MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 6100–6115 doi: 10.1111/mec.12556

Combining animal personalities with transcriptomics resolves individual variation within a wild-type zebrafish population and identifies underpinning molecular differences in brain function

S. REY,*[1] S. BOLTANA,*†[1] R. VARGAS,* N. ROHER* and S. MACKENZIE*†

Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Barcelona, 08193 Bellaterra, Spain, †Institute of Aquaculture, University of Stirling, Stirling, Stirlingshire FK9 4LA, UK

Abstract

Resolving phenotype variation within a population in response to environmental perturbation is central to understanding biological adaptation. Relating meaningful adaptive changes at the level of the transcriptome requires the identification of processes that have a functional significance for the individual. This remains a major objective towards understanding the complex interactions between environmental demand and an individual's capacity to respond to such demands. The interpretation of such interactions and the significance of biological variation between individuals from the same or different populations remain a difficult and under-addressed question. Here, we provide evidence that variation in gene expression between individuals in a zebrafish population can be partially resolved by a priori screening for animal personality and accounts for >9% of observed variation in the brain transcriptome. Proactive and reactive individuals within a wild-type population exhibit consistent behavioural responses over time and context that relates to underlying differences in regulated gene networks and predicted protein–protein interactions. These differences can be mapped to distinct regions of the brain and provide a foundation towards understanding the coordination of underpinning adaptive molecular events within populations.

Keywords: proactive, reactive, behaviour, gene expression, variation

Received 18 July 2013; accepted 23 September 2013

Introduction

Understanding phenotypical variation and physiological regulation during acclimation to external stimuli variation is a central issue in biology and underpins evolutionary studies. In most studies across the biological sciences, individuals within groups are considered as simple replicates, and interindividual variation is actively minimized, for example, using specific genetic backgrounds. Such practice is commonplace in biology and reflects the difficulties in understanding variation and the prevalence of an approach often centred upon differences between mean values of measured variables for a population. However, it is common to observe significant levels of variation, standard deviation, in many different types of data that more probably obscure the underlying individual differences that may have potential biological significance. Such an approach therefore cannot be used to understand the underlying interactions between individuals and their environment. Over the past few years, an increasing interest in the diversity of behavioural phenotypes and their consistency in and between individuals within a population has emerged (Sih *et al.* 2004; Wolf *et al.* 2007). This has been accompanied by a growing nomenclature including individual coping style, personality and behavioural

Correspondence: S. MacKenzie, Fax: + 44 (0) 1786 472133; E-mail: Simon.Mackenzie@stir.ac.uk

[1]These two authors contributed equally to the manuscript.

syndromes that is typical to an intense interdisciplinary effort (Dall *et al.* 2012). It has been suggested that the use of such consistent behavioural subgroups within a population greatly facilitates the interpretation of measured biological responses by helping to resolve variation in experimental measurements (MacKenzie *et al.* 2009).

The adaptive response to changing environmental conditions requires both genetic and epigenetic factors acting in concert, and individuals within a population have been shown to display significantly different responses at the physiological level (Øverli *et al.* 2007). To attain an adaptive response, a scaled suite of multi-directional regulatory processes, from the transcriptome to behaviour, interacts to optimize individual fitness. Such interdependent responses at multiple scales of biological regulation aka 'complex system' should be consistent over time and context. These requirements have been described at the level of animal behaviour and have been referred to as 'animal personality' (Sih *et al.* 2004; Dingemanse *et al.* 2010). Furthermore, consistent individual variability in physiological and behavioural responses to challenge named as individual or stress coping style has been reported within animals of the same species, sex and population across a wide range of vertebrates including fish (Korte *et al.* 2005; Ruiz Gomez *et al.* 2008; MacKenzie *et al.* 2009; Huntingford *et al.* 2010). Consistent behavioural responses that in this study we shall refer to as personality can be broadly described as proactive, reactive and intermediate individuals within a population. This is based upon consistency over time of specific individual behavioural traits such as risk taking, aggression, activity and feeding. The importance of understanding consistent differences between individual behaviour is manifest and importantly facilitates the understanding and effective measurement of an individual's response during environmental adaptation (Dingemanse *et al.* 2010).

In an initial study, we have shown how *a priori* screening for proactive and reactive personalities can significantly reduce within-population variation in gene expression studies and increase the interpretative value of physiological data (MacKenzie *et al.* 2009). However, targeted gene expression studies are limited and cannot describe a more complete system approach. A transcriptome or gene expression profile is a collection of mRNAs measured by microarray or RNA-seq within a cell, tissue or organism that represents the available transcripts at a specific point in time. Thus, measured transcript profiles represent phenotypes shaped by the genotype of an organism and the environment under which it exists. The transcriptome is dynamic and can be influenced by many factors, and in recent years, this approach has been used to address diverse ecological, evolutionary and environmental questions (Goetz & MacKenzie 2008). A major issue in such studies is to identify which changes in the transcriptome have a functional significance for the individual. This is mainly based upon the identification of differentially expressed mRNAs between defined experimental groups, for example, A vs B, where comparisons are time-, treatment- or environment-dependent. The interpretation of such interactions and the significance of biological variation between individuals from the same or different populations remain a difficult and under-addressed question (Whitehead & Crawford 2005, 2006a,b). Several studies have shown that within populations of distinct vertebrates including humans, mice and fish, significant variation in gene expression is common and in some cases may be higher than that observed among populations (Whitehead & Crawford 2006a). The reasons behind the observed variation are currently unknown although both the ecological context and genetic variation among individuals are two key contributory factors. As a result, the possibility to identify conserved responses with a functional relevance both within and across different vertebrate taxa *aka* bridging the phenotype-genotype gap remains a daunting task. Such considerations raise questions about how to relate differences at the transcriptome level to adaptive phenotypic variation (Crawford & Oleksiak 2007).

In this study, we describe animal personalities in a wild-type zebrafish population where proactive and reactive individuals display time- and context-specific differences in behaviour. This analysis is then used to resolve variation in a large-scale transcriptomic data set and provides insight into personality-dependent differences variation in the brain transcriptome. This increased resolution allows the identification and *in silico* localization of inter-related gene expression modules with biological significance, thus providing a foundation for future studies addressing adaptive variation within populations.

Materials and methods

A whole population of Zebrafish (*Danio rerio*), males and females, were obtained from a commercial supplier (PS®, Parets del Valles, Barcelona) as juveniles and maintained at standard conditions (12-h light/dark cycle, mean T°: 28 ± 1 °C) on a stocking tank in our aquarium facilities at a stocking density of one fish per litre. The experimental tank for most of the behavioural experiments was a 20-l glass aquarium (40 × 25 × 20 cm) lined on three sides with white paper; the front wall was not covered to allow the observer to record the behaviour and divided at 1/3 of its length with a black PVC screen with a 3-cm-diameter

hole in the middle. All tank surfaces around this third area of the tank were covered with dark paper and closed on the upper part with a removable lid to provide a shelter for the animals. For the first two tests, the hole was covered with the same PVC plastic material and removed once the screening started to allow the fish enter the novel environment. For the first activity test, there was no shelter, and the tank was divided into 16 equal squares. For the second activity test, we created a shelter, but the whole lid was removed, and the tank was divided into four equal squares to measure activity and square occupation (1 to 4 ordered in an inverted Z shape plus a 1/3 tank size shelter). For the mirror and the feeding test, a 9-litre aquarium (30 × 15 × 20 cm) was used. In the MIS test, the opposite face to the shelter was covered with a mirror (15 × 20 cm). All animals were fasted the day before the experiments.

Screening for risk taking in groups (boldness experiment)

Methods were adapted from Huntingford *et al.* 2010 (Huntingford *et al.* 2010). A total number of 280 juvenile fish were taken in groups of 9 and left for 10-min habituation in the sheltered area with the hole closed with a PVC screen, and the top of the sheltered area of the tank also covered to provide a complete shelter. Then, the lid covering the hole was gently removed. Either the first 3 fish to exit the shelter or fish with latency times inferior to 10 min were considered bold fish and were gently removed with a fish net from the test tank to avoid recruitment and provisionally placed in another tank (P subpopulation) with the same size and environmental conditions. Latency times of emergence from the sheltered area were individually recorded. The next 3 animals to emerge before 15 min were considered intermediate and also gently removed. Animals that still remained inside the sheltered area after this screening period were considered shy (R subpopulation). The screening lasted a maximum time of 35 min including the habituation time. Intermediate individuals were discarded, and selected fish were held in separate tanks (P and R populations) for *a posteriori* behavioural tests. A number of 10 randomly selected fish for both phenotypes were immediately killed by an overdose of MS-222, brains were sampled and frozen on liquid nitrogen for transcriptomic analysis.

The same screening test was performed with P and R subpopulations ($n = 36$ of each subpopulation) in adults, 10 months after the first screening for risk taking in groups, to test for consistency and to validate the previous first screening.

A consistent sample of fish screened for risk taking in groups ($n = 10$ for P and R, males and females; $N = 20$) was individually identified by visual external marks and pooled together in the same tank for posterior behavioural tests. The sequence of the behavioural tests was as follows: first week, the individual activity test; second week, the individual MIS tests directly followed by latency to feeding after a confinement event test. After each test, all water from the tank was changed to avoid chemical communication between experimental individuals.

Activity test

Two different activity tests were performed. The first test was an open-field activity test with no shelter. Animals were individually identified and placed in the test arena and left for 10 min to habituate. The test was then started, and animals were recorded with a colour video camera (SONY® SSC-DC378P) mounted over the test arena and connected by fire wire with a computer monitor with a specific software to record and analyse the data simultaneously (SMART 2.3, Panlab) for 5 min. Variables recorded were activity (ACT, number of transitions), total distance (Dist, cm), mean speed (cm/sec) and number of areas used (AR). We also calculated occupancy percentages and index of activity (Sq. Changes/Ttotal=300 sec).

The second test was visually recorded, and a shelter was used. Animals were individually placed in a shelter in an alternate pattern (P and R) and left for 10 min to habituate. Then, the shelter was completely opened by taking the lid off the tank, and the animals were recorded for 5 min for latency to emerge from shelter, total time in shelter (TTSh), activity (ACT, number of squares crossed) and number of squares used in total (SQR). Occupancy percentages and index of activity were also calculated (Sq. Changes/Ttotal=300sec).

Mirror test or mirror-image stimulation (MIS)

A week later, the same 20 fish were tested for mirror-image stimulation (MIS) or aggressive mirror test (Budaev 1998; Budaev *et al.* 1999; Oliveira *et al.* 2005; Archard & Braithwaite 2010). After 10-min habituation to the new tank, the lid that covered the mirror was lifted, and the test was run for 5 min. Variables measured were latency to realize the first approach (LFA), number of bites to the image (Bite), number of aggressive behaviours (AGR: parallel swimming, circles and strikes) and time spent freezing (FRZ: complete lack of movement only for eyes and gills).

Latency to first feeding after a confinement event

Directly after MIS test, the individual fish was caught and held in a submerged net for no more than 1 min,

while the mirror was removed. After this short confinement, the fish was released into the tank and fed at the same time with 3 red worms at the opposite side of the tank. Latency to reassume feeding (Moretz *et al.* 2007) was measured for each individual. Latency to first feeding (LRF) was defined as the time taken to capture the first worm. Additionally, time spent freezing (FRZ) for each individual was recorded for 5 min.

Morphometric analysis

The relationship between a set of defined morphometric variables was analysed, and additionally, the relationship with Fulton's condition index and cephalic index was explored (Sutton *et al.* 2000; Nash *et al.* 2006). Fulton's condition factor is widely used in fish biology studies (Brander 1995; Rätz & Stein 1999; Rätz & Lloret 2003). This factor is calculated from the relationship between the weight of a fish and its length, with the intention of describing the condition of that individual. The formula is $K = W/L^3$ where K = Fulton's condition factor, W = the weight of the fish and L is the length (total length). The cephalic index represents head length/total length. Individual fish were removed from P and R subpopulation holding tanks ($n = 68$/group) and measurements were taken. All weight measurements were taken with an analytical balance (±0.01 mg), and length measurements were made with a regular caliper (±0.05 mm). For statistical analysis, first, the relationship between morphometric measures and personality was tested (*t*-test). Following this, we then calculated the Fulton's condition factor and the cephalic index. Both indices were then compared with personality (*t*-test).

Tissue sampling

Selected animals were killed by an overdose of MS-22, and whole brains were carefully removed. Individual brain samples were homogenized in 0.3 mL of Tri-Reagent (Molecular Research Centre) and stored at −80 °C for further molecular analysis. The experimental protocols used for fish sampling have been reviewed and approved by the Ethics and Animal Welfare Committee of the Universitat Autònoma de Barcelona, Spain.

RNA isolation and microarray hybridization

Total RNA was extracted from individual zebrafish brains ($n = 10$ from each group; $N = 30$ animals in total) using the standard TriReagent-based method following manufacturer's instructions. Total RNA concentration was quantified (Nanodrop ND-1000), and RNA integrity and quality assessed (Bioanalyzer 2100, Agilent Technologies). The RNA integrity number (RIN) was calculated for each sample, and only RNAs with a RIN number >7.5 were processed. Total RNA (1 μg) was used to synthesize cDNA with SuperScript III Transcriptase (Invitrogen) and oligo-dT primers (Promega). Total RNA ($n = 30$ individual brain samples) was labelled and used for microarray hybridization. Microarray hybridizations were performed using the Zebrafish V2 (G2519F) 4x44K Agilent oligonucleotide microarray. Standard methods were used for all processes according to manufacturer's instructions (Agilent Technologies). Briefly, each amplified and labelled sample was hybridized at 65 °C for 17 h. Microarrays were scanned, and one-channel TIFF images (Feature Extraction software version 10.4.0.0) were imported into GeneSpring software (GX 11.0). Microarray data are described in accordance with MIAME guidelines and have been submitted to Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/ GEO Accession: GSE40615). For microarray analysis, standard analytical methods were applied to the data obtained. Briefly, array normalization (percentile shift normalization) was carried out and data filtered by standard deviation expression among groups (filter by expression). Signal intensities for unique probes within a probe set were averaged to obtain an expression value (median) for the probe set (Gene-level analysis). Statistical tests were implemented in the GeneSpring software GX 11.0 used to select transcripts differentially expressed ($P < 0.001$) between control and treatments. One-way ANOVAs were used to identify significant differences between treatments.

qPCR validation

Total RNA (1 μg) was used to synthesize cDNA with Superscript III (Invitrogen) and oligo-dT (Promega) according to manufacturer's instructions. cDNA (1 μL) was used as a template for PCR with specific primers for Gadph, FLI-1, Protocadherin, Crx, Opsin, Hox5a, Hox5c, Grk7 and Arrestin (Fig. S1). Primers were designed with Primer3 version 4.0 based on target sequences obtained from the Agilent database for each mRNA of interest. Products were separated on agarose gels, stained with SybSafe and purified with mini columns (Qiagen). Purified PCR products were ligated in pGEM-Teasy vector (Promega) and transformed in *E. coli* (DH5α strain). One selected transformant of each construct was grown to obtain plasmidic DNA (Miniprep kit, Macherey-Nagel). All constructs were verified by sequencing.

Absolute quantification was performed to validate the microarray expression data. The copy number of

each transcript, derived from the standard dilution curve obtained from target plasmid, was analysed using the MyIQ real-time PCR system (Bio-Rad, CA). Each sample was tested in triplicate in a 96-well plate. The reaction mix (15 μL final volume) consisted of 7.5 μL of EvoGreen mix (Bio-Rad), 0.75 μL of each primer (500 nM final concentration), 2.5 μL of H_2O and 3.75 μL of a 1/100 dilution of the cDNA sample. The thermocycling programme consisted of one hold at 95 °C for 4 min, followed by 40 cycles of 10 s at 95 °C and 45 s at 60 °C. After completion of these cycles, melting-curve data were then collected to verify PCR specificity and the absence of primer dimers.

PCA, Hierarchical clustering and ANOVA analysis

We performed a hierarchical clustering algorithm using the Euclidean distance matrix and complete linkage method. Analyses were conducted in the Cluster3/TreeView open software (Eisen *et al.* 1998; de Hoon *et al.* 2004; Saldanha 2004). Principal component analysis (PCA) from the GO package shows two-dimensional views, retrieved from SPSS 17 statistical software, and was used to visualize the relatedness of all 20 individual microarray samples. To identify the phenotypic transcriptional features of each personality, we conducted both one-way and two-way ANOVAS, in which the expression profile was modelled as a fixed effect and focused on genes that were highly significant ($P < 0.001$) for proactive and reactive individuals. All *P*-values were adjusted with a false discovery rate (FDR) correction for multiple testing by the Benjamini–Hochberg method (Benjamini & Hochberg 1995). All genes with FDR-corrected *P*-values <0.05 were considered significant. The expression of genes found to be significantly different between both personalities were further characterized by a hierarchical clustering analysis. Hierarchical clustering was based upon expression pattern across the sampled population, thereby identifying clusters of genes with common expression profiles. Sample variances were homogeneous (normal distribution).

Gene Ontology (GO-DAVID analysis)

Enrichment of specific gene ontology (GO) terms among the set of probes that are specific to personalities was assessed to correlate a specific set of mRNAs within a brain region. In all GO analyses, Ensembl Gene Identifiers were tested using DAVID Bioinformatics Resources (http://david.abcc.ncifcrf.gov/tools.jsp), (Huang *et al.* 2007; Sherman *et al.* 2007). Enrichment of each GO term was evaluated through use of the Fisher's exact test and corrected for multiple testing with FDR (Benjamini & Hochberg 1995). We applied a Bonferroni correction to account for multiple tests performed. To interpret our data, we used in GO-DAVID: the «Functional Annotation»: the first associated gene ID with a biological term that belongs to one of the 46 annotation anatomical categories available in DAVID. Each gene set comprised of at least 4 transcripts that shared the same GO biological process or annotation term.

Interactome analyses

Visualization of interactions and overlays of expression profiles were carried out using Cytoscape 2.8.2. (http://www.systemsbiology.org). The interactome network was obtained from all interactions with a FBS>6. The interactome backbone contains 5760 nodes (protein–protein and protein–DNA interactions) and 99 573 relationships between these proteins (interactions) (Table S1). The designation of protein properties was drawn from (Alexeyenko *et al.* 2010). NCBI gene name attributes were used to unify the protein list and were imported through the Biomart plugin. The network for the ZF_Personality was built from within the Danio_rerio_CS interactome. Topological analysis of individual and combined networks was performed with Network Analyzer, and jActiveModules 2.2 was used to analyse network characteristics (Montojo *et al.* 2010; Smoot *et al.* 2011). GO analyses were conducted with the Biological Network Gene Ontology (BinGO, version 2.0) plugin (Maere *et al.* 2005) used for statistical evaluation of groups of proteins with respect to the current annotations available at the Gene Ontology Consortium (http://www.geneontology.org). GO over-representation was calculated using the hypergeometric test with Benjamini and Hochberg false discovery rate (FDR) multiple testing correction and significance (pFDR < 0.05). In addition, we conducted a complementary analysis with ClusterMaker cytoscape plugin (Morris *et al.* 2011), using the MCL algorithm to search protein–protein interaction network modules derived from TAP/MAS (tandem affinity purification/mass spectrometry). This approach clustered the network into modules based on PE Score to indicate the strength of the node association and given a fixed set of genes with high protein–protein affinity (interactome cluster nodes).

Intra-individual variance

Agilent Zebrafish V2 (G2519F) (Agilent Technologies) arrays were used to collect individual brain transcriptomes for each selection strategy; proactive, reactive and random ($n = 10$ each category; $N = 30$). The raw gene expression data set was summarized as above. The data

set was filtered as described above, and 2 subsets of 32 761 probes common to coping styles were used for the comparison of intra-individual and personality variance analysis. A subset of significantly differentiated transcripts of both coping styles (nonparametric *t*-test with Benjamini and Hochberg FDR correction and *P*-value cut-off of 0.01) was also used to inspect the intra-individual and interpersonality variance. To gauge the effect of sample size on variance, the interindividual variance was calculated by computing CV distributions based on individuals from each personality group. These CV distributions were then compared between all groups. A CV value was calculated for each detected probe by dividing the standard deviation of its expression by its average group expression (proactive, reactive and random selected group). Low- and high-expression variance genes were identified as those with CV distribution below and above the cut-off 1. The relationship between standard deviation and average expression was examined and tested for significant differences between personalities with analysis of covariance (ANCOVA) using expression average as the covariable.

Results

Screening for risk taking in groups

From initial screening using risk taking in groups ($n = 9$/group) in a novel environment (base population (BP); $n = 206$), we obtained 26% proactive individuals (P) and 51% reactive individuals (R), and the remaining population classified as intermediate were discarded from the analysis. Animals were housed together as P and R subpopulations under the same environmental conditions. A rescreening for risk taking in groups was performed, after 10 months, with a random sample from the P and R populations ($n = 36$ animals in each group) to test for consistency over time and context. Rescreened animals in general were different for group composition with the BP ($\chi^2 = 37.01$; d.f. = 4; $P = 0.0001$). Both P and R subpopulation latencies to exit the shelter were significantly different with the BP ($\chi^2 = 74.86$; d.f. = 2; $P < 0.0001$ and $\chi^2 = 15$; d.f. = 2; $P = 0.0054$), respectively. The rescreening highlights that each subpopulation maintains at high levels the personality trait (bold/shy) selected by risk taking in groups through time and population context (Fig. 1A). Mean group latencies to exit the shelter were also different between BP and P and R (Kruskal–Wallis $H = 49.89$; d.f. = 2; $P < 0.0001$) and between BP and P, BP and R (post hoc Dunn's comparison: $P < 0.001$; $P < 0.0,1$ respectively) reflecting the enrichment of personality in the subpopulations. Habituation to the risk taking in groups test over time was addressed using mean latencies (post hoc Dunn's multiple comparisons), and no significant differences were observed. A significant interaction between personality and time was also identified (Kruskal–Wallis $H = 201.78$, d.f. = 3, $P < 0.0001$).

Activity test

None of the variables analysed were significantly different between proactive and reactive fish in the first activity test. This was probably due to both technical issues resulting from interference with the water, and the software was unable to detect differences and also stress induced by the experimental setup (S. Rey, pers. observation). Furthermore, recordings were in a 2 dimensional space, and some important behavioural patterns, such as vertical movement, were not detected. As a consequence, we decided to make a second activity test with a shelter to detect possible differences on vertical use of space and decrease the stress imposed upon the fish. However, no differences in latency to emerge from shelter, total time in shelter and activity indices were significant between P and R groups. However, SQR used in total, and SQR occupancy percentages and distribution were significantly different (one-way ANOVA, $F_{1,14} = 7$, $P = 0.019$; repeated-measures ANOVA, $F_{4,56} = 7.09$, $P < 0.001$, respectively). A completely different SQR use distribution between the two groups can be observed (Fig. 1B) where P individuals have a more homogenous distribution compared to R individuals with a higher preference for the bottom of the tank and spending more time in the shelter.

Mirror-image stimulation (MIS)

All variables measured: latency to realize the first approach (LFA, one-way ANOVA, $F_{1,18} = 14.240$, $P = 0.00139$**), number of bites to the image (Bite, one-way ANOVA, $F_{1,18} = 10.358$, $P = 0.00477$**) and number of aggressive behaviours (AGR, $F_{1,18} = 75.162$, $P = 0.01341$*) were significantly different between P and R individuals. Results showed that P individuals had a shorter latency to focus the mirror image, to initiate aggressive behaviour and a higher number of bites in comparison with R individuals. FRZ behaviour showed a marked difference between P and R individuals where P individuals minimally express this behaviour (Mann–Whitney *U*-test, $U = 0.00$, $P < 0.001$; see Fig. 1C).

Latency to first feeding after a confinement event

Results showed significant differences in LRF between P and R groups after a confinement event where P individuals always were faster to feed (one-way ANOVA;

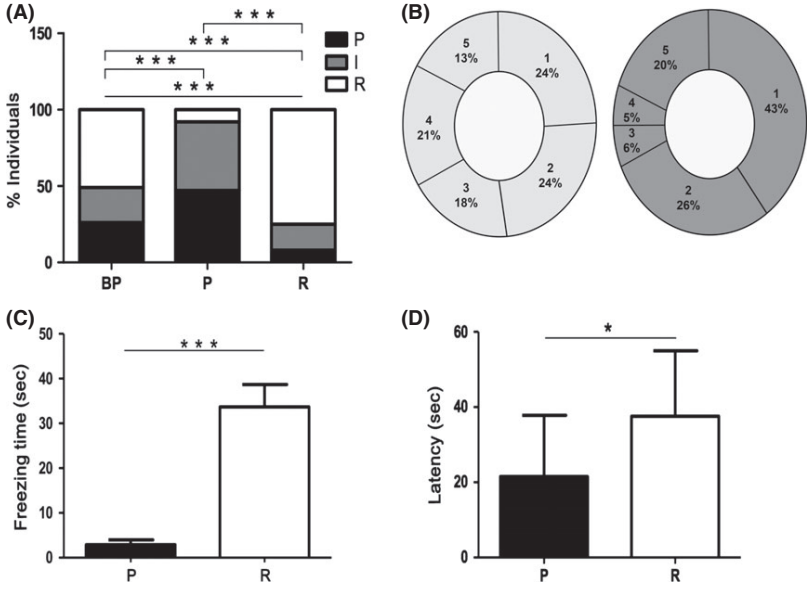

Fig. 1 Behavioural tests for personality analysis in a wild-type zebrafish population (A) Percentage of individuals in each subpopulation after rescreening, 10 months after initial screen, for risk taking in group: BP; base population (mixed), proactive (P) and reactive (R). ***$P < 0.001$. (B) Different occupancy percentages in shelter (segment nº 5) and square use (segments 1-4) between proactive (clear grey, P) and reactive (dark grey, R) individuals (Activity test: repeated-measures ANOVA, $F_{4,56} = 7.09$, $P < 0.001$). (C) Mean (± SE) freezing time for P and R individuals in MIS test (Mann–Whitney *U*-test, $U = 0.00$; ***$P < 0.001$). (D) Mean (± SE) of latency to refeeding after MIS test *$P < 0.05$.

$F_{1,18} = 46.785$, $P = 0.04425$; Fig. 1C, D) than P. Again when measuring FRZ, R individuals always spent more time freezing (Mann–Whitney *U*-test, $U = 27.5$, $P = 0.036$).

Consistency over time and context

In the context of risk taking, this behavioural trait remained consistent in both the groups (P and R) and over time (10 months between screenings) and context (mixed subpopulations) (see above for results, Fig 1A). Furthermore, across contexts, for different behavioural tests, Spearman rank correlation analysis using all the variables recorded across different contexts identified some significant correlations (Mann–Whitney *U*-test. * $P < 0.05$; **$P < 0.01$; Table S1). The strongest positive correlation identified ($r = 0.83$, $P < 0.01$) was between AGR in the MIS test and LRF on feeding measurements on the P group of animals. For R, no significant correlation was found between AGR and LRF. Curiously, a strong negative correlation was found between LFA in the aggression measurements and LRF ($r = -0.7$, $P < 0.05$, Fig 2A and B) both in P and R, where LFA always negatively correlated with all other aggression variables measured and strongly with FRZ in the MIS test ($r = -0.82$, $P < 0.01$) in R fish.

Morphometric comparisons between coping styles

Initial analyses for the measured morphometric variables (8 in total) and personality showed no significant differences. However, when we incorporated both the Fulton's condition factor (K) into the analyses, the total of DEO (diameter of eye orbit) length of P individuals that ranged from 2.20 ± 0.11 mm and R individuals from 2.13 ± 0.12 mm was significant ($P < 0.01$). This was also repeated when using cephalic index (Ci) where values ranged from 0.19 ± 0.013 and 0.18 ± 0.01 for P and R individuals, respectively. Both ratios are significantly smaller in the R population (Fig. 2C).

P and R individual brain transcriptome analyses

Global expression profiles of whole brains from 30 wild-type zebrafish screened *a priori* for personality traits were compared (proactive, reactive and random selection). 43 602 probes scored positively for reliable hybridization signals among all samples. 32 761 probes passed quality control for random effects between single-slide variability, leaving only transcripts that were either present or marginal in all experimental arrays. These 32 761 transcripts were used for statistical analysis to find mRNAs with abundance levels that significantly differed between zebrafish P and R subpopulations. (Table S2). Under constant aquaria conditions, the P and R populations exhibit different global profiles in the brain transcriptome. Both PCA and hierarchical cluster analysis highlighted the differences between the brain transcriptomes where P and R populations could easily be separated, and two major clusters were identified composed entirely of either P or R individuals (Fig 3 A,B). Analysis for mRNAs that significantly differed between brains of P and R individuals (nonparametric *t*-test with Benjamini and Hochberg FDR correction and $P < 0.01$) identified 3027 distinct mRNAs. These mRNAs are listed in the supplementary

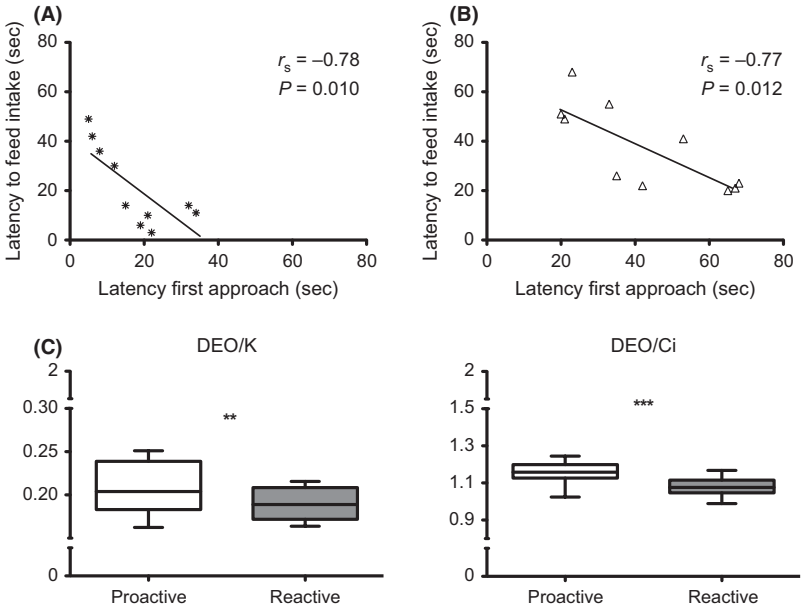

Fig. 2 Consistency of behavioural traits and morphometric analysis in P and R populations (A) Within-individual Spearman rank correlation analysis results ($P < 0.7$) for MIS and refeeding post-stress in proactive individuals and in (B) reactive individuals. (C) Mean ± SD of DEO/K (diameter of eye orbit/Fultons condition factor) and DEO/Ci (cephalic index) of proactive and reactive individuals; $n = 68$ individuals in each population (*t*-test ** $P < 0.01$, *** $P < 0.001$).

data table with their expression levels and functional annotations (Table S3a). Absolute QPCR validation data (Figure S1) show the reproducibility of the array measures for 5 mRNA targets identified in our analysis.

Of particular note from the set of mRNAs identified with differential relative abundance levels, high and low, between personalities related to key cellular processes (Table S3b). In the following section, we highlight some key mRNAs. We observed sets of mRNAs with high relative abundance in R individuals suggesting higher activity levels that were observed to have low abundance in P individuals relevant to growth and proliferation status including: type VII collagen alpha 1, Procollagen-proline, disintegrin, metalloproteinase domain 8, yippee-like 3, cyclin T2, progenitor cell differentiation and proliferation factor b (King *et al.* 2004; Lok *et al.* 2008; Sriraman *et al.* 2008; Oliver-Krasinski & Stoffers 2008; Thangjam *et al.* 2009; Tuttle *et al.* 2011; Umemoto *et al.* 2012; Marchesi *et al.* 2013). On the other hand, cellular trafficking and cellular communication function were observed to have a higher abundance in P individuals including Na+/K+ ATPase beta, calcyclin-binding protein, solute carrier family 25 member 47, catenin beta 1 (Jorgensen *et al.* 2003; Hediger *et al.* 2004; Shtutman *et al.* 2008; Chen *et al.* 2011). Both groups imply increased cellular proliferation associated with increased activity. Furthermore, in the high P/low R transcript set, a significant number of related mRNAs directly involved in regulation of circadian rhythm including Cryptochrome 1b, Cryptochrome 2a, Cryptochrome 3 (Delaunay *et al.* 2003; Ziv *et al.* 2005; Tamai *et al.* 2007) were observed. Interestingly, a correlated set involved in glycolytic metabolism including isovaleryl-coenzyme A dehydrogenase, insulin-like growth factor 1, succinate dehydrogenase (van Raamsdonk *et al.* 1993; Hwa *et al.* 1999; Lemeer *et al.* 2008) were also observed. These diametrically opposed observed abundances in key cellular processes suggest that the underpinning regulation of metabolism, biorhythm/cellular clock and cellular proliferation may be dependent upon personality and reflect physiological status/requirements of the brain in P and R individuals.

In silico analyses of brain regions and personality-dependent mRNA abundance

To identify underlying differences in function and localization of the sets of personality-dependent mRNAs (high and low), we tested the 3027 mRNAs for functional enrichment using GO-DAVID (Table S4). GO-DAVID identifies significant enrichment of functional GO categories and localizes these groups to distinct anatomical regions within the brain including specific cell types. 46 GO terms representing 1459 mRNAs were over-represented (Bonferroni corrected $P < 0.05$). The functions identified included as expected a wide range of biological processes (e.g. cell division, ATP synthesis, cell adhesion, extracellular matrix remodelling) and cell types reflecting the composition of the brain, that is, neurons and glial cells. Although multiple brain regions can be identified, we limited our analysis to the forebrain, midbrain and hindbrain regions to avoid hierarchical classification issues using the primary classification. We selected a total of 21 GO

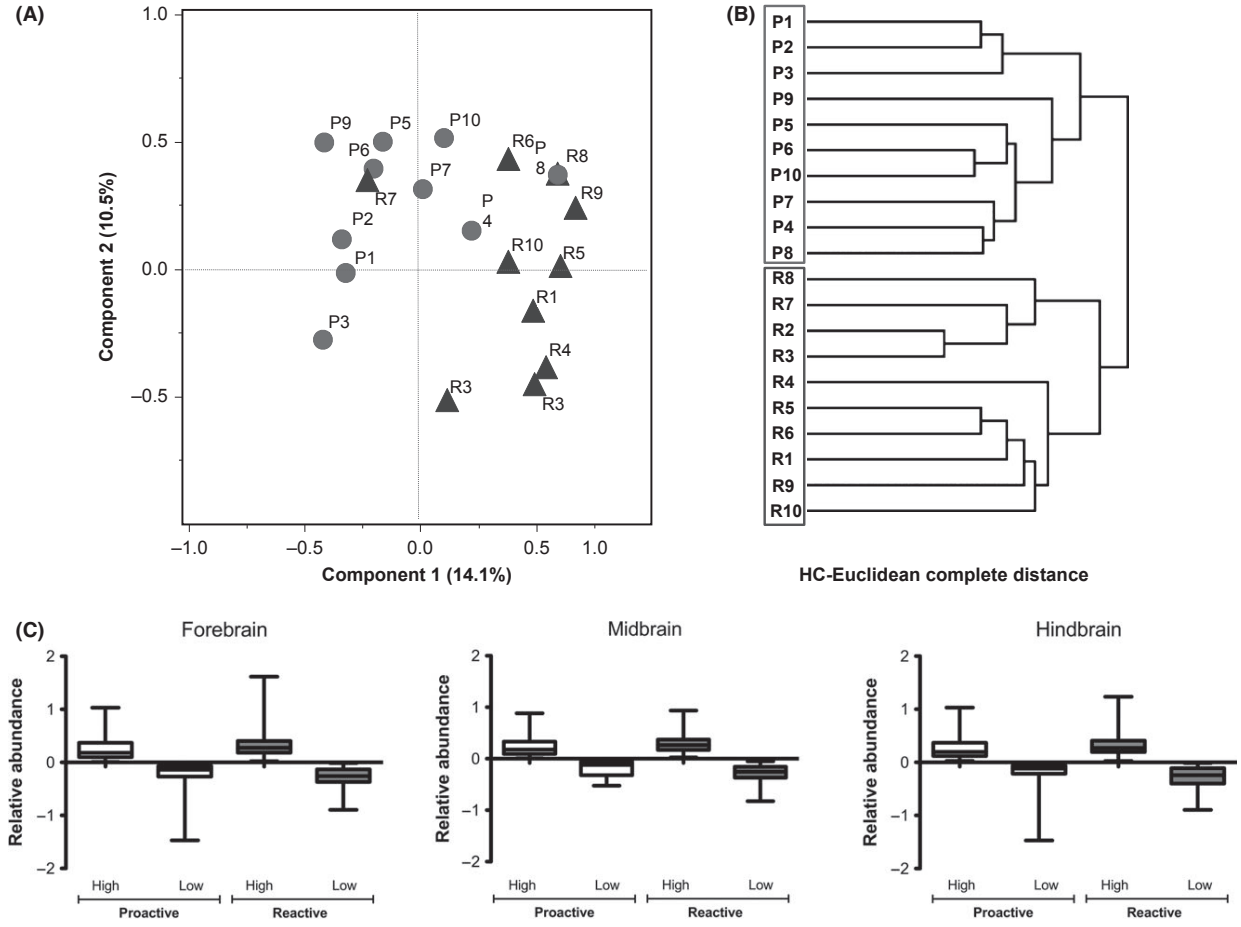

Fig. 3 Clustering and expression variance analysis of P and R brain transcriptomes (A) PCA; Factorial map of the principal component analysis (PCA) of P and R subpopulation transcriptomes. PCA was performed on data from all common expressed transcripts in each transcriptome represented as a single colour point for each personality. The portion of the variance explained by the principal component is indicated in parentheses and (B) Hierarchical clustering (Euclidean distance), sample size, $n = 10$ individuals for each personality. Refer to Table S2 for the annotated list of transcripts. (C) GO-DAVID analysis correlates the specific set of transcripts to gross brain regions (forebrain, midbrain and hindbrain). Sample size, $n = 167$ in forebrain, $n = 121$ in midbrain, $n = 110$ in hindbrain. Box and whiskers plots registered for each brain region ($n = 10$).

modules containing a total of 398 mRNA transcripts that could be allocated to one of the specific selected regions (Table S4). Transcripts that were ubiquitously distributed were removed from the analysis. In the forebrain, we found 9 GO classes containing 167 transcripts, and in the midbrain and hindbrain, we found 7 and 5 GO classes representing 121 and 110 transcripts, respectively. This suggests that there is a personality-dependent enrichment of function in distinct brain regions. Further analysis (interaction between mRNA abundance and personality in specific brain region) highlighted significant interaction for personality and mRNA abundance in the midbrain (two-way ANOVA $F_{3,\ 242} = 3.916$, $P < 0.049$), whereas significant difference in mRNA abundance was observed in forebrain (two-way ANOVA $F_{3,\ 354} = 3.945$, $P < 0.004$), and in the hindbrain, no significant differences were observed (Fig. 3C).

Zebrafish brain interactome

To further explore the functional significance of personality-dependent differences in the brain of P and R individuals, we developed a stringent analysis using selected clusters of transcripts that share a high protein–protein affinity (interactome analysis). For interactome-based analysis, we used the global network of functional coupling for zebrafish and combined these data with our microarray data set to create 3 personality-specific interactomes of each brain region using the Cytoscape platform. We identified network regions (subportions of the

full network) enriched in differentially expressed, interconnected mRNAs (nodes) in the forebrain, midbrain and hindbrain. In addition to further substantiate our models, we incorporated a protein–protein interaction network derived from tandem affinity purification/mass spectrometry (TAP/MAS) [ClusterMaker plugin, (Morris *et al*. 2011)]. This approach clustered modules (a fixed set of genes with high protein–protein affinity) in the network by the strength of their node association. We selected the modules with the highest representation of node–node interactions that were found in each brain region and conducted GO analysis with the BinGO plugin (Maere *et al*. 2005). Three distinct predicted biological outcomes were obtained (Fig. 4). As suggested and supporting the previous analysis, modules with significant differences were identified in the fore-, mid- and hindbrain interactomes (Table S4). This analysis, by identifying region-specific processes, further supports the observation that personality directly affects key cellular processes in zebrafish brain.

Personality as a resolving variable for variance in gene expression

The high level of observed natural variance in gene expression data sets obtained within populations is a major limitation for biological interpretation of results. In this section, we used personality as a resolving variable to reduce variance and therefore increase the biological value of data obtained in large-scale measurement of gene expression. As a measure of variance, we used the coefficient of variation (CV) that is computed for each transcript (microarray hybridization signal; raw data) by dividing the standard deviation of its

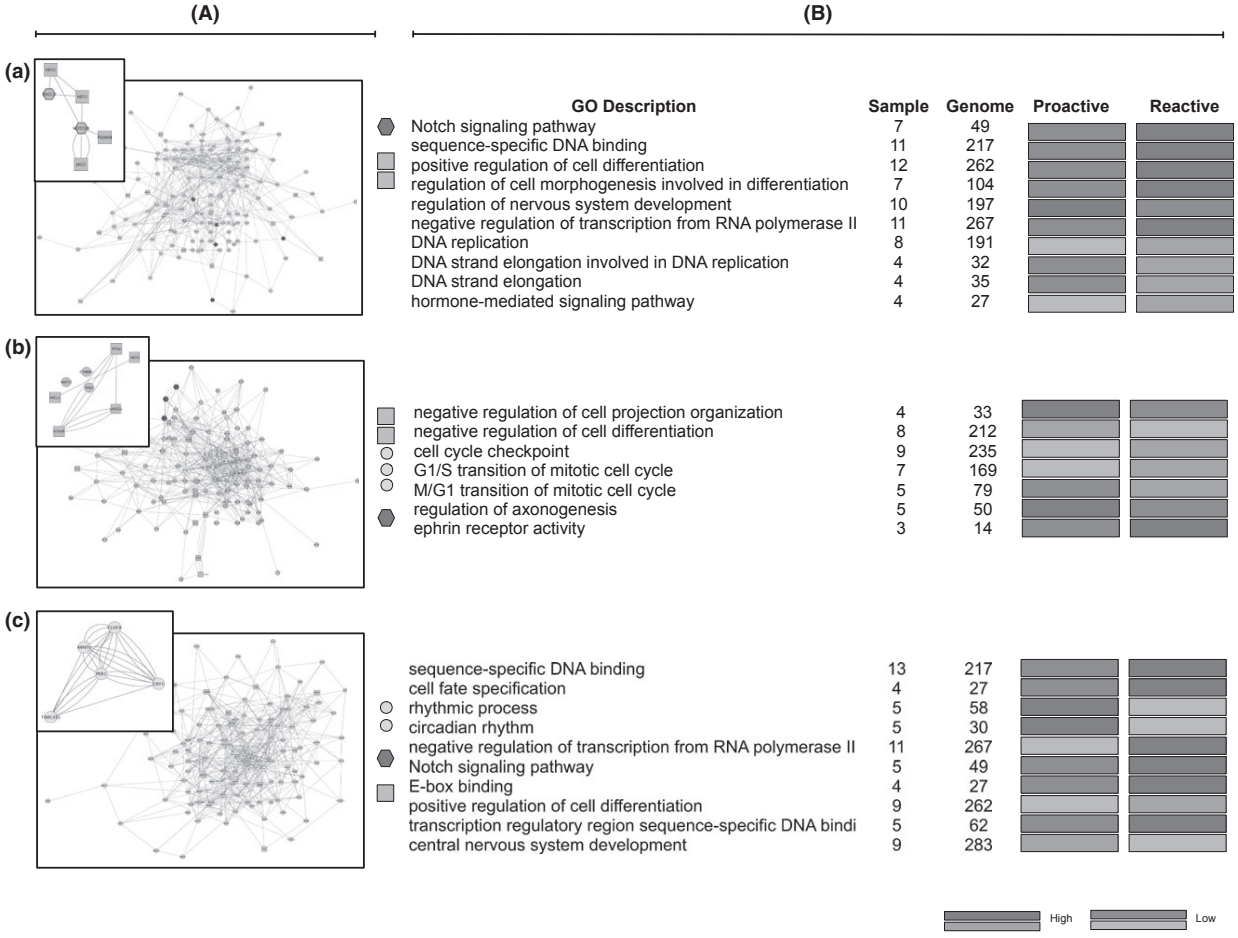

Fig. 4 Interactome mapping of personality-related gene networks in the zebrafish brain (A) Interactome modules of mRNAs expressed in each brain region based in MCL algorithm (ClusterMaker plugin) a; forebrain, b; midbrain, c; hindbrain. The inserts represent selected significant protein–protein interaction network modules expressed in the 3 brain regions. ($P < 0.01$). The symbols represent specific GO processes. (B) Gene ontology analysis (BinGO plugin) of each interactome-module of overexpressed GO categories ($P < 0.01$), the colour scale bar indicates relative abundance (high-low) of GO categories in each personality. Sample size, $n = 398$ in forebrain, $n = 121$ in midbrain, $n = 110$ in hindbrain.

measured abundance across the population by its mean expression value. We designated a threshold value for transcripts with a high variance as those falling above a CV distribution of >1 and low as <1. This analysis using CV values protects against detecting patterns in variability influenced by trends in relative expression based upon total mean values. We used both raw (all positive hybridization signals) and selected (personality-dependent) data sets for the P, R and the control populations (Fig. 5A, B). The distribution of expression values was consistent across both approaches; raw data (32 761 probes) and personality-dependent (3027 probes), which more probably reflects the high level of reproducibility of the expression data. The absolute numbers of transcripts in low and high variance categories varied slightly, CV_{error} of <2% between personalities when we used raw data (Fig 5A); however, in contrast, when selected transcripts are used, differences are significant where P and R data had CVs that were significantly smaller than the random selected group (one-way ANOVA $F_{2,\ 9075} = 5.074$ $P < 0.05$; Fig. 5B). Thus, we identified that the personality of the individual significantly contributes to the observed biological variation within the population. We then calculated the observed variance for each of our data sets (ANCOVA) and expressed this as a ratio for individual transcriptome versus mean population transcriptome. The slope of the regression was homogeneous (Table S5), and after adjusting for raw data, the ANCOVA revealed significant differences between personalities with respect to both standard deviation and mRNA abundance in the personality-dependent data set (Table S5). The variance for P and R individual transcriptomes was significantly smaller than the control (Fig. 5C). The observed effect on variation in the transcriptome attributable to each variable was 9.7% for personality and 0.2% for gene expression (ANCOVA on adjusted standard deviation followed by Tukey test: $\alpha = 0.05$; $N = 3.026$ probes Table S5).

Discussion

Resolving phenotype variation within a population in response to environmental perturbation is central to understanding adaptation and aims to bridge the genotype–phenotype gap. Such bridging of the 'genotype–phenotype gap' represents a truly transversal question spanning from biomedicine to behavioural ecology. Unprecedented opportunities to explore the phenotype–genotype gap from distinct biological perspectives have been unlocked by the genomics and system biology disciplines. Genomics driven 'big data' has provided an incomparable set of resources for biologists, and the main focus, highlighted in humans by initiatives such as HapMap (International HapMap Consortium *et al.* 2007), ENCODE (Birney *et al.* 2007) and the 1000 Genomes Project (Clarke *et al.* 2012), is now upon how nucleotide sequences drive specific traits and the genetic basis of phenotypic variation. In parallel to this exciting progress, the measurement of gene expression (functional genomics) has been greatly facilitated by the increasing availability of transcriptomic technologies including microarray and RNA-Seq methodologies. Gene expression is a complex trait that is influenced by cis- and trans-acting genetic and epigenetic variation

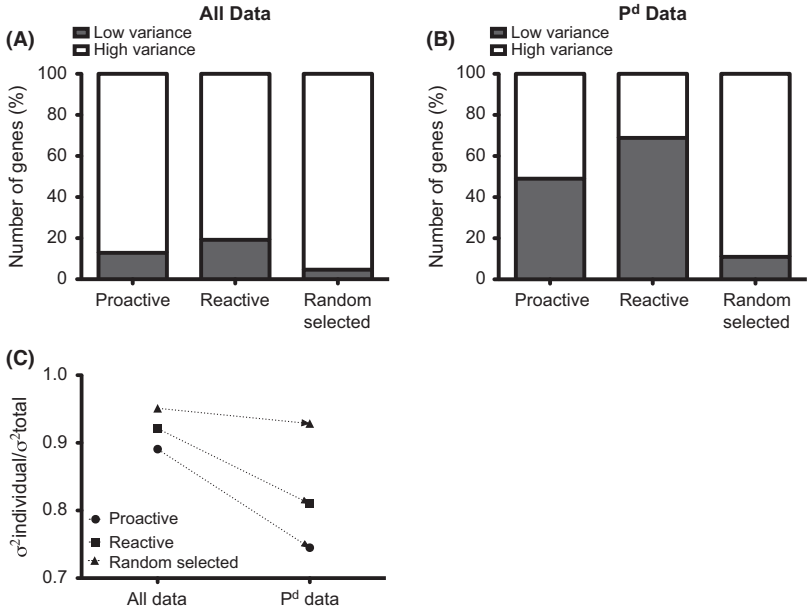

Fig. 5 Animal personality partially discriminates expression variance in transcriptomes. (A) Whole-brain transcriptome (all data, $n = 32.761$) coefficients of variation for individual transcripts for proactive, reactive and random (base population) selected individuals, CV distribution was set as >1 high and low as <1. (B) Personality-dependent (P^d, $n = 3.027$) coefficients of variation for proactive, reactive and random (base population) selected individuals, CV distribution was set as >1 high and low as <1. (C) Observed variance for all data and P^d data sets (ANCOVA) expressed as a ratio of individual transcriptome versus mean population transcriptome.

and also by environmental factors. Relating meaningful adaptive changes at the level of the transcriptome requires the identification processes that have a functional significance for the individual. This remains a major objective towards understanding the complex interactions between environmental demand and an individual's capacity to respond to such demands. The interpretation of such interactions and the significance of biological variation between individuals from the same or different populations remain a difficult and underaddressed question (Whitehead & Crawford 2006b). Although such studies have tremendous potential for unravelling in broad terms how patterns of gene expression change, questions have been raised about how differences in the transcriptome are related to adaptive phenotypic variation (Whitehead & Crawford 2006a). Although the concept of 'variation' is well accepted throughout the biological sciences, we remain in a paradigm that in biological systems, important differences between groups, identified by specific measureable parameters, are defined by significant average differences. However, this view, that is, reduction of expression heterogeneity has been challenged on numerous occasions and is counterintuitive if the aim is to understand adaptive phenotypic variation (Raser & O'Shea 2005; Rockman & Kruglyak 2006; Leek & Storey 2007; Mar *et al.* 2011). Variation in gene expression may arise due to multiple additive factors including cis- and trans-acting genetic variation, polygenic interaction, epigenetic modification and environmental factors. A significant body of evidence using a number of molecular approaches highlights that interindividual variation in gene expression within a population is common throughout the vertebrates (Stamatoyannopoulos 2004) and in humans may contribute to differences in disease susceptibility and other disorders. This variation is considered to represent a major source for phenotypic variation and has been shown to be tissue and tissue-region specific in both fish (Whitehead & Crawford 2005; Crawford & Oleksiak 2007) and mammals (Sandberg *et al.* 2000; Bray *et al.* 2003). A range of estimations of the actual contribution to variation from different animal models runs approximately from 10 to 30% (Stamatoyannopoulos 2004). It is accepted that much of this variation is the result of *cis-acting* influences on gene expression where allele-specific polymorphism in cis-regulatory regions is more probably a major determinant (Pastinen 2010). Furthermore, evidence now points towards variance as a key element in information flow across gene networks, and this has been suggested to be a major driver for disease susceptibility (Mar *et al.* 2011). It has been proposed that epigenetic variation via variable methylated regions (VMRs) plays a key role in disease susceptibility within a population and a predictive model highlighted affected evolutionary fitness (Feinberg & Irizarry 2010). Thus, to understand the underlying mechanisms for adaptation, it is necessary to address phenotypic variation within a population.

It is becoming increasingly recognized in the field of behavioural ecology that consistent interindividual variation in behaviour, known as animal personality, may represent an important source of phenotypic variation contributing to the adaptive potential within a population (Dingemanse *et al.* 2010; Dall *et al.* 2012). Animal personality is considered as a suite of behaviours that are consistent over time and context and coping style as a set of physiological and behavioural responses to external stressors (Dingemanse *et al.* 2010). In essence, both personality and coping style are directly related representing studies at different scales of biological research to the same question; understanding adaptation. Therefore, in this study, we use animal personality following consensus opinion in the current literature (Dall *et al.* 2012). Our results screened for animal personality using risk taking in groups of zebrafish (*Danio rerio*) as an a priori screen generates in comparison with individual studies more robust data (Sih *et al.* 2004; MacKenzie *et al.* 2009; Huntingford *et al.* 2010). Individual tests, especially involving isolation, provoked stressful physiological responses, that is, freezing over extended periods and contribute to observed variation in our intra-individual tests. Low correlations between tests at the individual level and low sample groups may favour artifactual bimodal distributions (Budaev *et al.* 1999; Budaev 2010). Such procedures are commonly used in personality tests including activity, exploration, novel environment, novel object or confinement and stressors resulting from these tests such as an inadequate environmental context may confound results (Killen *et al.* 2013). Risk-taking tests performed in groups over an extended time period (10 months) that incorporates normal environmental variation in the aquaria highlighted a consistent response over both developmental stage and social context as the P and R subpopulations were maintained separately over this period (both P and R were significantly different on latencies to exit the shelter, and only 8% of P fish were found in the R group after 10 months). Thus, bold–shy behaviour remained stable over time and under different social contexts. After rescreening for risk taking, individuals from P and R subgroups ($n = 20$; 10 animals of each group) were tested for activity, aggression and feeding recovery after stress. Overall consistency was mostly significant, and a group syndrome was detected. Animals classified as P in group screening tests were more active, distributed homogeneously throughout the tank, performed more aggressively in the MIS test and reassumed feeding faster after a

stressful event. On the contrary, R fish spend more time either in shelter or in freezing behaviours, preferentially occupied the bottom of the tank and were less aggressive and slower to eat after a stressful event. Aggressive behaviour, latency to first approach in the MIS test and latency to begin feeding after confinement provided the best correlation in P individuals, within the subpopulations. We were also able to correlate individually some behavioural measures on the tests performed. We detected, in proactive fish, a feeding and aggression syndrome: animals with shorter latency to reassume feeding (LRF) were strongly correlated ($r = 0.83$, $P < 0.05$) with general aggressive behaviours but negatively with LFA ($r = -0.78$, $P < 0.05$) Surprisingly, first focused attack seems to be a measure that correlates negatively with all other aggression variables measured (number of bites: $r = -0.60$ and aggressive behaviours: $r = -0.83$). R showed similar positive results in aggression and feeding tests with animals that resumed feeding faster spending less time freezing ($r = 0.72$) and the contrary. Thus, indicating higher levels of stress-related behaviours in these R individuals probably constrained their responses in other contexts (Martins *et al.* 2011). Also, LFA correlated negatively with LRF in R animals where first latencies were 10 times higher than the P fish. Limitations in the analysis more probably due to small sample size were detected, but basically, the consistency of observed behaviour across this suite of tests and contexts produced results similar to those described for proactive–reactive personalities (Wilson *et al.* 1994, Koolhas *et al.* 1999, Réale and Festa-Bianchet, 2003; Korte *et al.* 2005) for mammals (including humans) and other vertebrate groups including several other fish species (Felicity Ann Huntingford 1976, Øverli *et al.* 2007). Some contradictory results reflected the impossibility to decide upon the validity of the data where only the comparisons between the groups of individuals were clearly different, whereas within-individual comparisons in different contexts were not different. This observation is similar to that recently reported by Garamszegi *et al.* (2013). Therefore, we used individuals from the original P and R populations initially screened for risk taking in groups ($n = 10$ per personality and 10 controls; $N = 30$) to carry out individual brain transcriptomic analyses under stable environmental conditions to probe phenotypic variation within our zebrafish population.

The 9.7% reduction in variance that corresponds to the cohort of 3026 distinct mRNAs that are personality-dependent are expected to be directly related to underlying differences in genome regulation within the population. Taking into consideration the limitations of microarray analyses, that is, the influence of allelic variation, it is more probably a considerable portion of this observed dependence is related to both cis- and trans-acting determinants of gene expression and epistatic interactions. Recently, bold–shy behaviour was shown to have a strong heritable component in zebrafish (Oswald *et al.* 2013; Ariyomo *et al.* 2013), and we produced similar data (Vargas *et al.*, unpublished data). Thus, a significant genetic component can be expected although as all fish cohorts were reared in a common environment, we cannot discard the possibility that epigenetic regulation during development may be different between distinct personalities, and this may contribute to the observed differences in the brain transcriptome (Feinberg & Irizarry 2010; Killen *et al.* 2013). In this study, we have shown that there are significant differences in mRNA abundance related to both function, as described by interacting gene modules, for example, notch pathway, cellular clock and cellular proliferation/differentiation, and location, distinct brain regions, where specific functions are controlled within the brain and a portion of the observed variation is personality-dependent. Differences in brain region-specific gene expression has been shown to contribute to differences in behavioural profiles between different populations of in-bred mouse strains (Sandberg *et al.* 2000). Importantly, this is the first time that strong correlations between functional modules containing significant protein–protein affinities and personality have been made in individuals sourced from a wild-type genetic background. Interestingly, differences in the physiological functions identified by our analyses have been independently reported to contribute differences in animal personality (Careau *et al.* 2008; Huntingford *et al.* 2010). This encouraging support brings forward the possibility that a significant contribution to animal personality potentially through cis- and trans-driven regulation of gene expression is identifiable in our model. Although a significant portion of variation in expression heterogeneity can be resolved using a priori personality screening, the complexity of confounding factors such as epistasis, pleiotropic and environmental influence must not be ignored (Bray *et al.* 2003; Elmer & Meyer 2011).

Our data supported at different levels of analysis suggest that the inclusion of behavioural screening prior to a molecular analysis has the potential to identify important interactions that are more probably highly relevant towards understanding the underpinning molecular framework of adaptive variation that is expressed as an individuals' phenotype. Such interactions cannot be identified by a standard average mean approach within or between populations; thus, the biological significance of measured observations is more probably weakened or may in fact be misleading. In conclusion, the use of a priori screening with robust models of animal behaviour such as personality in a

coherent environmental context may provide a map towards understanding the underlying complexity of adaptive phenotypic variation in the genomic landscape.

Acknowledgements

This study was supported by the Consolider-Ingenio Programme 2010, project CSD2007-0002 funded by the Spanish Ministry of Science and Education, Spain to SM. NR (Ramon y Cajal Program) is funded by the Ministerio de Ciencia e Innovación, Spain. RV was funded by the Government of Panama as part of his PhD thesis.

References

Alexeyenko A, Wassenberg DM, Lobenhofer EK *et al.* (2010) Dynamic zebrafish interactome reveals transcriptional mechanisms of dioxin toxicity. *PLoS ONE*, **5**, e10465.

Archard GA, Braithwaite VA (2010) The importance of wild populations in studies of animal temperament. *Journal of Zoology*, **281**, 149–160.

Ariyomo TO, Carter M, Watt PJ (2013) Heritability of boldness and aggressiveness in the zebrafish. *Behavior genetics*, **43**, 161–167.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, **57**, 289–300.

Birney E, Stamatoyannopoulos JA, Dutta A *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.

Brander KM (1995) The effect of temperature on growth of Atlantic cod (Gadus morhua L.). *ICES Journal of Marine Science: Journal du Conseil*, **52**, 1–10.

Bray NJ, Buckland PR, Owen MJ, O'Donovan MC (2003) Cis-acting variation in the expression of a high proportion of genes in human brain. *Human genetics*, **113**, 149–153.

Budaev DS (1998) How many dimensions are needed to describe temperament in animals: a factor reanalysis of two data sets. *International Journal of Comparative Psychology*, **11**, 17–29.

Budaev SV (2010) Using principal components and factor analysis in animal behaviour research: caveats and guidelines. *Ethology*, **116**, 472–480.

Budaev SV, Zworykin DD, Mochek AD (1999) Consistency of individual differences in behaviour of the lion-headed cichlid, Steatocranus casuarius. *Behavioural Processes*, **48**, 49–55.

Careau V, Thomas D, Humphries MM, Réale D (2008) Energy metabolism and animal personality. *Oikos*, **117**, 641–653.

Chen X, Mo P, Li X *et al.* (2011) CacyBP/SIP protein promotes proliferation and G1/S transition of human pancreatic cancer cells. *Molecular Carcinogenesis*, **50**, 804–810.

Clarke L, Zheng-Bradley X, Smith R *et al.* (2012) The 1000 genomes project: data management and community acces. *Nature Methods*, **9**, 459–462.

Crawford DL, Oleksiak MF (2007) The biological importance of measuring individual variation. *Journal of Experimental Biology*, **210**, 1613–1621.

Dall SRX, Bell AM, Bolnick DI, Ratnieks FLW (2012) An evolutionary ecology of individual differences. *Ecology letters*, **15**, 1189–1198.

Delaunay F, Thisse C, Thisse B, Laudet V (2003) Differential regulation of Period 2 and Period 3 expression during development of the zebrafish circadian clock. *Gene Expression Patterns*, **3**, 319–324.

Dingemanse NJ, Kazem AJN, Réale D, Wright J (2010) Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, **25**, 81–89.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 14863–14868.

Elmer KR, Meyer A (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution*, **26**, 298–306.

Feinberg AP, Irizarry RA (2010) Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proceedings of the National Academy of Sciences*, **107** (Suppl 1), 1757–1764.

Felicity Ann Huntingford (1976) The relationship between anti-predator behaviour and aggression among conspecifics in the three-spined stickleback, Gasterosteus aculeatus. *Animal Behavior*, **24**, 245–260.

Garamszegi LZ, Marko G, Herczeg G (2013) A meta-analysis of correlated behaviors with implications for behavioral syndromes: Relationships between particular behavioral traits. *Behavioral Ecology*, **24**, 1068–1080.

Goetz FW, MacKenzie S (2008) Functional genomics with microarrays in fish biology and fisheries. *Fish and Fisheries*, **9**, 378–395.

Hediger MA, Romero MF, Peng J-B *et al.* (2004) The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflugers Archiv European Journal of Physiology*, **447**, 465–468.

de Hoon MJL, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics (Oxford, England)*, **20**, 1453–1454.

Huang DW, Sherman BT, Tan Q *et al.* (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Research*, **35**, W169–W175.

Huntingford FA, Andrew G, MacKenzie S *et al.* (2010) Coping strategies in a strongly schooling fish, the common carp Cyprinus carpio. *Journal of Fish Biology*, **76**, 1576–1591.

Hwa V, Oh Y, Rosenfeld RG (1999) The insulin-like growth factor-binding protein (IGFBP) superfamily. *Endocrine reviews*, **20**, 761–787.

International HapMap Consortium, Rosenfeld RG, Frazer KA *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.

Jorgensen PL, Håkansson KO, Karlish SJD (2003) Structure and mechanism of Na, K-ATPASE: functional sites and their interactions. *Annual Review of Physiology*, **65**, 817–849.

Killen SS, Marras S, Metcalfe NB, McKenzie DJ, Domenici P (2013) Environmental stressors alter relationships between physiology and behavior. *Trends in Ecology & Evolution*, **28**, 651–658.

King NE, Zimmermann N, Pope SM *et al.* (2004) Expression and regulation of a disintegrin and metalloproteinase (ADAM) 8 in experimental asthma. *American Journal of Respiratory Cell and Molecular Biology*, **31**, 257–265.

Koolhaas JM, Korte SM, Boer D *et al.* (1999) Coping styles in animals: current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*, **23**, 925–935.

Korte SM, Koolhaas JM, Wingfield JC, McEwen BS (2005) The Darwinian concept of stress: benefits of allostasis and costs of allostatic load and the trade-offs in health and disease. *Neuroscience & Biobehavioral Reviews*, **29**, 3–38.

Leek JT, Storey JD (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genetics*, **3**, 1724–1735.

Lemeer S, Jopling C, Gouw J *et al.* (2008) Comparative phosphoproteomics of zebrafish Fyn/Yes morpholino knockdown embryos. *Molecular & cellular proteomic*, **7**, 2176–2187.

Lok C-N, Ehrlich HP, White SL *et al.* (2008) Oligodeoxynucleotide decoy therapy blocks type 1 procollagen transcription and the prolyl hydroxylase β subunit translation. *Journal of Cellular Biochemistry*, **103**, 1066–1075.

MacKenzie S, Ribas L, Pilarczyk M *et al.* (2009) Screening for coping style increases the power of gene expression studies. *PLoS ONE*, **4**, e5314.

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)*, **21**, 3448–3449.

Mar JC, Matigian NA, Mackay-Sim A *et al.* (2011) Variance of gene expression identifies altered network constraints in neurological disease. *PLoS Genetics*, **7**, e1002207.

Marchesi I, Nieddu V, Caracciolo V *et al.* (2013) Activation and function of murine Cyclin T2A and Cyclin T2B during skeletal muscle differentiation. *Journal of Cellular Biochemistry*, **114**, 728–734.

Martins CIM, Silva PIM, Conceição LEC *et al.* (2011) Linking fearfulness and coping styles in fish. *PLoS ONE*, **6**, e28084.

Montojo J, Zuberi K, Rodriguez H *et al.* (2010) GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics (Oxford, England)*, **26**, 2927–2928.

Moretz JA, Martins EP, Robison BD (2007) Behavioral syndromes and the evolution of correlated behavior in zebrafish. *Behavioral Ecology*, **18**, 556–562.

Morris JH, Apeltsin L, Newman AM *et al.* (2011) clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*, **12**, 436.

Nash RD, Valencia AH, Geffen AJ (2006) The origin of Fulton's condition factor—setting the record straight. *Fisheries*, **31**, 236–238.

Oliveira RF, Carneiro LA, Canário AVM (2005) Behavioural endocrinology: no hormonal response in tied fights. *Nature*, **437**, 207–208.

Oliver-Krasinski JM, Stoffers DA (2008) On the origin of the cell. *Genes & Development*, **22**, 1998–2021.

Oswald ME, Singer M, Robison BD (2013) The quantitative genetic architecture of the bold-shy continuum in zebrafish, Danio rerio. *PLoS ONE*, **8**, e68828.

Øverli Ø, Sørensen C, Pulman KGT *et al.* (2007) Evolutionary background for stress-coping styles: relationships between physiological, behavioral, and cognitive traits in non-mammalian vertebrates. *Neuroscience & Biobehavioral Reviews*, **31**, 396–412.

Pastinen T (2010) Genome-wide allele-specific analysis: insights into regulatory variation. *Nature Reviews Genetics*, **11**, 533–538.

van Raamsdonk W, Smit-Onel MJ, Diegenbach PC (1993) Metabolic profiles of white and red-intermediate spinal motoneurons in the zebrafish. *Acta histochemica*, **95**, 129–138.

Raser JM, O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. *Science*, **309**, 2010–2013.

Rätz H-J, Lloret J (2003) Variation in fish condition between Atlantic cod (Gadus morhua) stocks, the effect on their productivity and management implications. *Fisheries Research*, **60**, 369–380.

Rätz HJ, Stein M (1999) Variation in Growth and Recruitment of Atlantic Cod (Gadus morhua) off greenland during the second half of the 20th Century. *Journal of Northwest Atlantic Fishery Science*, **25**, 161–170.

Réale D, Festa-bianchet M (2003) Predator-induced natural selection on temperament in bighorn ewes. *Animal Behaviour*, **65**, 463–470.

Rockman MV, Kruglyak L (2006) Genetics of global gene expression. *Nature Reviews Genetics*, **7**, 862–872.

Ruiz Gomez M, Kittlesen S, Hoglund E *et al.* (2008) Behavioral plasticity in rainbow trout (Oncorhynchus mykiss) with divergent coping styles: when doves become hawks. *Hormones and Behavior*, **54**, 534–538.

Saldanha AJ (2004) Java Treeview–extensible visualization of microarray data. *Bioinformatics (Oxford, England)*, **20**, 3246–3248.

Sandberg R, Yasuda R, Pankratz DG *et al.* (2000) Regional and strain-specific gene expression mapping in the adult mouse brain. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 11038–11043.

Sherman BT, Huang D, Tan Q *et al.* (2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics*, **8**, 426.

Shtutman M, Chausovsky A, Prager-Khoutorsky M *et al.* (2008) Signaling function of alpha-catenin in microtubule regulation. *Cell cycle (Georgetown, Tex.)*, **7**, 2377–2383.

Sih A, Bell A, Johnson JC (2004) Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*, **19**, 372–378.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics (Oxford, England)*, **27**, 431–432.

Sriraman V, Eichenlaub-Ritter U, Bartsch JW *et al.* (2008) Regulated expression of ADAM8 (a disintegrin and metalloprotease domain 8) in the mouse ovary: evidence for a regulatory role of luteinizing hormone, progesterone receptor, and epidermal growth factor-like growth factors. *Biology of reproduction*, **78**, 1038–1048.

Stamatoyannopoulos JA (2004) The genomics of gene expression. *Genomics*, **84**, 449–457.

Sutton SG, Bult TP, Haedrich RL (2000) Relationships among fat weight, body weight, water weight, and condition factors in wild Atlantic salmon parr. *Transactions of the American Fisheries Society*, **129**, 527–538.

Tamai TK, Young LC, Whitmore D (2007) Light signaling to the zebrafish circadian clock by Cryptochrome 1a. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 14712–14717.

Thangjam GS, Agarwal P, Balapure AK, Girish Rao S, Kondaiah P (2009) Regulation of extracellular matrix genes by arecoline in primary gingival fibroblasts requires epithelial factors. *Journal of Periodontal Research*, **44**, 736–743.

Tuttle R, Miller KR, Maiorano JN *et al.* (2011) Novel senescence associated gene, YPEL3, is repressed by estrogen in ER+

mammary tumor cells and required for tamoxifen-induced cellular senescence. *International Journal of Cancer*, **130**, 2291–2299.

Umemoto H, Akiyama M, Domon T *et al.* (2012) Type VII collagen deficiency causes defective tooth enamel formation due to poor differentiation of ameloblasts. *The American Journal of Pathology*, **181**, 1659–1671.

Whitehead A, Crawford DL (2005) Variation in tissue-specific gene expression among natural populations. *Genome Biology*, **6**, R13.

Whitehead A, Crawford DL (2006a) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*, **15**, 1197–1211.

Whitehead A, Crawford DL (2006b) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 5425–5430.

Wilson DS, Clark AB, Clark A, Coleman K, Dearstyne T (1994) Shyness and boldness in humans and other animals. *Trends in Ecology & Evolution*, **9**, 442–446.

Wolf M, van Doorn GS, Leimar O, Weissing FJ (2007) Life-history trade-offs favour the evolution of animal personalities. *Nature*, **447**, 581–584.

Ziv L, Levkovitz S, Toyama R, Falcon J, Gothilf Y (2005) Functional development of the zebrafish pineal gland: light-induced expression of period2 is required for onset of the circadian clock. *Journal of Neuroendocrinology*, **17**, 314–320.

SB, SR, RV and SM participated in the experimental design and all experiments. All authors participated in drafting the manuscript.

Data accessibility

Transcriptomic data: GEO; http://www.ncbi.nlm.nih.gov/geo/; GEO Accession: GSE40615: Individual behaviour, morphometric and QPCR validation data: Dryad doi:10.5061/dryad.b75t4.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 A. QPCR validation for Gadph, FLI-1, Protocadherin, Crx, Opsin, Hox5a, Hox5c, Grk7 and Arrestin mRNAs in P and R individuals ($n = 3$). Data show both array-derived data (black boxes) and absolute quantification (copy number; open boxes) of the 9 mRNA targets measured. Data are show as mean ± SD. B. Primer sequences for mRNA targets.

Table S1 Spearman correlation coefficients across different contexts for A) proactive and B) reactive individuals. (*$P < 0.05$; **$P < 0.01$).

Table S2 Whole transcriptome data (Gene ID, description, expression and regulation) from the three different experimental groups ($n = 32.761$); 1 proactive, 2 reactive, 3 random selected (base population) individuals.

Tables S3 A) Transcriptome data showing regulated transcripts that were specific to P and R subpopulations ($n = 3.027$).All data show Gene ID, description, raw expression and *P*-value. B) Gene ID, GO and relative abundance of 47 selected mRNA transcripts.

Tables S4 Transcriptome data showing regulated transcripts specific to P and R subpopulations found in each gross brain region.

Tables S5 Results of ANCOVA for common slope of regression and adjusted means examining differences in the variance of the expression data across P and R personalities in zebrafish (proactive, reactive and random selected; $n = 10$ individuals of each population).