Title

Improving anxiety research: novel approach to reveal trait anxiety through summary measures of multiple states

Zoltán K Varga*[,#], Diána Pejtsik*, Máté Tóth, Zoltán Balogh, Manó Aliczki, László Szente, Gyula Y Balla, Levente Kontra, Zsófia Eckert, Zsolt Borhegyi, Éva Mikics[#]

*equal contribution

[#]corresponding authors

Abstract

The reliability and validity of preclinical anxiety testing is essential for translational research outcomes. However, widely-used anxiety tests lack inter-test correlations and have repeatability difficulties that need clarification. Translational research seeks to capture individual variability and advance personalized medicine, which demands trait-like features reflecting the underlying neural characteristics. Here, we show that detailed sampling across multiple time-points and contexts covers various states of the individuals, which is needed to reliably capture trait anxiety (TA). We also propose a validated, optimized test battery to reveal TA in rats and mice. Instead of developing novel tests, we combined widely-used tests (elevated plus-maze, open field, light-dark box) to clarify current inter-test and repeatability issues and provide instantly applicable adjustments for better predictive validity. We repeated tests three times to capture multiple anxiety states in various paradigms that we combined to generate summary measures (SuMs). Using correlations and machine learning, we found that our approach resolves correlation issues and provides better predictions for subsequent outcomes under anxiogenic conditions or fear conditioning. Moreover, SuMs were more sensitive to detect anxiety differences in an etiological model of social isolation. Finally, we tested our sampling method's efficiency in discovering anxiety-related molecular pathways through RNA sequencing of the medial prefrontal cortex. We identified four times more molecular correlates of anxiety using SuMs, which pointed out functional gene clusters that had not emerged using single measurements applied by most studies. Overall, temporally stable SuMs are necessary to capture trait-like anxiety in rodents, providing better predictions for potential therapeutic targets.

Introduction

Most of us have experienced anxiety during our lives. The tendency to have such incidents stems from our individually characteristic, stable trait anxiety (TA)[1]. High levels of TA predict a high frequency of anxious states (SA) in different situations and are the determining core of anxiety disorders[2,3]. As the common root of the autonomic, emotional and behavioral manifestation of transient SAs, TA represents the primary psychiatric intervention target[4].

In pathological conditions of anxiety, anxiolytic therapy is ineffective in 4 out of 10 clinical cases, leaving anxiety disorders the most prevalent and burdening mental illnesses[5]. Prescribed compounds act through brain mechanisms that have been targeted for at least 30 years with moderate success, and new candidates are rare[6]. Unfortunately, animal models, our primary source of information on the molecular basis of anxiety, offer targets that fail in clinical trials in more than 60% of the cases[7]. Such alarming rates of failed medicine candidates at clinical and preclinical levels suggest that our approach to investigating anxiety is in a crisis and needs fundamental improvement.

This translational gap potentially stems from the fact that anti-anxiety therapies are considered satisfactory if they control the permanent TA, however, current animal tests only detect the fluctuating SA[4,8,9]. Consequently, we can only expect these tests to highlight markers and treatment targets of transient symptoms rather than drivers of the underlying phenomenon. Results from the most commonly used[7] rodent paradigms, the elevated plus-maze (EPM)[10,11], the light-dark (LD)[12,13] and the open field (OF)[14,15] tests, support this idea. Behavioral parameters measured in these tests are not correlated with each other[4,16] and are hardly reproducible[17–19], which indicates a lack of a shared basis for these, such as TA. Alternative approaches investigate TA using selectively bred highly anxious animal strains[20–22]. Unfortunately, these approaches i) are only suitable for the examination of TA of that specific strain, ii) their molecular profile is not necessarily in connection with the anxious phenotype and iii) are based on the low-validity tests mentioned above. Consequently, despite the alarming flaws of current methods, and the urgent need for a proper one, the field lacks an adequate tool to measure the most relevant subject of anxiety research: TA.

To bridge this gap, we developed an approach to measure TA that is based on refined sampling and analysis of conventional anxiety tests. We hypothesize that the behavior we measure in these tests is the direct consequence of SAs, which emerge as the interaction of the individually stable TA and the current environmental context (Fig.1 – hypothesis). Assuming this, improved sampling of SAs converge to TA. Our approach uses the conventional anxiety tests but expands the classical sampling method in time and depth to represent multiple states of our subjects (Fig.1 – sampling design) by measuring i) more variables, ii) involving serial testing events and even iii) more test types in our analysis. Particularly, we sampled SA in rodents with different sequences of the most frequently used anxiety tests, where each test was repeated three times with all animals. We measured a range of variables (single measures, SiMs), assessed the possibility of their summarisation, and created summary measures (SuMs) by averaging SiMs of serial tests and composite measures (COMPs) by averaging SiMs of different test types (Fig.1 – analysis design). In medical research, creating SuMs of serial sampling is one effective method for describing the background construct of fluctuating measures while decreasing environmental noise and avoiding over-parametisation[23]. Despite the advantages of such an approach, the field of behavioral neuroscience almost entirely misses using it. We validated SuMs as measures of TA with multiple novel approaches. i) Since TA is

considered the shared basis of different anxiety tests, we hypothesized that SuM-based inter-test correlations would be stronger than the classical SiM-based ones. ii) Since TA is stable across time and contexts, we hypothesized that SuMs better predict acute stress-induced future responses, such as behavior in anxiety or fear-triggering environments. iii) Since high levels of TA is a core symptom of anxiety disorders, we hypothesized that SuMs are more sensitive markers of chronic stress-induced anxiety in an etiological stress model. Following these validation steps, we examined the sensitivity of SuMs in discovering molecular markers and potential therapeutic candidates for anxiety by conducting an RNAseq experiment in the anxiety-mediator area medial prefrontal cortex (mPFC) of resting state animals.

Figure 1

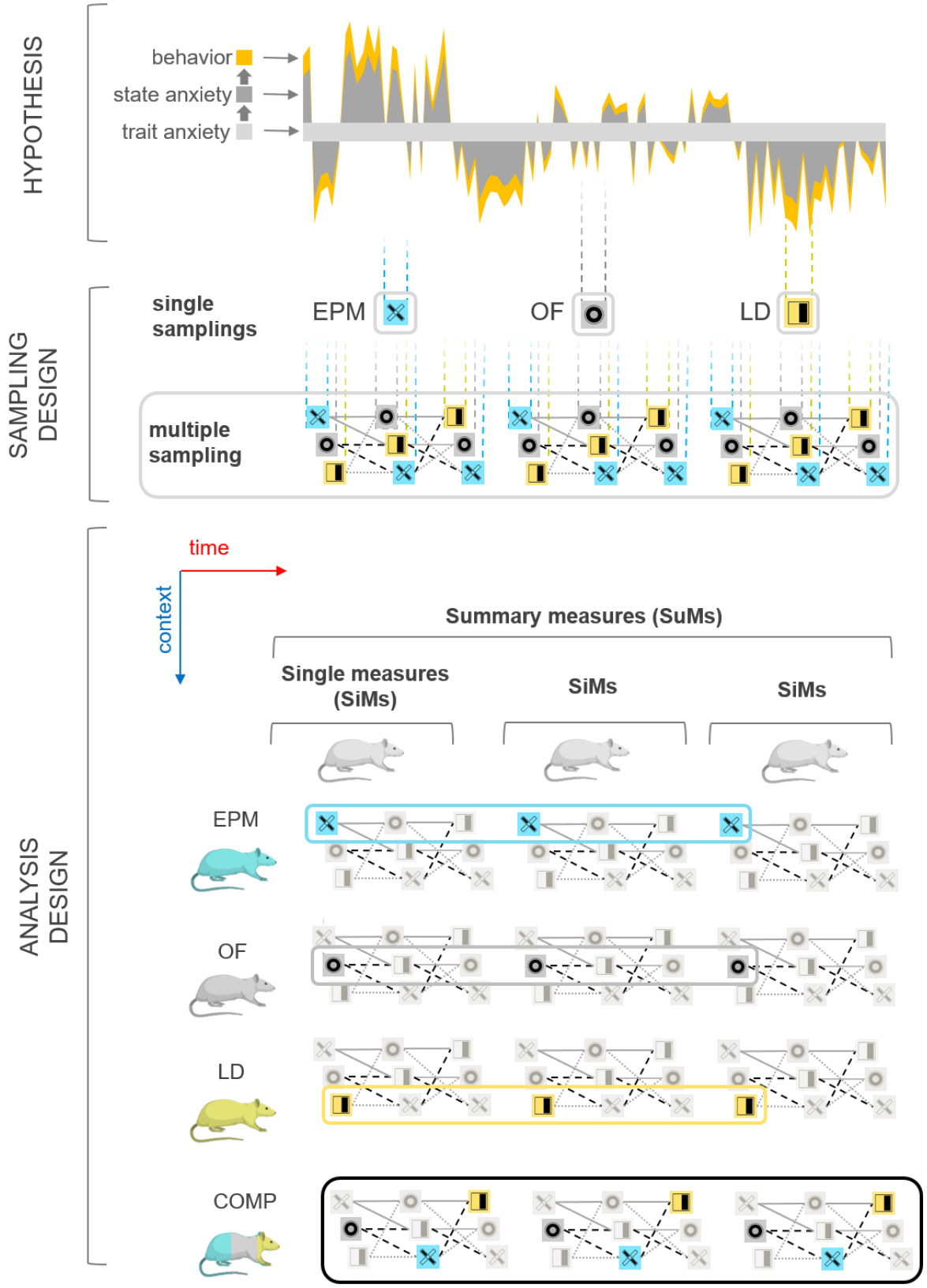

Figure 1. Hypothesis, sampling, and analysis design. Hypothesis: Schematic drawing at the top shows how an individual's behavior is expressed as a consequence of state and trait anxiety according to our working hypothesis. Measured anxiety-like behavior is an indirect consequence of trait anxiety and is influenced by time and context, hence using a single anxiety test leads to undersampled experiments. **Sampling design**: we used repeated sampling

with multiple test types in a semi-random order to cover more internal and external variability stemming from the individual's trait, state, environmental context, and their fluctuation in time. **Analysis design**: Behavioural data from each test or their composite score (COMP) were used to create classical single measures (SiM) and improved summary measures (SuM), by combining multiple variables, testing events, or test types to describe trait anxiety. **Abbreviations**: EPM: elevated plus-maze test, OF: open field test, LD: light-dark test, COMP: composite score.

Materials and Methods

Animals. Detailed methods are provided in the Supplementary materials. Six cohorts of animals were used throughout this study. Two cohorts of adult male Wistar rats (Charles-River, n=47) were used to examine trait anxiety and its neural correlates. To assess replicability, behavioral assessment was repeated with adult female Wistar rats (Charles-River, n=30), and adult male C57BL/6J mice (Jackson Laboratory, n=42). One cohort of adult male Long-Evans rats (Charles-River, n=47) was used to explore the correlations between acoustic startle responses, generalized fear behavior, and trait anxiety in the conditioned fear paradigm (CFP). Additionally, one cohort of adult male Wistar rats (Charles-River, n=30) was used to investigate the effects of social isolation on trait anxiety.

Experimental design. Wistar rats were tested according to a semi-randomized testing protocol for 3+2 weeks (Fig.1). The test battery involved a 3-week behavioral sampling with the most commonly used anxiety tests; the Elevated Plus-Maze[10,11] (EPM), Open Field[12,13] (OF), and Light-Dark test[14,15] (LD), each repeated three times for all individuals. Test order and time-of-day biases were eliminated by creating six different test combinations through semi-randomization of test orders, with an equal number of animals assigned to each combination. After the initial 3-week testing period, animals underwent an additional week of testing in a more aversive experimental condition within a brightly lit test arena. Two days after the aversive week, a different type of open field test (OF2) was performed as the final test for all animals. For the social isolation protocol, upon weaning at P21, animals were reared in groups of 4 (social), or alone in a cage (isolated) until testing in adulthood. Female rats, male rats in the social isolation cohort and mice underwent the 3-week baseline anxiety testing, and the additional week of testing in aversive conditions. Long-Evans rats first underwent a 3-week anxiety testing protocol, followed by an acoustic startle paradigm 3 days later, and finally a fear conditioning paradigm (CFP) a week after that. For anxiety tests, time spent in the aversive zone, frequency, and latency of aversive zone entries were analyzed automatically using Noldus EthoVision XT15[41], head-dips (HD) and rearing (R) were analyzed manually using Solomon Coder[42].

Behavioural measures. To attenuate the possibly state anxiety-driven fluctuation of behavioral variables across repeated testing, therefore to potentially detect stable traits, we introduced summary measures of behavior. Our enhancements to the conventional preclinical approach involved two key aspects: 1) Using the outcomes of the principal component analysis shown in Figure 2C, we calculated z-scores by summing variables that share the same dimension. Before summation, we scaled these variables by subtracting the population mean from each individual's values, then dividing by the standard deviation. We refer to the scaled behavioral variables characterized in the first tests (EPM, OF, LD performed during the first week of testing) as single measures (SiMs). 2) By averaging the scaled variables of the 2-, or 3-time repetition of each anxiety test type, we created summary measures (SuMs). To assess the effectiveness of SuMs in capturing the common underlying factors across different anxiety tests, we compared the strength of individual-level correlations using SiMs versus SuMs. Stronger inter-test correlations indicate that a variable better captures the shared background, potentially reflecting trait anxiety. Composite z-scores (COMP) were created by averaging the SiMs or SuMs of different tests.

Statistical analysis was done in R statistical environment[43]. Principal component analysis ('factoextra' package[44]) of all analyzed behavioral variables from the first week of testing was performed to investigate the shared variance between different behavioral variables. For correlations, Spearman correlation analysis ('Hmisc' package[45]) was used. The analysis of test-repeatability ('rptr' package[46]) was done for each test-type as described previously[47], by calculating the proportion of within-test variance divided by total variance[48]. We applied the random forest machine learning model ('randomForest' package)[49] to identify significant predictors of generalized fear in the CFP.

Blood corticosterone analysis of Wistar rats was carried out on baseline and stress-induced tail vein blood samples. The quantification of plasma corticosterone was carried out using radioimmunoassay similarly to previous work by our laboratory[50].

RNASeq analysis. Microdissection of bilateral medial prefrontal cortex (mPFC) tissue was performed 2 weeks after the last experiment (OF2). Whole transcriptome-analysis via RNASeq was performed on the homogenized mPFC tissue. RNA sequencing was performed on 27 animals, chosen by their COMP scores (Fig.5A). Isolation and mRNA library preparation was carried out by the Department of Biochemistry and Molecular Biology at University of Debrecen, using the NEBNext Ultra II RNA Library Prep with Purification Beads Kit (New England Biolabs) according to the manufacturer's protocol. Single end RNA sequencing was performed on an Illumina NextSeq500 sequencer. Bioinformatics analysis of RNASeq data was done in R statistical environment. Regression analysis to reveal differentially expressed genes (DEGs) was done using the R Bioconductor package DESeq2[51]. Analysis of gene-behavior regressions was carried out using a negative binomial generalized linear model, with anxiety-scales generated from behavior in different test-types (EPM/OF/LD) across test repetitions (SiM/SuM) as continuous covariates in the design. False discovery rate (FDR) correction was done using the Independent Hypothesis Weighting method[52]. Functional and pathway analysis of detected DEGs was done through the online functional annotation tool DAVID[53,54]. Connections were visualized using Cytoscape[55]. Cluster-determining functions, merging and associated genes are shown in Supplementary table 2.

Results

Overlaps between conventional anxiety-like variables support the use of SuMs

If variables carry significantly but not completely overlapping information, SuMs potentially represent their shared background while describing further context-specific information. Here, we assessed the level of overlap between EPM, LD, and OF test variables by repeatability (RA) and correlation analysis in rats and mice. We described repeatability with test-retest changes, within and between-test variability, and bootstrap analysis-based repeatability measures. Test-retest decrease was evident in the open-arm activity of EPM, but not in the center and light activity of the OF and LD tests in male and female rats (Fig.2A). Both OF and LD variables showed smaller between-test variances, while the latter also possessed high within-test variability (i.e. sensitivity for individual differences), making LD the most reliable test of the three (Fig.2B upper). Bootstrap analysis-based repeatability measures confirmed such ranking of the tests (Fig.2B lower). These trends were similar in mice, except that OF centrum activity was also slightly decreased in response to repeated testing (Supplementary fig1). Variables of the same test type showed similar repeatability and clustered together in both species. However, frequency variables were the most, while latency variables were the least reproducible variables in most of the cases, indicating a similarity between the three tests. PCA revealed further conformities, as variables of all tests were loaded in the same dimension (Fig.2C). In summary, there are test type and variable type similarities of these variables which are consistent in the investigated species. This indicates a shared but differently expressed basis of these anxiety tests and potential in the summarisation of their variables.

Figure 2

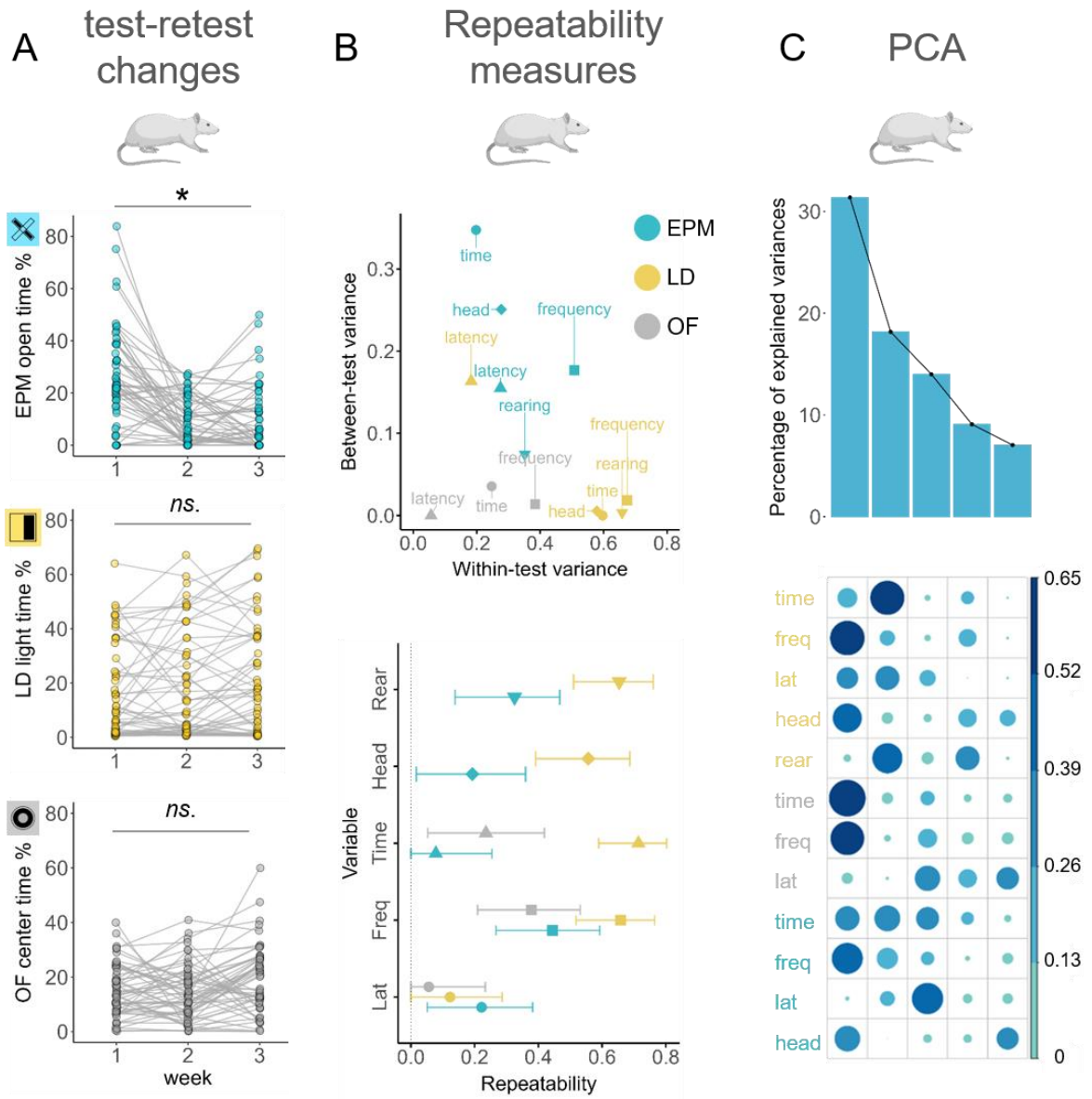

Figure 2. Shared and unique features of anxiety tests in rats. **A)** The effect of test repetitions on the most frequently used anxiety-like behavioral outcome (% of time spent in the aversive compartment) of the EPM, LD, and OF anxiety tests from top to bottom, respectively. **B) top:** all measured variables in variable space defined by a between-test / within-test variance. **bottom:** bootstrap analysis-based repeatability scores of variables with 95% confidence intervals (represented by error bars). Variables with confidence intervals not including a 0 repeatability score are considered repeatable. **C) top:** percentage of explained variance by the first 5 principal components (PC) of a PCA analysis on a single (first) anxiety test. **bottom:** variable representations on each PC. Size and color hue is proportional to the level of representation of a given loading. **Abbreviations:** EPM: elevated plus-maze test, OF: open field test, LD: light/dark test, Lat: latency to the aversive compartment (AC)(sec), freq: enter frequencies to AC (count), time: %time spent in AC, head: head dipping from the platform, rear: rearing behavior. All statistical results are presented in Table 2.

SuMs reveal shared variance of anxiety tests

Since variables of the same anxiety test were similar in PCA and RA, we averaged more of these from multiple testing events into SuMs to describe TA. As TA is the permanent basis of anxiety-like responses[2,3], TA markers should correlate between different tests. To test this, we created differentially complex SuMs of the EPM, LD, and OF tests and measured their inter-test correlations. Our least complex (SiM) variables were the time spent in, enter frequency or latency to the aversive compartment of an anxiety test sampled only one time. Our most complex SuMs consist of all of these variables sampled three times. We constructed all possible variables along this scale and described their complexity with a standard score (=number of variables x number of testing events). We found that SuMs of high complexity correlated stronger between tests than less complex SuMs or the SiMs in both rats and mice (Fig.3A,3B)(Table 1). According to our model, inter-test correlations can be explained by variable complexity as a fixed effect, and variable and test composition as interacting random effects. As Figure 3B shows, the weakest correlations come from the least complex SuMs, and SuMs that include latencies, while the strongest ones consist of more complex and/or frequency and time-included data. Given that, we excluded latencies from further analysis and used reduced models with time and frequency data which also outperformed the classical SiM variables. The association between variable complexity and inter-test correlations is apparent in our rat and mouse models, indicating the robustness of our approach.

Figure 3

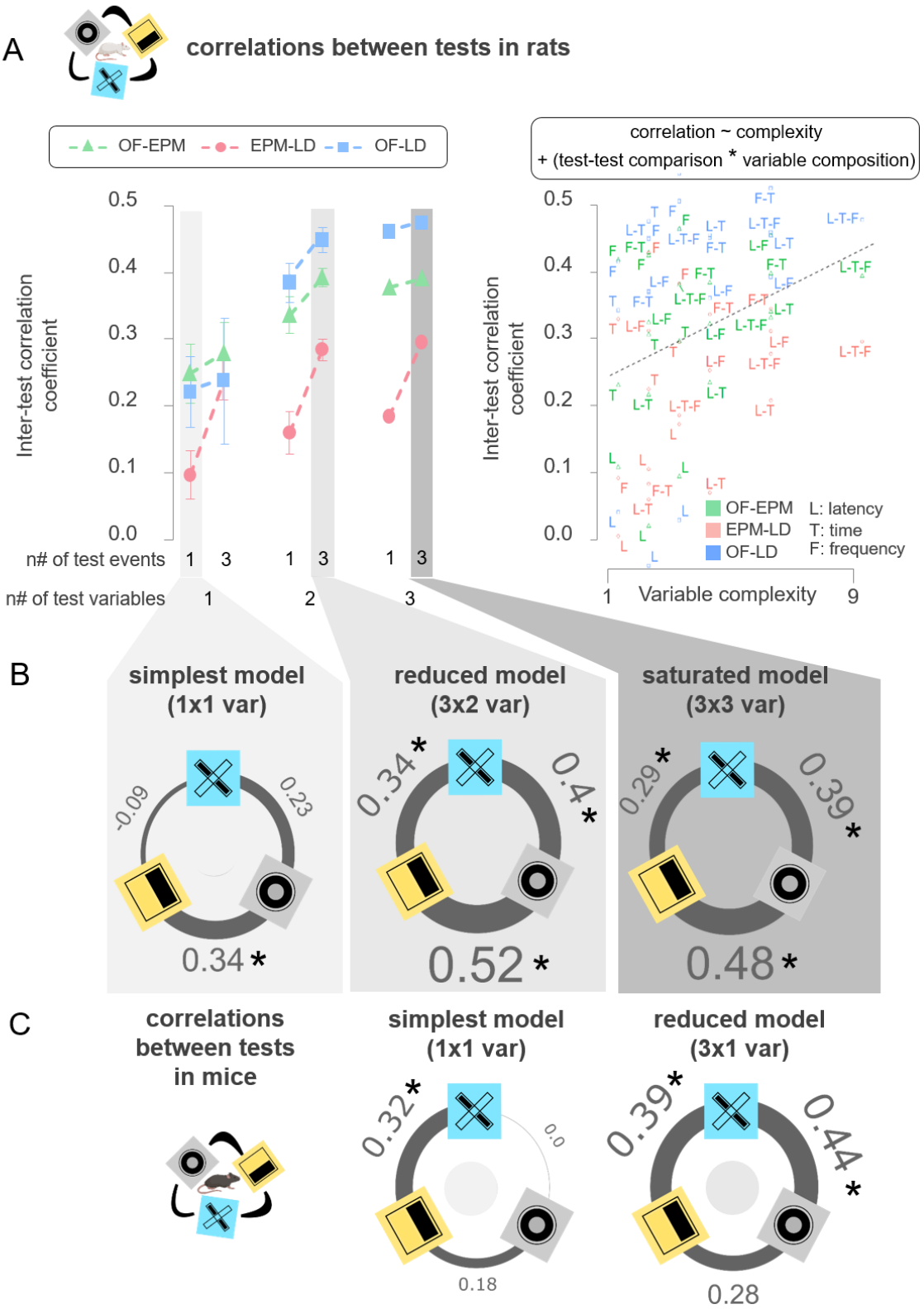

Figure 3. Inter-test correlations as a function of anxiety-measure complexity. **A)** Inter-test correlations between increasingly complex anxiety variables – SiMs and SuMs – were plotted in response to the number of test types, events, and variables that were included in each measure. According to our final statistical model, inter-test correlations were significantly influenced by variable complexity as a fixed effect, and variable composition and test composition as interacting random effects. The color code indicates different test-test comparisons, while the

abbreviations indicate which variables were used to compose a SuM. L: latency, T: time, F: frequency. **B-C)** network plot of inter-test correlations between different tests revealed by SiMs (simplest model) and SuMs (reduced and saturated models) in the rat and mouse experiments. All statistical results are presented in Table 2.

SuMs predict acute stress-induced future behavior

As TA is the basis of all anxiety-like responses, its markers should remain consistent through time and context. To test this, we introduced rats and mice to anxiety tests in an aversive context (Fig.4A), or rats to fear conditioning (CFP), or acoustic startle paradigms (Fig.4E) following our original baseline sampling. In aversive sampling, we detected enhanced anxiety-like behavior (Fig.4B) and corticosterone response compared to baseline conditions (Fig.4C). Correlation analysis between test-specific SiMs and SuMs versus their aversive counterparts revealed that more complex variables predict future behavioral responses better in both species (Fig.4D). Furthermore, creating composite scores from SiMs and SuMs of all tests (COMP) outperformed test-specific SiMs and SuMs, respectively. Our approach increased the strength of behavioral predictions by twofold, reaching a 0.78 correlation between baseline and aversive responses. Note that the predictions of the summarised data from multiple testing events (SuMs) are higher than the predictions of any of our SiMs (1st, $2^{nd,}$ or 3rd testing) indicating that the performance of SuMs is not due to a carry-over effect of repeated testing but these variables explain additional information. Likewise, the predictions of the summarised data from multiple tests (COMPs) are higher than the predictions of single tests which indicates that these variables also carry additional information to any test type that we used here. We confirmed these findings in female rats and male mice, showing a similar association between the number of sampling events and their predictive power (Table1).

Next, we aimed to predict the amount of freezing behavior in a safe context following CFP as a measure of generalized fear, the core symptom of post-traumatic stress disorder. Anxiety tests showed poor performance in predicting fear, but similarly to previous results, with increasing variable complexity, we observed an increase in correlations (Fig.4G)(Table 1). Despite such a trend, only COMPs correlated significantly with later fear responses. To find key predictors of fear responses we conducted Random Forest classification and reduced our saturated model to fewer variables (rCOMP)(sup.Fig.2). rCOMP classified subjects into high and low freezer groups with a 0.78 prediction accuracy and showed a 0.4 correlation with freezing (Supplementary Fig.2)(Fig.4G). The magnitude of acoustic startle responses could not be predicted by either variable, regardless of the sampling approach. These results indicate that SuMs and their composites can predict specific future states of the animals entering challenging environments.

Figure 4

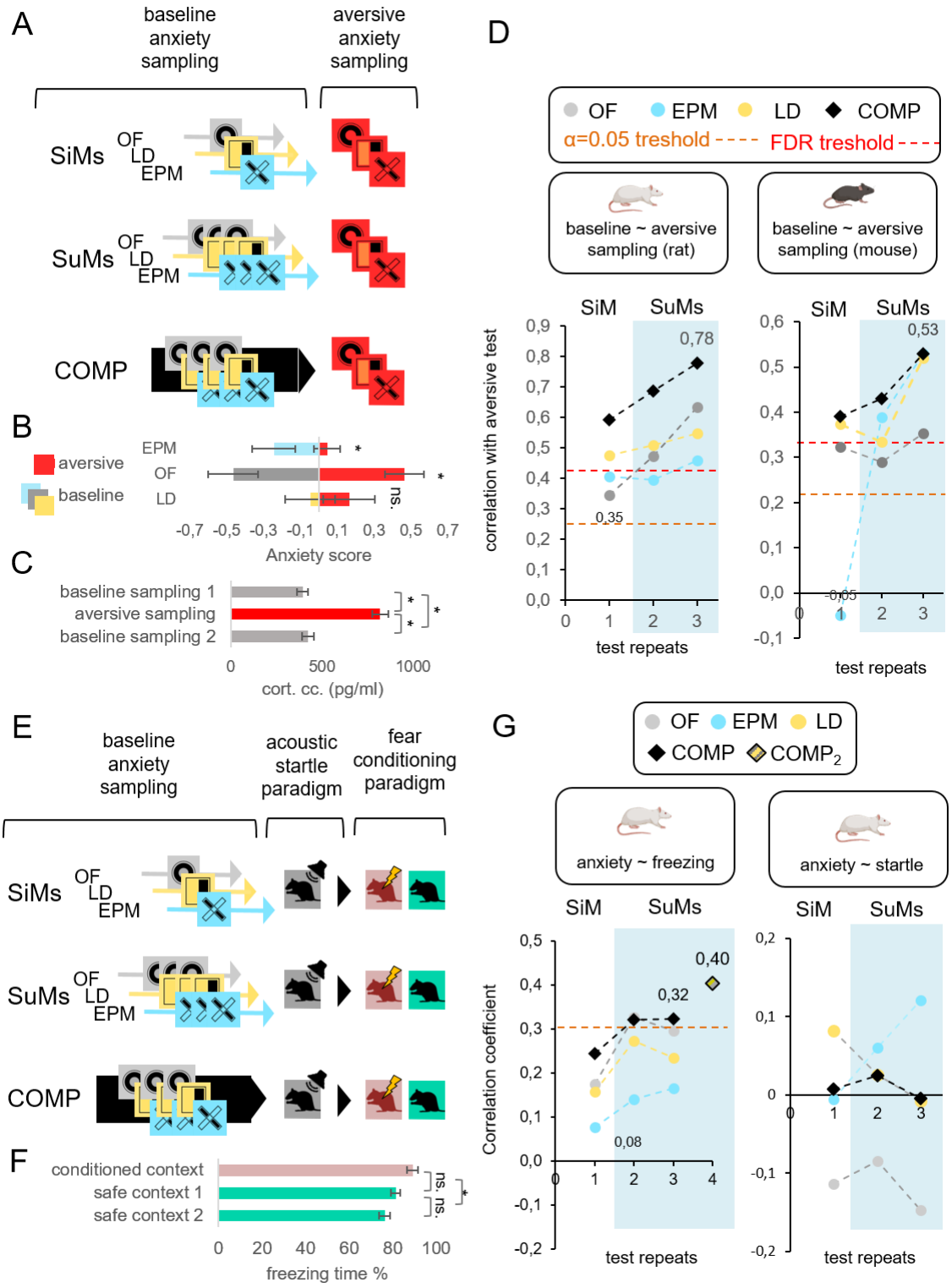

Figure 4. Behavioral predictions of anxiety measures. **A)** correlation coefficients were measured between identically composed variables in baseline and aversive contexts. SiMs represent one, and SuMs represent more testing events, while COMPs include SiMs or SuMs of all test types. **B)** Anxiety score from the last baseline and the aversive sampling. **C)** Blood plasma corticosterone levels in resting state before (1) and after (2) the test battery, and immediately following aversive anxiety sampling. **D)** Correlations between differently complex anxiety measures (SiMs, SuMs, or COMPs) from the baseline sampling and their aversive counterparts in rats (left) and

mice (right). **E)** Correlation of differently complex anxiety measures (SiMs, SuMs, or COMPs) with freezing behavior in novel context following fear conditioning (left) and startle response to acoustic stimuli (right). COMP represents the composite of all anxiety tests and measures, while rCOMP represents the reduced version of the former that was optimized by a Random Forest machine learning approach. Red dashed lines represent the threshold of significance on the scale of correlation coefficients. Additional numbers on the plots show the smallest and largest coefficients. **F)** Time spent with freezing (%) in the conditioned fear paradigm in the contextual reminder (conditioned) and safe (different) contexts. Freezing was sampled two times in the safe context, from which an average freezing value was used to be predicted. All statistical results are presented in Table 2.

			Correlation coefficients (rho)													p-values												
	sampling		1				1-2				1-2-3					1				1-2				1-2-3				
		tests compared	EPM	OF	LD	COMP	EPM	OF	LD	COMP	EPM	OF	LD	COMP	rCOMP	EPM	OF	LD	COMP	EPM	OF	LD	COMP	EPM	OF	LD	COMP	COMP
rat (male)	1	EPM		0.23	-0.09												0.095	0.511										
		OF	0.23		0.34											0.095		0.012										
		LD	-0.09	0.34												0.511	0.012											
	1-2	EPM						0.38	0.33												0.004	0.013						
		OF					0.38		0.45											0.004		0.001						
		LD					0.33	0.45												0.013	0.001							
	1-2-3	EPM										0.40	0.34												0.003	0.011		
		OF									0.40		0.52											0.003		0.000		
		LD									0.34	0.52												0.011	0.000			
mouse (male)	1	EPM		0.18	0.00												0.279	0.987										
		OF	0.18		0.32											0.279		0.050										
		LD	0.00	0.32												0.987	0.050											
	1-2	EPM						0.24	0.32												0.155	0.050						
		OF					0.24		0.24											0.155		0.142						
		LD					0.32	0.24												0.050	0.142							
	1-2-3	EPM										0.28	0.44												0.094	0.006		
		OF									0.28		0.39											0.094		0.015		
		LD									0.44	0.39												0.006	0.015			
rat (male)	aversive	EPM	0.34				0.39				0.46					0.013				0.003				0.001				
		OF		0.35				0.47				0.63					0.012				0.000				0.000			
		LD			0.44				0.51				0.55					0.001				0.000				0.000		
		COMP				0.59				0.69				0.78					0.000				0.000				0.000	
rat (female)	aversive	EPM	0.176				0.264				0.390					0.380				0.183				0.044				
		OF																										
		LD			0.384				0.570				0.659					0.048				0.002				0.000		
mouse (male)	aversive	EPM	-0.05				0.39				0.53					0.765				0.016				0.001				
		OF		0.32				0.29				0.35					0.047				0.077				0.030			
		LD			0.37				0.33				0.52					0.021				0.040				0.001		
		COMP				0.39				0.43				0.53					0.015				0.007				0.001	
rat (male)	safe	CPF	0.08	0.17	0.16	0.24	0.14	0.33	0.27	0.33	0.17	0.30	0.23	0.33	0.40	0.615	0.246	0.294	0.104	0.354	0.027	0.067	0.027	0.268	0.045	0.119	0.027	0.005
rat (male)		startle	-0.03	-0.10	0.08	0.01	0.06	-0.08	0.03	0.02	0.12	-0.15	-0.01	-0.01		0.833	0.495	0.590	0.960	0.689	0.575	0.865	0.869	0.424	0.329	0.959	0.973	

Table 1.: Correlation coefficients and corresponding p-values of the “inter-test correlations” and “behavioral predictions” experiments in rats and mice. The color code of the cells ranging from red to blue represents the level of correlations or p-values scaled to each parameter and experiment.

SuMs are more sensitive markers of social isolation-induced chronic stress

Since TA is the core symptom of anxiety disorders, TA markers should be responsive to etiological models of such conditions. Post-weaning social isolation (PWSI) is an extensively used animal model of human childhood neglect, one of the most prevalent maltreatment that facilitates abnormal anxiety[24]. To evaluate the sensitivity of our approach to characterize isolation-induced anxiety, we conducted a simplified version of our original protocol consisting of three baselines and one aversive sampling event of the EPM and LD tests following PWSI. In EPM, we found no isolation-related differences on neither sampling day (Fig.5B). However, SuMs of anxiety scores of baselines or aversive sampling days were higher following PWSI, indicating that two tests were necessary to show isolation-induced changes (Fig.5C). In the case of the LD test or COMPs, all sampling events and their SuMs showed significant differences (Fig.5E,5F,5H,5I). Since in these cases, statistical significance was not suitable to reveal differences between sampling approaches *per se* due to its binary nature, we also showed the effect sizes of social isolation on anxiety scores. We found that the effect sizes based on averaged sampling events (SuMs) were higher than the averages of effect sizes of those very same events. In other words, creating SuMs of repeated sampling was more sensitive to describe isolation-induced changes in anxiety than the single sampling events (Fig.5D,5G,5J).

Figure 5

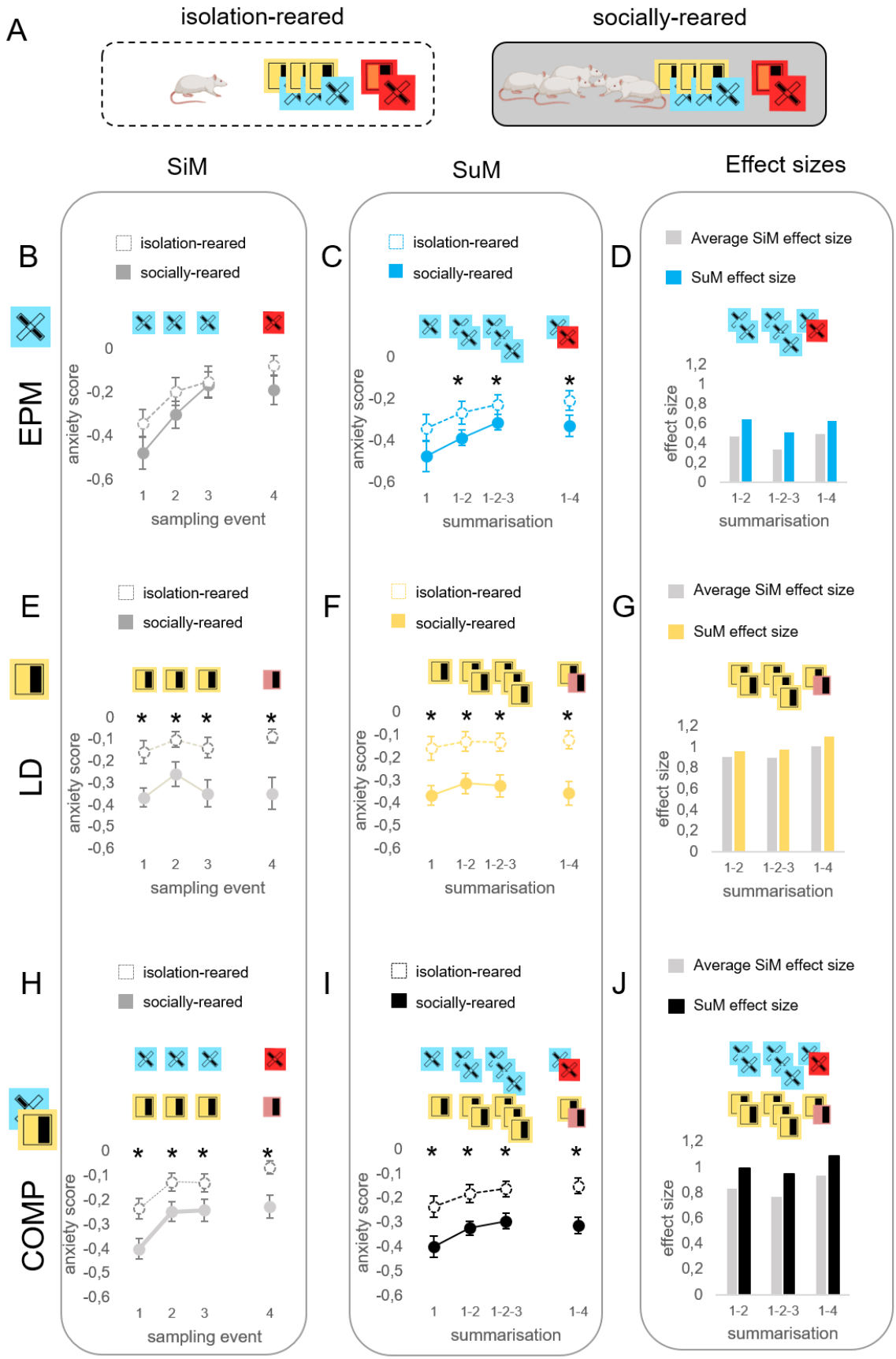
A
isolation-reared
socially-reared
SiM
SuM
Effect sizes
B
C
D
EPM
isolation-reared
socially-reared
anxiety score
sampling event
summarisation
Average SiM effect size
SuM effect size
effect size
E
F
G
LD
H
I
J
COMP

Figure 5.: **Summary measures as markers of a chronic stress state. A)** Socially or isolation-reared animals underwent three baseline and one aversive sampling events of the EPM and LD tests then summary measures of different complexity were calculated. **B-E-H)** SiMs (time and frequency-based anxiety scores) were calculated from each sampling events using the EPM (B) or LD tests (E), or from the composite of those (H). **C-F-I)** SiMs or SuMs were calculated from the first or more sampling events (1-2, 1-2-3, 1-4), respectively using the EPM (C), LD (F), or their composites (I). **D-G-J)** The effect sizes of isolation rearing based on averaged sampling events (SuMs) compared to the averages of effect sizes of those very same sampling events without summarization. Briefly, grey and colored bars represent the effect sizes of the comparisons from the first (B-E-H) and the second columns (C-F-I), respectively. All statistical results are presented in Table 2.

SuMs offer a superior ability to identify molecular markers of anxiety

Since SuMs represent robust measures of anxiety, we sought to investigate their molecular correlates too. We conducted an RNA-seq experiment on mPFC tissue – the area that represents a top-down control node of anxiety[25,26] – of rats in a resting state. Subjects (n=27) were homogeneously chosen based on their anxiety levels (Fig.6A) and gene expression data were explained by either SiMs or SuMs. Gene-behaviour associations were characterized by their robustness and the change in gene expression per one unit of change in the anxiety scale (logFC). We found that SuMs multiply the significant discoveries compared to SiMs: SuM of LD and SuM of EPM 3-fold and 8-fold increased the number of gene targets, respectively. Interestingly, most LD-SiM-associated genes are also associated with LD-SuM, while differently complex EPM measures are defined by non-overlapping gene sets (Fig.6C). We found that the type of sampling and test defines the strength of gene-behaviour associations interactively. Gene products that correlate with EPM-SiM showed the highest logFC values but also the lowest robustness of all (Fig.6D). The robustness of the relationship between LD-SuM and gene expression was slightly lower than that of the SiM-gene associations, however, most of these gene targets are represented in both sampling approach (Fig.6C). The least robust targets (EPM-SiMs) vanish in an enhanced sampling approach, while the robust associations, like LD-SiMs, are represented in both SiM and SuM sampling. These results indicate, that a SuM-based analysis increases both the specificity and sensitivity of finding molecular correlates of anxiety as it excludes non-robust false positive discoveries (like EPM SiMs) but includes false negative discoveries that would have never emerged in a SiM-based experiment as significant.

Figure 6

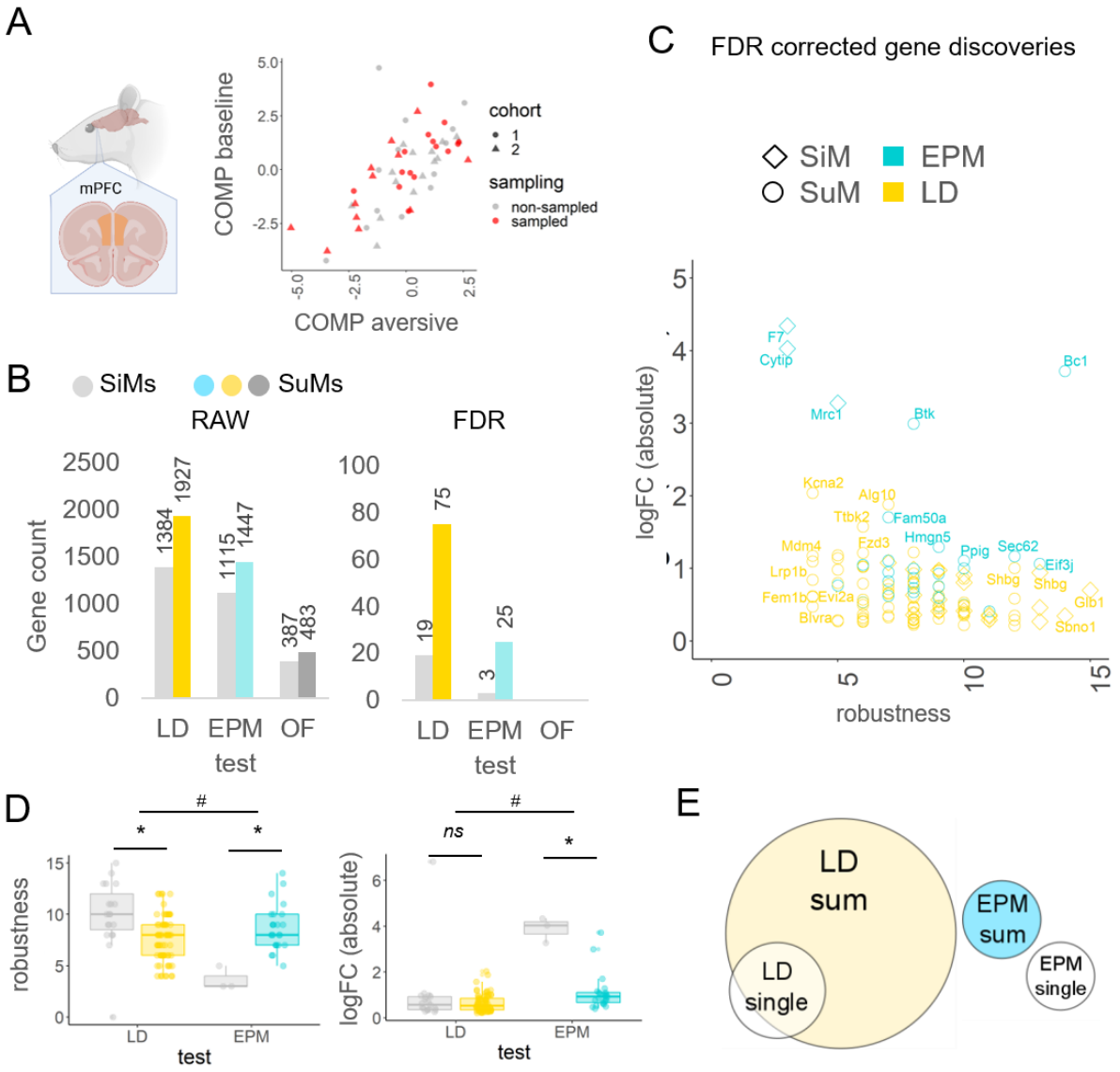

Figure 5. Sensitivity to gene-discovery of differently complex anxiety measures. A) left: Representative figure of bilateral mPFC sampling (orange areas). right: Samples for RNA-seq analysis were chosen based on their COMP level; every second experimental subjects on this scale were investigated. This way, predictive correlations remained closely intact. **B)** SiM and SuM-associated gene count following raw analysis (left) or FDR correction to multiple comparisons (right). **C)** The relation between the per unit fold change of expression (logFC / change in gene expression per one unit of change in the anxiety scale) and the robustness (the smallest number of randomly excludable samples that likely impair the statistical significance of a gene-behavior relationship) of genes in behaviour-gene associations. **D)** Robustness (left) and logFC (right) values of different test types and sampling approaches. **E)** The Euler diagram represents the proportionality of overlaps in gene sets of sampling approaches. All statistical results are presented in Table 2.

SiM and SuM-associated genes define distinct functional clusters

The solid contrast between the ability of sampling methods to unravel genes has led us to investigate the functional variability of differently discovered gene sets. We applied the DAVID functional annotation tool to cluster gene discoveries. We found that SiMs are associated with genes that are involved in metabolic and stress processes, while SuMs establish a plasticity-focused gene profile. The majority of genes associated with SuMs are also putatively involved in cellular localization, cellular transport, potassium signaling, and/or transcription. Plasticity-associated genes can be further structured to more specified functions, including but not limited to extracellular matrix shaping (ADAMTS4; HAPLN2/BRAL1), general modulation of neuronal development (FGF14; FAT3; FZD3), transcription (BTK; TCEB2; ASCL2; MDM4) and translation (BC1; EIFJ2; EIF2S2; UPF2; UPF3B). Also, a significant percentage of SuM-associated genes are direct effectors of neuronal excitability and neurotransmission, such as KCNK1, KCNK9, KCNA2, KCNS2, and KCNJ4. These data indicate that besides our sampling approach improves the gene discovery rate quantitatively, and it is also associated with a qualitatively distinct gene profile. We present the summary statistics of significant genes in Table 3.

Figure 7

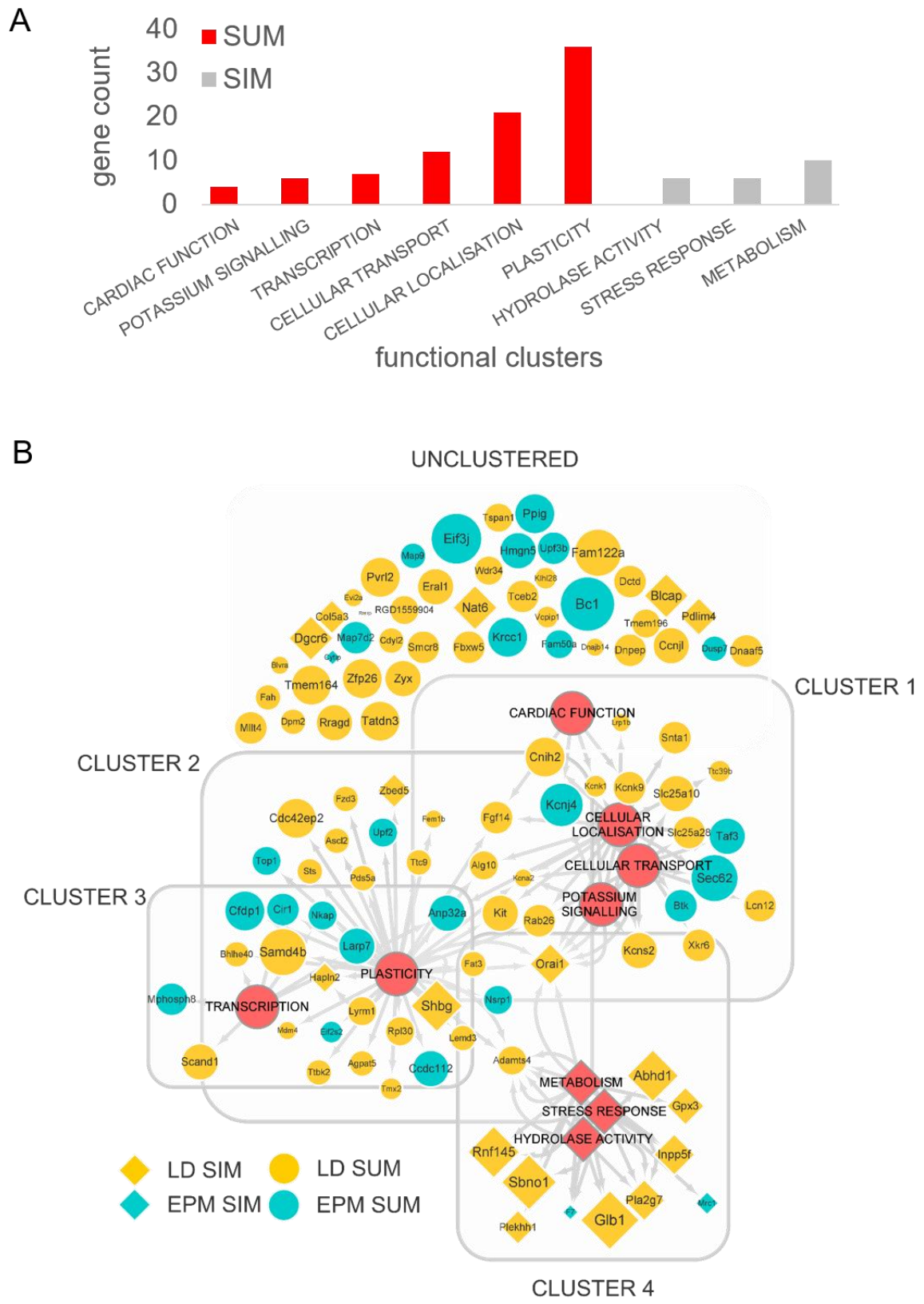

Figure 7. SiM and SuM-associated genes form functionally distinct clusters. A: Clusters unite genes with similar annotation profiles based on multiple databases. SiM and SuM-associated genes are grouped into completely non-overlapping clusters defined by metabolism or plasticity, respectively. **B)** Hub-plot of functional

annotation of significant genes. Due to the heuristic fuzzy clustering method, a gene does not have to belong to any clusters but can belong to more clusters, while clusters can be determined by multiple functional labels. Cluster 1 is formed by genes that share many functional labels that are associated with cardiac function, cellular localization and transport, and potassium signaling. Clusters 2 and 3 are solely characterized by the functions plasticity and transcription, respectively. Cluster 4 is characterized by metabolism, stress response, and hydrolase activity-associated functional terms and unites almost exclusively SiM-associated genes. Unclustered genes are not necessarily lacks a known function, but do not have the amount of relatedness with a certain amount of genes in our pool to form a separate cluster. Grey arrows indicate gene-function associations, while grey lines indicate the borders of clusters which together clearly outline the fuzziness of genes. Red circles indicate cluster-determining labels, and colour and the shape indicate sampling type. The size of hubs is proportional to the correlation coefficient (absolute value) between the gene expression and anxiety.

Discussion

Our study aimed to model TA in laboratory rodents to enhance the translational potential of preclinical testing. We hypothesized that the frequently used preclinical anxiety tests measure SAs that are transiently expressed consequences of TA, consequently, we under-sample in these paradigms. We approached this issue by improving the sampling of classical tests and introducing SuMs of multiple test events and variables.

Here we present the first experimentally validated markers of TA that improve behavioral and molecular predictions of anxiety tests. We found that anxiety tests possess partially overlapping features, which means that creating SuMs is possible and beneficial to explain TA. We validated SuMs in rats and mice as markers of TA by comparing their predictions to classical variables. We showed that SuMs enhance inter-test correlations proportional to their complexity, promote predictions of future behavior in threatening environments, and better describe the consequences of chronic stress. Also, these correlations between SuMs of different tests or between SuMs of different conditions are consistently higher than the means of correlations between SiMs, meaning that SuMs carry additional information over independent SiMs. In summary, we described a stable construct by measuring responses to multiple situations at different time points.

We argue that the above construct is TA, the shared basis of anxiety-like responses. The TA models of Eysenck, Belzung, and the Spielberger state/trait inventory-based clinical practice align with this concept; all describe SA as an interactional consequence of TA and situational stress[1,3,4,27,28]. These models all agree that the proper description of TA requires multiple sampling in multiple situations; however, our approach is the first that sought to achieve this in preclinical settings. Human phenomena that are explained by TA, such as the clustering of anxiety types in clinical cases or individual-level predictions of future responses are all inconsistent in preclinical models[4,16,18]. In our study, the gradual improvement in inter-test correlations, and the predictive power of our models by incorporating more complex sampling supports that SuMs converge to TA. To our best knowledge, this is the first TA paradigm that modifies experimental analysis instead of experimental subjects[20–22,29], giving way to individual phenotyping on a level that the time-and effort-consuming experiments of modern neuroscience desire.

Next, we aimed to physiologically validate SuMs as measures of TA, so we compared these with SiMs in describing whole-genome expression profiles in mPFC samples. SuMs revealed a quantitatively and qualitatively different genetic profile from SiMs. SuMs compared to SiMs enhanced gene discoveries up to an 8-fold improvement, and the detected genes were clustered into functionally different groups.

Higher discovery rates of SuMs indicate that such variables are more sensitive to detect gene-anxiety associations. The reason for that might be that SuMs describe a greater proportion of variance in gene expression data of the same subject. This allows us to use analytical approaches which require higher statistical power but also possess greater biological relevance, such as describing gene expression with continuous anxiety scores instead of categorical anxiety variables, e.g. high and low anxiety groups. Also, standard designs require fewer subjects to get valuable data which aligns with the 3R principle. Finally, greater sensitivity improves the discovery rate of new gene targets of anxiety[30].

Qualitatively different gene profiles of SuMs and SiMs indicate that SuMs, besides superior sensitivity, explain a previously undescribed type of variability. Given that our SuMs are validated markers of TA, additional variance in gene expression explained by these potentially represents such a trait. These previously unnoticed gene products are all potential candidates for the treatment of trait anxiety and thus anxiety disorders.

SuM genes show high functional relatedness and are associated with neural plasticity and psychiatric disorders through different mechanisms and pathways. SuMs are correlated positively with the expression of the perineuronal net (PNN) and perinodal extracellular matrix (ECM) shaping molecules that influence membrane currents, excitability, and synaptic reorganization and were described in mental disorders[31]. ADAMTS4 is responsible for the cleavage of PNN elements like brevican, aggrecan, and reelin[31,32], while HAPLN2/BRAL1 is an ECM composer that establishes the proper conductivity of neurons by stabilizing Ranvier nodes[31]. SuMs are negatively correlated with the trophic factor FGF14 that establishes proper excitability of neurons[33]. As a general modulator, FGF14 deficiency is also associated with psychiatric disorders, such as depression, bipolar disorder, epilepsy, and schizophrenia[34]. SuMs are in a positive relationship with many general modulators of gene transcription and translation that control the expression of thousands of genes with particular motifs. Examples of this are Bruton tyrosine kinase (BTK) and components of the nonsense-mediated mRNA decay, the up-frameshift factors UPF2 and UPF3b[35,36]. There are also gene discoveries that support transcription or translation by the composing and modulation of its machinery, such as the Elongin-B (TCEB2)[37] or brain cytoplasmic RNA 1 (BC1)[38] and subunits of the Eukaryotic translation initiation factors EIF2s2, EIFJ2, and EIF3J[39,40]. Finally, we found many direct modulators of membrane currents such as KCNK1, KCNK9, KCNA2, KCNS2, and KCNJ4 potassium channels. Interestingly, all the general mechanisms and pathways above converge to modify neuronal excitability and plasticity and many do it by controlling wide clusters of genes. This indicates that trait anxiety correlates with a consistently altered gene expression, excitability, and the capability of neuronal reorganization in resting conditions in the mPFC. These data represent a range of promising candidates that could trace out novel interventions to treat anxiety.

In summary, here we provide a novel sampling and analysis pipeline that enables to measure TA in preclinical research. Our paradigm is based on an extended sampling process, using popular anxiety tests, and consequently is readily applicable to any model and feasible to any laboratory. Also, most of these results are validated in both rats and mice indicating the robustness of our approach. Using SuMs in experiments boosts behavioral and molecular phenotyping and predictions, consequently reducing the minimally sufficient sample sizes while maximizing the discovery rate of novel treatment candidates. In addition, SuMs' gene discoveries functionally differ from SiM-associated genes, as the former revealed a plasticity-focused gene profile of trait anxiety. We encourage colleagues to use this approach to phenotype their subjects in a highly detailed manner, potentially bridging the gap between preclinical and clinical anxiety research.

figure / result	hypotheses testing / model	test statistics-value	df/ss	sig. level	further
Fig.2A	Kruskal-Wallis rank sum test **(EPM / OF / LD)**	**34.901 / 4.033 / 0.265** (chi-squared)	2 (df)	**2.6383e-08 / 0.133 / 0.876** (p-value)	
S.Fig. 1A	Kruskal-Wallis rank sum test **(EPM / OF / LD)**	**57.354 / 9.873 / 1.373** (chi-squared)	2 (df)	**3.513e-13 / 0.007 / 0.503** (p-value)	
Fig. 3B	Linear-mixed model's **(complexity)**	**0.019** (estimate), **0.005** (std.error)	54 (n)	**0.0008** (p-value), **3.597** (t-value)	**0.0065 / 0.0088 / 0.0042** (random effect's variances in variable **composition / test composition / residual**)
Fig.4B	Wilcoxon rank sum test **(baseline vs aversive EPM / OF / LD)**	**1512.5 / 1813 / 1229** (w)		**0.006 / 1.281e-06 / 0.575** (p-value)	
Fig.4C	ANOVA **(corticosterone)**	**7.000** (F)	2 (df)	**0.001** (p-value)	
Fig.4C	Tukey post-hoc **(aversive-resting1 / aversive-resting2 / resting1-resting2)**	**426 / 398 / -28** (estimate), **50.6** (std.error)	54 (n)	**0.01** **0.0001 / 0.845** (p-value), **8.407 / 7.854 / 0.554** (t-value)	
Fig. 5B	Wilcoxon rank sum test **(EPM sampling1 / sampling2 / sampling3 / sampling4aversive)**	**86 / 74 / 97 / 34** (w)		**0.143 / 0.056 / 0.263 / 0.023** (p-value)	FDR: **non-significant**
Fig. 5C	Wilcoxon rank sum test **(EPM sampling1-2 / sampling1-2-3 / sampling1-4aversive)**	**37 / 36 / 34** (w)		**0.023 / 0.034 / 0.020** (p-value)	FDR: **all significant**
Fig. 5E	Wilcoxon rank sum test **(LD sampling1 / sampling2 / sampling3 / sampling4aversive)**	**41 / 50 / 42 / 43** (w)	15-15 (n)	**0.01** **0.004 / 0.001 / 0.002** (p-value)	FDR: **all significant**
Fig. 5F	Wilcoxon rank sum test **(LD sampling1-2 / sampling1-2-3 / sampling1-4aversive)**	**38 / 43 / 33** (w)		**0.01** **0.002 / 0.000** (p-value)	FDR: **all significant**
Fig. 5H	Wilcoxon rank sum test **(COMP sampling1 / sampling2 / sampling3 / sampling4(aversive)**	**52 / 52 / 64 / 39** (w)		**0.006 / 0.006 / 0.023 / 0.001** (p-value)	FDR: **all significant**
Fig. 5I	Wilcoxon rank sum test **(COMP sampling1-2 / sampling1-2-3 / sampling1-4aversive)**	**37 / 36 / 34** (w)		**0.01** **0.000 / 0.000** (p-value)	FDR: **all significant**
Fig.6C	ANOVA **(robustness)**	**0.0128 / 4.8248 / 21.9427** (F)	1 (df)	**0.920 / 0.030 / 7.581e-06** (p-value)	
Fig.6C	ANOVA **(logFoldChange)**	**18.352 / 14.830 / 27.217** (F)	1 (df)	**3.765e-05 / 0.0002 / 7.874e-07** (p-value)	

Table 2.: Parameters of test statistics. The hypotheses testing/model, test statistics-value, df/ss, and sig. level sections shows the approach and assessed conditions, the values of the test statistics, sample size information and significance levels, respectively. Conditions and parameter values are shown in bold, parameter types are shown in plain text. We presented the conventional parameters of the particular statistical approaches. Section Further shows additional relevant information that is specific for a particular method, like random variances in mixed models or the results of FDR correction following multiple comparisions.

model	sampling	gene symbol	base means	log_2 FC	log_2 FC SE	robustness	FDR-adjusted p-value	IHW weight
EPM	SiM	Cytip	23.606	4.028	0.960	3	0.050	2.771
		F7	22.638	4.342	1.022	3	0.050	2.771
		Mrc1	155.711	3.277	0.780	5	0.050	4.962
	SuM	Anp32a	1803.473	0.481	0.113	9	0.013	2.500
		Bc1	25643.881	3.720	0.568	14	0.000	2.447
		Btk	38.921	2.995	0.763	8	0.045	1.325
		Ccdc112	176.938	0.929	0.218	9	0.026	0.650
		Cfdp1	806.939	0.499	0.126	10	0.026	2.500
		Cir1	476.013	0.870	0.191	8	0.004	2.500
		Dusp7	851.688	-0.534	0.130	6	0.016	3.026
		Eif2s2	1070.662	0.767	0.205	5	0.045	2.500
		Eif3j	696.432	1.062	0.205	13	0.001	1.200
		Fam50a	167.672	1.704	0.434	7	0.045	1.200
		Hmgn5	110.513	1.294	0.299	9	0.022	0.673
		Kcnj4	908.471	-0.407	0.110	11	0.045	3.026
		Krcc1	239.564	0.999	0.230	10	0.021	0.673
		Larp7	525.165	0.753	0.162	9	0.003	2.500
		Map7d2	1219.487	0.674	0.137	8	0.001	2.500
		Map9	582.845	1.049	0.270	6	0.033	2.500
		Mphosph8	1155.360	0.727	0.157	8	0.003	2.500
		Nkap	140.282	1.109	0.257	7	0.016	1.200
		Nsrp1	294.536	0.949	0.217	7	0.020	0.627
		Ppig	1241.829	1.107	0.196	10	0.000	2.500
		Sec62	1834.781	1.166	0.207	12	0.000	2.447
		Taf3	336.897	0.595	0.143	9	0.035	0.673
		Top1	420.345	0.825	0.192	7	0.017	1.200
		Upf2	529.868	0.623	0.142	7	0.008	2.500
		Upf3b	574.895	0.998	0.246	8	0.046	0.627
LD	SiM	Abhd1	145.172	0.460	0.112	13	0.034	0.999
		Adamts4	78.607	0.652	0.165	9	0.049	1.288
		Blcap	1577.023	0.291	0.065	11	0.024	0.449
		Col5a3	326.879	0.586	0.149	8	0.026	4.378
		Dgcr6	933.387	0.385	0.093	11	0.033	1.062
		Glb1	160.858	0.857	0.153	15	0.000	1.288
		Gpx3	181.445	0.983	0.234	9	0.024	1.464
		Hapln2	334.366	1.182	0.288	7	0.031	1.288
		Inpp5f	699.704	0.417	0.095	10	0.019	1.046
		Nat6	285.529	0.393	0.092	11	0.024	1.288
		Orai1	104.910	0.963	0.225	10	0.026	0.789
		Pdlim4	397.688	0.410	0.107	9	0.031	4.313
		Pla2g7	184.910	-0.932	0.209	10	0.018	0.999
		Plekhh1	406.534	0.988	0.235	8	0.015	4.378
		Rmrp	40.122	-6.810	1.370	0	0.004	1.288
		Rnf145	1136.341	-0.283	0.059	13	0.004	1.244
		Sbno1	1076.216	-0.406	0.081	14	0.004	0.825
		Shbg	53.145	1.004	0.207	13	0.004	1.252

	Zbed5	428.279	0.362	0.099	8	0.048	4.313
	Adamts4	78.607	0.570	0.159	7	0.047	1.587
	Agpat5	483.012	-0.263	0.070	6	0.035	1.652
	Alg10	32.986	-1.880	0.460	7	0.015	1.652
	Ascl2	66.192	1.032	0.245	6	0.013	1.959
	Bhlhe40	1299.835	-0.301	0.080	6	0.029	2.396
	Blcap	1577.023	0.291	0.057	12	0.004	0.640
	Blvra	296.247	-0.473	0.133	4	0.044	2.044
	Ccnjl	34.154	0.969	0.225	9	0.011	1.652
	Cdc42ep2	202.890	0.374	0.095	10	0.023	1.652
	Cdyl2	59.969	-0.790	0.210	6	0.034	1.959
	Cnih2	2437.705	0.352	0.076	10	0.004	2.396
	Col5a3	326.879	0.634	0.127	12	0.004	1.448
	Dctd	48.110	0.596	0.166	8	0.047	1.652
	Dgcr6	933.387	0.361	0.087	10	0.021	0.847
	Dnaaf5	117.892	0.461	0.128	8	0.046	1.476
	Dnajb14	67.919	-1.098	0.293	4	0.042	1.120
	Dnpep	964.720	0.254	0.067	8	0.044	0.913
	Dpm2	315.731	0.369	0.102	6	0.045	1.589
	Eral1	351.824	-0.308	0.084	9	0.037	2.044
	Evi2a	277.589	0.618	0.168	4	0.042	1.448
	Fah	200.247	0.497	0.122	6	0.015	1.652
	Fam122a	214.924	0.348	0.088	12	0.024	1.549
	Fat3	152.133	-1.018	0.277	6	0.042	1.476
SuM	Fbxw5	737.113	0.287	0.070	8	0.037	0.347
	Fem1b	221.242	-0.608	0.172	4	0.049	1.698
	Fgf14	31.569	-0.802	0.221	8	0.042	1.834
	Fzd3	55.025	-1.216	0.277	6	0.009	1.959
	Glb1	160.858	0.699	0.160	10	0.010	1.698
	Gpx3	181.445	0.974	0.213	9	0.006	1.652
	Hapln2	334.366	1.080	0.272	5	0.023	1.448
	Kcna2	36.710	-2.040	0.545	4	0.035	1.959
	Kcnk1	1531.718	-0.273	0.077	5	0.042	2.396
	Kcnk9	52.273	-1.220	0.334	8	0.042	1.834
	Kcns2	60.124	-0.745	0.180	9	0.014	1.959
	Kit	183.256	-0.620	0.148	9	0.014	1.652
	Klhl28	38.999	-1.140	0.323	5	0.048	1.834
	Lcn12	37.886	0.785	0.208	8	0.035	1.652
	Lemd3	277.339	-0.322	0.088	6	0.042	1.589
	Lrp1b	352.926	-0.844	0.236	4	0.049	1.409
	Lyrm1	182.250	0.463	0.126	7	0.042	1.698
	Mdm4	95.155	-1.173	0.319	4	0.042	1.587
	Mllt4	1188.392	-0.247	0.065	8	0.042	0.868
	Nat6	285.529	0.341	0.089	8	0.027	2.044
	Orai1	104.910	0.908	0.212	9	0.013	1.476
	Pds5a	192.446	-0.572	0.156	6	0.042	1.549
	Pla2g7	184.910	-0.807	0.205	8	0.024	1.549
	Pvrl2	86.300	0.495	0.132	10	0.037	1.597

Rab26	1010.801	0.305	0.081	8	0.044	0.847
RGD1559904	298.985	-0.327	0.091	7	0.045	1.652
Rnf145	1136.341	-0.268	0.054	11	0.004	0.347
Rpl30	472.323	0.662	0.164	7	0.019	1.507
Rragd	188.492	-0.452	0.107	9	0.013	1.698
Samd4b	829.421	0.215	0.051	12	0.029	0.347
Sbno1	1076.216	-0.347	0.081	10	0.014	0.847
Scand1	644.680	0.484	0.115	9	0.014	1.589
Shbg	53.145	0.942	0.195	12	0.004	1.652
Slc25a10	322.184	0.392	0.095	9	0.015	1.448
Slc25a28	536.981	0.295	0.071	8	0.015	1.507
Smcr8	96.412	-0.931	0.227	8	0.015	1.658
Snta1	641.738	0.223	0.062	8	0.047	1.467
Sts	173.874	0.531	0.147	6	0.044	1.698
Tatdn3	213.583	0.528	0.113	10	0.004	1.698
Tceb2	829.548	0.475	0.112	8	0.015	0.913
Tmem164	542.956	-0.220	0.061	10	0.045	1.652
Tmem196	166.953	-0.713	0.188	7	0.029	2.044
Tmx2	1251.797	-0.286	0.077	5	0.048	0.868
Tspan1	49.982	0.777	0.206	7	0.042	1.120
Ttbk2	69.348	-1.574	0.349	6	0.009	1.120
Ttc39b	71.332	-0.981	0.256	5	0.027	1.959
Ttc9	521.426	-0.276	0.073	6	0.035	1.652
Vcpip1	199.177	-0.799	0.222	5	0.045	1.652
Wdr34	166.398	0.350	0.098	7	0.047	1.549
Xkr6	81.662	0.814	0.193	8	0.013	1.587
Zfp26	155.941	-0.483	0.109	10	0.009	1.658
Zyx	607.518	0.409	0.095	9	0.011	1.589

Table 3.: Parameters of significant RNA expression – behavior associations sorted by different sampling and analysis approaches. **Abbreviations:** log2FC: log2 fold-change, SE: standard error, FDR: false discovery rate, IHW: independent hypothesis weighting

Acknowledgement

This work was supported by the Hungarian Scientific Research Fund (Grant No. K125390), Eötvös Loránd Research Network (Grant No. SA-49/2021), Hungarian Brain Research Program (Grant No. 2017-1.2.1-NKP-2017-00002), for dr. Éva Mikics, and the Bolyai Janos Research Fellowship for Mate Toth. We would like to thank Dr. Kornél Demeter, the head of the Unit of Behavioral Studies at IEM for providing the environment for keeping laboratory animals and conducting experiments. On behalf of the "NGS analysis in neurological and behavioral sciences" project we are grateful for the possibility to use ELKH Cloud (see Héder et al. 2022; https://science-cloud.hu/) which helped us achieve the results published in this paper.

In accordance with Elsevier guidelines and permissions, we chose to upload this preprint to the preprint server bioRxiv.

Conflict of Interest Statement

The authors decleare no conflict of interest of any kind.

References

1. Spielberger, C. D. & Reheiser, E. C. Assessment of Emotions: Anxiety, Anger, Depression, and Curiosity. *Appl. Psychol. Health Well-Being* **1**, 271–302 (2009).
2. Kennedy, B. L., Schwab, J. J., Morris, R. L. & Beldia, G. Assessment of state and trait anxiety in subjects with anxiety and depressive disorders. *Psychiatr. Q.* **72**, 263–276 (2001).
3. Vagg, P. R., Spielberger, C. D. & O'Hearn, T. P. Is the state-trait anxiety inventory multidimensional? *Personal. Individ. Differ.* **1**, 207–214 (1980).
4. Belzung, C. & Griebel, G. Measuring normal and pathological anxiety-like behaviour in mice: a review. *Behav. Brain Res.* **125**, 141–149 (2001).
5. Baxter, A. J., Scott, K. M., Vos, T. & Whiteford, H. A. Global prevalence of anxiety disorders: a systematic review and meta-regression. *Psychol. Med.* **43**, 897–910 (2013).
6. Sartori, S. B. & Singewald, N. Novel pharmacological targets in drug development for the treatment of anxiety and anxiety-related disorders. *Pharmacol. Ther.* **204**, 107402 (2019).
7. Haller, J., Aliczki, M. & Gyimesine Pelczer, K. Classical and novel approaches to the preclinical testing of anxiolytics: A critical evaluation. *Neurosci. Biobehav. Rev.* **37**, 2318–2330 (2013).
8. Lister, R. G. Ethologically-based animal models of anxiety disorders. *Pharmacol. Ther.* **46**, 321–340 (1990).
9. Goes, T. C., Antunes, F. D. & Teixeira-Silva, F. Trait and state anxiety in animal models: Is there correlation? *Neurosci. Lett.* **450**, 266–269 (2009).
10. Lister, R. G. The use of a plus-maze to measure anxiety in the mouse. *Psychopharmacology (Berl.)* **92**, 180–185 (1987).
11. Pellow, S., Chopin, P., File, S. E. & Briley, M. Validation of open:closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. *J. Neurosci. Methods* **14**, 149–167 (1985).
12. Crawley, J. N. Neuropharmacologic specificity of a simple animal model for the behavioral actions of benzodiazepines. *Pharmacol. Biochem. Behav.* **15**, 695–699 (1981).
13. Merlo Pich, E. & Samanin, R. A two-compartment exploratory model to study anxiolytic/anxiogenic effects of drugs in the rat. *Pharmacol. Res.* **21**, 595–602 (1989).
14. Carola, V., D'Olimpio, F., Brunamonti, E., Mangia, F. & Renzi, P. Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice. *Behav. Brain Res.* **134**, 49–57 (2002).
15. Hall, C. S. Emotional behavior in the rat. III. The relationship between emotionality and ambulatory activity. *J. Comp. Psychol.* **22**, 345–352 (1936).
16. Griebel, G., Belzung, C., Perrault, G. & Sanger, D. J. Differences in anxiety-related behaviours and in sensitivity to diazepam in inbred and outbred strains of mice. *Psychopharmacology (Berl.)* **148**, 164–170 (2000).
17. Beuzen, A. & Belzung, C. Link between emotional memory and anxiety states: A study by principal component analysis. *Physiol. Behav.* **58**, 111–118 (1995).
18. Andreatini, R. & Bacellar, L. F. Animal models: trait or state measure? The test-retest reliability of the elevated plus-maze and behavioral despair. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **24**, 549–560 (2000).
19. File, S. E. The interplay of learning and anxiety in the elevated plus-maze. *Behav. Brain Res.* **58**, 199–202 (1993).
20. Landgraf, R. & Wigger, A. High vs Low Anxiety-Related Behavior Rats: An Animal Model of Extremes in Trait Anxiety. *Behav. Genet.* **32**, 301–314 (2002).
21. Liebsch, G., Montkowski, A., Holsboer, F. & Landgraf, R. Behavioural profiles of two Wistar rat lines selectively bred for high or low anxiety-related behaviour. *Behav. Brain Res.* **94**, 301–310 (1998).
22. Landgraf, R. & Wigger, A. Born to be Anxious: Neuroendocrine and Genetic Correlates of Trait Anxiety in HAB Rats. *Stress* **6**, 111–119 (2003).
23. Matthews, J. N., Altman, D. G., Campbell, M. J. & Royston, P. Analysis of serial measurements in medical research. *Br. Med. J.* **300**, 230–235 (1990).
24. Dubowitz, H. & Bennett, S. Physical abuse and neglect of children. *The Lancet* **369**, 1891–1899 (2007).

25. Tovote, P., Fadok, J. P. & Lüthi, A. Neuronal circuits for fear and anxiety. *Nat. Rev. Neurosci.* **16**, 317–331 (2015).
26. Calhoon, G. G. & Tye, K. M. Resolving the neural circuits of anxiety. *Nat. Neurosci.* **18**, 11 (2015).
27. Eysenck, M. W., Derakshan, N., Santos, R. & Calvo, M. G. Anxiety and cognitive performance: Attentional control theory. *Emotion* **7**, 336–353 (2007).
28. Spielberger, C. D., Gorsuch, R. L. & Lushene, R. E. Manual for the State-trait Anxiety Inventory ("Self-evaluation Questionnaire"). *Calif. Consult. Psychol. Press* 36 pp (1970).
29. Filiou, M. D. *et al.* Proteomics and Metabolomics Analysis of a Trait Anxiety Mouse Model Reveals Divergent Mitochondrial Pathways. *Biol. Psychiatry* **70**, 1074–1082 (2011).
30. Lewis, D. I. Animal experimentation: implementation and application of the 3Rs. *Emerg. Top. Life Sci.* **3**, 675–679 (2019).
31. Fawcett, J. W., Oohashi, T. & Pizzorusso, T. The roles of perineuronal nets and the perinodal extracellular matrix in neuronal function. *Nat. Rev. Neurosci.* **20**, 451–465 (2019).
32. Sakai, K., Shoji, H., Kohno, T., Miyakawa, T. & Hattori, M. Mice that lack the C-terminal region of Reelin exhibit behavioral abnormalities related to neuropsychiatric disorders. *Sci. Rep.* **6**, 28636 (2016).
33. Lou, J.-Y. *et al.* Fibroblast growth factor 14 is an intracellular modulator of voltage-gated sodium channels. *J. Physiol.* **569**, 179–193 (2005).
34. Hoxha, E. *et al.* Emerging roles of Fgf14 in behavioral control. *Behav. Brain Res.* **356**, 257–265 (2019).
35. Raimondeau, E., Bufton, J. C. & Schaffitzel, C. New insights into the interplay between the translation machinery and nonsense-mediated mRNA decay factors. *Biochem. Soc. Trans.* **46**, 503–512 (2018).
36. Huang, L. *et al.* A Upf3b-mutant mouse model with behavioral and neurogenesis defects. *Mol. Psychiatry* **23**, 1773–1786 (2018).
37. Aso, T., Lane, W. S., Conaway, J. W. & Conaway, R. C. Elongin (SIII): A Multisubunit Regulator of Elongation by RNA Polymerase II. *Science* **269**, 1439–1443 (1995).
38. Briz, V. *et al.* The non-coding RNA BC1 regulates experience-dependent structural plasticity and learning. *Nat. Commun.* **8**, 293 (2017).
39. Placzek, A. N. *et al.* eIF2α-mediated translational control regulates the persistence of cocaine-induced LTP in midbrain dopamine neurons. *eLife* **5**, e17517 (2016).
40. Huang, W. *et al.* Translational control by eIF2α phosphorylation regulates vulnerability to the synaptic and behavioral effects of cocaine. *eLife* **5**, e12052 (2016).
41. Noldus, L. P. J. J., Spink, A. J. & Tegelenbosch, R. A. J. EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behav. Res. Methods Instrum. Comput.* **33**, 398–414 (2001).
42. Péter, A. Solomon coder. *Version Beta* **17**, 22 (2017).
43. RCoreTeam. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2019).
44. Kassambara, A. & Mundt, F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. (2019).
45. Harrell, F. E. Harrell Miscellaneous. (2021).
46. Stoffel, M. A., Nakagawa, S. & Schielzeth, H. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **8**, 1639–1644 (2017).
47. Anwer, H. *et al.* An efficient new assay for measuring zebrafish anxiety: Tall tanks that better characterize between-individual differences. *J. Neurosci. Methods* **356**, 109138 (2021).
48. Campbell, R. C. Review of Biometry: The Principles and Practice of Statistics in Biological Research. *J. R. Stat. Soc. Ser. Gen.* **133**, 102–102 (1970).
49. Svetnik, V. *et al.* Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **43**, 1947–1958 (2003).
50. Toth, M., Mikics, E., Tulogdi, A., Aliczki, M. & Haller, J. Post-weaning social isolation induces abnormal forms of aggression in conjunction with increased glucocorticoid and autonomic stress responses. *Horm. Behav.* **60**, 28–36 (2011).

51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
52. Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods* **13**, 577–580 (2016).
53. Huang, D. W. *et al.* The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* **8**, R183 (2007).
54. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* gkac194 (2022) doi:10.1093/nar/gkac194.
55. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).