

**Supplementary Table S1.** Synopsis of strain effects across labs and housing conditions.

	4-way ANOVA <sup>a</sup> all cases n=432 pooled data				Fisher's PLSD <sup>b</sup> enriched, n=216 pooled (split by lab)			Fisher's PLSD standard, n=216 pooled (split by lab)			Fisher's PLSD <sup>d</sup> all cases, n=432 enriched versus standard			Fisher's PLSD enriched, n=216 lab comparison		Fisher's PLSD standard, n=216 lab comparison	
	strain main effect		strain x enrichm.	strain x lab	strain ranking <sup>c</sup>			strain ranking			reliability of strain ranking			reliability of strain ranking		reliability of strain ranking	
	p<	ω <sup>2</sup>	p<	p<	F1 -B6	F1 -D2	B6 -D2	F1 -B6	F1 -D2	B6 -D2	missed <sup>e</sup> Δ enriched	missed <sup>f</sup> Δ standard	rank conflicts <sup>g</sup>	missed <sup>h</sup> Δ	rank conflicts <sup>i</sup>	missed Δ	rank conflicts
<i>O-maze</i>																	
% open arm entries	.0001	.152	.0127	.0029	↓(↓↓↓)	↓(↓-↓)	↑(↑↑-)	↓(-↓-)	↓(---)	↑(-↑-)	0/3	0/3	0/3	2/9	0/9	4/9	0/9
total head dips	.0001	.314	.0128	.0001	↓(↓↓↓)	↓(-↓-)	↑(↑↑↑)	↓(↓↓↓)	↑(-↑-)	↑(↑↑↑)	1/3	0/3	0/3	1/9	0/9	2/9	0/9
% protected head dips	.0014	.026	.0415	ns	↑(↑↑↑)	↑(↑↑-)	-(---)	-(-↑-)	-(-↑-)	-(---)	0/3	2/3	0/3	1/9	0/9	4/9	0/9
fecal bolus count	.0001	.259	.0001	.0108	↑(↑↑↑)	↑(-↑-)	↓(-↓↓)	↑(↑↑↑)	↑(↑↑↑)	↓(-↓↓)	1/3	0/3	0/3	2/9	1/9	1/9	0/9
path traveled	.0001	.207	.0006	ns	↓(↓↓↓)	↓(↓↓↓)	-(-↑-)	↓(↓↓↓)	↓(↓↓↓)	↓(↓-↓)	1/3	0/3	0/3	2/9	0/9	1/9	0/9
<i>open-field</i>																	
center avoidance (1st 10min)	.0001	.350	ns	.0001	↑(↑↑↑)	↑(-↑-)	↓(↓↓↓)	↑(↑↑↑)	↓(-↓-)	↓(↓↓↓)	1/3	0/3	0/3	1/9	1/9	2/9	0/9
center avoidance (habituation)	.0009	.028	ns	ns	-(---)	↓(---)	↓(---)	-(-↑-)	-(---)	-(-↓-)	0/3	2/3	0/3	0/9	0/9	4/9	0/9
path traveled (1st 10min)	.0001	.307	ns	.0002	↓(↓-↓)	↓(↓-↓)	-(-↑-)	↓(↓↓↓)	↓(↓↓↓)	-(-↑-)	0/3	0/3	0/3	4/9	0/9	2/9	0/9
path traveled (habituation)	.0001	.226	ns	.0106	↓(↓-↓)	↓(↓-↓)	↑(↑↑-)	↓(↓↓↓)	↓(↓-↓)	↑(↑↑↑)	0/3	0/3	0/3	3/9	0/9	1/9	0/9
% time running/walking	.0001	.180	ns	.0001	↓(↓-↓)	↓(↓-↓)	-(---)	↓(↓↓↓)	↓(↓-↓)	-(-↑)	0/3	0/3	0/3	2/9	0/9	2/9	1/9
<i>object exploration</i>																	
horizontal object exploration	.0001	.388	.0470	.0014	↑(↑↑↑)	↓(↓-↓)	↓(↓↓↓)	↑(↑↑↑)	↓(↓↓↓)	↓(↓↓↓)	0/3	0/3	0/3	1/9	0/9	0/9	0/9
vertical object exploration	.0001	.410	ns	.0001	↑(↑↑↑)	↓(↓-↓)	↓(↓↓↓)	↑(-↑↑)	↓(↓↓↓)	↓(↓↓↓)	0/3	0/3	0/3	1/9	0/9	1/9	0/9
% path in object zone	.0001	.333	.0118	.0097	↑(↑↑↑)	↓(↓-↓)	↓(↓↓↓)	↑(-↑↑)	↓(↓↓↓)	↓(↓↓↓)	0/3	0/3	0/3	1/9	0/9	1/9	0/9
object exploration distance	.0001	.358	ns	.0123	↓(↓↓↓)	↑(↑↑-)	↑(↑↑↑)	↓(-↓↓)	↑(↑↑↑)	↑(↑↑↑)	0/3	0/3	0/3	1/9	0/9	1/9	0/9
corner distance	.0001	.499	ns	.0001	↑(-↑↑)	↓(↓↓↓)	↓(↓↓↓)	↑(↑↑↑)	↓(↓↓↓)	↓(↓↓↓)	0/3	0/3	0/3	1/9	0/9	0/9	0/9
<i>water-maze</i>																	
swim path length	.0001	.147	.0129	ns	↓(↓↓↓)	↓(↓↓↓)	↑(-↑-)	↓(↓-↓)	↓(↓↓↓)	-(-↓)	0/3	1/3	0/3	2/9	0/9	3/9	0/9
% time near wall	.0001	.132	ns	.0001	↓(↓-↓)	↓(-↓)	↑(↑↑-)	↓(↓-↓)	↓(↓-↓)	↑(↑↑-)	0/3	0/3	0/3	3/9	0/9	4/9	0/9
average swim speed	.0001	.164	.0097	.0094	↑(↑↑↑)	↑(↑↑↑)	↑(↑↑↑)	↑(↑↑-)	↑(↑↑-)	-(---)	0/3	1/3	0/3	0/9	0/9	2/9	0/9
probe: annulus crossings	.0055	.020	ns	ns	-(---)	-(---)	-(---)	↑(---)	↑(↑↑-)	↑(↑↑-)	3/3	0/3	0/3	0/9	0/9	4/9	0/9
probe: target proximity	.0084	.018	ns	ns	-(---)	-(-↓)	-(-↓)	-(---)	-(-↓)	-(-↓)	0/3	0/3	0/3	4/9	0/9	4/9	0/9
<i>avearge</i>											11.7%	10.0%	0.0%	17.8%	1.1%	23.9%	0.6%

<sup>a</sup> between subject factors strain x housing x lab x replicate (only effects of interest are shown); first column: type-1 error p-values; second column for main effect: effect size as partial omega squared, the proportion of variance accounted for by strain if only this factor were in the design (range 0 to 1.0)

<sup>b</sup> post hoc analysis of strain factor in partial ANOVA model with between subject factors strain x lab x replicate, or strain x replicate, respectively

<sup>c</sup> established by pair-wise comparison of strains with data from all labs pooled or kept separately (arrows within parentheses: Lipp, Nitsch, Würbel lab), arrow up if strain listed at the top earned higher scores

<sup>d</sup> comparison of post hoc analyses of strain factor in enriched versus standard housing conditions

<sup>e</sup> failure under enriched housing conditions to differentiate two strains that were significantly different under standard housing conditions (in 3 strain combinations evaluated)

<sup>f</sup> failure under standard housing conditions to differentiate two strains that were significantly different under enriched housing conditions (in 3 strain combinations evaluated)

<sup>g</sup> strains showing significant but opposing differences under standard and enriched housing conditions (in 3 strain combinations evaluated)

<sup>h</sup> failure in one lab to differentiate two strains that were significantly different in another lab under the same housing conditions (in 3 strain combinations x 3 labs evaluated)

<sup>i</sup> strains showing significant but opposing differences in two labs under the same housing conditions (in 3 strain combinations x 3 labs evaluated)