Main title

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

Short title

Revealing trait anxiety through summary measures

Zoltán K Varga, PhD*[,#,1]; Diána Pejtsik, MSc*[,1,2]; Máté Tóth, PhD[1]; Zoltán Balogh, PhD[1]; Manó Aliczki, PhD[1]; László Szente, MSc[1,2]; Gyula Y Balla, MSc[1]; Levente Kontra, PhD[3]; Zsófia Eckert, HS[1]; Huba Szebik, MSc[1,2]; Zsolt Borhegyi, PhD[1]; Éva Mikics, PhD [#,1]

**equal contribution*

[#]*corresponding authors*

[1] Translational Behavioural Neuroscience Research Group, HUN-REN Institute of Experimental Medicine, Budapest, Hungary

[2] János Szentágothai Neurosciences Program, Semmelweis University, Budapest, Hungary

[3] Bioinformatics Core Facility, HUN-REN Institute of Experimental Medicine, Budapest, Hungary

Correspondence:

Dr. Éva Mikics
HUN-REN Institute of Experimental Medicine
43 Szigony street
1083 Budapest, Hungary
Phone: +36-1-2109400/ext.328
E-mail: mikics.eva@koki.hun-ren.hu

Abstract

The reliability and validity of preclinical anxiety testing is essential for translating animal research into clinical use. However, the commonly used anxiety tests lack inter-test correlations and face challenges with repeatability. While translational animal research should be able to capture stable individual anxiety traits, the current approach employs a single type of test at a single time that only measures transient states of animals, heavily influenced by experimental conditions. Here, we propose a validated, optimized test battery capable of reliably capturing trait anxiety in rats and mice of both sexes. Instead of developing novel tests, we combined widely used tests (elevated plus-maze, open field and light-dark test) to provide instantly applicable adjustments for better predictive validity. We repeated these tests three times to capture behaviour across multiple challenges, which we combined to generate summary measures (SuMs). Our approach resolved between-test correlation issues and provided better predictions for subsequent outcomes under anxiogenic conditions or fear conditioning. SuMs were also able to reveal differences in anxiety in an etiological stress model. Finally, we tested our method's efficacy in discovering anxiety-related molecular pathways through RNA sequencing of the medial prefrontal cortex. Using SuMs, we identified four-times more molecular correlates of trait anxiety as compared to transient anxiety states, pointing out novel functional gene clusters. Furthermore, 16% of these molecular correlates of anxiety were replicated in amygdalar samples, as well. In summary, we provide a novel approach to capture trait anxiety in rodents, offering improved predictions for potential therapeutic targets for personalized medicine.

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 optimal 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.

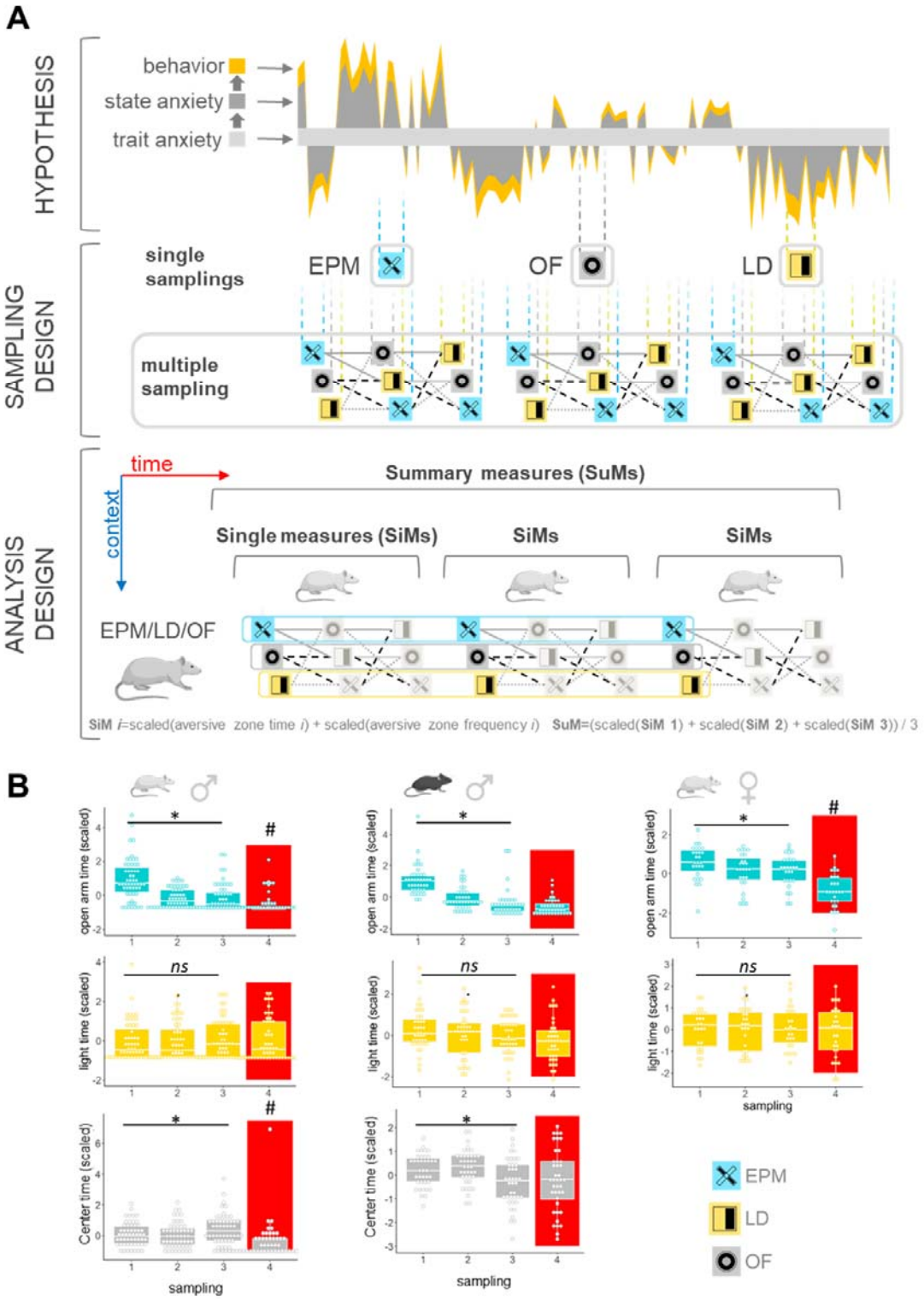
This translational gap potentially stems from the fact that anti-anxiety therapies are considered satisfactory if they control the permanent TA, however, current animal tests only detect the fluctuating SA[4,8,9]. Consequently, we can only expect these tests to highlight markers and treatment targets 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-parametrisation[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 assumed to be 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



106

107

Figure 1. A) Hypothesis, sampling, and analysis design. Hypothesis: Schematic drawing 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:** Behavioral data (time spent and frequency of enters to the aversive compartment) from each test were used to create classical single measures (SiM) and more stable summary measures (SuM), by combining multiple variables or testing events to describe trait anxiety. B) **Shared and unique features of anxiety tests in rats.** The effect of test repetitions on the most frequently used anxiety-like behavioral outcome (time spent in the aversive compartment) of the EPM, LD, and OF anxiety tests in male (left) and female (right) rats and male mice (middle). Red rectangles highlight the testing event in aversive conditions. All statistical results are presented in Table 1. **Abbreviations:** EPM: elevated plus-maze test, OF: open field test, LD: light-dark test * $p<0.05$ test repetition main effect. #: $p<0.05$ significant difference between the aversive and the last baseline sampling event

Results

Overlaps between anxiety-like variables support the use of SuMs

If variables carry partially overlapping information, summarisation is possible and also beneficial for describing their shared background. 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 in baseline and more aversive lighting conditions, within and between-test variability, and bootstrap analysis-based repeatability measures. Test-retest changes were evident in the EPM and the OF, but not in the LD test (Fig.1B). Similarly, aversivity increased anxiety in the EPM and OF, but not in the LD test. Between-test variance were smaller in OF and LD while within-test variance were the highest in LD variables indicating robustness and sensitivity for individual differences. Repeatability measures confirmed such ranking of the tests (supp.Fig.1A,2B). Variables of the same test type clustered together in repeatability. However, frequency variables were the most, while latency variables were the least reproducible variables in most cases, indicating a similarity between the tests. PCA revealed further conformities, as variables of all tests were loaded in the same dimension (supp.Fig.1B,2C). These trends were similar in both species and sexes. In summary, there are test type and variable type similarities of these variables which are consistent in the investigated species and sexes. This indicates a shared but differently expressed basis of these anxiety tests and potential in the summarisation of their variables.

SuMs reveal shared variance of anxiety tests and predict acute stress-induced behaviour

Since variables of the same anxiety test were similar in PCA and RA, we averaged 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 be expressed consistently through contexts and

time. To test this, we created differentially complex SuMs of the EPM, LD, and OF tests from a series of repeated design experiments. We assessed i) their inter-test correlations and their efficacy to predict behaviour in different stress-inducing contexts such as ii) the same tests in an aversive condition (Fig.2D-E), iii) fear conditioning (CFP)(Fig.2F-H), or iv) acoustic startle paradigms (Fig.2H). Our simplest variables (SiMs) were the time spent in, enter frequency or latency to the aversive compartment of a test sampled ones. Our most complex SuMs consist of all of these variables sampled three times. We constructed all possible variables along this scale and described those with a complexity 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.2A-C)(Fig.5A). Inter-test correlations could be explained by variable complexity as a fixed effect, and variable and test composition as interacting random effects. 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 (suppFig.3). We excluded latencies from further analysis and used reduced models with time and frequency data which also outperformed the SiM variables. The association between variable complexity and inter-test correlations were apparent in both species (Fig.2A-C, Fig.5A).

In aversive sampling, we detected enhanced anxiety-like behavior (Fig.2D-E) and corticosterone response compared to baseline conditions (Supplementary fig.4B). 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.2E). Furthermore, creating composite scores from SiMs and SuMs of all tests (COMP) outperformed test-specific SiMs and SuMs, respectively, reaching a 0.78 correlation between

baseline and aversive responses. Predictions of SuMs were higher than the predictions of any of our SiMs (1st, 2nd, or 3rd testing) indicating that their performance is not due to a carry-over effect of repeated testing but these variables explain additional information. Likewise, the predictions of COMPs were higher than the predictions of SuMs which indicates that those carry additional information to any test type that we used here. Similar to our previous studies, baseline or stress-induced corticosterone levels were not correlated with anxiety variables[38]. We confirmed these findings in both species and sexes, showing a similar association between the number of sampling events and their predictive power (Fig.5A).

We also aimed to predict the amount of freezing behavior in a safe context following CFP as a measure of generalized fear, a symptom of post-traumatic stress disorder[39]. Anxiety tests showed poor performance in predicting fear, but we observed a similar increase in correlations in response to using more complex variables, especially COMPs (Fig.2H)(Fig.5A). 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 2

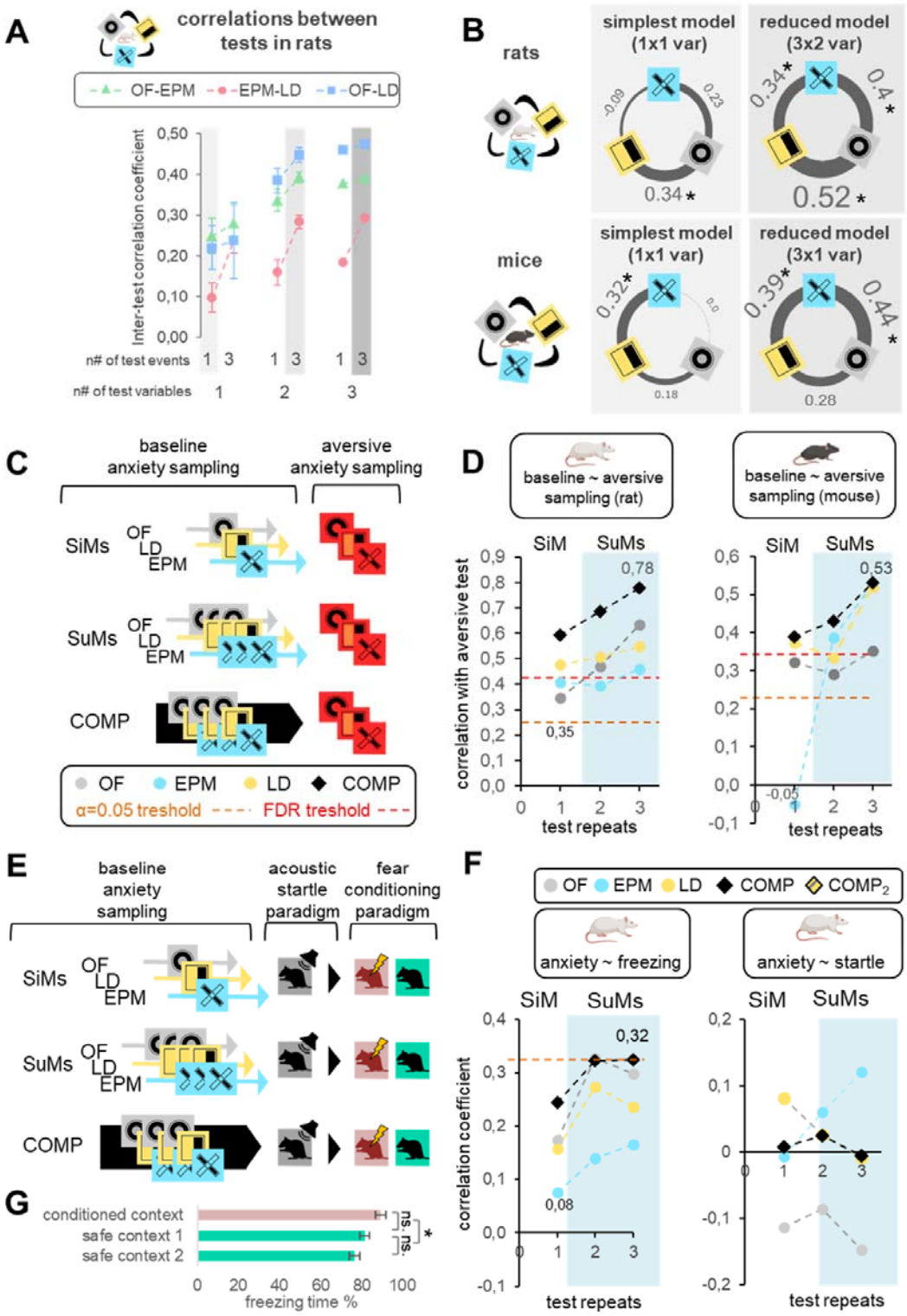

A
correlations between tests in rats
OF-EPM
EPM-LD
OF-LD
Inter-test correlation coefficient
n# of test events
n# of test variables
B
rats
simplest model (1x1 var)
reduced model (3x2 var)
mice
simplest model (1x1 var)
reduced model (3x1 var)
C
baseline anxiety sampling
aversive anxiety sampling
SiMs
SuMs
COMP
OF
EPM
LD
COMP
α=0.05 treshold
FDR treshold
D
baseline ~ aversive sampling (rat)
baseline ~ aversive sampling (mouse)
SiM
SuMs
correlation with aversive test
test repeats
E
baseline anxiety sampling
acoustic startle paradigm
fear conditioning paradigm
SiMs
SuMs
COMP
G
conditioned context
safe context 1
safe context 2
freezing time %
F
OF
EPM
LD
COMP
COMP2
anxiety ~ freezing
anxiety ~ startle
SiM
SuMs
correlation coefficient
test repeats

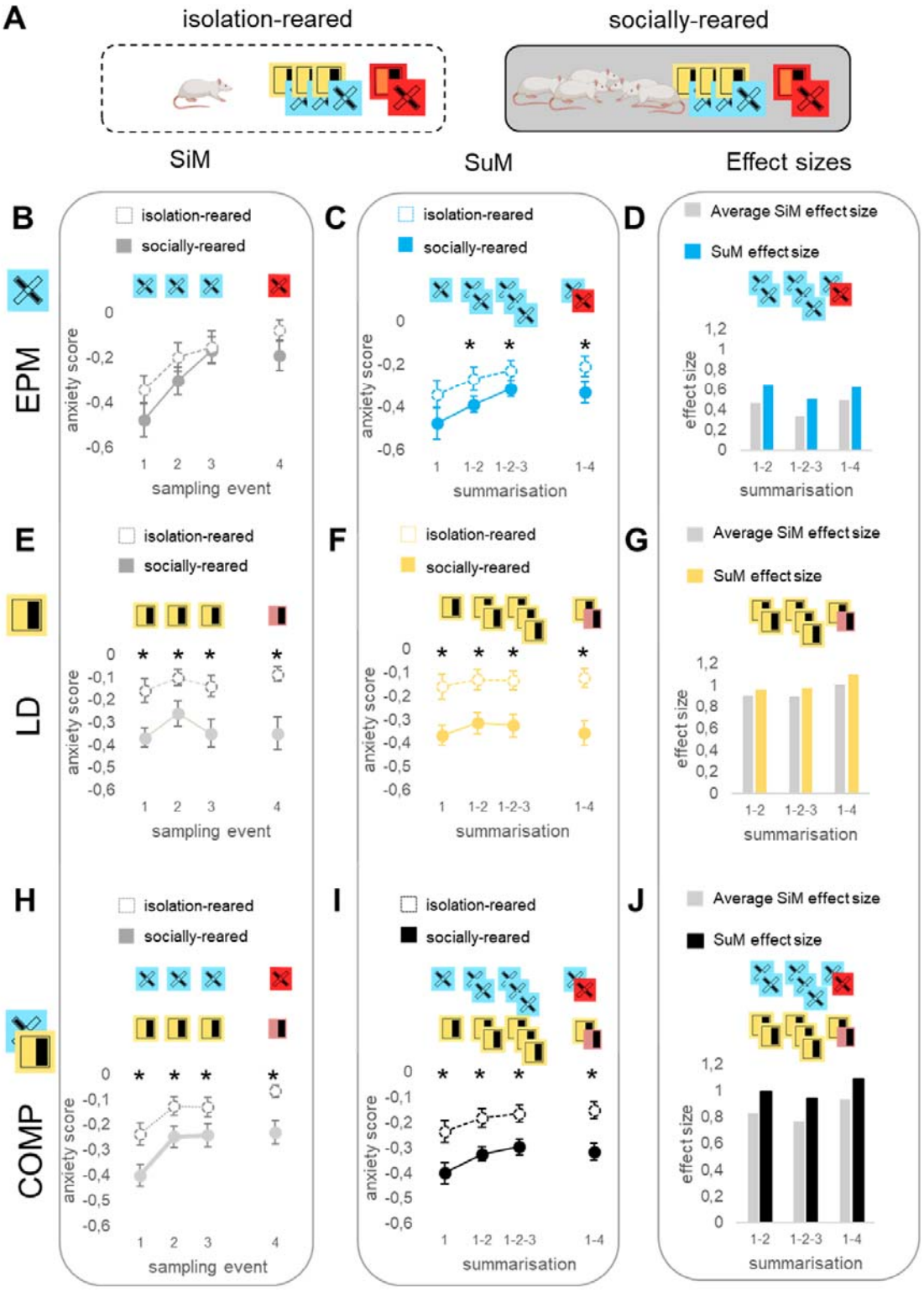
Figure 2. Inter-test correlations and behavioral predictions as a function of variable complexity. A) Inter-test correlations between increasingly complex variables of different anxiety tests. The correlation coefficients between OF and EPM (green), EPM and LD (red), and LD and OF (blue) were plotted at different variable combinations of these tests. The #n of test events and the #n of test variables refers to how many repeated test events (1 or 3 repeats) were included and to how many different variable types (time, frequency, latency) were used to create summary variables, respectively. SiMs are single variables, for example time spent in the open arms of an EPM, while SuMs consist of either multiple samplings of one variable type and/or multiple variable types measured in one or multiple events/repeats. **B)** Network plot of inter-test correlations between different tests revealed by SiMs (simplest model) and SuMs (reduced) in the rat (upper) and mouse (lower) experiments. **C) Behavioral predictions of anxiety measures.** 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. **D) Correlations between increasingly complex anxiety measures** (SiMs, SuMs, or COMPs) from the baseline sampling and their aversive counterparts in rats (left) and mice (right). **E)** Correlation coefficients were measured between baseline anxiety measures (SiMs, SuMs, COMPs), freezing in the fear conditioning paradigm, and startle in the acoustic startle paradigm. **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. **G) Correlation of increasingly 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. 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. All statistical results are presented in Figure 5A). Fig 2G) can be found in Table 1.

SuMs are more sensitive markers of chronic stress-induced behaviour

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 childhood neglect, a prevalent maltreatment that facilitates anxiety[40]. To evaluate the sensitivity of our approach in characterising isolation-induced anxiety, we reduced our protocol to three baselines and one aversive sampling with the EPM and LD tests following PWSI. In EPM, we found no isolation-related differences on any sampling days (Fig.3B). However, SuMs were higher following PWSI: two tests were necessary to show isolation-induced anxiety (Fig.3C). Since all LD and COMP variables were significantly increased by isolation (Fig.3E,3F,3H,3I), we also compared its effect sizes on these measures. 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, SuMs were more sensitive to isolation-induced anxiety than any SiMs (Fig.3D,3G,3J).

Figure 3



231
232

Figure 3.: **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 1.

SuMs trace out quantitatively and qualitatively different gene profiles compared to SiMs

Since SuMs represent robust measures of anxiety, we sought to investigate their molecular correlates too. We conducted an RNA-seq experiment on mPFC (exerting the top-down control of anxiety[41,42]) tissue of rats in a resting state. Samples from 27 subjects were chosen based on their anxiety levels. Briefly, we ranked the animals' COMP scores in ascending order, selecting every second animal for further analysis (Fig.4A). We found that SuMs outperformed SiMs in identifying gene targets, with LD SuMs and EPM SuMs increasing discoveries 3-fold and 8-fold, respectively (Fig.4B). Interestingly, LD-SiM associated genes largely overlap with LD-SuM, while EPM measures correlate with distinct gene sets (Fig.4C). Gene-behavior associations were further characterized by a robustness index showing that most of the molecular targets have a strong relationship with the expression of anxiety measures (Fig.4F).

The distinct gene discovery capabilities of different sampling methods prompted us to investigate their functional variability using the DAVID functional annotation tool for clustering gene discoveries. We found that SiMs correlated with metabolic and stress-related genes, while SuMs established a plasticity-focused gene profile (Fig.4D). The majority of genes associated with SuMs were also putatively involved in cellular localization, cellular transport, potassium signaling, or transcription. Plasticity-associated genes can be further structured to more specific functions, including but not limited to extracellular matrix shaping (ADAMTS4; HAPLN2), general modulation of neuronal development (FGF14; FAT3; FZD3), transcription (BTK; TCEB2; ASCL2; MDM4) or translation (BC1; EIFJ2; EIF2S2; EIF3J; 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 (Fig.4E). This suggests that our sampling approach not only quantitatively enhances gene discovery, but also yields a qualitatively distinct gene profile.

To confirm our RNASeq findings of differentially expressed genes with a differently sensitive method, we performed a qPCR analysis of the 43 most robust, and the 2 least robust protein-coding genes (Supplementary table 2). When comparing gene expressions between the RNASeq and qPCR data, 17.8% (8) of genes showed significant positive correlations (supplementary table 3). However, 94.6% of the genes that were not significantly correlated between the two different methods showed a fold change value lower than 1.2 (Fig.4F, Supplementary table 1) which is below the typical cut-off criteria for the qPCR analysis[43]. In addition, the remaining 2 non-correlating genes with a fold change higher than 1.2 showed a relatively low expression in the RNASeq and were not detectable with PCR. These were also the 2 least robust genes in our analysis (Supplementary table 2).

Furthermore, we have examined the differential expression of the aforementioned 45 genes in the amygdala, a key region in the regulation of anxiety, and found that after FDR correction, 15.6% (7) of them showed a significant correlation with anxiety (SiMs, SuMs, COMPs or behavior in an aversive environment) (Fig 4G, middle). When examining the correlations between amygdala and mPFC gene expressions measured by qPCR, we found that 17.8% (8) of genes showed a significant positive correlation between the two areas (Fig 4G, left). Furthermore, 6.7% (3) of the investigated amygdala genes showed significant correlations with both anxiety and mPFC gene expression. These genes were Glb1, Pla2g7 and Shbg. These findings suggest that some of the gene targets identified by our method may have a more general role in modulating trait anxiety.

Figure 4

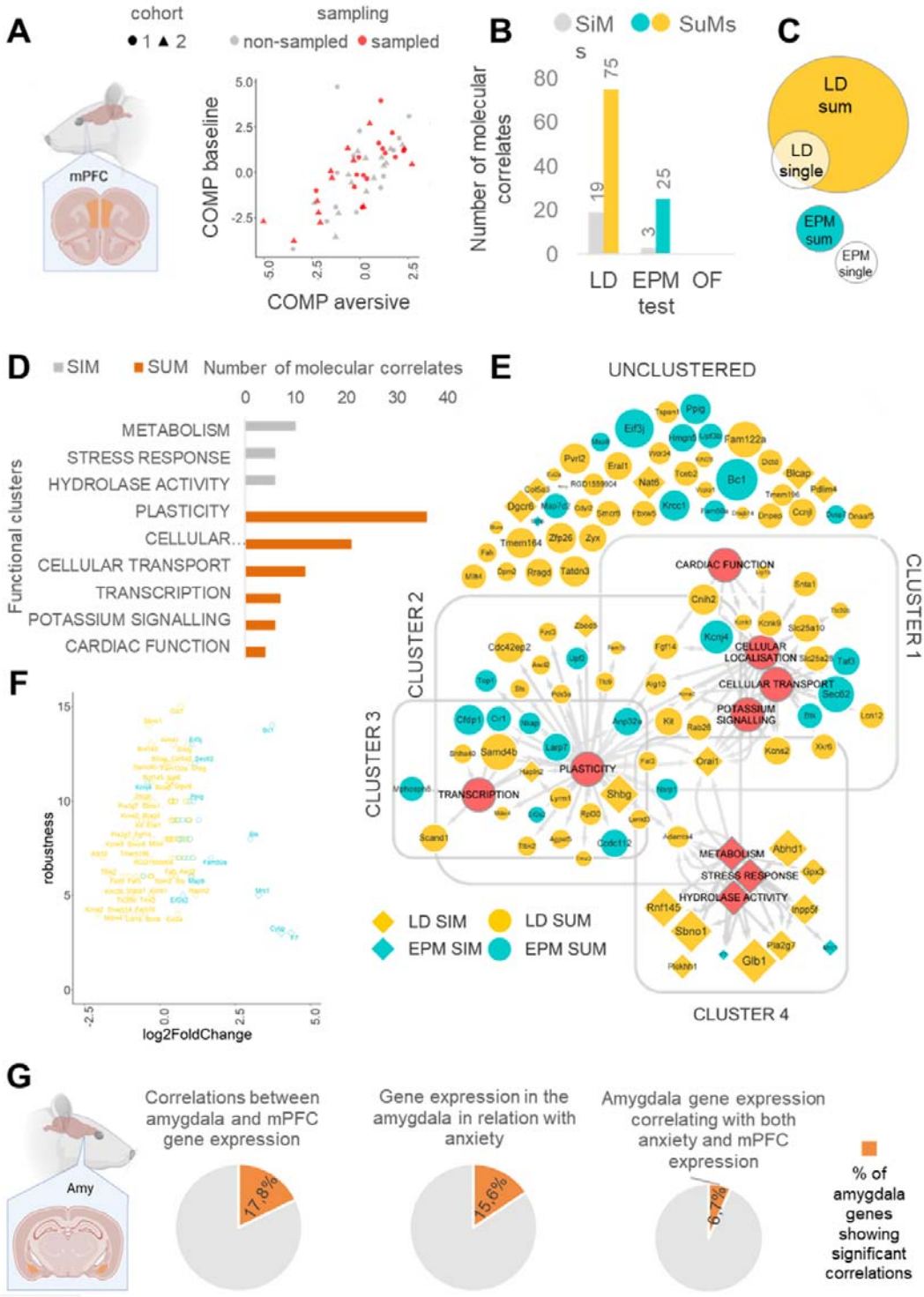

A
cohort
1 2
sampling
non-sampled
sampled
mPFC
COMP baseline
COMP aversive
B
SiMs
SuMs
Number of molecular correlates
LD
EPM
OF
test
C
LD sum
LD single
EPM sum
EPM single
D
SIM
SUM
Number of molecular correlates
Functional clusters
METABOLISM
STRESS RESPONSE
HYDROLASE ACTIVITY
PLASTICITY
CELLULAR...
CELLULAR TRANSPORT
TRANSCRIPTION
POTASSIUM SIGNALLING
CARDIAC FUNCTION
E
UNCLUSTERED
CLUSTER 1
CLUSTER 2
CLUSTER 3
CLUSTER 4
CARDIAC FUNCTION
CELLULAR LOCALISATION
CELLULAR TRANSPORT
POTASSIUM SIGNALLING
PLASTICITY
TRANSCRIPTION
METABOLISM
STRESS RESPONSE
HYDROLASE ACTIVITY
LD SIM
LD SUM
EPM SIM
EPM SUM
F
robustness
log2FoldChange
G
Amy
Correlations between amygdala and mPFC gene expression
17,8%
Gene expression in the amygdala in relation with anxiety
15,6%
Amygdala gene expression correlating with both anxiety and mPFC expression
6,7%
% of amygdala genes showing significant correlations

Figure 4. 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 **B)** SiM and SuM-associated gene count following FDR correction to multiple comparisons. **C)** Euler diagram representing the proportion of overlaps in gene sets of different sampling approaches. **D) SiM and SuM-associated genes form functionally distinct clusters.** Clusters contain 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. **E)** 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 includes genes that share many functional labels associated with cardiac function, cellular localization, transport, and potassium signalling. Clusters 2 and 3 are solely characterized by plasticity and transcription functions, respectively. Cluster 4 is characterized by metabolism, stress response, and hydrolase activity-associated functional terms and unites almost exclusively SiM-associated genes. Unclustered genes do not necessarily lack a known function, but do not have the amount of relatedness with enough genes in our pool to form a separate cluster. Gray arrows indicate gene-function associations, while gray lines indicate the borders of clusters. Red circles indicate cluster-determining labels, while color and shape indicate sampling type. The size of hubs is proportional to the correlation coefficient (absolute value) between the gene expression and anxiety. **F)** robustness values of genes as a function of their log2-fold change. **E) Correlations of amygdala gene expressions with PFC gene expression or anxiety variables**. Left piechart: percent of significant correlations between amygdala and mPFC qPCR results. middle: percent of significant correlations between amygdala gene expression and SiMs, SuMs, COMPs or aversive behavior. right: percent of amygdala genes that

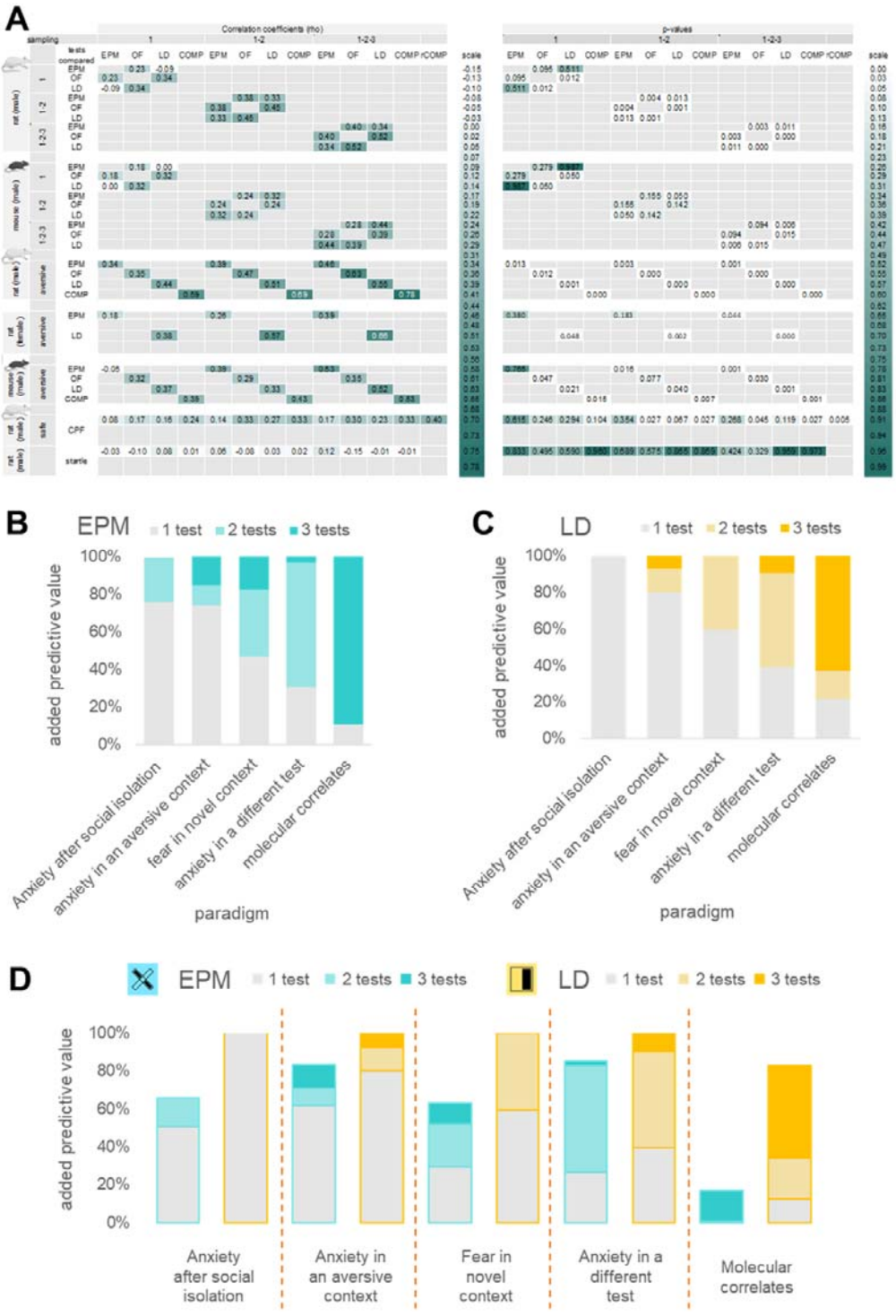
correlate with both anxiety and mPFC expression. All statistical results are presented in Table 1 and supplementary table 1 and 3.

A consensus score to measure trait anxiety

SuMs provide further predictive value compared to SiMs as a function of their complexity, e.g. the number of sampling points included in those. However, there is an inflection point in every paradigm where the addition of a further sampling point does not significantly enhance predictive value anymore. To offer a standardised approach to measure trait anxiety, we sought to determine a consensus inflection point by comparing the added predictive value of further testing in all paradigms of our study. In Figure 5, we show the improvement multiple testing offers in measured effect sizes of social isolation, correlations between similar anxiety tests under varying aversive conditions, correlations across different anxiety tests, or correlations between anxiety tests and generalized fear responses alongside molecular predictions from our RNASeq analysis. Values of different sampling depths (1, 2 or 3 tests used) are shown as a percentage of the highest prediction value that was achieved in each experiment with a particular test (Fig5B and C) or with any of the tests used (Fig5D). We found that the most easily predictable outcomes were the anxiety variables in aversive conditions, i.e. tests that were done in strong light or following social isolation. Here, a single use of either the EPM or the LD tests could explain above 80% of the maximum prediction value. However, a single EPM could not trace out the effect of isolation, in contrast to the more sensitive EPM SuMs, or an LD SiM that explained approximately the same amount of variability as multiple LD testing. Predicting different contexts was a bigger challenge as it was shown that significant correlations with anxiety in different tests or with freezing in the safe context of the CFP required multiple tests in most of the cases. Finally, finding molecular correlates of anxiety was the most challenging task, as the number of gene correlates continuously increased with every added test occasion, with the most complex SuM variables tracing out the highest number of gene targets in both tests. Finally, comparing the efficacy of the two tests by showing correlations as a percentage of the most accurate

predictions of all test variables (Fig.5D) shows that the LD test is superior to the EPM in every paradigm.

Figure 5



348

349

Figure 5. A) 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 white to teal represents the level of correlations or p-values scaled to all parameters and experiments. **B-C) Multiple testing in different paradigms.** Additional predictive value of further sampling points in different paradigms. Values are shown as the percentage of the highest prediction that was reached with a particular test in a particular paradigm. **D)** Additional predictive values are shown as a percentage of the highest prediction that was reached with any of the tests used in a particular paradigm.

Discussion

Our study aimed to model trait anxiety in rodents to enhance the translational potential of preclinical testing. We hypothesized that the frequently used preclinical anxiety tests measure state anxiety. Since the generally accepted definition of trait anxiety is the tendency of the subject to react with elevated state anxiety in situations[1], our approach to revealing trait anxiety was to repeat classical anxiety tests multiple times and introduce summary measures 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 beneficial to explain TA. We validated SuMs in male and female rats and mice as markers of TA by comparing their predictions to classical variables. Since anxiety-like behaviour, habituation, behavioural predictions based on SiM and SuM variables, test differences and test repeatability were similar in males and females, only males were used in the following experiments. We showed that, in proportion with their complexity, SuMs enhance inter-test correlations, strengthen predictions of behavior in threatening environments, and describe the consequences of chronic stress better.

The idea of repeating anxiety tests, especially the EPM has been a topic of debate due to the phenomenon of one-trial tolerance (OTT). OTT refers to two mechanically distinct phenomena; 1) the decreasing level of exploration in the EPM over repeated testing and 2) the decreasing sensitivity of the EPM to measure benzodiazepine (BDZ) effects in repeated testing[44]. The EPM in its original form was developed for assessing BDZ[11] effects, consequently, any experimental conditions that interfere with this purpose raised concerns about their validity. Therefore, the requirement of BDZ effects has limited the

representativity of the EPM test to very few aspects of trait anxiety. Indeed, there is a long list of conditions and drugs that fail to change the behavior expressed in the original EPM protocol, but anxiety-like effects become apparent in different testing conditions, like in more aversive environments, in response to prior stressful experiences, in a modified EPM or through the use of highly anxious animal strains[45–47]. We argue that while the carry-over effect is evident in EPM, additional information can be obtained through multiple testing, as we describe in detail below. Moreover, not all anxiety tests are burdened by a carry-over effect, as behavior in the LD test remains consistent through repetition.

We argue that SuMs represent 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[1,3,4,48,49]. 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 predictive power when increasing the complexity of 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[20–22,50], 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. The use of SuMs revealed a quantitatively and qualitatively different genetic profile from SiMs and

enhanced gene discoveries up to an 8-fold improvement. Higher discovery rates of SuMs indicate they 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 like 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[51].

Furthermore, genes detected with SuMs clustered into functionally different groups than SiMs indicating that besides having superior sensitivity, SuMs explain a previously undescribed type of variability. These previously undiscovered gene products are all potential candidates for the treatment of trait anxiety and thus anxiety disorders. We have also tested the robustness of the molecular findings, to further increase the translatability of our results.

SuM genes show high functional relatedness and are associated with neural plasticity and psychiatric disorders through different mechanisms and pathways. SuMs are positively correlated with the expression-levels of perineuronal net (PNN) and perinodal extracellular matrix (ECM) shaping molecules that influence membrane currents, excitability, and synaptic reorganization and were described in mental disorders[52]. ADAMTS4 is responsible for the cleavage of PNN elements[52,53], while HAPLN2 is an ECM composer that establishes the proper conductivity of neurons by stabilizing Ranvier nodes[52]. SuMs are negatively correlated with the trophic factor FGF14 that establishes proper excitability of neurons[54,55]. SuMs correlate positively with many general modulators of gene transcription and translation that control the expression of wide clusters of genes. Examples of this are Bruton tyrosine

kinase (BTK), the up-frameshift factors UPF2 and UPF3b[56], Elongin-B (TCEB2)[57], brain cytoplasmic RNA 1 (BC1)[58], or subunits of the Eukaryotic translation initiation factors EIF2s2, EIFJ2, and EIF3J[59,60]. Finally, we found multiple types of differently regulated potassium channels, which are direct modulators of membrane currents. 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. This data represents a range of promising candidates that could trace out novel interventions to treat anxiety.

To propose a versatile tool for measuring TA, we compared the predictive value of SiMs and SuMs of different tests in all the paradigms used in this study. We found that predicting behavior or molecular correlates varies in difficulty across paradigms. While we can get accurate predictions using a single test when predicting responses to an aversive environment, we need at least two tests for predicting behavior in a different context, and all three tests when searching for molecular correlates of anxiety. None of the variables, regardless of their complexity were able to predict acoustic startle, which shows that SuMs predicting future anxious states stronger than SiMs is not due to technical factors, but rather due to SuMs predicting specific emotional domains. Our analysis showed that the LD test outperformed the EPM in every paradigm, with a single LD test providing as much insight as multiple EPM tests, while two LD tests achieved 96% of the most accurate predictions. Additionally, LD SuMs identified four-times more anxiety-related genes than any EPM variable. Based on these findings, we recommend conducting at least two LD tests with an inter-test interval of one day, as an effective and precise approach for measuring rodent trait anxiety.

In summary, here we provide a novel sampling and analysis pipeline that enables us to measure TA in preclinical research. Our paradigm is based on an extended sampling process

using popular anxiety tests, consequently making it applicable to any model and feasible to any laboratory. We argue that the correlation between the complexity of sampling and the amount of information gained of the anxious phenotype of animals is robust and representative to various conditions. Such correlation was confirmed by applying multiple testing conditions: in baseline and aversive testing environment, in multiple test types, in naïve as well as chronic stress-exposed animals, and through using different cohorts, species and sexes of subjects. In line with this, the molecular profiling of TA in female rodents is a crucial further step to make due to potential differences at the level of the mPFC[61]. 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, SiMs reveal a distinct, plasticity-focused gene profile associated with trait anxiety. We encourage colleagues to adopt this detailed phenotyping approach, potentially bridging the gap between preclinical and clinical anxiety research.

Materials and Methods

Detailed methods are provided in the supplementary materials.

Animals. Six cohorts of animals were used throughout this study. After arrival at the facility, all cohorts were left to adapt to the housing conditions for 2 weeks. Two cohorts of adult male Wistar rats (Charles-River, n=54, age= 100 days) 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=27, age= 130 days), and adult male C57BL/6J mice (Jackson Laboratory, n=40, age= 95 days). One cohort of adult male Long-Evans rats (Charles-River, n=27, age= 110 days) 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, age at testing= 180 days) was used to investigate the effects of social isolation on trait anxiety. Sample sizes for the first rat experiment (n=54) were calculated using the pwr R package aiming to measure a minimum r=0.35 spearman linear correlation coefficient at alpha=0.05 significance level and beta=0.8 power. Following the first experiment we decreased sample sizes according to the robustness indices of behavior-behavior and behavior-gene expression correlations.

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[14,24] (OF), and Light-Dark test[12,13] (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 the experiments described above, animals were randomly chosen for each group. For anxiety tests, time spent in the aversive zone, frequency, and latency of aversive zone entries were analyzed automatically using Noldus EthoVision XT1541, head-dips (HD) and rearing (R) were analyzed manually using Solomon Coder[24] by experimenters that were blind to experimental groups.

Behavioral 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. SiMs and SuMs were measured in the same animals. 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. Means and standard errors are reported for behavioural measures.

Statistical analysis was done in R statistical environment[25]. Principal component analysis ('factoextra' package[26]) 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[27]) was used. The analysis of test-repeatability ('rptr' package[28]) was done for each test-type as described previously[29], by calculating the proportion of within-test variance divided by total variance[30]. False discovery rate (FDR) correction for multiple comparison testing was done using the Benjamini-Hochberg method[31].

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). Single-end RNA sequencing was performed on an Illumina NextSeq500 sequencer. Regression analysis to reveal differentially expressed genes (DEGs) was done using the R Bioconductor package DESeq2[32]. 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[31]. Only genes with FDR corrected p-values below 0.05 were considered to show significant differential

expression. Functional and pathway analysis of detected significant DEGs was done through the online functional annotation tool DAVID[33,34]. Connections were visualized using Cytoscape[35]. Cluster-determining functions, merging and associated genes are shown in Supplementary table 2.

qPCR analysis. We performed qPCR analysis of total RNA from mPFC and AMY. For mPFC, the RNA sample isolated for RNASeq was used for further analysis. Bilateral microdissection of the amygdala was performed as described for mPFC tissue. Gene expression patterns were analyzed using custom 384-well TaqMan Gene Expression Array Cards (Applied Biosystems, USA). These cards measured the expression of 45 genes of interest, which were selected based on the robustness of mPFC RNA sequencing experiments. Inter-run calibration values were calculated based on the run control samples, using Actb and Gapdh as reference genes, as per the method described by Hellemans[36]. The normalized relative gene expression values were calculated using the 2-ΔΔCT method[37]. These values were further normalized to the inter-run calibration values. Expression values with threshold cycles greater than 35 were excluded from the analyses.

Robustness. We characterised the stability of behaviour-behaviour and behaviour-gene expression correlations with the robustness index. The robustness index gives the smallest number of randomly excludable samples that likely impair the significance of such a relationship. The index is calculated by a custom-written script that reanalyses the Spearman correlations in randomly selected, different sample size sub-cohorts of the original population. We made 20000 correlations on all possible sample size sub-cohorts and calculated minimums, maximums, standard deviations and means of their p-values. We determined the cutoff criteria of a robustness index as the smallest sample size in which the

standard deviation of the correlations' p-values reaches the 0.05 alpha level. The function in R is available upon request.

Data Availability Statement

Raw count values of RNA sequencing data and the corresponding behavioural data is publicly available at the Zenodo online data repository through the accession number and doi 10.5281/zenodo.14236344.

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.

Conflict of Interest Statement

The authors declare no conflict of interest.

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. & Ballachey, E. L. A study of the rat's behavior in a field. A contribution to method in comparative psychology. *Univ. Calif. Publ. Psychol.* **6**, 1–12 (1932).
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. 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. *et al.* Behavioral, Physiological, and Neuroendocrine Stress Responses and Differential Sensitivity to Diazepam in Two Wistar Rat Lines Selectively Bred for High- and Low-Anxiety–Related Behavior. *Neuropsychopharmacology* **19**, 381–396 (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. Hall, C. S. Emotional behavior in the rat. III. The relationship between emotionality and ambulatory activity. *J. Comp. Psychol.* **22**, 345–352 (1936).
25. RCoreTeam. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2019).
26. Kassambara, A. & Mundt, F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. (2019).
27. Harrell, J. F. Hmisc: Harrell Miscellaneous. (2023).
28. 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).
29. 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).
30. 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).
31. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
32. 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).
33. 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).
34. 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.

35. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
36. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19 (2007).
37. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3**, 1101–1108 (2008).
38. Mikics, É., Barsy, B., Barsvári, B. & Haller, J. Behavioral specificity of non-genomic glucocorticoid effects in rats: Effects on risk assessment in the elevated plus-maze and the open-field. *Horm. Behav.* **48**, 152–162 (2005).
39. Morey, R. A. *et al.* Fear learning circuitry is biased toward generalization of fear associations in posttraumatic stress disorder. *Transl. Psychiatry* **5**, e700–e700 (2015).
40. Dubowitz, H. & Bennett, S. Physical abuse and neglect of children. *The Lancet* **369**, 1891–1899 (2007).
41. Tovote, P., Fadok, J. P. & Lüthi, A. Neuronal circuits for fear and anxiety. *Nat. Rev. Neurosci.* **16**, 317–331 (2015).
42. Calhoon, G. G. & Tye, K. M. Resolving the neural circuits of anxiety. *Nat. Neurosci.* **18**, 11 (2015).
43. Everaert, C. *et al.* Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci. Rep.* **7**, 1559 (2017).
44. File, S. E. & Zangrossi, H. ‘One-trial tolerance’ to the anxiolytic actions of benzodiazepines in the elevated plus-maze, or the development of a phobic state? *Psychopharmacology (Berl.)* **110**, 240–244 (1993).
45. Sorregotti, T., Mendes-Gomes, J., Rico, J. L., Rodgers, R. J. & Nunes-de-Souza, R. L. Ethopharmacological analysis of the open elevated plus-maze in mice. *Behav. Brain Res.* **246**, 76–85 (2013).
46. Haller, J. & Alicki, M. Current animal models of anxiety, anxiety disorders, and anxiolytic drugs: *Curr. Opin. Psychiatry* **25**, 59–64 (2012).

47. Haller, J., Baranyi, J., Bakos, N. & Hal□sz, J. Social instability in female rats: effects on anxiety and buspirone efficacy. *Psychopharmacology (Berl.)* **174**, (2004).
48. Eysenck, M. W., Derakshan, N., Santos, R. & Calvo, M. G. Anxiety and cognitive performance: Attentional control theory. *Emotion* **7**, 336–353 (2007).
49. 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).
50. 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).
51. Lewis, D. I. Animal experimentation: implementation and application of the 3Rs. *Emerg. Top. Life Sci.* **3**, 675–679 (2019).
52. 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).
53. 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).
54. Lou, J.-Y. *et al.* Fibroblast growth factor 14 is an intracellular modulator of voltage-gated sodium channels. *J. Physiol.* **569**, 179–193 (2005).
55. Hoxha, E. *et al.* Emerging roles of Fgf14 in behavioral control. *Behav. Brain Res.* **356**, 257–265 (2019).
56. Huang, L. *et al.* A Upf3b-mutant mouse model with behavioral and neurogenesis defects. *Mol. Psychiatry* **23**, 1773–1786 (2018).
57. 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).
58. Briz, V. *et al.* The non-coding RNA BC1 regulates experience-dependent structural plasticity and learning. *Nat. Commun.* **8**, 293 (2017).

59. Huang, W. *et al.* Translational control by eIF2α phosphorylation regulates vulnerability to the synaptic and behavioral effects of cocaine. *eLife* **5**, e12052 (2016).
60. Placzek, A. N. *et al.* eIF2α-mediated translational control regulates the persistence of cocaine-induced LTP in midbrain dopamine neurons. *eLife* **5**, e17517 (2016).
61. Laine, M. A., Greiner, E. M. & Shansky, R. M. Sex differences in the rodent medial prefrontal cortex – What Do and Don't we know? *Neuropharmacology* **248**, 109867 (2024).

Figure /result	Hypotheses testing / model	Comparison	Test statistics-value	n	P-value	Further information
Fig.1B	repeated measures ANOVA / rat male main effect	EPM	F=34.206	n=54	p=3.277e-12	
		OF	F=3.643		p=0.029	
		LD	F=0.801		p=0.452	
	aversive contrast paired t-test	EPM 1 vs 4 (aversive)	F=8.007		p=1.094e-11	
		EPM 3 vs 4 (aversive)	F=2.898		p=0.005	
		OF 1 vs 4 (aversive)	F=2.894		p=0.005	
		OF 3 vs 4 (aversive)	F=4.043		p=0.0001	
		LD 1 vs 4 (aversive)	F=8.007		p=0.753	
		LD 3 vs 4 (aversive)	F=0.356		p=0.722	
	repeated measures ANOVA / rat female main effect	EPM	F=3.419	n=27	p=0.040	
		LD	F=0.038		p=0.963	
	aversive contrast paired t-test	EPM 1 vs 4 (aversive)	t=6.152		p=1.105e-07	
		EPM 3 vs 4 (aversive)	t=4.431		p=4.883e-05	
		LD 1 vs 4 (aversive)	t=0.594		p=0.555	
		LD 3 vs 4 (aversive)	t=0.675		p=0.503	
	repeated measures ANOVA / mouse main effect	EPM	F=64.944	n=40	p=2.2e-16	
		OF	F=8.386		p=0.001	
		LD	F=1.411		p=0.250	
	aversive contrast paired t-test	EPM 1 vs 4 (aversive)	t=9.806		p=4.939e-14	
		EPM 3 vs 4 (aversive)	t=1.192		p=0.237	
		OF 1 vs 4 (aversive)	t=1.774		p=0.081	
		OF 3 vs 4 (aversive)	t=-0.289		p=0.773	
		LD 1 vs 4 (aversive)	t=2.468		p=0.016	

		LD 3 vs 4 (aversive)	t=1.054		p=0.295	
Fig. 2A	Linear-mixed model's	complexity	Est=0.019; SE=0.005	n=54	p=0.0008; t=3.597	variable composition: p=0.0065
						test composition: p=0.0088
						Residual: 0.0042
Fig. 2G	ANOVA	freezing	F=7.000	n=28	p=0.001	
	Tukey post-hoc	conditioned context (ctx) vs safe ctx1	Est=7.69; SE=3.42		p=0.067	
		cond ctx vs safe ctx2	Est=12.71 ; SE=3.42		p=0.001	
		safe ctx1 vs safe ctx2	Est=5.02; SE=3.42		p=0.309	
Fig. 3B	Wilcoxon rank sum test	EPM 1	w=86	n=15-15	p=0.143	FDR: non-significant
		EPM 2	w=74		p=0.056	
		EPM 3	w=97		p=0.263	
		EPM 4(aversive)	w=34		p=0.023	
Fig. 3C	Wilcoxon rank sum test	EPM 1-2	w=37		p=0.023	FDR: all significant
		EPM 1-2-3	w=36		p=0.034	
		EPM 1-4(aversive)	w=34		p=0.020	
Fig. 3E	Wilcoxon rank sum test	LD 1	w=41		p=0.001	FDR: all significant
		LD 2	w=50		p=0.004	
		LD 3	w=42		p=0.001	
		LD 4 (aversive)	w=43		p=0.002	

Fig. 3F	Wilcoxon rank sum test	LD 1-2	w=38		p=0.001	FDR: all significant
		LD 1-2-3	w=43		p=0.002	
		LD 1-4(aversive)	w=33		p=0.000	
Fig. 3H	Wilcoxon rank sum test	COMP 1	w=52		p=0.006	FDR: all significant
		COMP 2	w=52		p=0.006	
		COMP 3	w=64		p=0.023	
		COMP 4(aversive)	w=39		p=0.001	
Fig. 3I	Wilcoxon rank sum test	COMP 1-2	w=37		p=0.001	FDR: all significant
		COMP 1-2-3	w=36		p=0.000	
		COMP 1-4(aversive)	w=34		p=0.000	
Fig.4 D	ANOVA (**robustness**)	Test type main effect	F=0.0128	n=27	p=0.920	
		Sampling type main effect	F=4.8248		p=0.030	
		Test:sampling interaction	F=21.9427		p=7.581e-06	
	ANOVA (logFoldChange)	Test type main effect	F=18.352		p=3.765e-05	
		Sampling type main effect	F=14.830		p=0.0002	
		Test:sampling interaction	F=27.217		p=7.874e-07	
supp Fig.4 A	Wilcoxon rank sum test	EPM baseline vs aversive	w=1512.5	n=54	p=0.006	
		OF baseline vs aversive	w=1813		p=1.281e-06	

		LD baseline vs aversive	w=1229		p=0.575	
supp Fig.4 B	ANOVA	corticosterone	F=44.225	n=28	p=1.041e-13	
	Tukey post-hoc	aversive vs resting 1	Est=426; SE=50.6		p=0.0001; t=8.407	
		aversive vs resting 2	Est=398; SE=50.6		p=0.0001; t=7.854	
		Resting 1 vs resting 2	Est=-28; SE=50.6		p=0.845; t=0.554	

Table 1.: Parameters of test statistics. Table 1 shows (left-to-right) the result section/figure number corresponding to the statistical parameters, the hypotheses testing method/model that were used, the comparisons that were made, and the resulting test statistics-values, sample sizes, and significance levels. We presented the conventional parameters of the particular statistical approaches such as F values in the case of ANOVAs, t or w values in the case of t-tests or Wilcoxon tests, as well as estimates (est) and standard errors (SE) in the case of linear mixed models. The column entitled Further information 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 comparisons.