Title: Histone 4 acetylation regulates behavioral individuality in zebrafish

Angel-Carlos Román[1,2,*,#], Julián Vicente-Page[1,2,*], Alfonso Pérez-Escudero[2,3], Jose M. Carvajal-González[4] , Pedro M. Fernández-Salguero[4], Gonzalo G. de Polavieja[1,2,#]

[1]Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Avenida Brasília s/n, 1400-038 Lisbon, Portugal.

[2]Instituto Cajal, Consejo Superior de Investigaciones Científicas, Av. Doctor Arce, 37, 28002 Madrid, Spain.

[3]Physics Department, MIT, Cambridge, Massachusetts, USA

[4]Departamento de Bioquímica y Biología Molecular y Genética, Universidad de Extremadura. Av. de Elvas s/n, 06071 Badajoz, Spain.

[*]Contributed equally

[#]Correspondence to: angel.roman@neuro.fchampalimaud.org and gonzalo.polavieja@neuro.fchampalimaud.org

Keywords: Epigenetics, HDAC, YY1, behavioral variability, zebrafish

ABSTRACT

Behavioral individuality arises even in isogenic populations under identical environments, but its underlying mechanisms remain elusive. We found that inbred and isogenic zebrafish (*Danio rerio*) larvae showed consistent behavioral individuality when swimming freely in identical wells or in reaction to stimuli. We also found that behavioral individuality depends on the histone acetylation levels. Individuals with high levels of histone 4 acetylation behaved similar to the average of the population, but those with low levels deviated and showed behavioral individuality. More precisely, we found behavioral individuality to be related to individuality in histone 4 acetylation of a set of genomic regions related to neurodevelopment. We found evidence that this modulation depends on a complex of Yin Yang 1 (YY1) and histone deacetylase 1 (HDAC1). We suggest, using stochastic modelling, that this complex is part of the molecular machinery giving individuality in histone acetylation in neurodevelopmental genes ultimately responsible for behavioral individuality.

INTRODUCTION

Classically, the phenotypic diversity of a population is considered to be generated by the genetic differences between its members and the disparity of their environmental influences[1]. A simple prediction from this view alone would then be that isogenic populations would not show variability when the environment is constant. Nevertheless, a pioneering study showed that there was variability independent of genetic differences in some morphological traits in mice raised in identical environments[2]. In recent years, similar results have been obtained for behavioral variability in mice and flies[3,4]. There are several mechanisms that might contribute to this effect, including developmental noise[5], maternal and paternal effects[6], or the different experiences the individuals obtain by interacting with the environment or other animals[4], among others.

Our knowledge about behavioral variability independent of genetic differences has increased substantially, but its underlying mechanisms remain unclear. Neuronal mechanisms such as neurogenesis, or serotonin signaling have been shown to be final targets of behavioral individuality[3,4], but the molecular mechanisms remain elusive. Chromatin modifications could be a promising mechanism to encode stable differences among individuals and they have been hypothesized as a potential mechanism for the generation of experience-dependent behavioral individuality[4]. DNA methylation differences have been associated to behavioral castes in honeybees[7], and they are necessary and sufficient to mediate social defeat stress[8]. Histone acetylation is another of the main epigenetic modifications[9] and it has been shown to regulate different behaviors such as mating preference in prairie voles[10] or cast-mediated division of labor in ants[11]. We thus reasoned that molecular mechanisms linked to epigenetic modifications could lead to behavioral individuality.

We used zebrafish from 5 to 8 days post fertilization (dpf) to dissect the molecular substrates of behavioral individuality. Laboratory zebrafish larvae show individuality in behavior[12] and they present some advantages such as its wide genomic information, the simplicity of its pharmacological treatments and the possibility to do large-scale behavioral analysis. Additionally, it is relevant to use a species in which we can observe directly developmental changes, as some of the mechanisms responsible for behavioral individuality are likely accumulated during development[13]. Here we established zebrafish larvae as a model for the analysis of behavioral individuality to study individuality in free-swimming behavior. We found that in our experimental tests behavioral individuality of zebrafish larvae is independent of the genetic variability of the population but it is linked to histone acetylation differences.

RESULTS

Behavioral individuality in larval zebrafish is stable for days

We used three steps to establish zebrafish larvae as a model to study behavioral individuality using a high-throughput setup (see **Methods and Extended Data Figure 1a-k** for the custom-built video tracking software, downloadable from **www.multiwelltracker.es**. We first obtained that each larvae showed differences in their spontaneous behavior. Simple eye inspection of trajectories reveals this behavioral individuality (**Figure 1a-b,** left, 7-8 dpf, respectively; **Extended Data Figure 2a-b,** left, 5-6 dpf). This can be quantified using two parameters: overall activity (percentage of time in movement) and radial index (average relative distance from the border towards the center of the well). These two parameters were chosen because they are independent of each other (**Figure 1c**, P=0.98), while others like speed, bursting or tortuosity correlated with activity (**Methods, Extended Data Figure 1l**, P<0.006). We performed several control experiments to show that the observed individuality was not

affected by potential artifacts in the set-up (**Methods, Extended Data Figure 2c-e**). Also, we tested that these behavioral parameters describe individuality also in response to stimuli like light flashes, mechanical perturbation or at a novel tank (**Methods, Extended Data Figure 2f).**

In a second step, we showed that individual differences were robust along several days (**Figure 1d,** R=0.69 and R=0.58, $P<0.001$ for linear correlation of activity and radial index, respectively, 7 vs. 8 dpf; **Extended Data Figure 2g,** R=0.48 and R=0.41, $P<0.01$, 5 vs. 6 dpf). The third step consisted in showing that inter-individual variability is larger than intra-individual variability. In the two-dimensional phenotypic space defined by activity and radial index, it can be directly seen that the area covered by the behavior of one individual is smaller than the area of the whole population (**Figure 1e-f,** left, 7-8 dpf; **Extended Data Figure 2a-b** right, 5-6 dpf, see **Methods**). Measuring variability by the Coefficient of Variation (CV), we found this difference to be significant (**Methods, Figure 1g,** 8 dpf; **Extended Data Figure 2h-j**, 5-7 dpf respectively; $P<0.001$ in all cases). Intra-individual and inter-individual variability levels remained stable from 5 to 8 dpf (**Extended Data Figure 2k**, $P<0.01$ in all comparisons**)**.

The degree of individuality may be measured by variability in the population. This can be visualized using the probability density of finding an individual in a population with a given mean activity and radial index (**Methods, Figure 1e-f**, right, 7-8 dpf; **Extended Data Figure 2a-b,** right, 5-6 dpf). It can be quantified using *generalized variance*[14], a single parameter that summarizes this two-dimensional variability, that we used to compare two populations (**Methods**; **Extended Data Table 1** for summary of all results; other variability measures in **Extended Data Table 2)**.

Sources of behavioral variability in zebrafish

Our setup allowed us to perform high-throughput tests to study the possible origins of behavioral individuality. Behavioral individuality could in principle depend on environmental manipulations and the genetic differences across the population. Our experiments minimized environmental influences by isolating eggs in plates at pharyngula stage (24 hpf) and by keeping them at a controlled temperature (27-28ºC). Manual changes in water (24 hours before the experiment) or feeding did not affect individuality (**Figure 2a,** P=0.42 and **Figure 2b**, P=0.38, respectively).

We found that behavioral variability of a population did not depend on the genetic variability of its individuals. Our control inbred WIK zebrafish population (F1) resulted from a single batch of eggs retrieved from two adults with at least three cycles of inbreeding. We obtained the same behavioral variability after two more inbreeding cycles (WIK F3, **Figure 2c**, P=0.33) and in an isogenic population[15] (CG2, **Figure 2c**, P=0.44). Also, we did not find changes in the behavioral variability using groups of siblings from genetically diverse outbred parents (LPS line, **Figure 2c**, P=0.38).

Changes in chromatin acetylation alter behavioral variability

The absence of effects from genetic variability prompted us to test whether behavioral individuality could be modified by different epigenetic factors. To test the contribution of DNA methylation we used 5-azacytidine (AZA), an inhibitor of DNA-methyltransferases. We found that AZA added to the water did not alter the behavioral individuality of a population (15 mM AZA, **Figure 3a**, P=0.44) even if it reduced 3-methyl DNA in larval zebrafish (**Extended Data Figure 3a**). We then studied the role of histone deacetylation, a reversible molecular process in which an acetyl functional group is removed from specific residues of Histone 3 and 4[9]. This

system is regulated by a group of enzymes called Histone Deacetylases (HDACs). To test the effect of HDACs on behavioral individuality, we used sodium butyrate (NaBu, a class I HDAC inhibitor) at the standard concentration of 2 mM[16]. We first confirmed that NaBu increases the level of total acetyl-histone 4 in larval zebrafish (**Extended Data Figure 3b**). We then found that this treatment reduced the behavioral variability of a WIK F3 sibling population after 24 hours (2 mM NaBu, **Figure 3b**) compared to control PBS-treated larvae (PBS, **Figure 3b**, $P<0.001$). Note that this treatment only altered variability and not the mean of the population parameters ($P=0.63$). When we retired the treatment, behavioral variability was recovered after additional 24h (**Extended Data Figure 3c**, $P=0.71$). Similarly to the behavior, the total levels of acetyl-histone 4 increased with the treatment and were recovered 24 hours after retiring the treatment (**Extended Data Figure 3b**). We also studied transgenic larvae with alterations in the deacetylation pathway as an alternative more specific than the use of drugs. We found that heterozygotic mutant populations of the histone deacetylase *hdac1* (*hdac1 +/-*) showed a reduced behavioral variability compared to their AB controls, mirroring the results obtained with the drugs (**Figure 3c**, $P=0.008$). Our results suggest that histone deacetylation pathway modulates the behavior of zebrafish larvae without affecting its average behavior.

High acetylation levels result in a behavior close to the population average

We have shown that the degree of behavioral variability of a population depends on its average acetylation levels. Since an increment in the global acetylation decreased the behavioral variability without changing the average behavior, we reasoned that the individuals with higher mean acetylation should be placed near the average population behavior in the phenotypic space. To test this hypothesis, we performed an experiment with 90 individuals to obtain their acetylation state depending on their distance to the average behavior of the population. As we

needed at least five larvae in order to get enough tissue for the experiment, we pooled 5 larvae with very similar behavior and measured their acetylation state. We found that larvae whose behavior was placed near the average of the population had higher mean acetylation values (**Figure 4a**, left). To quantify the dependence between the average acetylation and the position in the phenotypic space of the individuals, we first defined a coordinate system (centered on the average behavior of the population) and then obtained two magnitudes for each pool of fish: their distance to the center (r) and their angle with the horizontal axis (θ), **Figure 4a**, right. We found that the acetylation levels of the individuals highly correlated with their phenotypic distance to the average, while we found no correlation with their angular position (see **Figure 4b**, blue dots, $P<0.001$, $P=0.53$, respectively). We found similar correlations when we analyzed the distance to the mean of each behavioral parameter separately (**Extended Data Figure 4a**, $P<0.001$).

We have seen that individuals with higher acetylation levels display a behavior similar to the average of the population, while the variability of the population behavior increases at lower acetylation levels. This is compatible with our previous experiments that reduced the behavioral variability of a population by increasing its acetylation levels. In fact, if we use fish treated with NaBu to perform the same analysis, we find that the individuals not only have a behavior similar to the average of the non-treated population, but also that their acetylation is at the same level of the non-treated individuals with more acetylation (**Figure 4b**, red dots). This shows that the treated animals present acetylation levels within the physiological range of the animals, consistent with NaBu having the global effect of increasing the acetylation levels of the population by bringing them close to the animals with more acetylation.

Differences in the acetylation of a set of genes are linked to behavioral variability

Our results link acetylation level of the individuals and their behavioral individuality. However, fish with similar (low) acetylation levels also can show very different behavior, so there must be other factors contributing to behavioral variability. We hypothesized that these factors could be the acetylation differences in specific genomic regions associated to behavior. To explore this possibility, we compared the acetylation variability between two groups of WIK zebrafish, one with high and the other with low behavioral variability. For the first population, we built four clusters of five sibling fish with low intra-cluster behavioral variability and high inter-cluster variability, so that they covered the complete phenotypic space of the population (clusters c1-c4 in **Extended Data Figure 4b**, see **Methods** for details). For the second population, we built four behavioral clusters each made of up of five individuals selected at random from the phenotypic space. This process eliminates any systematic difference in behavior across clusters. We then retrieved the acH4 epigenomic profiles of the clusters in each group using chIP-seq and computed the acetylation variability of each genomic region across behavioral clusters using techniques adopted from gene expression analysis (see **Methods** for details).

We found that there were more genomic regions with higher variability in acetylation across clusters in the first population, suggesting that acetylation variability in the regions correlated with behavioral variability (**Figure 5a,** $P<0.0001$). We then identified the epigenomic regions with high variability in histone acetylation as they are potentially related to behavioral variability (**Methods**, **Extended Data Table 3,** $P<0.01$). We found that genes located near these regions are enriched in different Gene Ontology (GO) terms ($P<0.001$) mainly related to neurobiological processes (**Figure 5b**). This chIP-seq analysis predicts that acetylation variability in specific regions can be associated to behavioral differences. In order to assess the

specificity of these hypervariable regions in behavioral individuality, we checked that our previous relation between histone acetylation levels and average behavior (**Figure 4b**) was maintained in the hypervariable regions but not in the rest (see **Figure 5c**). To find if these regions could have a causal action in behavioral individuality, we decided to affect them by impairing DNA-interacting proteins that significantly bind near these regions. We found that several DNA motifs that were enriched near the hypervariable regions were Yin-Yang binding sites (P<0.0001, **Extended Data Figure 5c**). Yin-Yang 1 (YY1) is a transcription factor that can activate or repress the same target gene depending on recruited co-factors[17], like HDAC1[18]. To study the impact of YY1 on hypervariable regions, we selected eight of these regions. We first confirmed that the acetylation variability seen in the chIP-seq experiment was maintained in these regions using conventional chIP (**Extended Data Figure 5d**). Remarkably, the mRNA expression of the genes located near these regions also showed a high variability (**Extended Data Figure 5d**). Then, we studied the interaction between YY1 and these eight regions, and we found that YY1 is significantly bound to them (**Figure 5d**). This is consistent with YY1 affecting our hypervariable regions, so we decided to analyze their acetylation variability in a *yy1 +/-* mutant population. We found that *yy1 +/-* fish showed a significant reduction in the acetylation variability (**Figures 5e,** P<0.001). In addition, mRNA expression variability in near genes was also reduced in *yy1 +/-* fish compared to controls (**Figure 5f,** P<0.001). This reduction in molecular variability led us to analyze the behavioral variability of the *yy1 +/-* population, and we found that the individuality was significantly reduced (**Figure 5g,** P=0.003). We thus confirmed that hypervariable genes are related to behavioral and acetylation individuality by using YY1 impairment as a method to alter the regulation of hypervariable regions.

YY1 and HDAC1 regulate histone acetylation individuality

We investigated further how YY1 acts on the hypervariable regions in search of a possible mechanism for molecular variability. HDAC1 is a known partner of YY1[18] and we thus tested how this partnership can be related to hypervariable regions. We first confirmed that HDAC1 binds to the eight selected regions (**Figure 6a**). We then tested, using rechIP experiments, that YY1 and HDAC1 bind together to these regions (**Figure 6b**), suggesting that these two proteins form a complex that interacts with the hypervariable regions. As YY1 is known to be regulated through deacetylation by HDAC1[18] , we assessed the effect of sodium butyrate (NaBu) on YY1 acetylation in zebrafish larvae. Using co-immunoprecipitation, we found that YY1 was acetylated in control conditions and that this acetylation was impaired after NaBu treatment (**Figure 6c**). Furthermore, YY1 binding to the selected hypervariable regions was decreased in NaBu-treated larvae (**Figure 6d**), suggesting that a functional YY1 is needed to bind the hypervariable regions. A potential mechanism for the acetylation modulation consists in YY1 guiding the YY1/HDAC1 complex to the hypervariable regions, which would then be deacetylated by HDAC. Using a simple model that simulates stochastic binding of this complex to the hypervariable regions (see **Figure 6e** and **Methods** for the details of the model), we found that the stochasticity of the interactions is sufficient to generate variability in a population composed by identical individuals. The model gives individuality at the level of mean acetylation level and also in the concrete patterns of acetylation, here measured simply by the acetylation variability of the genome of each individual. Specifically, there is not only individuality at the average acetylation level but also at the acetylation variability level, considering it as a proxy of the differences between individual acetylation profiles (**Figure 6f**).

DISCUSSION

In this paper, we have found that a histone acetylation pathway modulates individual behavior in a genetic-independent manner without affecting the global average behavior of the population. Histone acetylation levels of an individual correlated with and its individual behavior compared to the average of the population. Therefore, while the average behavior might depend more strongly on genetic background (as seen for different strains in **Figure 2**) or environmental changes (as seen for different responses in **Extended Data Figure 2f**), behavioral individuality could result from histone acetylation.

Several stochastic mechanisms can underlie behavioral individuality, such as paternal and maternal effects, differences in the experience received by individuals, developmental noise or stochastic DNA binding, among others. We propose the stochastic action of the complex formed by YY1 and HDAC1 as part of the molecular machinery that translates these factors into acetylation differences.

Another open question is how acetylation variability could lead to behavioral variability. A possibility is that histone acetylation is functionally transformed into changes in gene expression, as we have shown for eight of the hypervariable regions. Genes located near the hypervariable regions are significantly related to several neurodevelopmental processes, so differences in their expression might result in differences in brain development and then ultimately in behavior.

METHODS

Zebrafish lines and care

Zebrafish (*Danio rerio*) WIK strain[19] was kindly provided by Dr. Bovolenta (CBM-UAM) and inbred in our laboratory for at least three generations before the experiments. Afterwards, WIK F1 population was generated from a single batch of embryos from a single couple of adult fish. Two additional cycles of inbreeding were carried out, crossing a couple of siblings from the former generation. CG2 clone population, generated by double gymnogenetic heat-shock, and characterized by being pure isogenic zebrafish was kindly provided by Dr. Revskoy (Univ Northwestern) as a control of reduced genetic differences between siblings. The outbred LPS (Local Pet Store) strain was recently described[20], and used as a model of genetic heterogeneity. Heterozygotic *hdac1* and *yy1* mutant strains with wild-type counterparts were obtained from ZIRC.

Care and breeding of the zebrafish strains were as described[20], with specific additional details. Eggs were isolated after 24 hours post-fertilization, and maintained in custom multiwell plates until 10 days post-fertilization (dpf). They were fed (JBL NovoBaby) from 6 dpf and water was changed daily if it is not indicated specifically in the experiment.

All the experiments using animals were approved and performed following the guidelines of the CSIC (Spain) and the Fundaçao Champalimaud (Portugal) for animal bioethics.

Free-swimming setup and recording

The setup consists of a monochrome camera located over the wells at a distance of 70 cm and pointing downwards. The camera used was a 1.2 MPixel camera (Basler A622f, with a Pentax objective of focal length 16 mm). The wells are circular, carved on transparent PMMA (24 wells per plate, and typically two plates are recorded simultaneously), and have their walls

tilted so that even in the most lateral wells the wall never hides the larva from the camera. Each well is 15 mm deep, and has a diameter of 1.8 mm at the bottom and a diameter of 30 mm at the top (**Extended Data Figure 1a**). For the experiments, each well is filled with a volume of 3 ml. The dishes are supported by a white PMMA surface that is only partially opaque. Behind this white surface we place two infrared led arrays (830nm, TSHG8400 Vishay Semiconductors) pointing outwards (**Extended Data Figure 1a**). Two paper sheets stand between the lights and the central space that lies directly under the wells. With this disposition we ensure that only diffuse indirect light reaches the wells, so that the illumination is roughly uniform (most of the light comes from below the wells through the white surface). All the set-up is surrounded by white curtains. Video camera recorded at a 25 fps rate (**Extended Data Figure 1b-c** for examples of a single frame and final trajectories).

A larval population (5-8 dpf) consisted of at least 24 fish siblings from the same batch of embryos. After five minutes of acclimation to the new environment, the larvae were recorded for 20 minutes. Water temperature was maintained in a strict range (27-28 ºC) during each experiment.

Custom-built software tracking larvae

We developed multiwellTracker, a software to automatically track zebrafish larva in wells. The software is available at http://www.multiwelltracker.es.

Detection of wells

The program is prepared to auto-detect circular wells, regardless of their spatial arrangement. To detect the wells we use the circular Hough transform (we have modified the code of Tao Peng distributed by Matlab Central under BSD license). In order to estimate the diameter of the wells, it computes the image's Hough transform for 100 radii different in 5 pixels

and a rough estimate of the largest possible radius (length of the longer side of the image divided by the square root of the number of wells) (**Extended Data Figure 1d**). The system selects the highest point of this measure as an estimate of the radius of the wells (r_{est}). It is possible to skip this first step and instead specify manually a value for r_{est}. This may be advisable when many videos are recorded with the same set-up and the same wells.

In the second step the system locates the centers of the wells. To do this it performs a Hough transform of the original image, this time with radii only in the range between $0.8r_{est}$ and $1.2r_{est}$. The transformed image usually has clear peaks in the centers of the wells. Then it filters the transformed image with a Gaussian filter to increase its smoothness (the resulting transformed image is shown in **Extended Data Figure 1e**). Then, it selects the maximum of the transformed image as the center of the first well. To prevent selecting the same well twice, the system discards all the pixels of the transformed image that are within radius r_{est} of the selected center (**Extended Data Figure 1f**). It selects the new maximum as the center of the second well, and repeats the procedure until all wells have been found (**Extended Data Figure 1g**). The experimenter can correct the result by manually clicking on the center of the wells that have not been correctly located (<1% of cases).

Pre-processing of images

In order to control for fluctuations in illumination, each frame is normalized by dividing the intensity of each pixel by the average intensity across all pixels of the frame. After normalizing the frame, a 2D Gaussian filter is used to smooth the image (**Extended Data Figure 1h** shows the image before and **Extended Data Figure 1i** after filtering).

Background subtraction and detection of the larva

In order to extract the image of the larva from the background, the system finds the average of 1,000 frames equi-spaced along the whole video. This average image is what we will call "static background". By subtracting each frame by the static background, we obtain an image in which the larvae correspond to dark regions (**Extended Data Figure 1j**). However, because of relatively slow changes in the set-up over time, the system uses the static background in combination with a dynamic background, which is computed as the average of the previous 5 frames. The difference between the current frame and the dynamic background will only show larvae that are moving in that precise moment (**Extended Data Figure 1k**).

The specific algorithm to detect the larva is as follows. First, the difference between the current frame and the static background is thresholded keeping only pixels for which the difference is below -0.5. We then find connected components ("blobs") in this thresholded image, keeping those that are larger than 1 pixel. Because these blobs come from the difference with the static background, both static and moving larvae will be detected. But at this stage some blobs come simply from noise. In order to filter out noisy blobs, the system accepts a blob if it fulfills at least one of these two conditions: (a) It contains at least one pixel that was identified as part of the larva in the previous frame or (b) it contains at least one pixel for which the difference between the current frame minus the dynamic background is below the same threshold as before (-0.5).

Removal of reflections

In most cases only one blob is obtained after the process described in previous sections. But when the larva is close to the wall of the well, its reflection on the wall may also be selected. The system considers that a blob A is a reflection of blob B when all of the following conditions are met: (a) Blob B is bigger than blob A, (b) blob B is closer to the center of the well than blob

A and (c) the lines between the center of the well and the two blobs form an angle < 10°. When these three conditions are met, the system removes blob A.

Acquisition of the position of the larvae

If more than one blob remains in the same well after the previous steps, the system selects the one with highest contrast. For the selected blob, the system takes the position of its most contrasted pixel, and adds this position to the trajectory of the larva. If in a well no blob remains after the previous steps, the trajectory is left with a gap. When this happens, the program will not re-track the larva until it moves.

Behavioral Parameters

Different parameters reflecting the behavior of individual larvae were measured, and finally two of them were used through the paper: (i) activity (percentage of time in movement) and (ii) radial index (average position from the border towards the center of the well). We also studied three additional parameters: (i) Tortuosity in the trajectory was calculated as the scalar product of the velocity vectors between two consecutive frames and the value in **Extended Data Figure 1l** was obtained by averaging this parameter through the whole video, excluding the frames where the animal was immobile. (ii) Speed was calculated as the average distance (in pixels) travelled per frame, in those frames where the fish was active. (iii) Bursting was obtained as the total number of frames where fish changed from immobility to motion. We found that these three parameters correlated with activity.

The average of each individual parameter was tested from 5 to 8 days post-fertilization (dpf) to assess if individual behavior was significantly stable along the days using Pearson coefficient of correlation.

386 **Additional validation of the experimental setup**

387 Several controls were performed for possible experimental artifacts affecting wells
388 differently. Behavioral parameters were robust to 90 degrees counterclockwise rotations of the
389 multi-well plate (**Extended Data Figure 2c,** left**,** R=0.73, *P*<0.001, and R=0.68, *P*<0.001; for
390 linear correlation tests of individual behavior) or to interchanging the larvae between outer and
391 inner wells (**Extended Data Figure 2c,** right**,** R=0.65, *P*<0.001, and R=0.61, *P*<0.001; using the
392 same correlation test). Also, we found no correlation between the small differences in
393 illumination across wells and behavior (**Extended Data Figure 2d**). We further corroborated
394 using a significance test that the differences in behavior did not have an origin in systematic
395 differences across wells. For this, we found that the average behavioral parameters obtained in
396 fifteen individual experiments were not different between wells (**Extended Data Figure 2e**,
397 typically *P*=0.4 and always *P*>0.19 for both parameters).

398

399 **Stimulus response tests**

400 We studied the influence that our free-swimming behavioral parameters could have on
401 the performance of the individuals when they respond to three different stimuli.

402 *Response to mechanical disturbance*

403 We applied mechanical perturbations to each larva by pipetting up and down the water
404 content of the well for four times. Perturbations were applied at 6 dpf to previously recorded
405 animals, and the 20-minute recording was done at 7 dpf. The recording was performed in the
406 usual setup.

407 *Response to strong Light Pulse*

In complete darkness, we applied three different light flashes to the larvae and study their behavior in the 90 subsequent seconds. The flashes and the recording were performed in the usual setup. Pre-recording behavioral parameters were obtained the day before.

Novel tank with light/dark preference

In order to study the effect that a novel setup could have on the behavior of larvae we built a rectangular setup, which changed the geometry of the previous circular wells. The setup dimensions were 84 mm x 21 mm and it was built in transparent acrylic. To try to see if our parameters had any effect on the light-dark preference, half of the floor of the setup (42 mm x 21 mm) was white while the other half was black. The height of the setup was 5 mm. Larvae were placed in the center of the white part and recorded for 10 minutes. Activity was calculated as previously described and distance to the wall was represented by the average distance to the longest walls, normalized to 1 in the middle point of both walls and to 0 in the exact position of the walls.

The effect of our behavioral tests resulted in a decrease (increase) in mean activity (radial index), but maintaining the same individuality of the pre-recorded free-swimming experiments (**Figure S2f;** $P<0.04$ for changes in mean activity and radial index compared to control larvae of the same age; $P<0.02$ for linear correlation of activity and radial index. In the case of novel tank, radial index cannot be applied because the wells are elongated and was then replaced by the minimum distance to the longer walls. We note, however, that this parameter showed no correlation with the radial index of pre-recorded experiments in the same animals.

Inter-individual vs. Intra-individual differences

The behavioral parameters (activity and radial index) were also obtained from consecutive fragments of 30 seconds for each 20-minute experiment for each larva. This was fitted to a two-dimensional Gaussian, but for clarity when representing many animals (like in **Figure 1e-f,** left) an isocontour of the Gaussian for each animal was used. An isocontour is an ellipse with principal axes given by the eigenvectors of the covariance matrix. We chose the isocontour with length of each semiaxis given by the square root of the eigenvalue of the covariance matrix, as this reduces to the standard deviation in each direction for cases with no correlation between the two variables. Intra-individual variation distribution was obtained using the coefficients of variations (CVs) for activity and radial index separately. Inter-individual variation was calculated the same way but using fragments from different fish.

Comparing the behavioral variability between two animal groups

A simple visual method to characterize the variability in a population is to plot the bi-dimensional distribution of the activity and radial index of individuals (like in **Figure 1e-f, right**). To do so, we used Gaussian kernel smoothing that consists in adding up Gaussians centered at the data points as

$$P(x,y)=\frac{1}{N}\frac{1}{2\pi\sigma_x\sigma_y}\sum_{i=1}^{N}\exp\left(-\frac{1}{2}\left[\frac{(x-x_i)^2}{\sigma_x^2}+\frac{(y-y_i)^2}{\sigma_y^2}\right]\right)$$

with x_i and y_i the mean activity and radial index values of individual i of a total of N members of the population. An optimal smoothing uses standard deviations of each Gaussian given by $\sigma_x = N^{-1/6}\alpha_x$ with α_x the standard deviation in the x_i data values, and similarly for σ_y using the y_i values (see B.E. Hansen, unpublished manuscript, http://www.ssc.wisc.edu/~bhansen/718/NonParametrics1.pdf). The volume under the probability

surface has a value of 1, even when the values of the probability density are already up to 90. The probability surface sits on an area on the x-y plane of approximately 0.4x0.4, making the total volume under the surface to be 1.

While this distribution gives a visual and intuitive characterization of behavioral variability, an even simpler characterization is achieved using, for each group, a single parameter summarizing its two-dimensional variability. We used *generalized variance* as this single parameter, measured as the determinant of the covariance matrix (**Extended Data Table 1**), while other parameters like the standard deviation for each parameter gave similar statistical results (**Extended Data Table 2**).

ChIP-seq, conventional chIP, and rechIP analyses

Clusters of four fish were built for studying the molecular variability of a population, as the amount of tissue in a single larva was not sufficient to measure histone acetylation levels. In addition, the clustering also reduced the noise that could arise from the molecular analysis of a single larva. In the first (behavioral) cluster, a Hierarchical Clustering analysis using Euclidean distance as the metric and the average linkage clustering as the linkage criteria was used the clustering in the chIP-seq, as the total population consisted of 72 larvae. The selection in the NaBu experiment used the same algorithm. In the random experiment, fish were randomly selected from the population.

The samples were crosslinked with 1.8% formaldehyde for 30' and then quenched with 1% glycine for 5'. Extracts were lysed using a SDS Lysis buffer (50 mM Tris-HCl pH 8.1, 1% SDS, 10 mM EDTA) for 30' at 4ºC, and then diluted with a Dilution buffer (6.7 mM Tris-HCl pH 8.1, 0.01% SDS, 1.2 mM EDTA, 1.1% Triton X-100, 167 mM NaCl). 2 mM sodium butyrate

was added to avoid histone deacetylation activity during the preparation. Then, the fish were sonicated with two pulses (30" ON / 30" OFF) of 15' each with the Diagenode Bioruptor. Before pre-clearing the samples with protein A/G beads, an input sample was obtained. Then, the extracts were immunoprecipitated overnight using 1 µg of the anti- acetyl-Histone 4, anti-GAPDH, anti-HDAC1 or anti-YY1 antibodies. Bound DNA was recovered with protein A/G beads, then washed with Low-Salt (120 mM Tris-HCl pH 8, 0.1% SDS, 2 mM EDTA, 1% Triton X-100, 150 mM NaCl), High-Salt (120 mM Tris-HCl pH8, 0.1% SDS, 2 mM EDTA, 1% Triton X-100, 500 mM NaCl), LiCl (10 mM Tris-HCl pH 10, 1 mM EDTA, 0.25 M LiCl, 1% NP40, 1% sodium deoxycholate) and two times with 1X TE (10 mM Tris-HCl pH 8, 1 mM EDTA) buffers, and recovered with Elution (1% SDS, 0.1 M NaHCO3). DNA purified samples were de-crosslinked using sodium chloride, and cleared with Qiagen spin columns. In the case of rechIP, the samples were reincubated with the second antibody after elution, and another round of washes were performed.

The final samples were processed at the Genomics Unit at the Scientific Park of Madrid in the case of chIP-seq experiments. Libraries were built, and the samples were sequenced using an Illumina GAII. Raw data of the experiment can be obtained in the NCBI GEO repository (GSE IDXXX). Reads were aligned to *Danio rerio* genome sequence (Zv7) with BWA, and final reads in a 25-bp window were mapped to the reference genome using custom Perl scripts. The repetitive regions were removed from the analysis and the results were also normalized using the average number of reads in each sample. The standard deviation of every 25 bp region for each experiment (Behavior, Random and NaBu) was calculated, and then we quantified the probability for each region of being more variable than in the Random experiment. We selected the regions with $P<0.01$ and with the highest standard deviation (top 25%). Nearest genes were

retrieved, and their human orthologs were analyzed for Gene Ontology. In the case of conventional chIP or rechIP, qPCR analyses for amplification differences using specific primers were performed. In addition, hypervariable acetylated regions with their flanking sequences (300 bp) were used to predict enriched DNA motifs and their potential biological activity with MEME suite[21].

Reagents and antibodies

Sodium butyrate and AZA (Sigma-Aldrich) was dissolved in Phosphate-buffered saline (PBS), and used in a final 2 mM and 15 mM concentration of fish water, respectively. PBS alone was used as vehicle control. The pharmacological treatment lasted for 24 hours from 7 dpf to 8dpf. Acetyl-Histone 4 antibody was obtained from Promega, anti-HDAC1 from Epigentek, anti-YY1 antibody from SantaCruz, anti-GAPDH from Sigma-Aldrich, and McrBC enzyme from New England Biolabs.

Western Immunoblotting

Treated and untreated groups of fish (5-10) were frozen at different times, and then protein extracts were isolated from tissue using an extraction buffer (80 mM Tris-HCl pH 7.5, 2 mM EDTA, 2 mM EGTA, 0.27 M Saccharose, 10 mM β-glycerolphosphate, 5 mM Sodium pyrophosphate, 50 mM Sodium Fluoride, 1% Triton X-100, 0.1 mM Sodium vanadate, 0.1% β-Mercaptoethanol, 1X Complete protease inhibitor cocktail) during 30' at 4ºC with vortexing. After centrifugation, debris was removed, and the protein content was measured with Coomasie reagent (Pierce). Anti- acetyl-Histone 4, anti- ß-Actin and secondary antibodies were used following manufacturer recommendations.

521

522 **RNA isolation and qPCR quantification**

523 Total RNA was isolated using homogenized extracts from three fish per sample by Trizol
524 (Life Technologies) extraction and RNAeasy (Qiagen) purification. Retrotranscription was done
525 with iScript (Bio-Rad) following manufacturer recommendations. Finally, quantification of the
526 target genes was measured using qPCR with specific primers, and p-values obtained by using
527 Student's T-test.

528

529 **Quantification of Histone 4 acetylation levels**

530 Eighteen clusters of five fish from a total population of 90 were obtained from the
531 behavioral space (activity/radial index) using an *ad hoc* algorithm. First, 18 centroids were
532 randomly chosen, and 5 individuals were assigned to the nearest (not occupied) centroid. Then,
533 centroids were redefined using the average values of the new clusters, and a new round of
534 assignment of the fish to the centroids was done. This iteration was repeated until the centroids
535 were stable. Then, acetyl-H4 levels were quantified using Epigentek kit and following
536 manufacturer recommendations.

537

538 **Quantification of methylated DNA**

539 DNA methylation was quantified using larval DNA digested by MCrBC enzyme as
540 previously done[22] following kit instructions.

541

542 **Simulation of YY1/HDAC1 complex activity**

We simulated how the YY1/HDAC1 complex could stochastically deacetylate epigenomic regions. Regions can be considered as transcription factor binding sites (we used *Nreg=1,000* regions **Figure 6f**). Each region starts with an initial acetylation value given by a Gaussian distribution with a fixed mean acetylation and a standard deviation. The value of the mean μ is arbitrary as long as the relation of the rest of the parameters with this one is fixed and the value of the standard deviation is *μ/10* (in **Figure 6f** we used *μ=10*). We fixed the number molecules of the complex to be the same for each fish (*Complex molecules=3,000*). The complex binds to randomly selected regions and, once bound, the complex decreases the acetylation of each region by an amount of 0.3·μ. We allowed several complexes to bind at each region. The complex has an error rate of 10% so, once bound, the probability of deacetylating is 0.9. This error rate has the final consequence of generating different mean and variance acetylation values for the individuals of a population (**Figure 6f**). The minimum acetylation value for each region is 0 and, once reached, the complex cannot select that region for further binding. We repeated the whole process for 100 fish, each one representing a point in **Figure 6f**.

The pseudocode of the simulation would be as follows:

while *Complex molecules > 0*

Complex binds to a region r, p(binding)=1/Nregions

Complex deacetylates region r by an amount of 0.3·μ, with a probability p=0.9

YY1 molecules=YY1 molecules-1

endwhile

Statistical analysis

All statistical tests to compare the differences between two distributions were conducted by calculating the value of their representative parameter, shuffling randomly the data of both distributions for 1,000 - 10,000 times and computing a *P*-value given by the proportion of times in which the difference in the representative parameter of the random distributions was higher than their original value. All the experiments were done at least three times with different biological datasets, and *P*-values were calculated using the three replicas. Figures show a representative experiment of the triplicate. MATLAB was used for all the computations and the statistical analysis.

AUTHOR CONTRIBUTIONS

A.-C.R. designed project, performed experiments and analysis and wrote the paper, J.V.-P. performed experiments and analysis and wrote the paper, A.P.-E. developed the tracking software, J.M.C.-G. performed experiments, P.M.F.S provided reagents, and G.G.d.P. designed and supervised project, performed analysis and wrote the paper.

ACKNOWLEDGEMENTS

We thank Rui Costa, Carlos Ribeiro, Antonia Groneberg, and other members of the Champalimaud Neuroscience Programme for discussions, and Isidro Dompablo (CSIC) and Ana Catarina Certal (Champalimaud Foundation) for technical assistance and animal care. CG2 and WIK lines were kind gifts from Sergei Revskoy and Paola Bovolenta, respectively. This work was supported by a Juan de la Cierva and a FEBS Long-Term fellowship (to A.-C.R.), a JAE-CSIC Predoctoral Fellowship (to J.V.P.), Spanish Plan Nacional Ministerio de Economia y Competitividad grants BFU2014-54699-P (to J.M.C.-G.), BFU2011-22678 (to P.M.F.S),

BFU2012-33448 (to G.G.d.P.), ERASysBio+ initiative supported under the European Union European Research Area Networks (ERA-NET) Plus scheme in Framework Program 7 (to G.G.d.P.), Fundaçao para a Ciência e Tecnologia PTDC/NEU-SCC/0948/2014 (to G.G.d.P.) and Champalimaud Foundation (to G.G.d.P.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES:

1. Galton, F. *English men of science, their nature and nurture*. (1874).
2. Gärtner, K. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Lab. Anim.* **24,** 71–77 (1990).
3. Kain, J. S., Stokes, C. & de Bivort, B. L. Phototactic personality in fruit flies and its suppression by serotonin and white. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 19834–9 (2012).
4. Freund, J. *et al.* Emergence of individuality in genetically identical mice. *Science* **340,** 756–9 (2013).
5. Waddington, C. H. The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. *Strateg. genes A Discuss. some* … ix +-262 . (1957). doi:10.1007/3-540-32786-X_7
6. Seong, K. H., Li, D., Shimizu, H., Nakamura, R. & Ishii, S. Inheritance of stress-induced, ATF-2-dependent epigenetic change. *Cell* **145,** 1049–1061 (2011).
7. Herb, B. R. *et al.* Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nat. Neurosci.* **15,** 1371–1373 (2012).
8. Laplant, Q. *et al.* Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. *Nat. Neurosci.* **13,** 1137–1143 (2010).
9. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389,** 349–352 (1997).
10. Wang, H., Duclot, F., Liu, Y., Wang, Z. & Kabbaj, M. Histone deacetylase inhibitors facilitate partner preference formation in female prairie voles. *Nat. Neurosci.* **16,** 919–24 (2013).
11. Simola, D. F. *et al.* Epigenetic (re)programming of caste-specific behavior in the ant Camponotus floridanus. *Science (80-.).* **351,** (2015).
12. Pantoja, C. *et al.* Neuromodulatory Regulation of Behavioral Individuality in Zebrafish. *Neuron* **91,** 587–601 (2016).
13. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10604–10609 (2005).

14. Wilks, S. S. Certain generalizations in the analysis of variance. *Biometrika* **24,** 471–494 (1932).

15. Mizgirev, I. V. & Revskoy, S. A new zebrafish model for experimental leukemia therapy. *Cancer Biol. Ther.* **9,** 895–903 (2010).

16. Heruth, D. P., Zirnstein, G. W., Bradley, J. F. & Rothberg, P. G. Sodium butyrate causes an increase in the block to transcriptional elongation in the c-myc gene in SW837 rectal carcinoma cells. *J. Biol. Chem.* **268,** 20466–20472 (1993).

17. Shi, Y., Seto, E., Chang, L. S. & Shenk, T. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* **67,** 377–388 (1991).

18. Yao, Y. L., Yang, W. M. & Seto, E. Regulation of transcription factor YY1 by acetylation and deacetylation. *Mol. Cell. Biol.* **21,** 5979–5991 (2001).

19. Nechiporuk, A., Finney, J. E., Keating, M. T. & Johnson, S. L. Assessment of polymorphism in zebrafish mapping strains. *Genome Res.* **9,** 1231–1238 (1999).

20. Pérez-Escudero, A., Vicente-Page, J., Hinz, R. C., Arganda, S. & de Polavieja, G. G. idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nat. Methods* **11,** 743–8 (2014).

21. Bailey, T. L. *et al.* MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37,** (2009).

22. Torres-Núñez, E. *et al.* Matricellular protein SPARC/osteonectin expression is regulated by DNA methylation in its core promoter region. *Dev. Dyn.* **244,** 693–702 (2015).

FIGURES AND FIGURE CAPTIONS

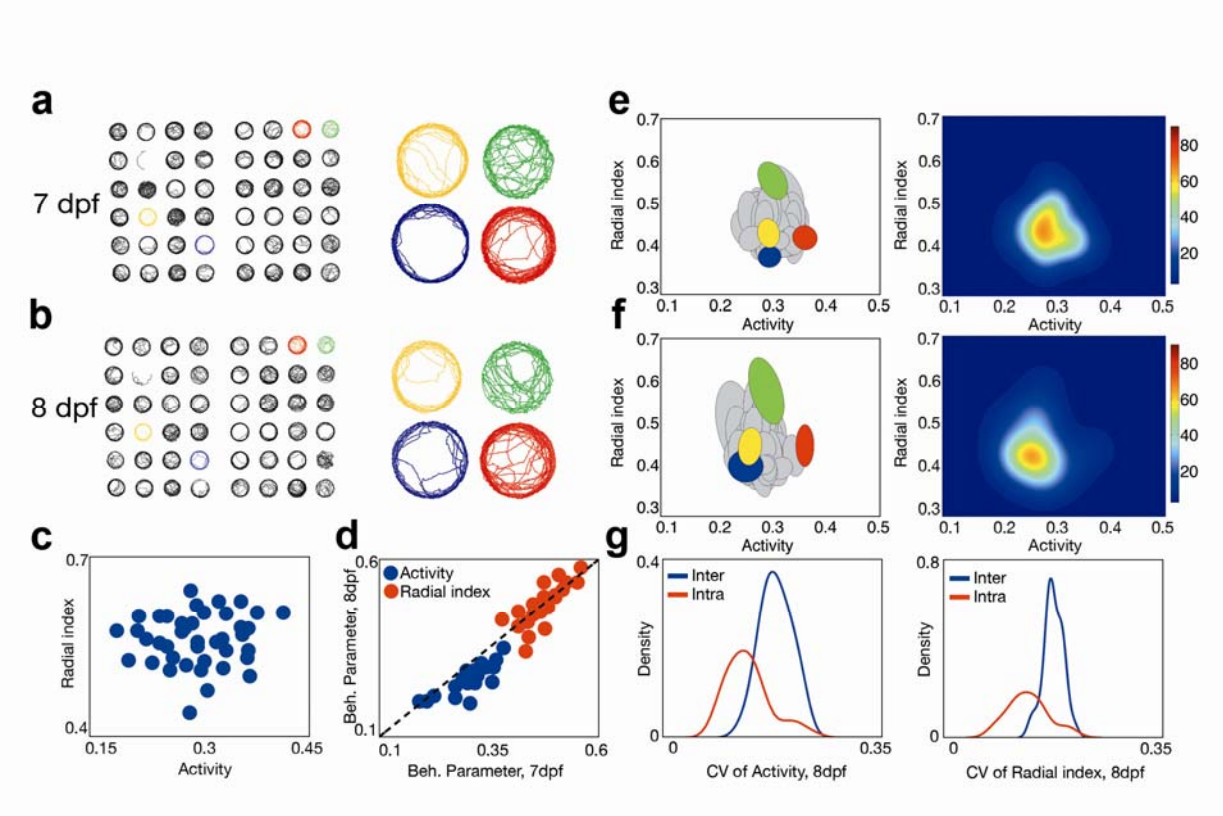
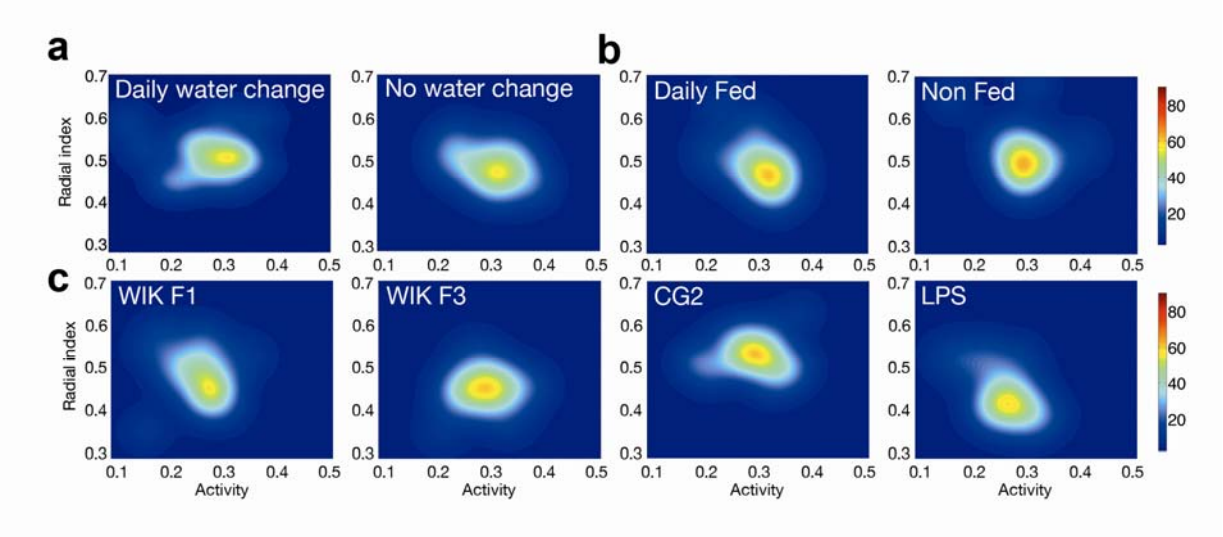


Figure 1. **Behavioral individuality in a population of 48 larval zebrafish. (a)** Example 20-minute trajectories for the group at 7 dpf (left), zoomed in for four fish (right). **(b)** The same group at 8 dpf. **(c)** Radial index *vs*. activity at 7 dpf. **(d)** Correlation of activity (blue) and radial index (red) between 7 dpf and 8 dpf for the same group. **(e)** Left: Population variability in activity and radial index of the same group at 7 dpf. Each ellipse represents the behavioral variability for each single fish as described in **Methods**. Colors as in **a**. Right: Probability density of finding an individual with a given mean activity and radial index. **(f)** Same as **e** but at 8 dpf. **(g)** Smoothed histogram of the intra-individual variability (red) and inter-individual

662 variability (blue) of the same group at 8 dpf, measured as CV of activity (left) and radial index

663 (right).

664



665

666 **Figure 2**. **Environmental changes and genetic background do not have an impact on**
667 **behavioral variability. (a)** Probability density of finding an individual with a given mean
668 activity and radial index for additional larval groups with and without daily water changes, at 7
669 dpf. **(b)** Same as **a**, but for additional daily fed and non-fed groups. **(c)** Same as **a** for additional
670 groups with different genetic backgrounds: WIK F1 (three inbreeding cycles), WIK F3 (five
671 inbreeding cycles), CG2 (gymnozygotic fish clones) and LPS (outbred parents).

672

673

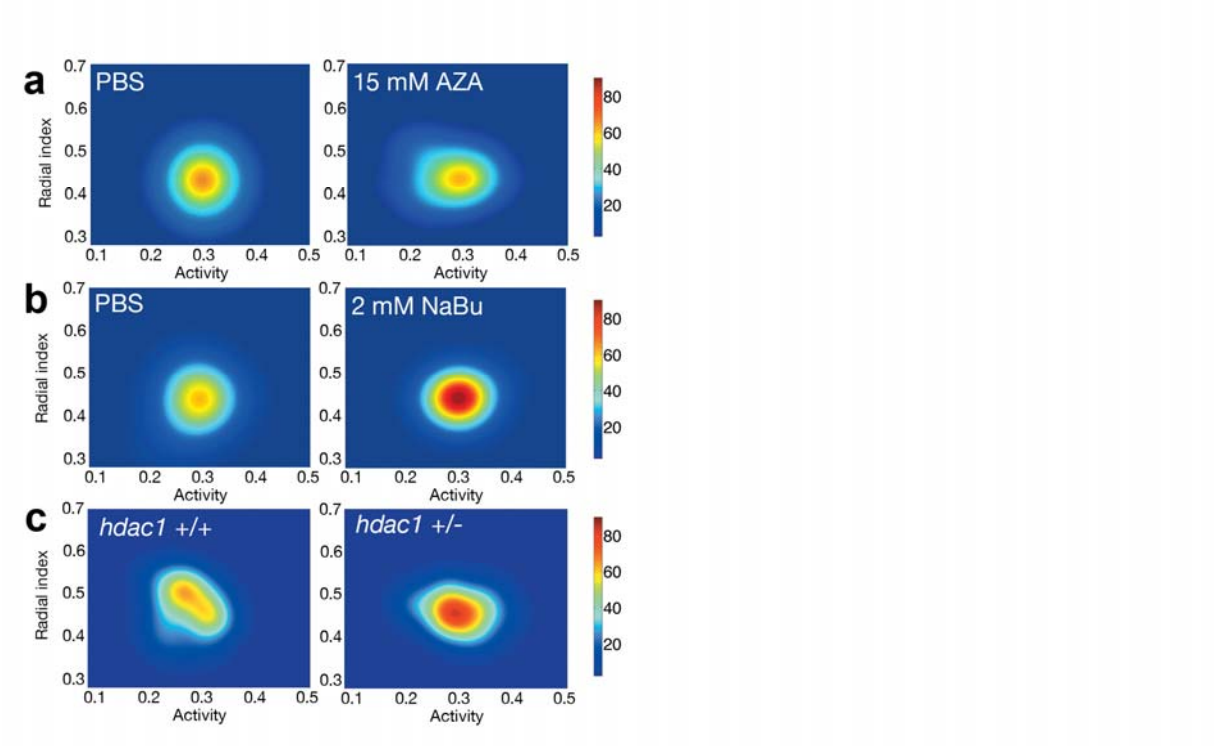


Figure 3. **Histone acetylation modulates behavioral variability. (a)** Probability density map for fish treated with a PBS solution as control and AZA. (**b**) The same for PBS and NaBu. **(c)** Probability density map for *hdac1* +/+ and *hdac1* +/- larvae.

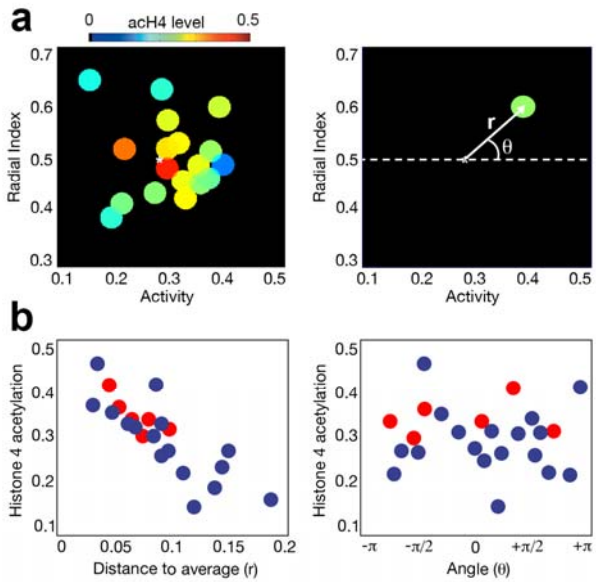


Figure 4. Relation between acetylation levels and behavior. (a) Average acetylation levels of fish depending on their behavior (left). Schematic representation of the two parameters used to analyze the dependence between acetylation and behavior (right). **(b)** Relation between the values of histone 4 acetylation and the two parameters of the coordinate system centered on the average behavior of the population: the distance to the average (left) and the angle with the horizontal axis (right).

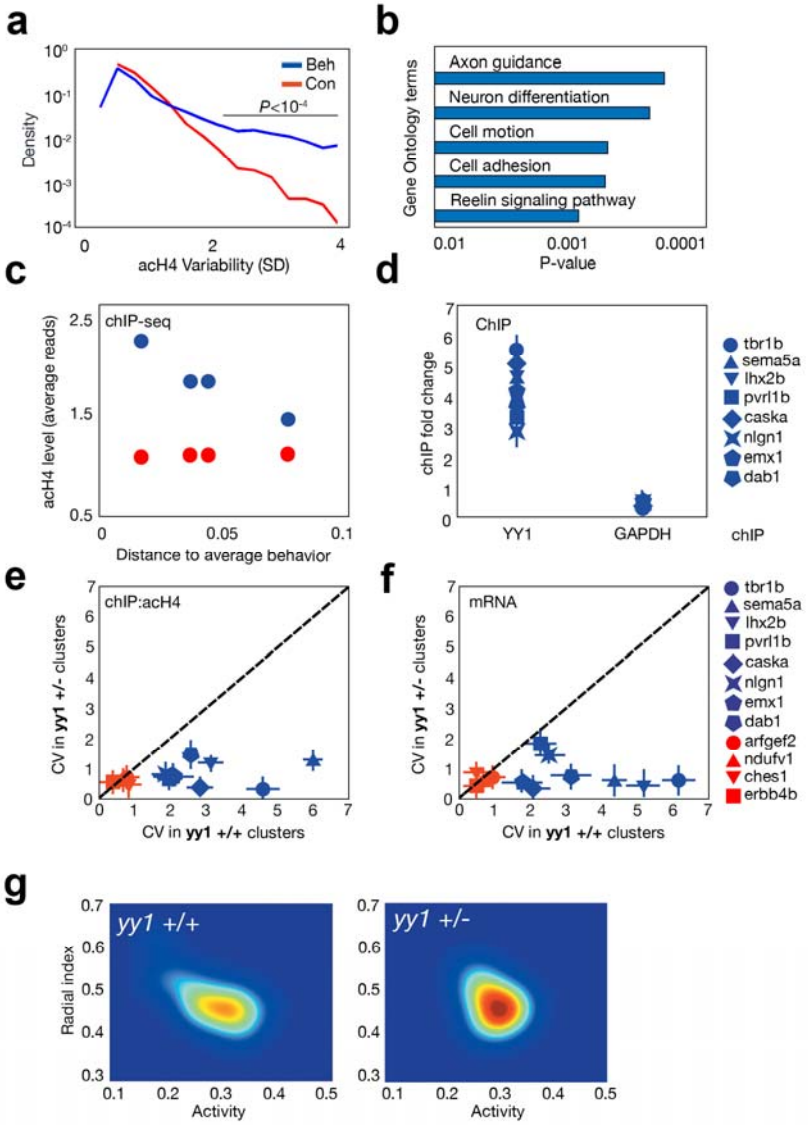


Figure 5. Hypervariable acetylation regions related to behavioral individuality. (a) Probability distribution of the variability (SD) in the acetylation of the behavioral (blue) and random control (red) clusters, obtained with chIP-seq. Horizontal line, $P<10^{-4}$. **(b)** Gene ontology of the hypervariably acetylated regions. **(c)** Relation between normalized number of reads in each cluster of the chIP-seq experiment and their distance to average behavior, as in **Figure 4b**. Blue dots represent the average of the hypervariable regions and red dots represent the average of the rest of the regions. **(d)** YY1 and GAPDH binding to eight selected hypervariable regions.

697 **(e)** Comparison between acH4 variability of the eight regions in *yy1* +/+ and *yy1* +/-
698 populations. **(f)** The same as **e**, but for mRNA expression variability of genes located near the
699 regions. **(g)** Probability density map for *yy1* +/+ and *yy1* +/- larvae.
700
701
702
703
704

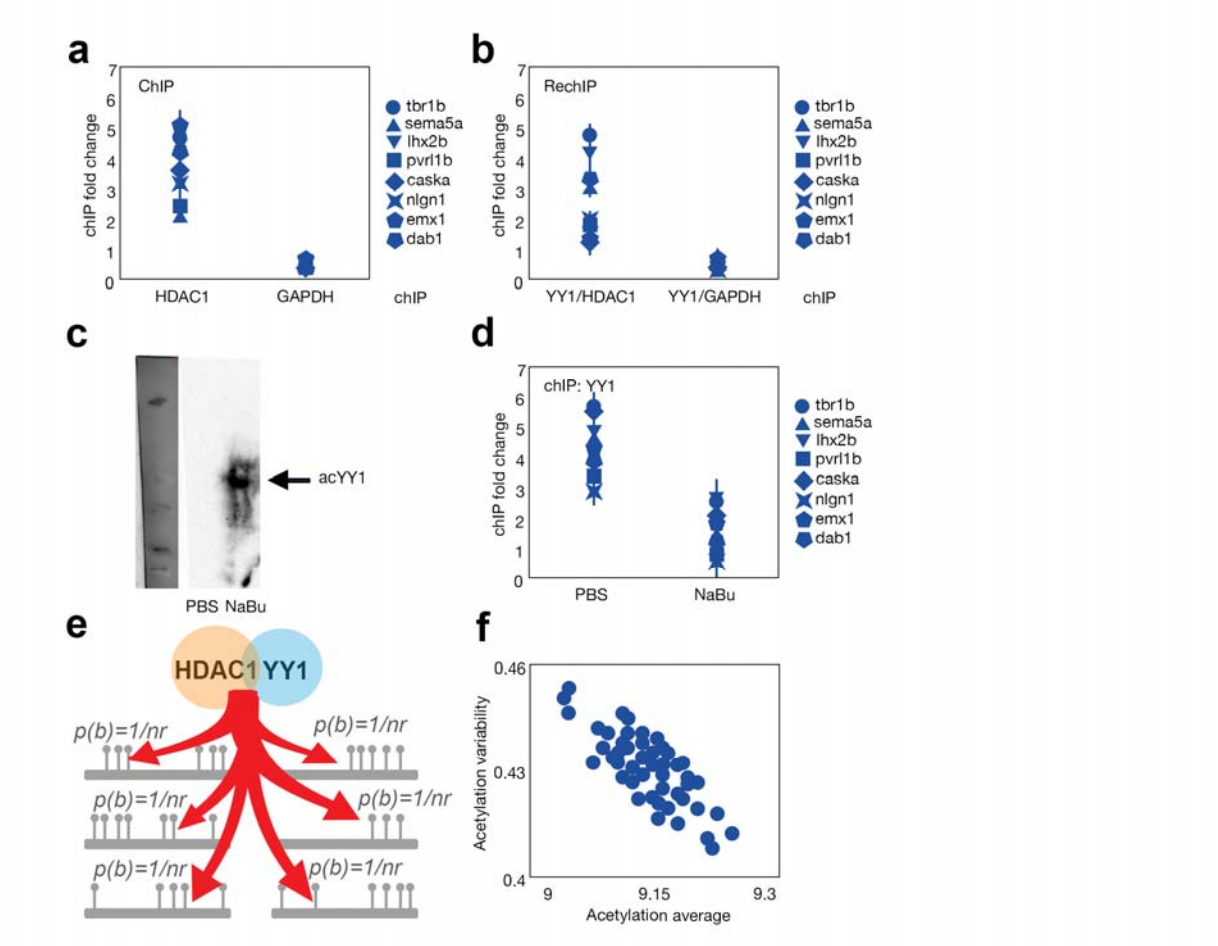


Figure 6. **YY1 and HDAC1 regulate acetylation individuality**. **(a)** YY1 and GAPDH binding to eight selected hypervariable regions. **(b)** RechIP binding of YY1/HDAC1 and YY1/GAPDH to the same regions. **(c)** Co-immunoprecipitation of acetyl-lysine and YY1 in larvae treated with PBS and NaBu, respectively. **(d)** YY1 binding to the eight regions in larvae treated with PBS and NaBu. **(e)** Schematic model of the stochastic action of YY1/HDAC1 complex on a set of epigenomic regions. All regions have the same binding probability, given by $p(b)=1/n_r$, where n_r is the total number of regions .**(f)** Correlation between the acetylation variability (measured as the coefficient of variation of the acetylation across all regions) and the average acetylation of the individuals of a population.