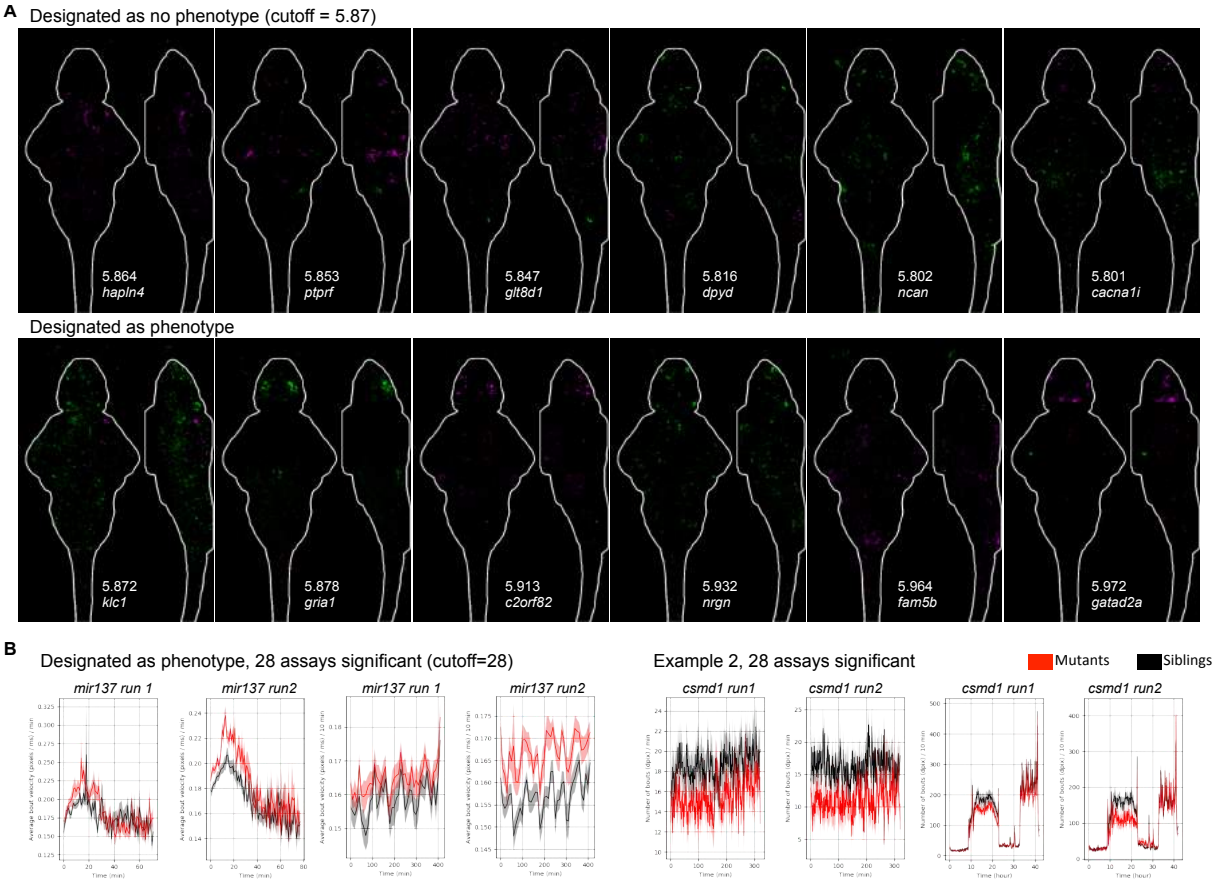

Figure S1. Related to Figure 1. Determination of cutoff for phenotype designations. (**A**) Determination of cutoff for brain activity signals. Images are sum-of-slices intensity projections (Z- and X- axis). This cutoff for the $\log_{10}$(sum of pixels) was made so that mutants with small, specific, repeatable signals were classified as having phenotypes (*gria1,* forebrain). Full stacks of all individual repeats are available on stackjoint.com/zbrain. Many mutants with diffuse and sparse signals fell below the cutoff. Exceptions included *klc1* and *nrgn* (shown here), which both had signals that reproduced in independent experiments (stackjoint.com/zbrain) and a high enough pixel count. (**B**) Determination of cutoff for number of significant behavior assays. Avoiding false positive phenotypes required determining a cutoff for the number of assays that must be significant for a mutant to be defined as having a phenotype. A total of 71 assays were analyzed (Figure S2, STAR Methods). Using randomized wild-type larvae, we established an analysis approach with an approximately 10% rate of false positives (Figure S2) for any comparison (i.e. heterozygous versus homozygous larvae). A total of 71 assays were performed; therefore 7 of those 71 assays could potentially be false positives for any comparison. We made the cutoff for designating phenotypes to more than three times this value, because multiple comparisons were performed for duplicated genes and when two heterozygous parents were crossed (i.e. homozygous larvae were compared both to heterozygous and to wild type). This stringent cutoff may result in false negatives, but it was chosen empirically to define mutants with specific and repeatable differences as having a phenotype. The mutant *mir137* was chosen as the lower end of the cutoff. This mutant has a repeatable increased velocity of motion, shown here for two time windows: heat stressor and dark flashes. A second example of a mutant at the cutoff is *csmd1*, which has repeatable differences in frequency of movement (shown here with day taps time window and entire assay). Data is expressed as mean ± s.e.m..

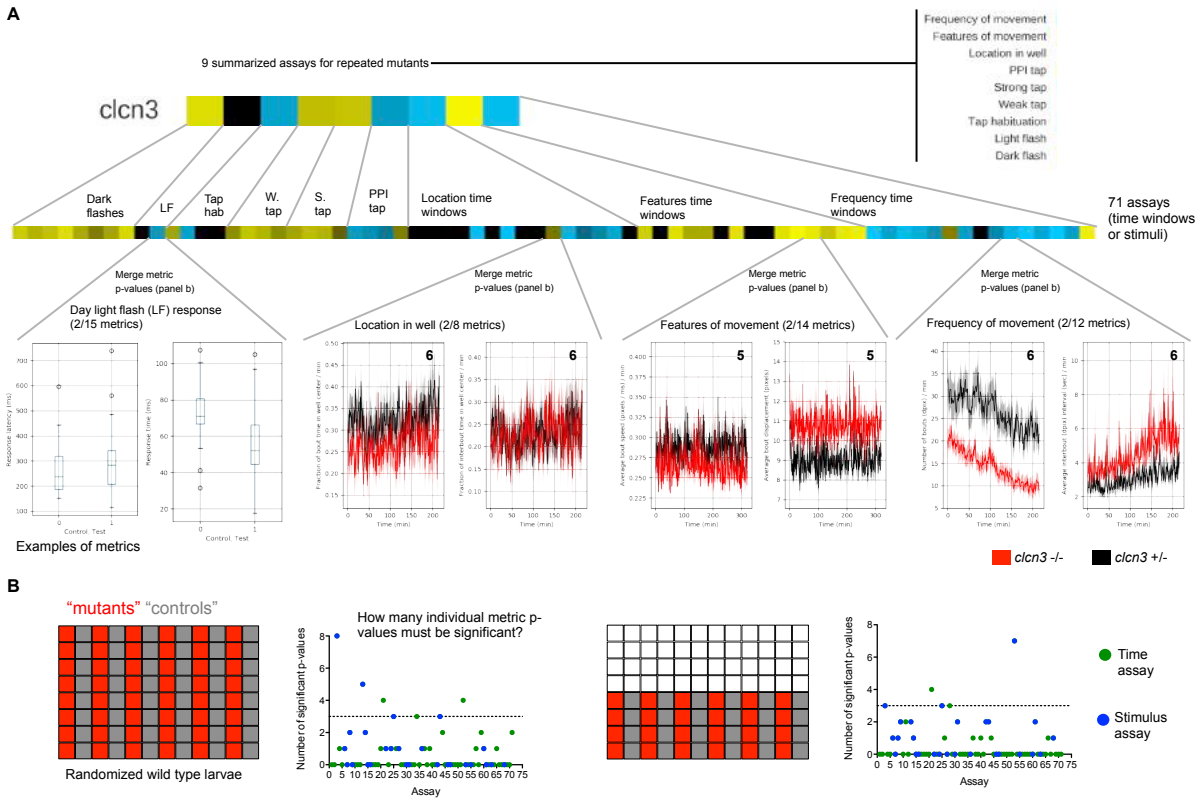

Figure S2. Related to Figure 2 and STAR Methods Quantification and Analysis. Pipeline for processing behavior data. a) The final reproducible p-values from 46 mutants (Figure 2F) determined to have a phenotype are based on two independent runs. If two runs had a significant p-value in the same time window or same stimulus assay of the 71 tested assays, then the phenotype was considered reproducible. The 71 assays (described in detail in the STAR Methods) included Frequency of movement, Features of movement, and Location in the well for each of the 14 time sections and full 2-day assay, for a total of 45 baseline assays. The remaining 26 assays were stimulus-driven: prepulse, weak, and strong taps during the night and day (12 assays), tap habituation (4 assays, 3 day and one night), light flashes (2 assays, one day and one night), and dark flashes (8 assays; 4 blocks, beginning and end ten flashes analyzed separately). Each of the p-values in the 71 assays is based on multiple metrics, which are merged to determine significance. Two examples of each type of metric are shown, and all metrics are described in the STAR Methods. **b**) We determined how many individual metrics in an assay must be significant for the assay (of the 71 assays) to be considered significant by comparing randomized groups of wild-type larvae. Three separate batches of wild-type larvae were put through the standard behavior experimental protocol and analysis. Three comparisons were conducted for each of the three batches (two comparisons with a total of 48 larvae and one with 96). We wanted a false-discovery rate of less than 10%, where 7 or less of the 71 assays would be significant false-positives. Based on all nine wild-type larvae comparisons, we determined that if more than three (grey dashed line) individual p-value metrics were significant, then less than 10% of 71 assays would be false positives. The merged p-values for stimulus response (blue points) included many more metrics than Frequency, Features, and Location (green points) making them more prone to false positive outcomes. These assays represent close to half of the false positives, while they make up 1/3 of the assays (26/71). To further minimize such false

positives in the stimulus response assays, the threshold was set to > 4 with the exception that it would be reduced to > 2 in the case that the exact same metric (e.g. velocity) were significant across two related assays (e.g. light flashes at night and during the day). This reduction in the cutoff was important, as highly specific phenotypes, such as change in only the latency of dark flash response, were designated as non-significant when the cutoff value was set to 4.

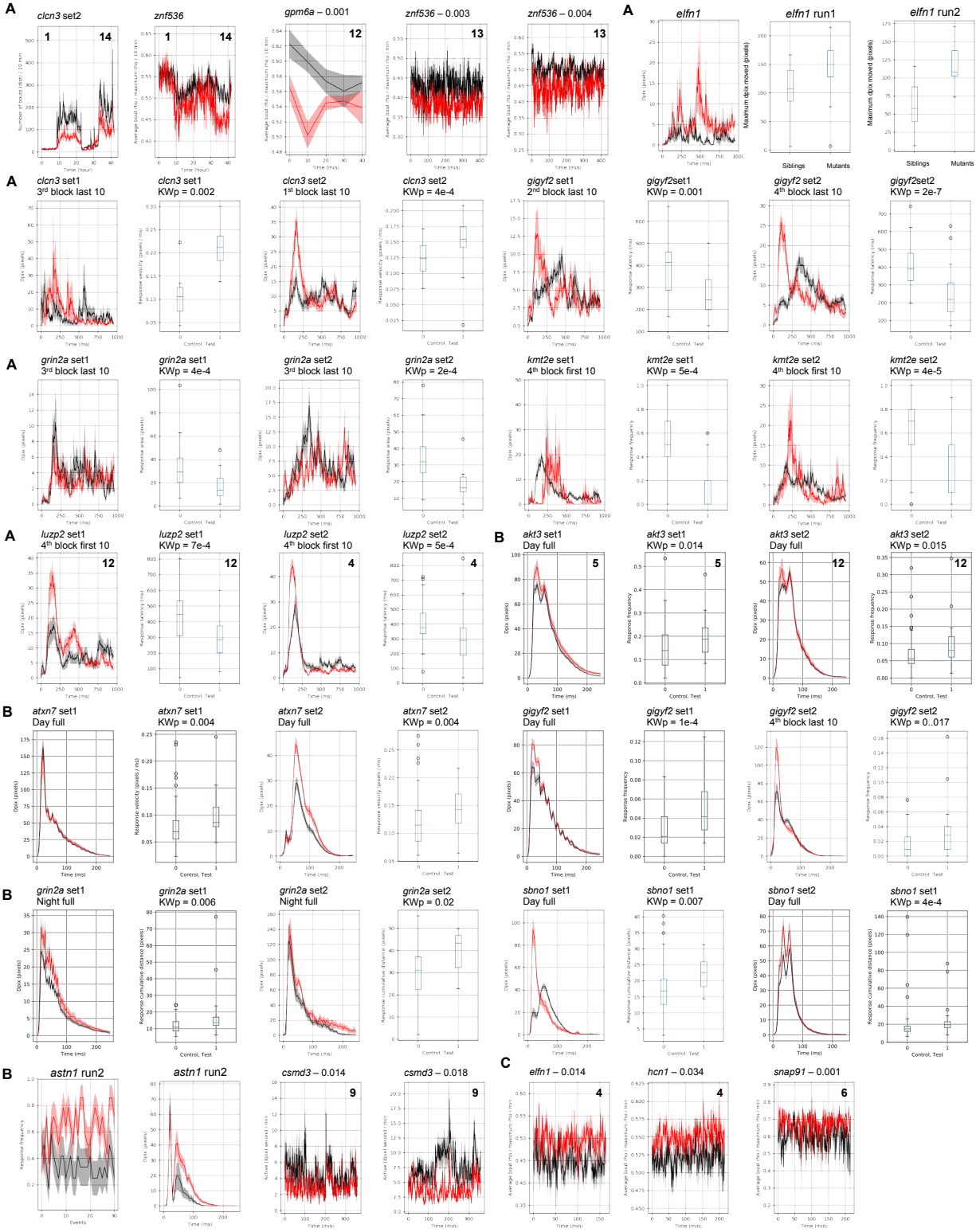
A
clcn3 set2
znf536
gpm6a – 0.001
znf536 – 0.003
znf536 – 0.004
A
elfn1
elfn1 run1
elfn1 run2
Siblings
Mutants
A
clcn3 set1
3rd block last 10
clcn3 set1
KWp = 0.002
clcn3 set2
1st block last 10
clcn3 set2
KWp = 4e-4
gigyf2 set1
2nd block last 10
gigyf2set1
KWp = 0.001
gigyf2 set2
4th block last 10
gigyf2set2
KWp = 2e-7
A
grin2a set1
3rd block last 10
grin2a set1
KWp = 4e-4
grin2a set2
3rd block last 10
grin2a set2
KWp = 2e-4
kmt2e set1
4th block first 10
kmt2e set1
KWp = 5e-4
kmt2e set2
4th block first 10
kmt2e set2
KWp = 4e-5
A
luzp2 set1
4th block first 10
luzp2 set2
KWp = 7e-4
luzp2 set2
4th block first 10
luzp2 set2
KWp = 5e-4
B
akt3 set1
Day full
akt3 set1
KWp = 0.014
akt3 set2
Day full
akt3 set2
KWp = 0.015
B
atxn7 set1
Day full
atxn7 set1
KWp = 0.004
atxn7 set2
Day full
atxn7 set2
KWp = 0.004
gigyf2 set1
Day full
gigyf2 set1
KWp = 1e-4
gigyf2 set2
4th block last 10
gigyf2 set2
KWp = 0..017
B
grin2a set1
Night full
grin2a set1
KWp = 0.006
grin2a set2
Night full
grin2a set2
KWp = 0.02
sbno1 set1
Day full
sbno1 set1
KWp = 0.007
sbno1 set2
Day full
sbno1 set1
KWp = 4e-4
B
astn1 run2
astn1 run2
csmd3 – 0.014
csmd3 – 0.018
C
elfn1 – 0.014
hcn1 – 0.034
snap91 – 0.001

Figure S3. Related to Figure 2; Figure 5; Figure 6. Supporting data for Figure 2, Figure 5, and Figure 6 behavior plots. (**A**) Repeats for behavior data in Figure 2. This in includes the plots for the heatmap of mutants with altered dark flash responses (Figure 2D). In Figure 2D, the significance for 8 sections of dark flash analysis were combined with fisher's method. Here, the most significant section is shown, with the p-value from Kruskal-Wallis one-way ANOVA. The movement trace for the response is the mean ± s.e.m for each mutants response trace for the ten flash events in the section. For the data that is not stimulus-driven, the p-values are from the linear mixed model. (**B**) Repeats for behavior data in Figure 5. This includes mutants with prepulse inhibition phenotypes from the heatmap in Figure 5D. Boxplots and associated Kruskal-Wallis one-way ANOVA p-values are shown for one modality of response that repeated across two assays. The movement trace for the response is the mean ± s.e.m for each mutants response trace for all prepulse tap (strong tap after the weak) events in the prepulse inhibition assay section. (**C**) Repeats for behavior data in Figure 6.

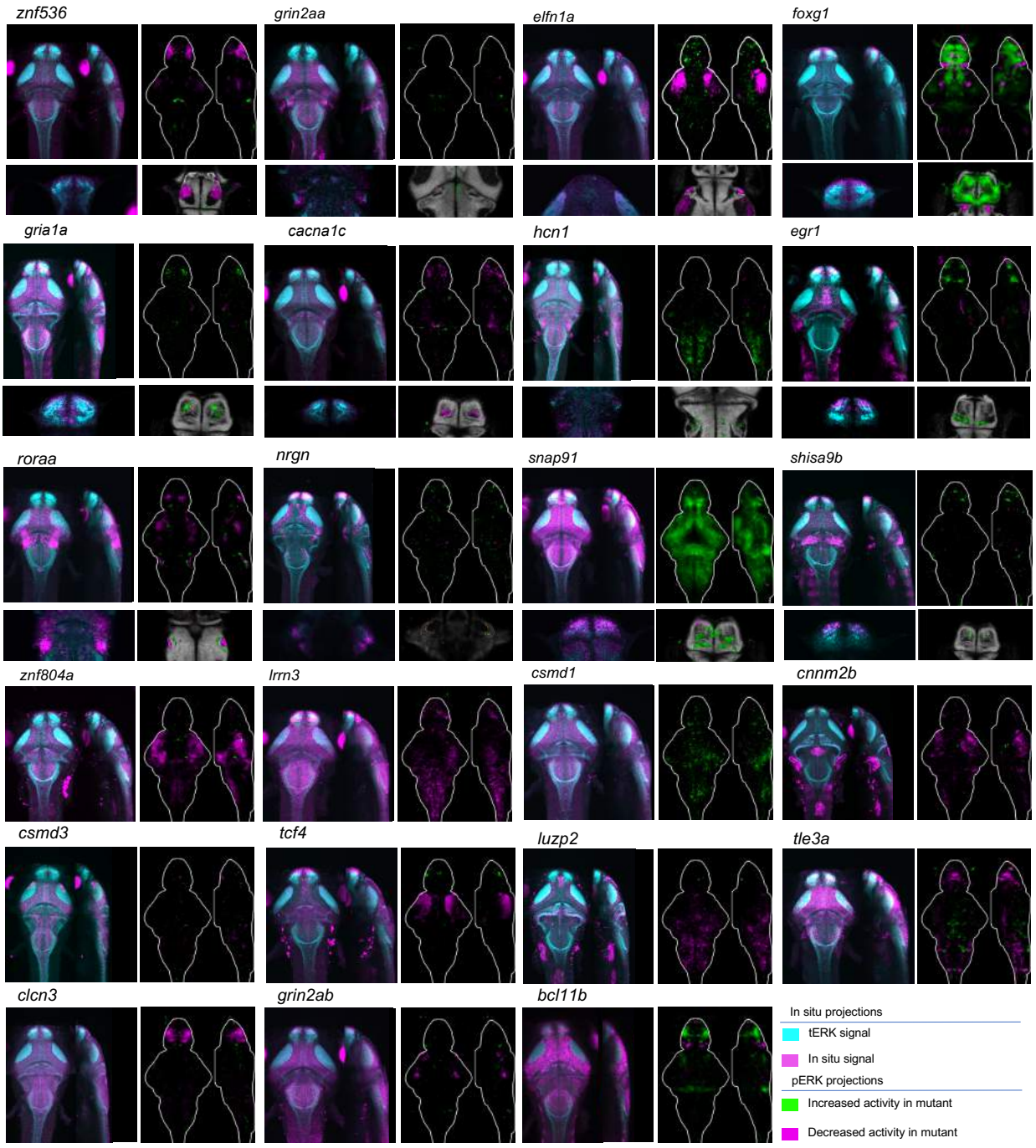

Figure S4. Related to Figure 4. Comparison between *in situ* RNA localization and brain activity. Stacks of all RNA-FISH images are available on http://stackjoint.com/basic/ and are tagged as "Thyme, 2018". Probe sequences are available on http://genepile.com/scz_gwas108. Brain area descriptions are based on the Z-Brain masks (stackjoint.com/zbrain). *znf536:* forebrain showed *in situ* and phosphorylated-Erk (pErk) brain activity signal. *grin2aa*: cerebellum showed *in situ* and activity signal. *elfn1a:* retinal arborization field AF7, tectum, and forebrain showed *in situ* and activity signal. *foxg1:* forebrain showed *in situ* and activity signal. *gria1a:* forebrain showed *in situ* and activity signal. *cacna1c:* cerebellum and forebrain showed *in situ* and activity signal. *hcn1:* hindbrain showed *in situ* and activity signal. *egr1:* forebrain

showed *in situ* and activity signal. *roraa:* tectum and hindbrain showed *in situ* and activity signal. *nrgn:* forebrain and hypothalamic cells showed *in situ* and activity signal. *snap91:* activity signal correlated to *in situ* signal showing expression throughout the brain. *shisa9b:* forebrain showed *in situ* and activity signal. *znf804a:* activity signal correlates to in situ signal showing expression throughout the brain. *lrrn3:* activity signal correlated to *in situ* signal showing expression throughout the brain. *csmd1*: activity signal correlated to *in situ* signal showing expression throughout the brain. *cnnm2b:* strong *in situ* signal in forebrain and torus semicircularis did not correlate to activity signal in tectum, retinal arborization field AF7, and hindbrain. *csmd3:* strong *in situ* signal in retinal arborization field AF7, tectum, and hindbrain did not correlate to diffuse and minimal activity signal. *tcf4:* strong *in situ* signal in forebrain and midbrain did not correlate to strong activity signal in tectum. *luzp2:* Strong *in situ* signal in tectum and cerebellum did not correlate to activity signal in hindbrain. *tle3a:* strong *in situ* signal in forebrain and tectum did not correlate to strong activity signal primarily in hindbrain. *clcn3:* strong *in situ* signal throughout the brain did not correlate to activity signal mainly in forebrain. *grin2ab:* strong *in situ* signal throughout the brain did not correlate to activity signal primarily in the tectum, retinal arborization field af7, and hindbrain. *bcl11b:* strong *in situ* signal throughout the brain did not correlate to activity signal in forebrain and cerebellum.

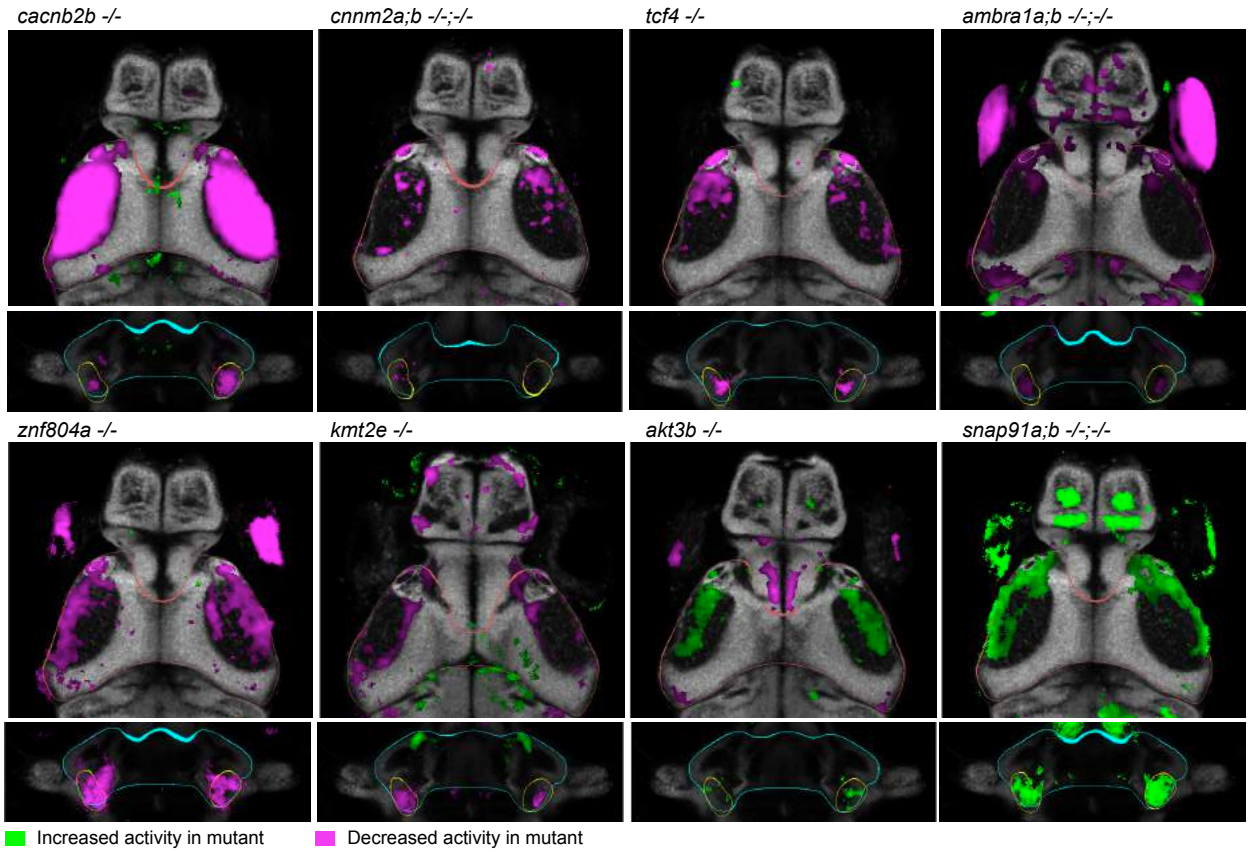

Figure S5. Related to Figure 5. Functional connections between hypothalamus and tectum revealed by phosphorylated-Erk activity mapping. Selected brain slices for the mesencephalon (orange outline) containing the tectum neuropil and retinal arborization field AF7, as well as the intermediate hypothalamus (cyan outline) and GABAergic cluster 3 sparse (yellow outline). Two genes, *tcf4* and *cnnm2*, are repeated from Figure 5C. The majority of genes with signal in the retinal arborization field AF7 and nearby tectum have the same direction of activity change and same approximate signal strength as in this small subregion of the hypothalamus.

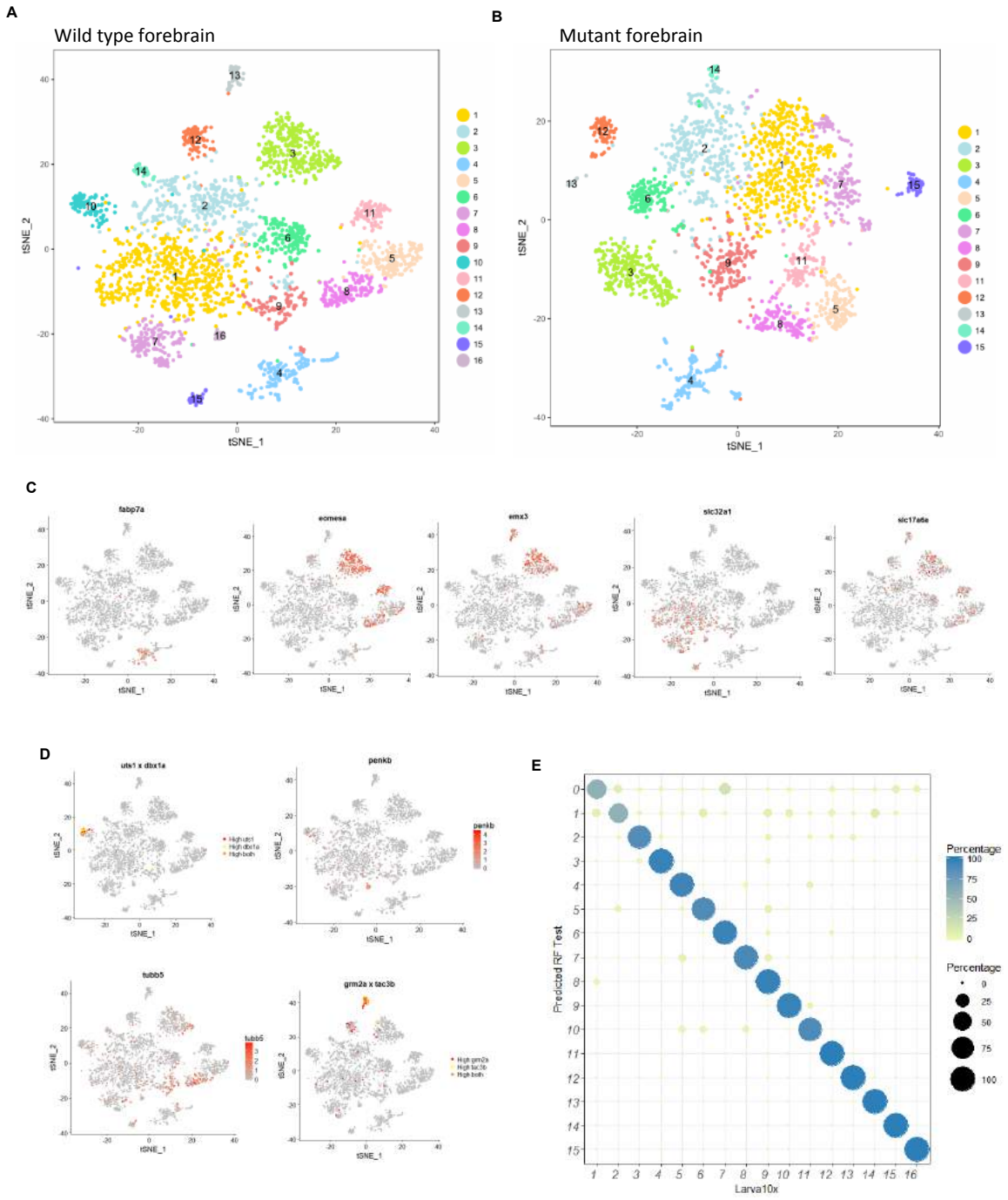

Figure S6. Related to Figure 7. Single cell Analysis of *znf536* Mutant and Wild Type forebrain cells. (**A**) 2D visualization of single-cell clusters in the wild type dataset using t-distributed stochastic neighbor embedding (t-SNE). Individual points correspond to single cells and are color coded according to their cluster membership determined by graph-based clustering. (**B**) 2D visualization of single-cell clusters in the *znf536* mutant dataset using t-distributed stochastic neighbor embedding (t-SNE). Mutant clusters are color-coded and labeled according to their correspondence to wild type forebrain clusters. (**C**) Gene expression profiles of key forebrain marker genes across the wild type dataset. (**D**) Gene expression profiles of marker genes of clusters that are 1) missing in mutant: (Cluster10: *uts1* and *dbx1a* and Cluster 16:

penkb) 2) reduced in mutant (Cluster 13: *grm2a* and *tac3b*) or 3) increased in mutant (Cluster 9: *tubb5*), a marker for immature neurons. (**E**) Performance of random forest (RF) model trained on the wild type dataset with graph clustering labels shown in Figure S6A. Training set was formed by choosing 70% of the cells from wild type dataset, with proportional representation from each cluster. The trained RF model was used to classify cells in the remaining (test) dataset. The resulting classification of the test set is shown as a confusion matrix. Cells are either assigned into their corresponding labels or left unassigned if they cannot be reliably assigned to any of the training labels.