Genetic architecture of collective behaviors in zebrafish

Wenlong Tang[a*], Guoqiang Zhang[a*], Fabrizio Serluca[a], Jingyao Li[a], Xiaorui Xiong[a], Matthew Coble[a], Tingwei Tsai[a], Zhuyun Li[a], Gregory Molind[a], Peixin Zhu[a] & Mark C. Fishman[b‡]

[a]Novartis Institutes for Biomedical Research, Chemical Biology & Therapeutics, 181 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.

[b]Harvard University, Dept. of Stem Cell and Regenerative Biology, 7 Divinity Avenue, Cambridge, MA 02138, U.S.A.

[*] These authors contributed equally to this work.

‡Corresponding author: M.C.F.: Mark_fishman@harvard.edu

Abstract Collective behaviors of groups of animals, such as schooling and shoaling of fish, are central to species survival, but genes that regulate these activities are not known. Here we parsed collective behavior of groups of adult zebrafish using computer vision and unsupervised machine learning into a set of highly reproducible, unitary, several hundred millisecond states and transitions, which together can account for the entirety of relative positions and postures of groups of fish. Using CRISPR-Cas9 we then targeted for knockout 35 genes associated with autism and schizophrenia. We found mutations in three genes had distinctive effects on the amount of time spent in the specific states or transitions between states. Mutation in *immp2l* (inner mitochondrial membrane peptidase 2-like gene) enhances states of cohesion, so increases shoaling; mutation in in the Nav1.1 sodium channel, *scn1lab*+/- causes the fish to remain scattered without evident social interaction; and mutation in the adrenergic receptor, *adra1aa-/-*, keeps fish close together and retards transitions between states, leaving fish motionless for long periods. Motor and visual functions seemed relatively well-preserved. This work shows that the behaviors of fish engaged in collective activities are built from a set of stereotypical states. Single gene mutations can alter propensities to collective actions by changing the proportion of time spent in these states or the tendency to transition between states. This provides an approach to begin dissection of the molecular pathways used to generate and guide collective actions of groups of animals.

Introduction

Darwin recognized that survival depends not only upon the attributes of individuals but also upon coordination of larger groups, such as schools of fish, flocks of birds, and armies of ants (Darwin, 1871). The benefits of collective movement include improved predator avoidance (Ioannou, Guttal, & Couzin, 2012; Ioannou, Morrell, Ruxton, & Krause, 2009), foraging success (Krause, Hartmann, & Pritchard, 1999; Torney, Berdahl, & Couzin, 2011), and reduction in energy consumption due to favorable local currents for fish and birds (Krause,

1994; Spedding, 2011). It is not known whether and how genetic changes in individuals affect group dynamics.

The schooling, shoaling, and other collective dynamics of fish have been beautifully described (Jolles, Boogert, Sridhar, Couzin, & Manica; Katz, Tunstrom, Ioannou, Huepe, & Couzin, 2011; Tunstrom et al., 2013). It appears that the sometimes rapid changes in direction and mode result from propagation of local interactions (Katz et al., 2011). The dynamics can be mathematically modelled as self-propelled particles subjected to forces such as attraction, repulsion, and alignment (Katz et al., 2011; Lukeman, Li, & Edelstein-Keshet, 2010; Sumpter, 2010). No genes regulating such forces have been identified, although QTL linkage analysis suggests that schooling at least is under genetic control (Greenwood, Wark, Yoshida, & Peichel, 2013).

Machine learning has begun to make more tractable the studies of complex behaviors (Anderson & Perona, 2014), such as *C. elegans* foraging (Greene, Brown, et al., 2016), hydra basal behaviors (Han, Taralova, Dupre, & Yuste, 2018), *Drosophila* mating and aggression (Dankert, Wang, Hoopfer, Anderson, & Perona, 2009; Kabra, Robie, Rivera-Alba, Branson, & Branson, 2013; Kravitz & Huber, 2003), and mouse social behaviors (Hong et al., 2015; Kabra et al., 2013; Wiltschko et al., 2015). Unsupervised machine learning approaches, in particular, have the power to capture unknown patterns with as few *a priori* assumptions as possible (Todd, Kain, & de Bivort, 2017), and are widely used in pattern learning, (Feng et al., 2017; Todd et al., 2017) even in cases where data is dense and not readily separable (Chuang, Tzeng, Chen, Wu, & Chen, 2006; Siew, 2013).

Here we have found that unsupervised machine learning can describe fairly completely the collective motion of groups of adult zebrafish as a set of 17 stereotypical states and transitions between states. (Although most studies of zebrafish behavior use larvae (Baier, 2000; Brockerhoff et al., 1995; Friedrich, Jacobson, & Zhu, 2010; Granato et al., 1996; Muto et al., 2005; Portugues & Engert, 2009; Rihel et al., 2010; Zhu et al., 2009), we had to use adult fish because we found robust social behaviors developed late.) We mutated by CRISPR-Cas9 35 genes linked to human disorders of social behavior, such as autism and schizophrenia, with the hope that these genes might play a conserved social behavioral role through evolution. Overall these mutations did not completely add or delete states. Three mutations in particular, however, changed overall usage of states or transitions, with dramatic and consistent effects upon the collective group dynamics.

Results

We first defined patterns of wild-type behavior. We find that the collective activity became more robust and the behavioral repertoire expanded as the animals grew (Supplemental Figure 1), so it was essential for our studies to use adult fish, at least 3 months of age. We tracked groups of adult fish by continuous video monitoring over half hour assays in an environmentally sheltered tank. As shown in the video attached to Figure 1 (Supplemental Video 1), wild-type fish tend to swim continuously and dynamically in relatively tight groups.

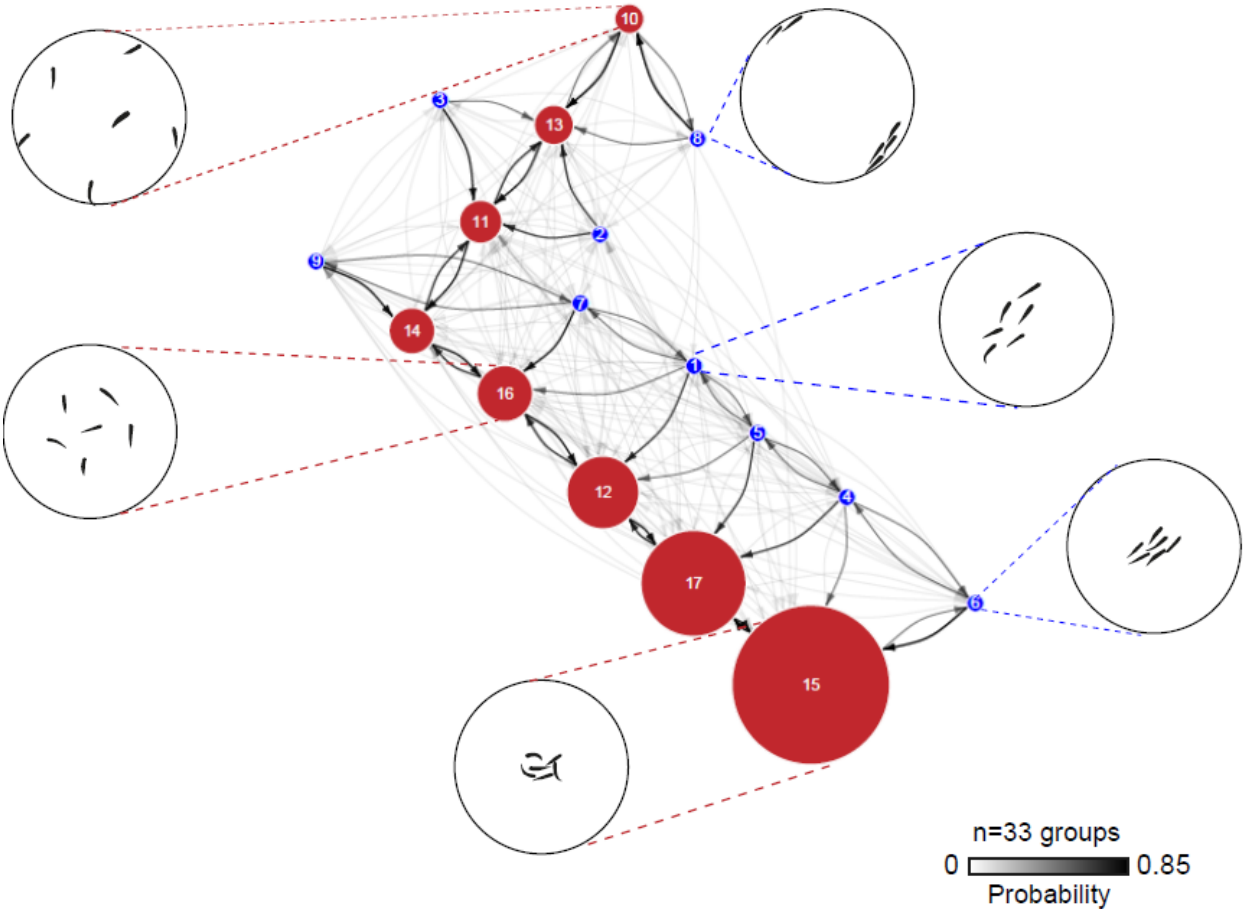


Figure 1. Ethogram of collective behavior in wild-type fish. Circle sizes correspond to state usage ratios, and density of arrows to transition probabilities, with polarized states in blue and unpolarized states in red (n= 33 groups). Dendrogram of state usage is shown in Supplemental Figure 2. Schematic representations (not to scale) of selected states from both categories are shown as blow-ups.

Description of collective activity

Ethograms are convenient ways to display the amount of time spent in each state as

well as the probability of transition between states (Anderson & Perona, 2014; Dankert et al., 2009). The ethogram of group behavior of wild-type fish is shown in Figure 1. Each circle represents a different state (with selected examples shown in blow-ups). For clarity, we have subdivided polarized (school) states (blue) from unpolarized (shoal) (red). The thickness of arrows shows transition probabilities between states. Certain states and transitions are used far more frequently than others (Supplemental Figures 2 and 3). The mean time spent in each state as a block is 0.78 seconds (Supplemental Figure 4). In this assay, fish spend far more time shoaling than schooling, and generally in more cohesive shoals, as shown by the prominence of states in the lower portion of the ethogram.

Mutations

For mutational analysis, we elected to mutate genes putatively related to autism or schizophrenia (Supplemental Table 1 and 2) because of the profound defects in social behavior that characterize these disorders, with the hope that their roles in social behaviors might be evolutionarily conserved, even if the behaviors themselves are species-specific. In order to provide sufficiently accurate information for design of effective CRISPR-Cas9 guide RNAs, we found it necessary to sequence and reassemble the complete genome of the AB strain (data available upon request).

We generated mutations in 35 genes (Supplemental file 1) and screened for changes to collective behavior in adult fish homozygous for the mutation (with the exception of *scn1lab* and *slc18a2* which were not viable as homozygous animals so were evaluated as heterozygous animals). Behavior for each mutation was quantified on ethograms generated as above, and then all 35 compared to wild type AB. Using the Kullback-Leibler divergence approach (Kullback & Leibler, 1951) we identified three mutations with particularly distinctive behavioral changes. Differences from wild-type were cross-validated by multi-class Support Vector Machines (SVM) (Supplemental Figure 5), and finally assessed by human expert validation using double blinded datasets. The ethograms for these three mutant fish are shown in Figure 2 and 3, along with exemplary videos. Some other mutations changed state usage in manner similar to the three described, but to a lesser degree (Supplemental Figure 6).

Enhanced cohesion Fish mutant in the inner mitochondrial membrane peptidase 2-like gene, *immp2l-/-*, tend to aggregate tightly in small shoals, as shown in the video attached to Figure 2b (Supplemental Video 2). The increased shoaling is reflected in enhancement of two tight cohesion states in the lower part of the

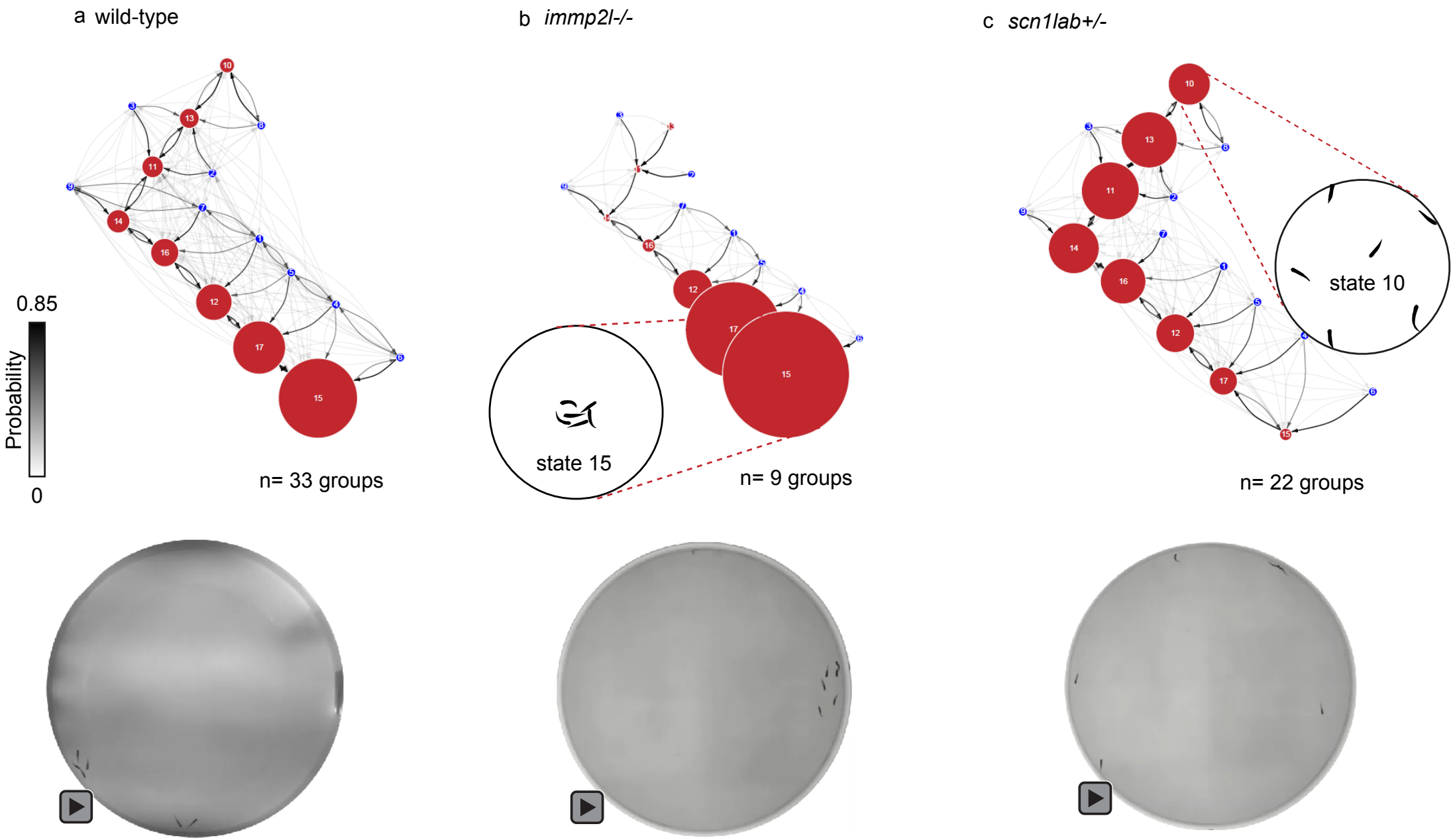
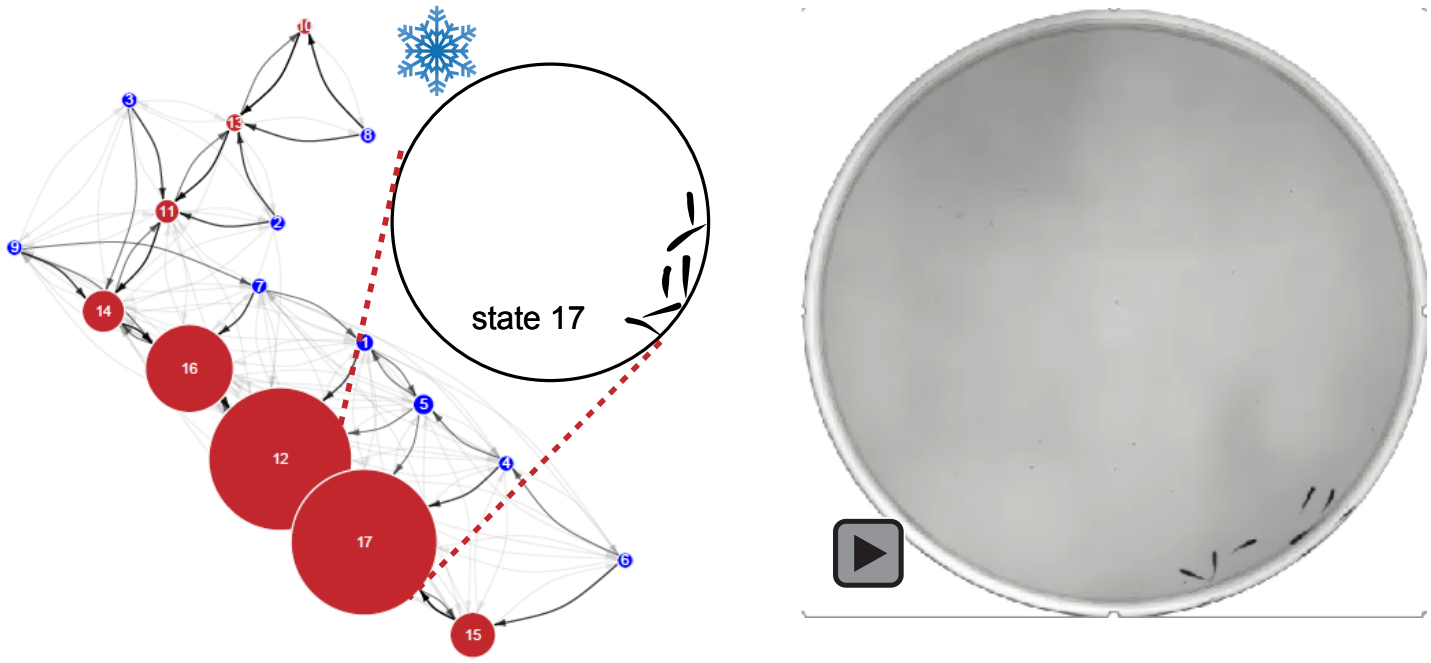


Figure 2. Ethograms of (a) wild-type fish; (b) *immp2l-/-* mutant fish, showing enhanced cohesion (note that dispersed polarized state 8 and unpolarized state10 are missing); (c) *scn1lab+/-* mutant fish, showing enhanced dispersion. Videos can be accessed by clicking where shown, or via Supplemental Video 1 (wild-type), Supplemental Video 2 (*immp2l-/-)*, and Supplemental Video 3 (*scn1lab +/-*). Usage ratios are quantified in Supplemental Figure 3 and 4. (wild-type, n=33 groups; *immp2l-/-*, n=9 groups; *scn1lab+/-*, n=22 groups). Color bar indicates transition probability: [0 0.85].

a *adra1aa-/-*

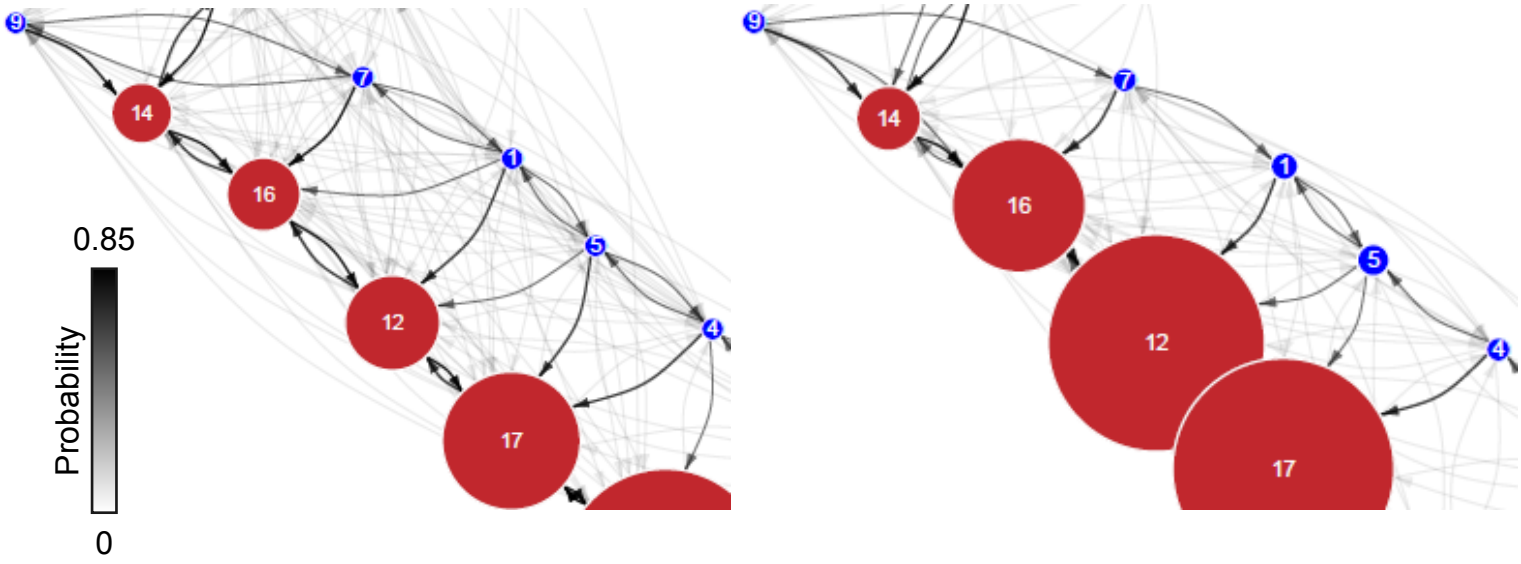


n= 8 groups

b

wild-type

adra1aa-/-



n= 33 groups

n= 8 groups

Figure 3. (a) Ethogram of *adra1aa-/-* mutant fish (n = 8 groups) showing reduced transitions. Video (Supplemental Video 4) is clickable, and shows bouts of "freezing". (b) blow-ups of the lower part of the ethograms, comparing wild-type to *adra1aa-/-* fish, to illustrate the relative paucity of transitions between states of *adra1aa-/-* fish. Color bar indicates transition probability: [0 0.85].

ethogram of Figure 2b (n=9 groups, $p<0.001$; Mann-Whitney U-test; Supplemental Figure 6a). This is the only mutation to completely abolish individual states (states 8 and 10 are missing). In the contrast with *scn1lab+/-*the two missing states are the most dispersed polarized and unpolarized states._The fish can move dynamically between the limited states, and rarely polarize (usage ratios shown in Supplemental Figure 3). Cross validation of this robust phenotype was performed by multi-class SVM (Prediction Accuracy (PA) =0.95; Leave-One-Out; Supplemental Figure 5). These fish feed well and spawn (Figure 4), and in the home tank movements are indistinguishable from normal.

Enhanced dispersion Fish mutant in the Nav1.1 sodium channel, *scn1lab+/-*, remain dispersed throughout the chamber, rarely aggregating, as shown in Figure 2c and its attached video (Supplemental Video 3). This is represented by the shift in predominant states of the ethogram to states to the upper right compared to wild-type (n=22 groups; $p<0.001$; Mann-Whitney U-test; Supplemental Figure 6b). In fact, the states lost in *immp2l-/-* (states 8 and 10) are enhanced in *scn1lab+/-*. Cross validation of this robust phenotype was performed by multi-class SVM (PA=0.88; Leave-One-Out; Supplemental Figure 5). The group centroid speed, pairwise distance, and polarization have little correlation with each other (Supplemental Figure 7). The fish feed, spawn and grow normally.

Enhanced freezing Fish mutant in the adrenergic receptor, *adra1aa-/-*, have many more periods of inactivity than do wild-type, especially towards the end of the half hour test. The state usage of the selected top two states for each comparison is not very distinct from wild type (n=8 groups; Tight shoaling: $p=0.59$; Medium schooling: $p=0.86$ and Dispersion: $p=0.07$; Supplemental Figure 6a-b) as shown in Figure 3a. Compared to wild-type, *adra1aa-/-* mutant fish tend to have significantly more “freeze” motifs in the last third of the recording time ($p<0.001$), as shown in Supplemental Figure 6c. . This difference in dynamics is shown by the relative paucity of arrows in the ethogram blow up of Figure 3b and the attached video (Supplemental video 4). Cross validation of this robust phenotype was performed by multi-class SVM (PA=0.92; Leave-One-Out; Supplemental Figure 5). The fish feed and spawn normally, and their behavior in the home tank is indistinguishable from wild-type.

Effects on other behaviors

We were curious as to whether the mutations that perturb collective activity also markedly affect motor control or vision. In terms of motor function, we measured speed, the frequency of acceleration, and turns of individual fish in mating

behavior. The three mutations with collective behavioral phenotypes noted above can reach the same level or even higher in all three motion measurements compared to wild type fish. We explored mating, another social behavior, a dance with strong and close proximate interactions, on the assumption that subtle visual or motor defects would perturb the effectiveness of this rapid and highly stereotyped interaction (Nasiadka & Clark, 2012; Spence, Gerlach, Lawrence, & Smith, 2008). It also has a quantitative output, i.e., number of fertilized eggs. The spawning consequences of the mutations were modest, *immp2l-/-* and *adra1aa-/-* slightly increasing and *scn1lab+/-* slightly reducing spawning (Figure 4). Hence the genes that dictate collective behavior have only modest effects on the individual behaviors or mating capabilities that we measured.

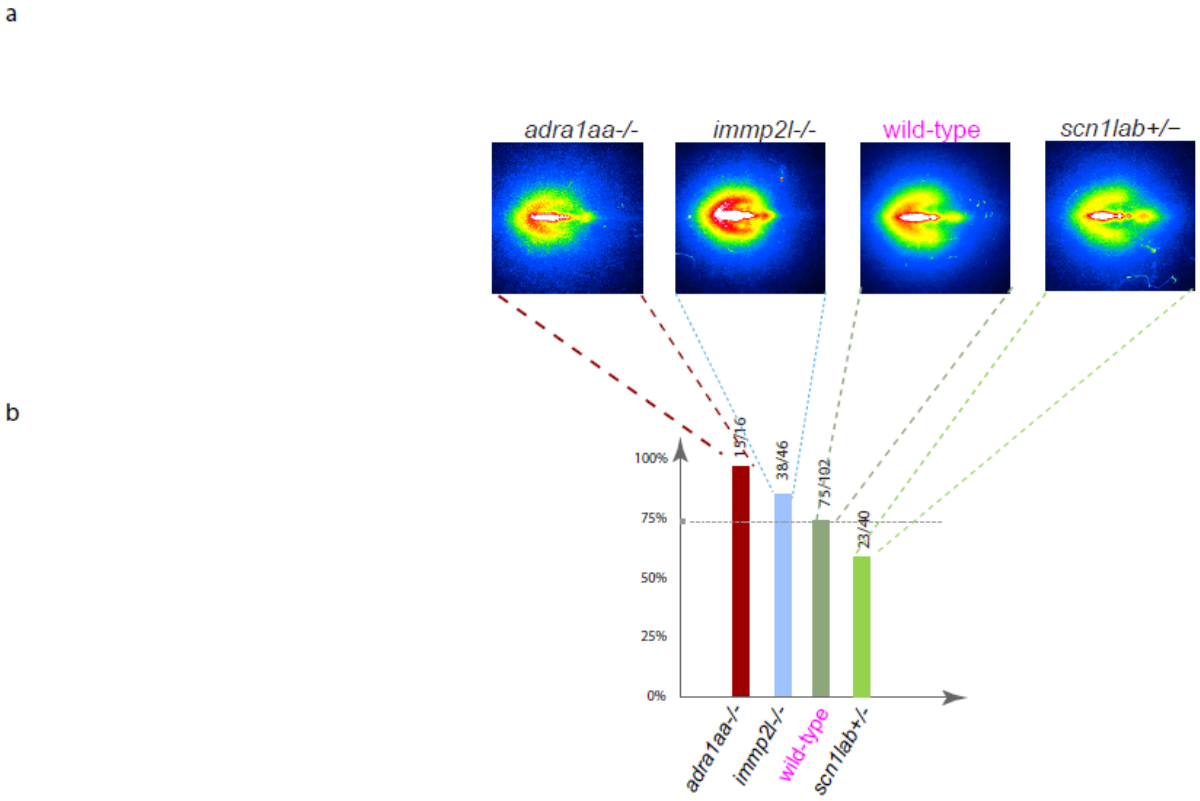


Figure 4. Courtship and spawning is relatively preserved in mutant fish. (a) Positional maps of chasing behavior, with female fish centered facing right and male fish chasing position frequency plotted as a heat map, shows chasing behavior is relatively retained in the mutations of interest. (b) Success of courtship behaviors from mutants and wild-type ranked by the ratio of pairs yielding fertilized embryos, showing modest effects on spawning of the mutations of interest.

Discussion

Darwin (Darwin, 1871) and Tinbergen (Tinbergen, 1952) speculated as to why

some species live in groups and others as solitary animals. Presumably, the benefits of living and moving in a collective provide some advantages (Couzin, 2009; Sumpter, 2010). It is believed that fish school and birds flock in part to help avoid or confuse predators, to improve foraging, and to reduce energy expenditure, especially during long haul migration. But such activities come at a cost to individuals, for example, the need to share food and the exposure of those positioned on the edge of the school to attack (Krause, 1994). Hence, how group activity has arisen though evolution, where selection acts on the genome of the individual, has been a source of intense debate.

Here we show that collective behavior is stereotypical and can be described as the consequence of time spent in, and transitions between, 17 different states, each lasting around. By clustering cost curve (Supplemental Figure 8b), we found these were the fewest number of states needed to define the full repertoire of collective actions for wild-type and mutant fish.

We found that single gene changes can modify collective activity in interpretable manner, using the same states. In other words, mutations did not add states. Mutation in the mitochondrial membrane peptidase 2-like gene *immp2l* increases the tendency to aggregate in tight cohesion, and caused the loss of two highly dispersed states. Mutation in the sodium channel Nav1.1 gene *scn1lab* appears to prevent any group interaction, abolishing cohesion and polarization. Mutation in the adrenergic receptor gene *adra1aa* reduces the transitions between states, and causes the animals often to freeze. Freezing and tight cohesion both are seen as responses to threat and conspecific injury (Chicoli & Paley, 2016; Faustino, Tacao-Monteiro, & Oliveira, 2017; Rennekamp et al., 2016).

It is noticeable that these effects correspond to some of the forces (repulsion, attraction, and spontaneity) applied to individual particles in mathematical models to generate coordinated movements of groups (Katz et al., 2011; Krause et al., 1999; Li & Xiao, 2010; Sumpter, 2010). (We have not seen, so far, mutations with selective effects on polarization, a fourth such force). In these models, individual fish need only to manifest these forces between immediate neighbors for the larger scale behavior to emerge as a consequence.

One might argue that widespread changes in neural function could affect a range of behaviors. Although we cannot rule out subtle changes in these parameters, these mutations did not dramatically affect vision or motor function. Of course,

other sensory cues might be changed, such as pheromones, so critical to mating (Spence et al., 2008; Yabuki et al., 2016) and to aggregation and dispersion of *C. elegans* (Greene, Dobosiewicz, Butcher, McGrath, & Bargmann, 2016). However, mating remains intact, in terms both of the ornate dance and number of fertilized eggs produced, and we noted no deficiencies in feeding. We also recognize that different environments will elicit distinct behaviors, as will other variables, such as numbers of fish participating (Tunstrom et al., 2013; Yates et al., 2009) and familiarity of the fish with the test chamber. In our studies, the assay was done on fish which had never experienced the test chamber before.

To our knowledge these are the first genes shown to regulate group collective behavior in vertebrates. Some of these have been analyzed for effects in mice, but for other attributes. A mutation in the Inner Mitochondrial Membrane Peptidase 2-Like Gene (*immp2l*) reduces fertility (Lu et al., 2008). One where there is some similarity is haploinsufficiency of the voltage-gated sodium channel, Nav1.1 (Han et al., 2012), which causes hyperactivity and loss of preference for another mouse over empty chambers and curiosity about a novel mouse.

We selected these genes for mutation because they have been reported, with differing degrees of certainty, to have genetic association with autism and schizophrenia, disorders marked by impairments in social interaction. Rare genetic mutations of the IMMP2L gene have been identified in Tourette syndrome and autism spectrum disorders (ASD), and this gene is in the 7q31 translocation region of a Tourette syndrome cohort (Bartnik et al., 2012; Petek et al., 2001). Haploinsufficiency of Nav1.1 in human causes Dravet syndrome, with autism proceeded in infancy by seizures. The gene for the adrenergic receptor ADRA1A is in a large genomic region associated with autism, but no behavioral role has previously been speculated.

Of course, whether these mutations are good models for behavioral deficits in humans remains to be seen. Heterozygous mutation of Nav1.1 in humans, fish, and mice all reduce social interactions. But with the exception of Nav1.1, we have evaluated homozygous effects of mutations, while in humans the mutations are studied in patients with heterozygous deficiencies. To the extent there is duplication of function that accompanies the duplication of many genes in teleost compared to humans or mice (Meyer & Schartl, 1999; Sztal, McKaige, Williams, Ruparelia, & Bryson-Richardson, 2018) certain effects may be compensated. In fact, adra1a is split into adra1aa and adra1ab, knockout of each which causes enhanced freezing, suggesting that there may be duplication of function and

compensation. This may also explain why certain candidate autism genes with striking behavioral disturbances when knocked out in mice, such as dlg4a and shank3b (Feyder et al., 2010; Peca et al., 2011), have less striking phenotypes in the fish. Additionally, our assays observe only a small biopsy of potential collective activities under specific controlled circumstances. Nonetheless, zebrafish are amenable to chemical screens for restoration of mutant to normal phenotype (Peterson et al., 2004),

It seems remarkable that some of the genes we discovered here to regulate collective behavior of fish also have been proposed to affect social behavior of humans. This suggests that, perhaps, the role of genes for social behavior, writ large, is conserved through evolution, of course playing out distinctively in different species. The relative convenience of the fish for therapeutic compound screening (Peterson et al., 2004), may be used to advantage in phenotypic screens for reparative agents for disorders such as autism and schizophrenia.

Materials and methods:

Animals

Zebrafish (*Danio rerio*; AB strain) were housed with mixed gender in 3L tanks on a recirculating Aquatic Habitats facility (Custom design, Pentair, USA). The fish were kept at a density of 12 fish per 3L tank. Facility water temperature was kept at 28±0.5°C, and water sourced from deionized water conditioned with sodium bicarbonate (Catalog #SC12-Pentair, USA) and Instant Ocean sea salt (Catalog #IS160-Pentair, USA) to a pH of 7.2±0.5 and conductivity of 420±50 µS. Fish were maintained on a 14-hour light/10-hour dark cycle with light turning on at 07:00 AM. Fish were fed a diet of brine shrimp (Catalog #BSEPCA-Brine Shrimp Direct, USA) twice daily and supplemented with flake fish food (Tetramin Catalog# 98525-Pentair, USA) every other day. All animals were maintained and procedures were performed in accordance with the Institutional Animal Care and Use Committees (IACUC) of NIBR.

Genome-wide CRISPR-Cas9 sgRNA design

Because the current work was carried out in the AB strain and the public genome assembly (GRCz11) is based on TU strain, we established a new genome assembly for the AB strain (unpublished, sequences available upon request). For designing CRISPR/Cas9 sgRNAs, we re-trained sgRNA-efficiency models using Random Forest and Naïve Bayes methods based on previously published sgRNAs. 1280 sgRNAs sequences and their editing efficiencies were obtained

from previous study in TU-strain by Moreno-Mateos, M. A. et al., (Moreno-Mateos et al., 2015), and mapped to the AB genomic sequence. About 150 features were used to train the two Machine Learning models, Random Forest and Naïve Bayes, in classification mode, including genomic strand of sgRNA, GC%, identity of ±4 bp of sgRNA targeting sequence, thermodynamics parameter, ΔG of sgRNA-genomic DNA heteroduplex for sliding windows of different sizes (Sugimoto, 1995), the free-energy of stem-loops 1, 2, and 3, tetraloop, repeat-anti repeat and linker structures of the full-length sgRNA predicted by UNA-fold (Hiroshi Nishimasu, 2014; Markham & Zuker, 2008) and etc. A training accuracy of > 0.7 was achieved by both models, and the efficacies of all sgRNAs in the AB genomic sequence was predicted. Guide RNAs with high predicted efficacy were selected to target the 5′ end of the coding exons or Pfam domains of the genes in current study (for the full gene list, citations for disease indications, please see Supplemental. Table 1 and for the guide RNA sequences used for CRISPR-cas9, please see Supplemental. Table 2).

Micro-injections, CRISPR/Cas9 mutant founder identification

sgRNAs were synthesized using T7 *in vitro* transcription using the MEGAshortscript™ T7 Transcription Kit (ThermoFisher, AM1354). Cas9/sgRNAs were co-injected into 1-cell stage fertilized zebrafish embryos. Conditions were optimized to maximize the CRISPR efficiency in our settings. High indel rates (>90%) were usually observed in fully developed embryos injected with 500 ng/μL Cas9 protein (PNABio, CP01) and 125 ng/μL sgRNA purified with MEGAclear™-96 Transcription Clean-Up Kit (ThermoFisher, AM1909).

Indel rate in the injected zebrafish were measured using NGS at 2 days post fertilization (dpf). PCR were performed on the genomic DNA extracted from zebrafish larvae using the HotSHOT method(Meeker, Hutchinson, Ho, & Trede, 2007), and NGS libraries for PCR products were generated using Nextera DNA Library Preparation Kit (Illumina, FC-121-1031). Subsequently, Nextera libraries were sequenced on Illumina MiSeq, and > avg. 2000 reads for each amplicon were obtained. Sequencing reads were mapped to the reference sequence using BLAT (Kent, 2002), and indels were extracted from the .psl file using a bioinformatic pipeline developed in-house (available upon request). The gene editing efficiency in the injected embryos were calculated as the maximum indel rate within the ±30 bp regions of PAM sites.

We usually sequence five larvae per CRISPR injected clutch to confirm the gene

editing efficiency. The rest fish of the confirmed clutch were raised to adulthood and crossed with wild-type AB fish, and founder fish carrying the desired frameshift mutation were screened from the F1 generation. F1 heterozygotes were inter-crossed, and homozygotes, if viable, were identified and inter-crossed again to obtain sufficient gene knockout fish for behavioral assays. Each CRISPR line was at least an F2 stable line before being run in any assay. All founders and homozygote identification were carried out using fin-clipping PCR, and sequencing the PCR product using NGS as described above. The indel sequence of homozygous mutations and heterozygous mutations that could not be bred to homozygous (*scn1lab* and *slc18a2*) are illustrated in Supplementary File 1, and also the predicted protein sequence aligned with that of the wild-type fish.

Experiment setup for behavioral assays

Attention was paid to ensure fish gender balance and matching of size and age, and to conduct experiments at similar times of days and feeding cycle, and to monitor by video without human presence. All behavioral rooms were fed with water directly from the main fish facility and room temperature and light cycle was consistent with the main facility. The behavioral setups consisted of 20" diameter acrylic circular arena (Custom Design: Acrylic Tank – Clear – 20" OD x 19.25" ID x 8" height – Open Top, Plastic Supply, Inc., USA) filled to a depth of 1 3/4" (~9 L total volume) with system water fed directly from the main housing unit to ensure all water parameters were identical to the housing conditions. The circular arenas were coated on the outside (I00810, Frosted Glass Finish, Krylon, USA) to prevent the fish from being able to see outside of the arenas but to allow IR light transfer. Underneath the tanks were 24 inch adjustable IR panels (with 940 nm IR LEDs made by Shenzhen VICO). Basler Ace 2040-90um Near Infrared (NIR) cameras (Order#-106541, Graftek Imaging, USA) were mounted 23 inches above the arena to collect a dorsal view of the fish. Infrared long pass filters (Midopt LP780-62, Graftek Imaging, USA) were attached to the lens (Scheider Cinegon 1.9, Graftek Imaging, USA) and were set to an aperture of 6 (Supplemental Figure 9a). All trials were recorded after 10 minutes habituation to allow recovery from any stress due to netting. Briefly, shoaling assays were run with six fish (3 males and 3 females), which were randomly combined from multiple tanks of siblings (with the exception of *chd8-/-* which had only male homozygous animals so were paired with 3 AB females) (Supplemental Figure 9c). Each trial was a recording of 30 min at 60 fps. Arenas were rinsed clean with system water at the end of the day and put through a cabinet washer (Type: 9LAV65, IWT Tecniplast Inc., Italy) once a week on a hot water only cycle.

Courtship assays (1 male and 1 female) were all conducted in Aquatic Habitat 2L group mating tanks (Catalog# Breeder2-Pentair, USA) positioned carefully within arenas used for the above shoaling assays (Supplemental Figure 9b). The main arena was still filled with water to ensure all fish's safety as they infrequently escape the mating tanks, which remain lidless due the need to eliminate optical obstructions. Image size was fitted to exactly cover the two tanks in the overnight assays (Supplemental Figure 9d). Three 30-min videos were recorded for each pair of animals: the first recording was initiated at 16:00 PM while fish were separated by a clear divider; the second recording at 22:00 PM while the lights were off and divider still in place; the third recording was done at 08:00 AM (the following day) with the divider removed. After each recording, spawning records were documented and confirmed only if more than 10 embryos survive till 1 dpf with normal course. All mating tanks used in assays were washed daily in under counter washers (Type: GG05/Model PG 8583; Miele AG, Germany).

Automated data collection

We used a virtual instrument console designed within LabVIEW (National Instruments, USA) to control all cameras. The acquisition software was designed with four prioritized functions: 1) Saving the recorded video; 2) Logging all relevant metadata accurately and automatically; 3) Enabling real time user visualization, verification and modification; 4) Generating associated files to streamline downstream analysis.

Non-default parameters were set within the NI-MAX (Measurement and Automation Explorer) including dimensions and offsets for centering (1880x1880 pixels (offset 84/84) for Shoaling assays, 1000x1000 pixels (offset 420/600) for Courtship assays) and frame rate (60 fps). One workstation (X2D65UT#ABA, Z440, Hewlett Packard Inc., USA) was dedicated for simultaneous recordings of two USB3.0 Basler ACE cameras. The resulting videos which are approximately 360 GB each for Shoaling assays, and 108 GB each for Courtship assays are saved on a local 4x2 TB SSD RAID0 (Samsung EVO 850, B&H Photos Inc., USA). All cameras were named uniquely to allow us to trace back which videos are generated by corresponding setups.

Relevant metadata includes time of trial, user, and fish information (genetic background, age, genetic knockouts, genotype, compound treatment (concentration/duration) and other manipulations), and assay information (duration, fps, number of male and female fish). All software developed in house is available upon request, excluding licenses.

Data preprocessing

The processing involved building of fingerprint libraries to distinguish fish accurately even after repeated crossings, using locomotion quantifications such as trajectory, angles, speeds, pairwise distances, and convex hull area. (software based on MATLAB 2013a; MathWorks, U.S.A.) inspired by Perez-Escudero *et al*. (Perez-Escudero, Vicente-Page, Hinz, Arganda, & de Polavieja, 2014). In addition, we used supervised behavioral annotation of simple motifs, such as nonsocial turn, dive, rest, and cruise, or social chase, cross, and parallel rotation, with previously trained classifiers as described by Kabra *et al* (Kabra et al., 2013). These processing procedures are integrated into an overarching program that manages data dependencies to automatically start the next stages when ready without needing additional user inputs (scripts available upon request).

Unsupervised collective behavior state hierarchical structure clustering

For this analysis we used 5 frame epochs. Polarization was determined by the swimming direction angle. For the i^{th} fish in epoch n, this was represented by the tangent angle of the i^{th} fish's trajectory from the 1st frame of the n^{th} epoch to the 1st frame of the $n+1^{th}$ epoch in the Cartesian coordinates where the trajectory was represented. If the standard deviation of the angles of the six fish in an epoch was lower than the threshold 0.6 radians and the minimum swimming speed of the 6 fish in the epoch faster than 1 inch per second, the epoch was classified as "polarized swimming"; otherwise the epoch will be classified as "unpolarized swimming".

The pairwise distances among the six fish's centroids was calculated from the 1st frame to represent the pairwise distances in each epoch. The pairwise distances within each epoch was ranked in a descend order. A principle component analysis (PCA) dimension reduction was applied to the ranked pairwise distance matrices (with a dimension of C_6^2 by number of epochs) in polarized swimming and un-polarized swimming epochs, respectively. The top 2 components that capture more than 95% of variance were kept and *fuzzy C*-means clustering was implemented on the top 2 PCA components with $k=9$ states, for polarized swimming and $k=8$ for unpolarized swimming components, respectively. The flow chart of the unsupervised learning approach we performed is shown in Supplemental Figure 9a. A majority filter is applied to the clustered labels to eliminate cases that only a single epoch has a different label from previous epoch labels and post epoch labels, since there is less interest in super short states with

duration less than 1/6 seconds. Change points identified state block duration distribution is shown in Supplemental Figure 9. The mean of block duration is 0.78 seconds with a standard deviation of 0.96 seconds.

A t-distributed stochastic neighbor embedding (t-SNE) (Laurens van der Maaten, 2008) plot was used to visualize the high dimensional data in a two dimensional space for the wild-type fish collective behavior (Supplemental Figure 10a-b). The collective behavior pattern is shown by clickable videos for the selected states.

The *k* values were determined by optimizing the clustering cost (D T Pham, 2005) (Supplemental Figure 9b). Moreover, to increase the resolution of the state quantification, the state number was pushed to *k=50* for both polarized and unpolarized swimming. The overall results remain consistent.

Self-Organizing map (SOM) clustering and k means were applied as orthogonal approaches to *fuzzy C*-means. The performance of SOM and k means is equivalent to *fuzzy C*-means (data not shown). All the quantifications are implemented in Matlab R2014a (Mathworks. Inc, U.S.A.). Figure 1 to 3 and supplementary figure 1, 2, 3, 5, and 10 were plotted in *R 3.4.1*, the rest figures were plotted in Matlab. Silhouette analysis was performed for both polarized and un-polarized swimming to ensure no significant over classification (data not shown). The Matlab and *R* codes to replot all the figures in this manuscript are available upon request.

Ethogram visualizations

Ethograms are used to represent the different states that occur during collective behaviors. The ethogram indicates the frequency and transition probability with which one state is followed by another state (Anderson & Perona, 2014; Dankert et al., 2009). To represent the relationship among behavioral states and probability distribution of states using the ethogram, two type of analysis were performed. First, we computed the one-step transition probability which is the conditional distribution of the current state given by the previous state (Bishop, 2006). Second, we computed the probability distribution of states which is the frequency of the occurrence. In practice, The ethogram is generated by *igraph R* package (Csardi, 2006). In the ethogram, the edge and vertex are corresponding to the one-step transition probability and probability distribution of states, respectively. The thicker and darker the edge is, the higher transition probability is; and the bigger the vertex is, the higher the state appearance probability is. All attributes (e.g., vertex color, vertex size, arrow size, edge size, and edge curvature, provided in

the scripts) were adjusted by parameters of the package to generate current figures.

Kullback-Leibler divergence

Kullback–Leibler divergence (relative entropy) (Kullback & Leibler, 1951) is a mathematical method to measure the differences of two probability distributions. The difference D is a function of the two discrete probability distributions P and Q as shown in the following equation (1):

$$D(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)} \quad (1)$$

Kullback–Leibler divergence was applied to the polarized and unpolarized state usages, respectively. The differences between each mutation and wild type were ranked in descending order (Supplementary File 2) and the top ranked mutations were selected.

Supervised behavioral phenotype category classification

Feature extraction

State usage, transition matrix and swimming speed in each trial are employed as features (also called measurements or attributes), including the averaged aggregation (unpolarized) state usages, dispersion (unpolarized) state usages, schooling (polarized) state usages, overall state usages, and transition matrix. Speed features include mean speed of all the 6 fish each trial in the last 10 minutes of the recording, number of stop epochs from all 6 fish in each trial. PCA then is implemented to reduce the dimensionality of the feature space. Top 12 PCs are kept capturing above 95% of the variance. In addition, an alternative feature pool is tested, including only swimming speed distribution in each unpolarized state from all the trials. PCA dimensional reduction is followed and 3 top PCs are selected to capture 95% variance of the data.

Support vector machine

The classifier we implement is the support vector machine (SVM) (Vapnick, 1995), which is widely used for pattern classification (Ma, Randolph, & Drish, 2001). For the classical binary formulation of SVM classification, the output of an SVM classifier yields:

$$y = sign(f(x)), \quad (2)$$

$$f(x) = \sum_{i=1}^{N_{SV}} \alpha_i y_i k(x, x_i) + b, \quad (3)$$

where α_i are the Lagrange multipliers from solving the quadratic optimization

problem, y_i are the class targets, x_i are the input data points, $k(x, x_i)$is a kernel function (Chang, 2011), b is the bias, and N_{SV} is the number of support vectors. Linear kernel is used here since it is more capable of avoiding possible overfitting in the classification than other kernels (Keerthi & Lin, 2003).

To extend a binary classification to a multi-class classification problem such as the problem at hand, several binary SVM classifiers need to be utilized. One-against-one classification method trains *k(k-1)/2* binary classifiers (where *k* is the number of classes) to solve *k* class classification problem. During prediction the class with the most votes becomes the winner. One-against-one outperforms other multi-class SVM classification methods (Hsu & Lin, 2002) and is the default choice in the libSVM package (Chang, 2011) that has been utilized in this study.

Leave-one-out

The Leave-One-Out (LOO) (Efron, 1993) validation method is used to evaluate the performance of the classifiers described in this study. Random one trial from each category is taken as the testing data and the remaining as the training data to train SVM multi-class classifiers. This procedure was applied for 100 times (most trial numbers are less than 10) by randomly choosing training and testing trials. All the test results are then combined into a cumulative confusion matrix.

Acknowledgements

We thank Ricardo Dolmetsch, Rainer Friedrich, Joseph Loureiro, Mark Borowsky, Brant Peterson, Gerald Sun, Lingling Shen, Ajeet Singh, Vibhas Aravamuthan, Stephen Litster, Guangliang Wang, and Jian Fang for discussion. We also want to thank Gerlinde Wussler, Kara Moloney, Meghan Aguirre, Franki Vetrano-Olsen, KarenJ Lee, Haley Clark, Johnathan Tobin, and Joseph Beaton for their support in the fish facility, also Stacey Gearin, Stephanie Wiessner, and Jessica Garver for CRISPR injection work, Caroline Fawcett for fish database management; thank Michael Steeves and Michael Derby for their help in computation resource; thank Michael Paolucci and Aaron Bickel for their help in hardware customizing; thank Andrea Schwarz and Lauren Goldfinger for project coordination. X.X. is supported by NIBR postdoctoral scholar program. Funding is provided by the Novartis Institutes for Biomedical Research (NIBR).

Competing interests

MCF: Consultant to NIBR; Boards of Directors of Semma Therapeutics and Beam Therapeutics; SAB of Tenaya Therapeutics.

Reference

Anderson, D. J., & Perona, P. (2014). Toward a science of computational ethology. *Neuron, 84*(1), 18-31. doi:10.1016/j.neuron.2014.09.005

Baier, H. (2000). Zebrafish on the move: towards a behavior-genetic analysis of vertebrate vision. *Curr Opin Neurobiol, 10*(4), 451-455.

Bartnik, M., Szczepanik, E., Derwinska, K., Wisniowiecka-Kowalnik, B., Gambin, T., Sykulski, M., . . . Stankiewicz, P. (2012). Application of array comparative genomic hybridization in 102 patients with epilepsy and additional neurodevelopmental disorders. *Am J Med Genet B Neuropsychiatr Genet, 159b*(7), 760-771. doi:10.1002/ajmg.b.32081

Bishop, Christopher (2006). *Pattern Recognition and Machine Learning*: Springer.

Brockerhoff, S. E., Hurley, J. B., Janssen-Bienhold, U., Neuhauss, S. C., Driever, W., & Dowling, J. E. (1995). A behavioral screen for isolating zebrafish mutants with visual system defects. *Proc Natl Acad Sci U S A, 92*(23), 10545-10549.

Chang, Chih-Chung, Lin, Chih-Jen. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol, 2*(3), 1-27.

Chicoli, Amanda, & Paley, Derek A. (2016). Probabilistic information transmission in a network of coupled oscillators reveals speed-accuracy trade-off in responding to threats. *Chaos, 26*(11), 116311. doi:10.1063/1.4966682

Chuang, K. S., Tzeng, H. L., Chen, S., Wu, J., & Chen, T. J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. *Comput Med Imaging Graph, 30*(1), 9-15. doi:10.1016/j.compmedimag.2005.10.001

Couzin, I. D. (2009). Collective cognition in animal groups. *Trends Cogn Sci, 13*(1), 36-43. doi:10.1016/j.tics.2008.10.002

Csardi, Gabor, & Nepusz, Tamas. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems, 1695*(5), 1-9.

D T Pham, S S Dimov, C D Nguyen. (2005). Selection of K in K-means clustering *Journal of Mechanical Engineering Science, 219*(1), 103-119.

Dankert, H., Wang, L., Hoopfer, E. D., Anderson, D. J., & Perona, P. (2009). Automated monitoring and analysis of social behavior in Drosophila. *Nat Methods, 6*(4), 297-303. doi:10.1038/nmeth.1310

Darwin, Charles. (1871). *The Descent of Man, and Selection in Relation to Sex* United Kingdom: John Murray.

Efron, B., Tibshirani, R.J. . (1993). *An introduction to the bootstrap*: Chapman&Hall.

Faustino, A. I., Tacao-Monteiro, A., & Oliveira, R. F. (2017). Mechanisms of social buffering of fear in zebrafish. *Sci Rep, 7*, 44329. doi:10.1038/srep44329

Feng, Y., Dong, F., Xia, X., Hu, C. H., Fan, Q., Hu, Y., . . . Mutic, S. (2017). An adaptive Fuzzy C-means method utilizing neighboring information for breast tumor segmentation in ultrasound images. *Med Phys, 44*(7), 3752-3760. doi:10.1002/mp.12350

Feyder, M., Karlsson, R. M., Mathur, P., Lyman, M., Bock, R., Momenan, R., . . . Holmes, A. (2010). Association of mouse Dlg4 (PSD-95) gene deletion and human DLG4 gene variation with phenotypes relevant to autism spectrum disorders and Williams' syndrome. *Am J Psychiatry, 167*(12), 1508-1517. doi:10.1176/appi.ajp.2010.10040484

Friedrich, R. W., Jacobson, G. A., & Zhu, P. (2010). Circuit neuroscience in zebrafish. *Curr Biol, 20*(8), R371-381. doi:10.1016/j.cub.2010.02.039

633 Granato, M., van Eeden, F. J., Schach, U., Trowe, T., Brand, M., Furutani-Seiki, M., . . . Nusslein-Volhard,
634 C. (1996). Genes controlling and mediating locomotion behavior of the zebrafish embryo and
635 larva. *Development, 123*, 399-413.
636 Greene, J. S., Brown, M., Dobosiewicz, M., Ishida, I. G., Macosko, E. Z., Zhang, X., . . . Bargmann, C. I.
637 (2016). Balancing selection shapes density-dependent foraging behaviour. *Nature, 539*(7628),
638 254-258. doi:10.1038/nature19848
639 Greene, J. S., Dobosiewicz, M., Butcher, R. A., McGrath, P. T., & Bargmann, C. I. (2016). Regulatory
640 changes in two chemoreceptor genes contribute to a Caenorhabditis elegans QTL for foraging
641 behavior. *Elife, 5*. doi:10.7554/eLife.21454
642 Greenwood, A. K., Wark, A. R., Yoshida, K., & Peichel, C. L. (2013). Genetic and neural modularity
643 underlie the evolution of schooling behavior in threespine sticklebacks. *Curr Biol, 23*(19), 1884-
644 1888. doi:10.1016/j.cub.2013.07.058
645 Han, S., Tai, C., Westenbroek, R. E., Yu, F. H., Cheah, C. S., Potter, G. B., . . . Catterall, W. A. (2012).
646 Autistic-like behaviour in Scn1a+/- mice and rescue by enhanced GABA-mediated
647 neurotransmission. *Nature, 489*(7416), 385-390. doi:10.1038/nature11356
648 Han, S., Taralova, E., Dupre, C., & Yuste, R. (2018). Comprehensive machine learning analysis of Hydra
649 behavior reveals a stable basal behavioral repertoire. *Elife, 7*. doi:10.7554/eLife.32605
650 Hiroshi Nishimasu, F. Ann Ran, Patrick D. Hsu, Silvana Konermann, Soraya Shehata, Naoshi Dohmae,
651 Ryuichiro Ishitani, Feng Zhang, and Osamu Nureki. (2014). Crystal Structure of Cas9 in Complex
652 with Guide RNA and Target DNA. *Cell, 156*(5), 935-949.
653 Hong, W., Kennedy, A., Burgos-Artizzu, X. P., Zelikowsky, M., Navonne, S. G., Perona, P., & Anderson, D.
654 J. (2015). Automated measurement of mouse social behaviors using depth sensing, video
655 tracking, and machine learning. *Proc Natl Acad Sci U S A, 112*(38), E5351-5360.
656 doi:10.1073/pnas.1515982112
657 Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE*
658 *Trans Neural Netw, 13*(2), 415-425. doi:10.1109/72.991427
659 Ioannou, C. C., Guttal, V., & Couzin, I. D. (2012). Predatory fish select for coordinated collective motion
660 in virtual prey. *Science, 337*(6099), 1212-1215. doi:10.1126/science.1218919
661 Ioannou, C. C., Morrell, L. J., Ruxton, G. D., & Krause, J. (2009). The effect of prey density on predators:
662 conspicuousness and attack success are sensitive to spatial scale. *Am Nat, 173*(4), 499-506.
663 doi:10.1086/597219
664 Jolles, Jolle W., Boogert, Neeltje J., Sridhar, Vivek H., Couzin, Iain D., & Manica, Andrea. (2017).
665 Consistent Individual Differences Drive Collective Behavior and Group Functioning of Schooling
666 Fish. *Current Biology, 27*(18), 2862-2868.e2867. doi:10.1016/j.cub.2017.08.004
667 Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S., & Branson, K. (2013). JAABA: interactive machine
668 learning for automatic annotation of animal behavior. *Nat Methods, 10*(1), 64-67.
669 doi:10.1038/nmeth.2281
670 Katz, Y., Tunstrom, K., Ioannou, C. C., Huepe, C., & Couzin, I. D. (2011). Inferring the structure and
671 dynamics of interactions in schooling fish. *Proc Natl Acad Sci U S A, 108*(46), 18720-18725.
672 doi:10.1073/pnas.1107583108
673 Keerthi, S. S., & Lin, C. J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel.
674 *Neural Comput, 15*(7), 1667-1689. doi:10.1162/089976603321891855
675 Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res, 12*(4), 656-664.
676 doi:10.1101/gr.229202
677 Krause, J. (1994). Differential fitness returns in relation to spatial position in groups. *Biol Rev Camb*
678 *Philos Soc, 69*(2), 187-206.
679 Krause, J., Hartmann, N., & Pritchard, V. L. (1999). The influence of nutritional state on shoal choice in
680 zebrafish, Danio rerio. *Anim Behav, 57*(4), 771-775. doi:10.1006/anbe.1998.1010

Kravitz, E. A., & Huber, R. (2003). Aggression in invertebrates. *Curr Opin Neurobiol, 13*(6), 736-743.

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79-86.

Laurens van der Maaten, Geoffrey Hinton. (2008). Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research, 9*, 2579–2605.

Li, X., & Xiao, J. (2010). Swarming in homogeneous environments: a social interaction based framework. *J Theor Biol, 264*(3), 747-759. doi:10.1016/j.jtbi.2010.02.016

Lu, B., Poirier, C., Gaspar, T., Gratzke, C., Harrison, W., Busija, D., . . . Bishop, C. E. (2008). A mutation in the inner mitochondrial membrane peptidase 2-like gene (Immp2l) affects mitochondrial function and impairs fertility in mice. *Biol Reprod, 78*(4), 601-610. doi:10.1095/biolreprod.107.065987

Lukeman, R., Li, Y. X., & Edelstein-Keshet, L. (2010). Inferring individual rules from collective behavior. *Proc Natl Acad Sci U S A, 107*(28), 12576-12580. doi:10.1073/pnas.1001763107

Ma, Changxue, Randolph, M. A., & Drish, J. (2001). *A support vector machines-based rejection technique for speech recognition*. Paper presented at the Acoustics, Speech, and Signal Processing, Proceedings. (ICASSP '01). 2001 IEEE International Conference on.

Markham, N. R., & Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol, 453*, 3-31. doi:10.1007/978-1-60327-429-6_1

Meeker, N. D., Hutchinson, S. A., Ho, L., & Trede, N. S. (2007). Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *Biotechniques, 43*(5), 610, 612, 614.

Meyer, A., & Schartl, M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol, 11*(6), 699-704.

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J. D., Fernandez, J. P., Mis, E. K., Khokha, M. K., & Giraldez, A. J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods, 12*(10), 982-988. doi:10.1038/nmeth.3543

Muto, A., Orger, M. B., Wehman, A. M., Smear, M. C., Kay, J. N., Page-McCaw, P. S., . . . Baier, H. (2005). Forward genetic analysis of visual behavior in zebrafish. *PLoS Genet, 1*(5), e66. doi:10.1371/journal.pgen.0010066

Nasiadka, A., & Clark, M. D. (2012). Zebrafish breeding in the laboratory environment. *Ilar j, 53*(2), 161-168. doi:10.1093/ilar.53.2.161

Peca, J., Feliciano, C., Ting, J. T., Wang, W., Wells, M. F., Venkatraman, T. N., . . . Feng, G. (2011). Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature, 472*(7344), 437-442. doi:10.1038/nature09965

Perez-Escudero, Alfonso, Vicente-Page, Julian, Hinz, Robert C., Arganda, Sara, & de Polavieja, Gonzalo G. (2014). idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nat Meth, 11*(7), 743-748. doi:10.1038/nmeth.2994

http://www.nature.com/nmeth/journal/v11/n7/abs/nmeth.2994.html#supplementary-information

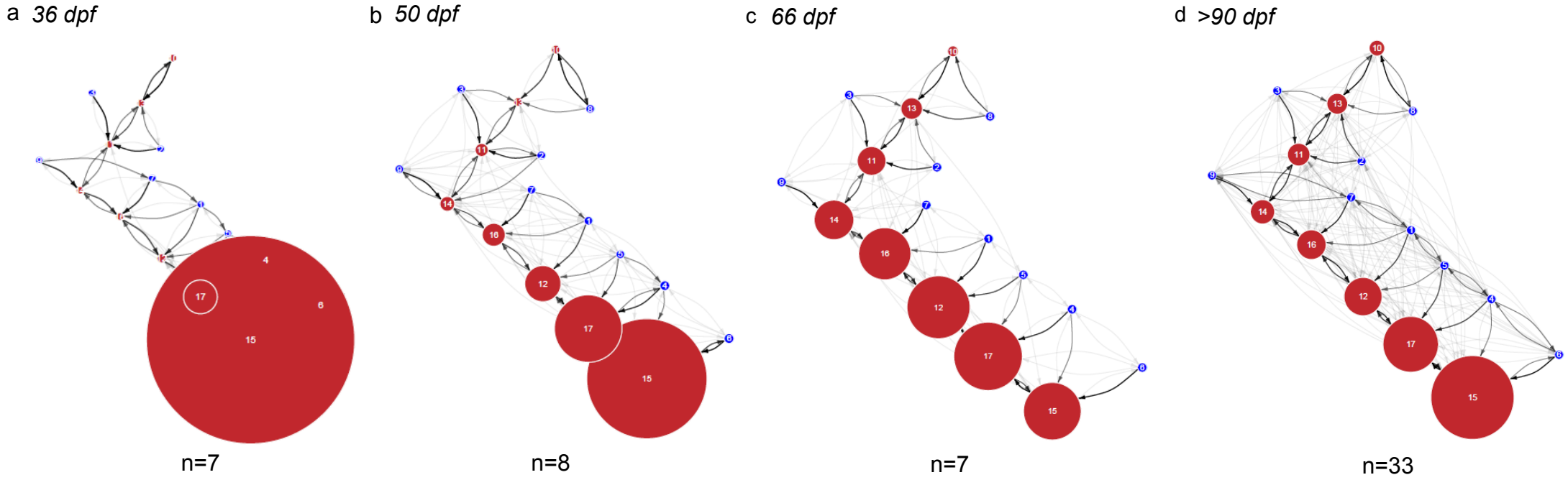
Petek, E., Windpassinger, C., Vincent, J. B., Cheung, J., Boright, A. P., Scherer, S. W., . . . Wagner, K. (2001). Disruption of a novel gene (IMMP2L) by a breakpoint in 7q31 associated with Tourette syndrome. *Am J Hum Genet, 68*(4), 848-858. doi:10.1086/319523

Peterson, R. T., Shaw, S. Y., Peterson, T. A., Milan, D. J., Zhong, T. P., Schreiber, S. L., . . . Fishman, M. C. (2004). Chemical suppression of a genetic mutation in a zebrafish model of aortic coarctation. *Nat Biotechnol, 22*(5), 595-599. doi:10.1038/nbt963

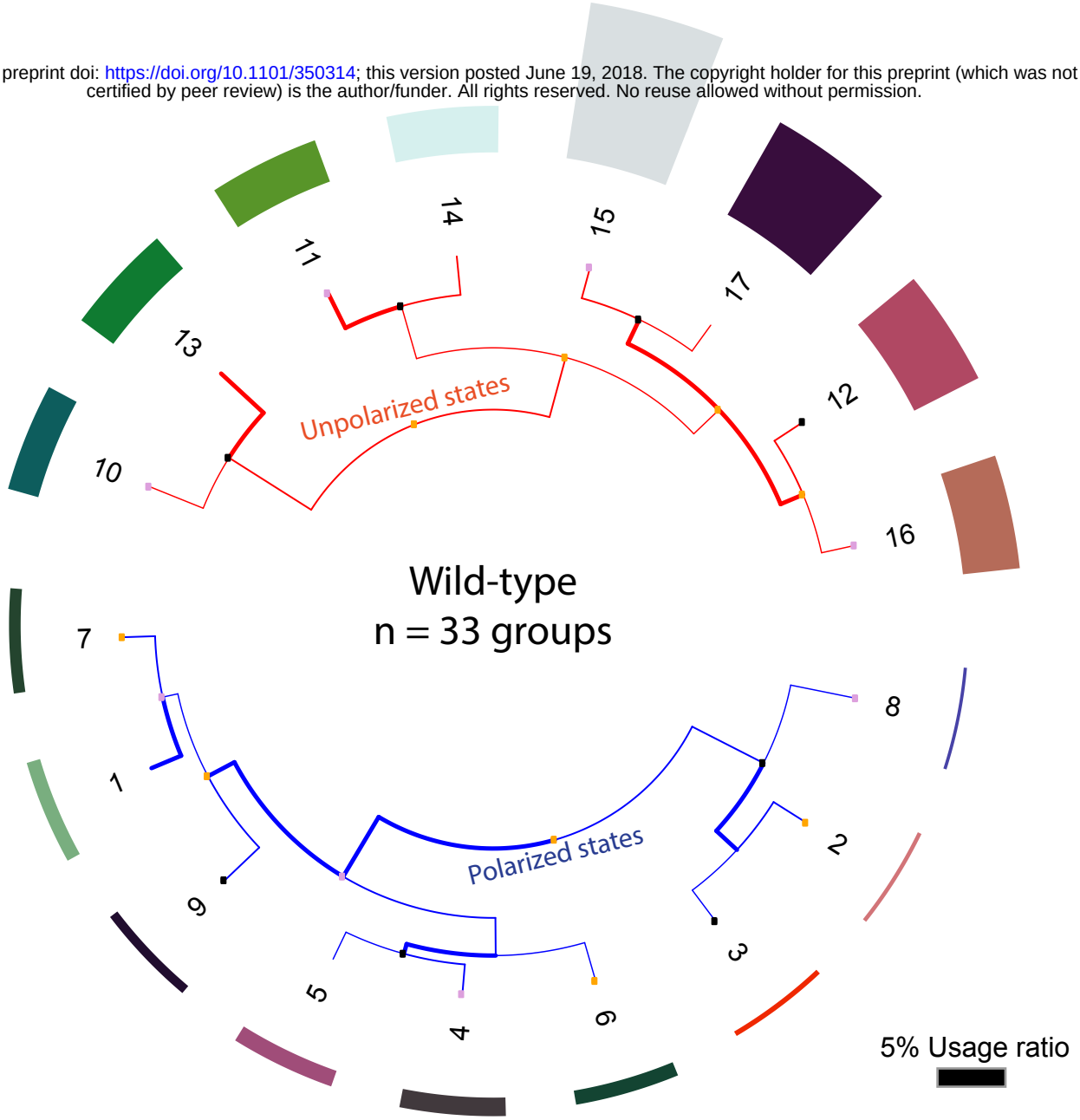
Portugues, R., & Engert, F. (2009). The neural basis of visual behaviors in the larval zebrafish. *Curr Opin Neurobiol, 19*(6), 644-647. doi:10.1016/j.conb.2009.10.007

Rennekamp, Andrew J., Huang, Xi-Ping, Wang, You, Patel, Samir, Lorello, Paul J., Cade, Lindsay, . . . Peterson, Randall T. (2016). Sigma-1 receptor ligands control a switch between passive and active threat responses. *Nature chemical biology, 12*(7), 552-558. doi:10.1038/nchembio.2089

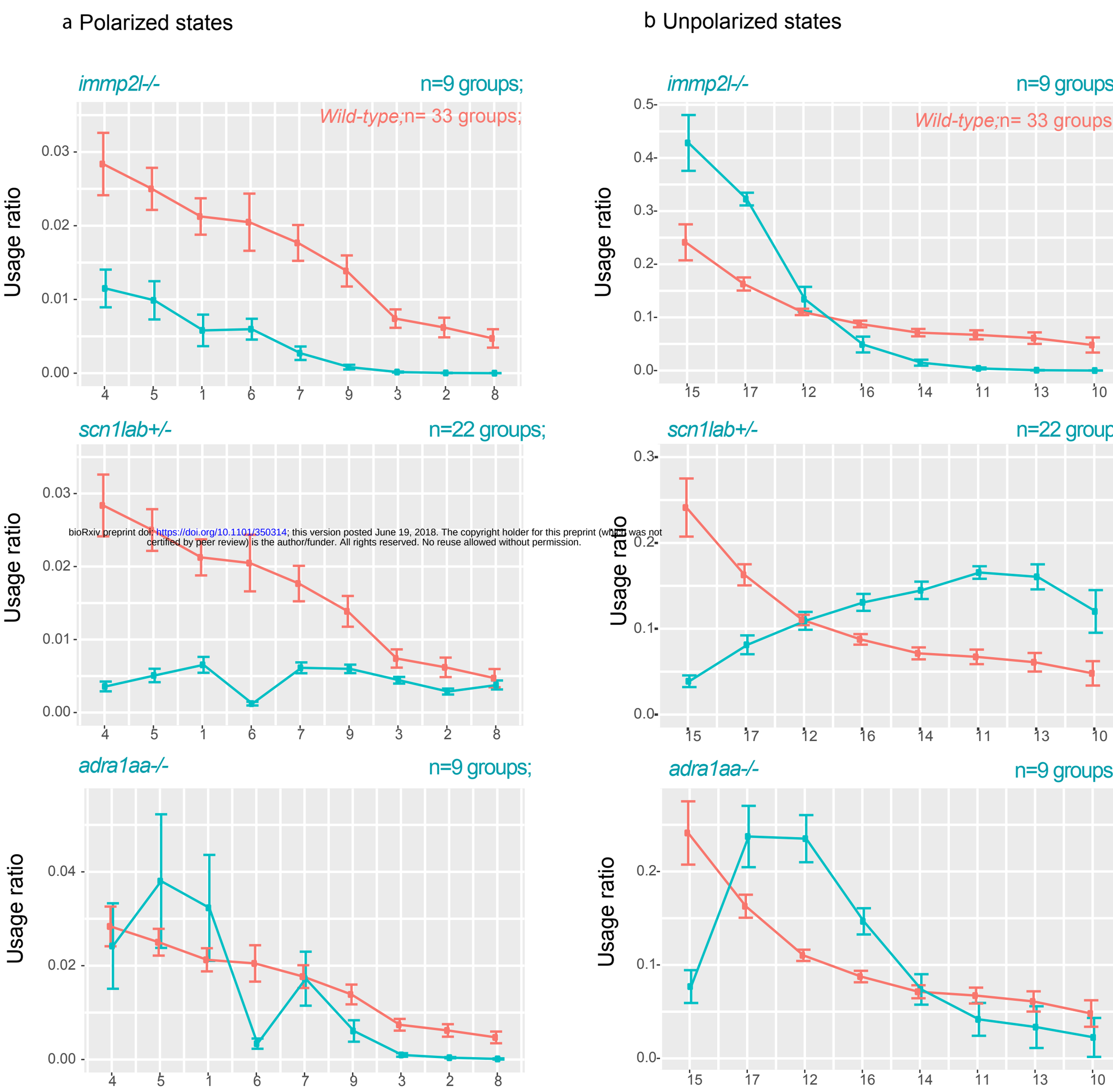
729 Rihel, J., Prober, D. A., Arvanites, A., Lam, K., Zimmerman, S., Jang, S., . . . Schier, A. F. (2010). Zebrafish
730 behavioral profiling links drugs to biological targets and rest/wake regulation. *Science,*
731 *327*(5963), 348-351. doi:10.1126/science.1183090
732 Siew, Xiaosheng Peng ; Chengke Zhou ;Donald M. Hepburn ; Martin D. Judd;W. H. (2013). Application of
733 K-Means method to pattern recognition in on-line cable partial discharge monitoring *IEEE*
734 *Transactions on Dielectrics and Electrical Insulation, 20*(3), 754 - 761
735 Spedding, G. (2011). Aerodynamics: The cost of flight in flocks. *Nature, 474*(7352), 458-459.
736 doi:10.1038/474458a
737 Spence, R., Gerlach, G., Lawrence, C., & Smith, C. (2008). The behaviour and ecology of the zebrafish,
738 Danio rerio. *Biol Rev Camb Philos Soc, 83*(1), 13-34. doi:10.1111/j.1469-185X.2007.00030.x
739 Sumpter, David J. T. (2010). *Collective Animal Behavior*: Princeton University Press.
740 Sztal, T. E., McKaige, E. A., Williams, C., Ruparelia, A. A., & Bryson-Richardson, R. J. (2018). Genetic
741 compensation triggered by actin mutation prevents the muscle damage caused by loss of actin
742 protein. *PLoS Genet, 14*(2), e1007212. doi:10.1371/journal.pgen.1007212
743 Tinbergen, N. (1952). The Study of Instict. *Oxford, ISBN-0198577222*.
744 Todd, J. G., Kain, J. S., & de Bivort, B. L. (2017). Systematic exploration of unsupervised methods for
745 mapping behavior. *Phys Biol, 14*(1), 015002. doi:10.1088/1478-3975/14/1/015002
746 Torney, C. J., Berdahl, A., & Couzin, I. D. (2011). Signalling and the evolution of cooperative foraging in
747 dynamic environments. *PLoS Comput Biol, 7*(9), e1002194. doi:10.1371/journal.pcbi.1002194
748 Tunstrom, K., Katz, Y., Ioannou, C. C., Huepe, C., Lutz, M. J., & Couzin, I. D. (2013). Collective states,
749 multistability and transitional behavior in schooling fish. *PLoS Comput Biol, 9*(2), e1002915.
750 doi:10.1371/journal.pcbi.1002915
751 Vapnick, V. (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.
752 Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., . . . Datta, S. R.
753 (2015). Mapping Sub-Second Structure in Mouse Behavior. *Neuron, 88*(6), 1121-1135.
754 doi:10.1016/j.neuron.2015.11.031
755 Yabuki, Y., Koide, T., Miyasaka, N., Wakisaka, N., Masuda, M., Ohkura, M., . . . Yoshihara, Y. (2016).
756 Olfactory receptor for prostaglandin F2alpha mediates male fish courtship behavior. *Nat*
757 *Neurosci, 19*(7), 897-904. doi:10.1038/nn.4314
758 Yates, C. A., Erban, R., Escudero, C., Couzin, I. D., Buhl, J., Kevrekidis, I. G., . . . Sumpter, D. J. (2009).
759 Inherent noise can facilitate coherence in collective swarm motion. *Proc Natl Acad Sci U S A,*
760 *106*(14), 5464-5469. doi:10.1073/pnas.0811195106
761 Zhu, P., Narita, Y., Bundschuh, S. T., Fajardo, O., Scharer, Y. P., Chattopadhyaya, B., . . . Friedrich, R. W.
762 (2009). Optogenetic Dissection of Neuronal Circuits in Zebrafish using Viral Gene Transfer and
763 the Tet System. *Front Neural Circuits, 3*, 21. doi:10.3389/neuro.04.021.2009
764



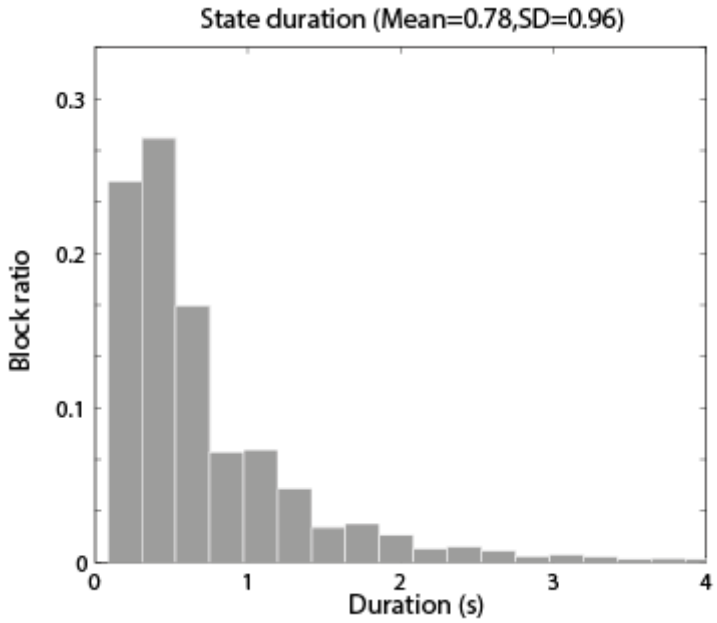
Supplemental Figure 1 Development of collective behavior, showing ethograms of wild-type fish (a) 36 dpf (note that most of the usage concentrates in state 15, the tightest cohesion state, and that the dispersed schooling state 8 is missing; (b) 50 dpf (tight cohesion states 15 and 17 are still the majority of state usage); (c) 66 dpf (the collective behavior state usage is similar to adult fish, although the transitions between states are less dense); (d) adult fish. Collective behavior is not mature until the fish are about 3 months old.



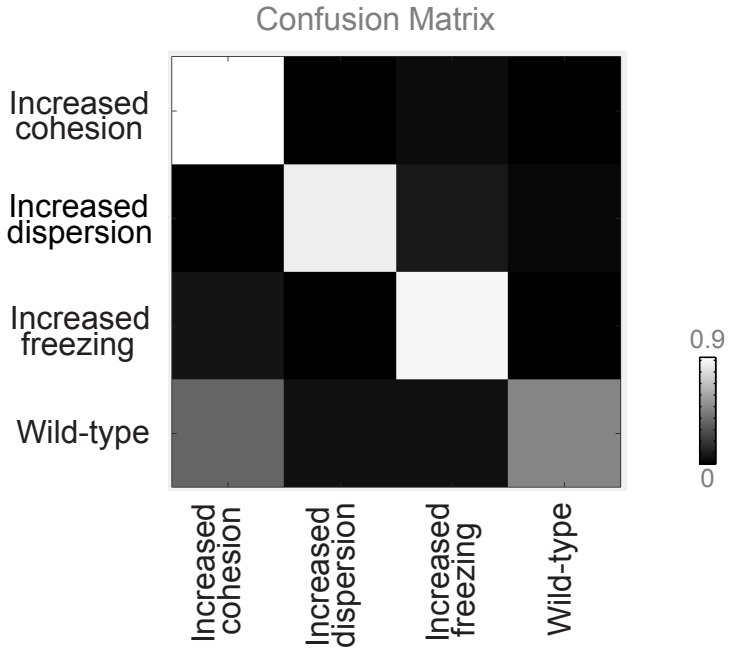
Supplemental Figure 2 Dendrogram of wild-type fish collective behavioral states defined using hierarchical clustering.



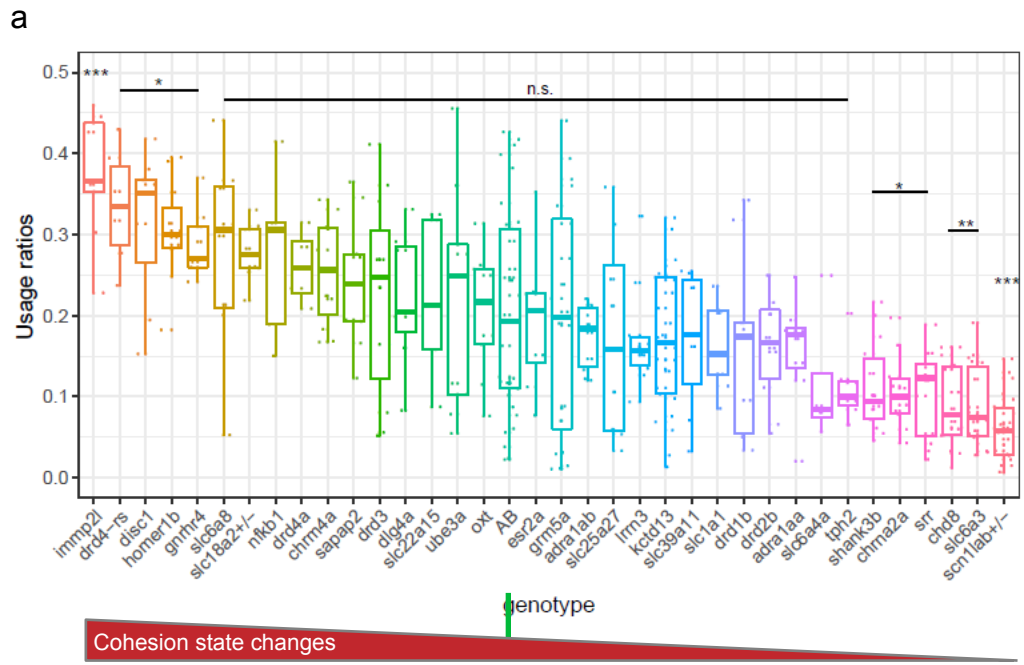
Supplemental Figure 3 Usage ratios of mutant states. Usage ratios of all (a) polarized and (b) unpolarized states are compared to corresponding states in wild-type (error bars indicate ±SEM). The usage ratios are compatible with observations described in the text for the ethograms of Figure 2, and with significance test in Supplemental Figure 6.

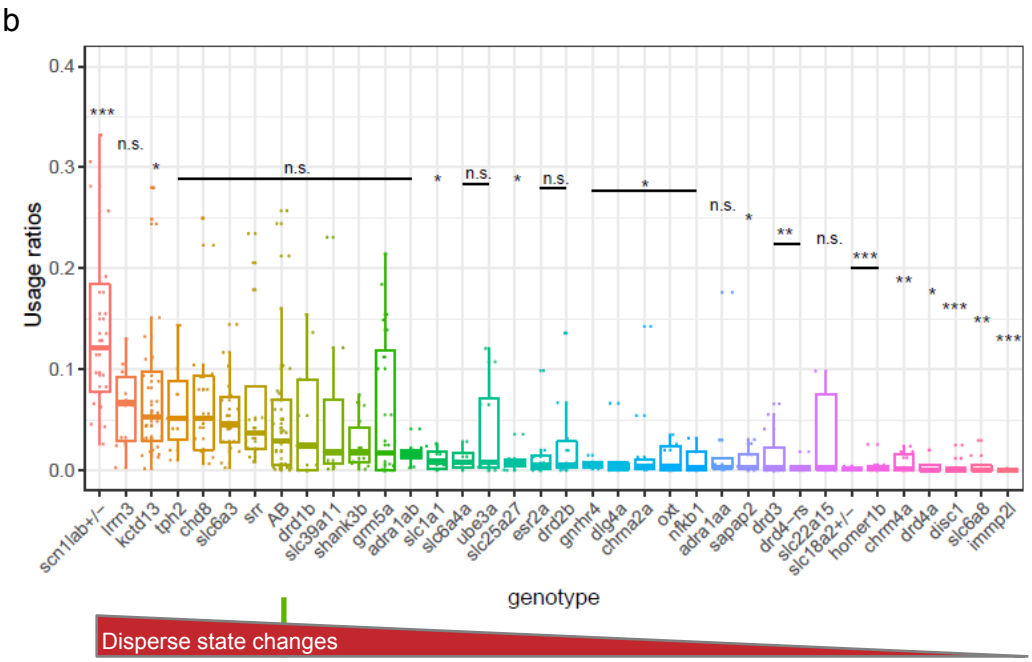


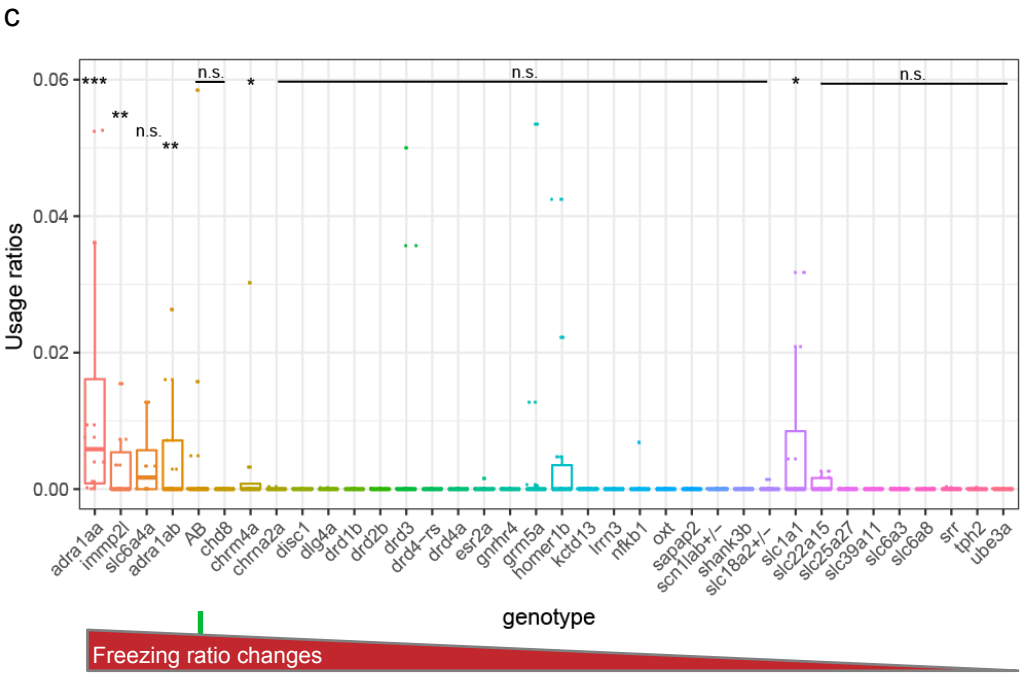
Supplemental Figure 4　Block duration of states (Fewer than 1.2% last longer than 4 seconds). The mean is 0.78±0.96 seconds.



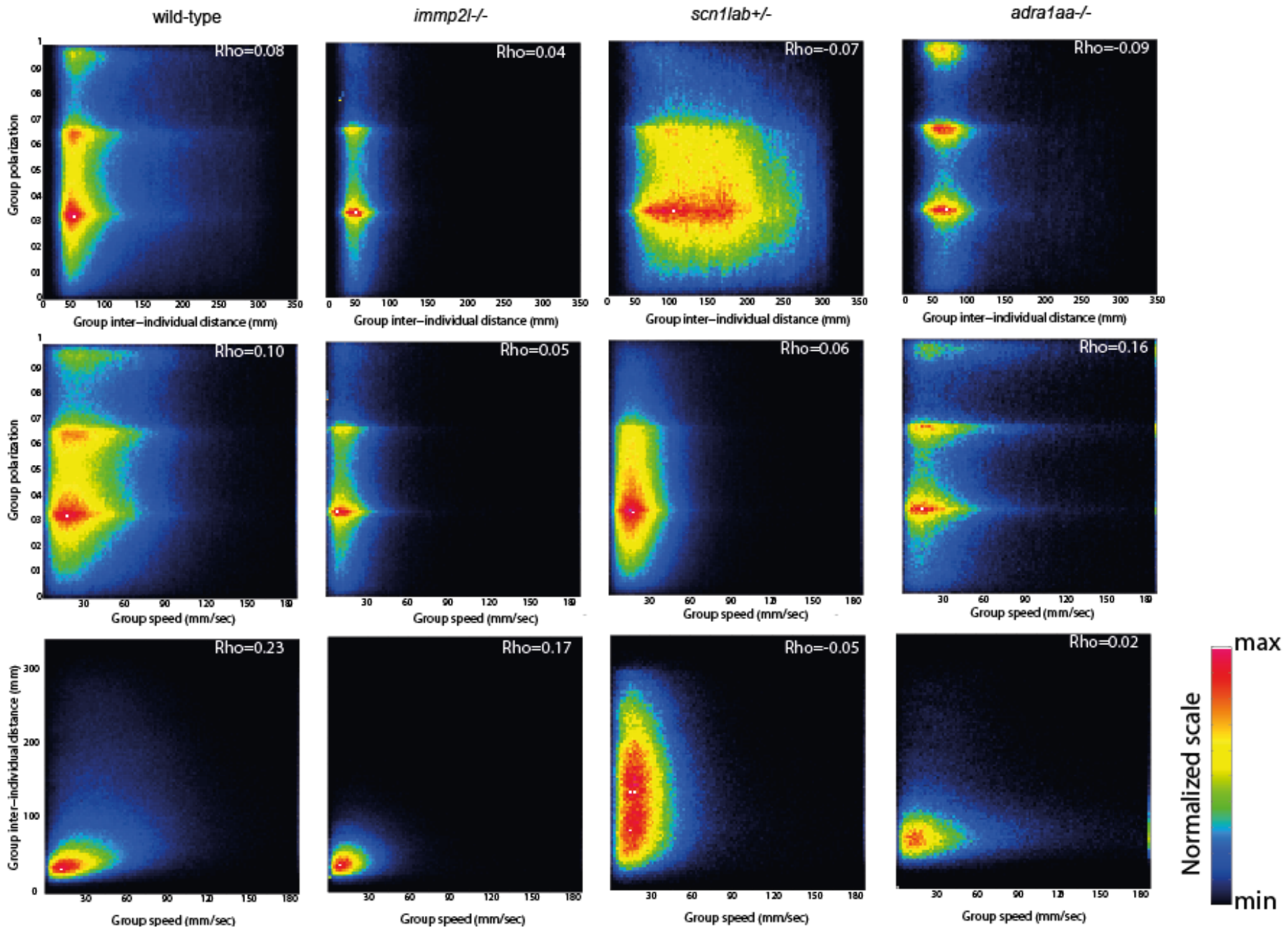
Supplemental Figure 5 Cross-validation by multi-class Support Vector Machines confirms robust phenotypes revealed in highlighted mutants. Confusion matrix computed using features, including state usage ratios, transitions, *etc*. *immp2l-/-* (n=9 groups, Prediction Accuracy (PA) = 0.95; Leave-One-Out); *scn1lab+/-* (n=22 groups, PA= 0.88; Leave-One-Out); *adra1aa-/-* (n=8 groups, PA= 0.92; Leave-One-Out).



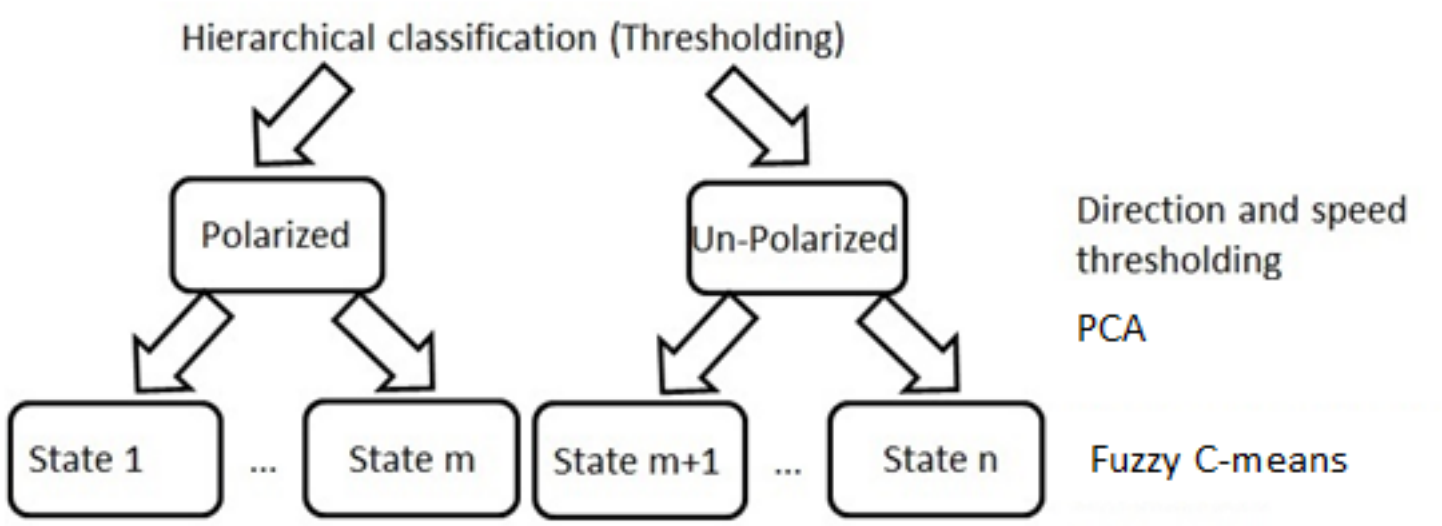




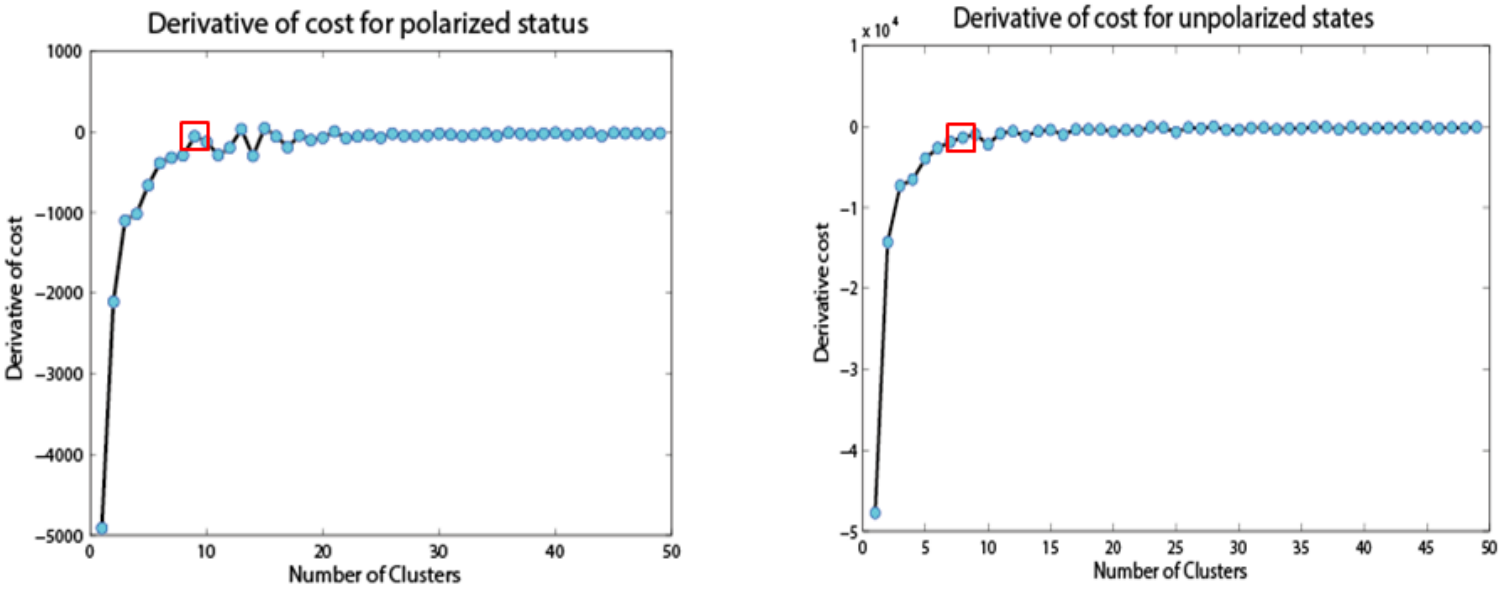
Supplemental Figure 6 Ranked usage ratios of fish having each mutation, compared to: (a) *immp2l-/-*'s tendency to cohesion; using the usage ratios of the top two tight cohesion sates which are significantly enhanced in *immp2l-/-* compared to wild-type (n=9 groups, p<0.001; Mann-Whitney U-test). By this comparison, significantly more cohesion is also demonstrated by *drd4-rs-/-* (n=6 groups; p<0.05; Mann-Whitney U-test), *disc1-/-* (n=9 groups; p<0.05; Mann-Whitney U-test), *homer1b-/-* (n=11 groups; p<0.05; Mann-Whitney U-test), and *gnrhr4-/-* (n=9 groups; p<0.05; Mann-Whitney U-test). (b) *scn1lab+/-*'s tendency to disperse, using the usage ratios of the top two dispersed states significantly enhanced in *scn1lab+/-* compared to wild-type (n=22 groups; p<0.001; Mann-Whitney U-test). By this comparison significantly more dispersion also is shown by *kctd13-/-* (n=28 groups; p<0.05; Mann-Whitney U-test). (c) *adra1aa-/-*'s tendency to freeze; the ratio is calculated from the duration of simultaneous freezing of 3 fish in the last 10 minutes of the recording (n=8 groups; p<0.001; Mann-Whitney U-test). This tendency also is shown by, *immp2l-/-* (n=9 groups, p<0.01; Mann-Whitney U-test) and *adra1ab-/-* (n=10 groups, p<0.01; Mann-Whitney U-test).



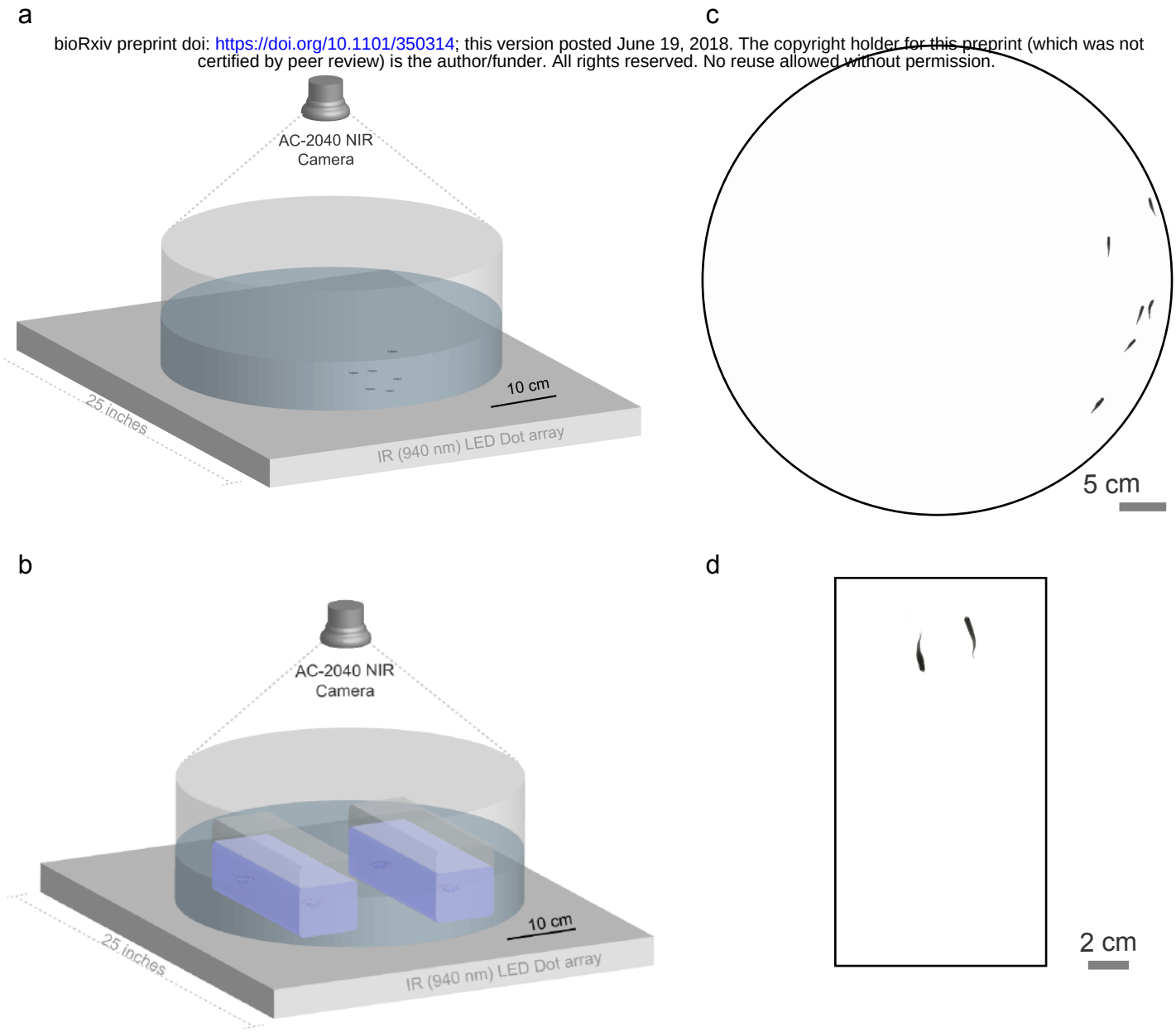
Supplemental Figure 7 Lack of relationship between group polarization and inter-individual distance (top panels) and speed (middle panels) or between inter-individual distance and speed (bottom panels), for the mutations discussed. The colors indicate the number of frames in each bin (100 bins in each dimension).



b

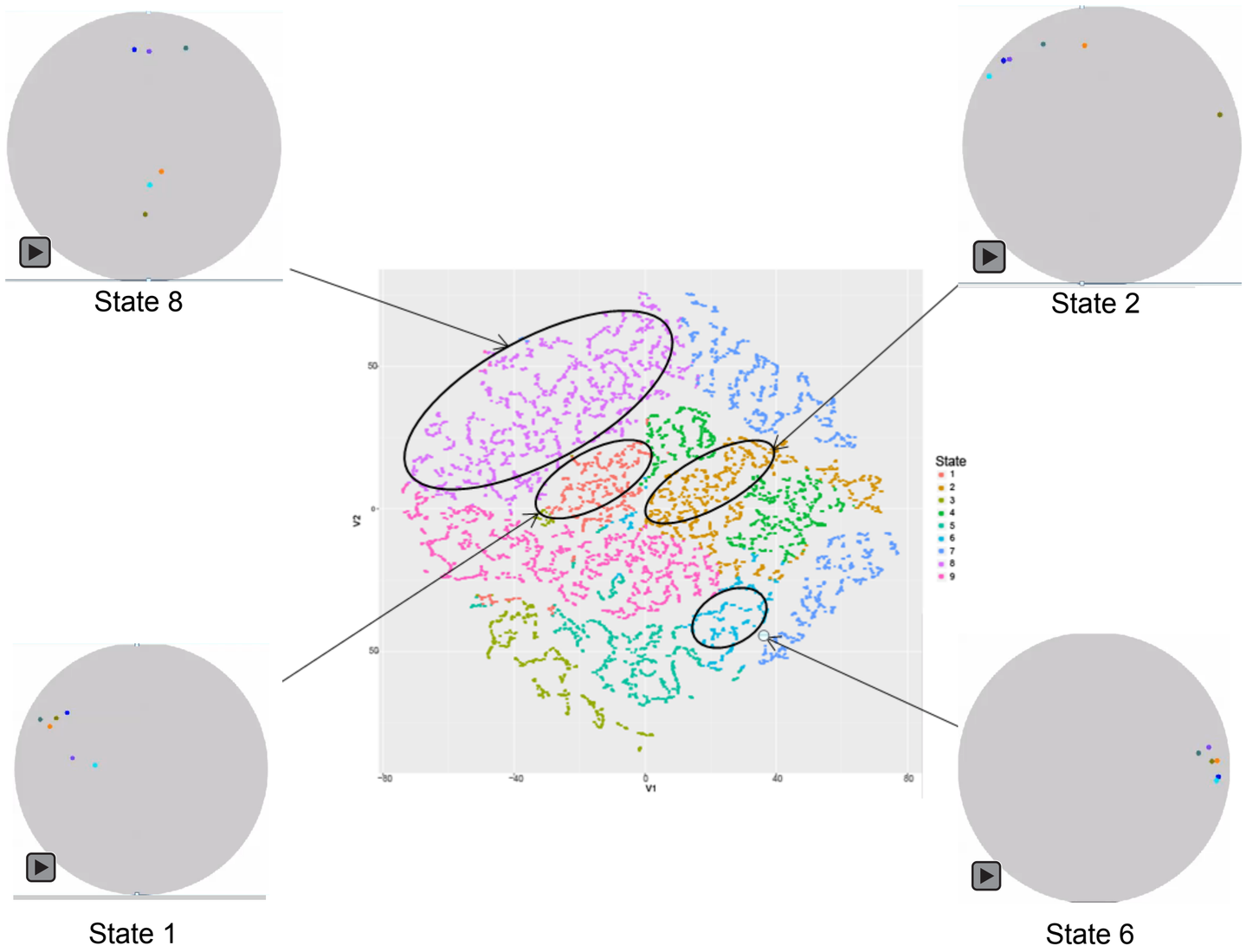


Supplemental Figure 8 (a) Hierarchical clustering flow chart of collective behavior unsupervised learning (b) The derivative of cost of the clustering for porlarized and unporlarized states, respectively. The red square boxed mark the cluster number we chose.

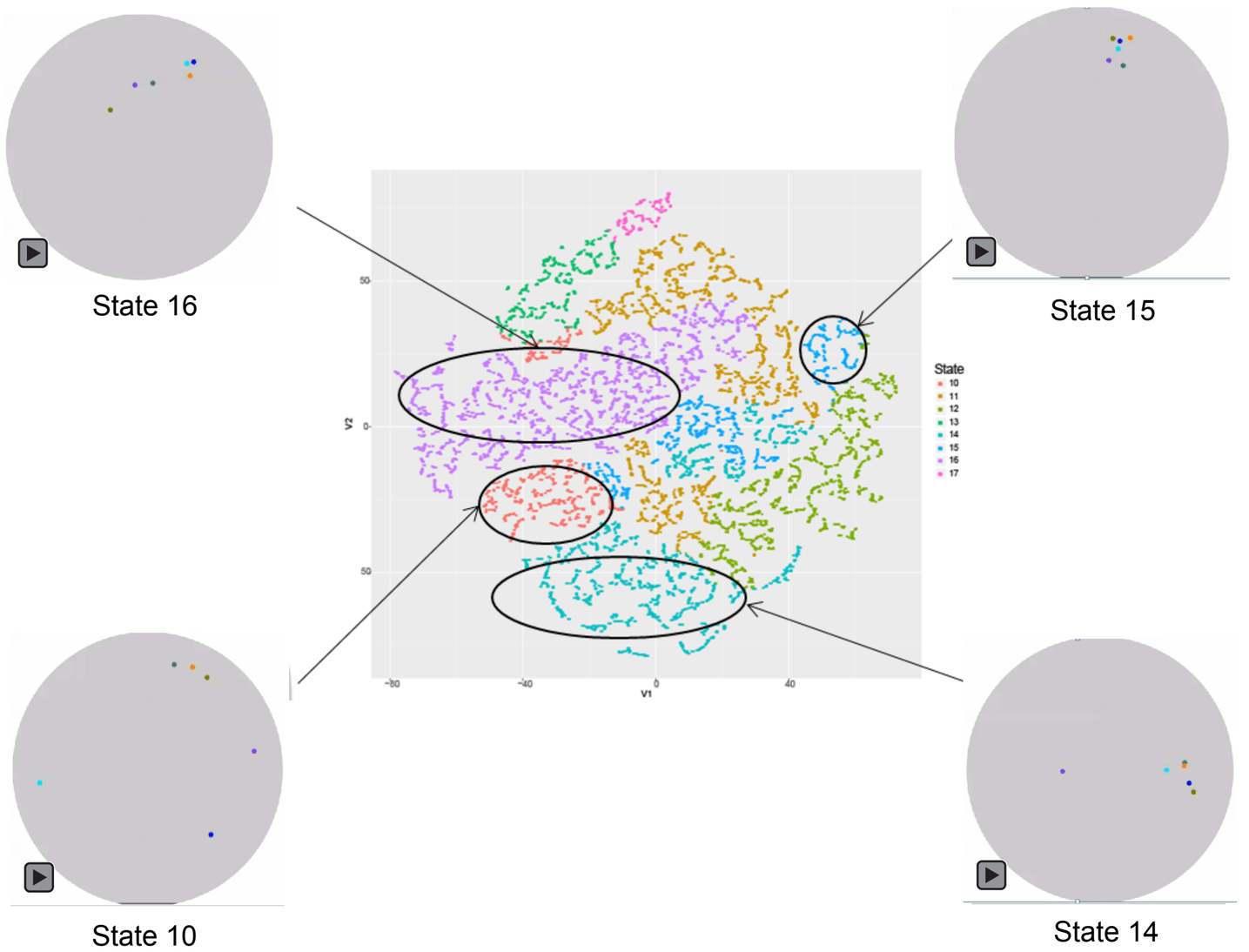


Supplemental Figure 9 Behavioral assay setup (a) Collective behavior assays and (b) courtship behavior assays were recorded from Top View, with bottom IR (940 nm) illumination. Examples of background removed images are shown for (c) collective behavior assays and (d) courtship behavior assays.

a t-SNE plot for polarized states



b t-SNE plot for unpolarized states



Supplemental Figure 10 t-SNE plots for (a) 9 polarized collective behavior states and (b) 8 unpolarized collective behavior states. Different colors in the plot represent different state clouds. Each plot shows the clustering of 5000 epochs (1 epoch=5 frames of recorded images) from polarized and unpolarized states, respectively. Videos show the collective behavior pattern of each corresponding state and can be accessed by clicking where shown. The colored dots in each video represent the centroid location of each individual fish.

Supplemental Table 1

Zebrafish gene	Human gene	Description
adra1aa	*ADRA1A*	Adrenoceptor Alpha 1A
adra1ab	*ADRA1A*	Adrenoceptor Alpha 1A
chd8	*CHD8*	Chromodomain Helicase DNA Binding Protein 8
chrm4a	*CHRM4*	Cholinergic Receptor Muscarinic 4
chrna2a	*CHRNA2*	Cholinergic Receptor Nicotinic Alpha 2 Subunit
disc1	*DISC1*	Disrupted In Schizophrenia 1
dlg4a	*DLG4*	Discs Large MAGUK Scaffold Protein 4
drd1b	*DRD1*	Dopamine Receptor D1
drd2b	*DRD2*	Dopamine Receptor D2
drd3	*DRD3*	Dopamine Receptor D3
drd4-rs	*DRD4*	Dopamine Receptor D4
drd4a	*DRD4*	Dopamine Receptor D4
esr2a	*ESR2*	Estrogen Receptor 2
gnrhr4	*GPR150*	gonadotropin releasing hormone receptor 4
grm5a	*GRM5*	Glutamate Metabotropic Receptor 5
homer1b	*HOMER1*	Homer Scaffolding Protein 1
immp2l	*IMMP2L*	Inner Mitochondrial Membrane Peptidase Subunit 2
kctd13	*KCTD13*	Potassium Channel Tetramerization Domain Containing 1
lrrn3	*LRRN3*	Leucine Rich Repeat Neuronal 3
nfkb1	*NFKB1*	Nuclear Factor Kappa B Subunit 1
oxt	*OXT*	Oxytocin/Neurophysin I Prepropeptide
sapap2(dlgap2a)	*DLGAP2*	DLG Associated Protein 2
scn1lab	*SCN1A*	Sodium Voltage-Gated Channel Alpha Subunit 1
shank3b	*SHANK3*	SH3 And Multiple Ankyrin Repeat Domains 3
slc18a2	*SLC18A2*	Solute Carrier Family 18 Member A2, Vesicular Monoamin
slc1a1	*SLC1A1*	Solute Carrier Family 1 Member 1, Sodium-Dependent Gl
slc22a15	*SLC22A15*	Solute Carrier Family 22 Member 15, Organic Cation Trar
slc25a27	*SLC25A27*	Solute Carrier Family 25 Member 27, Mitochondrial Uncou
slc39a11	*SLC39A11*	Solute Carrier Family 39 Member 11, Metal Ion Transport
slc6a3	*SLC6A3*	Solute Carrier Family 6 Member 3, Neurotransmitter Trans
slc6a4a	*SLC6A4*	Solute Carrier Family 6 Member 4, Neurotransmitter Trans
slc6a8	*SLC6A8*	Solute Carrier Family 6 Member 8, Neurotransmitter Trans
srr	*SRR*	Serine Racemase
tph2	*TPH2*	Tryptophan Hydroxylase 2
ube3a	*UBE3A*	Ubiquitin Protein Ligase E3A

Associated psychiatric disease	Rererence PubMed ID
autism, schizophrenia, ADHD	19204725, 22037178, 19918262
autism, schizophrenia, ADHD	19204725, 22037178, 19918262
autism	22495309, 22495311, 24998929
autism, schizophrenia	23490763, 14606695
autism, epilepsy	19204725
autism, schizophrenia, asperger syndrome, psy	15386212, 17579608
autism, schizophrenia, cerebral degeneration, h	25549968
autism, schizophrenia	18205172
autism, schizophrenia	22559203
autism, schizophrenia	25224105, 19897343
autism, ADHD	15892149
autism, ADHD	15892149
autism, asperger syndrome	19598235
autism, fragile X syndrome, epilepsy	22542183
autism	22558107
autism and Tourette syndrome	17043892
autism, schizophrenia	22596160
autism	20678249
schizophrenia	27992301, 27759212, 22479419
autism, asperger syndrome	27242401
autism, schizophrenia	27271353
autism, epilepsy	12610651
autism	26886798
schizophrenia	28094812
autism, schizophrenia, OCD	19360657
autism	22843504
autism	23116158
autism	22843504
autism, ADHD, epilepsy	23979605
autism, schizophrenia	23583772
autism, epilepsy, dystonia	16601898
schizophrenia	19483194
autism, schizophrenia	15768392
autism, epilepsy	11543639

Supplemental Table 2

Zebrafish gene	guide RNA 1
adra1aa	GGTTCCTTCCAGCCGAATAGAGG
adra1ab	GCAGCACAACACGTCAAGAGCGG
chd8	GGATGTTGATGAAGGCAAAGTGG
chrm4a	GCCAACTTCAGCTGTGACAGTGG
chrna2a	GTTCCAGCAGGAGACTGGAGTGG
disc1	GATTGTTTGTGTTCAGTTCTGGG
dlg4a	GGATGTGATGCATGAGGATGCGG
drd1b	TGGTCAGAATGAGGAGTGAGAGG
drd2b	GTCATAATAAGTGTCATTGTAGG
drd3	GCTAAACATTGCCATAAGTGTGG
drd4-rs	GCCCAGTATAGACCCAACGGCGG
drd4a	GTCGCTGGTCCACGTGTTTGCGG
esr2a	GGAAGTGGACTCAGGCCGTGTGG
gnrhr4	TGGCAGCCCACAACACGCCGAGG
grm5a	CCATCAGCCGCCAGCAGACAAGG
homer1b	AGAAGCAGCCAGAATAGCCAAGG
immp2l	AGATCATCAAACGGGTCATCGGG
kctd13	GGGTGGTGTTGGACCGATGCGGG
lrrn3	GGAAAGTCGGGGGTTGTTTGTGG
nfkb1	GGGGCTTCAGGTTCCGCTATGGG
oxt	GGAAAGGCCTGCGGTTATGAGGG
sapap2(dlgap2a)	GGAGACTCTCGCTCAGACAGCGG
scn1lab	GGTTAATCTCATCCTGGCTGTGG
shank3b	AGGAGCCCCGGGCCCAGATGG
slc18a2	GGACGACGAGGCTGCTCAGATGG
slc1a1	AGATGCTGGGGAAGAAAGAGCGG
slc22a15	GTCTGGAATGGATTTAGAGGAGG
slc25a27	TGGTTTTGGTAAGATCCAGGGGG
slc39a11	GCCTGGAAACATGCTGTCTGTGG
slc6a3	GGAGTACTAATTGAGGCCATCGG
slc6a4a	GATGATGAATCAAGAGTACGGGG
slc6a8	CGCCCTGAATATGGTCCTGGTGG
srr	TGTCGTACCCTATCCAGATGTGG
tph2	GGTGAGAGATATCCAGCCACAGG
ube3a	GGACCTGCCACCCACAAGGGAGG

guide RNA 2	guide RNA 3
AGAGAATCACCACTAGAGGGAGG	GTATTGTCGTGTGTATGTTGTGG
GATCCACAGAGATTACGCAGAGG	GCAAAGGAAGGGTATTGCAGCGG
TGGCTCATGTGCAGCGGTTGAGG	CGAGAGCCCGTACAAGACGGTGG
GGATGGAGAGGAGATGGATGTGG	AGACAAATGCAGAGAGGTTGCGG
TGTCCTCCTCATCATGCAAATGG	TGGCGAGAAATGGCTGACAGAGG
GATGCCATGGAAAGCCGCCACGG	GACGCACAGATTCAAGATGGAGG
GGCAATGTGTTGGTGTGTATGG	GACAATGAGGTAATTAGTGGTGG
TGAGCATTCATTCACCCCTGGGG	GACCACCACCAACTACCTGGTGG
AGACCCAACGGCGGCCCACGAGG	AGCGTTCGGTCAGCACACTGAGG
GGCTAAGGCCAGGATCCACGTGG	TGTTGATGCCGAACATCACGGGG
AGAGGATGAGTTGAAGATGGGGG	AGGGCGATGGGATGCAGATGGGG
CGGCGCCACACTTGCGCTCGTGG	GCAGTATGGCATCCAGCGGGTGG
GTCTCTAGAAAAGATGGAGCTGG	
GGGTGTGAGGGCCCGTCTCACGG	GGACTTGCGGTTCTTCTCGCTGG
GGTGCTCGATGGAGAAAGAGGGG	GCTGAAGCCTGTCACACGTGTGG
GCGGAGAAGCAGAGATGTGAAGG	AGGGTTTGCTGAAGAGGAACTGG
GAGGAGGCTTTCCAGGTGGTTGG	TGCATGTATCATTAGTTACCTGG
GGCAGATCTGGAAAGAATGGTGG	GCCCTTGAAACTGTGGCAAGGGG
GGGCCTGAACCAGCGGACTGAGG	GGAGAAGATGAAGACCAGTGCGG
AGAGAACGGCAGGCTGGTTGTGG	TGGAACATTTAGGACTCTGGTGG
GTCAACGGGACCCTCAACGGGGG	TGGGTTTTGCAGTGGGGCTGGGG
CGTACCCTATCCAGATGTGGTGG	GATGTGGTGGTTGTTTGCTGTGG

guide RNA 4	guide RNA 5
GAAACGCGAGGACTGATCAGCGG	CGCCAAGACTTTGGGTATCGTGG
GGTTCCCTCCATCCAAATAGTGG	GTACTGTAGGGTGTATGTAGTGG
GCCCAGTGGTGTGTGACCTGTGG	GAAGGACAGGATCCAGGCAGAGG
TGAGTCAGACAAATGCAGAGAGG	GCTCCAGCATTTGGGACATGTGG
GTTCTCCACAGCAGGCGACTTGG	GTCTGAGCCGATTAACCAGCTGG
GCGAAAGATGACACCCAAGGTGG	
TGGCCACATTAGTTATGCCCTGG	
TGCCCAAGGCATGACCAGTGAGG	
GCCGCTTTATGTCTACTCTGAGG	
GGGCCTGAGGAACTGGGAAGAGG	
AGGGCAGAGTGCTGAAGTCGTGG	TGTGCTGTACCCCAGCAGAGCGG
TGAGCGGGTCTGCGTTGATGCGG	GGACACTCTGGTCTCAGCCGAGG
GCACTTGTGGGTAGGACTTGCGG	
GGACAGTCGGACTGGTTTTGAGG	
GTGGTGGTTGTTTGCTGTGGTGG	GTTGTTTGCTGTGGTGGTGGAGG

PCR primer F	PCR primer R
TATGAGCTTGTGCGTGATTTCT	TCCCAACATTGACAGGTATTGA
ATCCTCCTTCCTGGTGTCATT	CCTTCGATCCGTGACTCATTAT
TGTTCTACTGTACGTTGCCAGC	CAGGCCGAGGCTGATGGAAGAGC
GGTTGGAAAGGAACTGGATGTAAC	TCATAATGAATGTCATGCTGTTTCC
AAAGTCTTCATCCTCAGCCCT	TCCATTGTCATTACCGTTTTTG
TGCATTTCCTCATTTGCAAGTTGG	GCTTGATTTTGATTGGCTTACC
CTATGACGATTGTTTCAGCCCCAA	TGCAAGATCCAGCAGTTCTAGAAA
AGGAGGAGATGGCGTATGTC	CAGCATAACTACGGACAAACCA
TATTTGCATACTCACAGGCTGGG	AATGACAACTCAAGATATGATGG
CTCAAGTCACCAAGCAAGCA	TTGCAACATAGGTCAAATGGTC
TGAGGGCGCTTGTTCATATT	GCTGCCATGAATCCATCTTT
AAGGGGCAAACTGAAGGACT	CCAACATGGACACAGGATCA
AAGGCTTGTGAAGGGTGAGTAAAC	GAAGTTCACCATGACACACATCAAC
CCAGAGGGTTGAGAATGGCCGAC	CCTCACTGTACAGCAGGGTGTCG
GGTTTTTCATTTGCTGGTGTTT	GGATGGCCACAGAGCTGGAT
CACAAAAAGAGAGTCAGAGGCAAATAC	ATTTGCAAAGGTAATCAGGGTGTT
CCTCCAATAGATAAGACCCATCTGG	TTTCAAGCTTTATTTTTATGGTGA
TGTAAATCCACCAATCATGCTTCTT	AAACATAGTAAGTCTAAGGGTACACGCC
TCAAGCTGCTCTCAGACAAGTGC	CCATTTCTCGTGAATGCACCTGC
CCGAGTTCAGAATACACGTTCAC	CCATATCAGTGCATCCCTAGCAT
AGGAACGGTCATAATAGGTTTCCC	TATCTAGCGCCTTTACAAAGTTGG
ATGCACTACAGTTCACATCAGGAC	AGTGAGTGATATGATCTCCGTGGT
AAACTCTTCCTCTTTCTGTTGAGC	TAGGCAAGTCTTCATCACTGACTC
TTTGTCTTCAGAGAGCACAGACTC	TATGGAAAGGCAGTTTAACCAGAT
TTTTCAGCACGGTGGTTATG	TTAAACTAGCGGCTGTACCTG
CCTAAAGATGGGCTATGTCTGAA	GGGATTTTACAGTGTCATACTGG
CAGTCTCATCCAAACAACTTCG	AAGACAACGTTTGACCCTCAAT
CTGGCCATGACTCACCATTAT	AAGTTCACTTTATCAGCCTGTGC
TGACACCAAATACTCCTCAAAA	CTGTGGTGTTTCATTGTCCTGT
CACTGTAATGATGGAGACGAGTGG	ACTTGGGGAAATGTTCATCGTAGG
ACTCTGTTCCACCCCTTAACC	ATTTGACCTTACCCCCTCCAT
GTTCTCCACCGTGTCTGTGAC	AGATCCCTCGCACCTGAAGA
TACCTCCTTCTGGTTCAACTCC	TTACTCTGACTGTGCTGGTGCT
CAGAAGTGGTGAAGGACAATGACA	ACAGGTGTAAAGCCAAGTCAAACT
ATGTCTGATACAGGAAGCGAGCG	CTGCTGGCCTCACTAACACCACT

Reference sequence

TATGAGCTTGTGCGTGATTTCTGTGGACCGATATATCGGCGTTAGCTATCCTCTACA
ATCCTCCTTCCTGGTGTCATTCTGGGCTTTTCGACAGTGGATACGTAGAGTCATTCC
TGTTCTACTGTACGTTGCCAGCAAAAGTGTGTGTGGATGTTTGTTTTGTGCTATAACG
GGTTGGAAAGGAACTGGATGTAACATTCCCCAGGTGCTACGGTTCGGGCTCCGACG
AAAGTCTTCATCCTCAGCCCTCAGATGATCAGCGATGTACCGCACCCCTTCCAGAGC
TGCATTTCCTCATTTGCAAGTTGGTTTCATTTCTCAACAGCTGTATTGCACGGATCAC
CTATGACGATTGTTTCAGCCCCAACTTCAGACACGCTTTCCGTCAAGCATTGACGCT
AGGAGGAGATGGCGTATGTCCGATTAAGGCTGGAGTCGCAGTTATCTGGGGGAAGT
TATTTGCATACTCACAGGCTGGGGTTGGGGTGTTGCTCAATGCTGTGTTCTTTTCTTT
CTCAAGTCACCAAGCAAGCAGCATGTCATCTTGAGCAGTGAATTGTTTACCGAGAAT
TGAGGGCGCTTGTTCATATTTATGGCTTCATCTTGTCATAATTCCTCTTCTTGTAACA
AAGGGGCAAACTGAAGGACTACCATAAAATATGGGTACAAACAGTTTAAATTCATGTT
AAGGCTTGTGAAGGGTGAGTAAACAGTGGCAGAAATTAAATTATGTGAACTATTCCT
CCAGAGGGTTGAGAATGGCCGACTGTCTGTCTAGGCTTATGACCACGGTCACAAATC
GGTTTTTCATTTGCTGGTGTTTGTGCTGATTGGGAACAGATTGGGTCCTACCCGGAT
CACAAAAAGAGAGTCAGAGGCAAATACTTTTTCACAACAATGCATCATTTGTCAGCTG
CCTCCAATAGATAAGACCCATCTGGTCTCTGAAAACTGTCCCTAATGCCCCATAATCC
TGTAAATCCACCAATCATGCTTCTTAATTTGATGTAGCGGTTAGCTAACTGAACTAAC
TCAAGCTGCTCTCAGACAAGTGCGGAACCTCAAATTTCTGGACCTTAACAAGAATCC
CCGAGTTCAGAATACACGTTCACACTAGCTAAACGTACCATACTATGACGTCAAACGA
AGGAACGGTCATAATAGGTTTCCCTTACCATCTGCAGTGTATGCCGTGTGGCCCCGC
ATGCACTACAGTTCACATCAGGACAGCTTGCGGGACGACGGCTCCACTTCTCATGGT
AAACTCTTCCTCTTTCTGTTGAGCCTCCTCTATAGTGGCCTGGTTCTGCTCATCATAC
TTTGTCTTCAGAGAGCACAGACTCCGCTCATGCAAGCAAACCACGTTGTGGGGTGGC
TTTTCAGCACGGTGGTTATGTTGTTTCCTCTATTTCATCTCTCTAGTGCCTATTATCCC
CCTAAAGATGGGCTATGTCTGAATTCAAACACACTAAAATGTGTTGTATTAACCCAAC
CAGTCTCATCCAAACAACTTCGTCCTTTGTATCACGTGGTTTCATTCCACGTCTGCGA
CTGGCCATGACTCACCATTATAACAGTATTCATACTGAATAATTATTACTAATTTAGCT
TGACACCAAATACTCCTCAAAATATCTTCTTAATTCCTCTAAAGAAACGAGCAGATTTC
CACTGTAATGATGGAGACGAGTGGCTGGTCACACTGTCTGAAAACACATCAGTCTCT
ACTCTGTTCCACCCCTTAACCCCCTTCTCCAGGCTCTCTATCCCCTCATGGGCACGA
AGATCCCTCGCACCTGAAGAGGAACAAAAAGGGACATCTTGGTGTGAAAAAAACTG
TACCTCCTTCTGGTTCAACTCCATAGATCTTTGTGTCTTCACAACCAGACAGTTTGAT
CAGAAGTGGTGAAGGACAATGACAGGATTGTGTGTGTGCTAGTACTCACGGCTCTGC
ATGTCTGATACAGGAAGCGAGCGGCTGCAAAGCATTTAATTGAGCGCTACTACCACC

GTACCCATCTATTGTGACAGAGCGACGTGCATTGCTTGCGCTCATAGCTTTATGGG
CTGCTCACTTAACCCATTCGTCTCCCTGCTCTTACTCATTGCACGAGTCTTGTGTCG
GTTTTACATCACATTTTATTTCAACTGCACACAAGCAACAAAAGCACAGATTCTGAAT
ATAAACTGCCAGAACAGGATAGCGGGAGCCCACAGGATGAAGGACAGGATCCAGC
CTCGCAGGACACACGGTGAAAACAGCAGACCAGGCTCTGACACCAGGTTACCAAGC
ACTAATTTTATTGCTCAGTCACATGAATGCAGCTGTCATGGGTCATTATTTGTTTAC
GAAGCCCCGTGTGAATGGGGCGTTAGACCAAAAAGCGTGGAGATGAATGTGTGCA
TCGCCATAGGTACCGTTAAGCTCTGTATAACTAGTCGTCTGGGCTTTGTGCCAATT
TTGTTTGCATGTTTTTGTAAGCTTTTGTGCATATCAGAATATTTATTCTGGTATTTGGC
TTTGCAGAATATAAAATTCAGAGATGTCAGTAGACAAACACAAGGATTCAAACAGAT
TATGGGCAAGAACACAAACACGTATTAGCATATATTACACACACAAAGAGCCATCCA
TAAATTTCAGTCTTTGTTACAGTAAAATATTAATAATAATAAATCACCTCATCTTTTTG
TTAAGAACATTCCATATGGAGTATTAAGCTTAAACCTACCTGTTTTCCTCTTCAACAT
GCACAGGAGTACATAGCCACGAGCTTCAGAAACATTAACAATCTGCAAGCCAGATC
TTTAGCCGAGGCCCAGGCTAGTGAGCGACGAGTTTTGGCCCACATTCCTGGAGAC
GCAAAAATTACAATGGACGCGAATCACATTTAATCTTTTCCTCTTCTCAGTTCGCAGA
CACTGAGCTCAGCAGCATCCTCAAATAAACCGCTAGCAGGTCTGTCATCACTATGCA
AGATGCAGAAACTAGCTCAGAGATGCACATTTAAACATGTGTCCATTTGTATTGAAG
CATCGAGAGGATCCAAAGGGGTGACTTTGTCGACATGATCCATTTGAAAGAACTTG
AACCCGGGTGCGGACCAAAAGTGCTAGTATGAAAGCACCCTAAATGAATGAACAAA
GGGACCGCGGACGCTGTTTCGGCCCCAGTATCTGCTGTGGTGAAGGCATCGGCTG
TGGAGGAGGTGGCGGAGGAGGGGTTAGCATTGTTAGCGGGGGTGGCGGATCCGG
GCCATGGCCACCACAGCCAGGATGAGATTAACCAGATAGAAGGAGCCCAGGAAGA
GGCGTGGTCGCCCCGCCCATCTCAGGGACCCACTGCTTAAACAGTCATCAGACAG
CAAGTTACCTGTACACGGTGGACGACGAGGCTGCTCAGATGGTTAAGAATCACTCC
CAATGAGTCAATAATGGGCAATAATTACAATTCTTTTCACCCAACTGTTGGGTTTATC
AAGGGAAGCCAGATTTACAGCGGAGCATCTGAAACCTCGACTGCTATCAAACCCCA
TTTTTTTGCACATGCAGCTAAAGATTAAGTTAATTTAGATTTTGGTTTGTGCATTGAT
GAGTGAACATTCTCATTTCTCAGAGTCTCTGCATGATATTAATGTGCATCAGCACAG
TAACGCAGCATCAGCTTGCTGGAGAAAATGCTGCAGCTCACCGTAAAACCAGGCGA
TATCAGTCAGTGCAAGAGGACCACCATCTCAACCCTAACAGCAGTCCTCACCATGG
GACTTAATCCAAAAACAATAAGAGATCAAATATTTATTTTTGACTGCTGATTGAAGAA
GGCAGCGGCCACACCAGAAAGTAACCCTCCACCACCACAGCAAACAACCACCACA
GTGTATAGAGAGGGTCTGTGCTGTGACGTATATACTGAGTGCAATTAAACACTCGA
CAGTTAACAGAGGGCTGTGGAAATGAGGCCTGCACGAATGAGGCTTGTGGCTCTTC

TACTGTCCATCACTATTTCCATCGGACCTCTATTCGGCTGGAAGGAACCGGCACCG
GGCCACTACATACACCCTACAGTACATTGCCAGAATGACTGCCAACGGAACATAGA
TCTATTTTTCCTTTTTACAGCCAGTTGTGTACTACCTGGTGAAGTGGTGCTCTCTAC
GCAGAGGCAATCATGAGGCCGGCCATCTTGGTGGTTCTGCGTGTGGGGTAGCTCA
CTTAGCAGGATGCTGACAAAGATTCACCACATGTCCCAAATGCTGGAGCTCATACTT
AAAAATGCCATTGTGAAAGTTTGCATGTCGGATATCCAGCTATAAATTACTTCTCAC
CATGTTGTACTCACTGCTGGTGAGGTCTGGTGGGTTGTAGTTGTTGGTAAGGTACA
TAGCTGCACAGGGATGAAGGAGATAAGGATGGACAGCGTCCATGCCACGCTTATC
GATTATTTTAAGTCTGCTTGATGACCAAATAAAACTTATTTGACAAAAAAGTAACATA
GTACAAAGATGTTACACTGCATGTTGTCAAGCATTTATTTTGCAAAAGTCAAAATCT
TCTCCATAAACAGCCCCCTCTGTGTCAATCTTTCCATCTCGCAGATGTTGAATCCG
TCTCCAGGTTCATCGCCGTCTCCATCCCTCTAAATTACAACCGCAAACACGTGGAC
GATCATGTGGTGTATGTTCTGTCCAAGTCCCACTAGTAGCATGGTTCTGTTGAAGT
TCCCGCCAGCCACTGAACTGTGATGTTCCAAGCTGCATCCACTGGCATCACGATGA
ATCATTATCGGGGCACTTTTCTCCGTCCACCATCAGCCGCCAGCAGACAAGGTCCA
TAAATTTGCAGAGTACAAAGAAGCAGCCAGAATAGCCAAGGAGAAGTCTCTAGAAA
AAATGTACAATCTCATATCTCTTCACAGGAGCCCAAAGAACCCTCAGCAGAAGATCA
GGCTGGGTGGTGTTGGACCGATGCGGGAGACACTTTTCCCTGGTGTTGAACTTCCT
GTATTAACAGCATGCCAGAGTTAGTTTCAATCGACAGCTTCGCCTTGCATAACCTAC
TGAATGAATGAATGTGTGATTTACTGAATGAATGAGTCATTAAACGAGTAAATGAGT
CTTGGTCGGCTCTCCAGAAACCCTGCGCTGCCTGGAGGAGGATTTTCTCCCTTCTC
GAAGATGAACCGTATTCCTGCGAACCTCCTTGACCAGTTCGAGAAGCAGCTACCGA
ATCACAAGAACAAAGAAGATCATATAGGGCTTTCCAGCTGCACGCAGAGTCTGTAAC
TGAGCTGCTTCCACACCCACCCAGCACAGGCCCCAGCCGTCCTCGCTACCTATTCC
ATGACCCCTCTTTCTCCATCGAGCACCTTTCAGAGCATTGTGTCTTTGTATGACAAC
CACATCTGACCCAACGTTTTGGTTAAAACAACTCATTTTTTGTCATGTGAGGAGGCG
ATTTAAAGATGAAAATATCCAGTCGGTGTCATGCAGTTGGATGAATTGAGTCTGGAA
TGTAGTTGTGAATAAAGTTCATAACACATTACCTATATGTCTGTAGATAGCTGGTGTC
CTGTGGAGATCAGGATCTCCTCACCTGTCTGCTGGAGAAGATGAAGACCAGTGCG
TGCCGATGGCCTCAATTAGTACTCCAAAGAGAATTGACGTCCCCGCAGCGAAGTGC
GATATGAAAGAGTCGATGATGATGAATCAAGAGTACGGGGGAGAGCAGCAGAAAGT
GTCCTTCCTTCTCGAGTCTCGTGAGGAGGTTCGTATTTGCTTTGTTTACACGAGAGA
TCTGGATAGGGTACGACATCCAGGATCTCAAAACCTAAGCTACAATTGGATTAAATA
TAAGCCAGTCCAGCCAGGAAGTCTCGTGGTGAGAGATATCCAGCCACAGGCCTTAC
CTCCGGGTTTCCAGCGCATGGATAACAATGCAGCAGCCATCAAGGCCCTGGAGTTA

GATGATGAATCCATATGCAAGATAACTGAGGAACCCGGTTATGCCATTTTCTCAGC
AAGGAGCATGCAGCAGAGAAGATGGCGTATCCTGGATCTTCGTTAACTCTGCACAC
CCTATGAAGATGCTACTTGGGAGCTGAAGGAGGATGTTGATGAAGGCAAAGTGGA
AGGGTTTGGTCACACAGAAGTATCGGTCGAAGCTGATGATTAGCAGGTTCATGACT
CTGGAGTAAAGGCTCTGAGTCAGACAAATGCAGAGAGGTTGCGGATGGAGAGGAC
AGTTGGTTATTGAGACTAAGAAGTTTAATCTGTAATCTGTGCTCAACAGGTGGTGTG
AGGTTATTTGGCTTGGCCACTCTGAGGTAGACCACCTCTGCTGTGTTTTTCAGCGCT
ATGATGAACGCCACCTTGGGTGTCATCTTTCGCTCATAGCGAAACGGGCTTGAAAT
CTATAAACAGGCATATTGGGATGCAGCAACAGTCTGTGGTTACTTCTAGCAACCCC
GTCCACCCTCTCCAGATCTTCCACACTTATGGCAATGTTTAGCAGTGGTGAATGGC
GGTTGGCAGCTGATCTTTGGGGCATCATGGTCAATGTGACGCCCAGTATAGACCCA
CAGCGACAGATTGTCCTGCTCTCCGCCACGTGGATCCTGGCCTTAGCCGTGGCCT
CGACTCTGTTGTTGCAATGTGAGTGAGGAAACATGGCTGTGAGGTGGCCAAAATAA
AAGTCACCAGCAGGTCGGCAACGGTGAGGTTCATGATTAAAATTCGTACATGGGA
ACGAGCGCAAGTGTGGCGCCGTCCGAGAGCAGTATGGCATCCAGCGGGTGGAGGC
AAGATGGAGCTGGCAAGTTCGCCCTCTCAGGTAAGTCGTGAAATGTCCTGCTACTT
ATCAAACGGGTCATCGGGATCGAAGGGGACTTCATTAAGTAAGTTTGTTGAACAATT
GCGGGACGGGACAGTTCCTTTGCCAGATAGCACGAGGGAGCTGGAGGAGGTGTT
CCAGAGCTTACCAAAATCGAAGCCACAAACAACCCCCGACTTTCCTATATTCACCCA
TTGATGAATGAGTGAATGATTAAATGAATGAGTCAATCAATAACAAACTAAATAAATC
CCGTGTGAGATGTCTGGAAAGGCCTGCGGTTATGAGGGACGCTGCGCTGCTCCTG
ATTCATAGGGATGGTTTCCACACGCTGCAGTACCAGCGCACGTCTGCCACCACCAC
GAGACACAACGCACATACATAAACAAACCAAATTAAAACAGTTTGTATCGTGCTTTAC
CAAAGACGTTCCAAACTGTGGGGGGAGGAGCCCCGGGCCCAGATGGGGTCATCAG
CACCACACGTGTGACAGGCTTCAGCCCGCAGATGAGCACAGCTGGTCCAATGAGCC
GAAATGAAGAAAAAGAAATGAATATAAACGACCGAGCAAAGGTTATCAAGAAAAGAG
ATGGATTTAGAGGAGGCTTTCCAGGTGGTTGGAGAGTTCGGAAGTCACCAGAAGCC
GACCCCTTGCCACAGTTTCAAGGGCCCCTCTTCTCGCACTATACCTGCAGCTGTGC
GCTCCAGCTGCAGTCAGCGCCCAGGTGAACAGAGTCCCCAGCAGGGCCTGAACCA
GTCCAACAGCGTAAACACATAGATTCCACCCTACAAAGAGAGCAGCAGAGTAAACA
GCCGGAGTCTCAAGAGAACGGCAGGCTGGTTGTGGATAGCGTTCCGGAGAAGGAT
ATGGAGAAATCCTCTCTGGACGCAGACTGCTGCGCCCTGAATATGGTCCTGGTGGA
ACATTACATTATTTGTGTTCAATATATTGCAAAAAATGGATCAAAAGCTGTTTCAAAAT
CGGTAAACCCTGACCTCTCTGTAAAGAGTCACAACCCAGACATCCATATTGAAGAGT
TATAAGAATAATGCCAAACTGTGTGATCCTCATCCCTCTAAGAAAGGAGCCAGTTCA

TTTGGGATCTTTTTACCTCCCTCTAGTGGTGATTCTCTCAATGTATTGTCGTGTGTA
CGACTCGTCTTCTGGCATGGGTTCCCTCCATCCAAATAGTGGCCCGACGGAGATA
AGAATTCAGGAAGATTGAAAGCCGCCAACCACGACTCAAGAGAACTGTGAGATACT
GAAGCGTTGCTGACCACGTAATCCAGCGCGAGCCACAGGTCACACACCACTGGG
GATGGATGTGGATAAACCAATCCCAAACTCCATATCCTGACAAACCAGACCATCCA
AATCCTTCGGGAAACCACCGCAGGAGGAGCTTCAGACGGCCCGGATACATGCGG
CCCACCGCATCCTCATGCATCACATCCTCCAAACACACATTATTCACCTGCAGGTC
GGCCCAGTAGCGGTCCACACTGATGACGCACAGATTCAAGATGGAGGCAGTGGAA
GAAATCTGGCTCCGTGTGACACAATTAATATTTCTTACCTCAAGATACACCACCCAG
TCTGGAATGACTCTGAGCATTCATTCACCCCTGGGGGAAACTACAGCCCAGCGTCA
ACGGCGGCCCACGAGGGGTACAACTACCTGGCTTTGATATGTGGCGTGCCTCTCA
CCCCCGTGATGTTCGGCATCAACAACGTCCCCAACCGGGACCACAGCGAATGCAA
AGTCGGGCTTAGCGACTGGCGCACAGACGAGCAGTCCGATAAAGGCGATGTGCT
CTTGCGCTTGTTGTTGGTGCTGGCAGCCCACAACACGCCGAGGTTGCACACAGCT
CCATGATGCACACCCTGGACCGCATCAACGCAGACCCGCTCATCCTTCCCAACATC
CAGTGTAATGTATGTGCATGAAATGATGAAAATCTGCTCAGATTATTTTAGCAACAC
GTTAAAAATAAACATACAGTTTAAATCTGATGAGGCCAAAAGGGTTTTAATGGCAGA
GAAGGAGGCGCAATACTACAGACTTCAGGGTCTTGTGCAGCACTGCCTCTCCACAT
AATGCTTTTTCTCAACTCCCAAGACTGGAGTCCTTGATGCTGAATAGCAATGCCCT
AGTAGATCAATAAGTAAATGAATCAGCTCCATAACATGAGAAGAATAAAATAAGTAA
GAGTCTGCTGCGACTCGGGTTAGTTTCAGCTCAAATTCAATTGACTCCAGAACAAA
CAACAGCGAGCAGCGCAACGAGAGTCCCGGGCGCATTCGCCACCTGGTGCACTC
CGCAATATAGATTGTTCCAAGGCAGCTTCCCAGTAACAGGAAGAATCACAGAGTCA
GATGAAAGTCGGCCTGCTGCTATGGGGGCAGAGTTGCTTAGCAAAGACACTCATTC
CTGGCTCCTACTTTTGTGAGCCCTCAAAACCAGTCCGACTGTCCCAAAGCAGATGA
GACGGAGAAAATAGAAACAGCAGCAGCAGATCATCATCACTTCACCCCTCCCTCTCT
GAATGATCACAGTGCTGGTGTTCCTCCAGGTAACTAATGATACATGCAAATGCATAA
TCAACATGCCCCTGTATTTCTGAGTCTGTACGCTTCCACCATTCTTTCCAGATCTGC
AGCGGACTGAGGCCTGGAAACATGCTGTCTGTGGAGAAAATCAGGACACACTGCAT
CGTAGATACCACCCTACAGAGAGAGCAACAGAGAGATGAAGACACAGTGATTAACC
TCAGAAATCTGGCTCTGGGCCTGGGCAAGTCTCCAATGGTTATCGCAGTACATCTC
AGGAGAAGAAGGGCCATCTCATTCCCAATGGAAACGCGCACCCAAGCGTCAACGGC
TGTACCGTTATGTACATTTTAGAAAAAAGCTTTAGGACTACAGCACCAGCACAGTCA
TGCTTATGTTTTTGTTAAAGAATAAATATATAGTTAATCAATCATTAGCAGAAAATAGC
AGCCTTCCCGGAGAACAGTGCCAAAGGAGCTCACAACTTCTCTGCTTGCAGCAACG

TGTTGTGGCGAGGAGGGAAACGCGAGGACTGATCAGCGGACAGAAAACTGAGAAG
GCAGCTGAAAGGGCCCAGAGAGCTGCCACTGCAGTCATTGCCCTTCGACCCGTAG
TGTTTTGTCTGGCTCTTTCTCTCTTCTCACTGGATCATTTTTATCCTGCCTGTATTTTT
CCAAGCGGCCAGTAACCCTTAATGATGTAGACGGTGTAAAGGTTCATGGAGAACAC
CTCCAGTCTCCTGCTGGAACCAGCTCTTGGATGTGCTCAGATTGGCTGTCTTTTTTG
TCTGAGCCGATTAACCAGCTGGACGTTGCAGAAACATCATGTGATTCTGAACATCA
ACAGCACACAACACAAAGAATGACCTTACTGTGTGCTTTTTGAAAAACAAGATTTAT
ACACATGATGTCAAAGGCCACCCAAACATCGCAGAAGGCGCCGAACGGCCAAAAAC
GGGCATAACTAATGTGGCCACGAGAAGATCTGCCACCGCCAAGCTGACAATGAGGT
AGGCGTTGAAGAAGCGAAGAGGAATTATTATGCCATGCTCTATTCCCTGCTCATCCT
ATCCTCATTATCATTTTGGGGAATGTCCTGGTATGCCTCAGTGTGCTGACCGAACGC
GCTGGAGGACAACAACTACGTTATCTACTCCTCCGTCTGTTCGTTTTTCGTGCCGTC
GTACCCCAGCAGAGCGGGACTGTAAAAGGGCAGAGTGCTGAAGTCGTGGCCAAGC
GACACTGCGCAAAGAGTGAAGGTAATTATCACCCTCACTTTTGCTGCCATGGAAAAT
ACCCTAGGCTGTGAAATCCGAGACTCCTGCTGGCATTCTGCCGTGGCTCTGGAGC
ACTAGCAACTGTTAAGTAAACACCCTGATTACCTTTGCAAAT
AGCGTGTTTTCTTGCCACTAAAATGTTGCCTGAACAGTGTTTACACAGACTAACAGG
TTACAGGTCAGGCGTGTACCCTTAGACTTACTATGTTT
TAGGGCCCTTCATCACATAACAGTGGAATCCTTGCCTAACCTTCAAGAGGTGAGCAT
AGAGCAGCAAAGTAAAAACACACAGTGGTTAAACATGTCTAGCTGTGTAACATATC
CCACTGTTATACAATGACTTGCCTAATTACCCTAACTTTACCCTAATTACCCTAGTGA
CTGTCCAGAAGCTCTTCACCAAATCGCACTCTCTTGAAGGCTCTTCCAAGATGAACG
GTGATGAAGACTTGCCTA
ACTGGGTGAGGAGCCACCAATGGGAGCACCACTTGACCCTGGCAGGAGATCGCCT
CCAGCTTCTGAACGAGAATGTGAAAGTGGGTCTGTTGTTTGCCTCTAAAGCAACAG
TACCGACAGCAGCTCCACCGTAAACCGCCTCGGTTTTTTATTATTCTCCGCCGGTCG
ACAGAGGACTGACATGTGTTTATGTAGGTAAATGTATTTGCATTTCCGATAGCACGA
CTTCACCCTGGATCTGGAGTCTGGTTTTGGTAAGATCCAGGGGGAATGTGACTAAA
TTCAATACACTCACAGAGAGAGACGTGTCTATATCTTTGTGTGTGTGTGTGTGTGTG
CCTACGATGAACATTTCCCCAAGT
CTCAAAGCCCCAAAGAAGGTGCGGGCACTGGGACGGACAGGGTCAATACCCCTGC
GACCCTCAACGGGGGGCCCATGTCAGCGGCTACCGGAGCCATCTCGGCGGTGGAC
GAGTAA
GATGAAACTCTACTTTTGTAAGTGTCTTCACTGGTGAGCGTAAAAAAAAAATGGACA
GGAAAGATGAACCATAAGGACCTGCCACCCACAAGGGAGGACTTTAGAGGTGAGTA

GTCGGATCACGCTGAGACCGTGACCTTGCGAATACACCGCGGAAACATGACGGTGT
CAAAGGAAGGGTATTGCAGCGGGTAACTGACGGCCATGCAGCGATCCACAGAGAT
TTGGTTTTCATTAAGAAACAGACGAAACCGTATGAATTATTAACAAGTCTTGCTCTTC
ACCGATGATCAGATCTGCACAGGCCAGGCTGAAAAGGAAGTAGTTGTTGACCGTC
GGTTTCTTTTCAGGCTCTGGACGTCTCATGAAGAGCCAGCGGGGAATGTGGTCCAG
CAGAAGTCCTATTTCCAAGTCGCCTGCTGTGGAGAACACTCAAAAGAGTGCTTCAG
TTCGGTATAATAACCACACACTATAAATCATTTATTAAACATGAGCAAATAGTGAATTC
CCTACAATCTCCGTGGCGGCTTTCCATGGCATCACCAAAATGGCCACAAGCAAGTC
TAATTAGTGGTGGTCTGCAGCGCTTTCTCTCGGGAGACTGCCATACACACCAACACA
TGGCTATCGTGTTTGGGAATGTTCTGGTTTGCATTGCTGTGCTTCGAGAGAGAGCTC
CTCACTCAAAACAGCAACCAACTACTTCATTGTCAGTCTGGCCGTGGCTGATCTCCT
GCCCCATCATGCTGCTCCTGTACTGCGGAATGTTCCGGGGCCTGAGGAACTGGGA
GTCTGTGTAGGGCGATGGGATGCAGATGGGGTGATTCTCCACTGGCAGAGATGGA
TGTGGGCAGCTGCAATGCCGCTTCACCTGTGTTATTGTTACACGTGGGCAAATCAC
AGAGCATTGAGTTCATCCGGGACACTCTGGTCTCAGCCGAGGAGGAGGAAGGCAT

CCTTGATATCACCATAAAAATAAAGCTTGAAA

TCCACTCCAACCCCATTTACTGTGACTGTGTCATTCGCTGGATCAATATGAACAATA
CTGCTTTCTGGAACCCTCTGATGTGTTAATCTGAACTCACTATATCAGCTCAACAGT
AAGCCTTTACATGTCACTTTAAGCTGAACACTAGTGTCTTGAAGAATATCTAGTCTAA
GCACCAAGGGAGACTCTCGCTCAGACAGCGGACACCACCACCATCACCACCATCA

TGTGGGAGGTGCCAGGTAGGACAGAGCTGCTCAAAACTGTGCCAACTTTAGCAAGC
TACAGCTAATCACTAACCCCTTCATAGGACCACTAACCAACAGGTACAGCCGCTAG
GACCGTCAGTCTCAAAGCTCCGAGCCTCGCCATATTCTCTCGCCTGCTCACAGATTA
AGCCGGCAATCTATGCAAAATGGATATATTGAGGGTCAAACGTTGTCTT
ACAAAAACATTATTAATAAAACAGAAAGTAGACAAAAAAAATATTCAAATACATTAGG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTC

GAACATTTAGGACTCTGGTGGTCCAACAGACGAGCCTTGATCCACCTAGGGAGACC
GAAGAAGAGGGAGGGCTACCCCGAGCGAGAGACGTGGACCAGACAGATGGACTTC

AGCTCAAAATTTTTTAGTTTGACTTGGCTTTACACCTGT
ACGGTGTTATGTTTGTAGACATTTTATCAGAACTAGTTTGATTAAAATTCCTTAAAGT

TCCGAGGACGAGGCGTTACGCAACCGGACACACTTCGCACTGCGTCTGCTCAAGTT
TTACGCAGAGGCTGAGGATGGATGCCGTGCAGCACAACACGTCAAGAGCGGTCCAC
CATCAGCCTCGGCCTG
TGCAGGTGTCGATTAACCTTGATTGAGAGCATGACGAGGATGTTGCCCACGACGGT
GAACACCGAATGAACCCAACATGGCATTGAATGCGTGCTAGGTGAACGATGGTGCA
AATTACTTGGCGAGAAATGGCTGACAGAGGGTTTCGAGAGAGACAACTCATCAAAG
CATTATTTATTAAGCATTAACTCTACATTAGTAGGTGTTAGTAAGCAGTTTATAACTAA
AGAAATAGCCAGTGATATAACAAAGAAGTTGGTGACTTTAGAGCGCAGATGGCGAA
ATTGCCAAAGACAATGACAAAGATGAGGAGAGTCAGGAGCATGGCATAGTAGTTATA
CTTCAGACCACCACCAACTACCTGGTGGTCAGTCTGGCAGTGGCTGATCTGTTGGT
TGCTGGCCATTCTCGTTCTGCCGCTTTATGTCTACTCTGAGGTGAGGACCAGAGATT
AGAGGCTCGGAAAGCTAAATTAAGGAGCAACATGGAGGCCTGCCGAAAGCTTCAG
GAGGATGAGTTGAAGATGGGGGAGAGGATGTGACCTCCCACACGGCCTGAGTCCA
CAACTGCCATTCAGAGTATCTGCAGTCAGCTGGTGAAACATGATGTTCTCAGATGTTC
TGGCTAAATGTTCCACCGAAGGCGGAGGGACTCCCATGAAAGGGAAGAAGCCTATT

CCAGGGTCCGCTTCATGGAACTGGATGCCCTTTTATGCGCAGGGCCTTCAGAGTTT
AAACCTGAGTTCTCTTCTCTGCAGAGGGGCTTCAGGTTCCGCTATGGGTGTGAGGC
ATATTATGTGCTGTCATCATGGAGAAGAGGAAACACATCAGCTATTAGAGATGAGTT
TCACGGCCATGACCATGGGAGCCACAAACACAGCAAACGAAGTAAGAGCAAGGAA

CACACGCACACCTGCATGAAACACTGCACAAATTTCAATATCATTTAGCAATACACC
TTTAA
AAACTCCATCATGGAGATGCTGGGGAAGAAAGAGCGGAGAAGCAGAGATGTGAAGC

CAGATATCAATGAATACAAATGTATAGATATTATTTGCTTAAAGATACATTTATTTAAT
GAGAGAGCAGATCCTGTTTATCCAGGGTGTCTGCAAATACAGGACAATGAAACACC

CTGGAGCAAAAAGATGGACTTCCTGTTGTCGGTCATCGGGTATGCAGTGGACCTGG
CATCATGTCCTGCGTGGGTTTTGCAGTGGGGCTGGGGAACGTCTGGCGCTTCCCG

TTTTTTACTGGAAATCCTCACGTATTAATAAACACGCATTTAAAACATATACCTTTCTA

TCTCAAGGGAAAAGAAGGCCGCCAAGACTTTGGGTATCGTGGTGGGATGCTTCGTA
GGCGTTACATAAATGACGACCGAACACCCAGCGACCCAAGGCTTCAGATACAGCAC

TGACGAAACTGAGAGATCCAGTGACCAGTGCGATGAAGACCATCTCCACCGTCTTG
ACGTTGAGCACAAAAACGGTAATGACAATGGA
GTCTTCAAATAAACACCATTTGCATGATGAGGAGGACAATTTACCAGTGCAATCACG
AATCTACAAATGCTGTATTCTTGACTTATAAGCTCATGTGGTTTTATACTTTGTTAATC
ACTTTGTGACTGCAGCACAGACCAACGTGTTTCCCAGTAAGGTGGTCAGAATGAGC
ACTGATGCTTGGCCTCACAGTCTGTGCAGTTGAGAGCCCCGGTCCCATTGTCATAA
CGCCTCACTGGTCATGCCTTGGGCAGTTTATTTGGAAGTAAGTACACTAGGAACTA
TGATGATGGAAGTAGGTGGTTTACAATCTCTGTTTGACTCTAAGCATGTGATGGTAT
GAGGCCGCAGCGTCGCTCCAGCCGCTGTCTGGCCTGCCGCCGCCGCTCCCGCCT
ACTTCCTGCAGCTGAAGAAGAGGACTGTCTCCTTCGGGATACTCGGACATGGTGAA
GGAGAGCTGTCATTCATCACTGAAGAGCGAAAGCGTGTGTTTGCTGCATTTGCTCG
TGTGGGTCTGATCGGGCCTGGATCCAGCTCTGTGGCCATCC

TGAAGGTCGTCTTGTCAAGCAGGTGCATTCACGAGAAATGG
GCCCGTCTCACGGAGGGCTTCCAGGGGCCTCCAGCGAGAAGAACCGCAAGTCCTA
ATTAAAATTAAAACAGAAATTGGGGAAAAACATAAAAGAATTAAGATTTAATTACTGA
CGCAAGTCAGACTCCAAGCAACGCTCAGGGATAGCAGGCTGGTGGAGCTCTGACC

CAATTGACTTTTGACACCCTTAGCCACACATCTGGTTAAACTGCCTTTCCATA

GGTTTGCTGAAGAGGAACTGGATTCTCATTGCAACCATCATTGCGGTCATTCTGGG

TTATACATATACATATTTTTTAACTACAAAGGTTAAGTAGCTATATGTTTCTGTTAAAT
ACAG

GCAATGTGTGGCGCTTTCCCTACATCTGCTACCAAAATGGAGGGGGTAAGGTCAA
TACCTGTGCTACAAAAACGGTGGAGGTAAGAACTGCCGGTGTCTCTCTGCAGCGC

AAATATAATTTGCCACAGCCATATCACCCTGCAGCCCAAGACCGGTTACTCACTGAA

ACTCTGTTGGCTTCCGTTTTTCCTGGTTCTTCCGATCGGTGAGTTGTTAGAACTTTTT
GAGAATGGAAGAACCACAGAGCTCAGAAGCAGGTCTGCGATTGCCAGATTTCCAAT

GTACGGGCTCTCGCCGTCGCCCTCAACCGCTGCACATGAGCCATTGGCGCTGCCAC

TGATGTTTTCAATTCCAGTTTCAGTTTCATTCAACAGTCGCTAGACACTAGCGATTTA
GATTCATTTTTCATTACTCAATTAAGTGTTAACGATTCATGACCAACTAAACTTCTCTC
GAGTGAGAGGAAACAACCAGTCAAAACTCGCTTTGATGAATCGCGCTGCGACACAC
ATAAGTGTCATTGTAGGGATACTCCGTGAGGAAATCCATGAGCTGTGGCCAGGGAA
CTGTAGTTTACCATGTTTAAGGAAAAAAACAATAAATTTTCCCATTCGAAAATCTGAT
ACTTTTTCCAATTGCTGAATCTTATTTGATGATATAGTCACTGCACTTTTATAAATTAC
TGTTATAGAGAAAGACATTACATTGGAGGAGCTGGACCAAGATCACTATCCCGATCC
GGCGGATGAGTTCAGAGCTGTCGCATCAGTAACGGGGGACTGTTGATGTGTGTCA
GACACCCTGCTGTACAGTGAGG

ACCCACAAGTGCAGGTAAAGTTTTTAATTATATATTTGTATTTAACATTTATAAGATTT
AAATAACATGTAATATTCATGTTTAGGTATTGTCTCTGCAATAGTTTCCAGTGGAGCT
GACAACCTGGACAGTGACAGCACCTACCGGACTCCCAGCATCATGAGTCGGCACCA

TAAGAGACACCGTTCATTTATTGAGCATGTTTCTCAATGCATATTTGTTGCTATCATT

TGAAATGAATGAATGTAAATGCATCTACCCAATTCTGCCACAGCTGCAGCACAGGCT

AT
GTTCATGTTCAGATCTTTCGGATGACACGCCGTCTGTCACTCCCCCTCCCCCTCTC

AGCTAAGCAGGGCTGAGCCTGGTCAGTACCTGGATGGGAGACCAATAGGGAACAC

TATATTTCATTATAGTCAATACCTGTCAATGTTGGGA
AAAATAATGAGTCACGGATCGAAGG

CTGTCACAGCTGAAGTTGGCATCACTGGACACGTTGCTCAAAGAGCCTGAAACACA

ATTAGATGTAAACACATGCTATTCACCCAGAACTGAACACAAACAATCTGAATCAGC
GACCACATTTCTAGAACTGCTGGATCTTGCA
CTGCTGTCCAGGACTGTGGAGTAATTCACATCCATGGCTTTTCACCGAGTCGAGTTA
GAAGCCTGAAACTGAGAGACATAATTCACTGTCAGTCAATGCAACAAACAAAGGAT
TGCATGGAAATATAAAAATGACAAACAAGTTTATATTTGAATTTGATCTATGTAAAAT
GAAAAAGAGGTTGAAAAGATGGATTCATGGCAGC
CCGACAGTCCTGATCCTGTGTCCATGTTGG
ATGGTGAACTTC

GTCGACGTTTATGTTTTCACTATTTCTCTTCATTTTTTTTTGTTTTATTACAATTTTAT
GATTGTGTGTGTGTGTGTGTGTGTCTCAGAGGGCTGCAGTGTGGACCAGTCGT
ACGGAGATCATATCACTCACT

TGTTGAGCTCTAATACGTCTCGTCGTTAATTATGTATCGTTTTATTATTGCTTCACTG

GATAAAGTGAACTT

GAGTGACGTCACAGACACGGTGGAGAAC

TAGGTTGCTGTTGGAAGTGGTGTTAGTGAGGCCAGCAG

AGCACACACCGTGATATAAATTACACAGCAGCAGATATGTCTGGATGCAGTCATTG

GTCAGGGCACCAGCTGAAGTCTAAAACTTCAAATTCAGGCTTTCTTAAACCTCCATC

GCAGTCACCCCATAATTCCAGACCAAACTAAGAGTGCTTACACATGAATCGGAGCA
GAAGGAAGCAGTAACACGCCAAGGCAAAAAGGAACGATATTGATCAGAAGCTTCCC
GTTATAAATGCAACATATTTCAAAATATGGCTAAATGTACGCTTAATTTAATTTCAGA

TTAATATTTTTCACTAACATGTTAATTTGTGTGTATAATTATTGACCTGTGATCTTCAA
GTGTGGACGGAGATGCTGACGCTGCAGCTGTCAATCAACCGGCCAACAGCCCAGA

TCTGTGGTGCTGCATGATGGCTTTACACCGCAAACAGTCGCTGGTTTGACTATAAG

AAAATAATCACAAGCTGATACGAATATAGCTGATACAAATATAGTGAGTGTATAGAA

CTGATTTAATGAATCACTTGAGCCAATCAGAAACCAGCATTGTTCAAATGAACCAATT

AACTGGTTGATCTGGTTTGTCCGTAGTTATGCTG
CCATCATATCTTGAGTTGTCATT
CCATTTGACCTATGTTGCAA

AGCTGCTGACTAATCTAATGTACATGCTAGGGATGCACTGATATGG
ATCTGCTGCTGAAGCTCCTGCACCTGTCAAGCCACGCACACCCCTCTAGAATCCAC

AGTAGGCTAATTAAAGTATTGCACATTTAATGCACGTATCTGCAGGTGTGTTGATGC

GGAAACAGCATGACATTCATTATGA

ΓAGAAACCAGAACTGTCCCGGTAAGCCAATCAAAATCAAGC

CAATGAAACGCCAGAGCTGTGACATCATCTCAGGAGGAGCCGCCCACCAACCACA

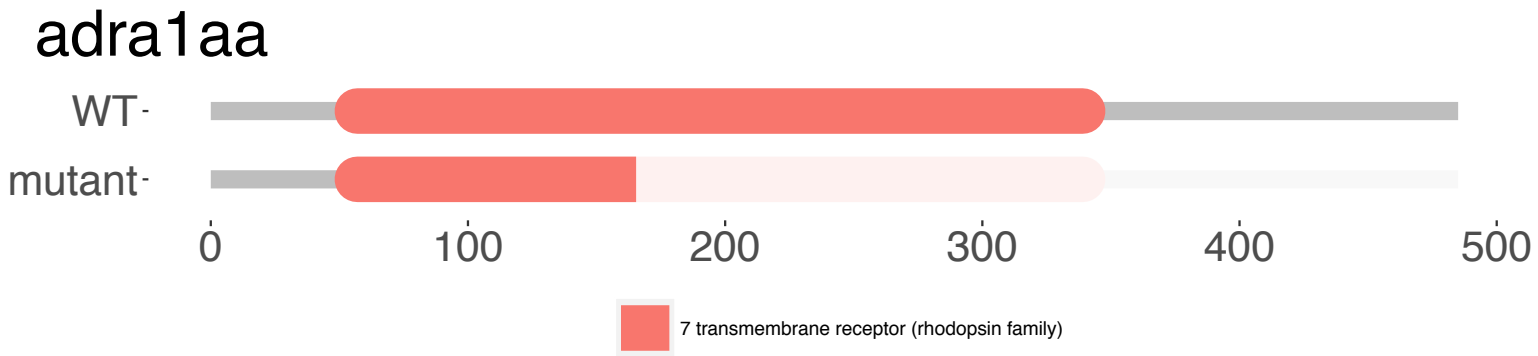
ɜGTGTATGTGATGAATGAAGGCATTGATCTGTATGTGTGTGTGTCCGTCTGTATATG

CGACTCGCCACCCACTTTTGTAAACCAACTTTGTAAAGGCGCTAGATA

GTGTGTGTGTGTCCAGATATTTTACTACTGCCACAGAGAAACGTGACCAGACTCCCA

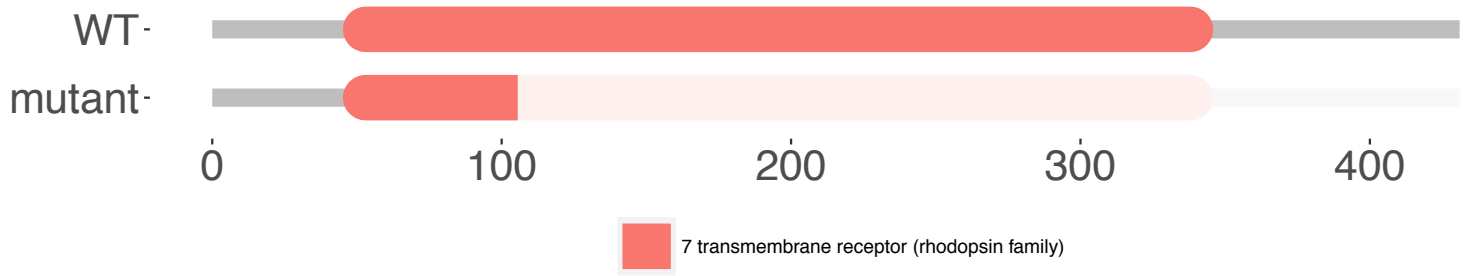
GTATGACACTGTAAAATCCC

Supplemental File 1



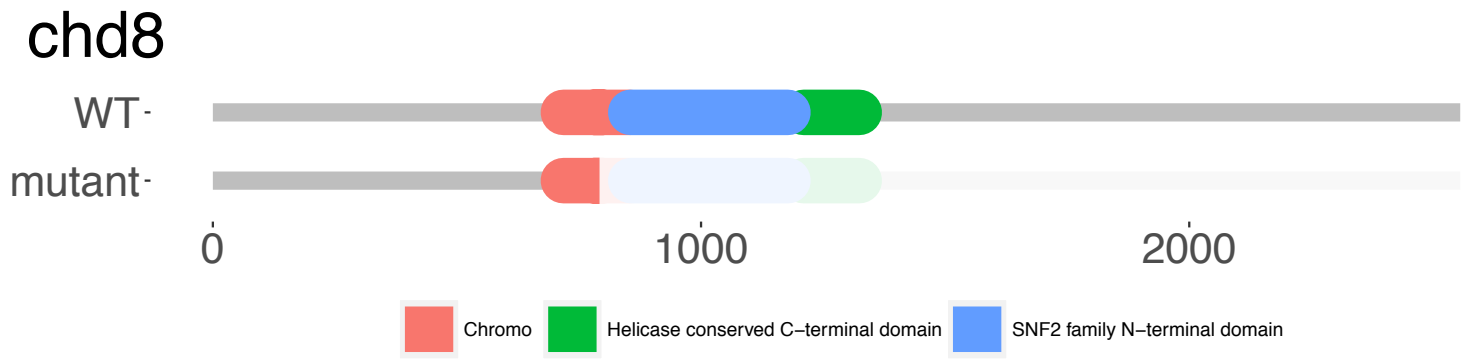

```
NM_001324454.1      gagcgacgtgcattgcttgcgctcatagctttatgggtactgtccatcactatttccatc      1560
NV1180              gagcgacgtgcattgcttgcgctcatagctttatgggtactgtccatcactatttccatc      1560
                    ************************************************************

NM_001324454.1      ggacctctattcggctggaaggaaccggcaccggatgatgaatccatatgcaagataact      1620
NV1180              ggacctc-------------------------------------------------------      1567
                    *******

NM_001324454.1      gaggaacccggttatgccattttctcagctttgggatcttttacctccctctagtggtg      1680
NV1180              ------------------------------------------------------------      1567

NM_001324454.1      atcctctcaatgtattgtcgtgtgtatgttgtggcgaggagggaaacgcgaggactgatc      1740
NV1180              ------------------------------------------------------------      1567

NM_001324454.1      agcggacagaagactgagaagtcggatcacgctgagaccgtgaccttgcgaatacaccgc      1800
NV1180              agcggacagaagactgagaagtcggatcacgctgagaccgtgaccttgcgaatacaccgc      1627
                    ************************************************************



NM_001324454.1      VLCCTASIMSLCVISVDRYIGVSYPLQYPSIVTERRALLALIALWVLSITISIGPLFGWK      180
NV1180              VLCCTASIMSLCVISVDRYIGVSYPLQYPSIVTERRALLALIALWVLSITISIGPQRTED      180
                    ************************************************** .
```

adra1ab

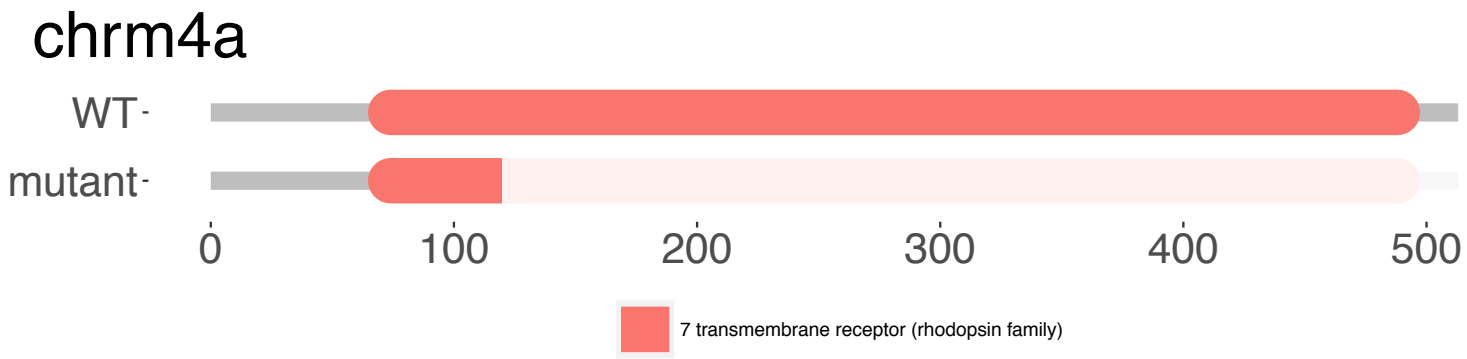



```
XM_680297.7      ctgctgtatctgaagccttgggtcgctgggtgttcggtcgtcatttatgtaacgcctgga   600
NV1175           ctgctgtatctgaagccttgggtcgctgggtgttcggtcgtcatttatgtaacgcctgga   600
                 ************************************************************
XM_680297.7      ccgctcttgacgtgttgtgctgcacggcgtccatcctcagcctctgtgtgatctctgtgg   660
NV1175           ATTTATGTAACccatcctca----------------------gcctcgtgatctctgtgg   638
                        * **   *  *                        *   *************
XM_680297.7      atcgctgcatggccgtcagttacccgctccaataccttcctttgctacgggtcgaaggg   720
NV1175           atcgctgcatggccg-----------------------------------------------   653
                 ***************
XM_680297.7      caatgactgcggtggcagctctctgggcccttcagctgctatctctgtcgggccactat   780
NV1175           ----------------------------------------------------tcagtta   660
                                                                      ** *
XM_680297.7      ttggatggagggaacccatgccagaagacgagtcggtgtgcagagttaacgaagatccag   840
NV1175           cccgctggagggaacccatgccagaagacgagtcggtgtgcagagttaacgaagatccag   720
                    * ************************************************
```

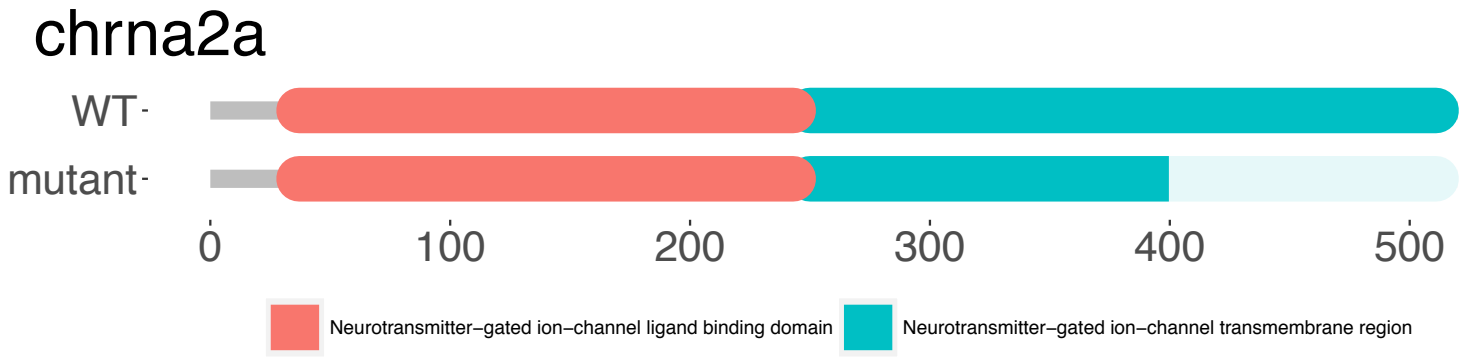
```
XM_680297.7      VACHRNLRSVTHYFIGNLAIADLLLSSVVLPFSAVSEALGRWVFGRHLCNAWTALDVLCC   120
NV1175           VACHRNLRSVTHYFIGNLAIADLLLSSVVLPFSAVSEALGRWVFGRHLCNAWNLCNPSSA   120
                 ************************************************************
```



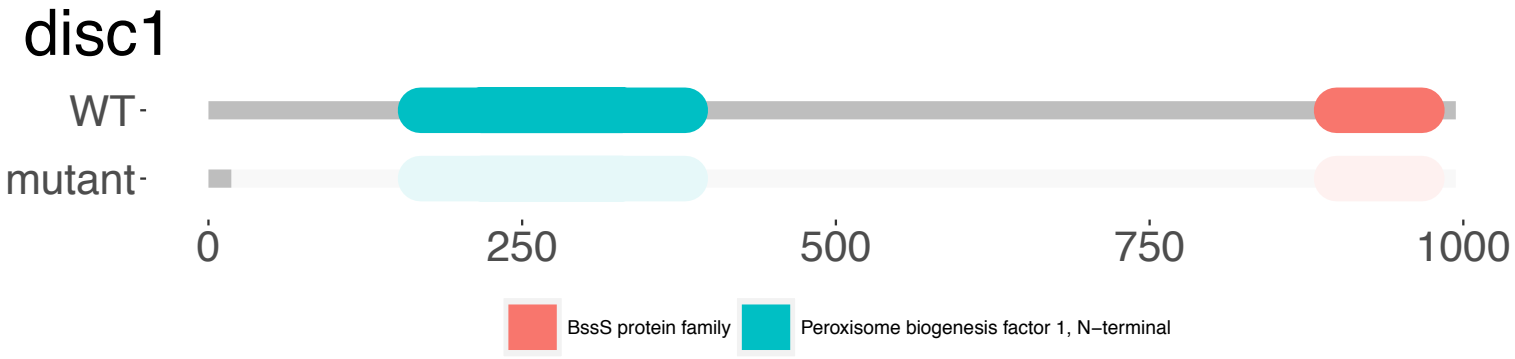

```
XM_005171312.3      tgttgatgaaggcaaagtggaagaattcaggaagattgaaagccgccaaccacgactcaa      2940
NV970               tgttgatgaaggcaa----gaagaattcaggaagattgaaagccgccaaccacgactcaa      2936


XM_005171312.3      EDVDEGKVEEFRKIESRQPR---LKRTPRPAASAWKKLDESTEYKNGNQLREYQLEGVNW      897
NV970               EDVDEGKKNSGRLKAANHDSRELLGLLQVPGKS----------------------------      873
                    ******* :. *   :.:     *     *. *
```



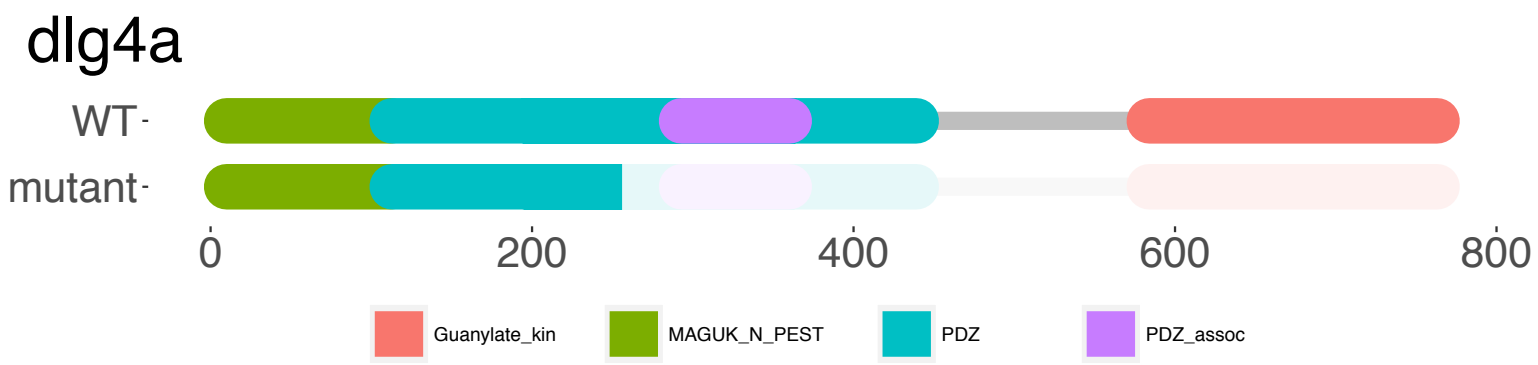

```
XM_001922407.4      accgtctacatcattaagggttactggccgcttggcccagtggtgtgtgacctgtggctc    960
nv1152              accgtctacatcattaagggttactggccgcttggcccagtggtgtg-------tggctc    953

XM_001922407.4      GYWPLGPVVCDL-WLALDYVVSNASVMNLLIISFDRYFCVTKPLSYPTRRTTKMAGLMIA        179
nv1152              GYWPLGPVVCGSRWITWSATLQS*-----------------------------------        143
                    **********.  *:: . .:..
```



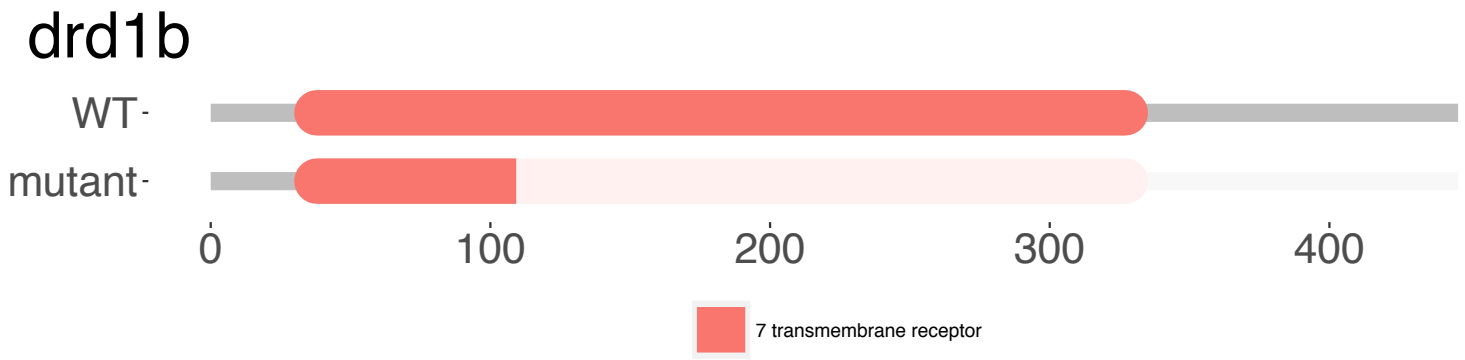

```
NM_001040327.1      gggattggtttatccacatccatctcctctccatccgcaacctctctgcatttgtctgac      1320
NV1187              gggattggtttatccacatccatctcctctccatccgcaacc--------tttgtctgac      1312
                    ****************************************        **********

NM_001040327.1      PEPEKKPKKTANLSTSKSWFQQETGVDGLVCQDMEFGIGLSTSISSPSATSLHLSDSEPL      420
NV1187              PEPEKKPKKTANLSTSKSWFQQETGVDGLVCQDMEFGIGLSTSISSPSATFV*-------      412
                    ************************************************** :
```



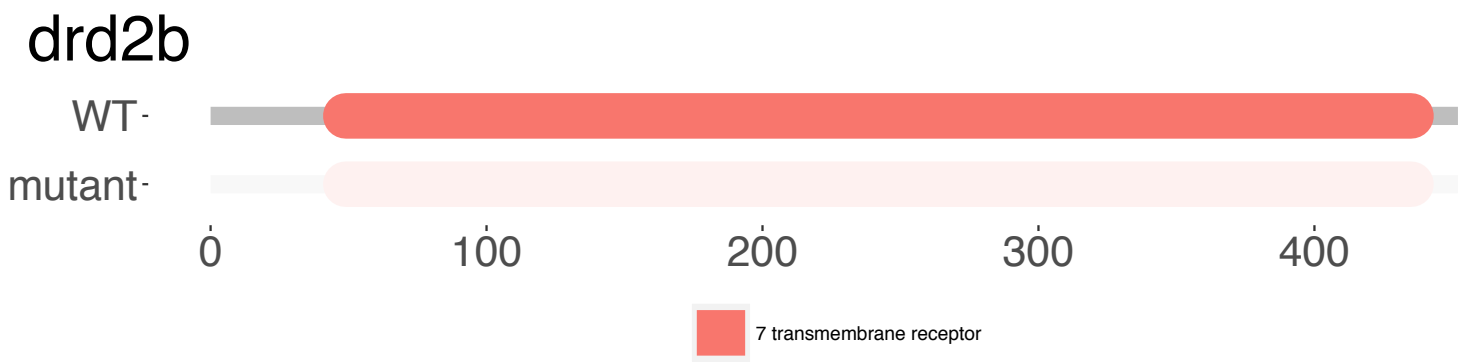

```
NM_001142263.1      tgtgaatccttcgggaaaccaccgcaggaggagcttcagacggcccggatacatgcggtc        240
NV1142              tgtgaatccttcgggaaacc-----aggaggagcttcagacggcccggatacatgcggtc        235

NM_001142263.1      MMFAGMVRVENTSKTLKTDIDSPCHRCAVRTGGVNPSGNHRRRSFRRPGYMRSEPINQLD        60
NV1142              MMFAGMVRVENTSKTLKTDIDSPCHRCAVRTGGVNPSGNQEELQTARIHAV*--------        51
                    ****************************************:.. .  *   :
```




```
XM_009302021.2      cggtgaataatgtgtgtttggaggatgtgatgcatgaggatgcggtgggagcgctgaaaa    1620
NV1149              cggtgaataatgtgtgtttggaggatgtgatg----------cggtgggagcgctgaaaa    1610

XM_009302021.2      KIIEGGAAHKDGRLQIGDKILAVNNVCLEDVMHEDAVGALKNTAEVVYLRVAKPNNLYLT    300
NV1149              KIIEGGAAHKDGRLQIGDKILAVNNVCLEDVMRWER*----------------------    276
                    ********************************: :
```

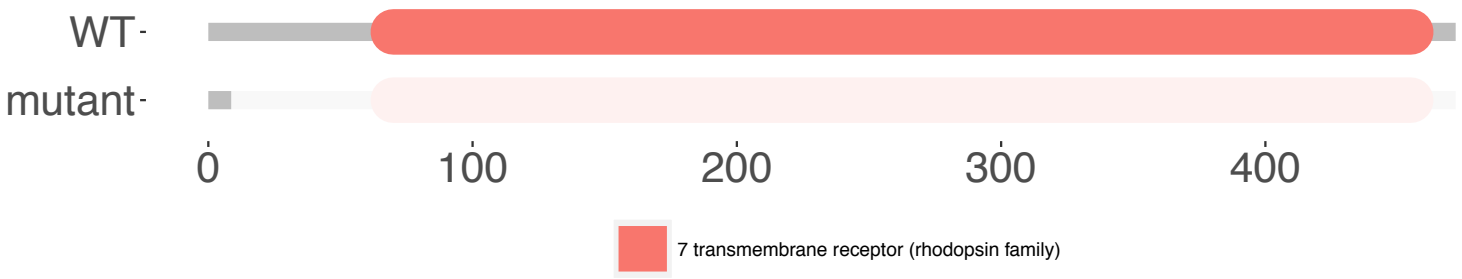



```
NM_001135976.2      tcagtgtggaccgctactgggccatttcaagcccgtttcgctatgagcgaaagatgacac    660
NV1145              tcagtgtggaccgc----gggccatttcaagcccgtttcgctatgagcgaaagatgacac    656


NM_001135976.2      VISLAISDLLVAILVMPWKAATEIVGFWPFGAFCDVWVAFDIMCSTASILNLCVISVDRY    120
NV1145              VISLAISDLLVAILVMPWKAATEIVGFWPFGAFCDVWVAFDIMCSTASILNLCVISVDRG    120
                    ************************************************************
NM_001135976.2      WAISSPFRYERKMTPKVAFIMISVAWTLSILISFIPVQLNWHKAQTTSYTELNGTYGELP    180
NV1145              -----PFQARFAMSER*-----------------------------------------    131
                         **: .  *: :
```




```
XM_009301838.2      tgacaatgggaccggggctctcaactgcacagactgtgaggccaagcat-----------      409
NV1147              tgacaatgggaccggggctctcaactgcacagactgtgaggccaagcaTAGTTAGCAAGt      420
XM_009301838.2      cagtataactactatgccatgctcctgactctcctcatctttgtcattgtctttggcaat      469
NV1147              cagtataactactatgccatgctcctgactctcctcatctttgtcattgtctttggcaat      480


XM_009301838.2      MDFLTEYPYNDTYYDNGTGALNCTDCEAKHQYNYYAMLLTLLIFVIVFGNVLVCMAVSRE      60
NV1147              MDSFTLLEEKMLSIKKLWI*                                              20
                    **
```

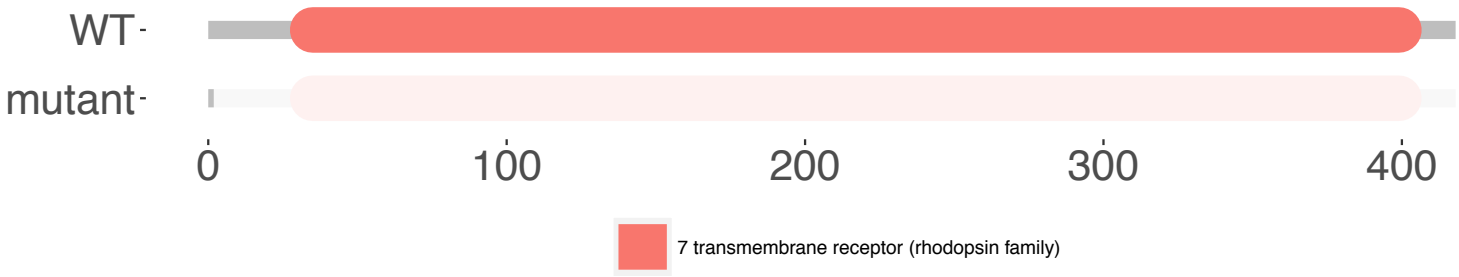
drd3




```
drd3-001 AAGTGCGTTCAGATGGGGTTCTGTTTCGGTTCTTCGCTGGAGAAGCTTAACAGATCTTCC 300
NV1227   AAGTGCGTTCAGATGGGGTTCTGTTTCGGTTCTTCGCTGGAGAAGCTTAACAGATCTTCC 300
         ************************************************************
drd3-001 ACACTTATGGCAATGTTTAGCAGTGGTGAATGGCTCTGGAATGACTCTGAGCATTCATTC 360
NV1227   ACAC------------------------------------------------------- 304
         ****
drd3-001 ACCCCTGGGGGAAACTACAGCCCAGCGTCAGGCGTTGAAGAAGCGAAGAGGAATTATTAT 420
NV1227   -----TGGGGGAAACTACAGCCCAGCGTCAGGCGTTGAAGAAGCGAAGAGGAATTATTAT 359
              ************************************************************
drd3-001 GCCATGCTCTATTCCCTGCTCATCCTGGCTATCGTGTTTGGGAATGTTCTGGTTTGCATT 480
NV1227   GCCATGCTCTATTCCCTGCTCATCCTGGCTATCGTGTTTGGGAATGTTCTGGTTTGCATT 419
         ************************************************************
drd3-001 GCTGTGCTTCGAGAGAGAGCTCTTCAGACCACCACCAACTACCTGGTGGTCAGTCTGGCA 540
NV1227   GCTGTGCTTCGAGAGAGAGCTCTTCAGACCACCACCAACTCAG----------TCTGGCA 469
         ************************************             *******
drd3-001 GTGGCTGATCTGTTGGTCGCCTCACTGGTCATGCCTTGGGCAGTTTATTTGGAAGTGGTC 600
NV1227   GTGGCTGATCTGTTGGTCGCC-----TCACATGCCTTGGGCAGTTTATTTGGAAGTGGTC 524
         **********************       ******************************
```

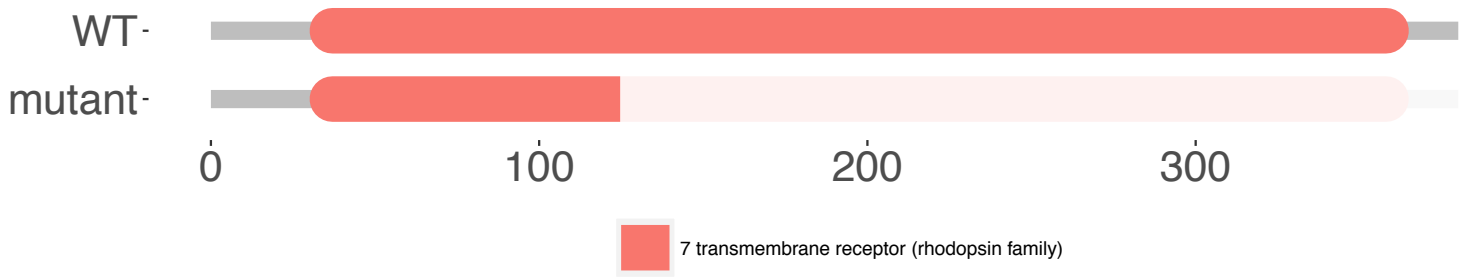
```
ENSDART00000130568      MGFCFGSSLEKLNRSSTLMAMFSSGEWLWNDSEHSFTPGGNYSPASGVEEAKRNYYAMLY          60
NV1227                  MGFCFGSSLEKLNRSSTLGETTAQRQALKKRR-------GIIMPCSIPCSSWLSCLGMF-          52
                        ******************    :. : * :         *   *.*   .:  .  .*:
```

drd4–rs

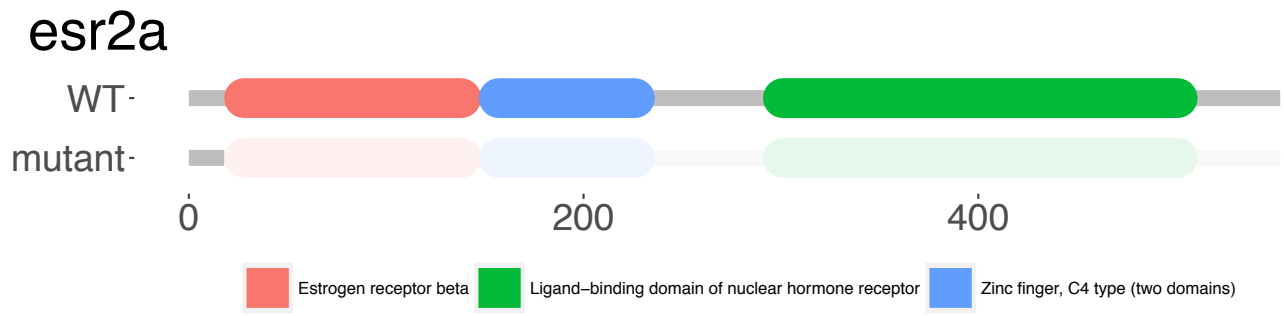



```
NM_001012620.2      agtatagacccaacggcggcccacgaggggtacaactacctggctttgatatgtggcgtg      480
NV1217              agtatagaccc-------------------------------------------------      431
                    ***********

NM_001012620.2      cctctcatcctcattatcattttggggaatgtcctggtatgcctcagtgtgctgaccgaa      540
NV1217              ------------------------------------------------------------      431

NM_001012620.2      cgctcactcaaaacagcaaccaactacttcattgtcagtctggccgtggctgatctcctg      600
NV1217              ---------aaaacagcaaccaactacttcattgtcagtctggccgtggctgatctcctg      482
                             **************************************************

NM_001012620.2      MVNVTPSIDPTAAHEGYNYLALICGVPLILIIILGNVLVCLSVLTERSLKTATNYFIVSL       60
NV1217              MVNVTPSIDPKQQPTTS-----------------------------------------       17
                    **********.
```

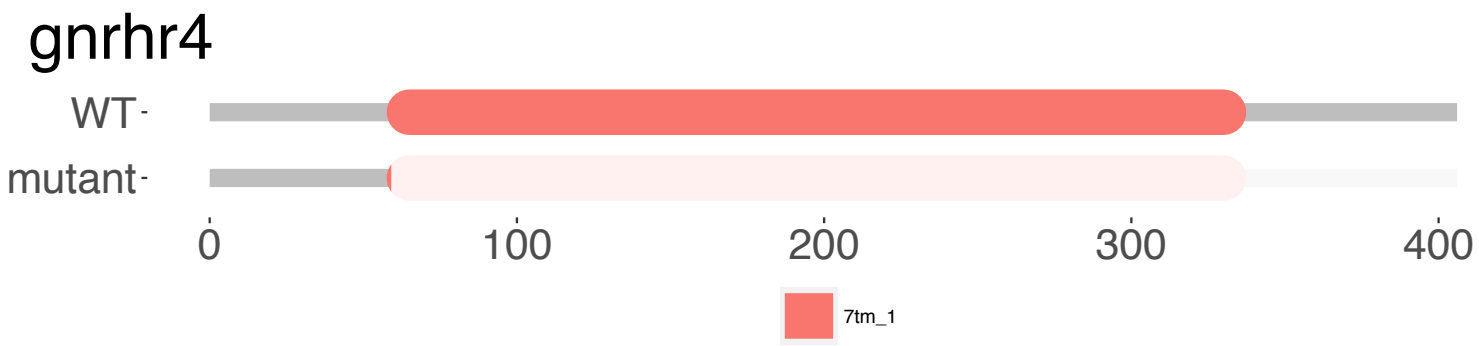
drd4a




```
NM_001012616.3      acaaccgcaaacacgtggaccagcgacagattgtcctgctctccgccacgtggatcctgg      660
NV1218              acaaccgcaa----------------------------------------------      610
                    **********
NM_001012616.3      ccttagccgtggcctcccccgtgatgttcggcatcaacaacgtccccaaccgggaccaca      720
NV1218              ----------------------------------------------------------      610

NM_001012616.3      gcgaatgcaagctggaggacaacaactacgttatctactcctccgtctgttcgtttttcg      780
NV1218              ----------------------------------------------------------      610

NM_001012616.3      tgccgtgccccatcatgctgctcctgtactgcggaatgttccggggcctgaggaactggg      840
NV1218              ----------------------------------------------------------      610

NM_001012616.3      aagaggctcggaaagctaaattaaggagcaacatggaggcctgccgaaagcttcaggagg      900
NV1218              --gaggctcggaaagctaaattaaggagcaacatggaggcctgccgaaagcttcaggagg      668
                      ************************************************************



NM_001012616.3      IAVSIPLNYNRKHVDQRQIVLLSATWILALAVASPVMFGINNVPNRDHSECKLEDNNYVI      180
NV1218              IAVSIPLNYNRKRLGKLN*----------------------------------------      138
                    ************:::.: :
```

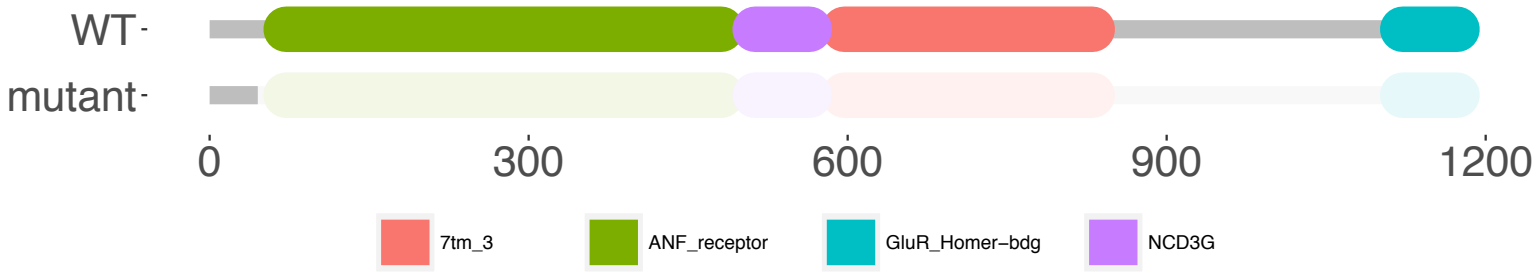



```
NM_180966.2      caggaagtggactcaggccgtgtgggaggtcacatcctctccccc-atcttcaactcatc      239
NV1182           caggaagtggactcaggccgtgtgggaggtcacatcctctcccccaTtcttcaactcatc      240
                 *****************************************  *************

NM_180966.2      MSEYPEGDSPLLQLQEVDSGRVGGHILSPIFNSSSP----SLPVENHPICIPSPYTDLGH      56
NV1182           MSEYPEGDSPLLQLQEVDSGRVGGHILSPILQLILSISASGESPHLHPIALHRPWPRLQH      60
                 ******************************:: .   . ***.:  *:  * *
```

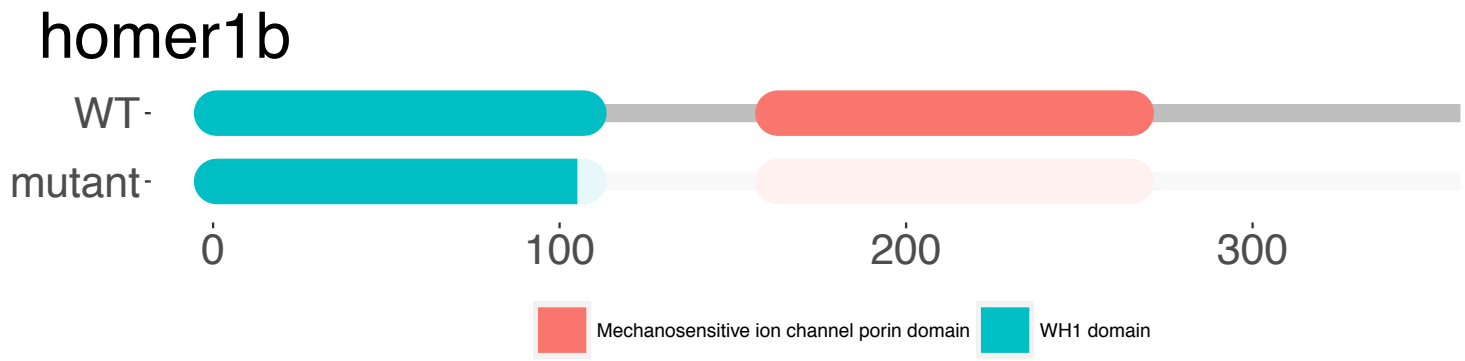



```
NM_001098193.1      agcaaaagtgagggtgataattaccttcactctttgcgcagtgtcagctgtgtgcaacct    240
NV1280              agcaaaagtgagggtgataattaccttcactctttgcgcagtgtcagctgtgtgcaacct    240
                    ************************************************************

NM_001098193.1      cggcg--tgttgtgggctgccagcaccaacaacaagcgcaagtcccatgtacgaatttta    298
NV1280              cggcAACGgttgtgggctgccagcaccaacaacaagcgcaagtcccatgtacgaatttta    300
                    ****    **************************************************



NM_001098193.1      VSAVCNLGVLWAASTNNKRKSHVRILIMNLTVADLLVTFIVMPVDAAWNITVQWLAGDLA    120
NV1280              VSAVCNLGNGCGLPAPTTSASPMYEF*--------------------------------    86
                    ********   .  : ..  * :  :
```

grm5a

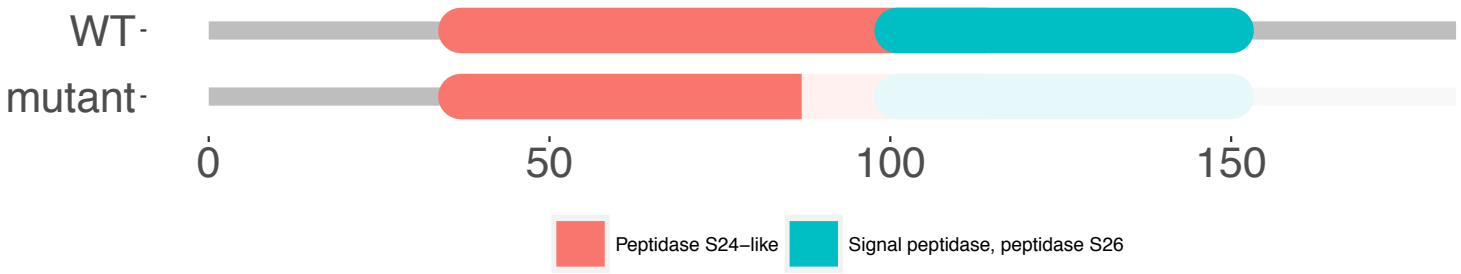



```
XM_009305305.2      caaggtccacgagcgcaagtgtggcgccgtccgagagcagtatggcatccagcgggtgga    360
NV1162              caaggtccacgagcgcaagtgtggcgccgtccgagagcagtatg----------ga    346

XM_009305305.2      KCGAVREQYGIQRVEAMMHTLDRINADPLILPNITLGCEIRDSCWHSAVALEQSIEFIRD    120
NV1162              KCGAVREQYGGHDAHPGPHQRRPAHPS-QHHPRL*----------------------    93
                    ********** : ..   *     : .    *.:
```

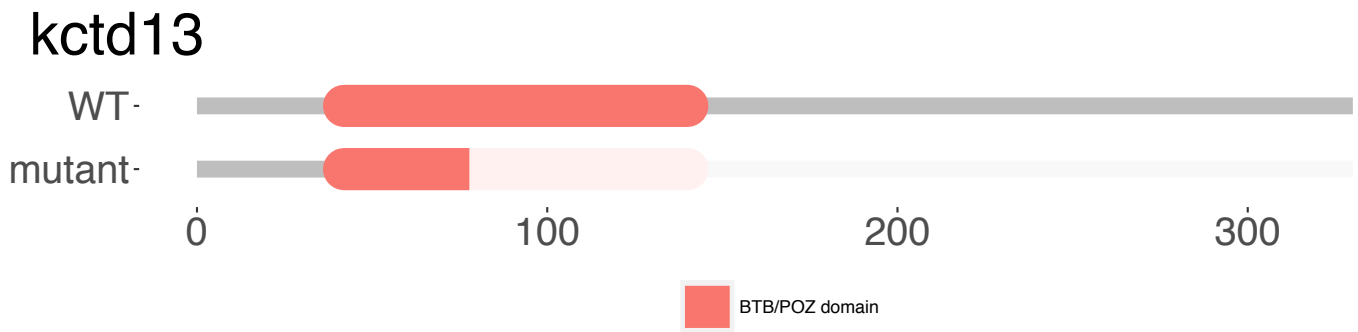



```
NM_001326290.1      tttgcagagtacaaagaagcagccagaatagccaaggagaagtctctagaaaagatggag        720
NV1165              tttgcagagtacaaagaagcagccagaatac-caaggagaagtctctagaaaagatggag        719
                    ****************************  ****************************

NM_001326290.1      ITPNMSFTKTSQKFGQWADSRANTVYGLGFSSEHHLAKFADKFAEYKEAARIAKEKSLEK       120
NV1165              ITPNMSFTKTSQKFGQWADSRANTVYGLGFSSEHHLAKFADKFAEYKEAARIPRRSL*--       117
                    ************************************************** :..
```

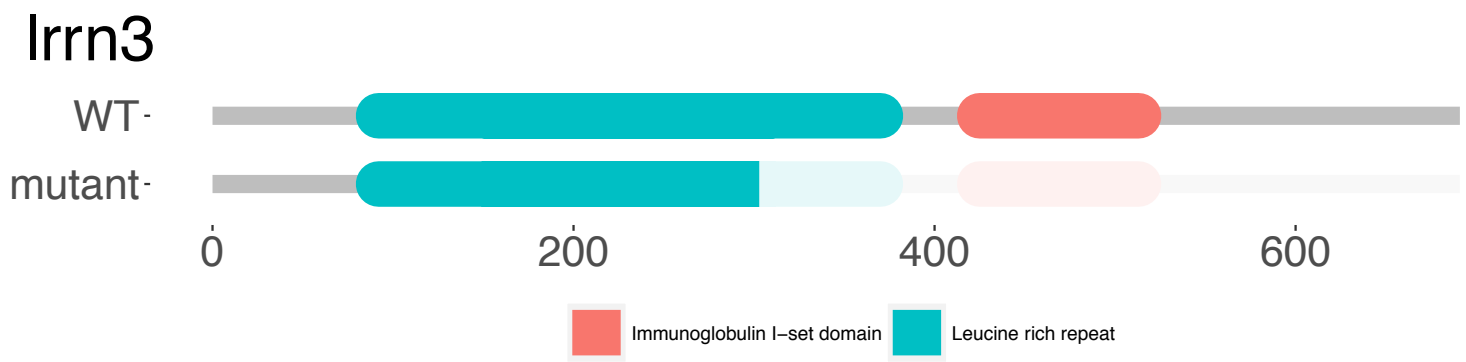
immp2l



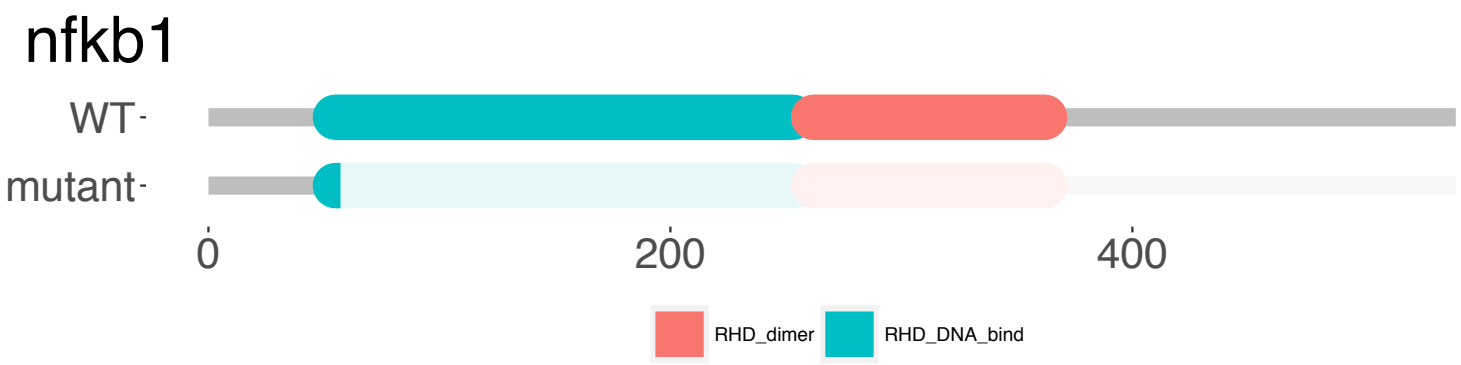

```
NM_001003755.2      gatcatcaaacgggtcatcgggatcgaaggggacttcattaaaacgctgggatataaaaa    360
NV1092              gatcatcaaa-------tcgggatcgaaggggacttcattaaaacgctgggatataaaaa    353

NM_001003755.2      NRWSVRNYHVQRGDIVSVLSPKNPQQKIIKRVIGIEGDFIKTLGYKNRYVRVPDGHLWIE    120
NV1092              NRWSVRNYHVQRGDIVSVLSPKNPQQKIIKSG--SKGTSLKRWDIKTVM*----------    107
                    ******************************     :*  :*  . *.
```




```
NM_001077151.1      agcatcgactctgaaggctgggtggtgttggaccgatgcgggagacactttccctggtg      480
NV966               agcatcgactctgaaggctgggtggtgtt--------gcgggagacactttccctggtg      472
                    ****************************        ***********************


NM_001077151.1      MSAEASGSSGGHAVTVSGSSPSSSSHVGDEKPGRSLVSSKYVKLNVGGTLHYTTVQTLSK      60
NV966               MSAEASGSSGGHAVTVSGSSPSSSSHVGDEKPGRSLVSSKYVKLNVGGTLHYTTVQTLSK      60
                    ************************************************************


NM_001077151.1      EDSLLRSICDGSTEVSIDSEGWVVLDRCGRHFSLVLNFLRDGTVPLPDSTRELEEVLKEA      120
NV966               EDSLLRSICDGSTEVSIDSEGWVVLRETLFPG-----------VELPAGRD--------S      101
                    ************************ .               * ** .          :


NM_001077151.1      QYYRLQGLVQHCLSTLQKRRDVCRGCHIPMITSAKEEQRMIATCRKPVVKLQNNRGNNKY      180
NV966               SFAR*-------------------------------------------------------      105
                    .: *
```

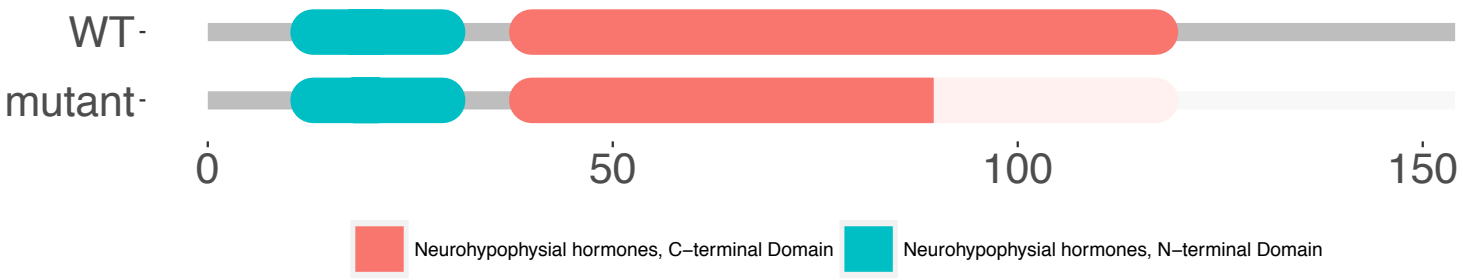



```
XM_005163071.3      aagccacaaacaacccccgactttcctatattcacccaaatgcttttctcaactcccaa      1620
NV971               aagccacaa-----ccccgactttcctatattcacccaaatgcttttctcaactcccaa      1615

XM_005163071.3      DSFALHNLPELTKIEATNNPRLSYIHPNAFSQLPRLESLMLNSNALRALHHITVESLPNL      360
NV971               DSFALHNLPELTKIEATTPTF-LY-SPKCFFSTPKTGVLDAE*----------------      340
                    *****************.     *  *:.* . *:   *  :
```



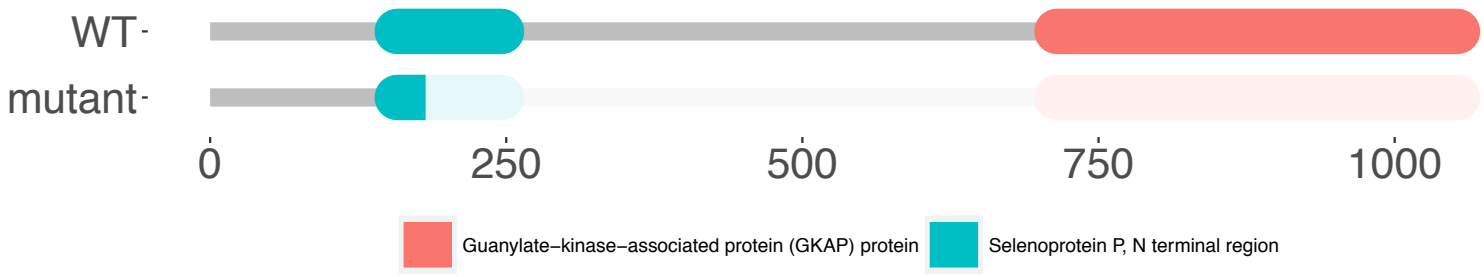

```
NM_001353873.1      GCTTACAAATCACAGAACAACCAAAACAGAGGGGCTTCAGGTTCCGCTATGGGTGTGAGG    540
NV1244              GCTTACAAATCACAGAACAACCAAAACAGAGGGGCTTCAGGTTCCGTGTGAGGGCACGGA    540
                    ********************************************     **     *
NM_001353873.1      GCCCGT---------CTCACGGAGGGCTTCCAGGGGCCTCCAGCGAGAAGAACCGCAAGT    591
NV1244              GGTGTGAGGTGTGAGGGCACGGAGGGCTTCCAGGGGCCTCC-------------AGCGAG    587
                    *               ************************
NM_001353873.1      CCTACCCACAAGTGCAGATCTGTAATTATCAGGGCCCTGCTCGTGTGGTGGTCCAGCTGG    651
NV1244              AAGAACCACAAGTGCAGATCTGTAATTATCAGGGCCCTGCTCGTGTGGTGGTCCAGCTGG    647
                       * ************************************************


NM_001353873.1      PKQRGFRFRYGCEGPSHGGLPGASSEKNRKSYPQVQICNYQGPARVVVQLVTNSQHPHLH    120
NV1244              PKQRGFRFRVRARRCEV*---------------------------------------      77
                    *********  ..  .
```

oxt



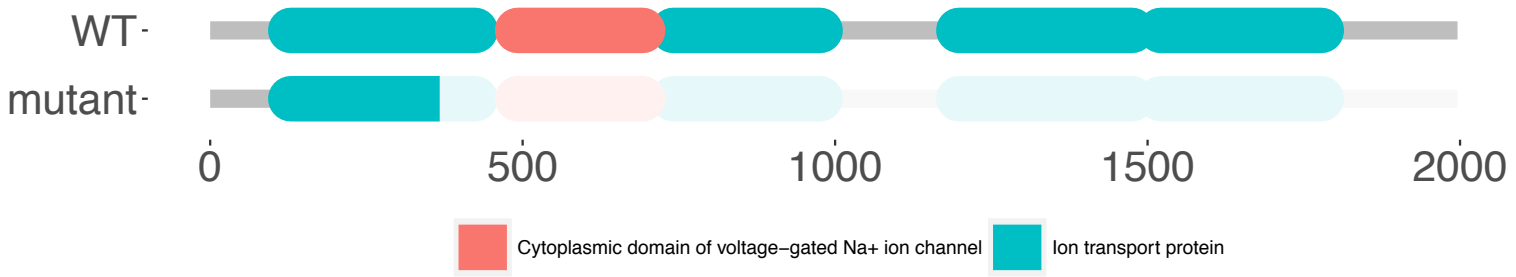

```
NM_178291.2      ccttctccgtgtgagatgtctggaaaggcctgcggttatgagggacgctgcgctgctcct   360
NV1081           ccttctccgtgtgagatgtctggaaaggcctgc-------------gctgcgctgctcct   347

NM_178291.2      GEGIGCLVGSPETLRCLEEDFLPSPCEMSGKACGYEGRCAAPGVCCDSEGCSVDQSCVDG   120
nv1081           GEGIGCLVGSPETLRCLEEDFLPSPCEMSGKACAALLLESAATRR---------------   105
                 *********************************.     :*
```

sapap2

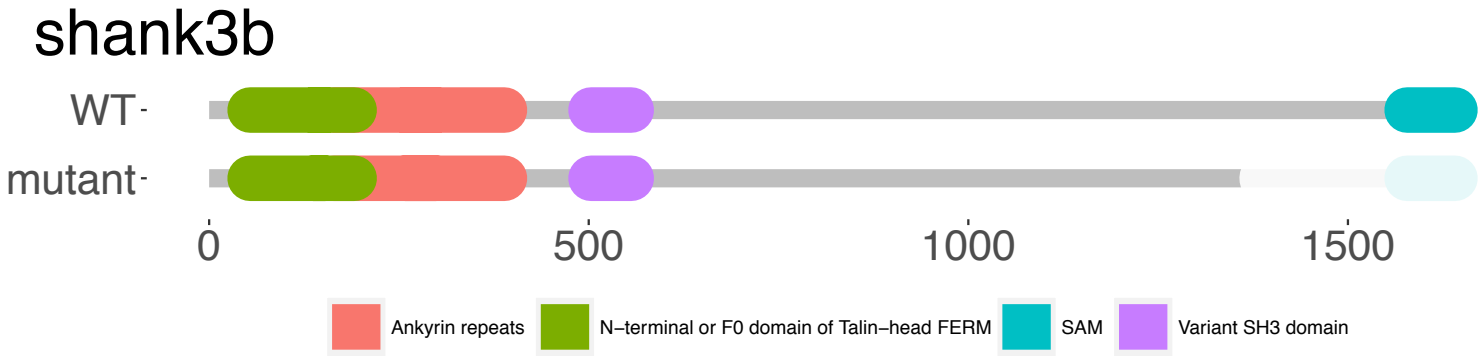



```
XM_680713.7      cttccaagatgaacggcaccaagggagactctcgctcagacagcggacaccaccaccatc   1020
NV1136           cttccaagatgaacggcaccaagggagactctcgctca----gcggacaccaccaccatc   1016
                 ************************************    ******************

XM_680713.7      KLFTKSHSLEGSSKMNGTKGDSRSDSGHHHHHHHHHGHDHGSHKHSKRSKSKERKSDSKQ   240
NV1136           KLFTKSHSLEGSSKMNGTKGDSRSADTTTITTIIT-AMTMGATNTANEVRARN---ASQT   236
                 ************************ .          .   *: : ::. ::::    *:
```

scn1lab

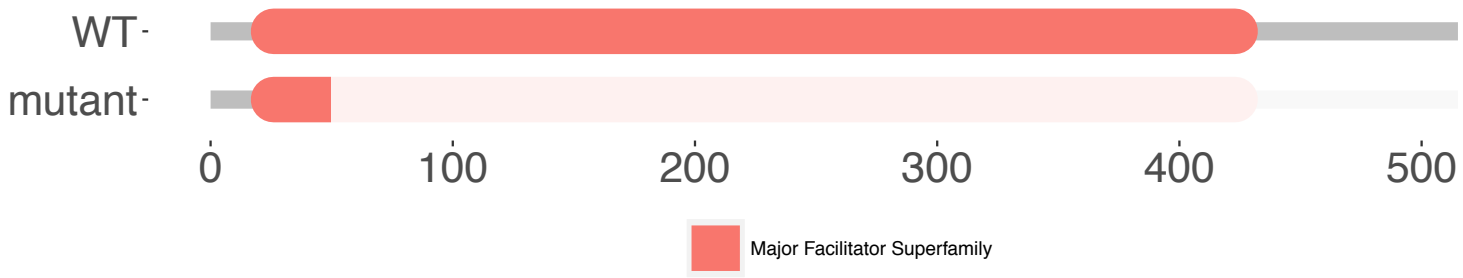



```
XM_005165770.3      tctatctggttaatctcatcctggctgtggtggccatggcgtatgatgagcagaaccagg    1740
NV978               tctatctggttaatctcatcc----tgtggtggccatggcgtatgatgagcagaaccagg    1736
                    **********************    ************************************

XM_005165770.3      WAFLSLFRLMTQDFWENLYQQTLRAAGKPYMIFFVLVIFLGSFYLVNLILAVVAMAYDEQ    420
NV978               WAFLSLFRLMTQDFWENLYQQTLRAAGKPYMIFFVLVIFLGSFYLVNLILWWPWRMMSRT    420
                    ************************************************        ..
```



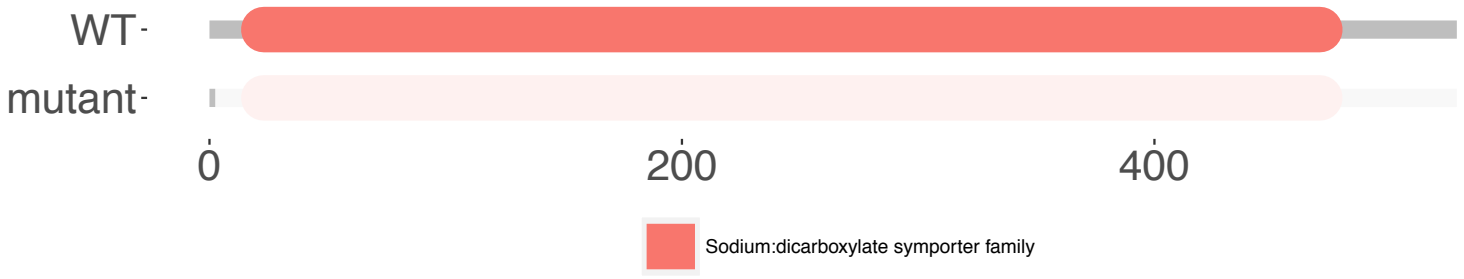

```
XM_009300184.2      ttccaaactgtgggggggaggagccccgggcccagatggggtcatcagatgaaagtcggcc      4440
NV1061              ttccaaactgtgggg----------------------ggtcatcagatgaaagtcggcc      4417

XM_009300184.2      PRYLFQRRSKLWGEEPRAQMGSSDESRPAAMGAELLSKDTHSLGEEPPMGAPLDPGRRSP      1440
NV1061              PRYLFQRRSKLWGVIR*-----------------------------------------      1396
                    *************
```

slc18a2



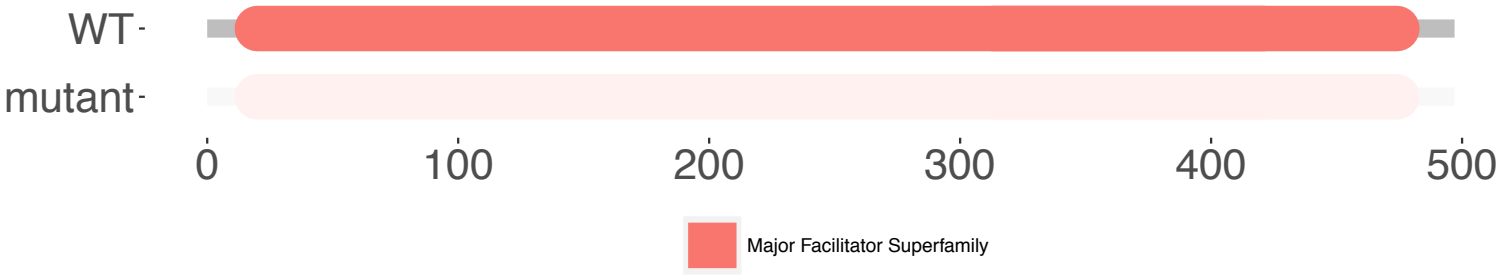

```
NM_001256225.2      ggtggacgacgaggctgctcagatggttaagaatcactccatgacccctctttctccatc      360
NV1206              ggtggacgacgaggctgctccatcg----------------------------------      325
NM_001256225.2      gagcacctttcagagcattgtgtctttgtatgacaacaccacacgtgtgacaggcttcag      420
NV1206              -agcacctttcagagcattgtgtctttgtatgacaacacc--------------------      364
NM_001256225.2      cccgcagatgagcacagctggtccaatgagcctggctcctacttttgtgagccctcaaaa      480
NV1206              ------------------------------------------------acacgaa      371
NM_001256225.2      ccagtccgactgtcccaaagcagatgaccagcttctgaacgagaatgtgaaagtgggtct      540
NV1206              ccagtccgactgtcccaaagcagatgaccagcttctgaacgagaatgtgaaagtgggtct      431


NM_001256225.2      MGLFDALRDFSLLTWLREERQSRRLILLIVFIALLLDNMLLTVVVPIIPSYLYTVDDEAA       60
NV1206              MGLFDALRDFSLLTWLREERQSRRLILLIVFIALLLDNMLLTVVVPIIPSYLYTVDDEAA       60
                    ************************************************************
NM_001256225.2      QMVKNHSMTPLSPSSTFQSIVSLYDNTTRVTGFSPQMSTAGPMSLAPTFVSPQNQSDCPK      120
NV1206              ------------PSSTFQSIVSLYDNTTRTSPTVPKQMTSF*------------------       89
                                ****************.:   *:  *:
```

slc1a1




```
NM_001002666.1      atgctggggaagaaagagcggagaagcagagatgtgaagggtttgctgaagaggaactgg        180
NV1211              atgctggggaagaaagagcggagaagcagag--atgaagggtttgctgaagaggaactgg        178
                    ******************************   *************************

NM_001002666.1      MEMLGKKERRSRDVKGLLK-------------RNWILIATIIAVILGIGLGVVVRDYTSL        47
NV1211              MEMLGKKERRSRDEGFAEEELDSHCNHHCGHSGDWSWCG-------GAG--LHLSDPAGE        51
                    *************     :              :*   .       * *  : : * :.
```

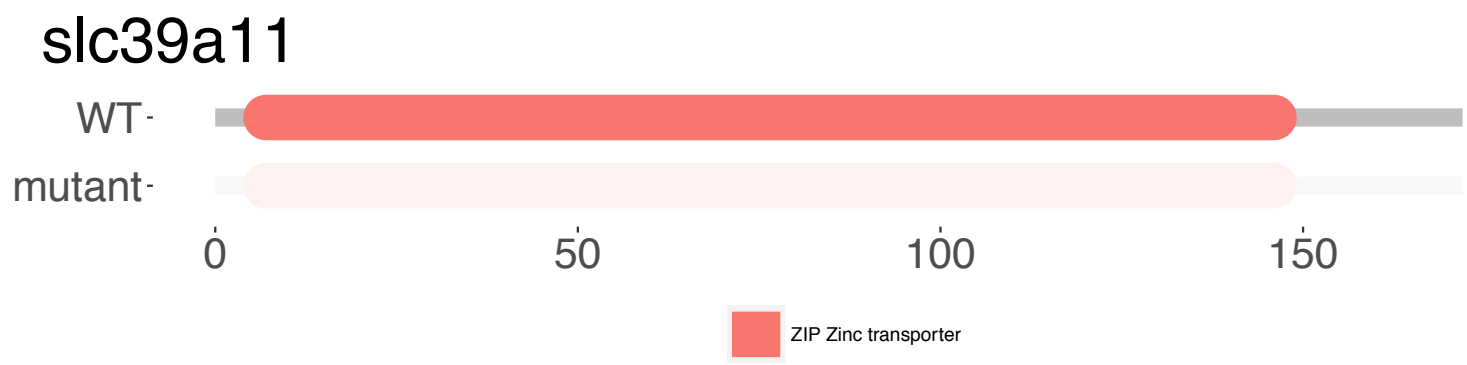
slc22a15




```
NM_001109699.1      tctggaatggatttagaggaggctttccaggt--------------------------      152
NV1209              tctggaatggatttagaggaggctttccaggtTGGAGGATGTTCCAGGATGTTCTCTCTA      180
NM_001109699.1      ---ggttggagagttcggaagtcaccagaagcgaatgatcacagtgctggtgttcctcca      209
NV1209              TCAggttggagagttcggaagtcaccagaagcgaatgatcacagtgctggtgttcctcca      240


NM_001109699.1      MDLEEAFQVVGEFGSHQKRMITVLVFLQIYMACQSMLIILVGAVPEYHIEPVTEGVDGDV       60
NV1209              MDLEEAFQVGGCSRMF-----------------------------------------       16
                    ********* *    .
```

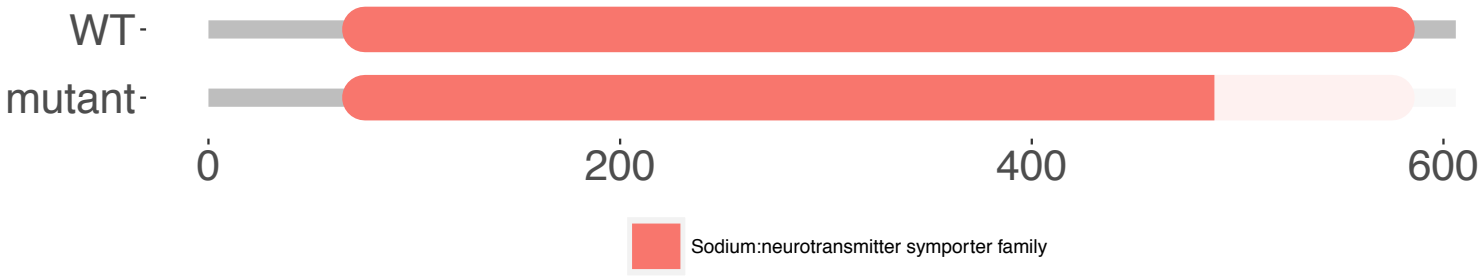
slc25a27


```
NM_200341.1      cagaattggtcacattccccctggatcttaccaaaaccagactccagatccagggtgaag    240
NV1195           cagaattggtcacattTTC----------------------------------------    199
                 ****************  *
NM_200341.1      gcagatctggaaagaatggtggaagcgtacagactcagaaatacaggggcatgttgagca    300
NV1195           ---------------------TGagcgtacagactcagaaatacaggggcatgttgagca    238
                                       ************************************


NM_200341.1      MSHLQENSRWPRVSKFTLSACAAAVAELVTFPLDLTKTRLQIQGEGRS----GKNGGSV-    55
NV1195           MSHLQENSRWPRVSKFTLSACAAAVAELVTFSERTDS---EIQGHVEHSCRYSARRGALE    57
                 ********************************     .   :***. .     . . *::
```



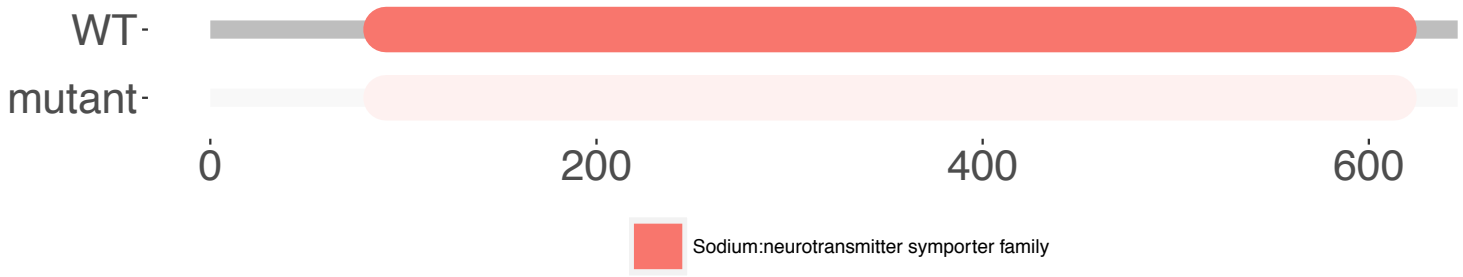

```
XM_005170120.3      cacagacacgtgtacaatgcgaagcaggaagttcttcactcggtgcctgtaaacaacagc        60
NV1200              cacagacacgtgtacaatgcgaagcaggaagttc-------------------------        34
                    **********************************
XM_005170120.3      atgtttccaggcctcagtccgctggttcaggccctgctggggactctgttcacctgggcg        120
NV1200              ------------------------------------------------ttcactcggtgc        46
                                                                     *****  **
XM_005170120.3      ctgactgcagctggagccgcactggtcttcatcttctccagcagacagaagcggattttg        180
NV1200              ctg--taaacaacagcatgCTCCACAGTCAAGcttctccagcagacagaagcggattttg        104
                    ***  *  *         ** *     *  * ****************************



XM_005170120.3      --------------------MFPGL-------------------SPLVQALLGTLFTWAL        21
NV1200              MLHSQASPADRSGFWMAVWDLQQGLCSQPRIGLCWLLLLRWQRNLGNMETLLSSQWLWAL        60
                                      :  **                     :::**.: : ***
```

slc6a3



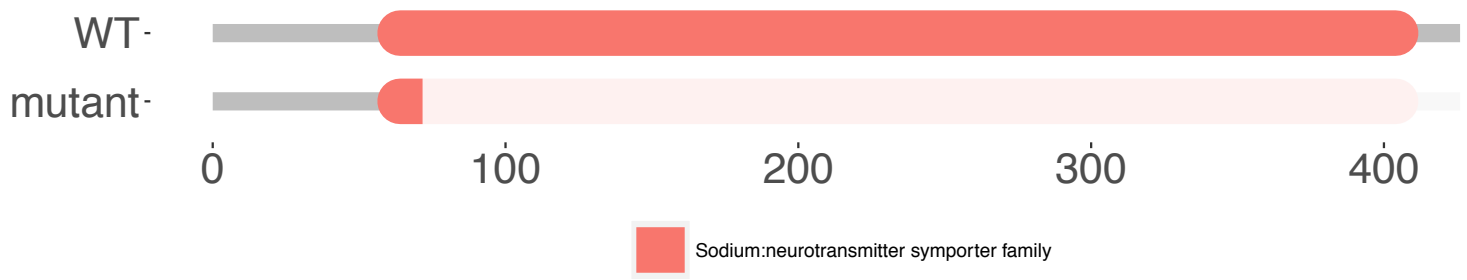

```
NM_131755.1      tttacgctgttggaccacttcgctgcggggacgtcaattctctttggagtactaattgag   1500
NV1087           tttacgctgttggaccacttcgctgcggggacgtcaattctctttggagtactaattgag   1500
                 ************************************************************

NM_131755.1      gccatcggcatcgcctggttttacggagtggatcgcttcagtgatgatatcgaggagatg   1560
NV1087           g-------catcgcctggttttacggagtggatcgcttcagtgatgatatcgaggagatg   1553
                 *       ************************************************

NM_131755.1      FTLLDHFAAGTSILFGVLIEAIGI--AWFYGVDRFSDDIEEMIGQRPGLYWRLCWKFVSP   538
NV1087           FTLLDHFAAGTSILFGVLIEASPGFTEWIASV---------MISRR*-------------   517
                 **********************      *: .*         **.:*
```

slc6a4a



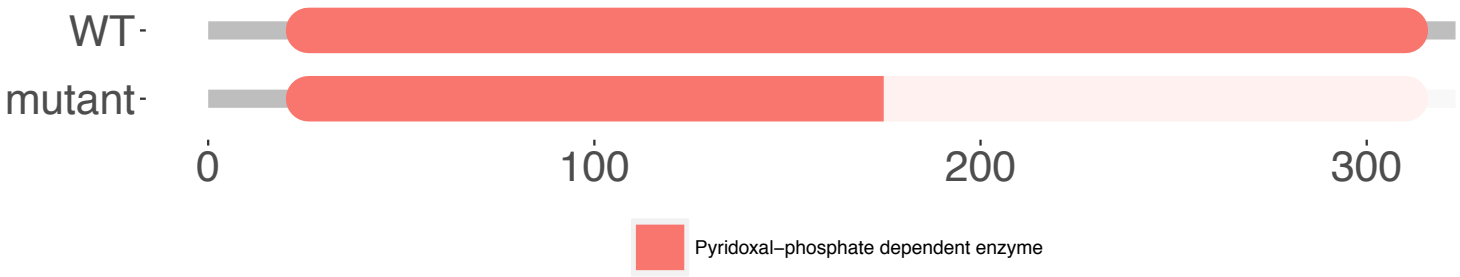

```
XM_009291669.2      aagagtacgggggagagcagcagaaagtgccggagtctcaagagaacggcaggctggttg      840
NV1208              aagag----ggggagagcagcagaaagtgccggagtctcaagagaacggcaggctggttg      836
                    *****    **************************************************

XM_009291669.2      MDMKESMMMNQEYGGEQQKVPESQENGRLVVDSVPEKDQKSGSGPGQVSNGYRSTSPQSP      60
NV1208              MDMKESMMMNQEGESSRKCRSLKRTAGWLWIAFR-RRIRNLALGLGKSPMVIAVHLLKAP      59
                    ************  ..::    .:  * * :    .: :: . * *:          ::*
```

slc6a8

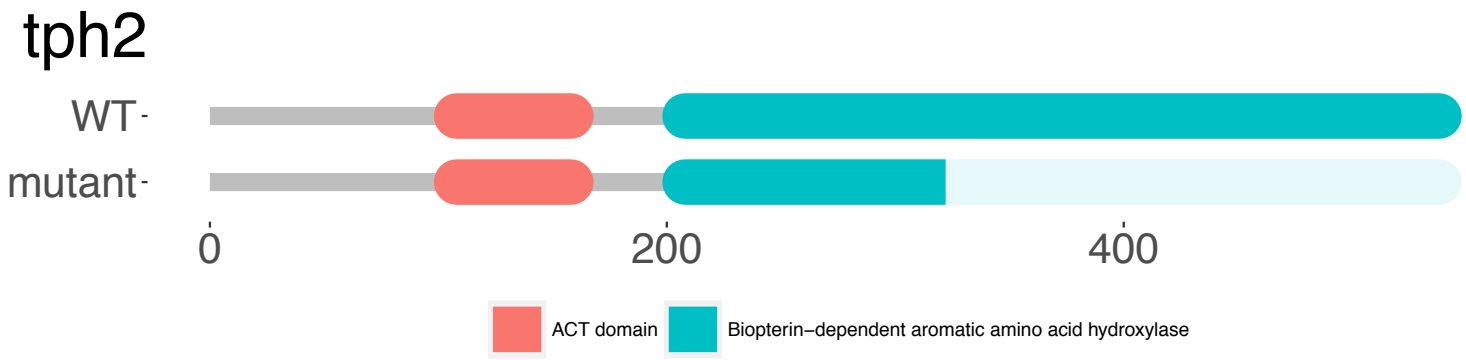



```
XM_005166721.2      aatggaaacgcgcacccaagcgtcaacgggaccctcaacgggggcccatgtcagcggct      840
NV1197              aatggaaacgcgcacccaagcgtcaacgggaccctcaTGTCggggcccatgtcagcggct      840
                    ************************************    ******************

XM_005166721.2      accggagccatctcggcggtggagaagaagagggagggctaccccgagcgagagacgtgg      900
NV1197              accggagccatctcggcggtggagaagaagagggagggctaccccgagcgagagacgtgg      900
                    ************************************************************

XM_005166721.2      accagacagatggacttcatcatgtcctgcgtgggttttgcagtggggctggggaacgtc      960
NV1197              accagacagatggacttcatcatgtcctgcgtgggttttt----------ggggaacgtc      950
                    ************************************          **********


XM_005166721.2      MEKSSLDADCCALNMVLVEEKKGHLIPNGNAHPSVNGTLNGGPMSAATGAISAVEKKREG       60
NV1197              MEKSSLDADCCALNMVLVEEKKGHLIPNGNAHPSVNGTLMSGPMSAATGAISAVEKKREG       60
                    ************************************ .******************

XM_005166721.2      YPERETWTRQMDFIMSCVGFAVGLGNVWRFPYLCYKNGGGVFLIPYVLFIFLGGIPIFFL      120
NV1197              YPERETWTRQMDFIMSCVGFWGTSGASRTCATKTVEE-SSSFPMCYSYFW----------      109
                    ********************    *          :: .. * : *   *

XM_005166721.2      EIALGQFMKAGSINVWNIAPLFKGLGYASMVIVFFCNTYYIMVLAWGFYYFIKSFNATLP      180
NV1197              --------EAFRYSSWRL-------------------------RWD-------------      122
                            :*   . *.:                          *.
```

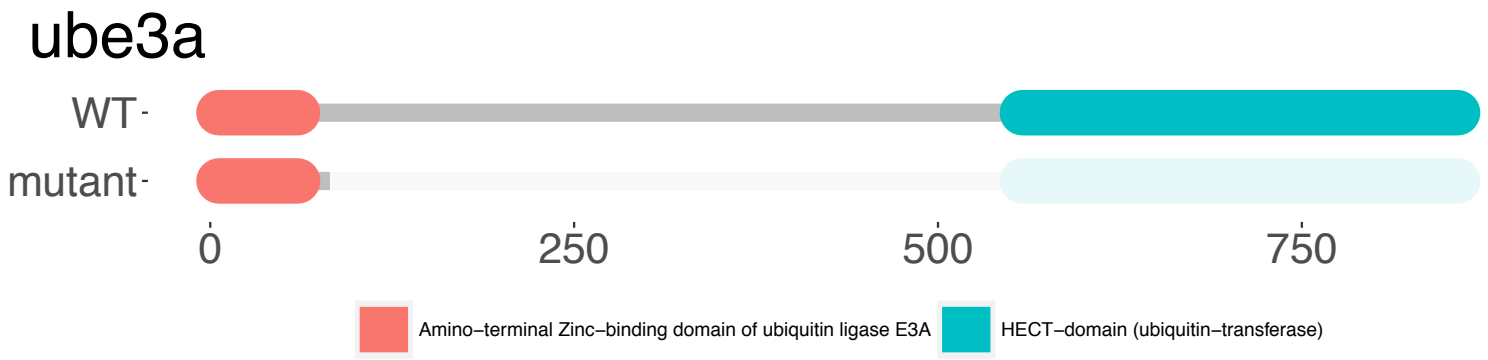
srr




```
XM_002661465.4      gacattccagcttaggttttgagatcctggatgtcgtaccctatccagatgtggtggttg      660
NV1157              gacattccagcttaggttttgagatcctggatgtcgtaccctatccagatgtggtggttg      660
                    ************************************************************

XM_002661465.4      tttgctgtggtggtggagggttactttctggtgtggccgctgccatcaaactgtctggtt      720
NV1157              t--------------------tactttctggtgtggccgctgccatcaaactgtctggtt      700
                    *                    ************************************

XM_002661465.4      NLGVEVERVPTTQLMGVVNRCVQEDGMTFLHSIDDPDLIAGHSSLGFEILDVVPYPDVVV      180
NV1157              NLGVEVERVPTTQLMGVVNRCVQEDGMTFLHSIDDPDLIAGHSSLGFEILDVVPYPDVVV      180
                    ************************************************************

XM_002661465.4      VCCGGGGLLSGVAAAIKLSGCEDTKIYGVEPEGACTMYKSFIEKRPVGMDAKSIASGLAP      240
NV1157              VTFWCGRCHQTVWL*-------------------------------------------      194
                    *    *   . *
```




```
NM_001310068.1      aaggcctgtggctggatatctctcaccacgagacttcctggctggactggcttatcgagt    1140
NV1089              aaggcctgt-gctggatatctctcaccacgagacttcctggctggactggcttatcgagt    1139

NM_001310068.1      LLTKHCGYREDNIPQLEDVSLFLRERSGFTVRPVAGYLSPRDFLAG-LAYRVFNCTQYIR    359
NV1089              LLTKHCGYREDNIPQLEDVSLFLRERSGFTVRPVLDISHHETSWLDWLIECLIALSIYVT    360
                    ********************************** .    .    . *   ::  : *:
```




```
NM_001007318.1      ggacctgccacccacaagggaggactttagagatctgaatttcttgactgaggataaggt      540
NV1064              ggacctgccacccac-----aggactttagagatctgaatttcttgactgaggataaggt      535


NM_001007318.1      DPHPSKKGASSAFPENSAKGAHNFSACSNGKMNHKDLPPTREDFRDLNFLTEDKVYEILS      120
NV1064              DPHPSKKGASSAFPENSAKGAHNFSACSNGKMNHKDLPPTGL*----------------      102
                    ************************************
```

Supplemental File 2

Unpolarized	Kullback–Leibler divergence	Polarized	Kullback–Leibler divergence
'immp2lHO'	Inf	'immp2lHO'	Inf
slc18a2HT'	0.503872588	'adra1aaHO'	0.408321982
'scn1abHT'	0.475243162	'slc25a27HO'	0.401431298
slc6a4aHO'	0.450082277	'drd4-rsHO'	0.380778006
'drd4-rsHO'	0.372634411	'chrna2aHO'	0.352634968
'disc1KO'	0.361275622	'slc18a2HT'	0.335771101
'drd4aHO'	0.360377693	'disc1KO'	0.317098761
'slc1a1HO'	0.344640154	'homer1bHO'	0.307501307
'slc6a8HO'	0.332613849	'slc6a8HO'	0.242433203
'homer1bHO'	0.328809872	'drd4aHO'	0.232985499
'chrna2aHO'	0.256689064	'scn1abHT'	0.229660386
'chrm4aHO'	0.254995398	'gnrhr4HO'	0.229206427
'gnrhr4HO'	0.249192827	'oxtKO'	0.170716837
'adra1aaHO'	0.245908343	'drd2bKO'	0.165339327
'slc6a3KO'	0.239132287	'slc6a4aHO'	0.135011208
'adra1abHO'	0.22832634	'slc1a1HO'	0.124852827
'drd2bKO'	0.223290031	'esr2aHO'	0.115764118
'sapap2KO'	0.215181308	'chrm4aHO'	0.115151629
'esr2aHO'	0.2130425	'srrHO'	0.099703579
'dlg4aHO'	0.210317791	'dlg4aHO'	0.097755979
'chd8KO_WT'	0.189007333	'drd3HO'	0.088199216
'tph2HO'	0.188094395	'slc22a15HO'	0.079772629
'srrHO'	0.186947659	'slc39a11HO'	0.070391332
'slc25a27HO'	0.185835058	'KCTD13KO'	0.065474473
'nfkb1HO'	0.180801278	'sapap2KO'	0.064423091
'oxtKO'	0.177612315	'drd1bHO'	0.053110873
'shank3bKO'	0.16671958	'lrrn3KO'	0.052257913
'drd3HO'	0.161913606	'adra1abHO'	0.045964875
'drd1bHO'	0.099974442	'shank3bKO'	0.045234588
'slc22a15HO'	0.049265198	'nfkb1HO'	0.035939197
'grm5aHO'	0.035304671	'slc6a3KO'	0.031863257
'slc39a11HO'	0.030706812	'tph2HO'	0.030858559
'KCTD13KO'	0.021098744	'grm5aHO'	0.018200187
'ube3aKO'	0.017664977	'ube3aKO'	0.006102855
'lrrn3KO'	0.013683683	'chd8KO_WT'	0.00384635
'WT'	0	'WT'	0