclinical research [27,28], but remain rarely used in basic research fields. Taking a mainstream sub-field of basic neuroscience as an example, a PubMed search in late 2015 with the phrase "meta-analysis AND (learning OR memory) AND mouse" identified fewer than ten studies in a field of >38,000 articles. We asked whether meta-analytic methods could be used to address the *Drosophila* mushroom body lobular specialization hypothesis. A particular strength of the olfactory T-maze is its use of hundreds or thousands of animals in a single experiment [29]. In addition, both the T-maze apparatus and the training regime are largely standardized between labs [29]. These advantages suggested that the published data would not be overwhelmed by weak statistical power or methodological heterogeneity, and thus suitable for meta-analysis.

In the present study, we aimed to evaluate the mainstream view that there is strong lobular specialization of STM function in the mushroom body, and to assess the extent to which the varying perspectives on this subject resulted from significance testing's dichotomization. We examined the proposals that restoration of *rut* function to the γ lobes alone is sufficient to rescue wild type STM and that only *shi* function in the γ lobes is necessary for STM. In both cases, meta-analysis of published studies spanning more than a decade found no evidence for strong lobular specialization. A subsequent analysis with multi-level meta-regression revealed that numbers of mushroom body cells explained nearly all transgenic effects. These results support the idea that associative olfactory information is initially processed in a distributed manner across the mushroom body. These results also confirm claims made by statistical texts that systematic review, meta-analysis and related estimation methods can be applied to resolve currently conflicting data and give new quantitative perspectives to basic research fields like experimental genetics.

Results

Systematic literature review of *rutabaga* and *shibire* interventions in short-term aversive olfactory memory

The review yielded ten studies that fulfilled the criteria (Fig 1A). Seven studies contained 81 experiments related to *rutabaga* restoration [6–8,14,30–32], with a total of 748 experimental iterations and 745 control iterations (see Table 1). Each iteration is the mean of two half-PI scores, which typically each use 50–100 flies, thus representing an estimated total of 150,000–300,000 assayed flies. Table 1 also lists the 5 studies that contained 37 experiments related to *shibire*-mediated inactivation [7,12,13,17,30], 263 experimental iterations and 265 control iterations, giving a total of 50,000–100,000 flies.

Experimental variability

Despite standardization of aspects of the T-maze, some methodological variation between studies was observed, including different control genotypes, varying odor pairs, temperatures, shock voltages, humidity and post-training delay times prior to testing (Table 1). These differences, along with other uncontrolled variables common to behavioral experiments, would explain the variability seen in data from control experiments (Fig 1B). We found considerable heterogeneity in several of the meta-analyses. In the six *rut* analyses, overall heterogeneity was

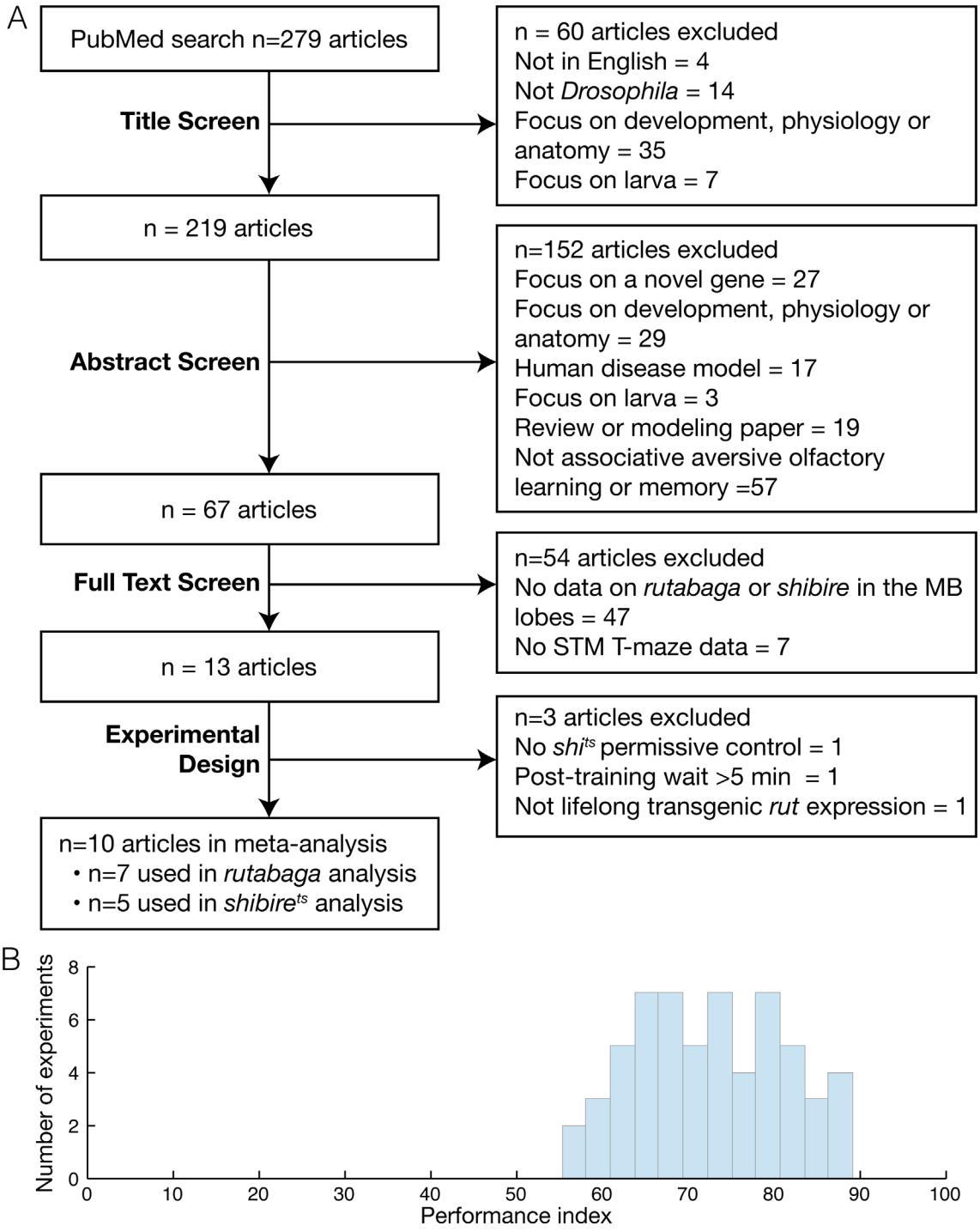

Fig 1. Review overview. A. Flow chart of systematic literature review procedure. The literature was reviewed in a five stage process, starting with a PubMed search that yielded 279 articles, followed by four screens of increasing detail, reviewing the article title, abstract full text and experimental design. A total of ten articles, two of which included relevant data for both rutabaga and shibirets experiments, were used in the meta-analyses. **B.** Histogram of performance indices for all control experiments identified by the review.

doi:10.1371/journal.pgen.1005718.g001

low in three ($I^2 < 50\%$), and high in three ($I^2 > 75\%$); subgroup heterogeneity (i.e. variance due to genotype differences) was low in four, and high in two. In the *shi* analyses, overall

Table 1. Characteristics of included experiments. All experiments are listed and identified by their study, figure panel and genotype/s. We name the most precise genotype possible based on the information given in the original article. Odor pair, range experimental temperature or temperature range, the nature of the conditioning shock and the relative humidity (RH) are also listed. The time delay between training and testing is listed in minutes; those labelled '0*' were reported as following training 'immediately.' Shock is listed in volts; current type is omitted if not reported in the original study. Cells containing a dash indicate that the information was not found in the original article.

Study	Fig.	Genotype, Experimental	Genotype, Control	N (E)	N (C)	Odor Pair	Experimental Temp. °C	Shock (V)	Time (min)	RH (%)
rutabaga rescue in the αβ (alphabeta) lobes										
Zars 2000	1	*rut^{2080}/Y; 17d; UAS-rut*	*17d/+*	6	6	MCH-BEN	25	120 AC	2	-
Zars 2000	1	*rut^{2080}/Y; 189Y; UAS-rut*	*189Y/+*	6	6	MCH-BEN	25	120 AC	2	-
McGuire 2003	S4	*rut^{2080}; c739; UAS-rut*	*c739/+*	7	7	OCT-BEN	25	90	3	-
Akalal 2006	3A	*rut^{2080}; c739; UAS-rut*	*c739/+*	12	12	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	3B	*rut^{2080}; c739; UAS-rut*	*c739/+*	10	10	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	3C	*rut^{2080}; c739; UAS-rut*	*c739/+*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	3D	*rut^{2080}; 17d; UAS-rut*	*17d/+*	24	24	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	3E	*rut^{2080}; 17d; UAS-rut*	*17d/+*	12	12	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	3F	*rut^{2080}; 17d; UAS-rut*	*17d/+*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Blum 2009	4B	*rut^{2080}/Y; c739; UAS-rut*	*+/rut^{2080};+;UAS-rut*	12	12	MCH-OCT	22	60	2	50
Blum 2009	6A	*rut^{2080}/Y; c739; UAS-rut*	*+/rut^{2080};+;UAS-rut*	6	6	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut^{1}; 17d; UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
rutabaga rescue in the α'β' (prime) lobes										
Blum 2009	4A	*rut^{2080}/Y; c305a; UAS-rut*	*+/rut^{2080};+;UAS-rut*	8	8	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut^{1}; c305a; UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
Scheunemann 2012	5A	*rut^{1}; c320; UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
rutabaga rescue in the γ (gamma) lobes										
Zars 2000	1	*rut^{2080}/Y; +; H24/UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
Zars 2000	1	*rut^{2080}/Y; 201Y; UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
McGuire 2003	S4	*rut^{2080}/Y; +; H24/UAS-rut*	*+; H24*	7	7	OCT-BEN	25	90	3	-
Akalal 2006	2A	*rut^{2080}/Y; +; H24/UAS-rut*	*+; H24*	18	18	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	2B	*rut^{2080}/Y; +; H24/UAS-rut*	*+; H24*	18	18	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	2C	*rut^{2080}/Y; +; H24/UAS-rut*	*+; H24*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	2D	*rut^{2080}; NP1131; UAS-rut*	*+; NP1131*	17	17	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	2E	*rut^{2080}; NP1131; UAS-rut*	*+; NP1131*	17	17	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	2F	*rut^{2080}; NP1131; UAS-rut*	*+; NP1131*	18	18	OCT-BEN	21–25	90 DC	3	60–68
Blum 2009	4A	*rut^{2080}/Y;* 201Y; *UAS-rut*	*+/rut^{2080}; +; UAS-rut*	8	8	MCH-OCT	22	60	2	50
Blum 2009	6A	*rut^{2080}/Y;* 201Y; *UAS-rut*	*+/rut^{2080}; +; UAS-rut*	6	6	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut^{1}; NP1131; UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
rutabaga rescue in all lobes										
Zars 2000	1	*rut^{2080}/Y; 30Y/UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
Zars 2000	1	*rut^{2080}/Y; 238Y; UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
Blum 2009	3A	*rut^{2080}; +; UAS-rut; OK107*	*+/rut^{2080};+;UAS-rut*	7	6	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut^{1}; +; UAS-rut; OK107*	wild type	8	8	EA-IA	24	120 AC	3	70

(*Continued*)

Table 1. (*Continued*)

Study	Fig.	Genotype, Experimental	Genotype, Control	N (E)	N (C)	Odor Pair	Experimental Temp. °C	Shock (V)	Time (min)	RH (%)
rutabaga rescue in the αβγ lobes										
Zars 2000	1	*rut^{2080}/Y; +; MB247/UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
Zars 2000	1	*rut^{2080}/Y; c772; UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
McGuire 2003	2A	*rut^{2080}; +; MB247/UAS-rut*	Canton-S	5	5	OCT-BEN	25	90	3	-
McGuire 2003	2A	*rut^{2080}; c772; UAS-rut*	Canton-S	5	5	OCT-BEN	25	90	3	-
McGuire 2003	S4	*rut^{2080}; c739; H24/UAS-rut*	*+; c739; H24*	7	7	OCT-BEN	25	90	3	-
Schwaerzel 2003	1C	*rut^{2080}; UAS-rut; MB247*	Canton-S	6	6	EA-IA	26	130	3	80
Akalal 2006	5A	*rut^{2080}; c739; H24/UAS-rut*	*+; c739; H24*	6	6	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	5B	*rut^{2080}; c739; H24/UAS-rut*	*+; c739; H24*	6	6	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	5C	*rut^{2080}; c739; H24/UAS-rut*	*+; c739; H24*	6	6	OCT-BEN	21–25	90 DC	3	60–68
Thum 2007	1D	*rut^{2080}; +; MB247/UAS-rut*	*MB247/+*	8	8	MCH-OCT	25	90 DC	0*	-
Blum 2009	3A	*rut^{2080}/Y; +; MB247/UAS-rut*	*+/rut^{2080}; +; UAS-rut*	7	6	MCH-OCT	22	60	2	50
Blum 2009	3A	*rut^{2080}/Y; c309; UAS-rut*	*+/rut^{2080}; +; UAS-rut*	7	6	MCH-OCT	22	60	2	50
Blum 2009	6A	*rut^{2080}/Y; c739/201Y; UAS-rut*	*+/rut^{2080}; +; UAS-rut*	6	6	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut^{1}; MB247/UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
rutabaga^{2080}										
Zars 2000	1	*rut^{2080}*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
McGuire 2003	2A	*rut^{2080}*	Canton-S	5	5	OCT-BEN	25	90	3	-
Schwaerzel 2003	1C	*rut^{2080}*	Canton-S	6	6	EA-IA	26	130	3	80
Blum 2009	1A	*rut^{2080}*	*rut^{2080}/ +*	6	6	MCH-OCT	22	60	2	50
rutabaga^{2080} with *Driver/s*										
Akalal 2006	3C	*rut^{2080}; c739*	*c739/+*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	2C	*rut^{2080}; H24*	*+; H24*	18	18	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	3F	*rut^{2080}; 17d*	*17d/+*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	5C	*rut^{2080}; c739; H24*	*+; c739; H24*	6	6	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	2A	*rut^{2080}; H24*	*+; H24*	12	12	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	2E	*rut^{2080}; NP1131*	*+; NP1131*	18	18	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	3A	*rut^{2080}; c739*	*c739/+*	12	12	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	5A	*rut^{2080}; c739; H24*	*+; c739; H24*	6	6	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	3B	*rut^{2080}; c739*	*c739/+*	10	10	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	2D	*rut^{2080}; NP1131*	*+; NP1131*	17	17	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	3E	*rut^{2080}; 17d*	*17d/+*	12	12	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	3D	*rut^{2080}; 17d*	*17d/+*	24	24	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	2F	*rut^{2080}; NP1131*	*+; NP1131*	18	18	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	5B	*rut^{2080}; c739; H24*	*+; c739; H24*	6	6	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	2B	*rut^{2080}; H24*	*+; H24*	18	18	MCH-OCT	21–25	90 DC	3	60–68
McGuire 2003	2A	*rut^{2080}; c772*	Canton-S	5	5	OCT-BEN	25	90	3	-
McGuire 2003	2A	*rut^{2080}; MB247*	Canton-S	5	5	OCT-BEN	25	90	3	-
McGuire 2003	S4	*rut^{2080}; c379*	*c739/+*	7	7	OCT-BEN	25	90	3	-
McGuire 2003	S4	*rut^{2080}; H24*	*+; H24*	7	7	OCT-BEN	25	90	3	-
McGuire 2003	S4	*rut^{2080}; c379; H24*	*+; c739; H24*	7	7	OCT-BEN	25	90	3	-
Schwaerzel 2003	1C	*rut^{2080}; MB247*	Canton-S	6	6	EA-IA	26	130	3	80

(*Continued*)

Table 1. (*Continued*)

Study	Fig.	Genotype, Experimental	Genotype, Control	N (E)	N (C)	Odor Pair	Experimental Temp. °C		Shock (V)	Time (min)	RH (%)
rutabaga^{2080}; UAS-rut											
Blum 2009	3A	*rut^{2080}; +; UAS-rut*	*+/rut^{2080}; +; UAS-rut*	6	6	MCH-OCT	22		60	2	50
Blum 2009	4A	*rut^{2080}; +; UAS-rut*	*+/rut^{2080}; +; UAS-rut*	8	8	MCH-OCT	22		60	2	50
Blum 2009	4B	*rut^{2080}; +; UAS-rut*	*+/rut^{2080}; +; UAS-rut*	12	12	MCH-OCT	22		60	2	50
Blum 2009	6A	*rut^{2080}; +; UAS-rut*	*+/rut^{2080}; +; UAS-rut*	6	6	MCH-OCT	22		60	2	50
McGuire 2003	2A	*rut^{2080}; +; UAS-rut*	Canton-S	5	5	OCT-BEN	25		90	3	-
Schwaerzel 2003	1C	*rut^{2080}; +; UAS-rut*	Canton-S	6	6	EA-IA	26		130	3	80
Thum 2007	1D	*rut^{2080}; +; UAS-rut*	*MB247/+*	6	6	MCH-OCT	25		90 DC	0*	-
Zars 2000	1	*rut^{2080}; +; UAS-rut*	Canton-S	8	8	MCH-BEN	25		120 AC	2	-
rutabaga^{1}; UAS-rut											
Scheunemann 2012	5A	*rut^{1}; +; UAS-rut*	wild type	8	8	EA-IA	24		120 AC	3	70
rutabaga^{1}											
Blum 2009	1A	*rut^{1}*	*rut^{1}/+*	6	6	MCH-OCT	22		60	2	50
Scheunemann 2012	5A	*rut^{1}*	wild type	8	8	EA-IA	24		120 AC	3	70
UAS-shibire^{ts} inactivation of the αβ (alphabeta) lobes							**Rest.**	**Perm.**			
McGuire 2001	2AB	*c739; UAS-shi^{ts1}*		6	6	OCT-BEN	32	25	90	3	-
Akalal 2006	4A	*c739; UAS-shi^{ts1}*		6	6	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4B	*c739; UAS-shi^{ts1}*		10	6	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4C	*c739; UAS-shi^{ts1}*		6	9	OCT-BEN	32–35	21–25	90DC	3	60–68
Akalal 2006	4D	*17d; UAS-shi^{ts1}*		6	10	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4E	*17d; UAS-shi^{ts1}*		10	10	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4F	*17d; UAS-shi^{ts1}*		13	10	OCT-BEN	32–35	21–25	90 DC	3	60–68
UAS-shibire^{ts} inactivation of the αβγ lobes											
Dubnau 2001	3A	*UAS-shi^{ts1}/c309*		6	6	MCH-OCT	30	20	-	0*	-
Dubnau 2001	3A	*UAS-shi^{ts1}/c747*		6	6	MCH-OCT	30	20	-	0*	-
McGuire 2001	2AB	*MB247; UAS-shi^{ts1}*		6	6	OCT-BEN	32	25	90	3	-
Schwaerzel 2002	3C	*MB247/UAS-shi^{ts1}*		6	6	OCT-BEN	34	26	-	3	85
Schwaerzel 2002	3C	*c772/UAS-shi^{ts2}*		6	6	OCT-BEN	34	26	-	3	85
Schwaerzel 2003	1E	*MB247/UAS-shi^{ts1}*		6	6	EA-IA	34	26	130	3	80
UAS-shibire^{ts} inactivation of the γ (gamma) lobes											
McGuire 2001	2AB	201Y; UAS-shi^{ts1}		3	4	OCT-BEN	32	25	90	3	-
Wild type heat effect controls											
McGuire 2001	2AB	*wCS10*		6	6	OCT-BEN	32	25	90	3	-
Schwaerzel 2002	3C	Canton-S		6	6	OCT-BEN	34	26	-	3	85
Driver heat effect controls											
Akalal 2006	4A	*17d/+*		6	6	MCH-BEN	32–35	21–25	90 DC	3	60–68

(*Continued*)

Table 1. (*Continued*)

Study	Fig.	Genotype, Experimental	Genotype, Control	N (E)	N (C)	Odor Pair	Experimental Temp. °C		Shock (V)	Time (min)	RH (%)
Akalal 2006	4D	*17d/+*		10	6	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4B	*c739/+*		6	10	OCT-BEN	32–35	21–25	90DC	3	60–68
Akalal 2006	4E	*17d/+*		10	10	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4F	*c739/+*		10	13	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4C	*c739/+*		9	6	OCT-BEN	32–35	21–25	90 DC	3	60–68
McGuire 2001	2AB	*201Y*		5	6	OCT-BEN	32	25	90	3	-
McGuire 2001	2AB	*c739*		6	6	OCT-BEN	32	25	90	3	-
McGuire 2001	2AB	*247*		6	6	OCT-BEN	32	25	90	3	-
Schwaerzel 2003	1E	*247/+*		6	6	EA-IA	34	26	130	3	80
UAS-shi^{ts} heat effect controls											
Akalal 2006	4E	*w; UAS-shi^{ts1}*		10	10	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4A	*w; UAS-shi^{ts1}*		6	6	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4F	*w; UAS-shi^{ts1}*		10	13	OCT-BEN	32–35	21–25	90DC	3	60–68
Akalal 2006	4D	*w; UAS-shi^{ts1}*		10	6	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4C	*w; UAS-shi^{ts1}*		9	6	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4B	*w; UAS-shi^{ts1}*		6	10	OCT-BEN	32–35	21–25	90 DC	3	60–68
Dubnau 2001	3A	*shi^{ts1}/+*		6	6	MCH-OCT	30	20	-	0*	-
McGuire 2001	2AB	*UAS-shi^{ts1}*		6	6	OCT-BEN	32	25	90	3	-
Schwaerzel 2002	3C	*UAS-shi^{ts2}/+*		6	6	OCT-BEN	34	26	-	3	85
Schwaerzel 2002	3C	*UAS-shi^{ts1}/+*		6	6	OCT-BEN	34	26	-	3	85
Schwaerzel 2003	1E	*UAS-shi^{ts1}/+*		6	6	EA-IA	34	26	130	3	80

doi:10.1371/journal.pgen.1005718.t001

heterogeneity was high in two and moderate in one, while their subgroup heterogeneity values were 34%, 64% and 80%.

Rutabaga function is required for 60% of wild type learning

We aimed to estimate the learning contribution made by restoring *rutabaga* function to each of the three lobes. The meta-analyses on *rutabaga* experiments produced 6 meta-analytical estimates of the effects of manipulating *rut* in the mushroom body lobes (Fig 2B). Data pooled from *rut^{1}* and *rut^{2080}* reveal that the strong *rut* hypomorphic alleles reduce learning to 40% of wild type (-60% [95CI -56, -64]). The forest plot summary in Fig 2A illustrates the individual effect sizes from 36 experiments and pooled effect sizes of the *rut* alleles (complete forest plot is shown in Fig 3). The data exhibit substantial overall heterogeneity (I^2 = 76%) and genotype subgroup heterogeneity (I^2 = 88%). This heterogeneity may derive from the methodological variation noted above, but in the case of the strong *rut* alleles we note that the weakest effect is seen in the *rut^{2080}; UAS-rut* subgroup (-45% [95CI -38, -52]), suggesting leaky expression from

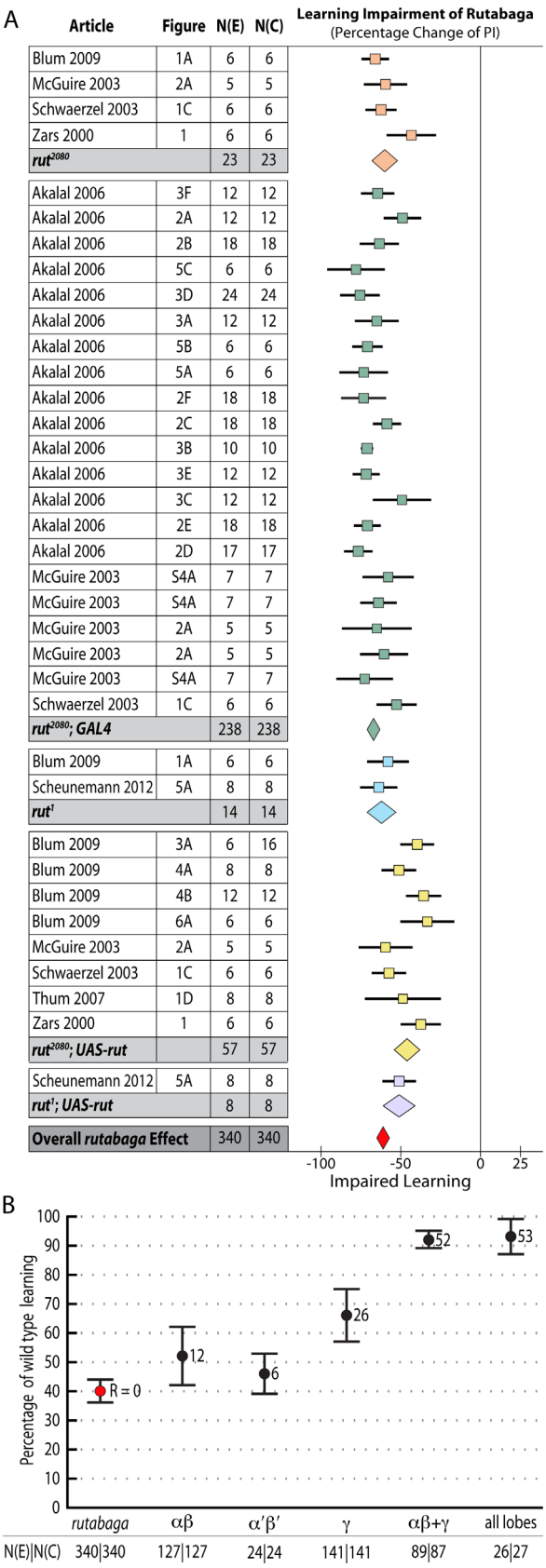
A
Article
Figure
N(E)
N(C)
Learning Impairment of Rutabaga
(Percentage Change of PI)
Blum 2009 1A 6 6
McGuire 2003 2A 5 5
Schwaerzel 2003 1C 6 6
Zars 2000 1 6 6
rut2080 23 23
Akalal 2006 3F 12 12
Akalal 2006 2A 12 12
Akalal 2006 2B 18 18
Akalal 2006 5C 6 6
Akalal 2006 3D 24 24
Akalal 2006 3A 12 12
Akalal 2006 5B 6 6
Akalal 2006 5A 6 6
Akalal 2006 2F 18 18
Akalal 2006 2C 18 18
Akalal 2006 3B 10 10
Akalal 2006 3E 12 12
Akalal 2006 3C 12 12
Akalal 2006 2E 18 18
Akalal 2006 2D 17 17
McGuire 2003 S4A 7 7
McGuire 2003 S4A 7 7
McGuire 2003 2A 5 5
McGuire 2003 2A 5 5
McGuire 2003 S4A 7 7
Schwaerzel 2003 1C 6 6
rut2080; GAL4 238 238
Blum 2009 1A 6 6
Scheunemann 2012 5A 8 8
rut1 14 14
Blum 2009 3A 6 16
Blum 2009 4A 8 8
Blum 2009 4B 12 12
Blum 2009 6A 6 6
McGuire 2003 2A 5 5
Schwaerzel 2003 1C 6 6
Thum 2007 1D 8 8
Zars 2000 1 6 6
rut2080; UAS-rut 57 57
Scheunemann 2012 5A 8 8
rut1; UAS-rut 8 8
Overall rutabaga Effect 340 340
-100
-50
0
25
Impaired Learning
B
Percentage of wild type learning
100
90
80
70
60
50
40
30
20
10
0
R = 0
12
6
26
52
53
rutabaga
αβ
α′β′
γ
αβ+γ
all lobes
N(E)|N(C) 340|340 127|127 24|24 141|141 89|87 26|27

Fig 2. Meta-analyses of *rutabaga* mutant lines and targeted transgenic restoration. Short-term memory data are expressed as percentages. **A.** A summary forest plot of learning changes observed in 340 experiments with *rut* mutant lines, with subgroups showing the differences between the various *rut* alleles and strains. Learning is expressed as a percentage change relative to wild type. The red diamond on the bottom line indicates that the overall impairment in learning in the *rut* hypomorphs relative to wild type controls is -60% [95CI -56%, -64%]. The complete forest plot is given in Fig 3. **B.** Summary estimates from the *rut* mutant meta-analysis and five meta-analyses of lobular restoration experiments. Learning is displayed as a percentage of wild type learning. The markers indicate the proportion of learning relative to wild type expressed as a percentage; error bars are 95% confidence intervals. To the right of the markers are numbers for the amount of rescue (R =) relative to the rut hypomorphs. N(E) and N(C) are the experimental and control iterations respectively. Except for the α′β′ lobes (p = 0.17), all lobe categories showed a statistically significant partial rescue of learning (αβ p = 0.029, γ p<1 x 10–45, αβ+γ p = 1.1 x 10–16, all lobes p<1 x 10–45) when compared with rut learning.

doi:10.1371/journal.pgen.1005718.g002

the transgene as one possible source (i.e. expression from the *UAS-rut* transgene independent of GAL4 transcriptional activation).

Rutabaga restoration to the γ lobes rescues 26% of wild type STM

Some studies have reported that complete rescue requires *rut* restoration in both αβ and γ lobes [7], while others report that restoring *rut* activity in the γ lobe is sufficient to rescue STM, and that the αβ lobes' *rut* activity has little or no STM role [8]. We used the meta-analytic data to specifically examine the lobular specialization hypothesis (Figs 3–8). The overall *rutabaga* loss-of-function effect was used as a reference point to which we compared the lobe restorations, shown in Fig 2B. Restoring *rut* function to each of the lobes revealed partial rescue: α′β′ rescues by 6% [95CI -1.5, 13.5], αβ rescues by 12% [95CI 2, 22] and γ rescues by 26% [95CI 17, 35]. When *rutabaga* was restored to both the αβ and the γ lobes, memory was rescued by 52% [95CI 50, 55]. Restoring *rutabaga* to all three lobes gave only 1% additional improvement (53% [95CI 47, 59]) compared to the rescue in the αβ + γ lobes, therefore *rut* in the α′β′ cells appears to have a minor effect on STM. Of the enhancer trap drivers included in the γ meta-analysis, 201Y contains a minority of αβ cells [33]. A variant analysis that removed 201Y from the γ group and reassigned it to the αβ + γ group resulted in weaker effects for both: only 20% [95CI 10, 31] γ rescue, while αβ + γ rescue was reduced to 49% [95CI 46, 52]. Taken together, these results are incompatible with the hypothesis that restoring *rut* activity to the γ lobe alone is sufficient to rescue the rut^{-} phenotype. From the lobe perspective, we conclude that normal STM requires *rut* function in both αβ + γ lobes.

Heating flies above 30°C impairs short-term memory

Using the temperature-sensitive alleles of *shibire* to block neurotransmission requires heating flies to over 30°C, which can lead to additional heat-related effects [20]. Researchers accommodate this possibility with separate 'heat control' flies that do not express shi^{ts}. We estimated the magnitude of this effect by meta-analysis, shown in Fig 9A (the complete forest plot is shown in Fig 10). Data pooled from 23 such experiments with three types of genotype (wild type, *Driver-GAL4*/+ and *UAS-*shi^{ts}/+) revealed that the overall effect of heating flies from the permissive temperature (20–26°C) to 30–35°C is a 17% [95CI 12, 22] reduction in memory. This decrement can be expected to affect the *UAS-*shi^{ts} inactivation data from the same studies, so we used 83% of wild type memory in Fig 9B as the zero reference point to estimate the specific effects of lobe inactivation.

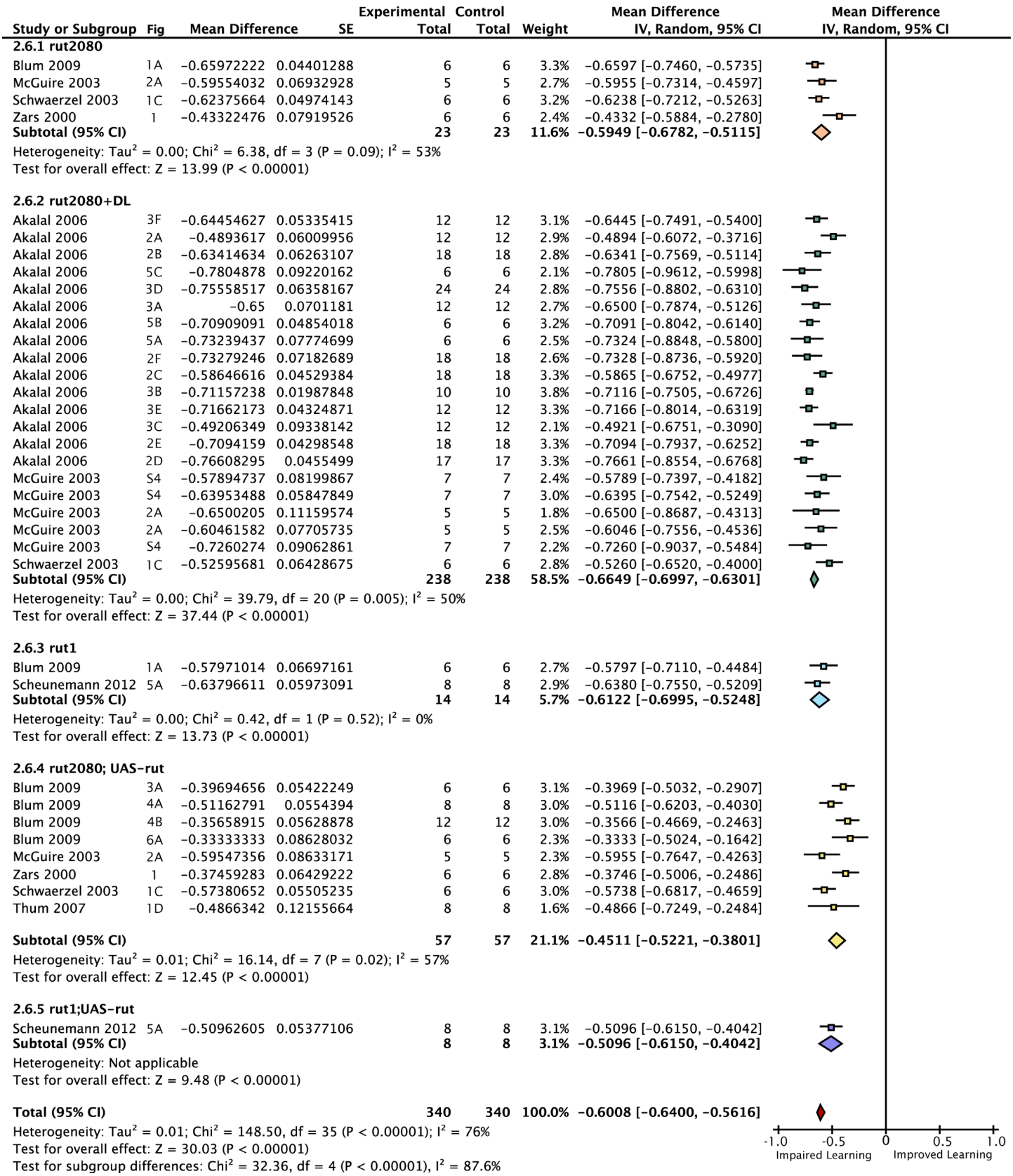
Study or Subgroup | Fig | Mean Difference | SE | Experimental Total | Control Total | Weight | Mean Difference IV, Random, 95% CI | Mean Difference IV, Random, 95% CI
2.6.1 rut2080
Blum 2009 | 1A | -0.65972222 | 0.04401288 | 6 | 6 | 3.3% | -0.6597 [-0.7460, -0.5735]
McGuire 2003 | 2A | -0.59554032 | 0.06932928 | 5 | 5 | 2.7% | -0.5955 [-0.7314, -0.4597]
Schwaerzel 2003 | 1C | -0.62375664 | 0.04974143 | 6 | 6 | 3.2% | -0.6238 [-0.7212, -0.5263]
Zars 2000 | 1 | -0.43322476 | 0.07919526 | 6 | 6 | 2.4% | -0.4332 [-0.5884, -0.2780]
Subtotal (95% CI) | | | | 23 | 23 | 11.6% | -0.5949 [-0.6782, -0.5115]
Heterogeneity: Tau² = 0.00; Chi² = 6.38, df = 3 (P = 0.09); I² = 53%
Test for overall effect: Z = 13.99 (P < 0.00001)
2.6.2 rut2080+DL
Akalal 2006 | 3F | -0.64454627 | 0.05335415 | 12 | 12 | 3.1% | -0.6445 [-0.7491, -0.5400]
Akalal 2006 | 2A | -0.4893617 | 0.06009956 | 12 | 12 | 2.9% | -0.4894 [-0.6072, -0.3716]
Akalal 2006 | 2B | -0.63414634 | 0.06263107 | 18 | 18 | 2.8% | -0.6341 [-0.7569, -0.5114]
Akalal 2006 | 5C | -0.7804878 | 0.09220162 | 6 | 6 | 2.1% | -0.7805 [-0.9612, -0.5998]
Akalal 2006 | 3D | -0.75558517 | 0.06358167 | 24 | 24 | 2.8% | -0.7556 [-0.8802, -0.6310]
Akalal 2006 | 3A | -0.65 | 0.0701181 | 12 | 12 | 2.7% | -0.6500 [-0.7874, -0.5126]
Akalal 2006 | 5B | -0.70909091 | 0.04854018 | 6 | 6 | 3.2% | -0.7091 [-0.8042, -0.6140]
Akalal 2006 | 5A | -0.73239437 | 0.07774699 | 6 | 6 | 2.5% | -0.7324 [-0.8848, -0.5800]
Akalal 2006 | 2F | -0.73279246 | 0.07182689 | 18 | 18 | 2.6% | -0.7328 [-0.8736, -0.5920]
Akalal 2006 | 2C | -0.58646616 | 0.04529384 | 18 | 18 | 3.3% | -0.5865 [-0.6752, -0.4977]
Akalal 2006 | 3B | -0.71157238 | 0.01987848 | 10 | 10 | 3.8% | -0.7116 [-0.7505, -0.6726]
Akalal 2006 | 3E | -0.71662173 | 0.04324871 | 12 | 12 | 3.3% | -0.7166 [-0.8014, -0.6319]
Akalal 2006 | 3C | -0.49206349 | 0.09338142 | 12 | 12 | 2.1% | -0.4921 [-0.6751, -0.3090]
Akalal 2006 | 2E | -0.7094159 | 0.04298548 | 18 | 18 | 3.3% | -0.7094 [-0.7937, -0.6252]
Akalal 2006 | 2D | -0.76608295 | 0.0455499 | 17 | 17 | 3.3% | -0.7661 [-0.8554, -0.6768]
McGuire 2003 | S4 | -0.57894737 | 0.08199867 | 7 | 7 | 2.4% | -0.5789 [-0.7397, -0.4182]
McGuire 2003 | S4 | -0.63953488 | 0.05847849 | 7 | 7 | 3.0% | -0.6395 [-0.7542, -0.5249]
McGuire 2003 | 2A | -0.6500205 | 0.11159574 | 5 | 5 | 1.8% | -0.6500 [-0.8687, -0.4313]
McGuire 2003 | 2A | -0.60461582 | 0.07705735 | 5 | 5 | 2.5% | -0.6046 [-0.7556, -0.4536]
McGuire 2003 | S4 | -0.7260274 | 0.09062861 | 7 | 7 | 2.2% | -0.7260 [-0.9037, -0.5484]
Schwaerzel 2003 | 1C | -0.52595681 | 0.06428675 | 6 | 6 | 2.8% | -0.5260 [-0.6520, -0.4000]
Subtotal (95% CI) | | | | 238 | 238 | 58.5% | -0.6649 [-0.6997, -0.6301]
Heterogeneity: Tau² = 0.00; Chi² = 39.79, df = 20 (P = 0.005); I² = 50%
Test for overall effect: Z = 37.44 (P < 0.00001)
2.6.3 rut1
Blum 2009 | 1A | -0.57971014 | 0.06697161 | 6 | 6 | 2.7% | -0.5797 [-0.7110, -0.4484]
Scheunemann 2012 | 5A | -0.63796611 | 0.05973091 | 8 | 8 | 2.9% | -0.6380 [-0.7550, -0.5209]
Subtotal (95% CI) | | | | 14 | 14 | 5.7% | -0.6122 [-0.6995, -0.5248]
Heterogeneity: Tau² = 0.00; Chi² = 0.42, df = 1 (P = 0.52); I² = 0%
Test for overall effect: Z = 13.73 (P < 0.00001)
2.6.4 rut2080; UAS-rut
Blum 2009 | 3A | -0.39694656 | 0.05422249 | 6 | 6 | 3.1% | -0.3969 [-0.5032, -0.2907]
Blum 2009 | 4A | -0.51162791 | 0.0554394 | 8 | 8 | 3.0% | -0.5116 [-0.6203, -0.4030]
Blum 2009 | 4B | -0.35658915 | 0.05628878 | 12 | 12 | 3.0% | -0.3566 [-0.4669, -0.2463]
Blum 2009 | 6A | -0.33333333 | 0.08628032 | 6 | 6 | 2.3% | -0.3333 [-0.5024, -0.1642]
McGuire 2003 | 2A | -0.59547356 | 0.08633171 | 5 | 5 | 2.3% | -0.5955 [-0.7647, -0.4263]
Zars 2000 | 1 | -0.37459283 | 0.06429222 | 6 | 6 | 2.8% | -0.3746 [-0.5006, -0.2486]
Schwaerzel 2003 | 1C | -0.57380652 | 0.05505235 | 6 | 6 | 3.0% | -0.5738 [-0.6817, -0.4659]
Thum 2007 | 1D | -0.4866342 | 0.12155664 | 8 | 8 | 1.6% | -0.4866 [-0.7249, -0.2484]
Subtotal (95% CI) | | | | 57 | 57 | 21.1% | -0.4511 [-0.5221, -0.3801]
Heterogeneity: Tau² = 0.01; Chi² = 16.14, df = 7 (P = 0.02); I² = 57%
Test for overall effect: Z = 12.45 (P < 0.00001)
2.6.5 rut1;UAS-rut
Scheunemann 2012 | 5A | -0.50962605 | 0.05377106 | 8 | 8 | 3.1% | -0.5096 [-0.6150, -0.4042]
Subtotal (95% CI) | | | | 8 | 8 | 3.1% | -0.5096 [-0.6150, -0.4042]
Heterogeneity: Not applicable
Test for overall effect: Z = 9.48 (P < 0.00001)
Total (95% CI) | | | | 340 | 340 | 100.0% | -0.6008 [-0.6400, -0.5616]
Heterogeneity: Tau² = 0.01; Chi² = 148.50, df = 35 (P < 0.00001); I² = 76%
Test for overall effect: Z = 30.03 (P < 0.00001)
Test for subgroup differences: Chi² = 32.36, df = 4 (P < 0.00001), I² = 87.6%
-1.0
-0.5
0
0.5
1.0
Impaired Learning
Improved Learning

Fig 3. Forest plot of *rut* mutant learning changes. Each data set is identified by the source article and figure panel. This figure is a detailed version of the same plot in the main article, but uses proportional reductions instead of percentage changes. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the percentage change relative to control.

doi:10.1371/journal.pgen.1005718.g003

Neurotransmission from the αβ + γ lobes accounts for 61% of STM

Inactivating the αβ lobes alone produced a 25% [95CI 14, 37] reduction in STM (Fig 11). Drivers that express in both the αβ and γ lobes reduced performance by 61% [95CI 50, 72] relative to heated control flies (Fig 12). The best estimate for γ lobe inactivation is a 6% reduction [95CI 35% reduction, 24% increase] relative to heated controls. This γ lobe estimate appears to be negligible, but has very wide confidence intervals and is drawn from only a single experiment with three iterations. Surprisingly, the literature review found no <5 min STM data on the impact of *shibire*^{ts} inactivation of either the entire mushroom body (All lobes) or the α′β′ lobes (empty columns in Fig 9B); at the time of the review the only studies reporting results for these interventions examined later memory, at 15 min or beyond [20]. The substantial decrement in the αβ lobe inactivation experiments (25% reduction) is incompatible with the idea that this lobe plays only a negligible role in STM. The paucity of data for γ, α′β′ and All lobes in STM highlights an area that would benefit from future experimental attention.

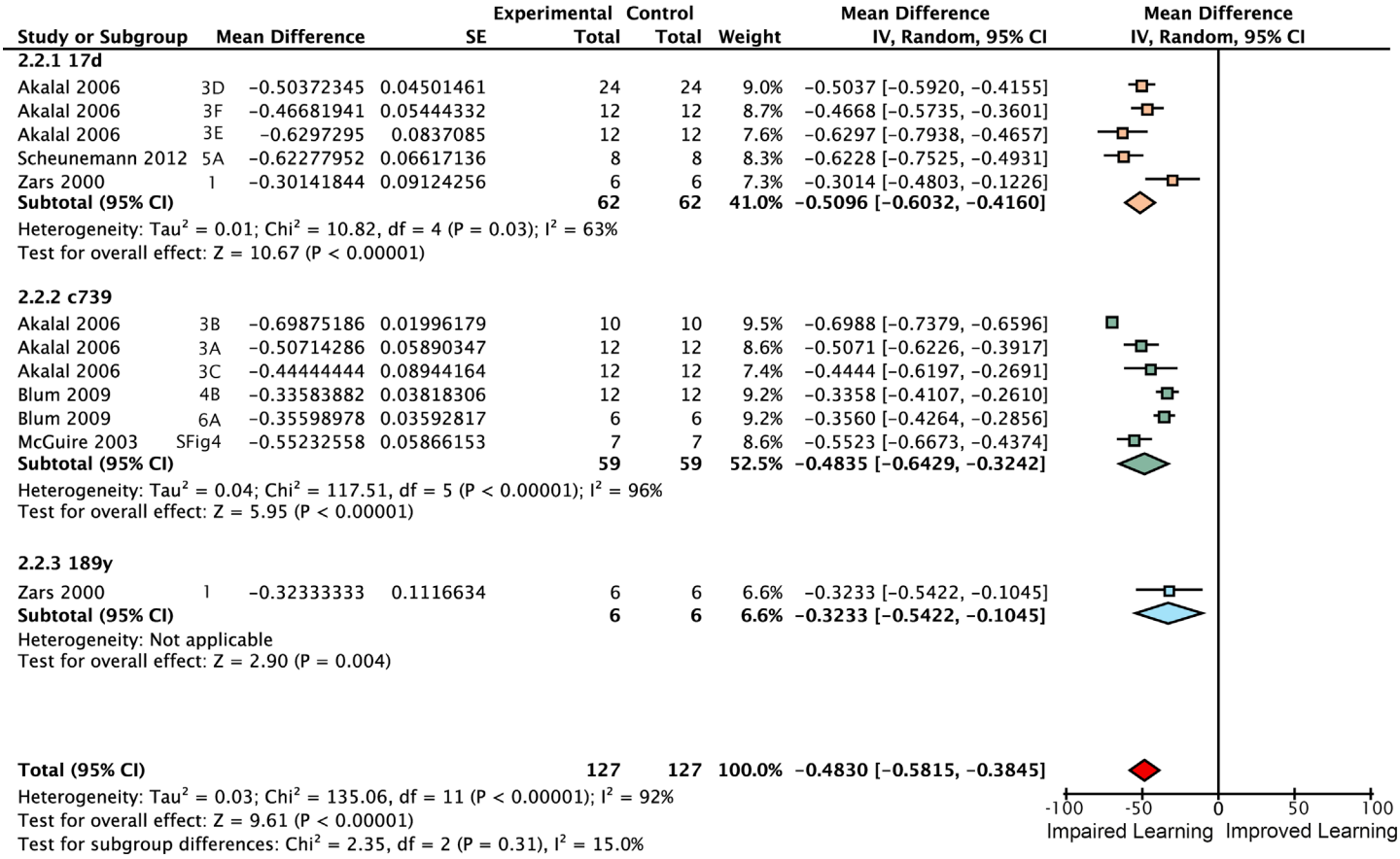

Fig 4. Forest plot of *rut* restoration in the αβ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

doi:10.1371/journal.pgen.1005718.g004

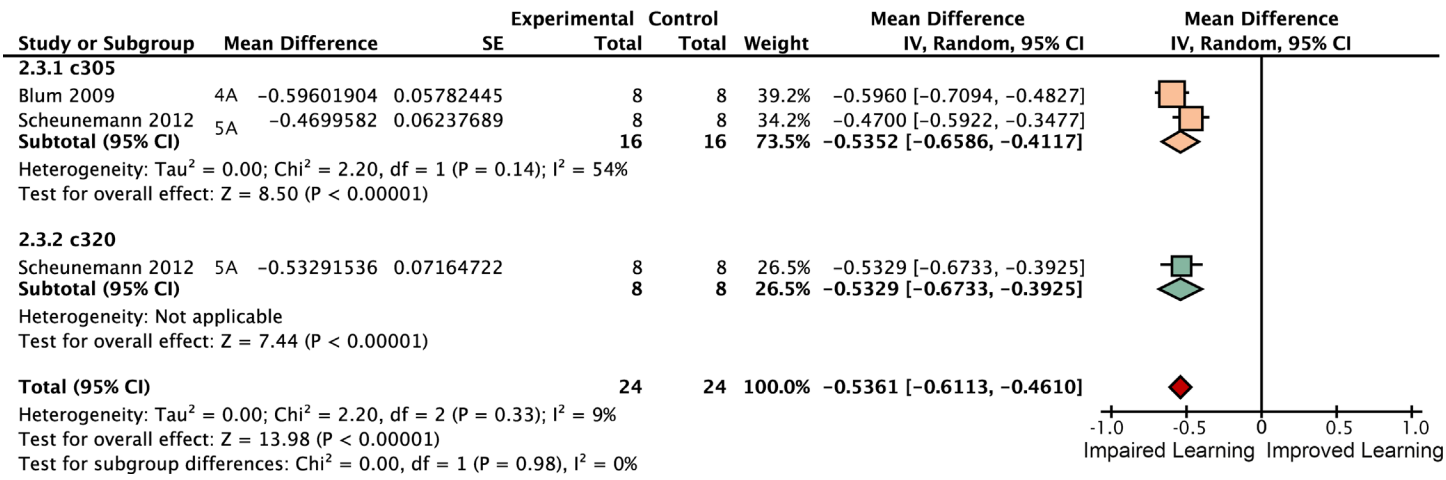

Fig 5. Forest plot of *rut* restoration in the α′β′ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

doi:10.1371/journal.pgen.1005718.g005

Cell number accounts for the majority of driver variation

Observing high heterogeneity (I^2) in some of the meta-analyses, we attempted to identify the source of variability, and examine the original hypothesis from a different perspective.

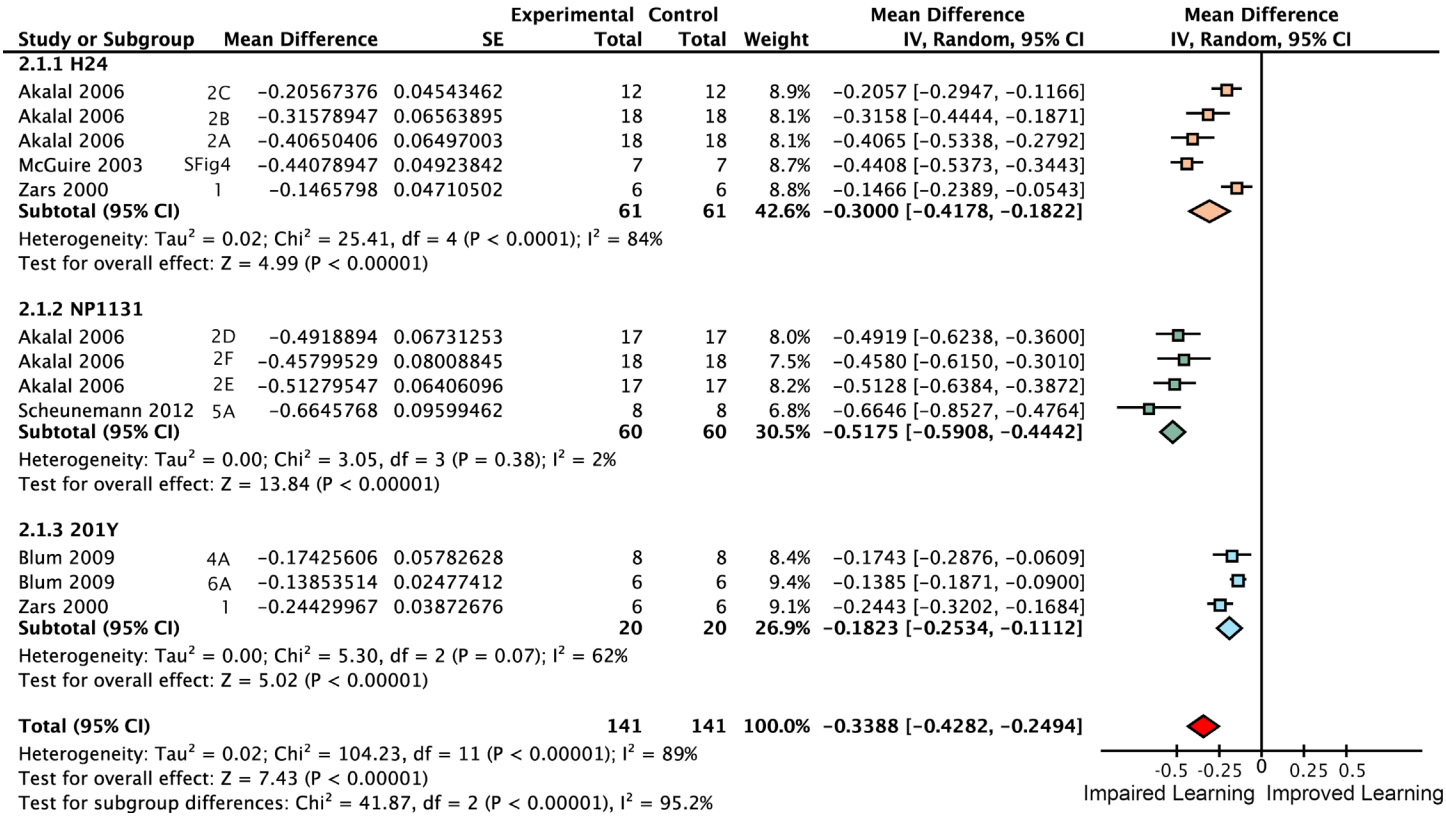

Fig 6. Forest plot of *rut* restoration in the γ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

doi:10.1371/journal.pgen.1005718.g006

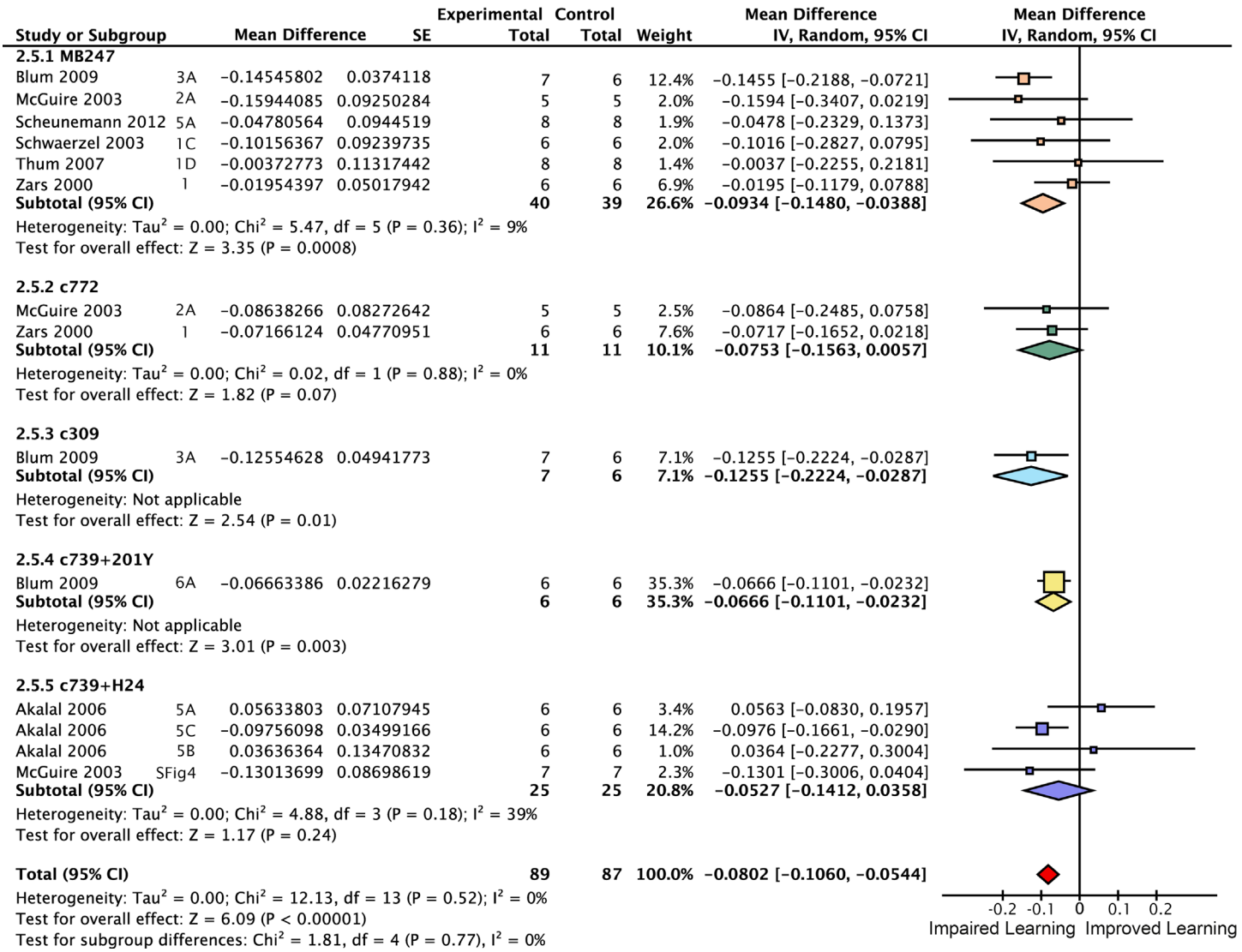

Fig 7. Forest plot of *rut* restoration in the αβ and γ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

doi:10.1371/journal.pgen.1005718.g007

Electrophysiological evidence [34] and anatomical connectivity analysis [35] indicate that the Kenyon cells, the intrinsic neurons of the mushroom body, are randomly connected to their olfactory input neurons. The lack of structured connectivity suggests that, for some or all odor-related functions, individual Kenyon cells are interchangeable; thus raising the possibility that a cell's lobular identity might be less important than its participation in a stochastically nominated odor-responsive ensemble. As three of the seven relevant meta-analyses showed driver heterogeneity as accounting for more than half of their variance, we asked whether the number of cells captured by a driver could explain some of the unaccounted variance. We extracted cell count data from an anatomical study that counted Kenyon cells for many of the drivers [33]. The driver-specific meta-analytic STM estimates were subjected to an initial simple linear regression against the drivers' available cell counts in both *rut* restoration and shi^{ts} inactivation. These indicated that cell numbers accounted for about 80% of the driver memory variance (*rut* $R^2 = 0.79$ [95CI 0.39, 0.94], $p = 2.5 \times 10^{-4}$; shi^{ts} $R^2 = 0.77$ [95CI 0.14, 0.96], $p = 8.4 \times 10^{-3}$). As

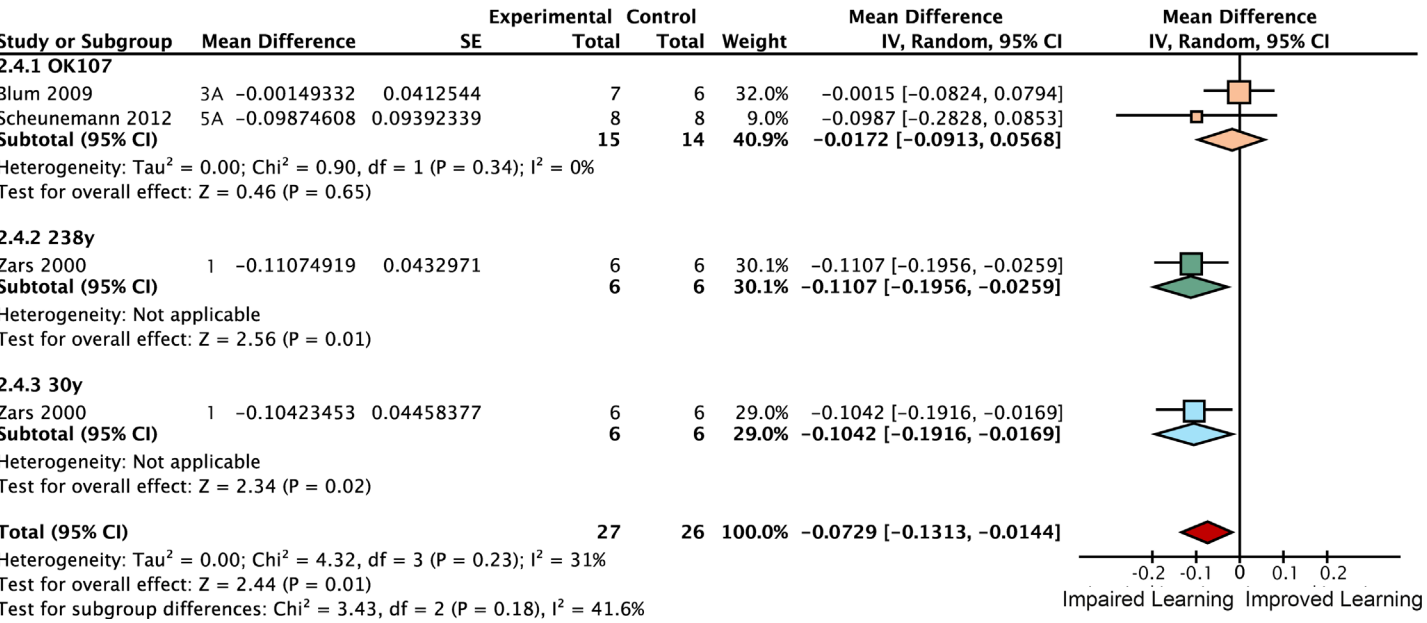

Fig 8. Forest plot of *rut* restoration in all lobes of the mushroom body. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

doi:10.1371/journal.pgen.1005718.g008

simple linear regression is unable to account for the full complexity of such hierarchical data, we constructed hierarchical, multivariate, weighted meta-regression models accommodating other variables that might explain some of the variance induced by differences in experimental design. These models were also able to account for the clustering of experiments within studies and for the shared control design in *rut* experiments, and included weighted estimates for each driver by the number of contributing experiments (described fully in Methods). The hierarchical meta-regression model of *rut* showed a strong relationship with driver cell count, generalized-R^2 = 0.84 [95CI 0.79, 0.89] (Fig 13A). The meta-regression model of *shi* data similarly revealed a large effect size for the cell count relationship, generalized-R^2 = 0.88 [95CI 0.84, 0.92] (Fig 13B). Compared with simple linear regression, the hierarchical models revealed stronger trends with substantially improved precision. These results are incompatible with the strong lobular specialization hypothesis of *rut* and *shi* function. Rather, drawing on data from thousands of T-maze iterations (N = 1008, 1006) while accounting for experimental heterogeneity, they constitute compelling evidence that each driver's extent of neuronal expression can account for the majority of that driver's short-term memory effect.

Kenyon cells in different lobes make equivalent contributions to STM

Different Kenyon cell drivers' varying impact on learning is primarily a result of how many cells they are expressed in: cell count as the overwhelmingly dominant factor therefore excludes highly specialized roles for *rut* and *shi* in different lobes' Kenyon cells. However, it is possible that minor quantitative differences explain the remaining unaccounted for 12–16% of STM variance in the meta-regression models. Within the overall memory-cell count trend in Fig 13A, several drivers' estimates do not fall on the regression line. To account for such deviations from the overall cell number trend, we aimed to factor out cell number and focus specifically on the potency of each neuron captured by a driver. We built new models in which the learning effect size of each driver line was first divided by the number of expressing cells, and weighted

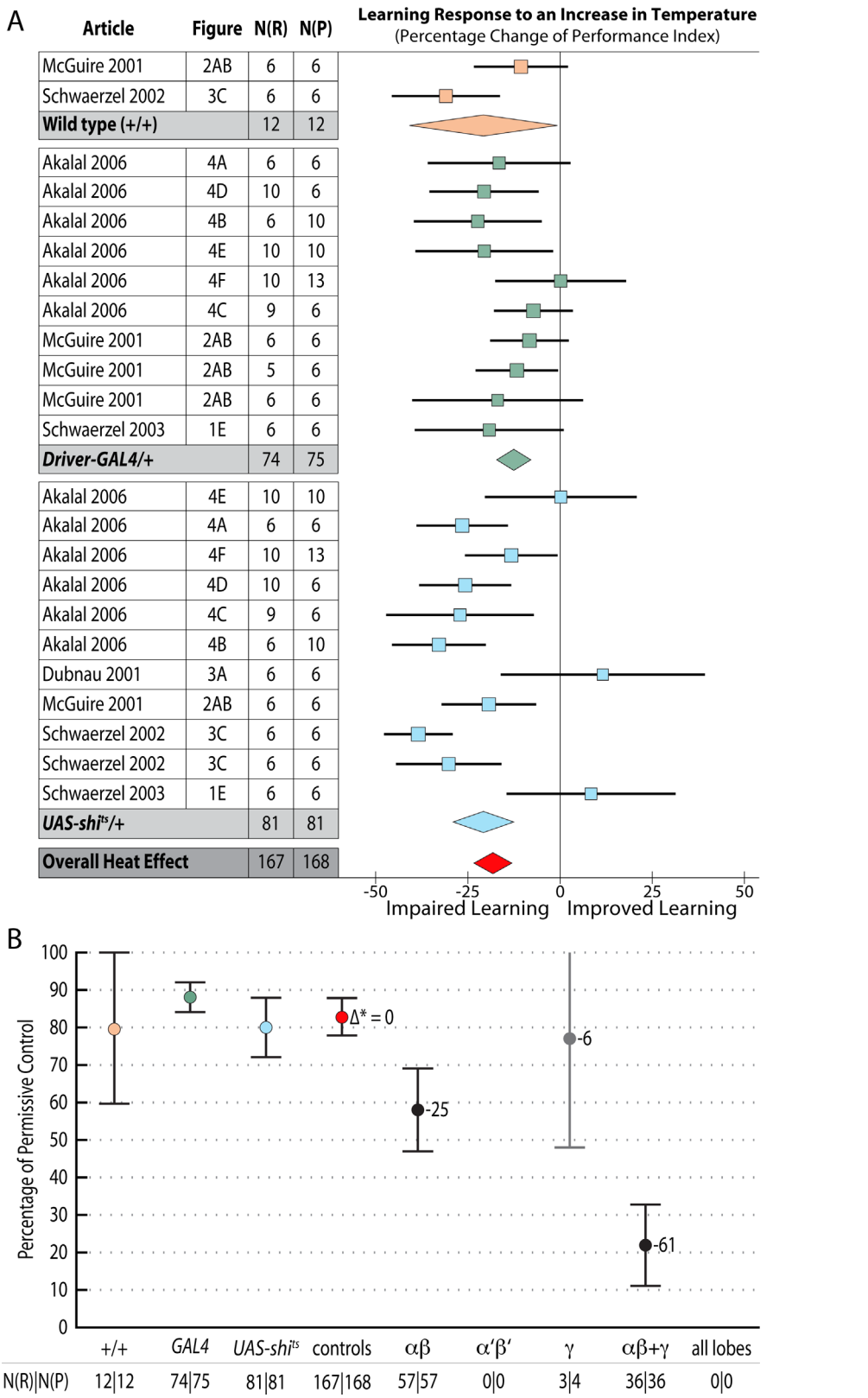

Fig 9. Meta-analyses of *shibire*^{ts} inhibition of neurotransmission in the mushroom body lobes. Learning data are expressed as percentages. **A.** A summary forest plot of learning changes in heat treatment

controls, with subgroups showing the differences between 3 types of controls. Learning is expressed as a percentage change relative to wild type. The red diamond on the bottom line indicates that the overall impairment in learning in flies exposed to elevated temperature is -17% [95CI -12%, -22%]. A complete forest plot is shown in Fig 10. **B.** Summary estimates from the heat exposure controls and three meta-analyses of lobular inactivation experiments. Colored markers correspond to diamonds in panel A. Learning at the restrictive temperature is shown as a percentage of learning at the permissive temperature; error bars are 95% confidence intervals. To the right of the markers are numbers learning impairment (* =) relative to the synthetic heat effect control. N(R) and N(P) are the restrictive and permissive iterations respectively. The αβ lobes (p = 0.0001) and the αβ+γ combination (p<1 x 10–45) show statistically significant impairment while the γ lobes do not (p = 0.7071). The γ lobe bar is in grey as it derives from only a single experiment with few replicates. There were no data in the literature on the α′β′ lobes or drivers that encompass all mushroom body lobes.

doi:10.1371/journal.pgen.1005718.g009

hierarchical meta-regression models were then used to perform synthesis by lobular category. These models produced estimates of a typical Kenyon cell's effectiveness within each lobe

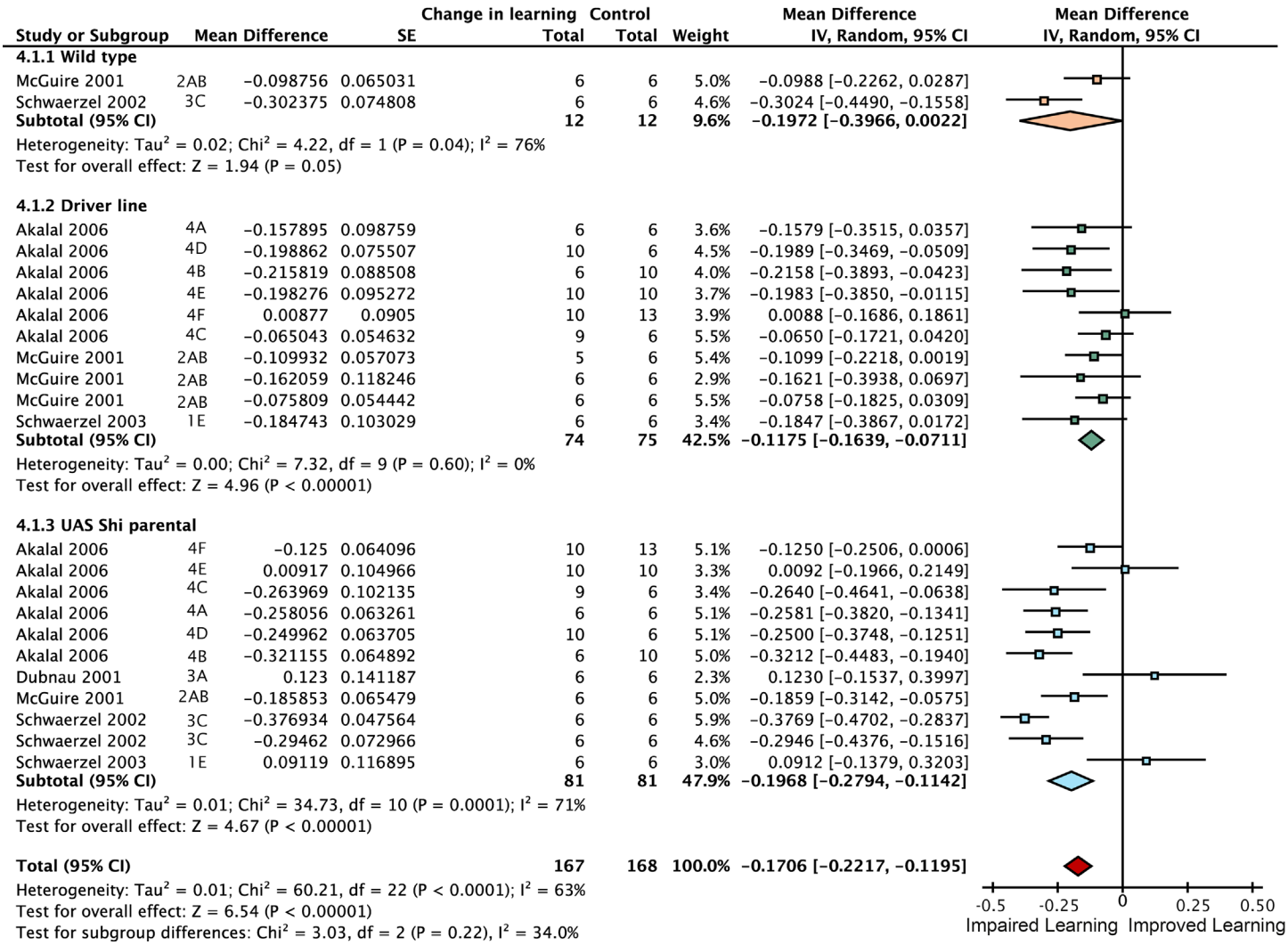

Fig 10. Forest plot of the effect on STM of elevating flies from permissive to restrictive temperatures. This figure is a detailed version of the same plot in the previous figure, but uses proportional reductions instead of percentage changes. The source article and figure panel identifies each data set. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

doi:10.1371/journal.pgen.1005718.g010

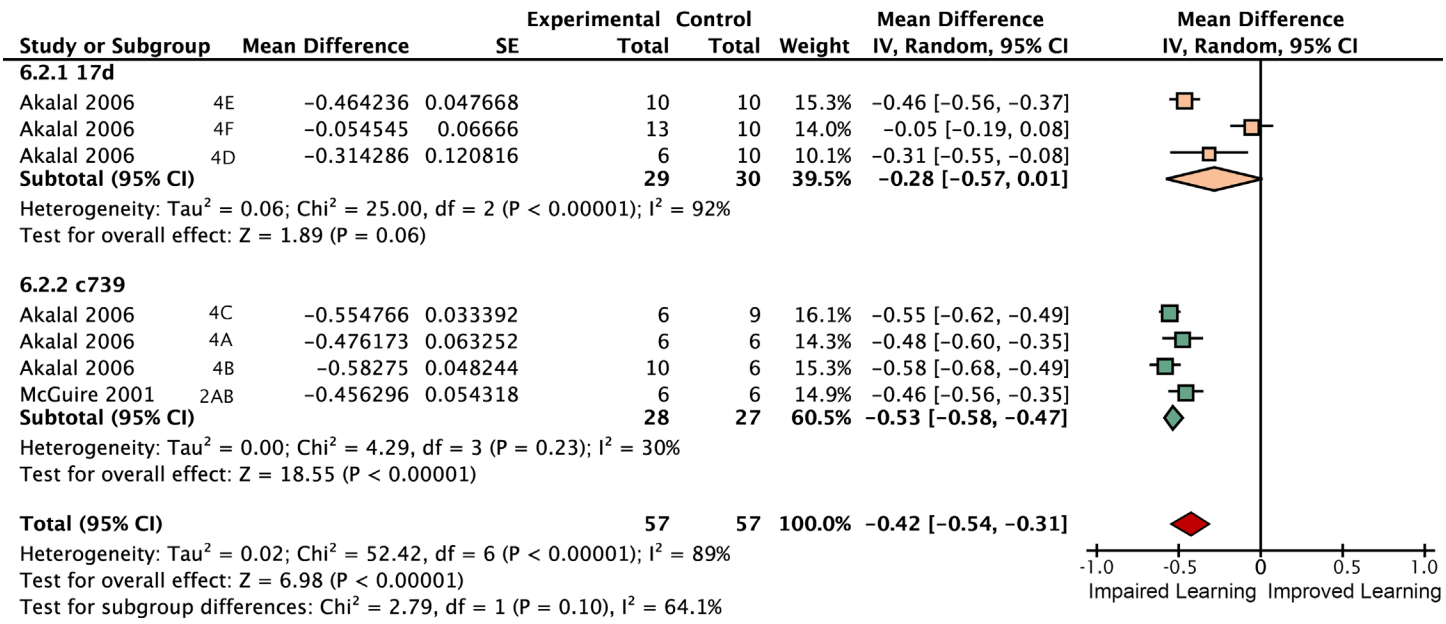

Fig 11. Forest plot of experiments using *shi*^{ts} to inactivate neurotransmission from the αβ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

doi:10.1371/journal.pgen.1005718.g011

category (Fig 13C and 13D). The *rut* rescue-per-cell data and the *shi* loss-per-cell data both show that there are no substantial differences between any lobe categories. In summary, when cell numbers are taken into account, the evidence does not support the strong lobular specialization hypothesis. Instead, it shows that lobular *rut* function is non-specialized and that STM makes use of all available functioning Kenyon cells.

Discussion

Previous studies concluded that differences between mushroom body lobes exist that reflect functional specializations in the various memory phases (STM, MTM and LTM). These conclusions about lobular specialization included the idea that γ lobe *rut* function is sufficient for STM formation. The aim of the present study was to specifically examine the strong lobular specialization STM hypothesis. Surprisingly, the synthetic evidence is incompatible with lobular specialization, and supports the alternative idea that STM function is generalized across lobes.

Meta-analysis of strong *rut* hypomorphic alleles confirmed that they cause a 60% reduction in STM. As previously reported in the literature, the other 40% must be mediated by other molecular factors either in the Kenyon cells or elsewhere. Restoring *rut* activity with lobe-targeting drivers revealed that partial rescue occurs in both the γ and αβ lobes (mean 26% and 12%), with a partial rescue even in the α′β′ lobes (mean 6%). To rescue the majority of lost function, *rut* had to be expressed in both αβ and γ lobes (Fig 2B). These data are incompatible with the hypothesis that the lobes' *rut* activity in the γ lobe is absolutely or strongly specialized for STM. With the synthesized evidence failing to support strong lobular specialization of *rut* in STM (Fig 2B), we considered an alternative hypothesis: that cell extent is the main predictor of a transgenic driver's STM impact. Indeed, multivariate meta-regression models incorporating cell count show that the dominant factor influencing STM is the number of Kenyon cells targeted by a specific driver line, for both *rut* and *shi* effects (Fig 13A and 13B). This result

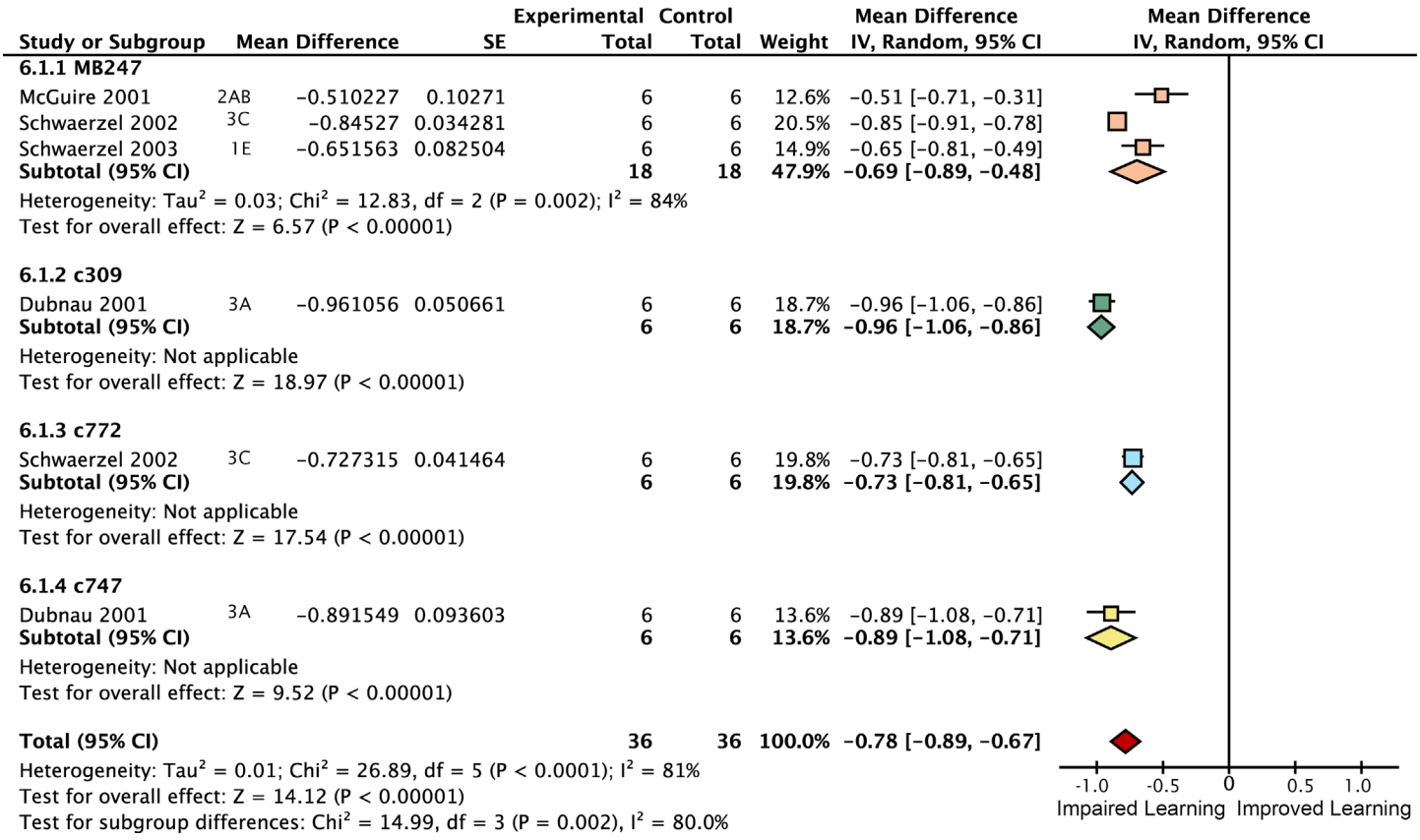

Fig 12. Forest plot of experiments using *shi*^{ts} to inactivate neurotransmission from the αβ + γ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

doi:10.1371/journal.pgen.1005718.g012

refutes the hypothesis that the mushroom lobes are specialized for aversive STM function. Rather, the linear relationships lead us to conclude that the different lobes' cells have similar potency for STM with regard to *rut* and *shi*-dependent memory processes.

Despite the paucity of experiments for *shi* in the γ, α′β′ and All lobes categories, the available data were sufficient to allow construction of a precise model of the relationship between driver cell count and memory. If STM relied on neurotransmission from a highly inter-dependent Kenyon cell ensemble, we would anticipate that *shi*^{ts} inhibition of small subsets of these cells would have a large effect. Instead, the observed linear trend between driver cell count and STM impact (Fig 13B) supports a model in which *shi*-dependent memory function in the αβ and γ cells occurs autonomously in individual cells or small groups of cells. It appears that associative olfactory information is initially processed in a distributed manner across the mushroom body. It appears that strong qualitative specialization of lobular neurotransmission emerges over the subsequent minutes and hours as later memory forms [18,20]. Further investigation of lobular specialization during the different memory phases could apply a combination of meta-analysis and experimental analysis. In the latter case important experiments would include examining genes beyond *rut* or *shi*, and the use of new driver lines with even more diverse lobe coverage to more thoroughly dissociate lobe identity from cell count.

The benefits of systematic review include gaining an estimate of statistical heterogeneity (I^2) in the data and an overview of the methodological variability. While the T-maze STM protocol is a largely standardized protocol, there is room for even greater standardization (Table 1) that

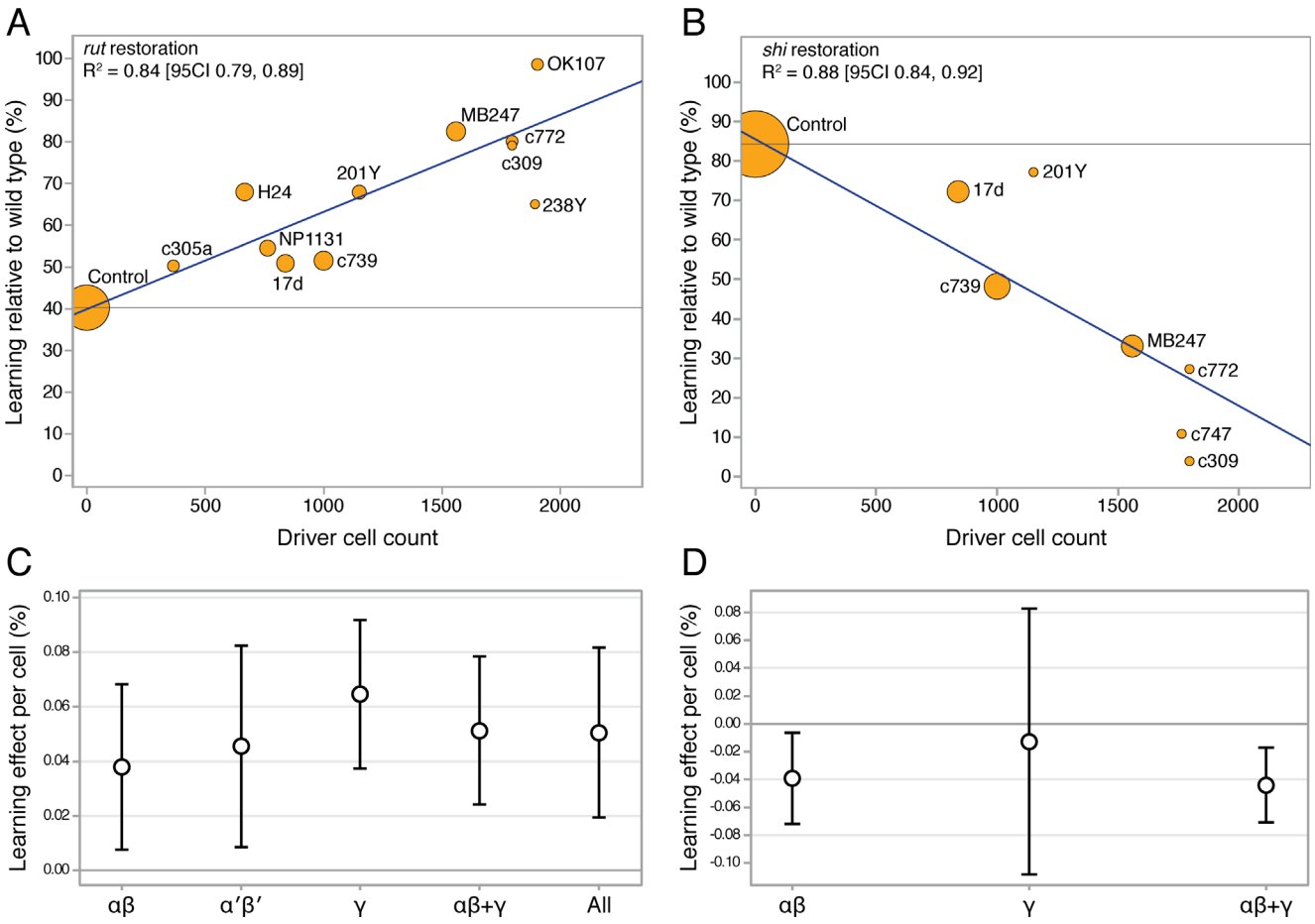

Fig 13. The extent of drivers' Kenyon cell expression accounts for the majority of short-term olfactory memory effects. The estimated Kenyon cell counts for drivers were taken from Aso et al. 2009. The memory effect sizes are derived from nested, weighted, multivariate meta-regression models that adjusted for confounding variables that contributed to heterogeneity. **A.** Bubble plot of *rut* restoration; the cell count of driver lines accounts for 84% of the variance of the learning effects of rut restoration ($p < 0.0001$). Each bubble's area indicates that estimate's weight in the regression model; the blue fit line has a slope of 0.023% per cell [95CI 0.016, 0.030]. The grey line indicates the level of no rescue, i.e. the learning level of *rut* mutants. **B.** For shi^{ts} inactivation, 88% of the learning variance is attributable to the number of cells encompassed by the driver ($p < 0.0001$). The blue fit line has a slope of -0.034% per cell [95CI -0.046, -0.0216]; the grey line indicates the level of no effect, i.e. the learning expected from the effect of heat alone. **C.** Learning effect per cell in mushroom body sub-regions from *rut* restoration in different lobes and combinations, adjusted for heterogeneity effects. Error bars are confidence intervals; there are no statistical differences between rut lobe categories. **D.** The shi^{ts} learning effect per cell in two lobes and their combination. There are no statistical differences between shi^{ts} lobe categories.

doi:10.1371/journal.pgen.1005718.g013

would likely help improve inter-study reproducibility and facilitate meta-analysis, perhaps reducing the need for complex modeling. Standardization would ideally involve adopting consistent values for all relevant experimental parameters (e.g. voltage, voltage type, relative humidity) that are currently sometimes omitted from published reports. Systematic methodological review is useful to identify censored and inconsistent experimental conditions.

This investigation serves as a case study in how systematic review, meta-analysis, and related estimation methods can help biological data analysis. Recent commentary has focused attention on reproducibility [21,36,37] and replication [38]; both of these issues are in part connected to significance testing. An encouraging aspect that was revealed as a part of this study is that the existing published data could support precise estimation with hierarchical modeling, suggesting firm data integrity in the fly memory neurogenetics field. Significance testing has been controversial in the behavioral sciences for half a century [39] but it remains the

dominant statistical methodology in neuroscience and other life sciences [40] while alternatives have yet to be widely adopted. Estimation is a data analysis framework that places the emphasis on effect sizes and the meta-analytic perspective [22,23,41]. This study shows how systematic review in conjunction with several meta-analytic techniques enable the synthesis of relevant available evidence so as to address inconsistencies in a field and reveal unexpected patterns in published data [28]. Estimation statistics is also appropriate for primary research; modern texts advise that reporting effect sizes with their confidence intervals, along with the use of graphical methods, are the rightful priorities of data analysis [23,25,41]. Hierarchical models can similarly be applied routinely to analyze primary data with complex experimental designs, such as experiments conducted in different labs [42] or by differing protocols within a lab [43,44], replacing basic methods such as ANOVA. Our results add further weight to the case that estimation is a superior statistical framework for the various phases of biological research: planning, analysis, interpretation and review.

Materials and Methods

Eligibility criteria and information sources

All information was sourced with searches of PubMed. To be eligible for consideration for inclusion in the systematic review each study was required to meet the following criteria: containing olfactory STM experiments on *Drosophila melanogaster* using the classic T-maze apparatus and a single training cycle [29]; reporting of the relevant control and experimental data as a Performance Index (PI); detailing the relevant genotypes and the number of experimental iterations (N or sample size). In addition, as STM is thought to begin to transition to MTM shortly after training [9], we defined STM as using a post-training delay of 5 minutes or less. All studies selected contained transgenic manipulations of the Kenyon cells targeted to one or more of the 3 lobes (αβ, α′β′, and γ). For the systematic review of *rut* function in the Kenyon cells, studies included use of a hypomorphic allele of the *rut* gene, transgenic drivers and *UAS-rut* expression constructs. Experiments using temporally controlled expression of *rut* were excluded to eliminate the possibility of heterogeneity associated with incomplete restoration due to variations in expression longevity or strength. For the systematic review of endocytosis-dependent neurotransmission in the Kenyon cells, studies included a *UAS-shi^{ts}* transgene in combination with transgenic drivers and heat treatment. Experiments that shifted *shi^{ts}* flies to different temperatures between training and testing were excluded to eliminate the possibility of heterogeneity due to these manipulations; only experiments using the conventional permissive-restrictive (cool-warm) comparison were included. Following the lead of the great majority of the STM literature, we did not attempt to analyze the acquisition, storage and retrieval phases of STM. This report contains the Preferred Reporting Items for Systematic reviews and Meta-Analyses guidelines [45], except for the structured summary and risk of bias analyses.

Database search

The systematic literature search was conducted as follows and is shown as a diagram in Fig 1A. On the 11th July 2013, the search phrase ((((Drosophila) AND (learning OR memory)) AND (mushroom OR Kenyon)) AND ("2000"[Date—Publication]: "3000"[Date—Publication]) NOT review[Publication Type] was used to query PubMed, and the resulting 279 records were downloaded as two.nbib files. These files were imported into Papers2 software, and then exported as EndNote.xml. This file was loaded into EndNote X4, copied into Excel, and then imported into Apple Numbers with all bibliographic information including Title and Abstract stored in one row per record. This was then used to screen the records' titles, abstracts and was

also used to record the results of the full text screen and the detailed experimental design screen.

Study selection

We designed the literature selection process to identify experiments that examined aversive olfactory STM (testing five minutes or less after training) in *Drosophila* as observed in the classic T-maze apparatus. We further aimed to focus the analysis on the two kinds of experiments most commonly used to understand the role of the three mushroom body lobes and the mushroom body intrinsic neurons (Kenyon cells). The first type of experiments was the usage of transgenic *rutabaga* (*rut*) to restore adenylyl cyclase function to one or more lobes in *rut* mutant flies; the second type included experiments targeting transgenic temperature-sensitive SHIBIRE (SHI^{TS}) protein to the lobes to disable dynamin-dependent neurotransmission. The SHI^{TS} proteins form part of the dynamin endocytosis complex and poison its function when flies are transferred to the restrictive temperature [46]. The exact odor pairs under investigation were explicitly disregarded in this analysis; rather, experiments containing the full variety odor pairs were included to enable us to arrive at the most general conclusion about mushroom body function.

Two investigators (TY and JMW) performed the literature review independently and discrepancies were resolved collaboratively with a third investigator (ACC). The 279 records yielded from the PubMed search were screened in four stages to systematically exclude studies: title review, abstract reading, full text scan and a detailed review of experimental design. This process is described in Fig 1A; we used title and abstract information to discover a set of *Drosophila* behavioral studies that were likely to include aversive olfactory conditioning in adult fly (n = 65 studies) and then scanned these full text articles to find *rutabaga* restoration or $shibire^{ts}$ experiments in the MB lobes. The final stage in the selection ("Experimental Design" in Fig 1) excluded three studies that did not meet the eligibility criteria listed above: one did not use or report an isogenic permissive control [47]; a second did not report sample sizes and used a post-training interval of 15 minutes [20], i.e. 10 minutes later than the original criterion and 12 minutes later than other studies included; a third used pharmacogenetic temporal control of *rut* restoration [48].

Data item extraction

Two investigators (TY and JMW) extracted data independently using the measuring tool in Adobe Acrobat Pro; any discrepancies between the two extractions were resolved collaboratively. The following data were collected from each of the included experiments: author, year of publication, figure and panel numbers, genotype, mean Performance Index (PI) [49] with corresponding SEMs and the number of experimental iterations (N) for each mean PI value for each intervention and its related control group. To calculate STM percentages we identified a non-intervention control for each experiment, using the control that was the most similar to the experimental animals. For the *rut* restorations the closest available controls ranged from otherwise isogenic rut^{+} siblings to generic wild type (e.g. Canton-S). For the shi^{ts} experiments, including the heat-effect experiments, we used the permissive temperature controls. We also extracted experimental conditions: time delay between training and testing, odor pair, temperature, voltage, current type and relative humidity. One study's *rut* restoration data were plotted with superimposed error bars, precluding their extraction and inclusion in the review [17].

Driver line classification

Driver lines were classified by lobe expression pattern according to the original studies themselves, except for the MB247 line, which was thought to drive expression in all lobes [13], but is

now characterized as primarily driving expression in the αβ and γ lobes [18,33]. In addition, while several studies used 201Y as a γ driver, there is more recent evidence that 201Y also drives in a minority of αβ cells [33]; we accommodated this by doing primary analysis counting 201Y as γ, but also doing a variation in which it was counted as αβ + γ.

Summary measures

For each experiment we calculated the intervention's effect as a percentage change relative to the control PI. All the meta-analyses were carried out for the percentage change metric as well as the raw change in PI; the results were equivalent. We chose to report data as percentage changes for easier interpretation. The histogram in Fig 1B shows that control PI scores vary considerably across experiments; using a percentage change re-scales the phenotypes to each experiment's wild type memory. A percentage not only reports how far a phenotype is from wild type memory but also sets a lower bound (0% memory). The standard error of each percentage change was calculated using the delta method approximation [50,51].

$$SE_{pooled} = \frac{mean_{experimental}}{mean_{control}} \sqrt{+(\frac{SE_{experimental}}{mean_{experimental}})^2 + (\frac{SE_{control}}{mean_{control}})^2}$$

Synthesis of results

Review Manager software, freely available at http://tech.cochrane.org/revman, was used to perform meta-analyses [52]. Nine meta-analyses were performed: six on the *rutabaga* data, three on the *shibire* data. One random effects model meta-analysis was carried out for each mushroom body lobe and any available combinations; within each meta-analysis a subgroup analysis was performed for each driver line, except for the *rut* mutant and heat effect controls analyses, where genotype subgroups were used. Table 1 gives full details. No meta-analysis was possible for *rut* restoration to the γ lobes as only one published experiment was found. Subgroup analysis of the driver lines was pre-specified. The I^2 statistic was used as a measure of the percentage contribution of heterogeneity to the total variance in each meta-analysis, including subgroup heterogeneity [53]. For ease of interpretation, summary plots showed learning as a percentage of wild type learning; these were calculated by addition of the impairment effect size to 100%. We report p-values from a two-sample t-test with unequal group variances in the *rut* and *shi* summary plots, and from a t-distribution transformation for the cell count regression. Otherwise, percentage effect sizes and their 95% confidence intervals were used to interpret all results [23]. All 95% confidence intervals are given in the form: [95CI lower, upper].

Meta-regression approach

Driver cell count data were extracted from a single anatomical study [33]. Initial examination of the relationship was done with MATLAB's simple linear regression function (LinearModel.fit.m) on the mean values. However, this method does not account for many important aspects of the data. To accommodate the complex nature of the data, we performed multivariate hierarchical weighted meta-regression analyses of the driver effects using generalized linear mixed models (GLMM) in SAS version 9.3 software (SAS Institute, Cary, North Carolina; PROC GLIMMIX). For experiment k with appropriate control group j in study i, the outcome PI_{ijk} (raw change or relative percentage change) was modeled using GLMM taking into account the following:

- *The meta-analytic nature of the data*: each PI_{ijk} was estimated with a certain level of precision in the primary study/experiment. PI_{ij} were weighted in the GLMM by their corresponding

precision or inverse variance ($1 / Var(PI_{ijk})$) with more weight assigned to more precise PI_{ijk}, as in the meta-analyses.

- *Relevant experimental design factors* X_{ij} were corrected for in the GLMM to reduce the variance induced by differences in design factors between individual experiments and studies. We developed univariate and multivariate GLMM models by including one and more-than-one design factors as independent variables in the GLMM respectively.
- *Clustering*: multiple experiments are clustered (nested) within each study and this clustering may introduce extra variability or dependence due to laboratory and personnel preferences (practice) in conducting experiments. Studies were modeled as clusters (b_i) through a random effect with variance τ. The cluster term in the model accounts for the correlation introduced by data produced by the same laboratory.
- *Shared Controls*: *rut* restorations within experiments were calculated based on a shared control, which created dependencies (correlation) between *rut* restoration effects that shared control groups. Therefore residuals (ε_{ijk}) based on the same (shared) controls were correlated and residuals based on different controls were independent. Due to convergence issues arising from a paucity of data we assumed a constant correlation (ρ) between residuals based on the same shared controls and modeled the residual variance-covariance matrix (Σ) with a block compound symmetry structure–blocked by shared controls, leading to conditionally independent residuals. A simple constant-variance diagonal variance-covariance matrix was used for the *shi* experiments, as matched controls were available, leading to independent residuals.

Coupling all these aspects together yielded the following univariate and multivariate weighted GLMM:

$$PI_{ijk} = \sum_{i,j} \beta_{ij} X_{ij} + b_i + \varepsilon_{ijk},$$

$$b_i \sim N(0, \tau^2),$$

$$\varepsilon_{ijk} \sim N(0, \Sigma), \text{ where } Corr(\varepsilon_{ijk}, \varepsilon_{ijk'}) \neq \rho,\ Corr(\varepsilon_{ijk}, \varepsilon_{ij'k}) = 0, \text{ and } Corr(\varepsilon_{ijk}, \varepsilon_{i'jk}) = 0.$$

Construction of models

Model construction started with inspection of all the available independent variables based on univariate GLMM. From Table 1, these variables included which pair of odors was used ('ODOR PAIR'), experimental temperature ('TEMPERATURE), delay time between testing and training ('TIME'), shock voltage ('VOLTAGE'), voltage type ('AC/DC') and relative humidity ('RH'). It was noted that the ODOR PAIR variable consisted of numerous categories, which would dramatically increase the degrees of freedom, so we considered replacing this with an approximation of the variable instead. Since benzaldehyde is known to stimulate gustatory receptors as well as olfactory receptors (and thus might have a different dependency on mushroom body function from other odorants), we used the presence or absence of benzaldehyde ('BENZALDEHYDE') as a proxy for ODOR PAIR. Of these variables, RH, AC/DC and VOLTAGE were both censored in a large proportion of experiments, and (for non-censored experiments) had mainly trivial and non-statistical effects on learning; these variables were excluded from subsequent models. TIME and BENZALDEHYDE data were available for all experiments. For *rut* experiments, both variables showed substantial and statistical influences on learning (TIME generalized-R^2 = 0.26 [95CI 0.15, 0.36]; BENZALDEHYDE

generalized-R^2 = 0.28 [95CI 0.17, 0.39]), so these were incorporated into further multivariate meta-regression models. For the *shi* experiments, only TIME had a substantial influence on learning outcome (TIME generalized-R^2 = 0.12 [95CI 0.04, 0.21]). Multivariate GLMM were used to account for and extract the effect of the relevant independent variables by obtaining residuals from the respective multivariate GLMM. We calculated a residual learning effect by summarizing the residuals by drivers and rescaling them by subtracting the wild type memory reference value (*shi* = 83%; *rut* = 40%). The residual learning effect was regressed against cell counts in a linear meta-regression that was weighted by sample size (the number of experiments contributing to each driver). The learning-per-cell model was built by first dividing each driver's effect (and standard error) by its cell counts, and then fitting a multivariate GLMM with lobe categories as the main independent variable, while adjusting for other relevant experimental design factors.

Supporting Information

S1 Dataset. A Cochrane Review Manager meta-analysis file shows the data and calculations performed to produce the forest plots.
(RM5)

S2 Dataset. An Excel spreadsheet contains the data that were used in the construction of the meta-regression model.
(XLSX)

Acknowledgments

We thank Jonathan Flint, Leslie Griffith, Ajay Mathuru, Joanne Yew, Gero Miesenböck, Scott Waddell, Daniel Stettler and members of the Claridge-Chang Lab for their helpful comments on earlier versions. We also wish to thank Lucy Robinson of Insight Editing London for assistance in manuscript preparation.

Author Contributions

Analyzed the data: TY JMW FM ACC PNA ESYC. Wrote the paper: TY JMW ACC PNA ESYC.

References

1. Keene AC, Waddell S. *Drosophila* olfactory memory: single genes to complex neural circuits. Nat Rev Neurosci. 2007; 8: 341–354. doi: 10.1038/nrn2098 PMID: 17453015
2. Busto GU, Cervantes-Sandoval I, Davis RL. Olfactory learning in Drosophila. Physiology (Bethesda, Md). 2010; 25: 338–346. doi: 10.1152/physiol.00026.2010
3. Kahsai L, Zars T. Learning and memory in *Drosophila*: behavior, genetics, and neural systems. Int Rev Neurobiol. 2011; 99: 139–167. doi: 10.1016/B978-0-12-387003-2.00006–9 PMID: 21906539
4. Davis RL. Traces of *Drosophila* memory. Neuron. 2011; 70: 8–19. doi: 10.1016/j.neuron.2011.03.012 PMID: 21482352
5. Perisse E, Burke C, Huetteroth W, Waddell S. Shocking revelations and saccharin sweetness in the study of *Drosophila* olfactory memory. Curr Biol. 2013; 23: R752–63. doi: 10.1016/j.cub.2013.07.060 PMID: 24028959
6. Zars T, Fischer M, Schulz R, Heisenberg M. Localization of a short-term memory in *Drosophila*. Science. 2000; 288: 672–675. PMID: 10784450
7. Akalal D- BG, Wilson CF, Zong L, Tanaka NK, Ito K, Davis RL. Roles for *Drosophila* mushroom body neurons in olfactory learning and memory. Learning & Memory. 2006; 13: 659–668. doi: 10.1101/lm.221206

8. Blum AL, Li W, Cressy M, Dubnau J. Short- and long-term memory in *Drosophila* require cAMP signaling in distinct neuron types. Curr Biol. 2009; 19: 1341–1350. doi: 10.1016/j.cub.2009.07.016 PMID: 19646879
9. Heisenberg M. Mushroom body memoir: from maps to models. Nat Rev Neurosci. 2003; 4: 266–275. doi: 10.1038/nrn1074 PMID: 12671643
10. Levin LR, Han PL, Hwang PM, Feinstein PG, Davis RL, Reed RR. The *Drosophila* learning and memory gene rutabaga encodes a Ca2+/Calmodulin-responsive adenylyl cyclase. Cell. 1992; 68: 479–489. PMID: 1739965
11. Kitamoto T. Conditional modification of behavior in *Drosophila* by targeted expression of a temperature-sensitive shibire allele in defined neurons. J Neurobiol. 2001; 47: 81–92. PMID: 11291099
12. Dubnau J, Grady L, Kitamoto T, Tully T. Disruption of neurotransmission in *Drosophila* mushroom body blocks retrieval but not acquisition of memory. Nature. 2001; 411: 476–480. doi: 10.1038/35078077 PMID: 11373680
13. McGuire SE, Le PT, Davis RL. The role of *Drosophila* mushroom body signaling in olfactory memory. Science. 2001; 293: 1330–1333. doi: 10.1126/science.1062622 PMID: 11397912
14. McGuire SE, Le PT, Osborn AJ, Matsumoto K, Davis RL. Spatiotemporal rescue of memory dysfunction in *Drosophila*. Science. 2003; 302: 1765–1768. doi: 10.1126/science.1089035 PMID: 14657498
15. Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 1993; 118: 401–415. PMID: 8223268
16. Crittenden JR, Skoulakis EM, Han KA, Kalderon D, Davis RL. Tripartite mushroom body architecture revealed by antigenic markers. Learn Mem. 1998; 5: 38–51. PMID: 10454371
17. Schwaerzel M, Heisenberg M, Zars T. Extinction antagonizes olfactory memory at the subcellular level. Neuron. 2002; 35: 951–960. PMID: 12372288
18. Krashes MJ, Keene AC, Leung B, Armstrong JD, Waddell S. Sequential use of mushroom body neuron subsets during *Drosophila* odor memory processing. Neuron. 2007; 53: 103–115. doi: 10.1016/j.neuron.2006.11.021 PMID: 17196534
19. Weislogel J- M, Bengtson CP, Muller MK, Hortzsch JN, Bujard M, Schuster CM, et al. Requirement for nuclear calcium signaling in *Drosophila* long-term memory. Sci Signal. 2013; 6: ra33. doi: 10.1126/scisignal.2003598 PMID: 23652205
20. Cervantes-Sandoval I, Martin-Pena A, Berry JA, Davis RL. System-like consolidation of olfactory memories in *Drosophila*. J Neurosci. 2013; 33: 9846–9854. doi: 10.1523/JNEUROSCI.0451-13.2013 PMID: 23739981
21. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013; 14: 365–376. doi: 10.1038/nrn3475 PMID: 23571845
22. Cohen J. The earth is round (p < .05). American Psychologist. 1994; 49: 997–1004.
23. Cumming G. Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. Multivariate Applications Series. 2012.
24. Halsey LG, Curran-Everett D, Vowler SL, Drummond GB. The fickle P value generates irreproducible results. Nat Methods. 2015; 12: 179–185. doi: 10.1038/nmeth.3288 PMID: 25719825
25. Ellis PD. The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge University Press; 2010.
26. Borenstein M, Hedges LV, Higgins J, Rothstein HR. Introduction to meta-analysis. John Wiley & Sons, Ltd; 2011.
27. Sena ES, van der Worp HB, Bath PMW, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS Biol. 2010; 8: e1000344. doi: 10.1371/journal.pbio.1000344 PMID: 20361022
28. Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, et al. Meta-analysis of data from animal studies: a practical guide. J Neurosci Methods. 2014; 221: 92–102. doi: 10.1016/j.jneumeth.2013.09.010 PMID: 24099992
29. Tully T, Quinn W. Classical conditioning and retention in normal and mutant *Drosophila* melanogaster. J Comp Physiol [A]. 1985; 157: 263–277.
30. Schwaerzel M, Monastirioti M, Scholz H, Friggi-Grelin F, Birman S, Heisenberg M. Dopamine and octopamine differentiate between aversive and appetitive olfactory memories in *Drosophila*. J Neurosci. 2003; 23: 10495–10502. PMID: 14627633
31. Thum AS, Jenett A, Ito K, Heisenberg M, Tanimoto H. Multiple memory traces for olfactory reward learning in *Drosophila*. J Neurosci. 2007; 27: 11132–11138. doi: 10.1523/JNEUROSCI.2712-07.2007 PMID: 17928455

32. Scheunemann L, Jost E, Richlitzki A, Day JP, Sebastian S, Thum AS, et al. Consolidated and labile odor memory are separately encoded within the *Drosophila* brain. J Neurosci. 2012; 32: 17163–17171. doi: 10.1523/JNEUROSCI.3286-12.2012 PMID: 23197709

33. Aso Y, Grübel K, Busch S, Friedrich AB, Siwanowicz I, Tanimoto H. The mushroom body of adult *Drosophila* characterized by GAL4 drivers. J Neurogenet. 2009; 23: 156–172. doi: 10.1080/01677060802471718 PMID: 19140035

34. Murthy M, Fiete I, Laurent G. Testing odor response stereotypy in the *Drosophila* mushroom body. Neuron. 2008; 59: 1009–1023. doi: 10.1016/j.neuron.2008.07.040 PMID: 18817738

35. Caron SJC, Ruta V, Abbott LF, Axel R. Random convergence of olfactory inputs in the *Drosophila* mushroom body. Nature. 2013; 497: 113–117. doi: 10.1038/nature12063 PMID: 23615618

36. Reducing our irreproducibility. Nature. 2013; 496.

37. Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Med. Public Library of Science; 2005; 2: e124. doi: 10.1371/journal.pmed.0020124 PMID: 16060722

38. Ioannidis JPA. Why science is not necessarily self-correcting. Perspectives on Psychological Science. 2012; 7: 645–654. Available: http://pps.sagepub.com/content/7/6/645.full doi: 10.1177/1745691612464056 PMID: 26168125

39. Morrison DE, Henkel RE, editors. The Significance Test Controversy. Chicago: Transaction Publishers; 1970.

40. Hentschke H, Stüttgen MC. Computation of measures of effect size for neuroscience data sets. Eur J Neurosci. 2011; 34: 1887–1894. doi: 10.1111/j.1460-9568.2011.07902.x PMID: 22082031

41. Altman DG, Machin D, Bryant TN, Gardner MJ, editors. Statistics with confidence: confidence intervals and statistical guidelines. 2nd ed. BMJ Books; 2000.

42. Crabbe JC, Wahlsten D, Dudek BC. Genetics of mouse behavior: interactions with laboratory environment. Science. 1999; 284: 1670–1672. PMID: 10356397

43. Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, Tuttle AH, et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. Nat Methods. 2014; 11: 629–632. doi: 10.1038/nmeth.2935 PMID: 24776635

44. Richter SH, Garner JP, Würbel H. Environmental standardization: cure or cause of poor reproducibility in animal experiments? Nat Methods. 2009; 6: 257–261. doi: 10.1038/nmeth.1312 PMID: 19333241

45. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. 2009. p. e1000100. doi: 10.1371/journal.pmed.1000100 PMID: 19621070

46. Narayanan R, Ramaswami M. Endocytosis in *Drosophila*: progress, possibilities, prognostications. Exp Cell Res. 2001; 271: 28–35. doi: 10.1006/excr.2001.5370 PMID: 11697879

47. Thum AS, Knapek S, Rister J, Dierichs-Schmitt E, Heisenberg M, Tanimoto H. Differential potencies of effector genes in adult *Drosophila*. J Comp Neurol. 2006; 498: 194–203. doi: 10.1002/cne.21022 PMID: 16856137

48. Mao Z, Roman G, Zong L, Davis RL. Pharmacogenetic rescue in time and space of the *rutabaga* memory impairment by using Gene-Switch. Proc Natl Acad Sci USA. 2004; 101: 198–203. doi: 10.1073/pnas.0306128101 PMID: 14684832

49. Quinn W, Harris W, Benzer S. Conditioned behavior in Drosophila melanogaster. Proc Natl Acad Sci U S A. 1974; 71: 708–712. PMID: 4207071

50. Oehlert GW. A Note on the Delta Method. The American Statistician. 1992; 46: 27–29.

51. Cramer H. Mathematical Models of Statistics. Princeton NJ: Princeton University Press; 1946.

52. The Nordic Cochrane Center. Review Manager (RevMan). 5 ed. Copenhagen: The Cochrane Collaboration; 2012.

53. Higgins J, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ: British Medical Journal. 2003; 327: 557–560. PMID: 12958120