

Article

Multimodal Behavior Scoring Quantifies Depression-Like Severity Across Chronic Stress Models and Identifies Stress-Resilient Mice

Yue Tu ^{1,2}, Qinghui Fu ^{1,2}, Canghao Sun ^{1,2}, Jing Deng ^{1,2}, Haoning Qin ^{1,2}, Xianlin Zeng ^{1,2}, Yuchuan Wang ^{1,2}, Shulan Qiu ^{1,2,*} and Weixiong Zhang ^{1,2,3,*}

¹ Hong Kong Jockey Club STEM Laboratory for Genomics and AI in Healthcare, Hong Kong, China

² Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, China

³ Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong, China

* Correspondence: shulan.qiu@polyu.edu.hk (S.Q.); weixiong.zhang@polyu.edu.hk (W.Z.)

Abstract

Rodent models, especially chronic stress paradigms, are essential in depression research, yet existing behavioral assessments often struggle to capture pathophysiological heterogeneity. We developed the Multimodal Behavior Scoring (MBS) algorithm, which integrates results from standard behavioral tests (sucrose preference, open field, and forced swim tests) into a single severity metric. Validated in two distinct chronic stress models (CUMS and CSDS), MBS reliably distinguished stressed from control mice and demonstrated high cross-cohort reproducibility. Furthermore, MBS successfully stratified stress-resilient subpopulations and revealed model-specific phenotypic patterns. By providing an integrative, biologically grounded framework, MBS enhances the resolution of chronic stress-induced behavioral phenotyping and may support future biomarker discovery and pharmacological validation studies.

Keywords: multimodal phenotyping; depression-like behavior algorithm; behavioral resilience stratification; cross-stress reliability; translational neuroscience

1. Introduction

Major depressive disorder (MDD) is a debilitating global health concern, affecting approximately 5% of adults worldwide and representing a leading cause of disability [1,2]. Despite decades of research, progress in developing novel antidepressants has stalled, with numerous clinical trials failing despite promising preclinical findings [3,4]. A major translational obstacle is the limited capacity of animal models to capture the multifaceted human depressive syndrome and the lack of objective, integrative quantification.

Rodent models, particularly mice subjected to chronic stress paradigms, are widely used in depression research. These models recapitulate core behavioral endophenotypes of depression—including anhedonia, despair, and social withdrawal—and share conserved underlying neurobiological mechanisms involving monoaminergic system disturbances, hypothalamic–pituitary–adrenal (HPA) axis dysregulation, and neuroplasticity deficits [5–8]. Among these, the Chronic Unpredictable Mild Stress (CUMS) paradigm induces depression-like states via varied, low-intensity stressors [9,10], whereas the Chronic Social Defeat Stress (CSDS) model effectively simulates the profound impact of

Academic Editor: Xerxes D. Arsiwalla

Received: 4 June 2026

Revised: 11 July 2026

Accepted: 14 July 2026

Published: 15 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

psychosocial trauma [11,12]. Both models support mechanistic studies, therapeutic screening, and the study of resilience among stress-exposed individuals.

Currently, depression-like behaviors in these models are quantified primarily through established behavioral assays: the Sucrose Preference Test (SPT) [13] measures anhedonia, the Open Field Test (OFT) [14] evaluates anxiety and locomotor activity, and the Forced Swim Test (FST) [15] and the Tail Suspension Test (TST) [16] assess behavioral despair [17]. Despite their widespread use, these assessments face significant limitations. Critically, depression manifests as a heterogeneous array of symptoms [18], yet conventional assays evaluate these modalities in isolation (e.g., SPT for anhedonia, FST for despair), lacking a comprehensive integrative measure of overall severity. This underscores the need for integrated, quantitative metrics capturing overall severity [19]. Additionally, manual scoring introduces subjectivity and inter-rater variability [20], while inherent behavioral noise and low sensitivity often obscure subtle phenotypic differences, particularly in mild cases or during recovery phases [21,22]. Furthermore, existing metrics are inadequate in predicting individual susceptibility, resilience, or longitudinal disease progression [23–25]. Reliable identification of stress-resistant individuals within exposed cohorts remains challenging but essential [26,27]. Variability in test performance and interpretation across different stress models [28,29] also raises concerns regarding cross-model validity and reproducibility. The lack of standardized, objective, and integrative assessment contributes significantly to reproducibility issues across laboratories or even experiments [30].

To address these issues, we developed Multimodal Behavior Scoring (MBS), a robust algorithm that integrates key metrics from established behavioral tests into a single continuous severity score, enabling a more holistic and objective assessment of depression-like behavior. The framework is flexible, allowing inclusion or exclusion of individual modalities (such as the TST) depending on experimental objectives. We validated MBS in both CUMS and CSDS models, evaluating cross-model reliability and generalizability, and demonstrated its utility for tracking progression, recovery, and recurrence after stress exposure and cessation. Using MBS, we identified stress-resistant mice that maintain normal behavioral function despite chronic stress exposure, enabling investigation of resilience mechanisms. In addition, we evaluated the Tail Suspension Test (TST) as an optional fourth modality to determine whether it provided incremental information beyond the shared SPT–OFT–FST core in the present chronic-stress phenotyping framework. Because TST is widely used in antidepressant-like pharmacological screening, including studies involving SSRIs, we treated it as an important comparator rather than as an excluded or invalid assay. The primary MBS implementation in this study was based on SPT, OFT, and FST to preserve cross-model comparability across CUMS and CSDS cohorts, while retaining the modular capacity to incorporate TST or other paradigm-specific behavioral measures when appropriate.

Collectively, we present a validated computational approach for integrating multimodal behavioral data into a unified quantitative measure of depression severity. By addressing fragmented and subjective assessments, MBS provides a standardized tool to enhance reproducibility, facilitate identification of stress-resistant phenotypes, and help bridge translational gaps in depression research.

2. Materials and Methods

2.1. Animals

C57BL/6J male mice (4–6 weeks, 18–22 g) were obtained from the Centralized Animal Facility (CAF) of The Hong Kong Polytechnic University (PolyU). All animal procedures were approved by the Animal Subjects Ethics Sub-committee of PolyU (No. 23-24/889-HTI-R-STG, No. 24-25/1083-HTI-R-STG) and conducted in accordance with the guidelines of the Department of Health of the Hong Kong SAR.

Mice were housed in barrier facilities under controlled conditions (22 ± 1 °C; $60 \pm 10\%$ humidity) with ad libitum access to food and water, and maintained on a 12:12 h light/dark cycle (lights on 8:00–20:00) unless stated otherwise.

2.2. Study Design and MBS Workflow

We employed two complementary yet mechanistically distinct stress paradigms: chronic unpredictable mild stress (CUMS), which models chronic environmental adversity through randomized low-intensity stressors, and chronic social defeat stress (CSDS), which models psychosocial trauma via resident–intruder confrontations. These paradigms capture different aspects of depressive etiologies and together provide a more comprehensive evaluation of depressive phenotypes, enabling delineation of shared and model-specific vulnerability markers and supporting characterization of stress-resistant phenotypes (Figure 1A).

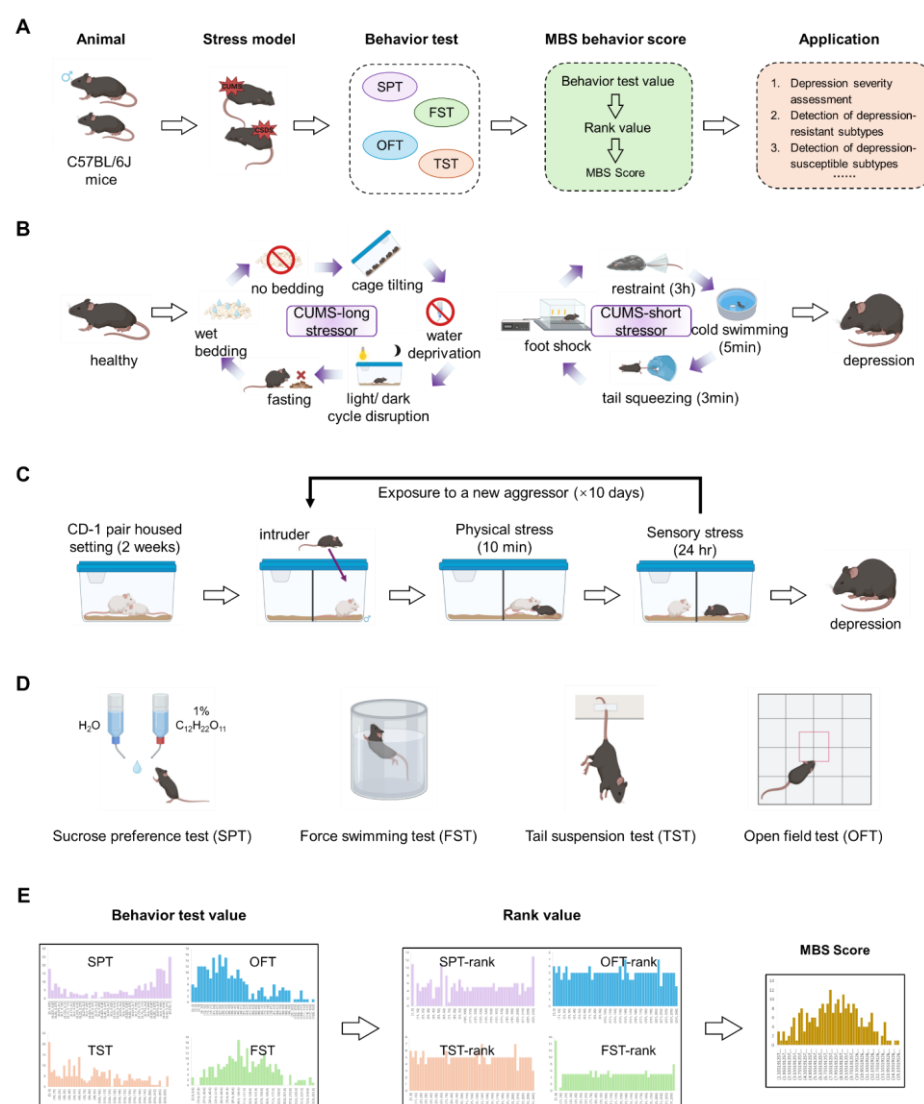


Figure 1. Overview of the methodological framework. (A) Flowchart depicting animal experiments and behavioral assessment procedures. (B) Chronic Unpredictable Mild Stress (CUMS) model. (C) Chronic Social Defeat Stress (CSDS) model. (D) Schematic representation of behavioral tests. (E) Principle of the Multimodal Behavior Scoring (MBS) algorithm.

The overall experimental design established a unified pipeline for multimodal behavioral phenotyping. CUMS (Figure 1B) and CSDS (Figure 1C) cohorts underwent a standardized behavioral battery (Figure 1D), and behavioral readouts were integrated by

the Multimodal Behavior Scoring (MBS) algorithm to generate a continuous severity score and support subsequent validation and phenotypic stratification. The computational architecture of MBS (Figure 1E) applies cohort-specific nonparametric rank transformation followed by linear integration of multimodal inputs, providing an analytical foundation for downstream analyses, including identification of stress-resilient and stress-susceptible subtypes and longitudinal tracking following stress cessation. The numbers of animals included in each cohort and analysis are specified in the corresponding figure legends and subgroup descriptions. In the principal cross-sectional analyses, the cohorts comprised CUMS control mice ($n = 54$) and CUMS-exposed mice ($n = 68$), as well as CSDS control mice ($n = 25$) and CSDS-exposed mice ($n = 43$). For the longitudinal CUMS experiment shown in Section 3.5, the same control cohort and the same stress-exposed cohort were followed across T1, T2, and T3, so each assessment window included contemporaneous control and stress-exposed mice of the same age and testing history.

2.3. Animal Models of Depressive-Like Behavior

2.3.1. Chronic Unpredictable Mild Stress (CUMS)

To induce depression-like behaviors, mice were subjected to a 4-week CUMS protocol with minor adaptations [31–33]. Each day, animals received one long-duration and one short-duration stressor, randomly selected from a predefined list to ensure unpredictability. Long-duration stressors included 24 h food or water deprivation, 16 h cage tilting (45°), 16 h strobe light exposure, 24 h housing without bedding, and 24 h housing with wet bedding. Short-duration stressors included 5 min cold water swim (4 °C, which is distinct from the forced swim test), 3 min tail pinch, 2 h restraint stress, and 10 min foot shock. Stressors varied daily in type and timing to prevent habituation and simulate chronic psychosocial stress.

2.3.2. Chronic Social Defeat Stress (CSDS)

The CSDS model was performed for 10 consecutive days using a standard resident–intruder paradigm. Male CD-1 mice (4–6 months old; Laboratory Animal Service Center, The Chinese University of Hong Kong) served as resident aggressors, and male C57BL/6J mice (6–7 weeks old) served as intruders. Aggressors were screened 3 days before CSDS to select mice with adequate aggressiveness.

One day before the first defeat session, selected aggressors were acclimated overnight in one side of a perforated, divided cage. Each day, a novel C57BL/6J intruder was introduced into the aggressor compartment for a 5–10 min defeat episode, then moved to the opposite compartment to allow continuous sensory contact through the divider for the next 24 h. A new aggressor was used each day.

Sessions were monitored to ensure consistent aggression (≥ 1 aggressive bout/min; each bout 5–10 s). Aggressors displaying affiliative behaviors (e.g., grooming) were excluded and replaced. Control mice were housed in identical divided cages with another control mouse but without physical interaction.

2.4. Behavioral Testing Procedures

To assess depression- and anxiety-like behaviors following chronic stress, mice underwent the sucrose preference test (SPT), open field test (OFT), forced swim test (FST), tail suspension test (TST), and the social interaction test (SIT; CSDS only). Except for SPT, all assays were conducted under dim light in a quiet testing room after 1–2 h habituation. Sessions were video-recorded and analyzed using Any-Maze (Stoelting, IL, USA) unless otherwise specified.

SPT was performed as previously described [34,35] to assess anhedonia. Mice were habituated to two bottles of 1% sucrose for 24 h, followed by two bottles of water for 24 h, with bottle positions switched every 12 h. After 24 h food and water deprivation, mice were

given one bottle of water and one bottle of 1% sucrose for 2 h starting at 20:00, with positions exchanged after 1 h. Sucrose preference was calculated as sucrose intake/total fluid intake.

OFT assessed locomotion and exploration in a 0.4 m × 0.4 m × 0.4 m opaque chamber. Mice were placed in a corner and recorded for 6 min, excluding the first minute from analysis. The arena was cleaned with 75% ethanol between trials. Time spent in the center and periphery was quantified.

FST assessed behavioral despair in a transparent cylinder (15 cm diameter, 30 cm height) filled with 23–25 °C water to 18 cm depth. Mice were tested for 6 min, and immobility time was quantified; animals were dried before returning to home cages.

TST assessed despair-like behavior by suspending mice via tail tape (~1 cm from the tip) with a rubber tube to prevent climbing. Mice were suspended 20 cm above the floor for 6 min, and immobility during the last 4 min was analyzed.

For CSDS mice, SIT was conducted 24 h after the final defeat to assess social avoidance. Mice were tested in a 0.44 m × 0.44 m × 0.38 m arena containing a wire-mesh enclosure (0.1 m × 0.065 m × 0.38 m) across two 2.5 min trials (target absent, then target present with a novel CD-1), separated by 30 s. Time in the interaction zone was recorded, and the social interaction ratio was calculated as target-present/target-absent interaction time.

All behavioral testing and analyses were conducted blindly. Sufficient recovery time was allowed between assays, and test order was counterbalanced across groups to minimize sequence effects.

2.5. Multimodal Behavior Scoring (MBS)

To quantitatively integrate heterogeneous behavioral metrics and overcome the modal fragmentation inherent to classical behavioral tests, we developed the Multimodal Behavior Scoring (MBS). MBS is a deterministic algorithm that projects individual-level behavioral data into a unified latent severity space, designed to capture depressive load across multiple modalities. The method operates via three sequential stages: metric transformation, rank space embedding, and normalized integrative projection.

Let $\mathcal{X} = \{x_i^{(j)} \in \mathbb{R} : i = 1, \dots, k; j = 1, \dots, m\}$ denote the raw behavioral data matrix, where k is the number of animals and m the number of behavioral features (e.g., SPT, OFT, FST, TST). Each variable $x_i^{(j)}$ represents the observed behavioral outcome for mouse i in test j . These features differ in range, directionality, and distribution, rendering them non-comparable in raw form.

We define a polarity function $\pi: \{1, \dots, m\} \rightarrow \{-1, 1\}$ such that

$$\pi(j) = \begin{cases} 1 & \text{if higher values indicate lower depression (e.g., SPT)} \\ -1 & \text{if higher values indicate higher depression (e.g., FST)} \end{cases} \quad (1)$$

Each variable is transformed into a relative rank space via:

$$r_i^{(j)} = \text{RANK}(\pi(j) \cdot x_i^{(j)}) \quad (2)$$

where $\text{RANK}(\cdot)$ denotes the position of the value in the sorted list of transformed scores within modality j , with rank 1 assigned to the most severe behavioral phenotype.

Let $\mathbf{r}_i = [r_i^{(1)}, r_i^{(2)}, \dots, r_i^{(m)}] \in \mathbb{N}^m$ be the rank vector of mouse i .

Definition (MBS Score) The raw score for subject i is completed as the mean normalized rank across selected modalities:

$$S_i = \frac{1}{m} \sum_{j=1}^m \frac{r_i^{(j)}}{k} \quad (3)$$

To normalize this to a bounded range and enhance comparability across experiments, we define the final MBS score as

$$\text{MDRS}_i = \gamma \cdot S_i \quad (4)$$

where $\gamma > 0$ is a tunable scaling constant (we set $\gamma = 16$ in practice), such that $MDRS_i \in \left[\frac{\gamma}{k}, \gamma\right]$.

This mapping ensures interpretability and comparability across experimental batches, regardless of cohort size or test modality. The MBS construction exhibits several desirable theoretical properties that support its use as a core metric in preclinical behavioral neuroscience.

The MBS framework, though algorithmically simple, possesses a set of well-defined mathematical properties that support its theoretical validity and practical robustness. These properties ensure the stability, interpretability, and extensibility of the score across varying behavioral inputs and experimental conditions. Below, we formally characterize several key invariance and distributional features of MBS.

Lemma 1 (Permutation Invariance). *Let $\sigma: \{1, \dots, k\} \rightarrow \{1, \dots, k\}$ be any permutation of the animal index. Then, the distribution of $\{MDRS_i\}$ is invariant under σ . That is, MBS scores depend solely on relative behavioral positioning, not animal identity.*

Lemma 2 (Scale Invariance). *Let behavioral feature j be subject to an affine transformation $x_i^{(j)} \leftarrow ax_i^{(j)} + b$ for all i , with $a > 0$. Then, the MBS scores remain unchanged. Thus, MBS is robust to unit scaling and shifts in measurement (e.g., converting immobility time from seconds to z-scores).*

Proposition 1 (Asymptotic Normality). *Let the entries $x_i^{(j)}$ be independently and continuously distributed. Then, by the Lyapunov central limit theorem, the distribution of $MDRS_i$ converges to a Gaussian as $m \rightarrow \infty$ and/or $k \rightarrow \infty$, facilitating parametric statistical inference.*

Remark 1 (Modality Modularity). *MBS is explicitly modular in the choice of behavioral modalities m , allowing for the dynamic inclusion or exclusion of specific tests. For instance, a 3D variant (SPT–OFT–FST) can be directly compared to a 4D variant (including TST) by rescaling via γ and using shared rank embedding. In the present study, the core MBS framework was intentionally based on the behavioral modalities shared by both CUMS and CSDS (SPT, OFT, and FST), thereby preserving direct cross-model comparability while retaining the flexibility to incorporate paradigm-specific assays, such as SIT, in future model-adapted implementations.*

The selection of behavioral modalities in MBS should be guided by the experimental objective rather than by an assumption that one behavioral assay is universally superior to another. In the present study, SPT, OFT, and FST were used as the shared core modalities because they were consistently available across both CUMS and CSDS cohorts and captured reward-related, exploratory/anxiety-related, and coping-related behavioral dimensions. TST was retained and analyzed as an optional fourth modality in the CUMS cohort to evaluate its incremental contribution to MBS-based severity ranking [36–39]. For pharmacological studies, particularly those designed to evaluate antidepressant-like activity or SSRI responsiveness, TST can be incorporated directly into the MBS framework as an additional modality [39,40].

This rank-based formulation provides robustness to outliers, skewed distributions, and inter-test heteroskedasticity—issues that commonly undermine raw-score-based composite indices. Furthermore, since all features are rank-normalized, their contribution to the final MBS score is equal by construction, regardless of their absolute variance or dynamic range.

In practical application across both CUMS and CSDS models, we used MBS to calculate depressive severity scores from behavioral data collected from standard tests. The resulting MBS distributions were approximately Gaussian, and animals with extreme scores could be reliably identified as highly stress-susceptible or stress-resilient. Notably, MBS allowed us to detect sub-threshold or compensatory behavioral patterns that were

not apparent in any individual test, offering a more nuanced phenotype stratification than categorical classification or test-by-test significance testing.

By embedding multimodal behavioral readouts into a uniform, interpretable space, MBS provides a scalable and theory-grounded alternative to conventional behavioral indices. It not only improves statistical power in the face of noisy and heterogeneous data but also facilitates the phenotyping of intermediate states, the detection of resilient individuals, and the reproducible tracking of behavioral recovery across models and time points. Thus, MBS should be regarded primarily as a continuous multimodal severity measure, whereas subsequent designation of resilient or susceptible subgroups represents an operational analytical stratification step rather than an intrinsic binary property of the score itself. The thresholds used in the present study were derived from the distribution of integrated behavioral trajectories and guided by cluster-level separation between control-like and clearly impaired phenotypic patterns, rather than being intended as universal biological cutoffs.

2.6. Statistical Analysis

All analyses were performed in GraphPad Prism 9.0. Data are presented as mean $\pm$ SEM unless otherwise stated. Two group comparisons (e.g., control vs. stress) were conducted using unpaired Welch's *t*-tests. Normality was assessed using Shapiro–Wilk tests; nonparametric tests were considered but not required. In Figures 2–5 and Supplementary Figures S1–S5, significance thresholds were $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***), not significant (ns). For analysis involving multiple simultaneous inferences within the same result family, Bonferroni-adjusted *p* values were reported within the defined family-wise testing set. Specifically, the endpoint-level comparisons in Figure 3A–D,H–K were adjusted within their respective four-comparison families, whereas the correlation analyses in Figure 3E–G,L–N, the pairwise phenotype contrasts in Figure 4, and the longitudinal time-point contrasts in Figure 5E–H were adjusted within their respective three-comparison families. Isolated preplanned contrasts that did not belong to a multi-test family were reported without additional adjustment. Associations between MBS and individual behavioral measures were evaluated using Pearson's correlation (*r*). No animal or data points were excluded unless predefined artifact criteria (e.g., incomplete trials or equipment error) were met. For the longitudinal dataset, the group differences reported at each assessment window were interpreted relative to contemporaneously evaluated control mice of the same age and testing history, whereas the individual trajectories were displayed to illustrate temporal evolution across repeated observations. Because the longitudinal measurements were obtained from the same mice across T1, T2, and T3, within-subject dependence was additionally analyzed using a linear mixed-effects model with group, time, and group $\times$ time as fixed effects and mouse identity as a random intercept. The preplanned directional between-group comparisons at each time point were retained as complementary pointwise contrasts.

3. Results

Building on the experimental design and MBS workflow (Figure 1), we evaluated the robustness and interpretability of MBS across independent CUMS and CSDS cohorts. We first tested whether rank transformation improves cross-cohort comparability (Section 3.1), then assessed MBS sensitivity and cross-model consistency in quantifying depression-like severity (Section 3.2). We next applied MBS to phenotype stratification to delineate susceptibility and resilience (Section 3.3), examined the contribution of TST to multimodal scoring (Section 3.4), and finally demonstrated longitudinal tracking of recovery and recurrence following stress cessation and re-exposure (Section 3.5).

3.1. Rank Transformation Enhances Cross-Cohort Behavioral Comparability

The integration of multimodal behavioral data across cohorts requires robust normalization to mitigate inherent variability. MBS addresses this by applying cohort-specific rank transformation, converting heterogeneous metrics (SPT, OFT, and FST) into uniform percentile distributions. In both CUMS (Figure 2A–C) and CSDS (Supplementary Figure S1A–C), rank-transformed values retained strong monotonic relationships with the original measures (Pearson $r > 0.92$, $p < 0.001$), indicating preservation of phenotypic information during rescaling. Importantly, rank transformation generated stable uniform distributions that neutralized baseline shifts across cohorts, enabling cross-cohort integration within a common metric space.

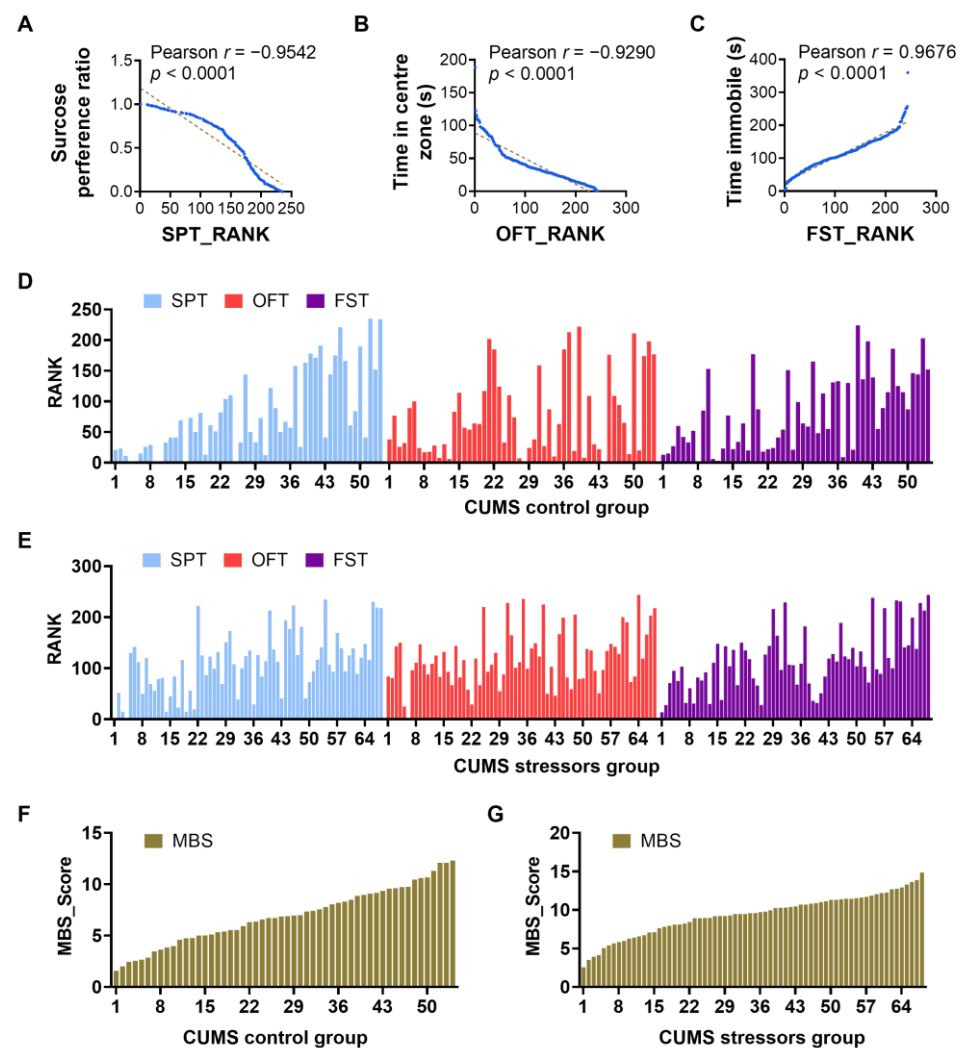


Figure 2. Rank transformation enables robust cross-cohort integration of behavioral data in the CUMS model. (A) Correlation between raw and rank-transformed scores in the sucrose preference test (SPT). (B) Correlation between raw and rank-transformed scores in the open field test (OFT). (C) Correlation between raw and rank-transformed scores in the forced swim test (FST). In (A–C), the yellow dashed lines represent the linear fit, which were calculated using least-squares regression. (D) Rank-based trajectories of individual behavioral metrics (SPT, OFT, FST) sorted by ascending MBS scores in CUMS control group. (E) Rank-based trajectories of individual behavioral metrics (SPT, OFT, FST) sorted by ascending MBS scores in CUMS stressors group. (F) Ordered bar chart of MBS in CUMS control group. (G) Ordered bar chart of MBS in CUMS stressors group.

Ordering mice along the MBS severity continuum further revealed close alignment between composite scoring and individual behavioral modalities. When stratified by ascending MBS values (CUMS: Figure 2F,G; CSDS: Supplementary Figure S1F,G), rank-transformed behavioral trajectories (Figure 2D,E; Supplementary Figure S1D,E) showed striking covariation with MBS. Individual-level analyses similarly demonstrated coordinated directional shifts in SPT, OFT, and FST ranks within MBS-defined subgroups (Supplementary Figures S2 and S3), supporting coherent multimodal integration without distorting phenotypic profiles.

Together, these findings establish rank transformation as the key mathematical step that suppresses batch-specific artifacts while preserving neurobehavioral signals, forming the basis for subsequent phenotype discovery and cross-model validation.

3.2. MBS Quantifies Depression Severity with High Sensitivity and Cross-Model Consistency

The MBS approach demonstrated exceptional discriminative power on CUMS cohorts. Stressed mice exhibited profound depression-like states characterized by a 39.8% increase in MBS scores compared to controls (9.070 ± 2.581 vs. 6.489 ± 2.445 ; $p < 0.001$, Figure 3A). This composite severity metric captured coordinated deterioration across core behavioral domains: sucrose preference (SPT) declined by 19.6% ($58.47\% \pm 0.3292$ vs. $72.73\% \pm 0.3222$ in controls; $p < 0.05$, Figure 3B), reflecting established anhedonia phenotypes; exploratory activity in the open field (OFT) diminished by 50.3% (28.49 ± 20.91 s center time vs. 57.35 ± 37.91 s in controls; $p < 0.001$, Figure 3C), indicating heightened anxiety; and immobility duration in forced swim tests (FSTs) increased by 32.8% (130.4 ± 51.74 s vs. 98.20 ± 50.17 s in controls; $p < 0.01$, Figure 3D), demonstrating behavioral despair. Critically, MBS scores maintained biologically coherent correlations with these core modalities, showing significant inverse associations with reward sensitivity (SPT: $r = -0.5898$, $p < 0.001$, Figure 3E) and exploratory drive (OFT: $r = -0.3967$, $p < 0.001$, Figure 3F), while exhibiting strong positive correlation with behavioral despair (FST: $r = 0.6077$, $p < 0.001$, Figure 3G). This integrated quantification outperformed isolated behavioral results in phenotypic resolution. Our CUMS mice were derived from two cohorts, with the individual results of MBS, SPT, OFT, and FST for each cohort shown in Figure S4.

On the etiologically distinct CSDS model, MBS similarly detected robust depressive phenotypes while revealing unique pathophysiological features. Mice subjected to chronic social defeat developed severe symptoms reflected in significantly elevated MBS scores (10.39 ± 2.086 vs. 4.684 ± 2.079 in controls; $p < 0.001$, Figure 3H). Detailed behavioral decomposition uncovered a distinct expression profile: sucrose consumption decreased to $71.41\% \pm 0.2114$, representing a 17.7% reduction versus controls ($p < 0.001$, Figure 3I); anxiety-related behaviors manifested as a 71.4% decline in OFT center exploration time (15.04 ± 16.30 s vs. 52.52 ± 25.86 s in controls; $p < 0.001$, Figure 3J); whereas behavioral despair exhibited comparatively moderate intensification with FST immobility increasing 55.8% (144.7 ± 64.52 s vs. 92.85 ± 58.32 s in controls; $p < 0.01$, Figure 3K). The correlation architecture of CSDS models diverged significantly from CUMS: social stress manifested stronger integration of anxiety-related behaviors (OFT: $r = -0.7397$, $p < 0.001$, Figure 3M), whereas unpredictable stress emphasized behavioral despair (FST: $r = 0.6077$ in CUMS vs. $r = 0.3674$ in CSDS, Figure 3G vs. Figure 3N). Similarly, the individual results for each cohort of CSDS-treated mice are presented in Figure S5.

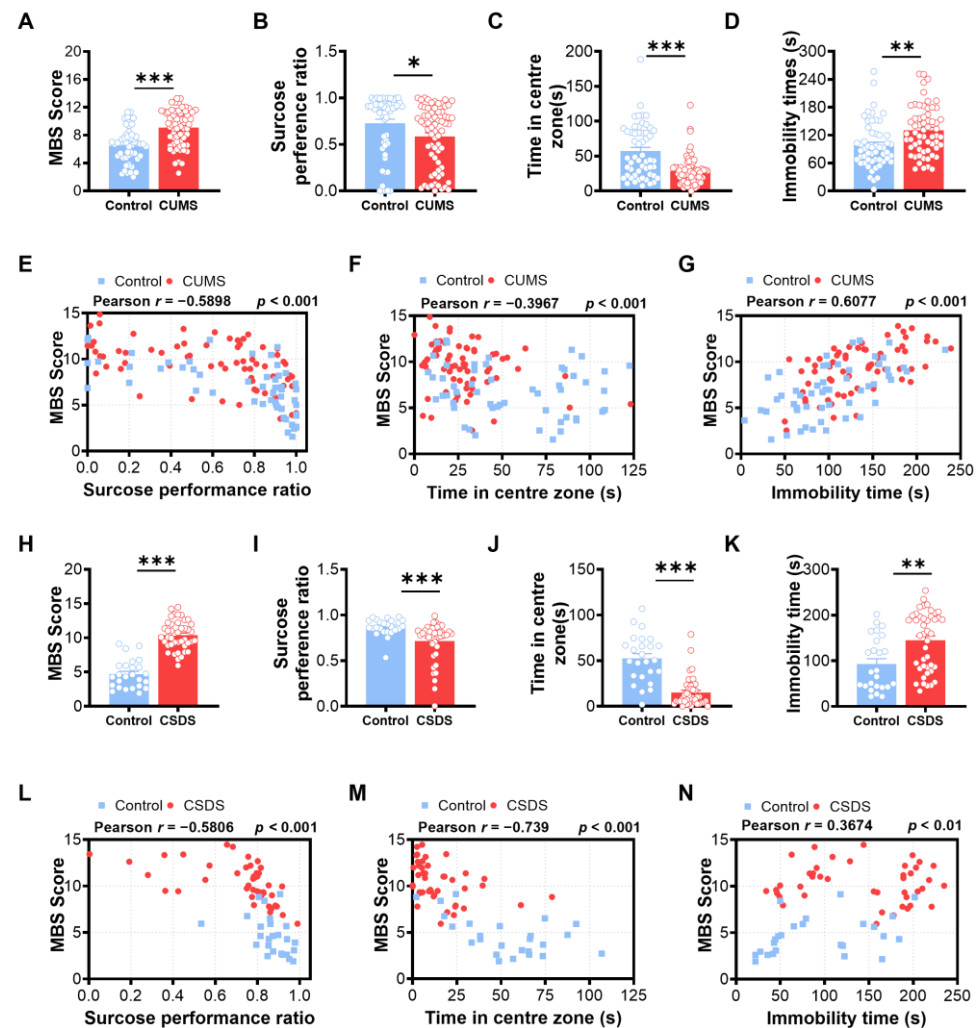


Figure 3. MBS quantifies depression severity with high sensitivity and stress model-specific behavioral integration. (A) Elevated MBS scores in CUMS-stressed mice indicate pronounced depression-like states. (B) Reduced SPT in CUMS model reflects anhedonia. (C) Decreased center exploration time in the OFT indicates heightened anxiety in CUMS mice. (D) Increased immobility time in the FST suggests enhanced behavioral despair in CUMS mice. (E) Negative correlation between MBS scores and sucrose preference (SPT) in the CUMS model. (F) Negative correlation between MBS scores and center time (OFT) in the CUMS model. (G) Positive correlation between MBS scores and immobility duration (FST) in the CUMS model. (H) Significantly elevated MBS scores in CSDS-stressed mice reflect robust depressive phenotypes. (I) Decreased sucrose preference in the CSDS model indicates anhedonia. (J) Reduced OFT center exploration in CSDS mice reflects severe anxiety-like behavior. (K) Increased FST immobility duration in the CSDS model reflects moderate behavioral despair. (L) Negative correlation between MBS scores and sucrose preference (SPT) in the CSDS model. (M) Strong negative correlation between MBS scores and OFT performance in the CSDS model. (N) Positive but attenuated correlation between MBS scores and FST immobility in the CSDS model. CUMS panels (A–G): control $n = 52$, CUMS $n = 65$. CSDS panels (H–N): control $n = 25$, CSDS $n = 43$. For panels (A–D,H–K), displayed p values are Bonferroni-adjusted within each four-comparison behavioral endpoint family. For panels (E–G,L–N), displayed p values are Bonferroni-adjusted within each three-comparison correlation family.

Cross-model comparative analysis revealed fundamental principles governing depression-like behavior expression across stress models. While both models exhibited conserved anhedonia pathology (SPT-MBS correlations: $r = -0.5898$ in CUMS vs. $r = -0.5806$

in CSDS; $|\Delta r| < 0.01$), their phenotypic architectures diverged in stressor-specific manifestations: CUMS produced despair-dominant phenotypes characterized by stronger FST-MBS associations ($r = 0.6077$ vs. $r = 0.3674$ in CSDS; $|\Delta r| = 0.2403$), whereas CSDS generated anxiety-centered pathophysiology showing enhanced OFT-MBS integration ($r = -0.7397$ vs. $r = -0.3967$ in CUMS; $|\Delta r| = 0.3430$). This differential sensitivity profile demonstrates MBS's capacity for intrinsic model adaptation—automatically amplifying contributions from dominant behavioral modalities without manual parameter adjustment—a critical feature for modeling heterogeneous psychiatric conditions. Because MBS is mathematically derived from SPT, OFT, and FST, these correlations should be interpreted primarily as evidence of internal coherence between the composite score and its constituent dimensions rather than as fully independent external validation. In addition, integrating MBS with independent readouts, such as molecular signatures, circuit-level measures, pharmacological responsiveness, or prospective prediction of later phenotypes, may provide an important direction for further extending the biological interpretability of this framework.

Robust reproducibility was confirmed through four independent biological replicates (Supplementary Figures S2 and S3, encompassing two CUMS cohorts and two CSDS cohorts). MBS maintained exceptional measurement consistency across cohorts, with inter-replicate intraclass correlation coefficients (CUMS: ICC = 0.88; CSDS: ICC = 0.93) substantially exceeding established reliability thresholds for preclinical research (ICC > 0.80 [41]). Crucially, the distinctive correlation topology characterizing each stress model remained stable across cohorts ($|\Delta r| < 0.12$ for equivalent modalities), confirming MBS's ability to preserve pathophysiological signatures despite experimental batch effects.

3.3. Phenotypic Stratification Delineates Susceptibility and Resilience

To integratively assess depressive behaviors and identify stress-resistant phenotypes, we established phenotypic stratification criteria based on multimodal behavioral profiles. Control mice (CON) were defined as unstressed animals with consistently low MBS scores (< 7). Within stress-exposed cohorts, resilient mice (RES) maintained control-like performance with MBS < 7 , whereas susceptible mice (SUS) showed maladaptive responses with elevated MBS scores (> 10). These thresholds were derived from cluster analysis of integrated behavioral trajectories, optimized to discriminate phenotypic categories while remaining consistent with established depressive endophenotypes. Accordingly, the interval between these bounds is better understood as an intermediate transition zone, in which animals may exhibit partial or mixed adaptive responses rather than a sharply defined categorical state. This design was intended to enrich for clearly preserved and clearly impaired phenotypes in downstream analyses, while retaining the flexibility to tighten or relax the thresholds according to cohort distribution and the desired level of selection stringency.

CUMS exposure produced heterogeneous responses across domains (Figure 4A–D). Resilient mice were indistinguishable from controls across core measures (MBS: RES 5.230 ± 1.315 vs. CON 4.878 ± 1.642 , $p > 0.05$; SPT: $p > 0.05$; FST: $p > 0.05$), whereas susceptible mice showed markedly increased severity (MBS 11.80 ± 1.140 ; $p < 0.001$ vs. CON and RES), severe anhedonia (SPT: SUS $41.23\% \pm 5.3$ vs. RES $79.06\% \pm 0.2177$, $p < 0.001$), and increased despair (FST: SUS 167.0 ± 54.83 s vs. RES 92.27 ± 26.70 s, $p < 0.001$). OFT measures suggested a mild anxiety-related profile in resilient mice (OFT: RES 35.68 ± 33.84 s vs. CON 62.57 ± 38.28 s, $p > 0.05$).

CSDS similarly revealed pronounced stratification (Figure 4E–H). Susceptible mice exhibited elevated MBS (SUS 10.88 ± 1.981) relative to controls (CON 4.124 ± 1.482 , $p < 0.001$) and resilient mice (RES 5.954 ± 0.5867 , $p < 0.001$). Resilient mice showed no significant impairment in sucrose preference, while anxiety-like behavior displayed a graded reduction in OFT center time from CON (57.75 ± 22.65 s) to RES (26.77 ± 11.89 s) to SUS (14.53 ± 16.10 s) (CON–SUS: $p < 0.001$; RES–CON: $p < 0.001$; SUS–RES: $p > 0.05$). FST immobility increased significantly only

in susceptible mice (SUS 162.1 ± 56.18 s vs. CON 88.71 ± 56.45 s, $p < 0.001$), whereas resilient mice were intermediate and not significantly different from either group.

Across paradigms, resilience profiles were model-specific: CUMS resilience emphasized preservation of reward and coping-related behaviors with partial retention of anxiety-related vigilance, whereas CSDS resilience was more strongly associated with normalized anxiety responses alongside partially retained passive coping.

Stratification was further validated against the social interaction test (SIT) in CSDS (Figure 4I–L). MBS-defined resilient mice showed comparable social approach to SIT-defined resilient mice (Figure 4M,N, SIT ratio: MBS-RES 0.6744 ± 0.3192 vs. SIT-RES 0.7261 ± 0.2342 , $p > 0.05$). Notably, MBS-RES represented a refined subset of SIT-RES with 37.5% overlap (Figure 4O), and mice classified as resilient only by SIT exhibited residual anxiety-related impairments. In the present study, SIT therefore served as a conventional CSDS benchmark rather than a mandatory element of the shared MBS core. Notably, the SIT-based and MBS-based classifications showed broadly comparable performance in distinguishing the resilient subgroup, while MBS additionally retained a continuous and extensible multimodal framework that can be adapted to include model-specific features when desired.

These results demonstrate that the integrative MBS approach robustly distinguishes stress-resilient and susceptible mice, effectively addressing behavioral heterogeneity, and providing a powerful framework for studying resilience mechanisms in depression research.

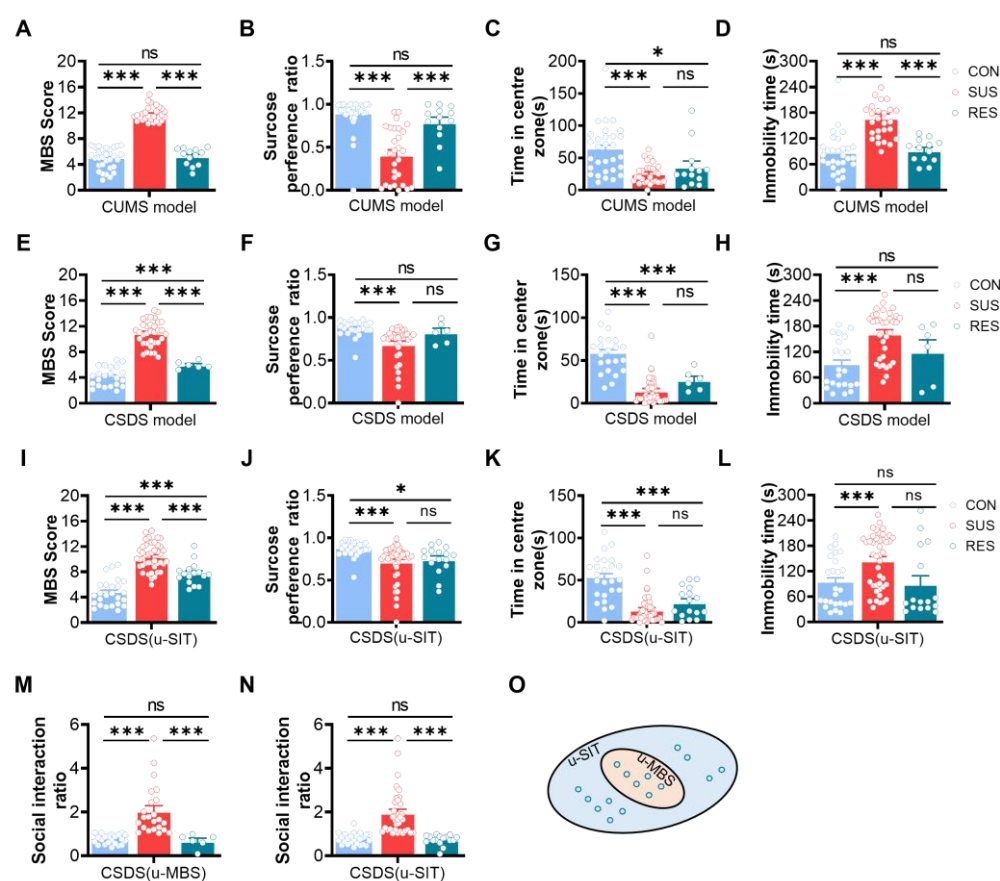


Figure 4. Phenotypic stratification using MBS distinguishes stress-susceptible and resilient individuals across depression models. CUMS model: (A) MBS scores differentiate control, resilient, and susceptible mice in the CUMS model. (B) SPT is preserved in resilient mice and reduced in susceptible mice. (C) OFT center time shows partial anxiety regulation in resilient mice and strong deficits in susceptible mice. (D) FST immobility duration is significantly increased in susceptible mice. CSDS model: (E) MBS-based stratification identifies control, resilience, and susceptible mice in the CSDS model. (F) Resilient mice maintain sucrose preference at control-like levels. (G) OFT performance

reveals a stepwise gradient of anxiety severity across phenotypes. (H) FST immobility is elevated only in susceptible mice, with resilient mice showing intermediate behavior. Validation with SIT: (I) SIT ratios of MBS-defined phenotypes show agreement with traditional SIT classifications. (J) MBS-defined resilient mice exhibit social behavior like SIT-defined resilient mice. (K) SIT-identified resilient mice with high MBS scores show residual behavioral impairments. (L) Scatter plot comparing SIT ratios with MBS scores across individuals. Model comparison and refinement: (M) Behavioral comparison of mice identified as resilient by utilizing MBS method. (N) Behavioral comparison of mice identified as resilient by utilizing SIT method. (O) Venn diagram showing overlap between MBS- and SIT-defined resilient subpopulations. CUMS panels (A–D): control $n = 30$, RES $n = 13$, SUS $n = 28$. CSDS panels (E–H): control $n = 22$, RES $n = 6$, SUS $n = 34$. For multi-group comparisons in Figure 4, reported pairwise p values are Bonferroni-adjusted within the corresponding three-comparison family.

3.4. TST-Inclusive Sensitivity Analysis Supports the Modularity of the MBS Framework

The Tail Suspension Test (TST) is a classical assay widely used to evaluate immobility-related coping behavior and antidepressant-like pharmacological activity in mice. Because the present MBS framework was primarily designed for chronic stress-induced severity stratification rather than antidepressant drug screening, we analyzed TST as an optional fourth modality to determine whether it provided incremental information beyond the shared SPT–OFT–FST core.

In the CUMS cohort, TST immobility was increased in stress-exposed mice compared with controls (89.81 ± 65.12 s vs. 70.09 ± 57.37 s; $p = 0.0391$, Figure 5A), indicating that TST captured a stress-related behavioral change. However, TST showed only a weak association with the core three-modality MBS score ($r = 0.09849$, $p = 0.2804$, Figure 5B), suggesting that, in this dataset, TST contributed limited additional information to the integrated chronic-stress severity axis.

We next compared the core three-modality MBS score with a four-modality MBS score incorporating TST. The two scores were highly concordant in both control and CUMS-exposed mice (control: $r = 0.8662$, $p < 0.0001$; CUMS: $r = 0.8836$, $p < 0.0001$; Figure 5C,D), indicating that the main individual-level severity ranking was preserved when TST was added. Because the four-modality score necessarily shares three components with the three-modality score, this analysis should be interpreted as a sensitivity analysis of score robustness rather than as independent validation.

These findings do not imply that TST is generally inferior to FST or that TST should be excluded from depression-related behavioral research. Rather, they indicate that, within the specific context of C57BL/6J chronic-stress severity stratification in this study, TST did not substantially alter the MBS-derived phenotypic structure. The modular design of MBS allows TST to be included when the experimental objective involves antidepressant-like pharmacological screening, SSRI responsiveness, or other contexts in which TST provides independent information [42,43].

3.5. MBS Dynamics Track Stress-Induced Recovery Trajectories

Our longitudinal study delineates a dynamic progression of stress-induced depressive phenotypes through sequential induction, recovery, and recurrence phases (Figure 5E–H). When subjected to a 4-week CUMS scheme (T1, age: 2.5 months), mice developed robust depressive-like behaviors evidenced by significantly elevated MBS scores relative to controls (10.39 ± 1.055 vs. 3.838 ± 1.623 ; $p < 0.001$). Following a 2.5-month stress-free recovery period at 5.5 months of age (T2), behavioral deficits were virtually abolished with no statistically distinguishable differences between groups (6.388 ± 1.826 vs. 5.499 ± 3.058 ; $p = 0.2788$), indicating phenotypic remission. Then, upon re-exposure to CUMS for a condensed 2-week rechallenge period (T3), previously stressed mice exhibited accelerated

symptom recurrence compared to control mice that had not been previously stressed, manifesting as moderately elevated MBS scores (10.49 ± 2.783 vs. 6.794 ± 3.244 ; $p = 0.09$). This temporal trajectory—characterized by initial susceptibility, full behavioral recovery, and preferential vulnerability to restress—demonstrates the capacity of MBS to detect distinct neuroadaptive transitions. Importantly, the longitudinal group differences at T1, T2, and T3 were evaluated against a contemporaneously followed control cohort of the same age, rather than being interpreted in the absence of age-matched reference animals. To account for within-subject dependence in the repeated-measures cohort, we additionally performed a linear mixed-effects analysis with mouse identity as a random intercept. This confirmed significant longitudinal effects for MBS, including main effects of group ($\chi^2(1) = 26.72$, $p < 0.0001$) and time ($\chi^2(2) = 19.99$, $p < 0.0001$), as well as a significant group $\times$ time interaction ($\chi^2(2) = 14.64$, $p = 0.0007$), supporting the induction-remission-rechallenge pattern described above. OFT also showed a significant group $\times$ time interaction ($\chi^2(2) = 9.13$, $p = 0.0104$), whereas SPT and FST did not show significant longitudinal interactions (SPT: $\chi^2(2) = 4.26$, $p = 0.1185$; FST: $\chi^2(2) = 1.14$, $p = 0.5659$).

Serial assessment at critical junctures revealed persistent pathophysiological alterations beneath phenotypic normalization. Longitudinal profiling of individual mice (Figure 5I) uncovered that despite group-level recovery at T2, CUMS subjects maintained elevated recurrence sensitivity during restress. The differential vulnerability was quantified through intragroup comparisons: the T1-to-T2 transition showed a dramatic improvement ($p < 0.001$), while the T2-to-T3 deterioration occurred with steeper kinetic progression ($p = 0.007$) relative to primary induction. More profoundly, baseline severity at T1 predicted restress susceptibility at T3 ($r = 0.8602$, $p < 0.001$), suggesting that MBS captures latent pathophysiology independent of overt symptom expression.

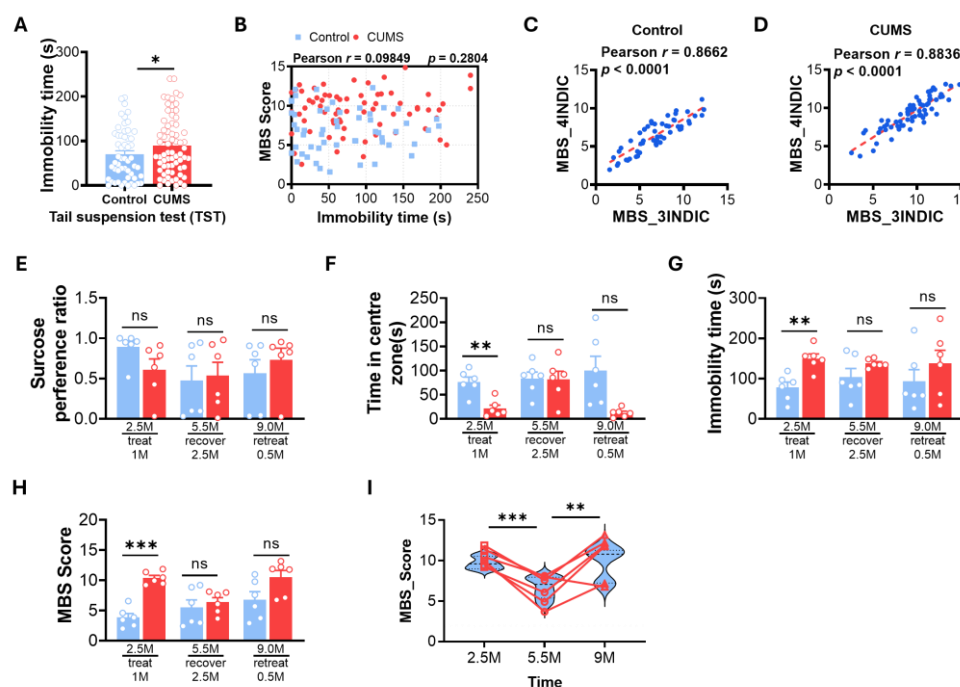


Figure 5. TST-inclusive sensitivity analysis and dynamic tracking of stress-induced behavioral trajectories via MBS. (A) TST immobility duration shows a modest increase in CUMS-exposed mice. (B) Association between TST immobility and the core three-modality MBS score. (C) Concordance between four-modality MBS (SPT, OFT, FST, TST) and core three-modality MBS (SPT, OFT, FST) in control mice. (D) Concordance between four-modality and core three-modality MBS in CUMS mice, supporting robustness of MBS-based severity ranking when TST is included as an optional modal-

ity. (E) SPT of Longitudinal trajectories. (F) OFT of Longitudinal trajectories. (G) FST of Longitudinal trajectories. (H) MBS score of Longitudinal trajectories. (I) Longitudinal MBS trajectories of individual mice. TST panels (A–D): control $n = 54$, CUMS $n = 68$. Longitudinal panels (E–I): control $n = 6$ and CUMS $n = 6$ across the tracked time points. For longitudinal panels (E–I), pointwise p values denote preplanned between-group Welch's t -tests at each time point, and repeated-measures dependence across T1–T3 was additionally evaluated using a linear mixed-effects model. For panels (E–H), these p values are Bonferroni-adjusted across the three contemporaneous time-point contrasts, and repeated-measures dependence across T1–T3 was additionally evaluated using a linear mixed-effects model.

4. Discussion

Identifying stress-resistant phenotypes in rodent models is critical for dissecting protective mechanisms against depression [26,27]. Using MBS, we consistently stratified resilient individuals—mice that maintained relatively normal behavioral function despite chronic stress—thereby providing a quantitative basis for probing resilience-associated molecular and circuit adaptations.

Conceptually, MBS treats depression-like behavior as an integrated, multimodal phenotype rather than a collection of isolated assay readouts. By linearly integrating rank-transformed SPT, OFT, and FST metrics, MBS captures coordinated dysregulation across reward, affective/exploratory, and coping-related behaviors within a single continuous severity axis. This continuous representation better reflects the clinical spectrum from sub-threshold symptoms to full disorder [6]. The high cross-cohort reproducibility observed here ($ICC > 0.88$) further supports MBS as a robust phenotypic readout rather than a cohort-specific construct [38].

More broadly, MBS is conceptually related to previously reported composite behavioral approaches, particularly integrated behavioral z-scoring, which combines multiple assays into a single emotionality-related summary measure [44]. Subsequent studies have also applied integrated emotionality or emotionality z-score-type summaries in chronic stress and mechanistic contexts [45,46]. However, MBS differs from these earlier frameworks in several practical respects. First, it uses within-cohort rank normalization rather than direct z-score standardization, which reduces sensitivity to assay scale differences and outlier-driven distortion across modalities. Second, it was deliberately built around a shared behavioral core that can be applied across both CUMS and CSDS, thereby facilitating direct cross-model comparison. Third, MBS preserves both a continuous severity axis and a flexible downstream stratification step for resilient and susceptible phenotypes. Finally, its modular structure allows future incorporation of paradigm-specific measures without changing the logic of the core framework.

Applying MBS across etiologically distinct stress paradigms revealed that “depression-like severity” is organized differently depending on the stress context. In CUMS, severity aligned more strongly with despair-like behavior (FST–MBS $r = 0.61$), whereas in CSDS it was more tightly coupled to anxiety-like exploration (OFT–MBS $r = -0.74$). At the same time, anhedonia-related changes were conserved across models (SPT–MBS $r \approx -0.59$). Together, these results frame cross-model differences as biologically meaningful heterogeneity rather than inconsistency [5].

The stratification enabled by MBS also sharpened phenotype definition beyond unimodal criteria. In CSDS, discrepancies between MBS-based stratification and SIT-based classification (Figure 4O) indicate that single-test thresholds can miss residual impairments outside the assayed domain. MBS effectively isolates a subset of animals with broadly preserved function across multiple behavioral dimensions, which is more appropriate for mechanistic studies of resilience and for reducing within-group heterogeneity in downstream analyses. This higher-resolution stratification is also relevant for modeling

partial recovery or latent vulnerability, where normalization in one dimension can coexist with persisting deficits in others [47].

The TST-related findings should be interpreted in a context-specific manner. We do not conclude that TST is generally inferior to FST or that TST lacks value in depression-related behavioral research. On the contrary, TST is a well-established assay for antidepressant-like pharmacological screening and has been widely used in studies of antidepressant mechanisms, including SSRI-related investigations. The present study did not administer antidepressants and therefore cannot evaluate whether FST or TST is more sensitive to SSRI responsiveness [36–39]. Our data only indicate that, in the current C57BL/6J chronic-stress cohorts, adding TST as a fourth MBS modality did not substantially change the integrated severity ranking derived from the SPT–OFT–FST core.

More generally, FST and TST should be regarded as related but non-identical assays. Both involve immobility-related coping behavior under acute inescapable stress, but they differ in physical context, motor demands, stress exposure, and pharmacological sensitivity [39,40]. Previous studies have also shown that antidepressant-like responses in FST and TST can depend on mouse strain and the specific SSRI tested. In addition, C57BL/6 mice may exhibit tail-climbing behavior during TST, which can introduce assay-specific noise. Therefore, the choice of FST, TST, or both should be determined by the experimental question. For chronic-stress severity stratification, a shared SPT–OFT–FST core was sufficient in the present study; for antidepressant-like drug screening, TST-inclusive MBS or direct TST analysis may be preferable and should be pharmacologically validated.

Several limitations should be acknowledged. While MBS provides a mathematically grounded framework for integration and stratification, the neurobiological mechanisms underlying the model-specific phenotypic architectures and resilience profiles require direct validation. In addition, although rank-based integration mitigates baseline shifts across cohorts, broader generalization across laboratories and testing conditions remains an important next step. Nevertheless, even though each testing window included a same-age control group, repeated behavioral testing and maturation across months may still influence later performance, and these factors should be considered when interpreting the longitudinal trajectories. Repeated exposure to the FST may influence subsequent immobility behavior through task adaptation or learned immobility. Therefore, although all cohorts were tested under the same schedule, repeated test exposure may contribute to shifts in the absolute immobility baseline over time and should be considered when interpreting the longitudinal trajectories.

MBS functions as a phenotypic integrator that suppresses batch-specific artifacts while preserving biologically informative variation, enabling standardized severity quantification and high-resolution identification of resilient and susceptible subtypes. This framework supports resilience-oriented, mechanism-based investigation in preclinical depression research and may improve the interpretability and reproducibility of chronic-stress behavioral phenotyping.

An additional direction for further development will be to examine how MBS relates to independent biological and functional readouts, including molecular alterations, circuit-level signatures, pharmacological responsiveness, and prospective behavioral outcomes. Such extensions may help broaden the interpretive scope of the current framework while preserving its utility as a multimodal behavioral severity measure.

5. Conclusions

The MBS framework establishes a novel approach for modal phenotyping in depression research by integrating individual behavioral domains into a unified severity metric. This computational approach overcomes the fragmentation in behavioral phenotyping, which has plagued conventional assessments in neuroscience and psychiatry, demon-

strating exceptional cross-model reliability ($ICC > 0.88$ across both CUMS and CSDS models). Algorithmic stratification resolves biologically distinct endophenotypes—exemplified by stress-resilient mice preserving normative reward function despite adversity—while exposing stressor-specific phenotypic architectures. Crucially, CUMS drove despair-dominated pathology characterized by heightened FST-MBS integration ($r = 0.61$), whereas CSDS manifested anxiety-centric dysfunction with superior OFT-MBS correlation ($r = -0.74$). TST-inclusive sensitivity analysis further supported the modular extensibility of MBS, indicating that the primary severity ranking was preserved when TST was added as an optional fourth modality. These findings emphasize that behavioral modality selection should be guided by experimental objectives. By establishing quantifiable behavioral homology across etiologically distinct models, MBS delivers a scalable phenotyping platform. Future studies leveraging this framework can dissect conserved adaptation mechanisms, accelerating the discovery of targeted interventions tailored to depression's heterogeneous manifestations.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/neurosci7040081/s1>, Supplementary Figure S1: Rank transformation enables robust cross-cohort integration of behavioral data in the CSDS model. (A) Correlation between raw and rank-transformed scores in the sucrose preference test (SPT). (B) Correlation between raw and rank-transformed scores in the open field test (OFT). (C) Correlation between raw and rank-transformed scores in the forced swim test (FST). (D) Rank-based trajectories of individual behavioral metrics (SPT, OFT, FST) sorted by ascending MBS scores in CSDS control group. (E) Rank-based trajectories of individual behavioral metrics (SPT, OFT, FST) sorted by ascending MBS scores in CSDS stressors group. (F) Ordered bar chart of MBS in CSDS control group. (G) Ordered bar chart of MBS in CSDS stressors group. Supplementary Figure S2: (A) Mouse-specific rank-based trajectories of individual behavioral metrics (SPT, OFT, FST) sorted by ascending MBS scores in the CUMS control group. (B) Mouse-specific rank-based trajectories of individual behavioral metrics (SPT, OFT, FST) sorted by ascending MBS scores in the CUMS control group. Supplementary Figure S3: (A) Mouse-specific rank-based trajectories of individual behavioral metrics (SPT, OFT, FST) sorted by ascending MBS scores in the CSDS control group. (B) Mouse-specific rank-based trajectories of individual behavioral metrics (SPT, OFT, FST) sorted by ascending MBS scores in the CSDS control group. Supplementary Figure S4: MBS scores and three behavioral indicators of two cohorts of mice in the CUMS model. Supplementary Figure S5: MBS scores and three behavioral indicators of two cohorts of mice in the CSDS model.

Author Contributions: Conceptualization, Methodology, Software, Formal Analysis Writing—Original Draft Preparation and Visualization: Y.T.; Resources, Validation, and Data Curation: Q.F., C.S., J.D., H.Q., X.Z. and Y.W.; Supervision and Project Administration: S.Q. and W.Z.; Funding Acquisition: W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Research Grants Council of Hong Kong through the Strategic Target Grant Scheme (STG1/M-501/23-N) and the Theme-based Research Scheme (T24-508/22-N); the Hong Kong Health and Medical Research Fund (HMRF10211696); the Hong Kong Global STEM Professor Scheme; and the Hong Kong Jockey Club Charities Trust.

Institutional Review Board Statement: All animal procedures were reviewed and approved by the Animal Subjects Ethics Sub-committee of The Hong Kong Polytechnic University (approval No. 23-24/889-HTI-R-STG (approval date: 10 April 2025) and No. 24-25/1083-HTI-R-STG (Approval date: 9 April 2025)) and were conducted in accordance with the guidelines of the Department of Health of the Hong Kong SAR.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available upon reasonable request to the corresponding author.

Acknowledgments: The authors gratefully acknowledge the technical support provided by the Central Animal Facility (CAF) and the ULS Animal Imaging Center (ULS-AIC) at The Hong Kong Polytechnic University. We specifically thank the staff of both facilities for their professional assistance in animal husbandry and imaging procedures.

Conflicts of Interest: The authors declare no competing interests.

References

1. Kirchner, L.; Kube, T.; Berg, M.; Eckert, A.L.; Straube, B.; Endres, D.; Rief, W. Social expectations in depression. *Nat. Rev. Psychol.* **2025**, *4*, 20–34.
2. Wu, D.; Qu, S.; Sun, H.; Zhou, S.; Qu, X.; Chen, Y.; Hu, H.; Li, X. Unveiling the Brain Mechanism underlying Depression: 12 Years of Insights from Bibliometric and Visualization Analysis. *Brain Res. Bull.* **2025**, *222*, 111246.
3. Hyman, S.E. Revitalizing psychiatric therapeutics. *Neuropsychopharmacology* **2014**, *39*, 220–229.
4. Brady, L.S.; Potter, W.Z.; Gordon, J.A. Redirecting the revolution: New developments in drug development for psychiatry. *Expert Opin. Drug Discov.* **2019**, *14*, 1213–1219.
5. Nestler, E.J.; Hyman, S.E. Animal models of neuropsychiatric disorders. *Nat. Neurosci.* **2010**, *13*, 1161–1169.
6. Krishnan, V.; Nestler, E.J. The molecular neurobiology of depression. *Nature* **2008**, *455*, 894–902.
7. Duman, R.S.; Aghajanian, G.K. Synaptic dysfunction in depression: Potential therapeutic targets. *Science* **2012**, *338*, 68–72.
8. Eliwa, H.; Brizard, B.; Le Guisquet, A.M.; Hen, R.; Belzung, C.; Surget, A. Adult neurogenesis augmentation attenuates anhedonia and HPA axis dysregulation in a mouse model of chronic stress and depression. *Psychoneuroendocrinology* **2021**, *124*, 105097.
9. Willner, P. Validity, reliability and utility of the chronic mild stress model of depression: A 10-year review and evaluation. *Psychopharmacology* **1997**, *134*, 319–329.
10. Sharma, S.; Chawla, S.; Kumar, P.; Ahmad, R.; Verma, P.K. The chronic unpredictable mild stress (CUMS) Paradigm: Bridging the gap in depression research from bench to bedside. *Brain Res.* **2024**, *1843*, 149123.
11. Golden, S.A.; Covington, H.E., III; Berton, O.; Russo, S.J. A standardized protocol for repeated social defeat stress in mice. *Nat. Protoc.* **2011**, *6*, 1183–1191.
12. Wang, W.; Liu, W.; Duan, D.; Bai, H.; Wang, Z.; Xing, Y. Chronic social defeat stress mouse model: Current view on its behavioral deficits and modifications. *Behav. Neurosci.* **2021**, *135*, 326.
13. Liu, M.Y.; Yin, C.Y.; Zhu, L.J.; Zhu, X.H.; Xu, C.; Luo, C.X.; Chen, H.; Zhu, D.Y.; Zhou, Q.G. Sucrose preference test for measurement of stress-induced anhedonia in mice. *Nat. Protoc.* **2018**, *13*, 1686–1698.
14. Voikar, V.; Stanford, S.C. The open field test. In *Psychiatric Vulnerability, Mood, and Anxiety Disorders: Tests and Models in Mice and Rats*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 9–29.
15. Yankelevitch-Yahav, R.; Franko, M.; Huly, A.; Doron, R. The forced swim test as a model of depressive-like behavior. *J. Vis. Exp. JoVE* **2015**, *97*, 52587. <https://doi.org/10.3791/52587>.
16. Can, A.; Dao, D.T.; Terrillion, C.E.; Piantadosi, S.C.; Bhat, S.; Gould, T.D. The tail suspension test. *J. Vis. Exp. JoVE* **2012**, *59*, 3769.
17. Porsolt, R.D.; Bertin, A.; Jalfre, M.J.A.I.P. Behavioral despair in mice: A primary screening test for antidepressants. *Arch. Int. Pharmacodyn. Ther.* **1977**, *229*, 327–336.
18. Insel, T.; Cuthbert, B.; Garvey, M.; Heinssen, R.; Pine, D.S.; Quinn, K.; Sanislow, C.; Wang, P. Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am. J. Psychiatr.* **2010**, *167*, 748–751.
19. Kalueff, A.V.; Stewart, A.M.; Song, C.; Berridge, K.C.; Graybiel, A.M.; Fentress, J.C. Neurobiology of rodent self-grooming and its value for translational neuroscience. *Nat. Rev. Neurosci.* **2016**, *17*, 45–59.
20. Wahlsten, D. Standardizing tests of mouse behavior: Reasons, recommendations, and reality. *Physiol. Behav.* **2001**, *73*, 695–704.
21. Strekalova, T.; Steinbusch, H.W. Measuring behavior in mice with chronic stress depression paradigm. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2010**, *34*, 348–361.
22. Lee, J.S.; Kang, J.Y.; Son, C.G. A comparison of isolation stress and unpredictable chronic mild stress for the establishment of mouse models of depressive disorder. *Front. Behav. Neurosci.* **2021**, *14*, 616389.
23. Berton, O.; McClung, C.A.; DiLeone, R.J.; Krishnan, V.; Renthal, W.; Russo, S.J.; Graham, D.; Tsankova, N.M.; Bolanos, C.A.; Rios, M.; et al. Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science* **2006**, *311*, 864–868.
24. Li, S.J.; Lo, Y.C.; Tseng, H.Y.; Lin, S.H.; Kuo, C.H.; Chen, T.C.; Chang, C.W.; Liang, Y.W.; Lin, Y.C.; Wang, C.Y.; et al. Nucleus accumbens deep brain stimulation improves depressive-like behaviors through BDNF-mediated alterations in brain functional connectivity of dopaminergic pathway. *Neurobiol. Stress* **2023**, *26*, 100566.

25. da Costa, V.F.; Ramírez, J.C.C.; Ramírez, S.V.; Avalo-Zuluaga, J.H.; Baptista-de-Souza, D.; Canto-de-Souza, L.; Planeta, C.S.; Rodríguez, J.L.R.; Nunes-de-Souza, R.L. Emotional and cognitive-like responses induced by social defeat stress in male mice are modulated by the BNST, amygdala, and hippocampus. *Front. Integr. Neurosci.* **2023**, *17*, 1168640.
26. Russo, S.J.; Murrough, J.W.; Han, M.H.; Charney, D.S.; Nestler, E.J. Neurobiology of resilience. *Nat. Neurosci.* **2012**, *15*, 1475–1484.
27. Osório, C.; Probert, T.; Jones, E.; Young, A.H.; Robbins, I. Adapting to stress: Understanding the neurobiology of resilience. *Behav. Med.* **2017**, *43*, 307–322.
28. Hodes, G.E.; Pfau, M.L.; Leboeuf, M.; Golden, S.A.; Christoffel, D.J.; Bregman, D.; Rebusi, N.; Heshmati, M.; Aleyasin, H.; Warren, B.L.; et al. Individual differences in the peripheral immune system promote resilience versus susceptibility to social stress. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 16136–16141.
29. Hollis, F.; Kabbaj, M. Social defeat as an animal model for depression. *ILAR J.* **2014**, *55*, 221–232.
30. Landis, S.C.; Amara, S.G.; Asadullah, K.; Austin, C.P.; Blumenstein, R.; Bradley, E.W.; Crystal, R.G.; Darnell, R.B.; Ferrante, R.J.; Fillit, H.; et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* **2012**, *490*, 187–191.
31. Sun, Z.W.; Wang, X.; Zhao, Y.; Sun, Z.X.; Wu, Y.H.; Hu, H.; Zhang, L.; Wang, S.D.; Li, F.; Wei, A.J.; et al. Blood-brain barrier dysfunction mediated by the EZH2-Claudin-5 axis drives stress-induced TNF- α infiltration and depression-like behaviors. *Brain Behav. Immun.* **2024**, *115*, 143–156.
32. Leng, L.; Zhuang, K.; Liu, Z.; Huang, C.; Gao, Y.; Chen, G.; Lin, H.; Hu, Y.; Wu, D.; Shi, M.; et al. Menin deficiency leads to depressive-like behaviors in mice by modulating astrocyte-mediated neuroinflammation. *Neuron* **2018**, *100*, 551–563.
33. Antoniuk, S.; Bijata, M.; Ponimaskin, E.; Włodarczyk, J. Chronic unpredictable mild stress for modeling depression in rodents: Meta-analysis of model reliability. *Neurosci. Biobehav. Rev.* **2019**, *99*, 101–116.
34. Lin, S.S.; Zhou, B.; Chen, B.J.; Jiang, R.T.; Li, B.; Illes, P.; Semyanov, A.; Tang, Y.; Verkhatsky, A. Electroacupuncture prevents astrocyte atrophy to alleviate depression. *Cell Death Dis.* **2023**, *14*, 343.
35. Cao, P.; Chen, C.; Liu, A.; Shan, Q.; Zhu, X.; Jia, C.; Peng, X.; Zhang, M.; Farzinpour, Z.; Zhou, W.; et al. Early-life inflammation promotes depressive symptoms in adolescence via microglial engulfment of dendritic spines. *Neuron* **2021**, *109*, 2573–2589.
36. Cryan, J.F.; Mombereau, C.; Vassout, A. The tail suspension test as a model for assessing antidepressant activity: Review of pharmacological and genetic studies in mice. *Neurosci. Biobehav. Rev.* **2005**, *29*, 571–625.
37. Stukalin, Y.; Lan, A.; Einat, H. Revisiting the validity of the mouse tail suspension test: Systematic review and meta-analysis of the effects of prototypic antidepressants. *Neurosci. Biobehav. Rev.* **2020**, *112*, 39–47.
38. Mayorga, A.J.; Lucki, I. Limitations on the use of the C57BL/6 mouse in the tail suspension test. *Psychopharmacology* **2001**, *155*, 110–112.
39. Kara, N.Z.; Stukalin, Y.; Einat, H. Revisiting the validity of the mouse forced swim test: Systematic review and meta-analysis of the effects of prototypic antidepressants. *Neurosci. Biobehav. Rev.* **2018**, *84*, 1–11.
40. Jin, Z.L.; Chen, X.F.; Ran, Y.H.; Li, X.R.; Xiong, J.; Zheng, Y.Y.; Gao, N.N.; Li, Y.F. Mouse strain differences in SSRI sensitivity correlate with serotonin transporter binding and function. *Sci. Rep.* **2017**, *7*, 8631.
41. Forkman, B.; Boissy, A.; Meunier-Salaün, M.C.; Canali, E.; Jones, R.B. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. *Physiol. Behav.* **2007**, *92*, 340–374.
42. Bourin, M.; Chenu, F.; Ripoll, N.; David, D.J.P. A proposal of decision tree to screen putative antidepressants using forced swim and tail suspension tests. *Behav. Brain Res.* **2005**, *164*, 266–269.
43. Chatterjee, M.; Jaiswal, M.; Palit, G. Comparative evaluation of forced swim test and tail suspension test as models of negative symptom of schizophrenia in rodents. *Int. Sch. Res. Not.* **2012**, *2012*, 595141.
44. Cuthbert, B.N. The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry* **2014**, *13*, 28–35.
45. Guilloux, J.P.; Seney, M.; Edgar, N.; Sibille, E. Integrated behavioral z-scoring increases the sensitivity and reliability of behavioral phenotyping in mice: Relevance to emotionality and sex. *J. Neurosci. Methods* **2011**, *197*, 21–31.
46. Soumier, A.; Sibille, E. Opposing effects of acute versus chronic blockade of frontal cortex somatostatin-positive inhibitory neurons on behavioral emotionality in mice. *Neuropsychopharmacology* **2014**, *39*, 2252–2262.
47. Lin, L.C.; Sibille, E. Somatostatin, neuronal vulnerability and behavioral emotionality. *Mol. Psychiatry* **2015**, *20*, 377–387.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.