Orphan cytochrome P450 20a1 CRISPR/Cas9 mutants and neurobehavioral phenotypes in zebrafish

Nadja R. Brun[1‡], Matthew C. Salanga[2‡], Francisco X. Mora-Zamorano[1], David C. Lamb[3]

Jared V. Goldstone[1], and John J. Stegeman[1*]

[1] Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA

[2] Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, 86011, USA

[3] Faculty of Health and Life Sciences, Swansea University, Swansea, SA2 8PP, UK

‡These authors contributed equally to this work

*Corresponding author: jstegeman@whoi.edu

ABSTRACT

Orphan cytochrome P450 (CYP) enzymes are those for which biological substrates and function(s) are unknown. Cytochrome P450 20A1 (CYP20A1) is the last human orphan P450 enzyme, and orthologs occur as single genes in every vertebrate genome sequenced to date. The occurrence of high levels of *CYP20A1* transcripts in human substantia nigra and hippocampus and abundant maternal transcripts in zebrafish eggs strongly suggest roles both in the brain and during early embryonic development. Patients with chromosome 2 microdeletions including *CYP20A1* show hyperactivity and bouts of anxiety, among other conditions. Here, we created zebrafish *cyp20a1* mutants using CRISPR/Cas9, providing vertebrate models with which to study the role of CYP20A1 in behavior and other neurodevelopmental functions. The homozygous *cyp20a1* null mutants exhibited significant behavioral differences from wild-type zebrafish, both in larval and adult animals. Larval *cyp20a1*-/- mutants exhibited a strong increase in light-simulated movement (i.e., light-dark assay), which was interpreted as hyperactivity. Further, the larvae exhibited mild hypoactivity during the adaptation period of the optomotor assays. Adult *cyp20a1* null fish showed a pronounced delay in adapting to new environments, which is consistent with an anxiety paradigm. Taken together with our earlier morpholino *cyp20a1* knockdown results, the results described herein suggest that the orphan CYP20A1 has a neurophysiological role.

Keywords: vertebrate; cytochrome P450; anxiety; hyperactivity; mental health disorder; *Danio rerio*

INTRODUCTION

Cytochromes P450 (CYP; P450), a superfamily of enzymes found in every branch of life, catalyze a vast array of oxidation reactions, as well as the reduction and rearrangement of endogenous and exogenous compounds [1]. In vertebrates, including humans, CYP enzymes catalyze both physiological and toxicological reactions and play critical roles in many developmental stages.

When the physiological substrate(s) and function of a CYP are unknown, it is defined as an “orphan” P450. The functions of the majority of human and (by extrapolation) other mammalian P450s are known, although a few remain mysterious despite decades of intensive research [2-4]. Notable among these orphan CYPs is CYP20A1, the sole member of the CYP20 family, found in a single copy in all vertebrate genomes sequenced to date. CYP20A1 is the last human orphan P450 for which no biological or catalytic function is known.

While the activity of recombinant human CYP20A1 has been tested with possible substrates, no oxidation reaction was found to occur with steroids or selected biogenic amines [5]. Likewise, the activity of recombinant zebrafish Cyp20a1 has been tested with several different substrates without success [6]. Recently, human CYP20A1 expressed in yeast was observed to be weakly active with luminogenic substrates, as well as aniline [7, 8], suggesting that endogenous substrates may yet be identified.

Tissue and organ-specific expression patterns of genes such as *CYP20A1* can provide insights into function. In humans, the expression of *CYP20A1* transcripts varies in an organ-dependent manner. Expression is especially abundant in the hippocampus and

substantia nigra regions of the brain [5], regions that are prominently associated with learning and memory, and which are involved in neurodegenerative diseases including hyperactivity disorders (e.g., ADHD), panic disorders, social anxiety, and bipolar disorders. Such disorders affect >10% of the global population (~748 million people) [9, 10]. In other vertebrates, high levels of *CYP20A1* transcript occur in the brain and gonads of adult zebrafish [6] as well as in unfertilized eggs [11] and the notochord [12] of developing zebrafish, and during embryonic development of mice [13]. These findings suggest the participation of CYP20A1 in vertebrate development, as well as its potential involvement in endocrine and neuronal processes.

We have previously demonstrated that transient morpholino knockdown of *cyp20a1* in zebrafish resulted in behavioral abnormalities, including increased latency or reduced responsiveness to a visual stimulus in larvae at 6 days post-fertilization (dpf). Morphants also exhibited a higher level of total physical activity and more bursts of movement than the control larvae; zebrafish behaviors that are consistently interpreted as hyperactivity [6]. Now we have developed *cyp20a1* mutant zebrafish to further interrogate the relationship between Cyp20a1 and behavioral phenotypes. Using CRISPR/Cas9, we generated zebrafish with lesions in the *cyp20a1* coding locus, resulting in a *cyp20a1*(-/-) crispant line following additional standard breeding. The *cyp20a1* crispants were examined for behavioral phenotypes in both larval and adult zebrafish. Ultimately, this *cyp20a1*(-/-) zebrafish may enable further characterization of genetic involvement in behavioral disorders, functions for this protein, and the discovery of potential therapies at the molecular level.

RESULTS

CYP20A1 mutant lines

CYP20A1 was simultaneously targeted by two different sgRNAs in the 2nd and 3rd exons, resulting in multiple INDEL mutations (**Figure 1A**). Standard F_0 outcrossing and sibling incrossing resulted in stable *cyp20a1*-/- mutant lines in the AB background (**Figure 1B, C**). Two separate *cyp20a1*-/- mutant lines were isolated: line 60 (wh^{60}), with a 5 bp deletion and 4 bp insertion in exon 2, and line 61 (wh^{61}), with a 1 bp insertion in exon 2, and a 7 bp deletion in exon 3 (**Figure 1D**). In both cases, apparent nonsense mutations were created and computational translation of the mutant alleles showed the predicted amino acid sequence (**Figure 1E**). Due to the unavailability of specific antibodies, we were unable to confirm that the Cyp20a1 protein was completely absent from these lines, although without the heme-binding domain any P450 protein would be inactive. The behavior concordance (see below) suggests that both lines are missing active Cyp20a1 protein.

We observed mild morphological differences between the wh^{61} mutant line and wild-type (WT; of the AB strain) fish in our facility. Fewer mutant fish exhibited swim bladder inflation at 6 dpf than control fish (**Supplementary Figure S1**). Collating the three trials to assess swim bladder inflation showed that the unpaired mean difference of *cyp20a1*-/- wh^{61} (n = 18) minus WT (n = 18) was -30.6% (95 CI; -38.3, -22.8), $p < 0.001$. In adults, there also was a consistent color difference, with the wh^{61} line exhibiting an overall paler pigmentation. (The wh^{60} line was not available to be observed for swim bladder or color at the time this was noted.)

Larval behavior

We assessed the optomotor response (OMR) of larval zebrafish by analyzing the swimming responses (entrainment) to repeated sinewave gratings moving in one direction, and then reversing the direction. The OMR is essential for many animals to correct for deviation from an intended track direction requiring integration of both visual and movement functions. Changes in OMR are indicative of altered motor control, which can originate from altered muscular or retinal sensitivity or neuronal function of the underlying circuit [14, 15]. WT and *cyp20a1*-/- wh^{61} fish were subjected to two instances of OMR visual stimulation (first to the right, then to the left). The *cyp20a1*-/- larvae were less active compared to the WT strain during the 60 seconds prior to the beginning of the sinewave movement in both the right and left directions (**Figure 2A**). For example, in the 15 seconds before the sinewave moment to the right, the cyp20a1-/- mutant larvae moved on average 0.555 cm less (95CI; -0.841, -0.283), p = 0.003. Once the sinewave movement was started, however, both the *cyp20a1*-/- mutant and WT larvae responded equally to the sinewave movement in both directions. **Supplementary Figure S2** indicates the parameters calculated from the larval movement in the 5 minutes prior to the OMR assay. *cyp20a1*-/- mutant larvae showed decreases in average speed by -0.499 mm s^{-1} (95CI; -0.712, -0.287), p = 0.001 (**Figure S2A**), in distance traveled by -149 mm 5 min^{-1} (95CI; -213, -85.6), p = 0.001 (**Figure S2B**), and in overall activity prior to the sinewave movement by -14.1% (95CI; -21.3, -6.88), p = 0.001 (**Figure S2C**) compared to the WT strain. However, both WT and *cyp20a1*-/- exhibited an equal capacity to engage in high-speed swimming activity after the

sinewave movement was initiated (**Figure S2D**). Collectively, these observations reveal that *cyp20a1*-/- fish are far more reactive to the OMR visual stimulus, despite being less active in the absence of it in this assay. In our earlier study (Lemaire et al., 2016) we measured CYP20 mRNA expression in the eye and optic nerve of adult fish. Levels of expression in the eye were similar to those in the brain. Levels in the optic nerve were somewhat greater than those in the forebrain and midbrain. We did not measure expression in eye or optic nerve at different stages of development, which would be valuable to do. However, observing the shoaling and other behaviors of adult knockout fish, we do not anticipate major visual deficits resulting from cyp20a1 deletion but more subtle effects cannot be ruled out at this time.

Larval locomotion during daylight in some fish species is driven by a natural need for hunting and exploring. Upon sudden darkness, zebrafish larvae respond with hyperactivity, potentially in response to an overshadowing predator. We used a light-dark assay consisting of a 30-minute light acclimation period followed by repeated 10-minute dark and light exposures. Compared to the WT strain, the locomotor activity in *cyp20a1*-/- wh^{60} and wh^{61} mutant larvae was higher during the acclimation period, as well as during the dark stimulations (both wh^{60} and wh^{61} at $p < .0001$), whereas the wh^{61} mutants also exhibited hyperactivity in the light phases following the dark stimulations ($p < .0001$, **Figure 2B, Supplementary Figure S3**). The difference in response of wh^{60} and wh^{61} mutant larvae suggests that in one of the mutants some residual gene product is being produced, contributing to the difference in response during the light phase. Both mutants exhibit hyperactivity, suggesting that *cyp20a1*-/- behavioral differences may not

be attributed to muscle impairments but rather to neurological or other effects. This is further supported by the fact that the *cyp20a1*-/- fish remained less active than the WT fish during the OMR assays just prior to any visual stimulus (**Figure 2A**), but significantly increased their locomotor activity during the first 15 seconds of the OMR stimulus.

The startle response in fish is triggered by sensory stimuli (visual or vibro-acoustic) to rapidly escape from predators and changes in this response can be indicative of altered neuronal cell development or transmission. The startle latency exhibited by the *cyp20a1*-/- wh^{61} mutant larvae did not differ from that of the WT larvae at the highest two stimulus intensities (**Figure 2C, Supplementary Figures S4**) but showed on average a more rapid response at the two lowest stimulus intensities ($p < .017$, **Supplementary Figures S4**). Typically, with increasing stimulus intensity, more larvae will exhibit a startle response. However, in comparison to WT larvae, *cyp20a1*-/- wh^{61} mutant larvae were less responsive at the lowest and highest stimulus intensity ($p < .002$, $p < .008$, **Supplementary Figures S4C**). Repeated stimulation within a short period often leads to habituation, indicating that the nervous system is capable of filtering out irrelevant information. However, this can be impaired in several psychiatric and neurological diseases including schizophrenia and autism. Both the WT larvae and the *cyp20a1*-/- wh^{61} mutant larvae appeared to adapt to the highest auditory stimulus, suggesting habituation (**Figure 2D**).

Adult behavior

We also examined adult behaviors that are related to anxiety disorders, using the novel tank assay [16]. This behavioral assay involves an anxiety response to a novel environment, and by repeating the assay a measure of acclimation or, conversely, a buildup of stress can also occur. In all three trials, the *cyp20a1*-/- fish (wh^{61}) spent more time in the bottom third of the novel tank (**Figure 3A**), which is an indication of anxiety-like behavior. Both the *cyp20a1*-/- and the WT fish showed a tendency toward increased bottom-dwelling when the assay was repeated on days 7 and 14, suggesting a long-lasting stress effect from the handling in the previous week. Consistent with the increased time spent in the bottom third of the tank, the *cyp20a1*-/- fish also showed a delay in moving to the top half of the tank, for the first (**Supplementary Figure S5A**) and the second entry (**Supplementary Figure S5B**). There also was a decreased number of transitions to the top half (**Supplementary Figure S5C**). For statistical results of all three trials please see **Table S2**.

In terms of distance moved, both the *cyp20a1*-/- and the WT fish moved about the same (**Figure 3B**). The total duration of time spent freezing (displacement of ≤ 3 mm/s, **Supplementary Figure S5D**), the number of freezing episodes (at least 1 s of immobility, **Supplementary Figure S5E**), and the number of erratic swimming movements (darting, **Supplementary Figure S5F**) differed based on the *p*-value in one out of three trials between the *cyp20a1*-/- and the WT zebrafish. Estimation statistics only indicated a decrease in the number of erratic swimming movements in *cyp20a1*-/- fish. No difference between males and females was observed for the endpoints

measured except for the distance traveled in the first trial, in which both *cyp20a1*-/- and the WT females generally moved less than males.

The novel tank assay was performed using two tanks to record a *cyp20a1*-/- fish and a WT fish of the same sex at the same time. This setup allowed for direct visual comparison of the adult morphology, which in every case indicated a paler appearance of the *cyp20a1*-/- zebrafish in comparison to the WT.

DISCUSSION

Our specific focus on behavior was prompted by the possible neurological implications of *CYP20A1* RNA expression levels in the hippocampus and substantia nigra in the human brain [5], early larval zebrafish, [11], and in the developing mouse brain [13]. Moreover, our prior studies with transient morpholino knockdown of *cyp20a1* resulted in behavioral phenotypes involving visual responses and overall activity, akin to hyperactivity [6]. The results from our CRISPR/Cas9 *cyp20a1*-/- mutant experiments further support the idea that the function(s) of CYP20A1 are involved in neurological processes that when disrupted lead to behavioral changes.

In an earlier study [6], we gleaned information from case reports of interstitial micro-deletions in the human Chr2q33.1-2q33.2 region, including *CYP20A1* gene loss, which resulted in a suite of neurological defects among other adverse effects [6]. Patients with 2q33 microdeletion syndrome display developmental delays, psychomotor retardation, hyperactivity and bouts of anxiety, and in some cases delayed visuomotor coordination [17, 18]. However, hyperactivity, particularly in children, was observed primarily in patients in which the deletions in this region included the locus for *CYP20A1*. Recent

examination of additional case studies [19] has now strengthened this observation of possible involvement in human neurobehavioral disorders.

Zebrafish inherently exhibit many different types of behavior, some of which are analogous to mammalian behaviors. These include anxiety and hyperactivity [16, 20, 21]. These cross-species behavior analogies are cemented by the observations of identical outcomes resulting from pharmacological manipulations. For instance, ethanol reduces stress and anxiety behaviors, resulting in increased exploration and reduced erratic movements, whereas caffeine increases stress-associated behaviors, resulting in irregular movements [22]. Such observations often occur in parallel with shifts in cortisol levels, which are used as a physiological marker of anxiety and stress [23, 24]. In our study, the dark-induced hyperactivity in *cyp20a1*-/- mutant larvae, and the finding that *cyp20a1*-/- adults spent more time in the bottom third of a novel tank compared to wild-type fish, suggest that the absence of *Cyp20a1* gene product may dysregulate steroid hormones such as cortisol. The resemblance in anxiety and hyperactivity responses between humans and zebrafish with deletions in the *Cyp20a1* locus suggests that our zebrafish *cyp20a1*-/- crispants can serve as a disease model organism. The endogenous catalytic function of CYP20A1 remains unknown. Earlier, based on predicted protein structural features, we speculated that substrates of CYP20A1 may carry their own oxygen for catalysis and that these might include oxysterols or related compounds [6]. Human CYP20A1 expressed in yeast has been reported to weakly act on non-physiological luminogenic substrates and can be inhibited by azoles, suggesting that this enzyme may catalyze typical P450-type transformations, albeit at low reaction

rates [8]. However, this observation may aid in our search to determine whether candidate biological substrates are detectably metabolized by recombinant proteins.

Although the catalytic function of CYP20A1 remains elusive, its broad tissue distribution suggests that CYP20A1 likely possesses multiple catalytic activities or that the activity with some substrate may be relevant in multiple organs, including the brain. CYP20A1 is widely distributed in the animal kingdom, including in early-diverging groups such as sponges [25]. Although CYP20A1 appears to be ubiquitous among deuterostomes, its presence is sparse among arthropods [25], apparently having been lost in some groups. Nevertheless, the broad distribution suggests that this orphan P450 may serve functions that are critical in vertebrate biochemistry and that these may be conserved among animals, especially in the deuterostome lineage.

As with function, the regulation of CYP20A1 expression is not understood. Most human and macaque tissues exhibit some level of expression at the RNA level [5, 26]. We also found *cyp20a1* expression in most tissues of adult zebrafish [6], and widespread expression has been found in mice [13]. Unusual among non-mitochondrial P450s, the N-termini of the predicted CYP20A1 protein sequences are nearly identical across mammals [6], suggesting a conserved targeting or functioning of this protein region.

Although we believe that CYP20A1 has role(s) in neural tissues, the expression patterns clearly imply functions in other tissues. Tissue expression and promoter analysis also suggest reproductive, immune, hematopoietic, and neural involvement. Previously, we reported that *cyp20a1* transcript expression in zebrafish embryos is modestly affected by steroids and other nuclear receptor agonists, and was suppressed by the neurotoxicant methylmercury [6]. In any case, the behavioral alterations in

zebrafish in which *cyp20a1* has been knocked down [6] or knocked out (*cyp20a1*-/-; this study) imply that if there are multiple functions for this protein, these would include function(s) in the brain and steroid hormone synthesizing gonads. While beyond the scope of this study, future studies will address the levels of dopamine and 5HT, as well as cortisol in mutant larvae to further explore the underlying mechanisms and the potential function of CYP20A1.

The expression of *CYP20A1* transcript during development and in multiple adult organs in mammals and zebrafish implies endogenous regulation. In a human tissue screen, high levels of CYP20A1 expression were observed in endocrine tissues (as a group) and the pancreas [5], in addition to the hippocampus and substantia nigra. We previously found the highest expression level in adult zebrafish gonads [6]. Multiple other lines of experimental evidence point to endocrine participation involving steroids, which is consistent with the expression patterns in fish and humans. The hyperactivity in larvae and the anxiety-like behavior in adults may indicate a dysregulation of glucocorticoid biochemistry as previously described for these specific behaviors [27, 28]

In summary, we report on a *cyp20a1*(-/-) crispant zebrafish and the results obtained substantiate the specific involvement of Cyp20a1 in behavioral phenotypes in this vertebrate model. However, the broader significance of CYP20A1 to vertebrate physiology and disease processes remains unclear. The fact that the *cyp20a1*-/- null strain grows and reproduces with few defects suggests that *cyp20a1* is not an essential gene, barring some escape from the mutant condition or low-level redundancy as seen with some other genes, including in zebrafish (e.g., [29]). A comprehensive search for substrates is underway with recombinant zebrafish Cyp20a1 expressed in *E. coli,* and

metabolomics studies. The features of CYP20A1 structure, regulation, and biological correlations should aid in the deciphering of the molecular functions and roles of this orphan P450 in health and disease, as well as the evolution of these functions. The mutant strains we have developed are being explored to determine the functional and metabolic significance of CYP20A1. The CRISPR/Cas generated *cyp20a1*-/- zebrafish described herein will enable the functional characterization of this last human orphan P450, potentially advancing our understanding of the molecular mechanisms related to human mental health and the search for potential therapies.

MATERIALS AND METHODS

All Methods and Analyses reported here are reported in accordance with ARRIVE guidelines (https://arriveguidelines.org).

Animal husbandry. Experimental and husbandry procedures using zebrafish were approved by the Woods Hole Oceanographic Institution's Animal Care and Use Committee, and followed the NIH and American Veterinary Association (AMVA) guidelines and regulations. AB strain wild-type zebrafish were used in these studies. Embryos were obtained through pairwise or group breeding of adults using standard methods, rinsed with system water, and moved to clean polystyrene Petri dishes with 0.3X Danieau's solution (17.4 mM NaCl, 0.21 mM KCl, 0.12 mM $MgSO_4$, 0.18 mM $Ca(NO_3)_2$, and 1.5 mM HEPES at pH 7.6). Embryos were cultured at 28.5 °C and a 14 hr light – 10 hr dark diurnal cycle. The 0.3X Danieau's solution was replaced at 24 hours post-fertilization (hpf) and all dead or defective embryos were removed. Larvae were fed daily with a diet according to their age starting with rotifers (*Brachionus rotundiformis)* at

5 days post-fertilization (dpf), then rotifers coupled with brine shrimp (*Artemia franciscana*) at 9 dpf, adding pellet feed (Gemma Micro 300, Skretting) at 21 dpf. The fish were then exclusively fed with brine shrimp and pellets from 30 dpf onward. To anesthetize the adult fish to obtain fin biopsies, the fish were immersed in fresh buffered Tricaine (3-amino benzoic acid ethyl ester; Sigma A-5040) diluted in system water (0.016%$^{w/v}$) until motionless. Following fin biopsy, the adults were returned to their aquatic habitat and fed brine. The biopsied fish were allowed 7-10 days to recover before any additional handling.

sgRNA site selection and synthesis. The coding sequence of exons 2 and 3 (reference sequence ZDB-GENE-030903-3) were queried for putative targets using the "CHOPCHOP" web tool [30]. Based on this analysis, we selected two targets, opting for sequences that contained a G nucleotide within the first three nucleotides of the target sequence and no predicted off-target site.

Briefly, transcription was conducted using the MEGAscript (Ambion, AM1330) or MAXIscript (Ambion, AM1309) *in vitro* transcription reaction kits according to the manufacturer's instructions using 80-200 ng of purified PCR products (see *PCR - sgRNA template preparation*). The samples were then incubated at 37°C between 4 and 5 hours; 80 ng of template DNA was used for the MAXIscript reaction and 200 ng of template DNA was used for the MEGAscript reactions.

Microinjection equipment. Embryos were injected using a pneumatic microinjector (Model PV-820, World Precision Instruments). Injection needles were pulled from borosilicate capillary tubes (TW100F-4, WPI) using a vertical pipette puller (Model P-30, Sutter Instruments Inc.).

Microinjection solutions. 1-2 nl of injection solution was targeted to the yolk compartment of one-cell embryos immediately below the developing zygote. Injection solutions consisted of combinations of Cas9 recombinant protein (PNA Bio, CP-01) 1 µg µl^{-1}, Cas9 mRNA (from Addgene plasmid #51307 [31]) 200-400 ng µl^{-1}, H2B-RFP mRNA 200-400 ng µl^{-1}, and pooled sgRNA 50-200 ng µl^{-1} (**Supplementary Table 1**).

***mRNA synthesis*.** 1-5 µg of CS2-plasmid containing the ORF for Cas9 or H2B-RFP was linearized via Not1 endonuclease digestion followed by phenol:$CHCl_3$:IAA extraction and EtOH precipitation. Next, 1 µg linearized plasmid was used as a template in the SP6 mMessage mMachine *in vitro* transcription reaction (Ambion, AM1344) according to the manufacturer's instructions.

PCR. Endpoint PCR for genotyping or single guide RNA template preparation was carried out using Q5 (M0491 NEB) or Taq (M0267 NEB) polymerase and the corresponding reaction buffers. Genotype PCR assembly reactions included a template (20-200 ng gDNA or cDNA), dNTPs at a 200 µM final concentration, forward and reverse primers at a final concentration of 300 nM (for Taq reaction) or 500 nM (for Q5 reaction), a polymerase-specific reaction buffer at a 1x final concentration, and Q5 at 0.02 U µl^{-1} or Taq at 0.025 U µl^{-1}. These components were scaled to 25 µl reaction volumes. See **Supplementary Table 1** for primer sequences and cycling conditions. sgRNA templates were prepared as described in [32-34] [29]. Briefly, a universal reverse primer was combined with a forward primer containing a 5' T7 polymerase binding site, a gene-specific target sequence, and approximately 20 nucleotides of a 3' sequence complementary to the universal reverser primer in a 100 µl reaction at a 500 nM final concentration for each primer, dNTPs at a 200 µM final concentration, Q5

reaction buffer at a 1x final concentration, and 2U of Q5. PCR products were visualized via agarose gel electrophoresis and nucleic acid staining with SYBR safe DNA stain (S33102, Thermo Fisher Scientific), and imaged using an EZ Gel Documentation System (Bio-Rad, 1708270 and 1708273). The PCR products were purified using the PCR QIAquick PCR cleanup kit (Qiagen, 28106) according to the manufacturer's instructions.

RNA isolation. Total RNA was isolated from embryonic or larval tissue by mechanically homogenizing the tissue at room temperature in 200-500 µl TRIzol (Ambion, 15596-018) followed by RNA isolation according to the TRIzol product instructions or using a Direct-zol RNA MiniPrep Plus kit (ZYMO Research Corp, 2072). DNA contamination was removed from the TRIzol-isolated RNA via enzymatic digestion with 10 U of Turbo DNase (Ambion, AM2239) at 37°C for 15 minutes in a reaction tube for TRIzol-mediated extraction or on a ZYMO RNA MiniPrep spin column. DNase was removed from the RNA via organic extraction with phenol:CHCl3:IAA (isoamyl alcohol) (125:24:1) followed by CHCl3:IAA (24:1), then precipitated by adding 10% (v/v) 3M pH 5.2 sodium acetate solution and 2.5 volumes of 100% ice-cold ethanol and cooled to -20°C for ≥20 minutes, then centrifuged at 16,000-20,000 RCF for 20 minutes. The RNA pellet was washed twice with 70% (v/v) EtOH, air-dried, and dissolved in 20-50 µl DNase/RNase-Free water. The RNA isolated using the ZYMO Direct-zol RNA MiniPrep columns was eluted in 50 µl of DNase/RNase-Free water. The final concentrations were measured at a 260 nm/280 nm absorbance on a Nanodrop 2000 spectrophotometer.

cDNA synthesis for cloning. Up to 1 µg of DNA-free RNA was reverse transcribed using ProtoScript II Reverse Transcriptase (NEB, M0368) and anchored oligo dT primers according to the product instructions.

T7E1 mutant survey (F_0). T7 endonuclease 1 (#E3321, New England BioLabs) was used to survey for heteroduplexed PCR products as a result of mutagenized target loci. 200 ng of PCR product was denatured and reannealed by heating to 95°C for 5 minutes followed by gradual cooling to 85°C at a rate of 0.5°C/second and then to 25°C at a rate of 0.1°C/second. Annealed DNA was exposed to T7E1 for 15 minutes at 37°C followed immediately by cooling ice. Products were separated and visualized on 2% agarose gel alongside 200 ng of undigested product for comparison.

Outcross and T7E1 mutant survey (F_1). Sibling larvae (to the injected embryos positive in the T7E1 mutant survey) were raised to sexual maturity and five adult individuals were crossed with wild-type AB adults. Fifteen embryos from each cross were pooled and gDNA was isolated and cleaned as done previously, and dissolved in 100 µl nanopure water. PCR amplification of target loci was done as previously described and products were column purified and eluted in 20 µl Elution Buffer (Qiagen). T7E1 survey was performed as described above. Sibling embryos to T7E1 positive extracts were reared as putative *cyp20a1* heterozygotes, whereas those that were T7E1 negative were euthanized.

Morphological observations. Zebrafish larvae of the WT and *cyp20a1*-/- mutant line were kept until 6 dpf in 35 mm culture dishes (Falcon) containing 10 larvae in 10 mL per dish. At 6 dpf, the larvae were visually compared using a stereomicroscope, scored based on swim bladder inflation, and imaged. This experiment was independently

repeated three times with six dishes per line (WT, wh^{61}) and experiment (total n = 18). The morphological appearances of adult WT and mutant fish were also compared during the novel tank assay.

Behavioral assays. Optomotor response (OMR) assays were performed in "raceway"-shaped arenas created with 2% (w/v) agarose in deionized water with 60 mg L^{-1} Instant Ocean using a custom plastic mold. This mold was modified from a previously published design [35]. Each mold would cast a 7.5 cm × 11.6 cm gel containing 10 individual 7 cm × 0.8 cm raceways. Especially developed plastic molds measuring 11.7 cm × 7.6 cm × 5 mm were custom-built in-house. The molds were then used to create lanes using agarose poured into single-well plastic plates measuring 12.4 cm × 8.1 cm × 1.2 cm (Thermo Scientific). The molds contained five lanes in which the sides were angled at 60° to facilitate visualization. The lanes in the molds were 3.5 mm high with a base of 18 mm at the top, which tapered to 14 mm at the bottom of the lane. There was a 4 mm gap between the lanes in the mold. The agarose lanes were only used once per experiment and were discarded after each use. Videos of sinewave gratings for entrainment were provided by Dr. Elwood Linney. Prior to the video recordings, individual fish were transferred into each raceway and allowed to acclimate for 5 minutes in lighted conditions. Video recordings were acquired with two Logitech C920 USB webcams at a resolution of 960 × 720 pixels and a frame rate of 30 frames $second^{-1}$ (fps) as described in a previous study [36]. A total of 120 larvae per fish line were recorded before the videos were analyzed using custom R scripts.

Standard light-dark locomotor assays were performed using a DanioVision™ observation chamber (Noldus Inc.; Wageningen, Netherlands). At 6 dpf, zebrafish WT

413 and *cyp20a1*-/- mutant larvae (wh^{60} and wh^{61}) were randomly distributed in 48-well
414 plates and acclimated in the light for 30 minutes prior to the start of the light-dark
415 transitions. Three 10-minute dark periods were each followed by a 10-minute light
416 period. Each replicate experiment was run at approximately the same time of day (early
417 afternoon). The experiments were repeated at least three times with cohorts from
418 separate breeding events (total n = 120 for wh$^{60;}$ total n = 72 for wh^{61}), and the data from
419 the replicate experiments were pooled for final analysis. Videos were recorded at 30 fps
420 and analyzed with EthoVision XT® 12 (Noldus Inc.).

421 Vibroacoustic startle latency was assessed as described previously [37, 38] and the
422 same set-up was used to test for startle habituation as a form of non-associative
423 learning in 6 dpf larval zebrafish. For each trial, 16 larvae with inflated swim bladders
424 were distributed in a 4×4 acrylic well-plate which was mounted on a minishaker (Brüel &
425 Kjaer, Vibration Exciter 4810) connected to an amplifier (Brüel & Kjaer, Power Amplifier
426 Type 2718). For the startle response assay, vibro-acoustic stimuli were delivered at four
427 different amplitudes (32, 38, 41, 43 dB) and for each amplitude, the stimulus was
428 delivered four times spaced 20 seconds apart. For the habituation assay, vibro-acoustic
429 stimuli were delivered at 43 dB only. To establish a baseline response in the startle
430 habituation assay, the interval of the first three stimuli was set to 2 minutes. The interval
431 of the following 30 stimuli was set at 10 seconds to test for habituation. After an
432 additional 5 minutes of rest, responsiveness recovery was tested after a single stimulus.
433 The startle response was tracked at 1000 fps using a high-speed video camera
434 (Edgertronic, CA) and analyzed using FLOTE [39] and the analysis pipeline developed
435 by [40]. To assess habituation, the fraction of the 16 larvae per plate and stimulus that

responded with a short-latency C-bend (SLC; within 15 ms) was calculated. Both the startle response assay (total *n* = 144 larvae) and the habituation assay (total *n* = 11 plates) were repeated three times with cohorts from separate breeding events.

The novel tank assay assesses anxiety-like behaviors and was performed using adult (10-month-old) zebrafish. The experimental room was heated to 26 °C before the assay, which was performed between 11 am and 3 pm. Two narrow tanks (H: 15.1 cm; L: 21.5 cm; W: 5.1 cm) filled with system water were placed next to each other. For each round, two zebrafish (one fish of each line and of the same sex) were placed individually in a 50 mL glass beaker with 2 mL of system water for 30 seconds prior to releasing the fish simultaneously in the novel tank environment [41]. Videos were recorded with a Sony HD HDR-CX5 for 10 minutes. The temperature of the tank surfaces was regularly checked using an infrared thermometer and kept between 24.2 and 27.2 °C. DeepLabCut (version 2.2.b8) was used to track the zebrafish in the novel tank assay [42, 43]. To enhance the tracking performance, several body parts were labeled, including the snout, left eye, right eye, left gills, right gills dorsal fin, upper caudal fin, base caudal fin, and lower caudal fin. The residual neural network ResNet-50 was trained using 62 manually labeled frames from 5 randomly selected videos, after which 95% of the frames were used for 100,000 training iterations. We validated the training dataset and found the Root Mean Square Error for test was 20.9 pixels and for train: 2.8 pixels (the image resolution was 1920 by 1080 pixels). We then used a p-cutoff of 0.9 to condition the x,y coordinates for future analysis. Ultimately, the x,y values for the snout generated by DeepLabCut were processed using the NTD analysis script to evaluate 'Total distance moved', 'Time percent in bottom third', 'Latency for first entry to upper

459 half', 'Latency for second entry to upper half', 'Number of transitions to top half',
460 'Number of erratic swimming episodes', 'Number of freezing episodes', and 'Total freeze
461 time' [44].

462 **Statistical analysis.** Biological data, in particular behavioral data, exhibit inherently
463 wide sample-to-sample variability, and therefore many samples are required to achieve
464 sufficient statistical power for a reliable *p*-value interpretation [45]. As an alternative to
465 null hypothesis significance testing, which focuses on a dichotomous reject-nonreject
466 decision strategy based on *p* values, estimation statistics report on the estimation of
467 effect sizes (point estimates) and their confidence intervals (precision estimates). In this
468 study, we used estimation statistics and depicted effect size using Gardner-Altman plots
469 [46]. For those unfamiliar with interpreting effect sizes, *p* values from unpaired t-tests
470 (parametric) or Mann-Whitney tests (nonparametric) were also calculated and are
471 reported alongside confidence intervals in the following format: 'mean difference' (95%
472 confidence intervals; upper limit, lower limit), *p*-value. Normality was determined using
473 the D'Agostino & Pearson test. The statistical results of all assays are listed in **Table**
474 **S2**. All experimental animals were included in each analysis.

475 **REFERENCES**

476 1. Lamb, D.C. and M.R. Waterman, *Unusual properties of the cytochrome P450*
477 *superfamily.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1612): p. 20120434.
478 2. Guengerich, F.P., *Fifty years of progress in drug metabolism and toxicology:*
479 *What do we still need to know about cytochrome P450 enzymes?*, in *Fifty years*
480 *of cytochrome P450 research*, H. Yamazaki, Editor. 2014, Springer: Japan.
481 3. Cheng, Q. and F.P. Guengerich, *Identification of endogenous substrates of*
482 *orphan cytochrome P450 enzymes through the use of untargeted metabolomics*
483 *approaches.* Methods Mol Biol, 2013. **987**: p. 71-7.

4. Guengerich, F.P. and Q. Cheng, *Orphans in the human cytochrome P450 superfamily: approaches to discovering functions and relevance in pharmacology.* Pharmacol Rev, 2011. **63**(3): p. 684-99.
5. Stark, K., et al., *mRNA distribution and heterologous expression of orphan cytochrome P450 20A1.* Drug Metab Dispos, 2008. **36**(9): p. 1930-7.
6. Lemaire, B., et al., *Cytochrome P450 20A1 in zebrafish: Cloning, regulation and potential involvement in hyperactivity disorders.* Toxicol Appl Pharmacol, 2016. **296**: p. 73-84.
7. Durairaj, P., et al., *Functional expression and activity screening of all human cytochrome P450 enzymes in fission yeast.* FEBS Lett, 2019. **593**(12): p. 1372-1380.
8. Durairaj, P., et al., *Identification of new probe substrates for human CYP20A1.* Biol Chem, 2020. **401**(3): p. 361-365.
9. Wirdefeldt, K., et al., *Epidemiology and etiology of Parkinson's disease: a review of the evidence.* Eur J Epidemiol, 2011. **26 Suppl 1**: p. S1-58.
10. Zhou, Y., et al., *Abnormal connectivity in the posterior cingulate and hippocampus in early Alzheimer's disease and mild cognitive impairment.* Alzheimers Dement, 2008. **4**(4): p. 265-70.
11. Goldstone, J.V., et al., *Identification and developmental expression of the full complement of Cytochrome P450 genes in Zebrafish.* BMC Genomics, 2010. **11**: p. 643.
12. Thisse, B. and C. Thisse, *Fast Release Clones: A High Throughput Expression Analysis.*, in *ZFIN Direct Data Submission (http://zfin.org).*2004.
13. Choudhary, D., et al., *Comparative expression profiling of 40 mouse cytochrome P450 genes in embryonic and adult tissues.* Arch Biochem Biophys, 2003. **414**(1): p. 91-100.
14. Ahrens, M.B., et al., *Brain-wide neuronal dynamics during motor adaptation in zebrafish.* Nature, 2012. **485**(7399): p. 471-7.
15. Naumann, E.A., et al., *From Whole-Brain Data to Functional Circuit Models: The Zebrafish Optomotor Response.* Cell, 2016. **167**(4): p. 947-960 e20.
16. Stewart, A., et al., *Modeling anxiety using adult zebrafish: a conceptual review.* Neuropharmacology, 2012. **62**(1): p. 135-43.
17. Balasubramanian, M., et al., *Case series: 2q33.1 microdeletion syndrome--further delineation of the phenotype.* J Med Genet, 2011. **48**(5): p. 290-8.
18. Tomaszewska, A., et al., *Deletion of 14.7 Mb 2q32.3q33.3 with a marfanoid phenotype and hypothyroidism.* Am J Med Genet A, 2013. **161**(9): p. 2347-51.
19. Zarate, Y.A., et al., *Growth, development, and phenotypic spectrum of individuals with deletions of 2q33.1 involving SATB2.* Clin Genet, 2021. **99**(4): p. 547-557.
20. Champagne, D.L., et al., *Translating rodent behavioral repertoire to zebrafish (Danio rerio): relevance for stress research.* Behav Brain Res, 2010. **214**(2): p. 332-42.
21. Burgess, H.A. and M. Granato, *Modulation of locomotor activity in larval zebrafish during light adaptation.* J Exp Biol, 2007. **210**(Pt 14): p. 2526-39.
22. Egan, R.J., et al., *Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish.* Behav Brain Res, 2009. **205**(1): p. 38-44.

23. Cachat, J., et al., *Measuring behavioral and endocrine responses to novelty stress in adult zebrafish.* Nat Protoc, 2010. **5**(11): p. 1786-99.
24. Brun, N.R., et al., *Polystyrene nanoplastics disrupt glucose metabolism and cortisol levels with a possible link to behavioural changes in larval zebrafish.* Commun Biol, 2019. **2**: p. 382.
25. Nelson, D.R., J.V. Goldstone, and J.J. Stegeman, *The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1612): p. 20120474.
26. Uno, Y., S. Hosaka, and H. Yamazaki, *Identification and analysis of CYP7A1, CYP17A1, CYP20A1, CYP27A1 and CYP51A1 in cynomolgus macaques.* J Vet Med Sci, 2014. **76**(12): p. 1647-50.
27. Best, C., D.M. Kurrasch, and M.M. Vijayan, *Maternal cortisol stimulates neurogenesis and affects larval behaviour in zebrafish.* Sci Rep, 2017. **7**: p. 40905.
28. Sireeni, J., et al., *Profound effects of glucocorticoid resistance on anxiety-related behavior in zebrafish adults but not in larvae.* Gen Comp Endocrinol, 2020. **292**: p. 113461.
29. Salanga, M.C., et al., *CRISPR-Cas9-Mutated Pregnane X Receptor (pxr) Retains Pregnenolone-induced Expression of cyp3a65 in Zebrafish (Danio rerio) Larvae.* Toxicol Sci, 2020. **174**(1): p. 51-62.
30. Montague, T.G., et al., *CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W401-7.
31. Guo, X., et al., *Efficient RNA/Cas9-mediated genome editing in Xenopus tropicalis.* Development, 2014. **141**(3): p. 707-14.
32. Bassett, A.R. and J.L. Liu, *CRISPR/Cas9 and genome editing in Drosophila.* J Genet Genomics, 2014. **41**(1): p. 7-19.
33. Gagnon, J.A., et al., *Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs.* PLoS One, 2014. **9**(5): p. e98186.
34. Nakayama, T., et al., *Cas9-based genome editing in Xenopus tropicalis.* Methods Enzymol, 2014. **546**: p. 355-75.
35. Richendrfer, H. and R. Creton, *Automated high-throughput behavioral analyses in zebrafish larvae.* J Vis Exp, 2013(77): p. e50622.
36. Mora-Zamorano, F.X., et al., *Parental Whole Life Cycle Exposure to Dietary Methylmercury in Zebrafish (Danio rerio) Affects the Behavior of Offspring.* Environ Sci Technol, 2016. **50**(9): p. 4808-16.
37. Panlilio, J.M., et al., *Developmental exposure to domoic acid disrupts startle response behavior and circuitry in zebrafish.* Toxicol Sci, 2021.
38. Brun, N.R., et al., *Neurological mechanism of sensory deficits after developmental exposure to non-dioxin-like polychlorinated biphenyls (PCBs)* Nat Commun Biol, in review.
39. Burgess, H.A. and M. Granato, *Sensorimotor gating in larval zebrafish.* J Neurosci, 2007. **27**(18): p. 4984-94.
40. Panlilio, J.M., N. Aluru, and M.E. Hahn, *Developmental Neurotoxicity of the Harmful Algal Bloom Toxin Domoic Acid: Cellular and Molecular Mechanisms*

Underlying Altered Behavior in the Zebrafish Model. Environ Health Perspect, 2020. **128**(11): p. 117002.
41. Tran, S. and R.T. Gerlai, *The Novel Tank Test: Handling Stress and the Context Specific Psychopharmacology of Anxiety.* Current Psychopharmacology, 2016. **5**(2).
42. Mathis, A., et al., *DeepLabCut: markerless pose estimation of user-defined body parts with deep learning.* Nat Neurosci, 2018. **21**(9): p. 1281-1289.
43. Nath, T., et al., *Using DeepLabCut for 3D markerless pose estimation across species and behaviors.* Nat Protoc, 2019. **14**(7): p. 2152-2176.
44. Haghani, S., et al., *An Automated Assay System to Study Novel Tank Induced Anxiety.* Front Behav Neurosci, 2019. **13**: p. 180.
45. Halsey, L.G., et al., *The fickle P value generates irreproducible results.* Nat Methods, 2015. **12**(3): p. 179-85.
46. Ho, J., et al., *Moving beyond P values: data analysis with estimation graphics.* Nat Methods, 2019. **16**(7): p. 565-566.

Acknowledgments

These studies were supported in part by the Boston University Superfund Research Program NIH 5P42ES007381 (MCS, NRB, FXM, JVG, JJS), the Woods Hole Center for Oceans and Human Health (NIH: P01ES021923 and P01ES028938; NSF: OCE-1314642 and OCE-1840381; NRB and JJS), and EBI/EMBL Medakatox NIEHS R01ES029917 (JVG). DCL was funded by a UK-US Fulbright Scholarship. This paper is dedicated in memory of Dr. Michael Waterman.

Author contributions

N.R.B. performed, analyzed, and interpreted larval and adult behavioral experiments, statistical analysis, contributed to draft the manuscript; **M.C.S.** generated the zebrafish line, contributed to draft the manuscript; **F.X.M-Z.** performed, analyzed, and interpreted larval behavioral assays, contributed to draft the manuscript; **D.C.L.** performed laboratory experiments and contributed to draft the manuscript; **J.V.G.** helped with every step of the research, overcoming obstacles, and in writing the manuscript**; J.J.S.** helped with every step of the research, overcoming obstacles, and wrote most of the manuscript. All authors edited and accepted the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Figure legends

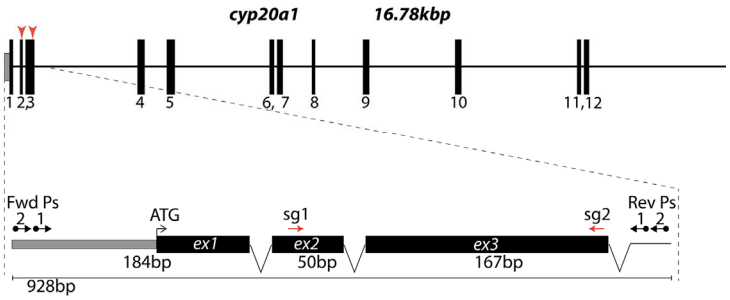
Figure 1. Zebrafish *cyp20a1* gene map and allele sequences. (A) Gene models **(B)** Gel image showing PCR products derived from T7E1 mutant survey (F_0). Lower gel shows the positive (heteroduplexed) T7D1 signature. **(C)** Chromatogram from F_1 heterozygous embryos (wh^{60} and wh^{61}) beginning near the sg1 site. (Note the appearance of double peaks). **(D)** cDNA sequences for exons 1-3 for WT, wh^{60}, wh^{61}. Note wh^{60} is a 5 bp deletion and 4 bp insertion in exon 2 and wh^{61} is a 1 bp insertion in exon 2, and a 7 bp deletion in exon 3. **(E)** Putative translation of cDNAs.

Figure 2. Larval behavior. (**A**) Optomotor response of wild-type (WT) and *cyp20a1* wh^{61} mutant larvae (*n* = 120). (**B**) The locomotor activity of *cyp20a1*-/- wh^{61} mutant larvae (*n* = 65) during the dark and the light phases in comparison to the WT larvae (*n* = 71). (**C**) Rapid startle response to the highest acoustic stimulus (43 dB) of WT (*n* = 138) and *cyp20a1* wh^{61} (*n* = 127) larvae. (**D**) Habituation to the highest acoustic stimulus measured as short-latency C-bend response (< 15 ms) per plate (*n* = 11) and depicted as mean ± 95 CI. All individual data points represent biologically independent replicates from three independent experiments.

Figure 3. Adult behavior in the novel tank assay. (**A**) *cyp20a1* wh^{61} mutant zebrafish spend more time in the bottom third of the novel tank in the first 10 minutes in comparison to wild-type (WT) zebrafish. (**B**) The distance moved in the novel tank does not differ between *cyp20a1* wh^{61} mutant zebrafish and WT zebrafish. The experiment was repeated on day 7 and day 14 with the same fish.

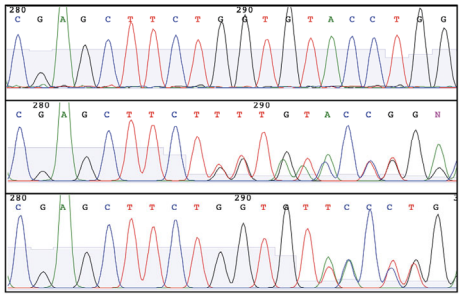
Figures

A



B

C



D

cDNA sequence

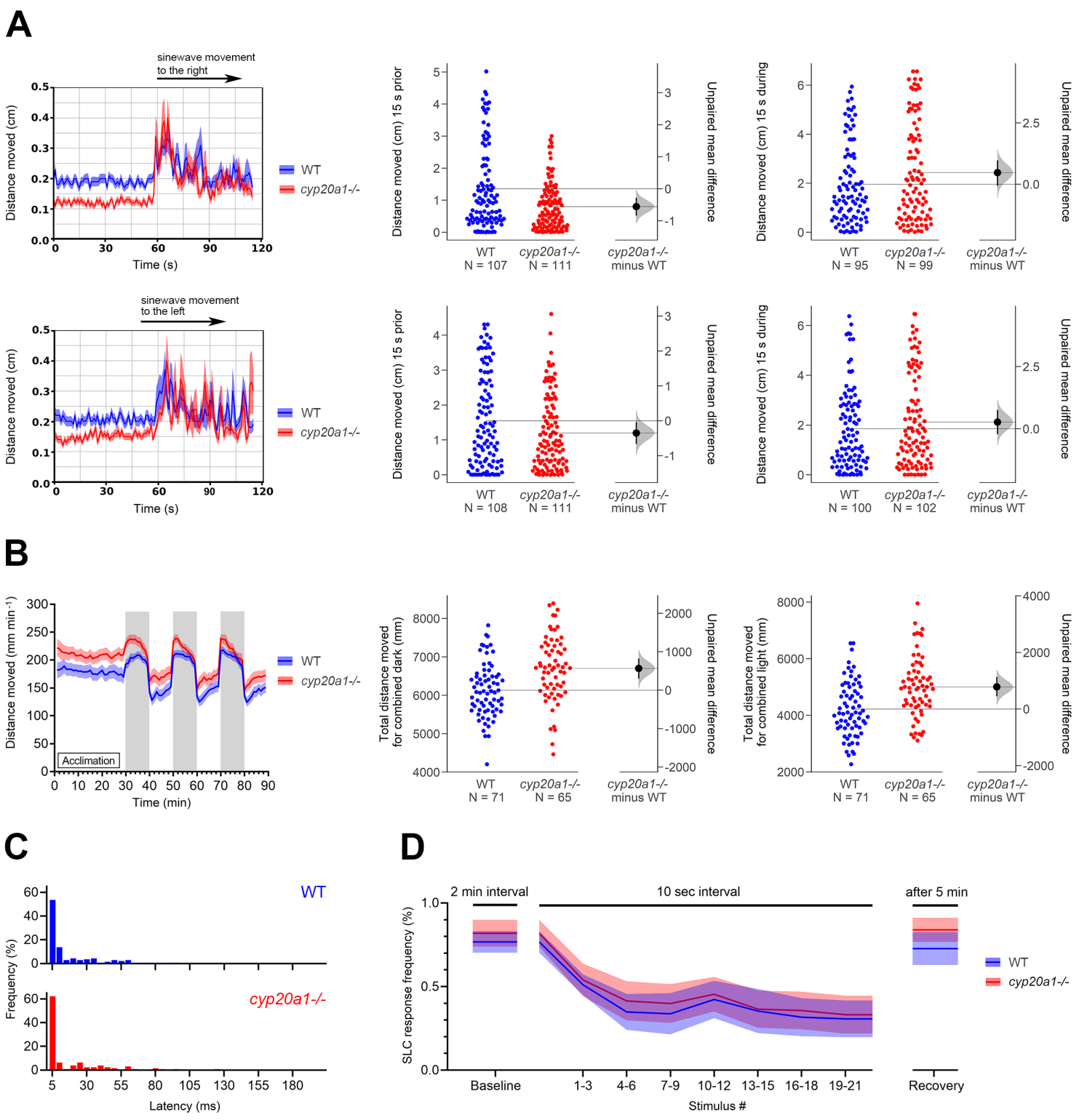
```
WT   ATGCTAGATTTTGCCATTTTTGCTGTGACATTTGTCATCATTCTGATTGGTGCCGTCCTG
60   ATGCTAGATTTTGCCATTTTTGCTGTGACATTTGTCATCATTCTGATTGGTGCCGTCCTG
61   ATGCTAGATTTTGCCATTTTTGCTGTGACATTTGTCATCATTCTGATTGGTGCCGTCCTG
                ex1|2                  sg1
WT   TATTTATATCCGTCATCTAGACGAGCTTCTGGTGT-ACCTGGACTAAACCCAACAGAAGA
60   TATTTATATCCGTCATCTAGACGAGCTTCTTTTA--ACCTGGACTAAACCCAACAGAAGA -5,+4
61   TATTTATATCCGTCATCTAGACGAGCTTCTGGTGTTACCTGGACTAAACCCAACAGAAGA +1
       ex2|3
WT   GAAAGATGGGAATCTTCAAGACATCGTGAACAAAGGAAGTCTCCATGAGTTTCTGGTGGG
60   GAAAGATGGGAATCTTCAAGACATCGTGAACAAAGGAAGTCTCCATGAGTTTCTGGTGGG
61   GAAAGATGGGAATCTTCAAGACATCGTGAACAAAGGAAGTCTCCATGAGTTTCTGGTGGG

WT   TCTTCATGATGAGTTTGGGTCTGTGGCATCTTTCTGGTTCGGGGCGAGACCAGTGGTGAG
60   TCTTCATGATGAGTTTGGGTCTGTGGCATCTTTCTGGTTCGGGGCGAGACCAGTGGTGAG
61   TCTTCATGATGAGTTTGGGTCTGTGGCATCTTTCTGGTTCGGGGCGAGACCAGTGGTGAG
                          sg2                  ex3|4
WT   CCTGGGTGCTGTGAACCAACTACGACAACACATCAACCCTAACTGGACCACGGATTCATT
60   CCTGGGTGCTGTGAACCAACTACGACAACACATCAACCCTAACTGGACCACGGATTCATT
61   CCTGGGTGCTGTGAAC-------GACAACACATCAACCCTAACTGGACCACGGATTCATT -7
```

E

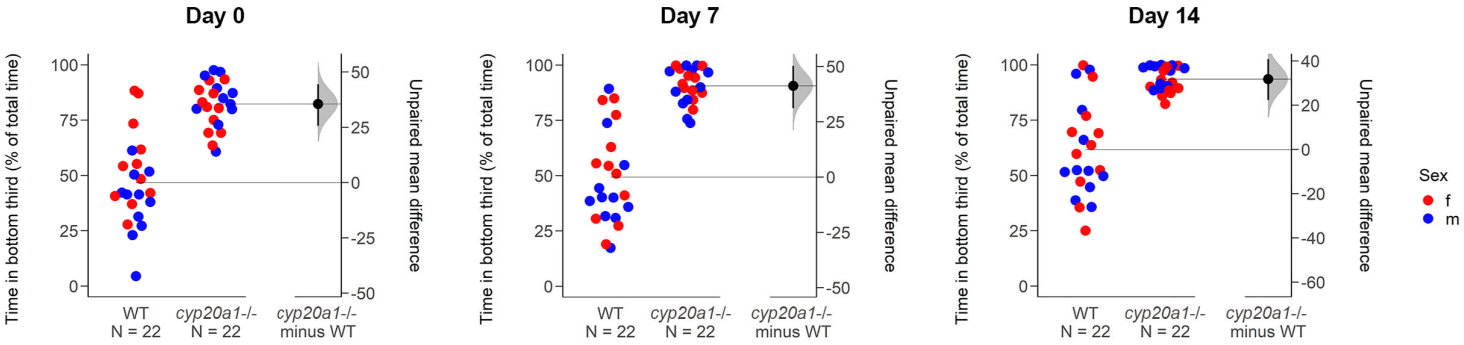
Predicted amino acid sequence

```
Wt   MLDFAIFAVTFVIILIGAVLYLYPSSRRASGVPGLNPTEEKDGNLQDIVNKGSLHEFLVG…
61   MLDFAIFAVTFVIILIGAVLYLYPSSRRASGVTWTKPNRRERWESSRHREQRKSP*
60   MLDFAIFAVTFVIILIGAVLYLYPSSRRASFNLD*
```

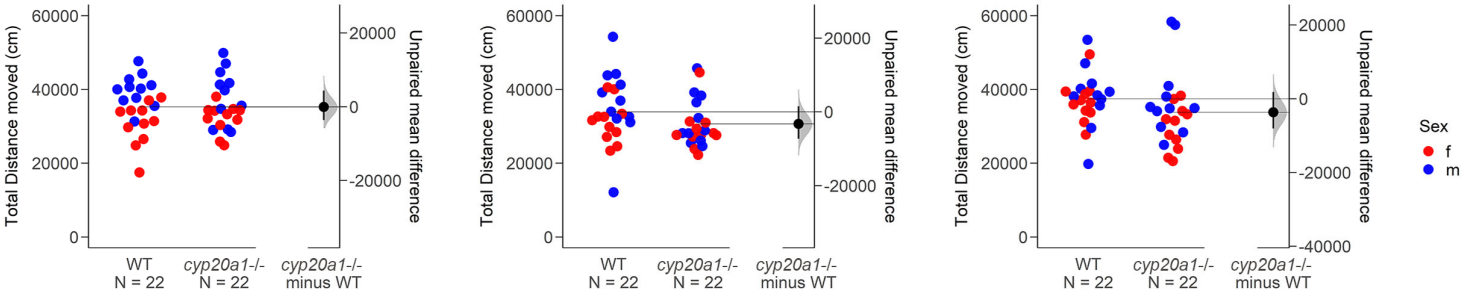
Formatted: Font: (Default) Arial


A
sinewave movement to the right
Distance moved (cm)
Time (s)
WT
cyp20a1-/-
Distance moved (cm) 15 s prior
Distance moved (cm) 15 s during
Unpaired mean difference
WT N = 107
cyp20a1-/- N = 111
cyp20a1-/- minus WT
WT N = 95
cyp20a1-/- N = 99
sinewave movement to the left
WT N = 108
cyp20a1-/- N = 111
WT N = 100
cyp20a1-/- N = 102
B
Distance moved (mm min-1)
Acclimation
Time (min)
Total distance moved for combined dark (mm)
Total distance moved for combined light (mm)
WT N = 71
cyp20a1-/- N = 65
C
WT
cyp20a1-/-
Frequency (%)
Latency (ms)
D
2 min interval
10 sec interval
after 5 min
SLC response frequency (%)
Baseline
1-3
4-6
7-9
10-12
13-15
16-18
19-21
Recovery
Stimulus #

A


Day 0
Day 7
Day 14
Time in bottom third (% of total time)
Unpaired mean difference
WT
N = 22
cyp20a1-/-
N = 22
cyp20a1-/-
minus WT
Sex
f
m

B


Total Distance moved (cm)
Unpaired mean difference
WT
N = 22
cyp20a1-/-
N = 22
cyp20a1-/-
minus WT
Sex
f
m

Supplementary Information

Orphan cytochrome P450 20A1 CRISPR/Cas9 mutants and neurobehavioral phenotypes in zebrafish

Nadja R. Brun, Matthew C. Salanga, Francisco X. Mora-Zamorano, David C. Lamb

Jared V. Goldstone, and John J. Stegeman

Table S1. Primer sequences for endpoint PCR and CRISPR-Cas sgRNA target and synthesis sequences.

Sequence Name	Barcode ID	Primer sequence (5')	Cycling condition	Polymerase	Product length (bp)
cyp20-wtP2-F	1025475438	TGTCATCATTCTGATTGGTGCC	98˚ - 3 min - (1x); 98˚ - 10 sec, 65˚ - 10 sec, 72˚ - 20 sec (35x); 72˚ - 5 min (1x)	NEB Q5	437
cyp20-wtP2-R	1025479906	GTTGATGTGTTGTCGTAGTTGG			
cyp20-ms61P1-F	1025475440	TAGACGAGCTTCTGGTGTTACC	98˚ - 3 min - (1x); 98˚ - 10 sec, 65˚ - 10 sec, 72˚ - 20 sec (35x); 72˚ - 5 min (1x)	NEB Q5	277
cyp20-ms61P1-R	1025475441	CAGTTAGGGTTGATGTGTTGTCG			
cyp20_60_wtF	1033095856	GACGAGCTTCTGGTGTACCTGG	98˚ - 3 min - (1x); 98˚ - 10 sec, 68˚ - 10 sec, 72˚ - 20 sec (35x); 72˚ - 5 min (1x)	NEB Q5	304
cyp20_60_wtR	1033095857	CACTGCGGCTACTCACTGGT			
cyp20_60_mutF	1033095860	TTACGCCCCTGTCTTGCAGT	98˚ - 3 min - (1x); 98˚ - 10 sec, 65˚ - 10 sec, 72˚ - 30 sec (35x); 72˚ - 5 min (1x)	NEB Q5	744
cyp20_60_mutR	1033095861	CTGTTGGGTTTAGTCCAGGTTAAAAG			
dr.cyp20a1-F	1023898079	ACCATGCTAGATTTTGCCATTTTTGCTGTG			
dr.cyp20a1-R	1023898080	TCAGTTTCTCTTGCTGACCGTG			
z.cyp20a1-5'utr-F	1023898081	GTAGTCGAGTACCGATCTAGAGG			
z.cyp20a1-3'utr-R	1023898082	GTGTAATTCCCATCCTCCAGAGG			
dr.cyp20a1.T7.sg1	1023342952	GATTAATACGACTCACTATAGGACGAGCTTCTGGTGTACCGTTTTAGAGCTAGAAATAGC			
dr.cyp20a1.T7.sg2	1023342953	GATTAATACGACTCACTATAGGTTGATGTGTTGTCGTAGTGTTTTAGAGCTAGAAATAGC			
sgRNA Universal reverse primer		AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC			

dr.cyp20a1.seqpF1	1023342954	ATCGCCAGCTCGTAGTTCAC
dr.cyp20a1.seqpR1	1023342955	CAGTCTTCAACTGTAAATGCAGC
dr.cyp20a1.seqpF2	1023342956	TCCTGATGGTCATTGTAGACG
dr.cyp20a1.seqpR2	1023342957	CAGGCGGACTGATAATTCAGG
CYP20A1_Dr_pENTRF	1018501550	CACCATGCTAGATTTTGCCA
CYP20A1_Dr_pENTRR	1018501551	TCTCTTGCTGACCGTGATCCA
zf_cyp20_f4	1017216446	TACAGGAGGTGGAAGGAAAGGTG
zf_cyp20_r4	1017216447	GACGACACCAAGGGCATAGATAAC

Table S2. Statistical results.

Assay	**Endpoint**	***cyp20a1-/-* Line**	**# of Trials**	**Total *n* (WT)**	**Total *n* (Cyp20A1)**	**Estimation Stats**	**Passed Normality Test**	**Unpaired t-test[1]**	**Mann-Whitney Test[2]**
Morphology	Swim Bladder Inflation	wh^{61}	3	18	18	Unpaired mean difference of *cyp20a1-/-* (n = 18) minus WT (*n* = 18) -30.6 [95CI -38.3; -22.8]	Some		U = 6, $p < .001$***
OMR	Right Grating (prior)	wh^{61}	3	107	111	Unpaired mean difference of *cyp20a1-/-* (*n* = 111) minus WT (*n* = 107) -0.555 [95CI -0.841; -0.283]	No		U = 4577, $p = .003$**
OMR	Right Grating (during)	wh^{61}	3	95	99	Unpaired mean difference of *cyp20a1-/-* (*n* = 99) minus WT (*n* = 95) 0.472 [95CI -0.0656; 0.973]	No		U = 4189, $p = .190$
OMR	Left Grating (prior)	wh^{61}	3	108	111	Unpaired mean difference of *cyp20a1-/-* (*n* = 111) minus WT (*n* = 108) -0.349 [95CI -0.668; -0.0377]	No		U = 5266, $p = .121$
OMR	Left Grating (during)	wh^{61}	3	100	102	Unpaired mean difference of *cyp20a1-/-* (*n* = 102) minus WT (*n* = 100) 0.267 [95CI -0.223; 0.753]	No		U = 4760, $p = .414$

OMR	Average Speed	wh[61]	3	115	115	Unpaired mean difference of *cyp20a1-/-* (*n* = 115) minus WT (*n* = 115) -0.499 [95CI -0.712; -0.287]	Yes	t(228) = 4.711, *p* < .001***	
OMR	Maximum Speed	wh[61]	3	115	115	Unpaired mean difference of *cyp20a1-/-* (*n* = 115) minus WT (*n* = 115) 1.11 [95CI -0.293; 3.13]	Some		U =5778, *p* = .098
OMR	Activity	wh[61]	3	115	115	Unpaired mean difference of *cyp20a1-/-* (*n* = 115) minus WT (*n* = 115) -14.1 [95CI -21.3; -6.88]	No		U = 4492, *p* < .001***
OMR	Total Distance Traveled	wh[61]	3	115	115	Unpaired mean difference of *cyp20a1-/-* (*n* = 115) minus WT (*n* = 115) -149 [95CI -213; -85.6]	Yes	t(228) = 4.706, *p* < .001***	
Ligth-Dark	Total Activity Dark	wh[61]	3	71	65	Unpaired mean difference of *cyp20a1-/-* (*n* = 65) minus WT (*n* = 71) 566 [95CI 299; 824]	Yes	t(134) = 4.251, *p* < .0001****	
Ligth-Dark	Total Activity Light	wh[61]	3	71	65	Unpaired mean difference of *cyp20a1-/-* (*n* = 65) minus WT (*n* = 71) 782 [95CI 448; 1130]	Yes	t(134) = 4.447, *p* < .0001****	
Ligth-Dark	Total Activity Dark	wh[61]	5	120	120	Unpaired mean difference of *cyp20a1-/-* (*n* = 120) minus WT (*n* = 120) 659 [95CI 433; 883]	Some		U = 4200, *p* < .0001****
Ligth-Dark	Total Activity Light	wh[61]	5	120	120	Unpaired mean difference of *cyp20a1-/-* (*n* = 120) minus WT (*n* = 120) 58.9 [95CI -213; 312]	Yes	t(238) = 0.44399, *p* = .6604	
Startle Response	Startle Latency 32dB	wh[61]	3	111	83	Unpaired mean difference of *cyp20a1-/-* (*n* = 83) minus WT (*n* = 111) -3.34 [95CI -14.1; 8.58]	No		U = 3685, *p* = .017*
Startle Response	Startle Latency 38dB	wh[61]	3	133	123	Unpaired mean difference of *cyp20a1-/-* (*n* = 123) minus WT (*n* = 133) 2.02 [95CI -5.13; 8.93]	No		U = 7395, *p* = .185
Startle Response	Startle Latency 41dB	wh[61]	3	135	124	Unpaired mean difference of *cyp20a1-/-* (*n* = 124) minus WT (*n* = 135) 3.02 [95CI -1.96; 8.68]	No		U = 7805, *p* = .348
Startle Response	Startle Latency 43dB	wh[61]	3	138	127	Unpaired mean difference of *cyp20a1-/-* (*n* = 127) minus WT (*n* = 138) 0.0176 [95CI -4.88; 5.28]	No		U = 7977, *p* = .207
Startle Response	Short Latency C-Bend Bias 32dB	wh[61]	3	111	83	Unpaired mean difference of *cyp20a1-/-* (*n* = 83) minus WT (*n* = 111) 0.344 [95CI 0.0934; 0.582]			
Startle Response	Short Latency C-Bend Bias 38dB	wh[61]	3	133	123	Unpaired mean difference of *cyp20a1-/-* (*n* = 123) minus WT (*n* = 133) 0.137 [95CI -0.0474; 0.34]			
Startle Response	Short Latency C-Bend Bias 41dB	wh[61]	3	135	124	Unpaired mean difference of *cyp20a1-/-* (*n* = 124) minus WT (*n* = 135) 0.0345 [95CI -0.131; 0.212]			
Startle Response	Short Latency C-Bend Bias 43dB	wh[61]	3	138	127	Unpaired mean difference of *cyp20a1-/-* (*n* = 127) minus WT (*n* = 138) 0.0327 [95CI -0.141; 0.201]			

Startle Response	Fraction Responding 32dB	wh^{61}	3	140	129	Unpaired mean difference of *cyp20a1-/-* (*n* = 129) minus WT (*n* = 140) -0.148 [95CI -0.24; -0.055]	No		U = 7092, *p* = .002**
Startle Response	Fraction Responding 38dB	wh^{61}	3	140	131	Unpaired mean difference of *cyp20a1-/-* (*n* = 131) minus WT (*n* = 140) -0.0154 [95CI -0.0901; 0.0587]	No		U = 8913, *p* = .637
Startle Response	Fraction Responding 41dB	wh^{61}	3	140	129	Unpaired mean difference of *cyp20a1-/-* (*n* = 129) minus WT (*n* = 140) -0.0444 [95CI -0.101; 0.00953]	No		U = 8122, *p* = .036*
Startle Response	Fraction Responding 43dB	wh^{61}	3	140	130	Unpaired mean difference of *cyp20a1-/-* (*n* = 130) minus WT (*n* = 140) -0.0536 [95CI -0.103; -0.00462]	No		U = 8002, *p* = .008**
Novel Tank Assay (Trial1)	Time In Bottom Third	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 35.6 [95CI 25.7; 44.4]	Yes	t(42) = 7.423, *p* < .001***	
Novel Tank Assay (Trial2)	Time In Bottom Third	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 41.3 [95CI 31.2; 50.3]	Yes	t(42) = 8.430, *p* < .001***	
Novel Tank Assay (Trial3)	Time In Bottom Third	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 31.8 [95CI 22.3; 40.8]	Yes	t(42) = 6.646, *p* < .001***	
Novel Tank Assay (Trial1)	Latency For First Entry	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 118 [95CI 73.1; 183]	Some		U = 54, *p* < .001***
Novel Tank Assay (Trial2)	Latency For First Entry	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 151 [95CI 80; 225]	Some		U = 103.5, *p* < .001***
Novel Tank Assay (Trial3)	Latency For First Entry	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 69.9 [95CI -8.18; 142]	Some		U = 176, *p* = .123
Novel Tank Assay (Trial1)	Latency For Second Entry	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 128 [95CI 75.1; 193]	Some		U = 72.5, *p* < .001***
Novel Tank Assay (Trial2)	Latency For Second Entry	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 76.6 [95CI -8.86; 161]	Some		U = 178, *p* = .135
Novel Tank Assay (Trial3)	Latency For Second Entry	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 107 [95CI 9.49; 192]	Some		U = 159, *p* = .051
Novel Tank Assay (Trial1)	Total Distance Traveled	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -52.6 [95CI -3600; 4390]	Yes	t(42) = 0.02575, *p* = .980	
Novel Tank Assay (Trial2)	Total Distance Traveled	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -3260 [95CI -7310; 1510]	Some		U = 158, *p* = .049*
Novel Tank Assay (Trial3)	Total Distance Traveled	wh^{61}	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -3650 [95CI -8090; 1830]	Some		U = 147, *p* = .025*

Novel Tank Assay (Trial1)	No Transitions Top Half	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -17.2 [95CI -30; -5.28]	Yes	t(42) = 2.595, *p* = .013*	
Novel Tank Assay (Trial2)	No Transitions Top Half	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -26.1 [95CI -38.6; -12.9]	Some		U = 81, *p* < .001***
Novel Tank Assay (Trial3)	No Transitions Top Half	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -26.9 [95CI -37.8; -17]	Some		U = 74.50, *p* < .001***
Novel Tank Assay (Trial1)	Total Freeze Time	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 1.27 [95CI -5.32; 13.9]	No		U = 218.5, *p* = .588
Novel Tank Assay (Trial2)	Total Freeze Time	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -0.182 [95CI -9.91; 24.7]	No		U = 123, *p* = .004**
Novel Tank Assay (Trial3)	Total Freeze Time	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -1.32 [95CI -17.5; 16.2]	No		U = 197, *p* = .297
Novel Tank Assay (Trial1)	No Freezing Episodes	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) 1.36 [95CI -4.64; 9.41]	No		U = 241.5, *p* = .995
Novel Tank Assay (Trial2)	No Freezing Episodes	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -4.23 [95CI -12.1; 6.06]	No		U = 155.5, *p* = .042*
Novel Tank Assay (Trial3)	No Freezing Episodes	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -4.45 [95CI -13.8; 3.23]	Some		U = 217, *p* = .564
Novel Tank Assay (Trial1)	No Darting Episodes	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -7.68 [95CI -14.8; -1.93]	Yes	t(42) = 2.294, *p* = .027*	
Novel Tank Assay (Trial2)	No Darting Episodes	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -6.68 [95CI -13.1; 0]	Yes	t(42) = 1.930, *p* = .060	
Novel Tank Assay (Trial3)	No Darting Episodes	wh[61]	1	22	22	Unpaired mean difference of *cyp20a1-/-* (*n* = 22) minus WT (*n* = 22) -5.82 [95CI -12.8; 0.853]	Yes	t(42) = 1.643, *p* = .108	

[1] t(degrees of freedom) = the t statistic, *p* = p-value

[2] U = the U statistic, *p* = p value

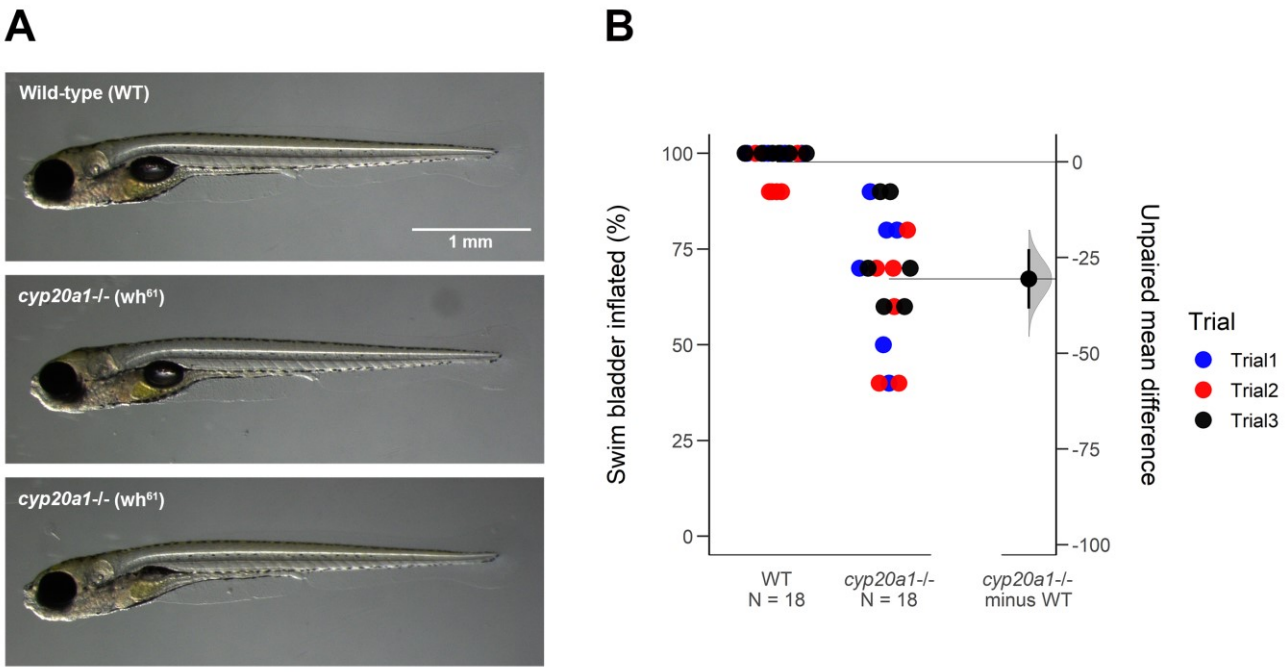


Figure S1. Morphology of *cyp20a1-/-* larvae at 6 dpf. (**A**) *cyp20a1-/-* (wh^{61}) mutant larvae show no apparent morphological differences in comparison to the wild-type (WT) line except for (**B**) swim bladder inflation, which was reduced in the mutant larvae. The experiment was repeated three times independently with 6 dishes containing 10 larvae per experiment (total *n* of 3 experiments = 18).

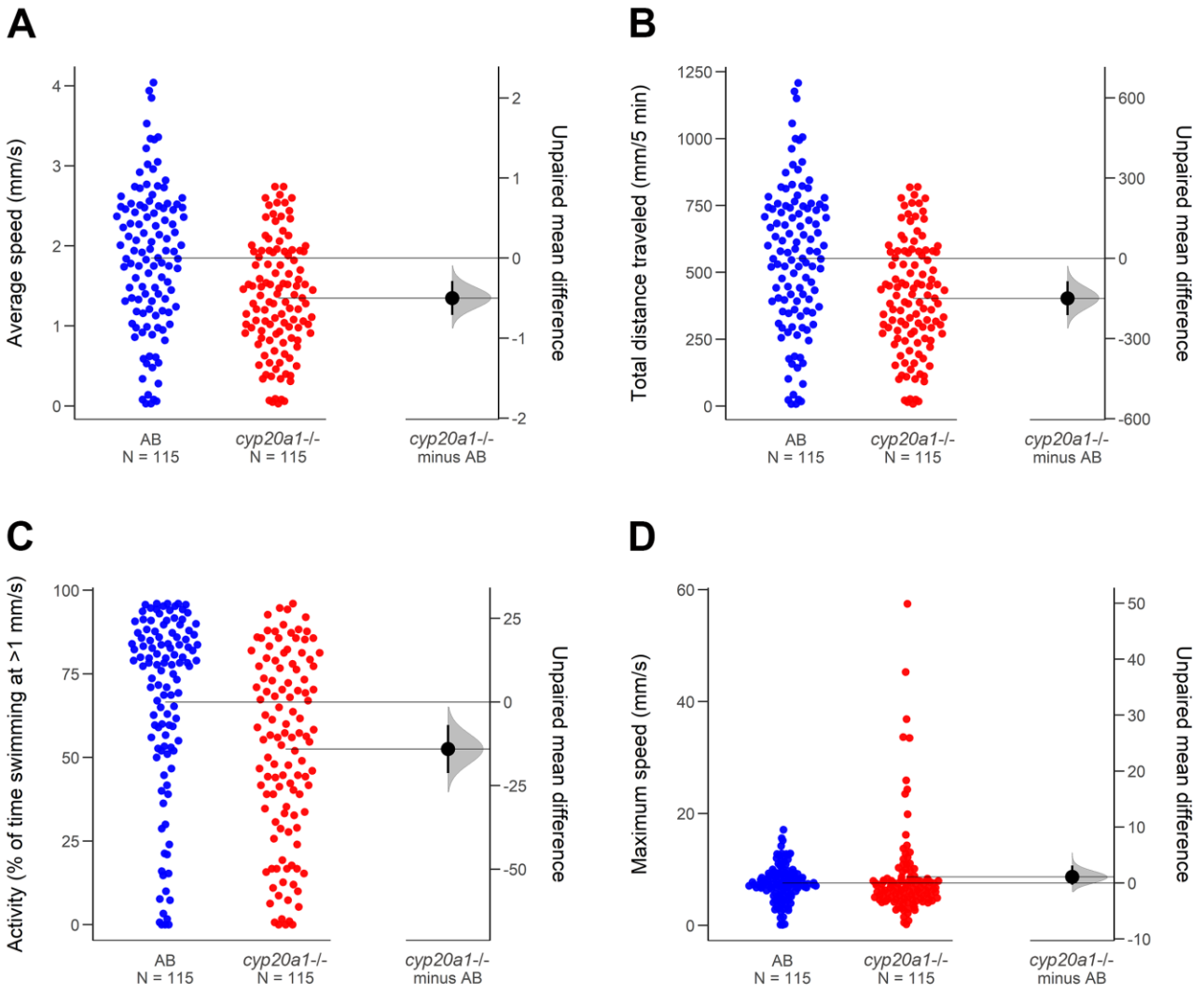


Figure S2. Activity of larvae prior to optomotor response (OMR) assay. (**A**) Average speed, (**B**) total distance traveled, (**C**) activity, and (**D**) maximum speed was measured in the 5 minutes prior to the OMR assay. A total of 120 larvae per fish line (WT and *cyp20a1-/-* wh^{61}) were recorded.

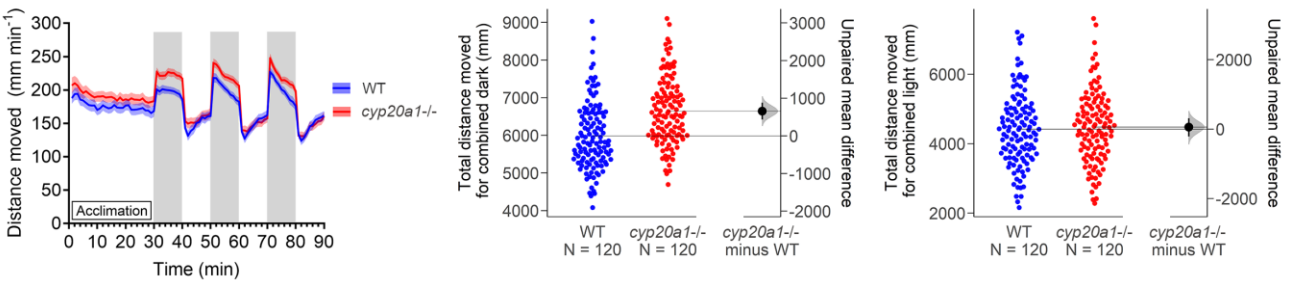


Figure S3. Locomotion during the light-dark assay of *cyp20a1-/-* (wh^{60}) mutant larvae in comparison to WT. *cyp20a1-/-* (wh^{60}) mutant larvae show hyperactivity in the dark phase but not in the light phase. The experiment was repeated five times independently (total *n* = 120).

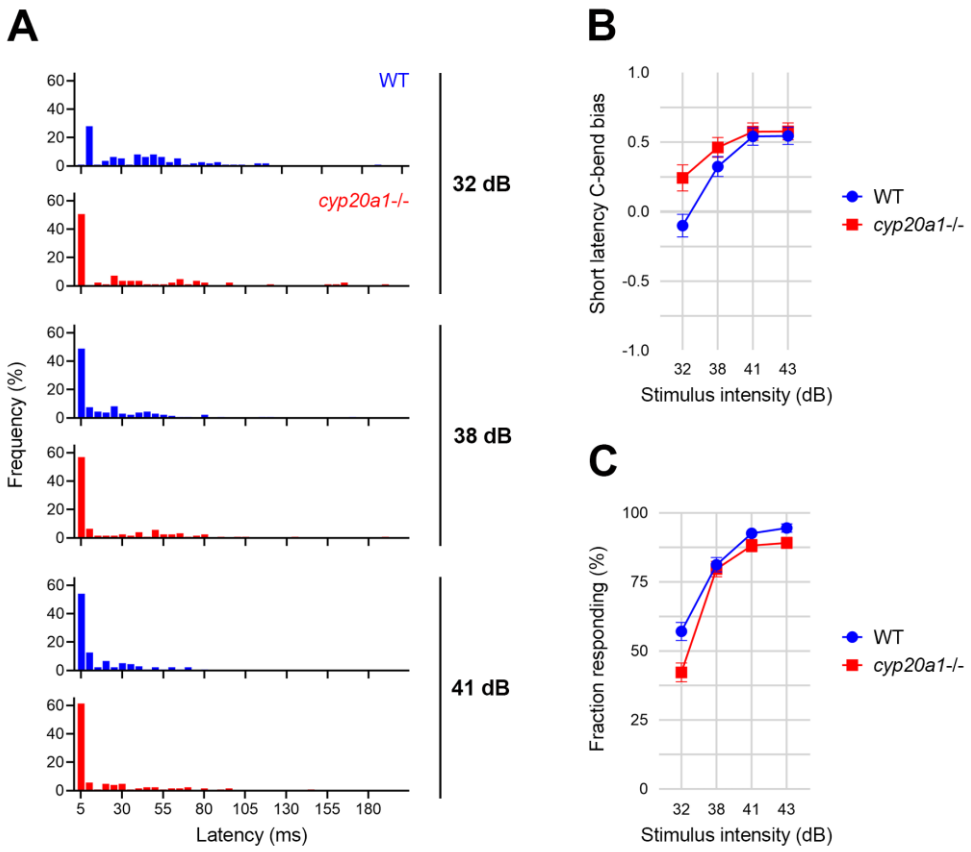
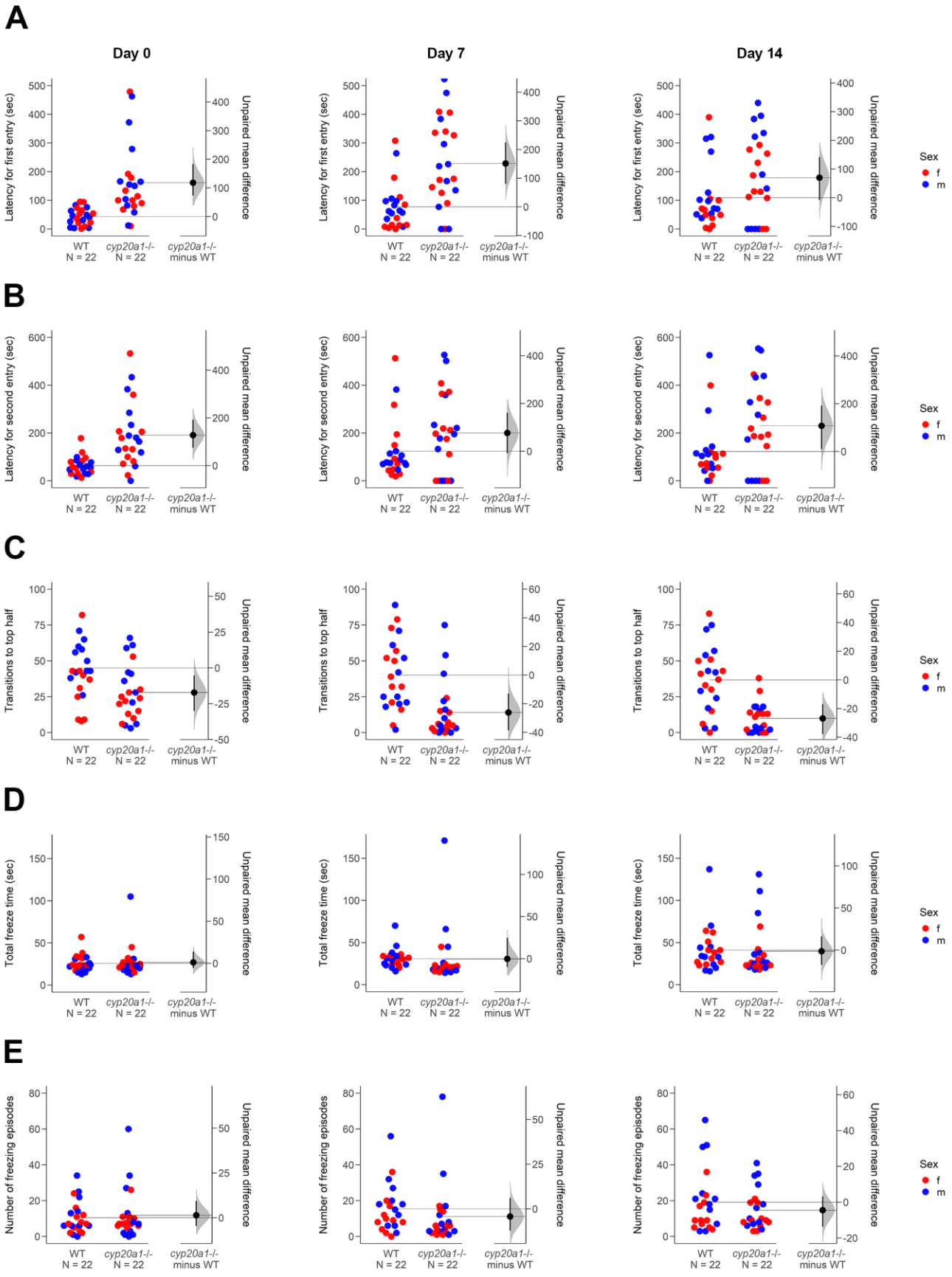


Figure S4. Startle response. (**A**) Latency of WT and *cyp20a1-/-* (wh^{61}) mutant larvae at 32 dB, 38 dB, and 41 dB. For 43 dB see Figure 2. (**B**) Bias toward short-latency C-bend (< 15 ms) at all stimulus intensities. (**C**) Fraction of larvae responding at different stimulus intensities. All data from three independent experiments.


A
Day 0
Day 7
Day 14
Latency for first entry (sec)
Unpaired mean difference
B
Latency for second entry (sec)
C
Transitions to top half
D
Total freeze time (sec)
E
Number of freezing episodes
WT
N = 22
cyp20a1-/-
N = 22
cyp20a1-/-
minus WT
Sex
f
m

F

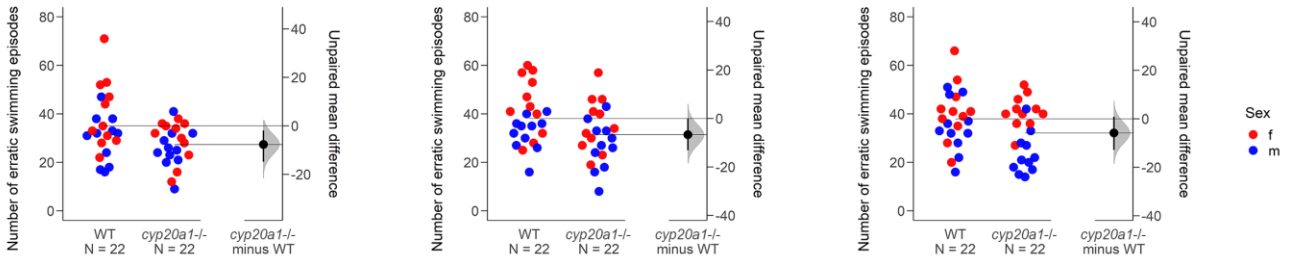


Figure S5. Adult WT and *cyp20a1-/-* (wh^{61}) behavior in the novel tank assay. (**A**) Latency for first entry to the top half, (**B**) latency for second entry to the top half, (**C**) number of transitions to the top half, (**D**) total freeze time, (**E**) number of freezing episodes, (**F**) number of erratic swimming episodes (also called darting). The experiment was repeated on day 7 and day 14 with the same fish.