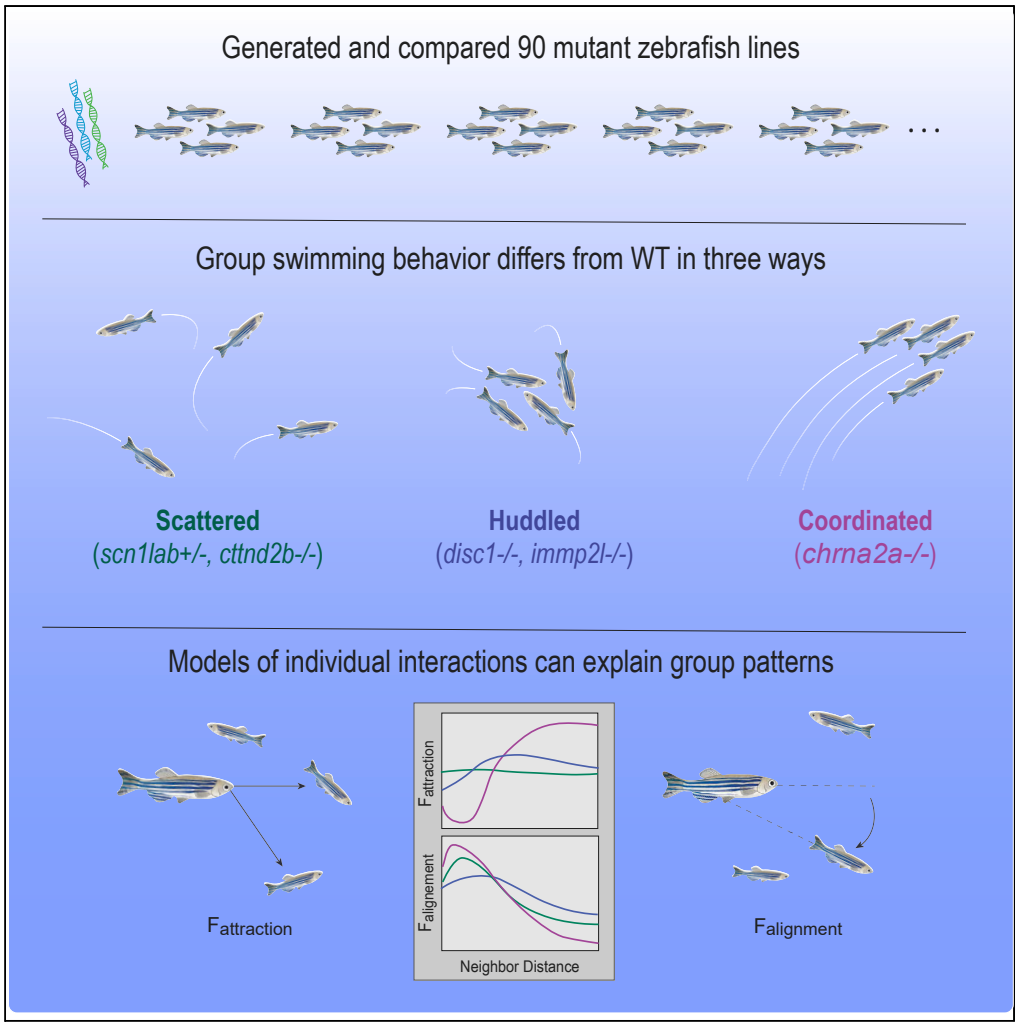
iScience

CellPress

Article

Genetic Control of Collective Behavior in Zebrafish



Wenlong Tang, Jacob D. Davidson, Guoqiang Zhang, ..., Peixin Zhu, Iain D. Couzin, Mark C. Fishman

icouzin@ab.mpg.de (I.D.C.)
mark_fishman@harvard.edu (M.C.F.)

HIGHLIGHTS

Genes linked to human psychiatric disorders can alter zebrafish collective behavior

Differences from wild-type lead to "scattered," "coordinated," and "huddled" behavior

Changes in individual interaction rules can explain emergent group level patterns

Tang et al., iScience *23*, 100942
March 27, 2020
https://doi.org/10.1016/j.isci.2020.100942

iScience

CellPress

Article

Genetic Control of Collective Behavior in Zebrafish

Wenlong Tang,[1,6] Jacob D. Davidson,[2,3,4,6] Guoqiang Zhang,[1,6] Katherine E. Conen,[2,3,4] Jian Fang,[1] Fabrizio Serluca,[1] Jingyao Li,[1] Xiaorui Xiong,[1] Matthew Coble,[1] Tingwei Tsai,[1] Gregory Molind,[1] Caroline H. Fawcett,[1] Ellen Sanchez,[1] Peixin Zhu,[1] Iain D. Couzin,[2,3,4,*] and Mark C. Fishman[5,7,*]

SUMMARY

Many animals, including humans, have evolved to live and move in groups. In humans, disrupted social interactions are a fundamental feature of many psychiatric disorders. However, we know little about how genes regulate social behavior. Zebrafish may serve as a powerful model to explore this question. By comparing the behavior of wild-type fish with 90 mutant lines, we show that mutations of genes associated with human psychiatric disorders can alter the collective behavior of adult zebrafish. We identify three categories of behavioral variation across mutants: "scattered," in which fish show reduced cohesion; "coordinated," in which fish swim more in aligned schools; and "huddled," in which fish form dense but disordered groups. Changes in individual interaction rules can explain these differences. This work demonstrates how emergent patterns in animal groups can be altered by genetic changes in individuals and establishes a framework for understanding the fundamentals of social information processing.

INTRODUCTION

Across many species, survival depends on coordination of individuals in a group, such as in a school of fish, flock of birds, or colony of ants. Group dynamics serve, for example, to enhance foraging abilities and confuse predators (Krause et al., 2010). Social interactions are also essential to humans, as shown by the profound dysfunction caused by social processing deficits in psychiatric disorders such as autism and schizophrenia (Henry et al., 2016; Torosyan and Bota, 2017; Gur and Gur, 2016). Work across several species has shown that motion patterns formed by individuals in animal groups can emerge as a consequence of relatively local interactions (Couzin et al., 2002; Tunstrøm et al., 2013; Calovi et al., 2018). Although previous work has provided evidence that schooling behavior in stickleback fish is under genetic control (Peichel Catherine and Marques David, 2017), it remains unclear what attributes of collective behavior might be genetically regulated, and by which specific genes. The genetic, neural, and behavioral techniques developed for zebrafish make them a valuable model for beginning to address these questions (Norton and Bally-Cuif, 2010; Stewart et al., 2014). Zebrafish are highly social animals and exhibit a range of complex behavior (Geng and Peterson, 2019). Several recent studies have investigated a single mutant line in comparison with wild-type (WT) fish, providing evidence that individual genes can alter the interactions and collective behavior of zebrafish (Huang et al., 2019; Liu et al., 2018; Gutiérrez et al., 2019; Kim et al., 2017).

RESULTS

Mutant Zebrafish Groups Exhibit Differences in Swimming Behavior

We used CRISPR-Cas9 to generate mutations in 90 genes associated with psychiatric disorders (Data S1) and performed experiments to ask how these mutations affect the collective behavior of freely swimming zebrafish. We evaluated mutant fish as adults, after the development of the full range of social interactions has matured (Buske and Gerlai, 2011), comparing their collective behavior as they swam in an open circular arena (Figure 1A, Video S1). All mutants were tested as homozygous adults, except for *scn1lab* and *slc18a2*, which could not be raised to adulthood as homozygous and thus were tested as heterozygous fish. We performed multiple trials, with each trial featuring different fish, for each line (Table S1).

Zebrafish tend to swim in groups, sometimes aligning and moving together with others (Figure 1B) and at other times swimming closely and in a disordered configuration (Figure 1C). We found that mutant fish vary in their swimming speed, group spacing, and polarization. Although there was considerable variation between trials performed with a given line, many lines showed consistent trends (Figures 1D–1F).

[1]Novartis Institutes for BioMedical Research, 181 Massachusetts Avenue, Cambridge, MA 02139, USA

[2]Department of Collective Behavior, Max Planck Institute of Animal Behavior, Universitätstraße 10, 78764 Konstanz, Germany

[3]Centre for the Advanced Study of Collective Behavior, University of Konstanz, Universitätstraße 10, 78764 Konstanz, Germany

[4]Department of Biology, University of Konstanz, Universitätstraße 10, 78764 Konstanz, Germany

[5]Department of Stem Cell and Regenerative Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA

[6]These authors contributed equally

[7]Lead Contact

*Correspondence: icouzin@ab.mpg.de (I.D.C.), mark_fishman@harvard.edu (M.C.F.)

https://doi.org/10.1016/j.isci.2020.100942

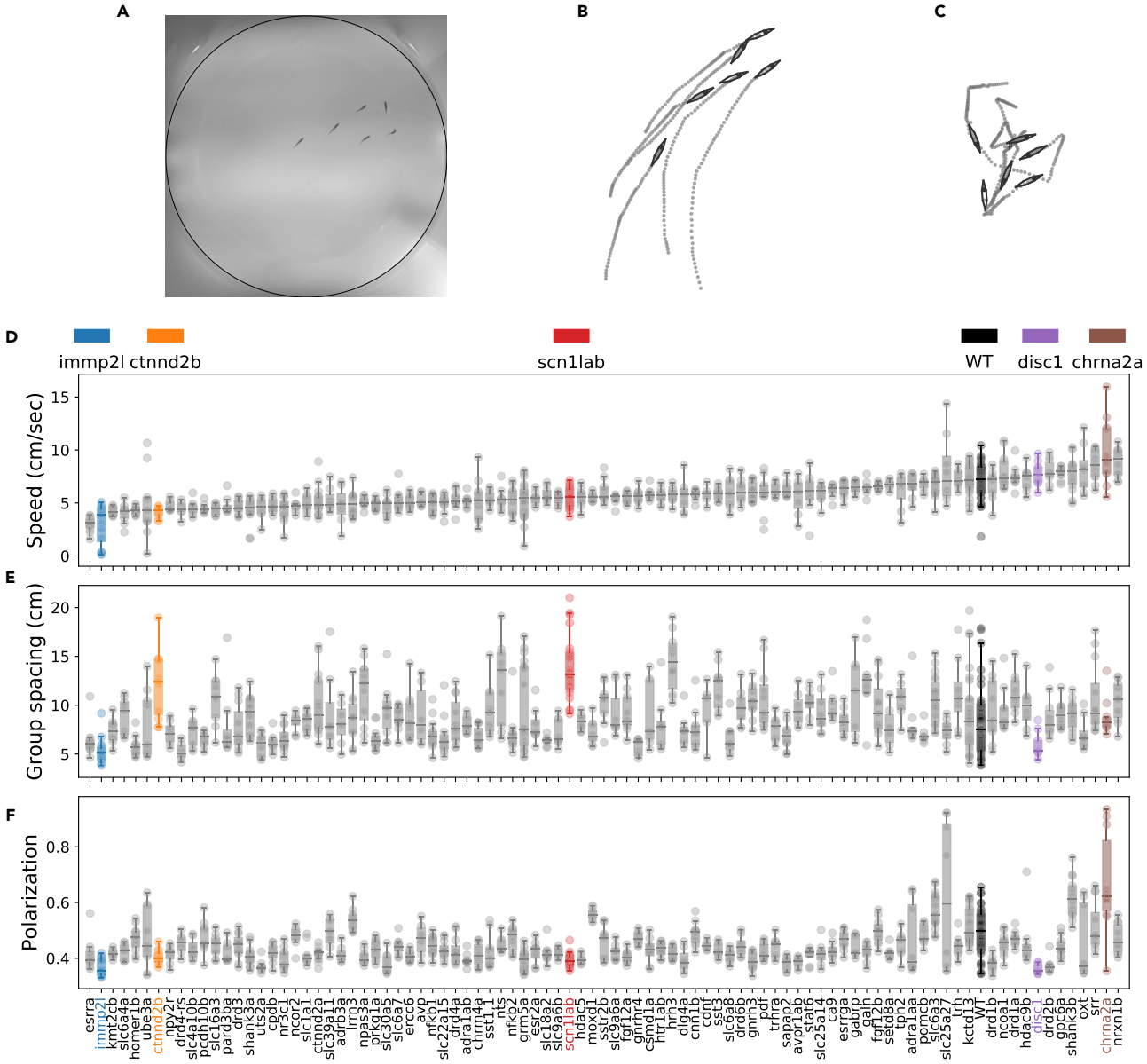


Figure 1. Patterns of Group Behavior Vary Across Mutant Lines

(A) In each trial, six adult fish were filmed from above as they swam in a circular arena.

(B and C) Example trajectories from WT fish, showing groups in aligned (B) and disordered (C) configurations. Each trace shows 1 s of swimming.

(D–F) Box plots of median speed (D), group spacing (E), and polarization (F) for all genetic lines. Each point shows data from one trial. Lines are ordered from lowest to highest speed. Colored points highlight examples of lines that differ from WT in certain aspects of behavior. Figure S1 contains analogous plots for other behavioral metrics.

See also Figure S1, Table S1, Video S1, and Data S1.

Previous work in several fish species has shown that individual speed directly affects motion dynamics, with higher speed being associated with both wider spacing and increased alignment between the individuals (Couzin et al., 2002; Tunstrøm et al., 2013; Jolles et al., 2017). Consistent with this work, most mutant lines with higher swimming speed have larger inter-individual spacing and group polarization (Figures 1D–1F, 2A, and S2). Linear and exponential regression reveals that speed can explain approximately 30% of the total variance in inter-individual spacing, polarization, speed inter-quartile range (IQR), time moving, nearest neighbor distance, and group centroid speed (Figure S2; see Transparent Methods). However, after subtracting the effects of speed and accounting for trial variability, multiple lines continue to show differences from WT (Figure 2B).

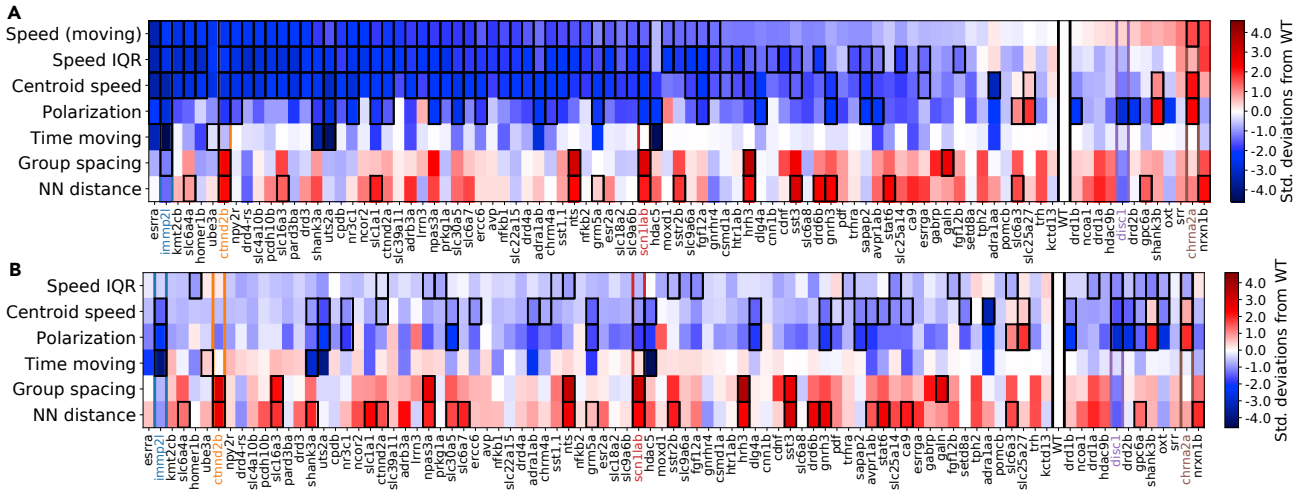


Figure 2. Speed Only Partially Explains Altered Group Behavior across Lines.

Differences between mutant lines and WT in the median values of seven behavioral metrics before (A) and after (B) correcting for speed differences. Lines are listed in order of increasing speed. Colors of five highlighted lines have the same conventions as in Figure 1. Bold outlines indicate statistically significant differences from WT, determined by considering the variability of each quantity across trials for a given line (Dunnett's test, $p < 0.05$). Units refer to the standard deviation of median values across lines.

See also Figure S2.

Three Patterns Describe Distinct Behavioral Differences

We find that three general patterns, scattered, coordinated, and huddled, describe the most distinct differences in movement of the mutant fish. The scattered phenotype has high spacing and low polarization among individuals in the group, exemplified by the mutant lines *scn1lab+/–* (encoding the Nav1.1 protein [Schoonheim et al., 2010]) and *ctnnd2b–/–* (encoding δ-catenin [Turner et al., 2015]). These mutants have high inter-individual spacing, and although they occasionally form groups, they tend to dissociate from each other more frequently and show less collective coordination. The coordinated phenotype describes individuals that exhibit an increased tendency to align their direction of travel and to move coherently as a group, exemplified by *chrna2a–/–* (encoding the α2-nicotinic acetylcholine receptor [Demontis et al., 2019]). The huddled phenotype is characterized by groups of individuals exhibiting low polarization and tight spacing. These groups are generally dense but disordered, and fish spend more time swimming in a relatively local area, as exemplified by *disc1–/–* (encoding disrupted-in-schizophrenia [Blackwood et al., 2001]) and *immp2l–/–* (encoding the inner mitochondrial membrane peptidase2-like protein [Casey et al., 2012]).

We use principal component analysis (PCA) on the median values of the behavioral metrics for each line after controlling for speed (Figure 2B) to describe and quantify the different patterns of group behavior (Figure 3). We find that the first two components reveal much of the relative behavioral differences between lines, showing differences between the scattered, coordinated, or huddled phenotypes. Note that, although Figure 3 highlights five mutant lines that exhibit clear differences along the PCA dimensions, other lines also exhibit differences from WT (Figures 3C and S3). Accounting for trial variability and limited sampling using a bootstrap procedure shows that some lines are separated from WT in one or both of the first two PCA components, whereas other lines overlap with WT (Figure S3). Although the first two components reveal the largest fraction of variation, some lines show distinct differences in the third PCA component. For example, although both display the huddled phenotype, unlike *disc1–/–* fish, *immp2l–/–* fish spend time without moving, or are "frozen." Because of this these two lines have opposite signs for PC3 (Figure 3B) and are separated in the "freezing component" by using a modified PCA procedure (Figure S4; see Transparent Methods).

Model Fit Connects Group Behavior to Individual Interaction Rules

Evidence suggests that collective motion dynamics can be explained by relatively simple rules of interaction between individuals, such as how an individual fish alters its trajectory depending on the locations of its

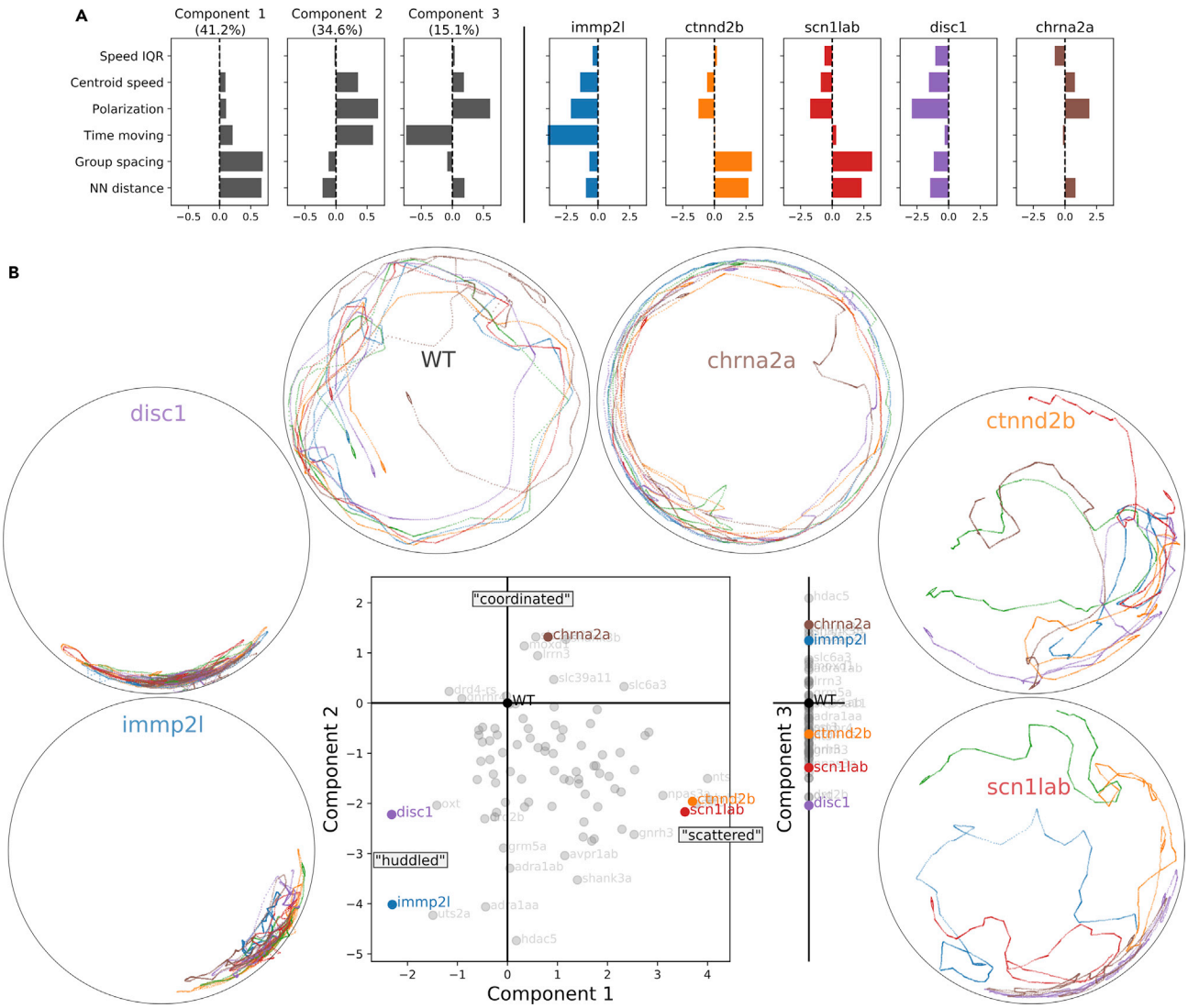


Figure 3. PCA Reveals Categories of Behavioral Variability in Genetic Mutants

(A) Left: After accounting for speed, 90.8% of the remaining variation across lines is described by three orthogonal components. The first component reflects a change in group spacing. The second and third components reflect combinations of polarization and fraction of time moving: positive values of PC2 correspond to high polarization and increased time moving, whereas high values of PC3 correspond to high polarization and frequent freezing. Right: Values of the input metrics for five highlighted lines, relative to WT.

(B) Center: projection of the data for all genetic lines onto the first two components; projection onto the third component shown separately on the right-hand side. Edges: example trajectories for WT and five highlighted lines (1,000 frames, ~17 s). The "scattered" phenotype is described by positive values of the first PCA component (PC1) and negative values of the second component (PC2). The "coordinated" phenotype is described by positive values of PC2. The "huddled" phenotype is described by negative values of PC1 and PC2. Table S2 lists each line in order of position along each PC.

See also Figures S3 and S4 and Table S2.

neighbors (Couzin et al., 2002; Calovi et al., 2018; Heras et al., 2018; Harpaz et al., 2017; Katz et al., 2011; Zienkiewicz et al., 2018). We use a model to ask if salient observed differences in collective motion (i.e., the scattered, huddled, and coherent collective motion patterns) can be explained by differences in how individuals turn in response to neighboring fish. The model uses neighbor positions and velocities to determine the effective attraction and alignment forces that best predict whether a fish will turn left or right after a specified time delay (see Transparent Methods for details). We examined differences in the distance-dependence of attraction ($G_{att}(r_j)$), the distance-dependence of alignment ($G_{ali}(r_j)$), and the relative overall strength of attraction to alignment (α) for WT fish and for mutant lines that exemplified the three distinctive group patterns: scattered (*scn1lab*+/-), coordinated (*chrna2a*−/−), and huddled (*disc1*−/−).

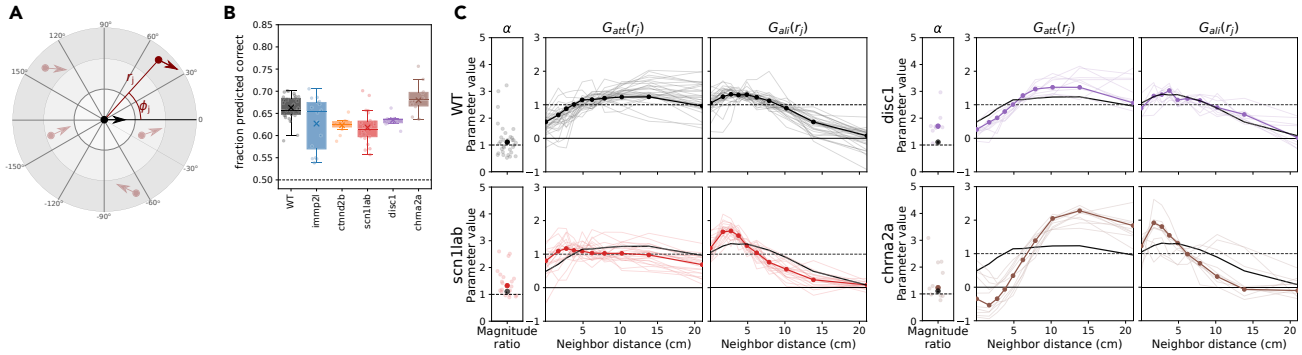


Figure 4. A Simple Model Reveals Differences in Individual-Level Interaction Rules across Lines

(A) Schematic of one focal fish (black) at a given point in time, with five neighbors (red) at different locations and around it. Arrows indicate heading. For each fish in the group, the model takes relative neighbor positions and headings as input to predict whether the individual will turn left or right after a specified time delay of 1 s for WT (delay time adjusted for different speeds of the mutant lines—see Transparent Methods). The fit yields the interaction functions for effective attraction and alignment forces that best predict movement decisions across all fish in a trial.

(B) Predictive ability of the model, showing boxplots for the distribution of fraction of turns predicted correctly for different trials with each line.

(C) Fits to WT and mutant lines *scn1lab+/-*, *disc1−/−*, and *chrna2a−/−*. The attraction/alignment ratio α is shown along with the distance dependence of attraction $G_{att}(r_j)$ and the distance dependence of alignment $G_{ali}(r_j)$. The bold line shows the fit to all trials together, and the thin lines show fits to individual trials. For fits to the mutant lines, the overall fit to WT is shown in solid black for comparison. Compared with WT, *scn1lab+/−* fish exhibit a weaker distance-dependence for attraction and a more rapid decrease in alignment as a function of distance. In contrast, *chrna2a−/−* mutants exhibit a relatively strong attraction to distant neighbors. The interaction functions for *disc1−/−* are similar to WT overall but yield a higher ratio of attraction to alignment.

See also Figure S5.

For WT fish, the effective "force" of attraction to neighbors increases with distance, whereas alignment decreases with distance (Figure 4C). These trends agree with previous work on fish (Calovi et al., 2018). Compared with WT, *scn1lab+/−* fish exhibit a weaker distance-dependence for attraction and a more rapid decrease in alignment as a function of distance (Figure 4C). This suggests that *scn1lab+/−* fish interact predominantly with close neighbors and tend to ignore those at a distance, leading to the group instability and scattered behavior we observe in these lines. Model fits to *ctnnd2b−/−* mutants show a similar trend (Figure S5B). Notably, the model predicts turning decisions of *scn1lab+/−* fish with lower accuracy than WT fish, indicating that the motor decisions of *scn1lab+/−* fish are less consistently affected by neighboring fish (Figure 4B). In contrast, *chrna2a−/−* mutants move more predictably than do WT and exhibit a relatively strong attraction to distant neighbors (Figure 4C), which accounts for the cohesive nature of groups formed by these lines. The interaction functions for *disc1−/−* are similar to WT overall but yield a higher ratio of attraction to alignment (Figure 4C), and fits to *immp2l−/−* mutants show a similar trend (Figure S5B). This change in social responsiveness results in individuals that stay close to one another but do not move together as a coherent group. Overall, our results suggest that differences in how individual fish respond to their neighbors can lead to the different observed group-level patterns.

DISCUSSION

This work uses quantitative behavioral metrics to show how genetics may direct patterns of collective behavior. The patterns that arise in groups—their structure, cohesion, leadership, and dynamics—contribute to species fitness and adaptation to environmental changes and hence to their evolution. This study establishes a fundamental framework for understanding the relationship between genes, social interaction, and sensorimotor transformations. Prior work has demonstrated that swimming speed alone can drive changes in the shoal structure and dynamics (Couzin et al., 2002; Tunstrøm et al., 2013; Jolles et al., 2017), and we find here that mutations that alter speed do, in general, change the behavior of the group in predictable ways. However, we also discovered mutations with effects on the group pattern and dynamics dissociated from the effects of speed. These fall into three patterns, coordinated, scattered, and huddled, which each describe the behavior of several mutant zebrafish lines and can arise from differences in interaction rules among individuals.

In our results we highlighted five lines with distinctive behavioral differences from WT: *scn1lab+/−*, *ctnnd2b−/−*, *chrna2a−/−*, *immp2l−/−*, and *disc1−/−*. However, we note that, in addition to these, other mutant lines also exhibited distinct differences from WT (Figures 2, 3, and S3). We also note the possibility

that some genes did not manifest a behavioral phenotype because of the activity of paralogs (Hoffman et al., 2016) or the transcriptional upregulation of paralogs or other genes triggered by the CRISPR-Cas9 (El-Brolosy et al., 2019).

The effect of the mutations upon the individual fish may be upon sensory, motor, or integrative processes, and we cannot suggest that the affected genes are specifically "social" in their roles, but rather that such effects upon individuals are manifest as changes in group behavior. For example, it is known that larvae of homozygous deficient *scn1lab* mutant fish (as opposed to the heterozygous fish studied here) partially lose the ability to maintain eye positions following saccades associated with the optokinetic reflex (Schoonheim et al., 2010) and have increased levels of swim activity as individual larvae (Baraban et al., 2013). In future work it will be important to define the nature of the sensorimotor transformations and corresponding neural activity that underlie collective and social behavior. With transparent larval zebrafish, whole-brain imaging can examine neural activity from stimulus onset to the subsequent motor output (Chen et al., 2018). New tools, including the use of fixed fish combined with virtual reality, may permit similar evaluations related to collective behavior in adult fish (Huang et al., 2019; Stowers et al., 2017).

Several mutations that generate distinctive patterns of collective behavior are in genes that have been linked to human psychiatric illnesses. For example, variants of *CTNND2* have been associated with autism (Turner et al., 2015); *DISC1* with schizophrenia, bipolar disease, and major depressive disorder (Blackwood et al., 2001); *CHRNA2* with nicotine and cannabis abuse (Demontis et al., 2019); *SCN1A* with autism and epilepsy ("Dravet syndrome") (Catterall et al., 2010); and, in some studies, *IMMP2L* with autism (Casey et al., 2012). Although we are cautious in relating collective behavior in fish to any particular social phenotype in humans, we note that elements of the "social brain" are known to be conserved from fish to human (O'Connell and Hofmann, 2012; Geng and Peterson, 2019). In addition, social interaction assays may provide sensitive means to evaluate effects of pharmacological agents upon behavioral deficits in genetically defined animals. Drug screening for reversal of other behavioral phenotypes in fish has revealed agents used successfully to treat human psychiatric disorders (Baraban et al., 2013; Wang et al., 2019; Hoffman et al., 2016).

Limitations of the Study

In this study, we identify mutations in several zebrafish genes that lead to altered collective swimming behavior. Although we show that changes in swimming speed do not account for all behavioral variation, our study does not test whether other changes in sensory processing or affective response interact with social motivation to give rise to these changes. In addition, although we note three general patterns that describe how behavior varies across lines, this study does not test whether similar behavioral variations reflect changes in the same neuroanatomical and cellular pathways. Future experiments examining the anatomical changes and neural activity in distinctive mutant lines will be needed to resolve this question. Although we examined variability across trials for each line, we did not examine variability among individual fish due to challenges in maintaining individual identities in the tracking. Finally, although we generate an individual-level model that accounts for group-level differences in behavior, this study does not directly test how well the effective forces in the model reflect the underlying decision processes in fish. Subsequent work will be needed to test the specific forms of the interaction functions in the model and to ask if differences for mutant fish are due to social response or due to differences in sensory, motor, or integrative processes.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY

Data are available at doi.org/10.5061/dryad.hx3ffbg9n. Code is available at github.com/jacobdavidson/zfish_mutants_analysis/.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.100942.

ACKNOWLEDGMENTS

Funding provided by research grants from FBRI (F-Prime Capital) and Novartis Institutes of Biomedical Research to M.C.F. J.D.D., K.E.C., I.D.C. acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2117 - 422037984. J.D.D. acknowledges support from the Heidelberg Academy of Science. I.D.C. acknowledges support from NSF Grant IOS-1355061 and Office of Naval Research Grants N00014-09-1-1074 and N00014-14-1-0635.

AUTHOR CONTRIBUTIONS

M.C.F. conceptualized and supervised the study; M.C.F, W.T., and G.Z. designed experiments; W.T., G.Z., F.S., G.M., J.L., X.X., M.C., T.T., C.H.F., E.S., J.F., and P.Z. performed experiments; W.T., J.D.D., G.Z., K.E.C., I.D.C., and M.C.F. analyzed the data; J.D.D., K.E.C., and I.D.C. developed the model; W.T., G.Z., and M.C.F wrote an early version of the manuscript; J.D.D., K.E.C., I.D.C., and M.C.F. wrote the final version of the manuscript.

DECLARATION OF INTERESTS

Mark Fishman: Boards of Directors of Beam Therapeutics and Geneception; Founder of AditumBio; SAB of Tenaya Therapeutics. Other authors have no competing interests to declare.

Received: November 14, 2019
Revised: January 17, 2020
Accepted: February 21, 2020
Published: March 27, 2020

REFERENCES

Baraban, S.C., Dinday, M.T., and Hortopan, G.A. (2013). Drug screening in scn1a zebrafish mutant identifies clemizole as a potential dravet syndrome treatment. Nat. Commun. *4*, 2410.

Blackwood, D.H., Fordyce, A., Walker, M.T., St Clair, D.M., Porteous, D.J., and Muir, W.J. (2001). Schizophrenia and affective disorders–cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and p300 findings in a family. Am. J. Hum. Genet. *69*, 428–433.

Buske, C., and Gerlai, R. (2011). Shoaling develops with age in Zebrafish (*Danio rerio*). Prog. Neuro Psychopharmacol. Biol. Psychiatry *35*, 1409–1415.

Calovi, D.S., Litchinko, A., Lecheval, V., Lopez, U., Escudero, A.P., Chaté, H., Sire, C., and Theraulaz, G. (2018). Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors. PLoS Comput. Biol. *14*, e1005933.

Casey, J.P., Magalhaes, T., Conroy, J.M., Regan, R., Shah, N., Anney, R., Shields, D.C., Abrahams, B.S., Almeida, J., Bacchelli, E., et al. (2012). A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. Hum. Genet. *131*, 565–579.

Catterall, W.A., Kalume, F., and Oakley, J.C. (2010). Nav1.1 channels and epilepsy. J. Physiol. *588*, 1849–1859.

Chen, X., Mu, Y., Hu, Y., Kuan, A.T., Nikitchenko, M., Randlett, O., Chen, A.B., Gavornik, J.P., Sompolinsky, H., Engert, F., and Ahrens, M.B. (2018). Brain-wide organization of neuronal activity and convergent sensorimotor transformations in larval zebrafish. Neuron *100*, 876–890.e5.

Couzin, I.D., Krause, J., James, R., Ruxton, G.M., and Franks, N.R. (2002). Collective memory and spatial sorting in animal groups. J. Theor. Biol. *218*, 1–11.

Demontis, D., Rajagopal, V.M., Thorgeirsson, T.E., Als, T.D., Grove, J., Leppala, K., Gudbjartsson, D.F., Pallesen, J., Hjorthoj, C., Reginsson, G.W., et al. (2019). Genome-wide association study implicates chrna2 in cannabis use disorder. Nat. Neurosci. *22*, 1066–1074.

El-Brolosy, M.A., Kontarakis, Z., Rossi, A., Kuenne, C., Günther, S., Fukuda, N., Kikhi, K., Boezio, G.L.M., Takacs, C.M., Lai, S.L., et al. (2019). Genetic compensation triggered by mutant mRNA degradation. Nature *568*, 193–197.

Geng, Y., and Peterson, R.T. (2019). The zebrafish subcortical social brain as a model for studying social behavior disorders. Dis. Models Mech. *12*, dmm039446.

Gur, R.C., and Gur, R.E. (2016). Social cognition as an RDoC domain. Am. J. Med. Genet. B: Neuropsychiatr. Genet. *171*, 132–141.

Gutiérrez, H.C., Colanesi, S., Cooper, B., Reichmann, F., Young, A.M.J., Kelsh, R.N., and Norton, W.H.J. (2019). Endothelin neurotransmitter signalling controls zebrafish social behaviour. Sci. Rep. *9*, 1–17.

Harpaz, R., Tkačik, G., and Schneidman, E. (2017). Discrete modes of social information processing predict individual behavior of fish in a group. Proc. Natl. Acad. Sci. U S A *114*, 10149–10154.

Henry, J.D., von Hippel, W., Molenberghs, P., Lee, T., and Sachdev, P.S. (2016). Clinical assessment of social cognitive function in neurological disorders. Nat. Rev. Neurol. *12*, 28–39.

Heras, F.J.H., Romero-Ferrero, F., Hinz, R.C., and Polavieja, G.G.d. (2018). Deep attention networks reveal the rules of collective motion in zebrafish. bioRxiv, 400747, https://doi.org/10.1101/400747.

Hoffman, E.J., Turner, K.J., Fernandez, J.M., Cifuentes, D., Ghosh, M., Ijaz, S., Jain, R.A., Kubo, F., Bill, B.R., Baier, H., et al. (2016). Estrogens suppress a behavioral phenotype in zebrafish mutants of the autism risk gene, CNTNAP2. Neuron *89*, 725–733.

Huang, K.H., Rupprecht, P., Schebesta, M., Serluca, F., Kitamura, K., Bouwmeester, T., and Friedrich, R.W. (2019). Predictive neural processing in adult zebrafish depends on shank3b. bioRxiv, 546457, https://doi.org/10.1101/546457.

Jolles, J.W., Boogert, N.J., Sridhar, V.H., Couzin, I.D., and Manica, A. (2017). Consistent individual differences drive collective behavior and group functioning of schooling fish. Curr. Biol. *27*, 2862–2868.e7.

Katz, Y., Tunstrøm, K., Ioannou, C.C., Huepe, C., and Couzin, I.D. (2011). Inferring the structure and dynamics of interactions in schooling fish. Proc. Natl. Acad. Sci. U S A *108*, 18720–18725.

Kim, O.H., Cho, H.J., Han, E., Hong, T.I., Ariyasiri, K., Choi, J.H., Hwang, K.S., Jeong, Y.M., Yang, S.Y., Yu, K., et al. (2017). Zebrafish knockout of Down syndrome gene, DYRK1a, shows social impairments relevant to autism. Mol. Autism *8*, 50.

Krause, J., Ruxton, G.D., and Krause, S. (2010). Swarm intelligence in animals and humans. Trends Ecol. Evol. *25*, 28–34.

Liu, C.x., Li, C.y., Hu, C.c., Wang, Y., Lin, J., Jiang, Y.h., Li, Q., and Xu, X. (2018). CRISPR/Cas9-induced shank3b mutant zebrafish display autism-like behaviors. Mol. Autism *9*, 23.

Norton, W., and Bally-Cuif, L. (2010). Adult zebrafish as a model organism for behavioural genetics. BMC Neurosci. *11*, 90.

O'Connell, L.A., and Hofmann, H.A. (2012). Evolution of a vertebrate social decision-making network. Science *336*, 1154–1157.

Peichel Catherine, L., and Marques David, A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. Philos. Trans. R. Soc. B Biol. Sci. *372*, 20150486.

Schoonheim, P.J., Arrenberg, A.B., Bene, F.D., and Baier, H. (2010). Optogenetic localization and genetic perturbation of saccade-generating neurons in zebrafish. J. Neurosci. *30*, 7111–7120.

Stewart, A.M., Nguyen, M., Wong, K., Poudel, M.K., and Kalueff, A.V. (2014). Developing zebrafish models of autism spectrum disorder (ASD). Prog. Neuro Psychopharmacol. Biol. Psychiatry *50*, 27–36.

Stowers, J.R., Hofbauer, M., Bastien, R., Griessner, J., Higgins, P., Farooqui, S., Fischer, R.M., Nowikovsky, K., Haubensak, W., Couzin, I.D., et al. (2017). Virtual reality for freely moving animals. Nat. Methods *14*, 995–1002.

Torosyan, N., and Bota, R.G. (2017). Social cognition in schizophrenia. Ment. Illn. *9*, https://doi.org/10.4081/mi.2017.7228.

Tunstrøm, K., Katz, Y., Ioannou, C.C., Huepe, C., Lutz, M.J., and Couzin, I.D. (2013). Collective states, multistability and transitional behavior in schooling fish. PLoS Comput. Biol. *9*, e1002915.

Turner, T.N., Sharma, K., Oh, E.C., Liu, Y.P., Collins, R.L., Sosa, M.X., Auer, D.R., Brand, H., Sanders, S.J., Moreno-De-Luca, D., et al. (2015). Loss of delta-catenin function in severe autism. Nature *520*, 51–56.

Wang, G., Zhang, G., Li, Z., Fawcett, C.H., Coble, M., Sosa, M.X., Tsai, T., Malesky, K., Thibodeaux, S.J., Zhu, P., et al. (2019). Abnormal behavior of zebrafish mutant in dopamine transporter is rescued by clozapine. iScience *17*, 325–333.

Zienkiewicz, A.K., Ladu, F., Barton, D.A.W., Porfiri, M., and Bernardo, M.D. (2018). Data-driven modelling of social forces and collective behaviour in zebrafish. J. Theor. Biol. *443*, 39–51.

iScience, Volume *23*

Supplemental Information

Genetic Control of Collective Behavior in Zebrafish

Wenlong Tang, Jacob D. Davidson, Guoqiang Zhang, Katherine E. Conen, Jian Fang, Fabrizio Serluca, Jingyao Li, Xiaorui Xiong, Matthew Coble, Tingwei Tsai, Gregory Molind, Caroline H. Fawcett, Ellen Sanchez, Peixin Zhu, Iain D. Couzin, and Mark C. Fishman

Supplemental figures and tables

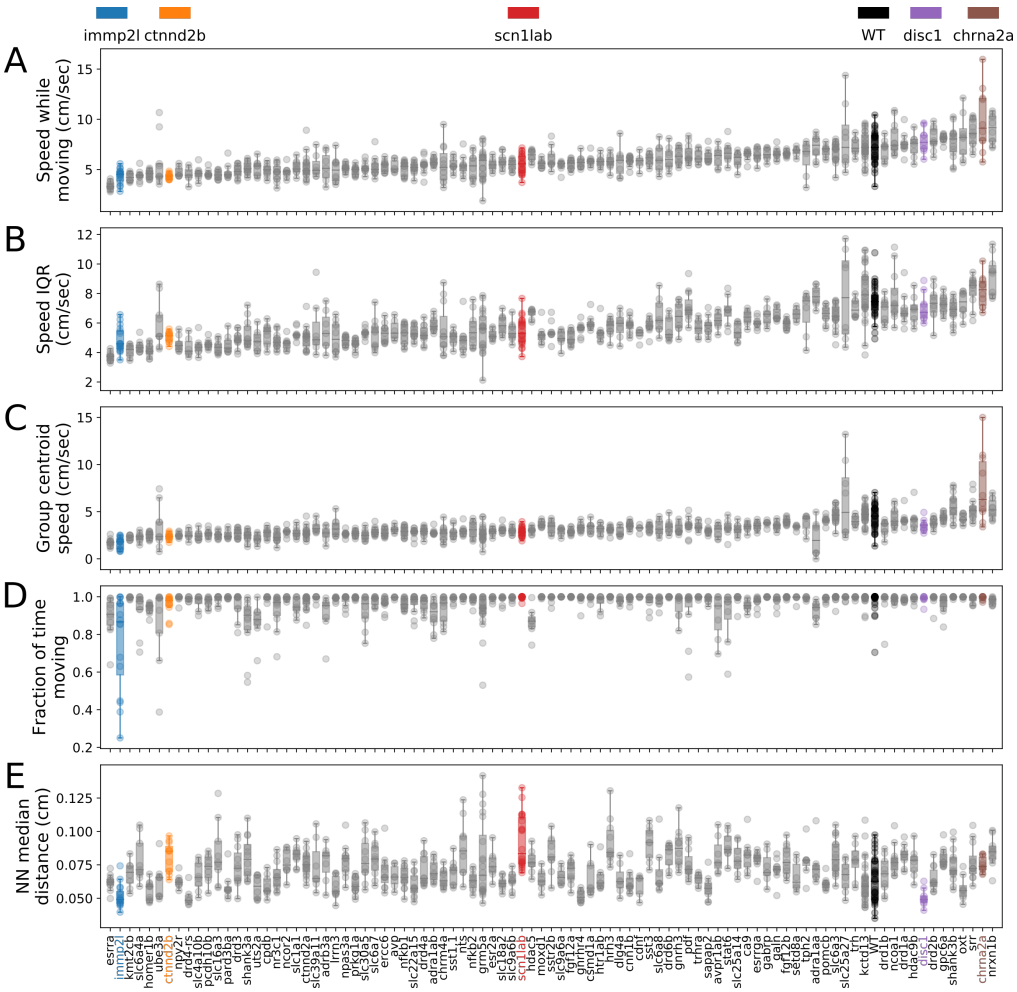


Figure S1. Boxplots of other behavioral metrics. Related to Fig 1. Analogous box plots to Fig 1D-F are shown for the behavioral metrics of *speed while moving* (A), *speed inter-quartile range* (B), *group centroid speed* (C), *fraction of time moving* (D), and *nearest neighbor median distance* (E). See Methods for definitions of these quantities. Each point shows data from one trial. Lines are ordered from lowest to highest speed. As in the main text, colored points highlight examples of lines that differ from WT in certain aspects of behavior.

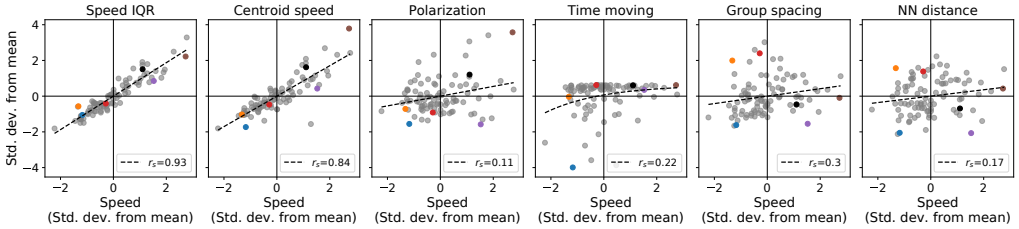


Figure S2. Relationship of speed with each behavioral metric. Related to Fig 2. Units refer to the standard deviation of the median values of each quantity across lines. A linear regression was used for all quantities except time moving, which uses an exponential function (see Methods). The value of Spearman's correlation, r_s, is shown on each plot. Colored points correspond to highlighted lines, using the same color conventions as Fig S1 and Figs 1-4 in the main text.

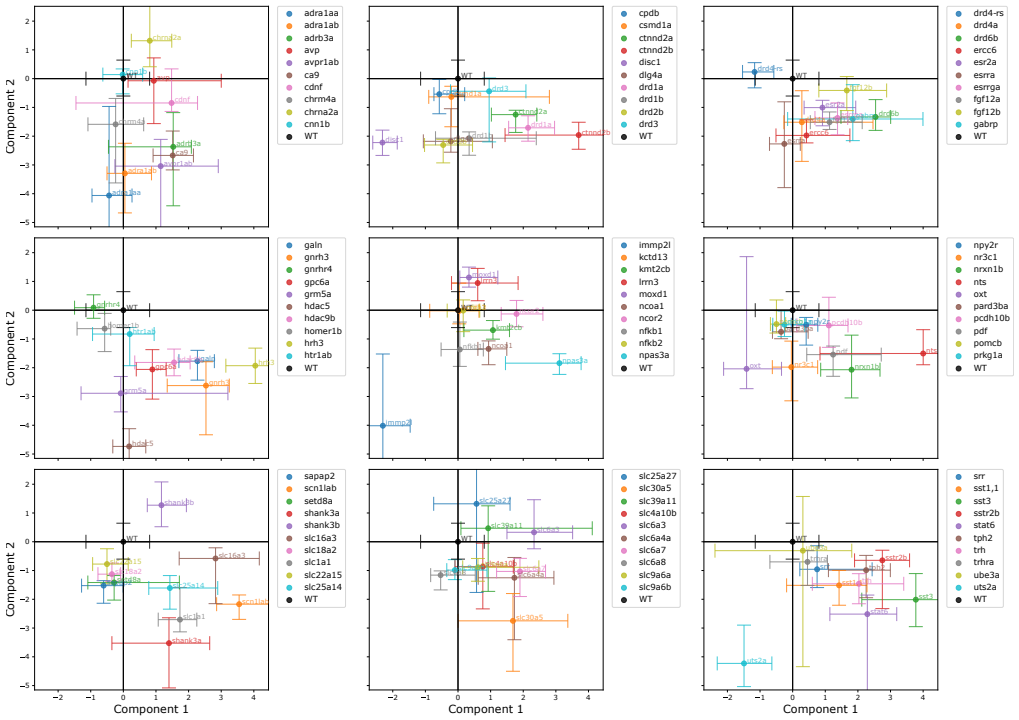


Figure S3. Expanded images of PCA embedding with confidence intervals. Related to Fig 3. Plots of the first two PCA components for mutants lines compared to WT, with each plot showing a subset of 10 mutants compared to WT (sorted alphabetically). Error bars show the 90% confidence intervals calculated using a bootstrap procedure (see Methods).

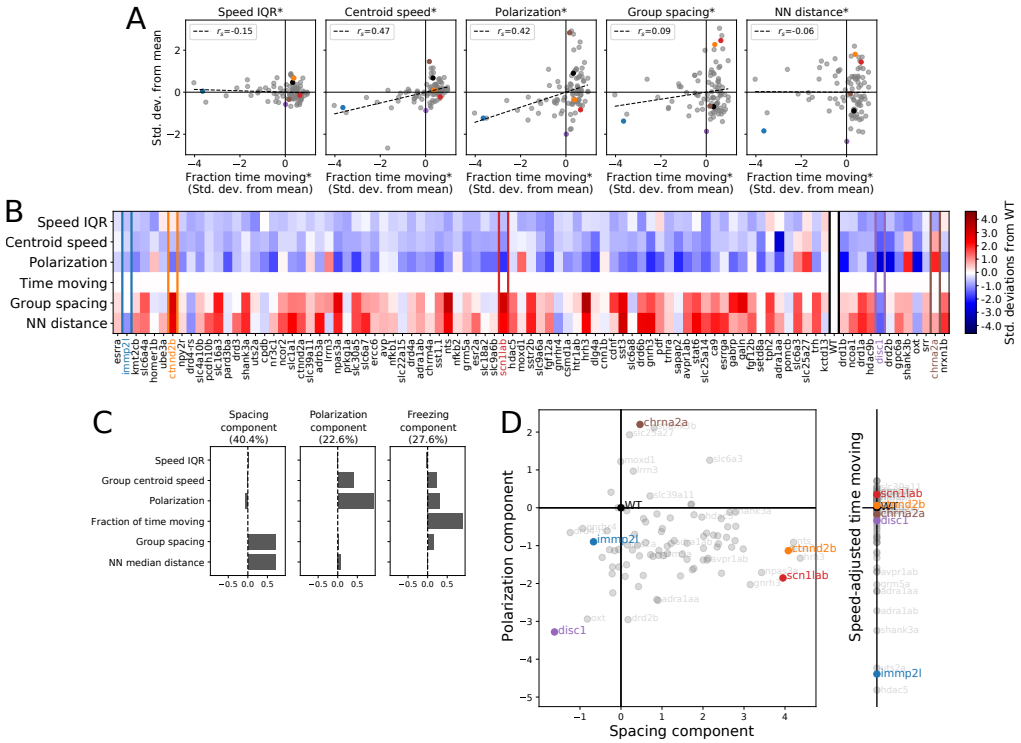


Figure S4. PCA with freezing-independent components. Related to Fig 3. Similar to speed, some of the components showed a correlation with fraction of time moving (Fig S4). We used a modified procedure to account for this (see Methods). (A) After subtracting the regression fit to speed from each metric (Fig S2), some of adjusted metrics still have correlations with the adjusted time moving. A linear relationship is used to capture changes of each adjusted metric with adjusted time moving. The value of Spearman's correlation, r_s, is shown on each plot. Colored points correspond to highlighted lines, using the same color conventions as Fig S1 and Figs 1-4 in the main text. (B) Remaining difference from WT, after subtracting correlations with adjusted time moving. Units refer to the standard deviation of median values across lines. (C) The freezing component is defined by the correlations shown in (A), and the spacing and polarization components explain the remaining variance. (D) Embedding patterns of the freezing-independent spacing and polarization components, along with the freezing component.

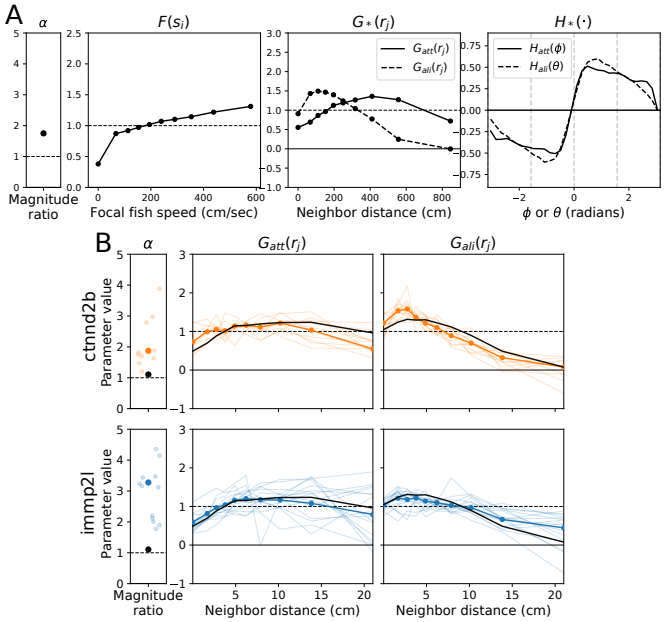


Figure S5. Model fit to combined data set and other lines. Related to Fig 4. The model was first fit to a dataset contain all data from the 5 highlighted lines plus WT (A). The functions $F(s_i)$ and $H_*(\cdot)$ were used from this fit and held constant for the fits that compare the distance-dependent interactions functions across lines. See Methods for details on the fit procedure. (B) Fit results analogous to Fig 4 are shown for two other highlighted lines, *ctnnd2b-/-* and *immp2l-/-*.

Line	No. trials	Line	No. trials	Line	No. trials	Line	No. trials	Line	No. trials	Line	No. trials
WT	33	dlg4a	7	galn	9	ncoa1	12	sapap2	9	slc6a4a	9
adra1aa	8	drd1a	9	gnrh3	10	ncor2	8	scn1lab	22	slc6a7	11
adra1ab	10	drd1b	9	gnrhr4	9	nfkb1	13	setd8a	6	slc6a8	8
adrb3a	7	drd2b	10	gpc6a	12	nfkb2	10	shank3a	12	slc9a6a	7
avp	6	drd3	11	grm5a	20	npas3a	11	shank3b	14	slc9a6b	9
avpr1ab	7	drd4-rs	6	hdac5	8	npy2r	8	slc16a3	9	srr	12
ca9	8	drd4a	11	hdac9b	8	nr3c1	12	slc18a2	9	sst1,1	9
cdnf	5	drd6b	8	homer1b	11	nrxn1b	10	slc1a1	9	sst3	9
chrm4a	11	ercc6	11	hrh3	10	nts	9	slc22a15	12	sstr2b	9
chrna2a	10	esr2a	8	htr1ab	8	oxt	8	slc25a14	8	stat6	8
cnn1b	11	esrra	11	immp2l	16	pard3ba	8	slc25a27	8	tph2	7
cpdb	9	esrrga	9	kctd13	28	pcdh10b	9	slc30a5	10	trh	8
csmd1a	8	fgf12a	11	kmt2cb	8	pdf	11	slc39a11	7	trhra	7
ctnnd2a	17	fgf12b	11	lrrn3	10	pomcb	9	slc4a10b	8	ube3a	9
ctnnd2b	11	gabrp	7	moxd1	8	prkg1a	12	slc6a3	18	uts2a	8
disc1	9										

Table S1. Number of trial performed for each mutant line. Related to Fig 1. Lines are sorted alphabetically.

Component 1		Component 2		Component 3	
hrh3	4.04	slc25a27	1.32	hdac5	2.09
nts	4.0	**chrna2a**	**1.32**	**chrna2a**	**1.56**
sst3	3.78	shank3b	1.27	shank3a	1.47
ctnnd2b	**3.7**	moxd1	1.14	shank3b	1.4
scn1lab	**3.55**	lrrn3	0.94	slc25a27	1.32
npas3a	3.11	slc39a11	0.47	**immp2l**	**1.25**
slc16a3	2.83	slc6a3	0.33	uts2a	1.22
sstr2b	2.75	drd4-rs	0.23	slc6a3	0.87
drd6b	2.54	cnn1b	0.14	homer1b	0.83
gnrh3	2.53	gnrhr4	0.09	esrra	0.78
slc6a3	2.33	**WT**	**0.0**	moxd1	0.7
stat6	2.29	nfkb2	-0.01	adra1ab	0.65
galn	2.27	kctd13	-0.02	ca9	0.46
tph2	2.25	avp	-0.07	esrrga	0.44
drd1a	2.15	ncor2	-0.13	slc1a1	0.43
trh	2.03	ube3a	-0.31	nfkb1	0.41
slc6a7	1.9	fgf12b	-0.41	lrrn3	0.36
gabrp	1.84	drd3	-0.44	adrb3a	0.36
nrxn1b	1.81	pomcb	-0.48	grm5a	0.17
ncor2	1.79	prkg1a	-0.5	fgf12b	0.15
ctnnd2a	1.77	npy2r	-0.51	drd1a	0.11
slc1a1	1.73	pcdh10b	-0.53	ncor2	0.02
slc6a4a	1.73	cpdb	-0.54	**WT**	**0.0**
slc30a5	1.69	slc16a3	-0.58	slc39a11	-0.01
fgf12b	1.66	csmd1a	-0.63	pcdh10b	-0.03
hdac9b	1.56	homer1b	-0.64	avpr1ab	-0.03
adrb3a	1.53	sstr2b	-0.65	nfkb2	-0.04
ca9	1.51	kmt2cb	-0.69	stat6	-0.04
cdnf	1.48	trhra	-0.7	kctd13	-0.06
slc25a14	1.43	pard3ba	-0.75	cnn1b	-0.07
sst1,1	1.42	slc22a15	-0.78	tph2	-0.07
shank3a	1.4	htr1ab	-0.83	sstr2b	-0.1
esrrga	1.38	cdnf	-0.85	nrxn1b	-0.11
pdf	1.24	slc4a10b	-0.87	srr	-0.16
shank3b	1.17	slc9a6a	-0.89	avp	-0.23
avpr1ab	1.15	srr	-0.96	ncoa1	-0.24
fgf12a	1.13	slc9a6b	-0.99	slc6a7	-0.29
pcdh10b	1.11	tph2	-0.99	gpc6a	-0.3
kmt2cb	1.07	esr2a	-1.01	adra1aa	-0.3
drd3	0.96	slc6a7	-1.03	chrm4a	-0.31
ncoa1	0.94	slc18a2	-1.13	pomcb	-0.32
avp	0.93	slc6a8	-1.16	slc6a4a	-0.36
slc39a11	0.93	ctnnd2a	-1.25	ercc6	-0.38
esr2a	0.91	slc6a4a	-1.25	drd3	-0.42
gpc6a	0.89	drd6b	-1.33	drd6b	-0.43
chrna2a	**0.81**	ncoa1	-1.35	ube3a	-0.5
slc4a10b	0.76	nfkb1	-1.37	sst3	-0.52
srr	0.75	esrrga	-1.37	gnrhr4	-0.54
slc9a6a	0.61	gabrp	-1.4	slc16a3	-0.54
lrrn3	0.6	setd8a	-1.42	esr2a	-0.55
slc25a27	0.57	trh	-1.46	drd4a	-0.56
trhra	0.46	nts	-1.51	trh	-0.57
ercc6	0.42	fgf12a	-1.51	galn	-0.58
npy2r	0.41	drd4a	-1.52	slc30a5	-0.6
drd1b	0.34	sst1,1	-1.52	**ctnnd2b**	**-0.62**
moxd1	0.34	sapap2	-1.52	trhra	-0.64
ube3a	0.31	pdf	-1.54	ctnnd2a	-0.64
drd4a	0.28	chrm4a	-1.59	htr1ab	-0.66
htr1ab	0.2	slc25a14	-1.61	drd4-rs	-0.67
hdac5	0.18	drd1a	-1.71	hdac9b	-0.68
nfkb2	0.16	galn	-1.77	slc22a15	-0.69
nfkb1	0.06	hdac9b	-1.82	nts	-0.7
adra1ab	0.05	npas3a	-1.84	pdf	-0.78
kctd13	0.04	hrh3	-1.93	nr3c1	-0.8
WT	**0.0**	**ctnnd2b**	**-1.96**	slc4a10b	-0.8
cnn1b	-0.02	ercc6	-1.97	cdnf	-0.82
nr3c1	-0.04	nr3c1	-1.98	csmd1a	-0.83
grm5a	-0.08	sst3	-2.01	prkg1a	-0.87
slc9a6b	-0.1	oxt	-2.04	slc6a8	-0.92
csmd1a	-0.2	gpc6a	-2.06	slc25a14	-0.93
dlg4a	-0.22	nrxn1b	-2.07	kmt2cb	-0.96
chrm4a	-0.24	drd1b	-2.07	slc9a6a	-0.96
prkg1a	-0.25	**scn1lab**	**-2.17**	hrh3	-0.96
esrra	-0.25	dlg4a	-2.18	slc18a2	-0.98
setd8a	-0.28	**disc1**	**-2.22**	gnrh3	-0.98
slc18a2	-0.35	esrra	-2.27	gabrp	-1.0
pard3ba	-0.35	drd2b	-2.31	cpdb	-1.01
adra1aa	-0.44	adrb3a	-2.37	slc9a6b	-1.01
drd2b	-0.45	stat6	-2.52	dlg4a	-1.02
pomcb	-0.5	gnrh3	-2.62	npy2r	-1.08
slc22a15	-0.5	ca9	-2.67	sst1,1	-1.1
slc6a8	-0.53	slc1a1	-2.71	setd8a	-1.12
cpdb	-0.57	slc30a5	-2.75	fgf12a	-1.19
homer1b	-0.57	grm5a	-2.89	npas3a	-1.25
sapap2	-0.6	avpr1ab	-3.04	**scn1lab**	**-1.29**
gnrhr4	-0.91	adra1ab	-3.29	pard3ba	-1.39
drd4-rs	-1.17	shank3a	-3.53	drd1b	-1.5
oxt	-1.41	**immp2l**	**-4.02**	sapap2	-1.5
uts2a	-1.49	adra1aa	-4.06	drd2b	-1.87
immp2l	**-2.3**	uts2a	-4.23	oxt	-1.88
disc1	**-2.32**	hdac5	-4.74	**disc1**	**-2.04**

Table S2. Ordering of lines by PCA component projection values. Related to Fig 3; values correspond to the PCA results plotted in Fig 3.

Transparent methods

Animal use

Zebrafish (Danio rerio; AB strain) were housed in mixed gender 3L tanks in 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^{o}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 $420 \pm 50\mu$S. 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 Novartis Institute for Biomedical Research (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. (NCBI genome database QQSU00000000, BioSample #SAMN09717249). For designing CRISPR/Cas9 sgRNAs, we re-trained sgRNA-efficiency models using random forest and naive Bayes methods based on previously published sgRNAs. 1280 sgRNAs sequences and their editing efficiencies were obtained from a previous study in the TU-strain by Moreno-Mateos et al. (2015), and mapped to the AB genomic sequence. About 150 features were used to train the two machine learning models in classification mode, including genomic strand of sgRNA, GC%, identity of $\pm$4 bp of sgRNA targeting sequence, thermodynamics parameter, ΔG of sgRNA-genomic DNA heteroduplex for sliding windows of different sizes (Sugimoto et al., 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 (Nishimasu et al., 2014; Markham and Zuker, 2008), etc. Training accuracy of $>$0.7 was achieved by both models, and the efficacies of all sgRNAs in the AB genomic sequence were 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 this study.

Micro-injections, CRISPR/Cas9 mutant founder identification

sgRNAs were synthesized using T7 *in vitro* transcription with the MEGAshortscript^{TM} T7 Transcription Kit (ThermoFisher, AM1354). Cas9/sgRNAs were co-injected into one-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^{TM}-96 Transcription Clean-Up Kit (ThermoFisher, AM1909).

Indel rate in the injected zebrafish was measured using NGS at 2 days post fertilization

(dpf). PCR was performed on the genomic DNA extracted from zebrafish larvae using the HotSHOT method (Meeker et al., 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 $>$2000 avg. 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. The gene editing efficiency in the injected embryos was calculated as the maximum indel rate within the $\pm$30 bp regions of PAM sites. It is possible for the effects of CRISPR-Cas 9 to be obviated or changed from that predicted due to splicing around the targeted site. Hence, we performed RNAseq on brains from 22 homozygous lines (*adra1aa-/-*, *adra1ab-/-*, *chrm4a-/-*, *chrna2a-/-*, *disc1-/-*, *dlg4a-/-*, *drd1b-/-*, *drd2b-/-*, *drd3-/-*, *drd4a-/-*, *esr2a-/-*, *fmr1-/-*, *grm5a-/-*, *homer1b-/-*, *kctd13-/-*, *lrrn3-/-*, *oxt-/-*, *sapap2-/-*, *shank3b-/-*, *slc6a3-/-*, *srr-/-* and *tph2-/-*) and found, in all cases, that the mRNA was modified as predicted from the targeting, yielding a stop codon near the 5'-end without anomalous splicing. Also, in several of the lines (*chrna2a-/-*, *drd4a-/-*, *fmr1-/-*, *homer1b-/-*, *kctd13-/-*, *slc6a3-/-* and *tph2-/-*) there was dramatic reduction in mRNA level, presumably by nonsense-mediated decay, which further indicates that the mutations had the desired effects.

We sequenced five larvae per CRISPR-injected clutch to confirm the gene editing efficiency. The rest of the fish from 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 Data S1, 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, 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" ($\sim$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 2040-90μm 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. All trials were recorded after 10 minutes habituation to allow recovery from any stress due to netting. Shoaling assays were run with six fish (3 males and 3 females), which were randomly combined from multiple tanks of siblings. Each trial consisted of a 30 minute recording 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. Multiple trials were performed for each line (Table S1).

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) 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 were 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 were generated by corresponding setups.

Fish trajectories were obtained from videos using automated tracking inspired by idtracker (Pérez-Escudero et al., 2014). Tracking involved building fingerprint libraries to minimize switching of fish IDs after repeated crossings, using locomotion quantifications such as trajectory, angle, speed, pairwise distance, and convex hull area. Although this reduces the switching of fish IDs during tracking, it does not guarantee that individual identities are maintained over the course of a trial. Thus in our analysis we examine variability across trials instead of the variability among individuals within a trial.

Data processing

Before further analysis, we filtered tracked data to remove errors and smoothed the data to decrease noise in speed calculations. Using the (x, y) coordinates of the fish from the raw output files (see Data Availability), we calculated the velocity vector, speed, and scalar acceleration values. Initial data exploration found that fish identities were occasionally switched for brief periods in the tracking. This resulted in either a quick 'jump' in the trajectory or a multi-frame high-speed linear movement. To correct for this, errors were defined as sections of $\geq$3 consecutive frames with either unusually high acceleration ($>$671 cm/sec^2, $\sim$10 times the inter-quartile range of the acceleration distribution for WT), or very low acceleration values combined with sufficiently high speed ($<$0.0015 cm/sec^2 acceleration and $>$231 cm/sec speed, the median speed across all data). Error frames were marked with NaN values and excluded from calculations.

For WT, 4.5% of the trajectory points were marked as having tracking errors (overall average was 3.7% across all lines). We used a strict filter on the data and only use frames where no identified tracking errors were present for any of the fish in the group;

with this criteria, 21% of the frames for WT were marked as having tracking errors (overall average 16.5% across all lines). Loosening filter criteria did not substantially alter results.

To reduce noise in the velocities calculated from the tracking output, we smoothed the trajectory coordinates using a Gaussian filter with a standard deviation of 1 frame (cut off to include 5 frames total). NaN values were replaced by linear interpolation to replace segments of the trajectories identified as tracking errors. We calculated velocity and heading from these smoothed and interpolated trajectories. Although this introduces some errors (for example, when identity is switched), it greatly decreases the noise in velocity calculations. The smoothing procedure was only used to calculate speed and heading; unsmoothed trajectory coordinates were used for calculating distances.

A second type of tracking error was an 'overlap' error, when two nearby fish were identified as having the same position, along with nearly identical velocity and heading. This generally occurred when fish were close together but not necessarily moving in the same direction. Since this type of error could artificially increase estimates of group polarization, we removed frames where overlap errors occurred. Frames where a pair of fish were $<$ 1cm apart and had a heading difference of $<$ .2 radians were labeled as 'overlap frames', and were omitted from further calculations. For example, in WT, 5.8% of the trajectory points were identified by these criteria as having tracking overlap (overall average 5.5% across all lines). We again used a strict filter on the data and removed frames where overlap among any of the fish was detected; for WT, this led to 14.0% of the remaining data frames after the other filtering described above (overall average 13.5% across all lines).

Behavioral metrics

We analyzed seven behavioral features: median speed, polarization, group spacing, speed inter-quartile range (IQR), group centroid speed, nearest neighbor distance and fraction of time moving. Metrics were calculated as follows.

Group spacing is the per-trial median of the mean pairwise distances across the whole group:

$$\textit{median}\left(\langle r_{ij}(t)\rangle_{t,i\neq j}\right),$$

where r_{ij} is the pairwise distance between fish i and j at time t. We use the the notation $\langle Q(\{x\})\rangle_{\{x\}}$ to denote the mean of the quantity Q, taken over the set of variables $\{x\}$.

Polarization is calculated as in Tunstrøm et al. (2013):

$$P(t) = \frac{1}{N}\left|\sum_{i=1}^{N}\hat{v}_i(t)\right| = \sqrt{\langle\cos\theta_i(t)\rangle_i^2 + \langle\sin\theta_i(t)\rangle_i^2},$$

where $N = 6$ is the number of fish in the group, θ_i is the heading of fish i and $\hat{v}_i = (\cos\theta_i, \sin\theta_i)$ is the unit velocity vector of fish i.

Centroid speed reflects the movement of the group as a whole, and arises from a combination of individual speed and the coherence of the group motion. If individuals are moving together in the same direction, centroid speed is high; if they are not moving or are going opposite directions, centroid speed is close to zero. To obtain centroid speed, we calculated the group centroid as $\mathbf{c}(t) = \langle\mathbf{r}_i(t)\rangle_i$, where $\mathbf{r}_i(t)$ is the coordinate

vector of fish i at time t. We then calculated the centroid speed as distance traveled between frames divided by the time step, and took the median of this:

$$\textit{median}\left(\left|\mathbf{c}(t+dt)-\mathbf{c}(t)\right|/dt\right).$$

Nearest neighbor distance was calculated by selecting the closest neighbor for each fish at each time step, and taking the median of these values for a given trial.

To calculate *fraction of time moving*, note that we use 'not moving' to refer to times when fish move very slow for consecutive periods of time. We therefore define not moving, which we also refer to as 'freezing', by first taking a moving median of the speeds of an individual fish:

$$\eta_i(t,\tau) = \textit{median}(\{s_i(t-\tau/2)\ldots s_i(t+\tau/2)\}),$$

where $\eta_i(t,\tau)$ is the moving median speed of fish i at time t, calculated using a sliding window of $\tau = 10$ sec. We defined the fish to be not moving when the moving median speed was less than a threshold of U. The value used, $U = 0.87$ cm/sec, is the 0.05 quantile of the combined speed distribution of all the filtered data for all of the lines. The fraction of time moving is then simply the overall fraction of time where $\eta_i(t,\tau) > U$ for all fish at all times during a trial, i.e. $\langle \eta_i(t,\tau) > U \rangle_{i,t}$.

Speed while moving was then calculated as the median speed during frames where the fish was moving, i.e when $\eta_i(t,\tau) > U$. Similarly, *speed IQR* (inter-quartile range) was calculated as the difference between the 0.75 and 0.25 speed quantiles, using speed values only at times when $\eta_i(t,\tau) > U$.

Statistical tests on the behavioral metrics were used to evaluate differences in median values obtained for the set of trials for each line. A Kruskal-Wallis test yielded $p < 10^{-10}$ for all behavioral metrics. Comparisons between WT and mutant lines were done using Dunnett's test, which controls the family-wise error rate for many-to-one comparisons. Calculations and statistical analyses on behavioral metrics were conducted in Python and Matlab (Mathworks).

Principal component analysis

We performed principal component analysis (PCA) to find the primary axes of variation across lines. PCA provides a dimensionality reduction technique that preserves global structure, in contrast to methods such as tSNE that maximize clustering but distort global patterns (see, for example, (Kobak and Berens, 2019)).

To find dominant axes of variation, we first create a data matrix X_{iQ}, where the first (lowercase) index $i = 1..91$ is for the different lines plus WT, and the second (uppercase) index represents the 7 different quantities used as behavioral metrics. Following standard procedures, we normalized the data matrix so that the column mean is zero and the column standard deviation is 1, i.e. $\textit{mean}(X_{iQ})_i = 0$ and $\textit{std}(X_{iQ})_i = 1$, where the notation $\textit{fn}(X_{iQ})_i$ refers to the function *fn* evaluated over all indices of i for a given value of Q. This data matrix is shown in Fig 2A.

Because many of the behavioral metrics depend on speed, we used the data to create a 'speed model' for each quantity. For linear relationships, this can simply be calculated using the correlations of the (normalized) data matrix. However, because fraction of time moving is a bounded quantity, with many lines having values at or near 1, we instead

used a generalized procedure which can account for this. For all quantities Q except fraction of time moving, we fit a function

$$f_Q(S_i) = a_Q S_i, \tag{1}$$

where a_Q is a slope parameter for quantity Q and $S_i = X_{i1}$ is normalized speed for each line, which is the first column of the data matrix. For fraction of time moving, we fit the function

$$f_Q(S_i) = \max(X_{iQ})_i - e^{-a_Q(S_i - b_Q)}, \tag{2}$$

where a_Q and b_Q are parameters, and we note that the maximum is not 1 because of the data normalization. The parameter values for either function were determined using a least-squares fit with data input X_{iQ}, i.e. minimizing the cost function

$$E_Q = \left\langle (f_Q(S_i) - X_{iQ})^2 \right\rangle_i. \tag{3}$$

We again note that Eq. 3 has an analytical solution for the specific functional form of Eq. 1, but instead we use the least squares fit (via `curve_fit` in the Python package `scipy`) because it is more general. These functional fits for all quantities, along with values of the Spearman (rank) correlation coefficient of each quantity with speed are shown in Fig S2.

Using these functions, we then make a 'speed model prediction' for each quantity by calculating

$$L_{iQ} = f_Q(S_i),$$

and then create a new data matrix by subtracting this:

$$X^{(s)}_{iQ} = X_{iQ} - L_{iQ}. \tag{4}$$

We then use this to calculate the remaining variance in the data as

$$e = \frac{||X^{(s)}_{iQ}||^2}{||X_{iQ}||^2}. \tag{5}$$

Note that if the data for a line does not follow predictions of the speed model, its variance in the speed-adjusted data matrix $X^{(s)}_{iQ}$ could actually increase. Therefore the variance fraction e in Eq. 5 is best described as the variance remaining after accounting for speed correlations. We obtained $e = 0.60$, or, equivalently, that the overall variance is reduced by an amount $1 - e = 0.40$ using the regression fits of each quantity to speed. However, since the speed model of course perfectly removes variance in speed, an alternative description is that a fraction $7/6e = 0.70$ of the variance remains in quantities other than speed, or that the overall variance in quantities other than speed is reduced by 0.30 using the regression fits. We use this value of 0.30 in the main text.

The speed-adjusted data matrix $X^{(s)}_{iQ}$ is shown in Fig 2B, and contains the six quantities with remaining variance after omitting speed. To provide a quantitative description of how the different lines vary, we performed principal component analysis (PCA). This showed that three components explain 90.8% of the variance in $X^{(s)}_{iQ}$. The first component is driven by spacing, and the other two components are a mixture of quantities:

component 2 has positive changes in both polarization and fraction of time moving, and component 3 has opposing signs for the changes in polarization and fraction of time moving (Fig 3).

To account for sampling error due to the limited number of trials performed for each line, and how this affects the PCA embeddings, we used a bootstrap procedure to calculate confidence intervals on the projection of the each line onto each principal component (Efron, 1982). To perform this procedure, first note that the speed-adjusted data matrix $X_{iQ}^{(s)}$ contains the median value of a quantity Q calculated across trials for a line i. We thus use bootstrap sampling of the trials for each line, repeated $m = 1000$ times, to calculate a distribution of line-level median values and to form a 'sampled' data matrix X_{iQm}, where each index m refers to median values from a specific bootstrapped sample set. Denote the sampled median speed values as S_{im}. We use the functions and parameters from the speed model as described above to create a speed-adjusted data matrix for each set of sampled median values:

$$X_{iQm}^{(s)} = X_{iQm} - f_Q(S_{im}). \tag{6}$$

We then project the values of $X_{iQm}^{(s)}$ for each line and each sample set onto the principal components shown in Fig. 3A by calculating a dot product, and calculate the 0.05 and 0.95 quantiles of this distribution in order to determine the 90% confidence intervals shown in Fig. S3.

Components orthogonal to time moving

Because time moving could be considered a physiological difference (similar to speed), we asked how how much of the variance in $X_{iQ}^{(s)}$ is explained by time moving. The process for this calculation is similar to adjusting the base data matrix for speed correlations, except now we start with the speed-adjusted data matrix $X_{iQ}^{(s)}$ and calculate the correlations of each quantity with the column of this matrix that contains speed-adjusted fraction of time moving (M_i). Using a similar procedure as above, we fit simple slope functions,

$$f_Q(M_i) = m_Q M_i,$$

where m_Q is the slope parameter corresponding to quantity Q, to each column of $X_{iQ}^{(s)}$. To define the two components orthogonal to time moving, we first make a 'freezing-adjusted' data matrix

$$X_{iQ}^{(f)} = X_{iQ}^{(s)} - m_Q M_i.$$

We then perform PCA on the matrix $X_{iQ}^{(f)}$ to obtain the spacing and polarization components shown in Fig S4C-D.

Denote the spacing, polarization, and freezing components as $h_Q^{(1)}$, $h_Q^{(2)}$, and $h_Q^{(3)}$, respectively, and the projections onto each axis as $u_i^{(1)}$, $u_i^{(2)}$, and $u_i^{(3)}$. The components $h_Q^{(1)}$, $h_Q^{(2)}$ and the projections $u_i^{(1)}$, $u_i^{(2)}$ come from PCA on the freezing-adjusted data matrix $X_{iQ}^{(f)}$. The 'freezing' component is defined as the normalized vector of the values of m_Q, i.e. $h_Q^{(3)} = m_Q/||m_Q||$, and the projection onto this axis is simply a scaling of

the values of M_i: $u_i^{(3)} = M_i \cdot ||m_Q||$. The reconstructed data matrix is calculated as

$$X_{iQ}^{(s*)} = \sum_k u_i^{(k)} h_Q^{(k)},$$

from which we can calculate that the variance of the speed-adjusted data matrix is reduced by an amount $1 - ||X_{iQ}^{(s)} - X_{iQ}^{(s*)}||^2 / ||X_{iQ}^{(s)}||^2 = 0.906$ using this alternative dimensionality reduction procedure, which is nearly identical to the value of 0.908 obtained by simply performing PCA on $X_{iQ}^{(s)}$. The variance remaining after subtracting each component $k = 1..3$ is

$$e_k = \frac{||X_{iQ}^{(s)} - u_i^{(k)} h_Q^{(k)}||^2}{||X_{iQ}^{(s)}||^2}.$$

Fig S4C shows the values of $1 - e_k$ for the three components.

Model of individual turning decisions

We fit a model to examine how the turning decisions of individual fish depend on speed and the surrounding distribution of neighbors. Building on previous modeling work (Couzin et al., 2002; Calovi et al., 2018; Harpaz et al., 2017; Katz et al., 2011; Zienkiewicz et al., 2018), we use effective forces of attraction and alignment to describe the response of a focal fish to its neighbors. The model has a form of

$$z_i = \underbrace{\alpha F(s_i) \frac{1}{N} \sum_{j \neq i} G_{att}(r_j) H_{att}(\phi_j)}_{\text{attraction}} + \underbrace{\frac{1}{N} \sum_{j \neq i} G_{ali}(r_j) H_{ali}(\theta_j)}_{\text{alignment}} \tag{7}$$

$$P_i = 1/(1 + e^{-w z_i}), \tag{8}$$

where s_i is the speed of the focal individual, r_j, ϕ_j, and θ_j are the neighbor distance, angular position, and relative heading, respectively (Fig 4A), and $N = 6$ is the number of fish. Following Heras et al. (2018), we formulate the model as a 'turn-classifier' and calculate P_i, the probability of turning left after a specified time delay, using a logistic function with weight parameter w. Positive values of z_i predict left-turns, and negative values predict right turns, with the magnitude setting the probability of the prediction. The terms on the right-hand side of Eq. 7 correspond to effective attraction and alignment interactions, which are defined by an overall attraction-alignment ratio (α), a speed dependent attraction-alignment ratio ($F(s_i)$), and neighbor interactions that depend on distance ($G_{att}(r_j)$, $G_{ali}(r_j)$), relative angular position ($H_{att}(\phi_j)$), and relative heading ($H_{ali}(\theta_j)$). Each interaction function is defined by a discrete set of parameters, which correspond to discretizations of the input. The bins for discretization were defined by the relevant quantile distributions (i.e. either for speed or neighbor distance) of the combined dataset of all lines and all trials. The separable form used in Eq. 7 is similar to the model developed by Calovi et al. (2018), and therefore the functional forms of $G_{att}(r_j)$, $H_{att}(\phi_j)$, $G_{ali}(r_j)$, and $H_{ali}(\theta_j)$ can be compared with those in Calovi et al. (2018), where a model was fit to the burst-glide motion of juvenile zebrafish.

We fit the model as a binary classifier which predicts whether the focal fish moved left or right after a specified time delay of τ. Because of the speed differences, we use different values of τ for each line. For WT, we set $\tau = 1$ sec (this value was used in other modeling work that examined motion prediction of adult zebrafish (Heras et al., 2018)). For the other lines we use a scaled value $\tau S_{\text{WT}}/S_L$, where S_{WT} is the median speed for WT and S_L is the median speed for line L. With this scaling, predictions are made after approximately the same distance traveled for lines that moved at different speed. The angle of movement of a focal fish i after a time delay τ is calculated as

$$\gamma_i(t,\tau) = \arctan2\left(y_i(t+\tau) - y_i(t), x_i(t+\tau) - x_i(t)\right) - \theta_i(t). \tag{9}$$

Note that alternatively the model can be fit to predict the heading change $\Delta\theta_i(t,\tau) = \theta_i(t+\tau) - \theta_i(t)$, instead of the angle of movement $\gamma_i(t,\tau)$. The values of $\gamma_i(t,\tau)$ and $\Delta\theta_i(t,\tau)$ are equivalent in the limit $\tau \to 0$, but can differ as τ increases. We use the position-based movement change $\gamma_i(t,\tau)$, because it is a more robust measure of changes in movement. However, the results are qualitatively similar using either quantity.

The overall magnitude parameter w is related to the predictive ability of the model; higher values of w mean a higher predictive ability. For ease of interpretation we instead report the fraction of turns predicted correctly, not the values of w. The fraction of turns predicted correctly is the fraction of cases where $\text{sgn}(P_i(t) - 0.5) = \text{sgn}(\gamma_i(t,\tau))$, where the implied threshold of 0.5 reflects left-right symmetry.

The combination of Eqs. 7-8 is a standard formulation for binary classification, and we fit using a soft-margin loss function (Heras et al., 2018) (Goodfellow et al., 2017) with additional normalization constraints on the interaction functions implemented via a Lagrange-multiplier approach. The loss function for line L is

$$\begin{aligned} E = &\left\langle \ln\left(1 + \exp\left[-P_i(t)\,\text{sgn}(\gamma_i(t,\tau S_{\text{WT}}/S_L))\right]\right)\right\rangle_{i,t} \\ &+ \lambda\Big[\left(\langle|F(s_k)|\rangle_k - 1\right)^2 + \left(\langle|G_{att}(r_k)|\rangle_k - 1\right)^2 + \left(\langle|G_{ali}(r_k)|\rangle_k - 1\right)^2 \\ &\quad + \left(2\pi\left\langle H_{att}(\phi_k)^2\right\rangle_k - 1\right)^2 + \left(2\pi\left\langle H_{ali}(\theta_k)^2\right\rangle_k - 1\right)^2 \Big], \end{aligned} \tag{10}$$

where the Lagrange multiplier value of $\lambda = 0.01$ was used. 32 bins were used for the angular functions $H_{att}(\phi_k)$ and $H_{ali}(\theta_k)$ (i.e. $k = 1..32$), and 10 bins were used for the other functions $F(s_k)$ and $G_*(r_k)$ (i.e. $k = 1..10$). the normalization is a soft constraint to enforce $\langle|F(s_k)|\rangle_k = 1$, $\langle|G_{att}(r_k)|\rangle_k = 1$, $(\langle|G_{ali}(r_k)|\rangle_k = 1$, $\int_{-\pi}^{\pi} H_{att}(\phi)^2 d\phi = 1$, and $\int_{-\pi}^{\pi} H_{ali}(\theta)^2 d\theta = 1$. The model was fit using PyTorch optimization libraries and the Adam optimizer using the learning rate of 0.1. Convergence of the fit was defined to be when the loss function changed by less than a threshold of 10^{-5} for 50 steps.

Before fitting the model, we performed a mirror reflection of the data to enforce symmetry, and filtered the data so that the interactions functions were only fit to cases when the focal fish was sufficiently far from the boundary. To enforce left-right symmetry, we used the same procedure as Heras et al. (2018), and fit to a combined dataset featuring the original data and a left-right reflection in the frame of the focal fish. We use this procedure because we find no evidence for handedness or left/right turn preferences in zebrafish. Conceptually, the mirror reflection represents that a left turn towards neighbors on the left is equivalent to a right turn towards neighbors on the right. By enforcing

this symmetry, the input for each model fit contains exactly 50% left and 50% right turns. Therefore, random predictions would achieve an average of 50% correct predictions. To fit the model, we used data only when the focal fish was above a threshold distance from the boundary (22.9 cm, $\sim$10% of the radius of the arena) (Heras et al., 2018). To perform this filtering we first performed a procedure to estimate the location of the boundary from the trajectory data. This was done by first thresholding the trajectories to remove outlier coordinates, and then calculating a midpoint and a radius on thresholded data points. However, since this can lead to incorrect values when the group did not use the whole arena during a trial, we set the boundary equal to the median value across all trials for that line for trials where the midpoint or radius differed sufficiently (threshold set by visual inspection) from the median values. Note that boundary identification is only used to filter frames where the fish are sufficiently far away from the boundary for model fits, and therefore the results are not sensitive to the exact procedure. On the other hand, we note that due to this approximate procedure in identifying the boundary, fitting a model to the response of the fish to the boundary itself (as done in Calovi et al. (2018)) yielded inconsistent results.

We first fit the full form of the model to a combined dataset featuring subsampled data points from all of the highlighted lines. This fit is shown in Fig S5A. The fits of the angular functions $H_{att}(\phi_j)$ and $H_{ali}(\theta_j)$ represent that fish turn towards to and tend to align with their neighbors. Since fits to individual lines showed little variation in the H_* and $F(s_i)$ functions, we simplified the model comparison by holding these functions constant, set equal to the combined fit shown in Fig S5A and only fit α and $G_*(r_j)$ to compare the response of the different lines. The fits for WT, *scn1lab+/-*, *disc1-/-*, and *chrna2a-/-* are shown in Fig 4, and the fits for the other highlighted lines (*immp2l-/-*, *ctnnd2b-/-*) are shown in Fig S5B. For each line, we fit the interaction functions to either all of the data for that line (bold lines in Figs 4 and S5), or to individual trials (thin lines).

The predictive ability of the model varied for each line. Fig 4B shows the fraction of turn movements predicted correctly for fits to each line, using all data (the model was fit to all data, regardless of the magnitude of the movement change $\gamma_i(t,\tau)$). In Fig S5 we also show the fraction predicted correct for turns in the range where $20^o < |\gamma_i(t,\tau)| < 160^o$, where the change in heading is more predictable (Heras et al., 2018).

References

Calovi, D.S., Litchinko, A., Lecheval, V., Lopez, U., Escudero, A.P., Chaté, H., Sire, C., Theraulaz, G., 2018. Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors. PLOS Computational Biology 14, e1005933. doi:10.1371/journal.pcbi.1005933.

Couzin, I.D., Krause, J., James, R., Ruxton, G.M., Franks, N.R., 2002. Collective Memory and Spatial Sorting in Animal Groups. Journal of Theoretical Biology 218, 1–11. doi:10.1006/jtbi.2002.3065.

Efron, B., 1982. The jackknife, the bootstrap and other resampling plans. volume 38. SIAM, Philadelphia.

Goodfellow, I., Bengio, Y., Courville, A., 2017. Deep Learning. The MIT Press, Cambridge, Massachusetts.

Harpaz, R., Tkačik, G., Schneidman, E., 2017. Discrete modes of social information processing predict individual behavior of fish in a group. Proceedings of the National Academy of Sciences 114, 10149–10154. doi:10.1073/pnas.1703817114.

Heras, F.J.H., Romero-Ferrero, F., Hinz, R.C., Polavieja, G.G.d., 2018. Deep attention networks reveal the rules of collective motion in zebrafish. bioRxiv , 400747doi:10.1101/400747.

Katz, Y., Tunstrøm, K., Ioannou, C.C., Huepe, C., Couzin, I.D., 2011. Inferring the structure and dynamics of interactions in schooling fish. Proceedings of the National Academy of Sciences 108, 18720–18725.

Kent, W.J., 2002. BLAT–the BLAST-like alignment tool. Genome research 12, 656–64. doi:10.1101/gr.229202.

Kobak, D., Berens, P., 2019. The art of using t-SNE for single-cell transcriptomics. Nature Communications 10, 1–14. doi:10.1038/s41467-019-13056-x.

Markham, N.R., Zuker, M., 2008. UNAFold: Software for nucleic acid folding and hybridization. Methods in Molecular Biology 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, 610–614. doi:10.2144/000112619.

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. Nature methods 12, 982–8. doi:10.1038/nmeth.3543.

Nishimasu, H., Ran, F., Hsu, P., Konermann, S., Shehata, S., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O., 2014. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. Cell 156, 935–949. doi:10.1016/J.CELL.2014.02.001.

Pérez-Escudero, A., Vicente-Page, J., Hinz, R.C., Arganda, S., de Polavieja, G.G., 2014. idTracker: tracking individuals in a group by automatic identification of unmarked animals. Nature Methods 11, 743–748. doi:10.1038/nmeth.2994.

Sugimoto, N., Nakano, S.i., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., Sasaki, M., 1995. Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. Biochemistry 34, 11211–11216. doi:10.1021/bi00035a029.

Tunstrøm, K., Katz, Y., Ioannou, C.C., Huepe, C., Lutz, M.J., Couzin, I.D., 2013. Collective States, Multistability and Transitional Behavior in Schooling Fish. PLOS Comput Biol 9, e1002915. doi:10.1371/journal.pcbi.1002915.

Zienkiewicz, A.K., Ladu, F., Barton, D.A.W., Porfiri, M., Bernardo, M.D., 2018. Data-driven modelling of social forces and collective behaviour in zebrafish. Journal of Theoretical Biology 443, 39–51. doi:10.1016/j.jtbi.2018.01.011.