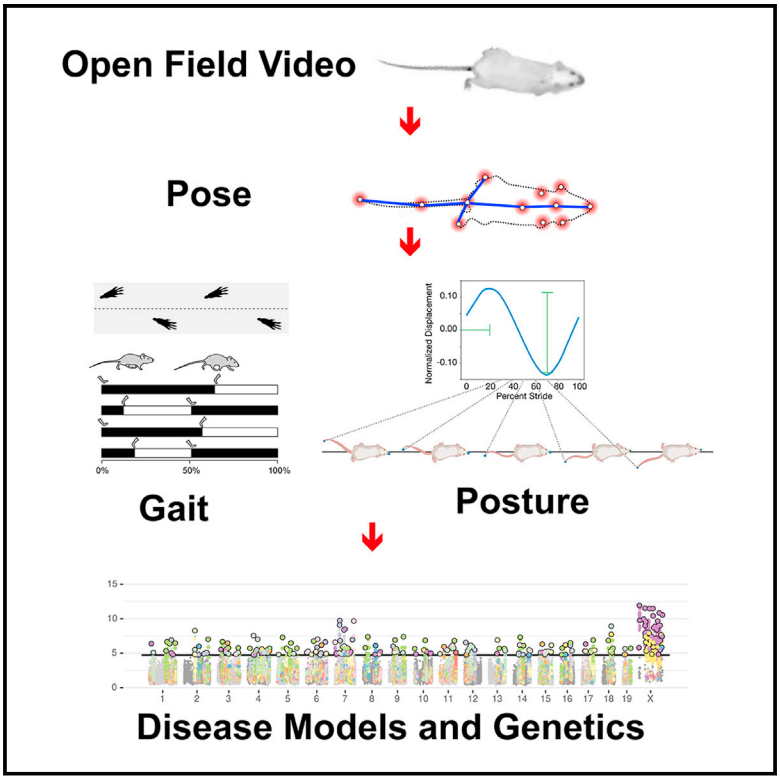
Cell Reports

Resource

Stride-level analysis of mouse open field behavior using deep-learning-based pose estimation

Graphical abstract



Authors

Keith Sheppard, Justin Gardin, Gautam S. Sabnis, ..., Brian Geuther, Cathleen M. Lutz, Vivek Kumar

Correspondence

vivek.kumar@jax.org

In brief

Sheppard et al. present a method for gait and posture analysis in the common open field apparatus using neural-network-based pose estimation. They apply this high-throughput method to dissect the genetic architecture of mouse movement.

Highlights

- A method to determine mouse pose in an open field to extract key gait and posture metrics
- These methods are genetically validated with known gait mutants
- Mouse models of autism spectrum disorder have gait and posture deficits
- GWAS describes the genetic architecture of gait and posture in 62 mouse strains

Sheppard et al., 2022, Cell Reports *38*, 110231
January 11, 2022
https://doi.org/10.1016/j.celrep.2021.110231

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

Stride-level analysis of mouse open field behavior using deep-learning-based pose estimation

Keith Sheppard,[1] Justin Gardin,[1] Gautam S. Sabnis,[1] Asaf Peer,[1] Megan Darrell,[1] Sean Deats,[1] Brian Geuther,[1] Cathleen M. Lutz,[1] and Vivek Kumar[1,2,*]

[1]The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

[2]Lead contact

*Correspondence: vivek.kumar@jax.org

https://doi.org/10.1016/j.celrep.2021.110231

SUMMARY

Gait and posture are often perturbed in many neurological, neuromuscular, and neuropsychiatric conditions. Rodents provide a tractable model for elucidating disease mechanisms and interventions. Here, we develop a neural-network-based assay that adopts the commonly used open field apparatus for mouse gait and posture analysis. We quantitate both with high precision across 62 strains of mice. We characterize four mutants with known gait deficits and demonstrate that multiple autism spectrum disorder (ASD) models show gait and posture deficits, implying this is a general feature of ASD. Mouse gait and posture measures are highly heritable and fall into three distinct classes. We conduct a genome-wide association study to define the genetic architecture of stride-level mouse movement in the open field. We provide a method for gait and posture extraction from the open field and one of the largest laboratory mouse gait and posture data resources for the research community.

INTRODUCTION

In humans, the ability to quantitate gait and posture at high precision and sensitivity allows the determination of the proper function of numerous neural and muscular systems (Nutt et al., 1993; Sanders and Gillig, 2010). Many psychiatric, neurodegenerative, and neuromuscular illnesses are associated with alterations in gait and posture (Verghese et al., 2002; Allan et al., 2005; Licari et al., 2020; Green et al., 2009; Flyckt et al., 1999; Walther and Strik, 2012; Baldaçara et al., 2008; Hausdorff et al., 2004; Scherder et al., 2007; McIntosh et al., 1997). This is because proper gait, balance, and posture are under the control of multiple nervous system processes (Takakusaki, 2013, 2017), which include critical sensory centers that process visual, vestibular, auditory, proprioceptive, and visceral inputs. Regions of the brain that directly control movement, such as the cerebellum, motor cortex, and brainstem, respond to cognitive and emotionality cues. Thus, gait and posture integrity reflects proper functioning of many neural systems in humans (Takakusaki, 2013, 2017). Mice offer genetically tractable models for mechanistic and interventional studies. The ability to measure gait and posture in an accurate and scalable manner enhances the utility of existing models and may also lead to the development of better models of diseases.

Analysis of human and animal movement, including gait, has a storied past (Baker, 2007). Aristotle wrote a philosophical treatise on animal movement and gait using physical and metaphysical principles (Aristotle, 2004). During the Renaissance, Borelli applied the laws of physics and biomechanics to muscles, tendons, and joints of the entire body to understand gait (Borelli and Maquet, 2012). The application of imaging technologies to the study of gait is credited to the work of Muybridge and Marey, who took sequential photographic images of humans and animals in motion to derive quantitative measurements of gait (Lanska, 2016; Manjila et al., 2015; Braun, 1992). Modern animal gait analysis methods are credited to Hildebrand (1977), who in the 1970s classified gait based on quantified metrics. He defined a gait cycle in terms of contact of the limb to the ground (stance and swing phases). This is in contrast to human gait and posture analysis, which, since the time of Borelli, has focused on body posture and is akin to the quantitation of whole-body movement rather than simply contact with the ground (Kirtley, 2006). In rodents, recent methods have fomented progress by the incorporation of speed in gait analysis (Batka et al., 2014; Bellardita and Kiehn, 2015; Broom et al., 2017) and determination of whole-body posture (Machado et al., 2015, 2020).

The open field assay is one of the oldest and most commonly used assays in behavioral neurogenetics (Greenberg and Haraway, 1998; Hall, 1934). In rodents, it has classically been used to measure endophenotypes associated with emotionality, such as hyperactivity, anxiety, exploration, and habituation (Crawley, 2007). For video-based open field assays, rich and complex behaviors of animal movement are often abstracted to a simple point to extract behavioral measures (Dell et al., 2014). This oversimplified abstraction is necessary mainly due to technological limitations that have prohibited the accurate extraction of complex poses from video data (Egnor and Branson, 2016). Recent technology has started to overcome this limitation (Mathis et al., 2018; Pereira et al., 2019); Wiltschko et al., 2015) and has enabled a new era of

animal behavior analysis. Gait, an important indicator of neural function, is not typically analyzed in the open field mainly due to the technical difficulty of determining limb position when animals are moving freely (Lakes and Allen, 2016). The ability to combine open field measures with gait and posture analysis would offer key insights into the neural and genetic regulation of animal behavior. Here, we leverage modern neural network methods to carry out mouse gait and posture analysis in the open field. We develop and apply a system to measure gait and posture from a top-down perspective that is invariant to the high level of visual diversity seen in the mouse, including coat color, fur differences, and size differences (Geuther et al., 2019). We characterize a set of neurodegenerative, neurodevelopmental, and autism spectrum disorder (ASD) models using our approach. We apply our methods to carry out a strain survey and genome-wide association study (GWAS) analysis of 62 mouse strains and find that gait and posture traits are highly heritable. We also find that variance of gait and posture phenotypes are heritable and regulated by distinct genetic architecture. These method and strain data are a community resource for mouse movement in the open field for the behavioral neurogenetics community.

RESULTS

Our approach to gait and posture analysis is composed of several modular components. At the base of our toolkit is a deep convolutional neural network that has been trained to perform pose estimation on top-down video of an open field. This network provides 12 two-dimensional markers of mouse anatomical location, or "key points," for each frame of video describing the pose of the mouse at each time point. We have also developed downstream components that are capable of processing the time series of poses and identifying intervals that represent individual strides. These strides form the basis of almost all of the phenotypic and statistical analyses that follow. We can extract several important gait metrics on a per-stride basis because we have pose information for each stride interval (see Table 1 for a list of metrics). This gives us significant power to perform statistical analysis on stride metrics as well as allowing us to aggregate large amounts of data to provide consensus views of the structure of the mouse gait.

Pose estimation

We selected 12 key points to capture mouse pose: nose, left ear, right ear, base of neck, left forepaw, right forepaw, mid-spine, left hind paw, right hind paw, base of tail, mid-tail, and tip of tail (Figure S1B). Much effort has been spent developing and refining pose estimation techniques for the human pose (Moeslund et al., 2006; Dang et al., 2019; Insafutdinov et al., 2016; Newell et al., 2016). These advances in pose estimation techniques have also successfully been applied to pose estimation in animals. Prominent examples of this include DeepLabCut (Mathis et al., 2018) and LEAP (Pereira et al., 2019). We selected the HRNet architecture (Sun et al., 2019), which maintains high-resolution features throughout the network stack, thereby preserving spatial precision (Figure S1A), and implemented modifications for our experimental configuration (see STAR Methods). We used this network to generate 12 480 × 480 heatmaps (one heatmap per key point) for each 480 × 480 frame of video (Figure S1A). The maximum value in each heatmap represents the highest confidence location for each respective point. Thus, after taking the argmax of each of the 12 heatmaps, we have 12 (x, y) coordinates representing the animal's pose at that frame (Figure S1B). We labeled ~8,000 images across a diverse set of strains to train a network that operates across 62 mouse strains with high visual diversity (Geuther et al., 2019) (Figure S1C; Video S1).

Stride detection

Our approach to detecting stride intervals is based on the cyclic structure of gait as described by Hildebrand (1977, 1989) (Figures 1A and 1B). During a stride cycle, each of the paws has a stance phase and a swing phase (Lakes and Allen, 2016). During the stance phase, the paw of the mouse supports the weight of the mouse and is in static contact with the ground. During the swing phase, the paw moves forward and does not support the weight of the mouse. Following Hildebrand, we refer to the transition from stance phase to swing phase as the toe-off event and the transition from swing phase to stance phase as the foot-strike event.

To calculate stride intervals, we determined stance and swing phases for the hind paws. We calculated paw speed and infer that a paw is in stance phase when the speed falls below a threshold and that it is in swing phase when it exceeds that threshold (Figures 1C–1F). We can now determine that foot-strike events occur at the transition frame from swing phase to stance phase (Figure 1C). We defined the left hind foot strike as the event that separates stride cycles. An example of the relationship between paw speed and foot strike events is shown in Figure 1D for hind paws. We find clean, high-amplitude oscillations of the hind paws, but not forepaws, as shown in Figure 1E. This difference in inference quality between the forepaws and hind paws is likely due to the fact that forepaws are occluded more often than hind paws from the top-down view and are therefore more difficult to locate accurately. We observe a corresponding decrease in the confidence of forepaw inferences (Figure 1G). For this reason, we exclude forepaws from consideration when deriving stride intervals and focus instead on hind paws. We also perform a significant amount of filtering on strides to remove spurious or low-quality stride cycles from our dataset (Figure 1G). The criteria for removing strides include low-confidence or physiologically unrealistic pose estimates, missing right hind paw strike event, and insufficient overall body speed of mouse, which is any speed <10 cm/s. Figure 1G shows the distribution of confidences for each key point. Our filtering method uses 0.3 as a confidence threshold. Very-high-confidence key points are close to 1.0. We always remove the first and last strides in a continuous sequence of strides to avoid starting and stopping behaviors from adding noise to our stride data (Figures 1C and 1D, labeled A and D, in tracks A and B). This means that a sequence of seven strides will result in at most five strides being used for analysis. The distribution of key point confidence varies by key point type (Figure 1G). Key points that tend to be occluded in a top-down view such as fore paws have confidence distributions shifted down compared to other key points. We find

Table 1. Gait metrics definitions

Measure	Definition of measure	Units
Angular velocity	the current angle of a mouse is determined by the vector connecting the mouse's base of tail to its base of neck; the first derivative of this value gives us angular velocity; for strides, angular velocity is averaged over the duration of the stride	degrees/s
Stride speed	the speed of a mouse is determined by tracking the movement speed of the base of tail key point; stride speed is the average speed for all frames over the duration of a stride; we shortened "stride speed" to "speed" in some figure labels for compactness	cm/s
Limb duty factor	the stance time of a paw (the amount of time that the paw is in contact with the ground) divided by the full stride time; duty factor is calculated for each of the hind paws and averaged	none
Temporal symmetry	where *l* is the duty factor of the left hind paw and *r* is the duty factor of the right hind paw; temporal symmetry is calculated as $(l - r)/(l + r)$	none
Step length	the distance that the right hind paw travels past the previous left hind paw strike	cm
Step width	the averaged lateral distance separating hind paws; this is calculated as length of the shortest line segment that connects the right hind paw strike to the line that connects the left hind paw's toe-off location to its subsequent foot-strike position	cm
Stride length	the full distance that the left hind paw travels for a stride, from toe-off to foot-strike.	cm
Lateral displacement of nose	to calculate lateral displacement, we first calculate the mouse's displacement vector for a stride; we then measure the nose's perpendicular distance from this vector for each frame of a stride; we then subtract the minimum distance from the maximum and divide by the mouse's body length so that the displacement measured in larger mice will be comparable to the distance measured in smaller mice	none
Lateral displacement of base of tail	calculated using the same approach that is applied to the nose lateral displacement, except that we are using the base of tail key point	none
Lateral displacement of tip of tail	calculated using the same approach that is applied to the nose lateral displacement, except that we are using the tip of tail key point	none
Nose lateral displacement phase offset	the lateral displacement is calculated for each frame of a stride, as described for nose lateral displacement, above; we then perform a cubic spline interpolation to generate a smooth curve for displacement; then, we determine the point in time at which maximum displacement occurs; note that because of cubic interpolation, this can occur at time points between frames	percent stride cycle
Base of tail displacement phase offset	calculated using the same approach that is applied to the nose lateral displacement phase offset, except that we are using the base of tail key point	percent stride cycle
Tip of tail displacement phase offset	calculated using the same approach that is applied to the nose lateral displacement phase offset, except that we are using the tip of tail key point	percent stride cycle

that key points that are not visually salient, such as the spine center, have lower confidence since they are more difficult to locate precisely. Finally, we also calculated an instantaneous angular velocity, which allows us to determine the turning direction of each stride (Figure 1F). The angular velocity is calculated by taking the first derivative of the angle formed by the line that connects the base of the mouse's tail to the base of its neck. In summary, this approach allows us to identify individual high-quality strides of a mouse in the open field.

To validate that our gait quantitation is functioning properly, we analyzed data from a commonly used inbred strain, C57BL/6NJ. We calculated percentage of stance and swing from 15,667 strides of 31 animals using ~1 h of open field video per mouse. We analyzed data from hind paws since these showed the highest amplitude oscillations during stance and swing (Figures 1D and 1E). We stratified the data into 9 angular velocity and 8 stride speed bins based on the tail base point (Figures 1H and 1I, respectively). As expected, we find an increase in stance percentage over a stride of the left hind paw when the animal turns left. Reciprocally, when the animal turns right, the stance percentage of the right hind paw increases (Figure 1H). We then analyzed strides in the central angular velocity bin (−20° to

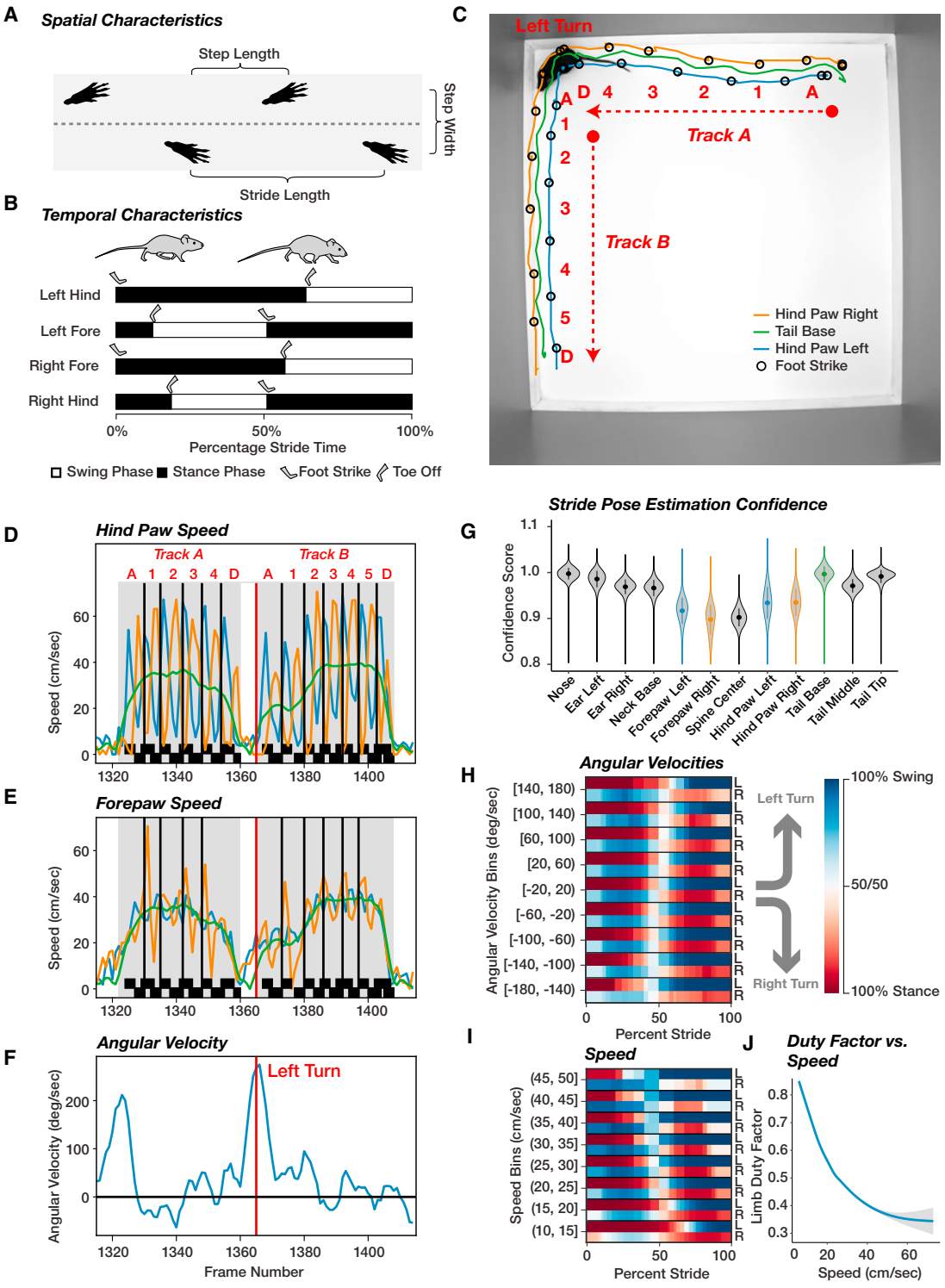


Figure 1. Extraction of gait metrics from video pose estimation

(A and B) Spatial and temporal characteristics of gait. (A) We derived three spatial stride metrics from hind paw foot-strike positions: step length, step width, and stride length. (B) All of the metrics shown in this Hildebrand plot have percent stride time for units. We see here the relationship between foot-strike and toe-off events with the stance and swing phases of stride. (A) and (B) are adapted from Lakes and Allen, 2016.

(C) A single frame of input video with hind paw tracks plotted 50 frames in the past and 50 frames in the future. The location of hind foot-strike events is indicated with black circles, and paths are shown (left hind paw [blue], right hind paw [orange], and base of tail [green]) for two sequences (tracks A and B).

(legend continued on next page)

20°/s) to determine whether stance percentage during a stride cycle decreases as the speed of the stride increases. We find that the stance time decreases as the stride speed increases (Figure 1I). We generated the same plots for five other mouse strains and see similar results for all five (Figure S2A). We calculated a duty factor for the hind paws to quantitate this relationship with stride speed (Figure 1J). We conclude that our methods are able to quantitatively and accurately extract strides from these open field videos from a top-down perspective.

After the stride intervals have been determined, we use frame poses in conjunction with stance and swing phase intervals to derive several stride metrics as defined in Table 1. We are able to extract the most relevant spatiotemporal metrics from the hind paws, which serve as the primary data source for our statistical analyses (Lakes and Allen, 2016).

Posture estimation during gait cycle

Our top-down videos allow us to determine the relative position of the spine with six key points (nose, neck base, spine center, tail base, tail middle, and tail tip). With these, we extracted the pose during a stride cycle, similar to previous work, which carried this out with nose and tail pose only (Machado et al., 2015). We used three points (nose, base of tail, and tip of tail) to capture the lateral movement during a stride cycle (Figures 2A–2C; Video S2). These measures are circular, with opposite phases of the nose and the tip of tail. For display, we use C57BL/6J and NOR/LtJ (Video S3), which have different tip-of-tail phases during a stride cycle. We are able to extract these phase plots for each stride (Figures 2D and 2E; Video S4). Since we have captured several hours of video across each strain, we are able to extract thousands of strides, enabling high levels of sensitivity. We can combine these at one stride speed and angular velocity bin at which we constrain the speed range from 20 to 25 cm/s and angular velocity from −20° to 20°/s to determine a consensus stride phase plot for each animal and strain (Figures 2F and 2G). Finally, we compared these phase plots between several strains and find striking diversity among posture during the gait cycle (Figures 2H and 2I). The diversity in posture across mouse strains is evident and implies high heritability of this phenotype.

Several of our metrics relate to the cyclic lateral displacement we observe in pose key points (Figure 2). Our measures of lateral displacement are defined as an orthogonal offset from the relevant stride displacement vector. We define the displacement vector as the line connecting the mouse's center of spine on the first frame of a stride to the mouse's center of spine on the last frame of stride. We calculate this offset at each frame of a stride and then perform a cubic interpolation to generate a smooth displacement curve. The phase offset of displacement is defined as the percent stride location where maximum displacement occurs on this smoothed curve. As an example, if we observe a value of 90 for phase offset, then it indicates that the peak lateral displacement occurs at the point at which a stride cycle is 90% complete. The lateral displacement metric assigned to stride is the difference between maximum displacement value and minimum displacement value observed during a stride (Figure 2A). This analysis is sensitive and allows us to detect subtle, but highly significant differences in overall posture during a stride (Videos S3 and S4). We used the previous classical spatiotemporal measures based on Hildebrand's methods with the combined posture metrics for our analysis. Because of the cyclic nature of phase offset metrics, care was taken to apply circular statistics to these metrics in our analysis. The other measures are analyzed using linear methods.

Next, we determined whether stride metrics changed depending on the location of the animal. For instance, animals displaying thigmotaxis are considered to be more anxious and in a lower state of arousal than those that are in the center (Crawley, 2007). To determine whether these differing emotional states affect gait and posture metrics, we analyzed the strides based on the location at which they occur. We partitioned each stride into center or periphery. To carry this out, we trained a new neural network to detect corners of our open field. We defined the periphery as the outermost 10% of the matrix (Figure S7A, blueversus purple). We only analyzed strides in 20–25 and 25–30 cm/s speed bins with angular velocity in (−20,20)°/s. Both groups contained an approximately equal number of strides for both strains (Figure S7B, red versus blue). Analysis of the gait and posture metrics showed no difference between the center and periphery (Figures S7C and S7D). Surprisingly, this analysis indicates that these measures do not change in response to location, and by extension, emotional state of the animal.

Statistical analysis of gait measures on three mutant strains

We phenotyped three mouse models that have previously been shown to have gait defects and are preclinical models of human diseases: Rett's syndrome, amyotrophic lateral sclerosis (ALS, or Lou Gehrig's disease), and Down syndrome. The three models, *Mecp2* knockout, SOD1 G93A transgene, and *Ts65Dn Trisomic*, respectively, were tested with appropriate controls at two ages in a 1-h open field assay (see STAR Methods). Gait metrics are highly correlated with animal size and stride speed (Batka et al., 2014; Bellardita and Kiehn, 2015; Machado et al., 2015, 2020; Hildebrand, 1989) (Figures 1I and J). However, in

(D–F) Three plots showing different aspects of the mouse's movement over the same 100-frame interval. The centered red vertical line indicates the current frame (displayed in C). The top plot shows three lines indicating speed of the left hind paw (blue), the right hind paw (orange), and the base of tail (green). The vertical black lines in the plot indicate the inferred start frame of each stride.

(G) The distribution of confidence values for each of the 12 key points we estimate.

(H) Aggregate view of Hildebrand plot for hind paws binned according to angular velocity (left [L] and right [R]) shows changes in strike duration based on direction of turning.

(I) Similar to (H), except binned by increasing stride speed and a fixed angular velocity (−20° to 20°/s).

(J) Limb duty factor changes as a function of stride speed.

Data for (H–J) are derived from 15,667 strides from 31 C57BL/6NJ animals.

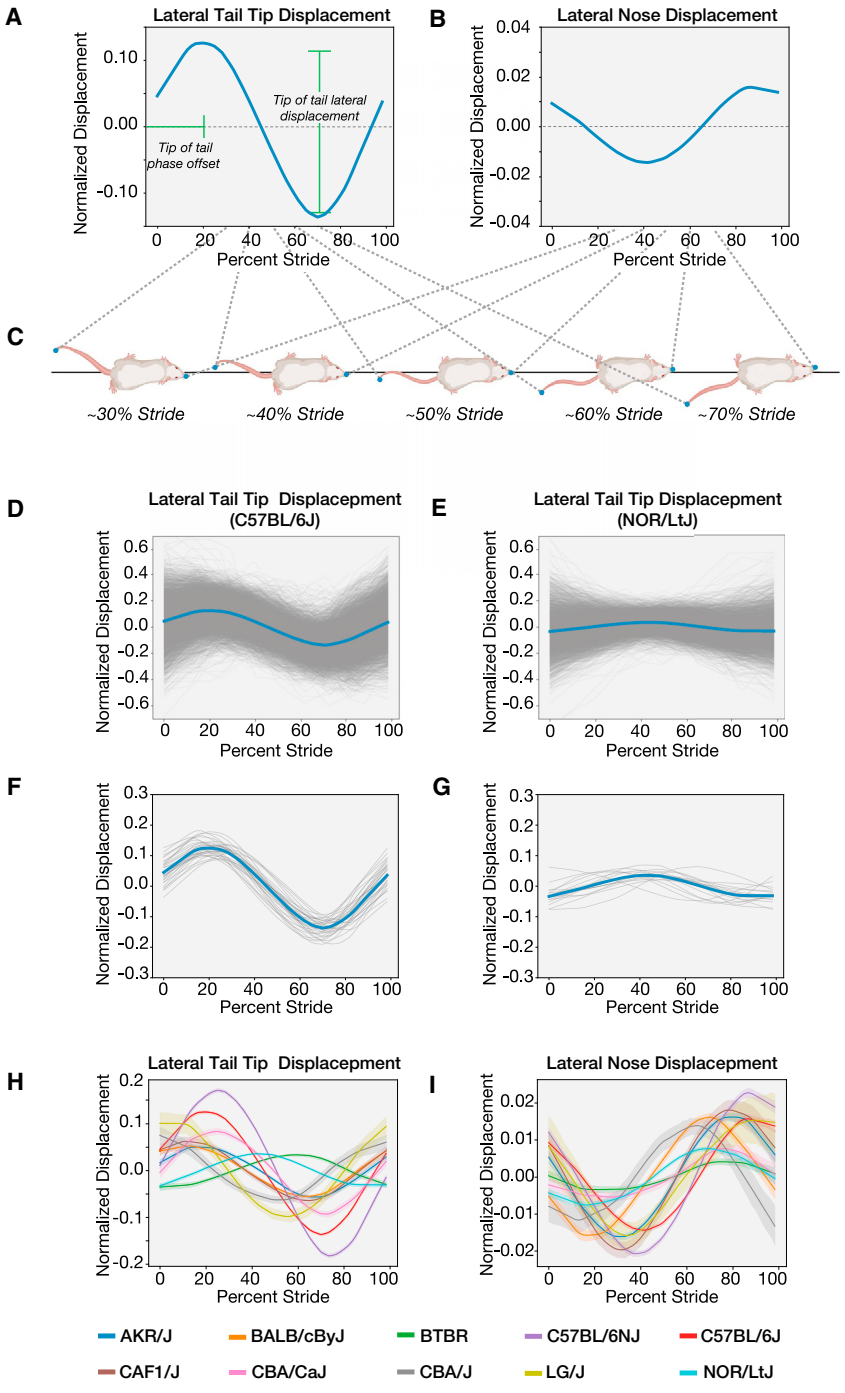


Figure 2. Extraction of cyclic posture metrics during gait cycle

(A and B) We measured lateral displacement of (A) the tail tip and (B) the nose. Positive values are to the animal's left and negative values are to its right. We labeled this "normalized displacement" because displacement values are divided by the respective animal's body length. We did this so that differences in amplitude could not simply be attributed to animal size. (C) A cartoon of the mouse during a gait cycle. Tail and nose points are shown at various positions during one cycle.

(D–G) We could also average displacement across many strides within a cohort to form a consensus view such as (D) C57BL/6J versus (E) NOR/LtJ, or we could average many strides within individuals: (F) C57BL/6J versus (G) NOR/LtJ.

(H and I) For tail (H) and nose (I), we note the diversity of lateral displacement between a set of strains selected from our strain survey. The translucent bands for these two plots represent the 95% confidence interval of the mean for each respective strain.

many cases changes in stride speed are a defining feature of gait change due to genetic or pharmacological perturbation. To compare our results with previously published data that do not take animal size and sometimes stride speed into account, we analyzed our data with three models that take only age and body length (M1), age and stride speed (M2), and age, stride speed, and body length (M3) as covariates (see STAR Methods). Since sex of the animal and its body length are highly collinear, we do not include sex as a term in the final model (measured using ANOVA and denoted by η, is strong for both SOD1 [η = 0.81] and *Ts65Dn* [$\eta = 0.16$ overall, $\eta = 0.89$ for controls, $\eta = 0.61$ for mutants]). We model the posture-based circular phase variables (Table 1) as a function of linear variables using a circular-linear regression model (Fisher and Lee, 1992) (see STAR Methods). The results are displayed in Figures 3, 4, S3, and S4, and exact statistics are reported in Figure S11B.

Characterization of gait in a Rett's syndrome model

Rett's syndrome, an inherited neurodevelopmental disorder, is caused by mutations in the X-linked *MECP2* gene (Amir et al., 1999). We tested a commonly studied deletion of *Mecp2* that recapitulates many of the features of Rett's syndrome, including reduced movement, abnormal gait, limb clasping, low birth weight, and lethality (Guy et al., 2001). We tested hemizygous males (n = 8), heterozygous females (n = 8), and littermate controls (n = 8 of each sex) (see STAR Methods). Null males are normal at birth and have an expected lifespan of ~50–60 days. They start to show age-dependent phenotypes by 3–8 weeks and lethality by 10 weeks. Heterozygous females have mild symptoms at a much older age (Guy et al., 2001). We tested male mice twice at 43 and 56 days and females at 43 and 86 days.

Previous gait studies of this knockout did not take animal size, and in some cases, changes in stride speed, into account. These studies have shown changes in stride length and stance width in an age-dependent manner in hemizygous males (Kerr et al., 2010;

CellPress
OPEN ACCESS

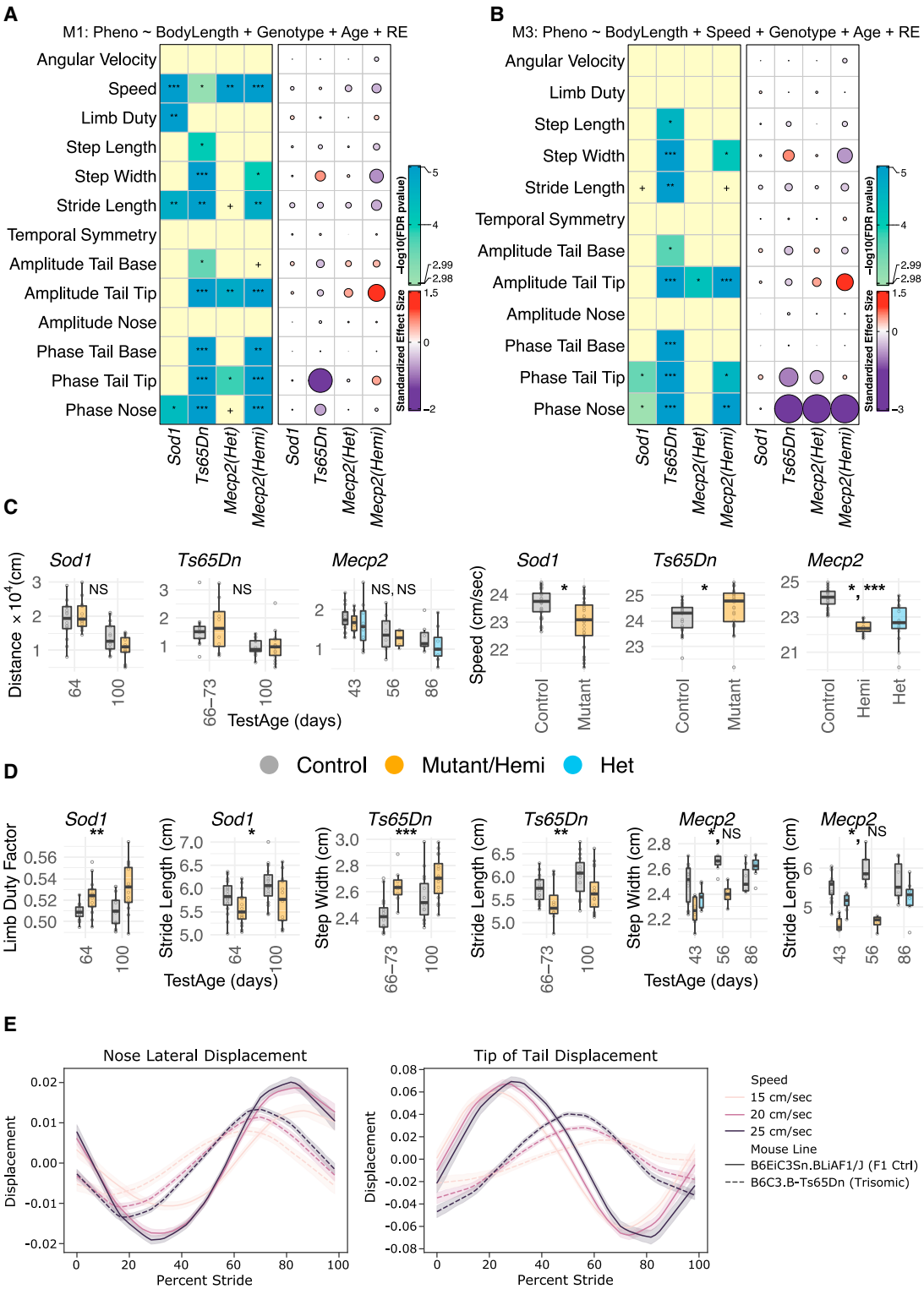


Figure 3. Analysis of gait in three mutant strains

(A) We found previously reported differences using M1, which adjusts only for body length, test age (Age), genotype, and random effects (RE). The LOD ($-\log_{10}(q_{value})$) scores and effect sizes are shown in the left and right vertical blocks, respectively. In the left block, the number of "★" and heat represent the strength of evidence against the null hypothesis of no genotype-based effect, while + represents a suggestive effect. In the right block, the color (red for positive and blue for negative) and area of the circle (area ∝ size of the effect) represent the direction and magnitude of the effect size.

(legend continued on next page)

Santos et al., 2010; Robinson et al., 2012). Recent analysis showed increased step width, reduced stride length, changes in stride time, step angle, and overlap distance (Gadalla et al., 2014). *Mecp2* hemizygous males show 13% reduced body (Figures S11A and S4C) (Guy et al., 2001) and progressive changes in movement speed that should be taken into account when modeling gait parameters. We limit our analysis to stride speeds between 20 and 30 cm/s, which allows us to reduce the variation introduced by differences in speed and compare a model that includes body length instead of stride speed as a covariate (M1, Figure 3A) and one in which both body length and stride speed are included (M3, Figure 3B). We placed the results from M2 into the supplemental information for comparison with previously published data (Figure S4), and all of the statistics (M1, M2, M3) are reported in Figure S11B. Model M3 that includes both stride speed and body length showed a significant decrease in step width and suggestive difference in stride length, as well as robust differences in posture metrics (tail tip amplitude, phase of tail tip, and nose) (Figure 3B). We also note a decrease in total distance traveled in the open field, stride speed, stride length, and step width in the mutants after adjusting for body length (M1) (Figures 3C and 3D). Even though we limit the analysis to one angular and speed bin, we see differences in the distribution of stride speed (Figure S3B). We observe very few significant differences in *Mecp2* heterozygous females that are consistent across all three models. All three models consistently find tail tip amplitude to be significantly higher, suggesting more lateral movement in the females (Figures 3A, 3B, and S3). In summary, these results demonstrate that we are able to accurately detect previously described differences in *Mecp2*. In addition, our posture metrics are able to detect differences that have not been previously described. All three models consistently find tail tip amplitude to be significantly higher suggesting more lateral movement in the females (Figures 3A, 3B, and S3A).

Characterization of gait in an ALS model

Mice carrying the SOD1-G93A transgene are a preclinical model of ALS with progressive loss of motor neurons (Gurney et al., 1994; Rosen et al., 1993). The SOD1-G93A model has been shown to exhibit changes in gait phenotypes, particularly of hindlimbs (Wooley et al., 2005; Amende et al., 2005; Preisig et al., 2016; Tesla et al., 2012; Mead et al., 2011; Vergouts et al., 2015; Mancuso et al., 2011). The most salient phenotypes are an increase in stance time (duty factor) and decreased stride length in an age-dependent manner. However, several other studies have observed opposite results (Wooley et al., 2005; Amende et al., 2005; Mead et al., 2011; Vergouts et al., 2015), and some have not seen significant gait effects (Guillot et al., 2008). These studies did not adjust for body size difference or in some cases for stride speed. We tested SOD1-G93A transgenes and appropriate controls at 64 and 100 days, during time of disease onset (Wooley et al., 2005; Preisig et al., 2016; Vergouts et al., 2015; Mancuso et al., 2011; Knippenberg et al., 2010). We do not see significant differences in body length or weight (Figures S3C and S11), but changes are seen in stride speed (Figure S3B).

Using model M3, we find small changes in the phase of tail tip and nose (Figure 3B). Otherwise, we see significant changes in M1 in stride speed, limb duty factor, and stride length (Figures 3A and 3D). These results argue that the major effect of the SOD1 transgene is on stride speed, which leads to changes in stride time and duty factor. Our results are congruent with reports that gait changes may not be the most sensitive preclinical phenotype in this ALS model, and other phenotypes such as visible clinical signs and motor learning tasks such as rotarod are more sensitive measures (Guillot et al., 2008; Mead et al., 2011). Our results validate that our statistical model is able to detect known gait defects in this model and may help explain some of the discordant results in the literature.

Characterization of gait in a Down syndrome model

Down syndrome, caused by trisomy of all or part of chromosome 21, has complex neurological and neurosensorial phenotypes (Haslam, 1995). Although there are a spectrum of phenotypes such as intellectual disability, seizures, strabismus, nystagmus, and hypoacusis, the more noticeable phenotypes are developmental delays in fine motor skills (Shumway-Cook and Woollacott, 1985; Morris et al., 1982). These are often described as clumsiness or uncoordinated movements (Vimercati et al., 2015; Latash, 2000). One of the best studied models, Ts65Dn, trisomic for a region of mouse chromosome 16 that is syntenic to human chromosome 21, recapitulates many of the features of Down syndrome (Reeves, 1995; Herault et al., 2017). Ts65Dn mice have been studied for gait phenotypes using traditional inkblot footprint analysis or treadmill methods (Hampton and Amende, 2009; Costa et al., 1999; Faizi et al., 2011). The inkblot analysis showed mice with shorter and more ‘‘erratic and ‘‘irregular gaits, similar to motor coordination deficits seen in human patients (Costa et al., 1999). Treadmill-based analysis revealed further changes in stride length, frequency, some kinetic parameters, and footprint size (Hampton et al., 2004; Faizi et al., 2011). These previous analyses have not studied the posture of these mice.

We analyzed Ts65Dn mice along with control mice at ~10 and 14 weeks (see STAR Methods), and all three linear mixed models (M1–M3) found consistent changes. The Ts65Dn mice are not hyperactive in the open field (Figure 3C), although they have increased stride speed (Figures 3A and 3C). This indicates that the Ts65Dn mice take quicker steps but travel the same distance as controls. After adjusting stride speed and animal size, step width was increased and step and stride lengths were significantly reduced (Figure 3B). In particular, posture phenotypes were highly affected in the Ts65Dn mice. The amplitude of tail base and tip and the phase of tail base, tip, and nose were

(B) Same as (A), except that we used model M3, which adjusts for body length, stride speed (speed), genotype, test age (Age), and RE.
(C) We plotted distance $\times 10^4$ (cm) across test ages (x axis) and stride speed (cm/s) across gait mutants (*Sod1*, *Ts65Dn*, *Mecp2*) to compare mutants with controls. Each dot represents a tested animal.
(D) We plotted the most significant gait parameters from (A) for different gait mutants to compare mutants with controls across test ages (x axis).
(E) Lateral displacement of nose and tail tip for Ts65Dn strain. The solid lines represent the mean displacement of stride, while the translucent bands provide a 95% confidence interval for the mean.

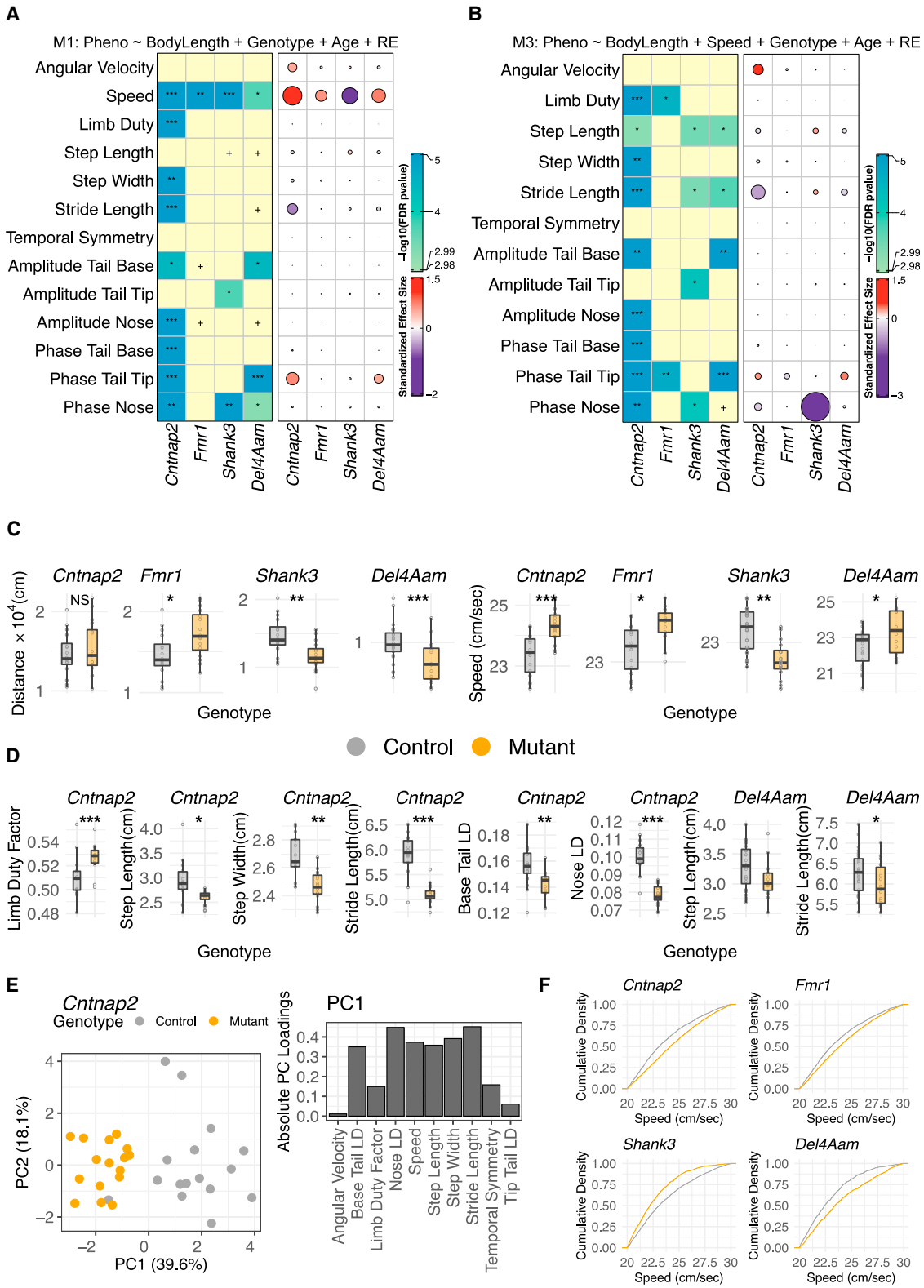


Figure 4. Characterization of gait and posture in mouse genetic models of ASD

(A) We found previously reported differences using M1, which adjusts only for body length. The LOD scores and effect sizes are shown in the left and right vertical blocks, respectively. In the left block, the number of ★s represents the strength of evidence against the null hypothesis of no genotype-based effect, while + represents a

(legend continued on next page)

significantly decreased (Figure 3B). We confirmed this with a phase plot of nose and tail tip (Figure 3E). Surprisingly, we found that there were large differences in phase. The tail tip phase peak is near 30% of the stride cycle in controls and close to 60% in mutants at multiple stride speeds (Figure 3E). Similar changes are seen in the phase plot for the nose. These results confirm previously reported differences in traditional gait measures, and highlight the utility of our open field posture measures in broadening the assayable phenotypic features in models of human disease. The most salient feature of the Ts65Dn gait is the alteration of posture, which previously was reported as a qualitative trait using inkblot analysis (Costa et al., 1999) and is now quantifiable using our methods.

Characterization of ASD-related mutants

To further validate our approach, we investigated gait in four ASD mouse models, in addition to *Mecp2*. In humans, gait and posture defects are often seen in ASD and sometimes gait and motor defects precede classical deficiencies in verbal and social communication and stereotyped behaviors (Licari et al., 2020; Green et al., 2009). Recent studies indicate that motor changes are often undiagnosed in ASD cases (Hughes, 2011). It is unclear whether these differences have genetic etiologies or are secondary to the lack of social interactions that may help children develop learned motor coordination (Zeliadt, 2017). In mouse models of ASD, gait defects have been poorly characterized, and thus we sought to determine whether any gait phenotypes occur in four commonly used ASD genetic models, which we characterized with appropriate controls at 10 weeks (see STAR Methods). Similar to the three models with known gait defects, we tested these mutants and controls in the 1-h open field assay and extracted gait and posture metrics (Table 1). We modeled the results using the same approach used for gait mutants (M1 and M3 results are presented in Figure 4, M2 results are found in Figure S4).

Cntnap2 is a member of the neurexin gene family, which functions as a cell adhesion molecule (Poliak et al., 1999). Mutations in *Cntnap2* have been linked to ASD, schizophrenia, bipolar disorder, and epilepsy (Toma et al., 2018). *Cntnap2* knockout mice have previously been shown to have mild gait effects, with increased stride speed leading to decreased stride duration (Brunner et al., 2015). These mice are significantly smaller in body length and weight than controls (Figures S11 and S4C). We used model M2 to compare our results to the previous study and found that *Cntnap2* mice show significant differences in a majority of the gait measures (Figure S4A). In the open field, *Cntnap2* mice were not hyperactive (Figure 4C) but showed a markedly increased stride speed (M1, Figures 4A and 4C). These results argue that the *Cntnap2* mice do not travel more, but take quicker steps when moving, similar to Ts65Dn mice.

Since *Cntnap2* mice are smaller and have faster stride speeds (Figure 4F), we used results from M3 to determine whether gait parameters are altered after adjusting for body size and stride speed (Figure S11). We found that *Cntnap2* mice were significantly different from controls for a majority of the traditional gait metrics as well as posture measures (Figures 4A and 4B). The *Cntnap2* mice have reduced limb duty factor, step length, and step width, and highly reduced stride length (Figures 4B and 4D). The mice also show altered phase of tail tip, base, and nose, as well as significant but small changes in amplitude of tail tip base and nose. Another salient feature of gait in *Cntnap2* mice is the decrease in interanimal variance compared to controls, particularly for limb duty factor (Fligner-Killeen test, $p < 0.01$), step length (Fligner-Killeen test, $p < 0.01$), and stride length (Fligner-Killeen test, $p < 0.02$) (Figure 4D). This may indicate a more stereotyped gait in these mutants. Combined, these results imply that *Cntnap2* mice are not hyperactive as measured by total distance traveled in the open field, but are hyperactive at the individual stride level. They take quicker steps with shorter stride and step length and narrower step width. Next, we asked whether there is a lower-dimensional gait space where the Cntnap2 mutants separate from the controls. We performed principal-component analysis (PCA) on *Z* score transformed gait metrics and embedded the animals in a two-dimensional (2D) space for visualization. We found the first PC that explained 40% of the total variance to separate the mutants and controls effectively. We plotted the absolute PC loadings to shed light on contributions of gait and posture metrics to PC1. The loadings revealed that most gait metrics contributed to PC1. We found that the gait metrics allow us to distinguish *Cntnap2* from controls (Figure 4E). This analysis shows that *Cntnap2* mice can be distinguished from controls based on its gait patterns in the open field. We report similar analyses with body length and body length + speed adjusted residuals in Figures S6A and S6B).

Mutations in *Shank3*, a scaffolding postsynaptic protein, have been found in multiple cases of ASD (Durand et al., 2007). Mutations in *Fmr1*, an RNA-binding protein that functions as a translational regulator, are associated with fragile X syndrome, the most commonly inherited form of mental illness in humans (Crawford et al., 2001). Fragile X syndrome has a broad spectrum of phenotypes that overlaps with ASD features (Belmonte and Bourgeron, 2006). *Del4Aam* mice contain a deletion of 0.39 Mb on mouse chromosome 7 that is syntenic to human chromosome 16p11.2 (Horev et al., 2011). Copy-number variations (CNVs) of human 16p11.2 have been associated with ASD features,

suggestive effect. In the right block, the color (red for positive and blue for negative) and area of the circle (area $\propto$ size of the effect) represent the direction and magnitude of the effect size.

(B) Same as (A), except that we used model M3, which adjusts for both body length and stride speed.

(C) We plotted distance $\times 10^4$ (cm) across test ages (x axis), and stride speed (cm/s) across ASD mutants (*Cntnap2*, *Fmr1*, *Shank3*, *Del4Aam*) to compare mutants with controls. Each dot represents a tested animal.

(D) We plotted the most significant gait parameters from (B) for different gait mutants to compare mutants with controls.

(E) We performed PCA on *Z* score-transformed gait data for *Cntnap2* mutants and their controls. We used the first two PCs to plot a 2D representation of the multidimensional gait space, which separates controls from the mutants. The dots represent individual animals.

(F) We plotted the stride speed cumulative distributions between mutants and controls across ASD mutants.

CellPress
OPEN ACCESS

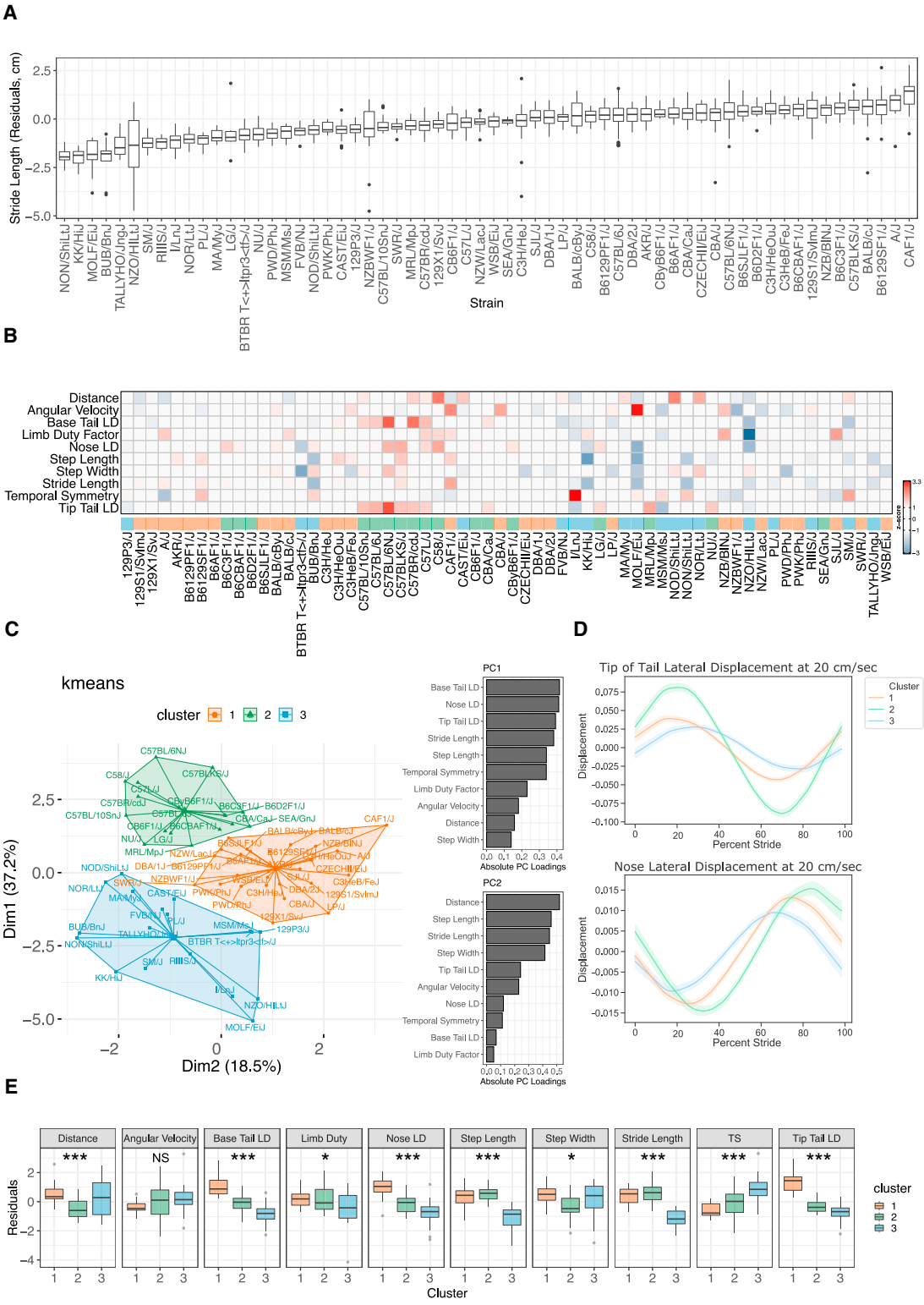


Figure 5. Strain survey results

(A) We plotted body-length-adjusted residuals for gait parameter stride length (y axis) across 62 strains in the strain survey (x axis). We arranged the boxplots in increasing order of medians from left to right.

(legend continued on next page)

including intellectual disability, stereotypy, and social and language deficits (Weiss et al., 2008). *Fmr1* mutant mice travel more in the open field (Figure 4C) and have higher stride speed (Figures 4A and 4C). When adjusted for stride speed and body length (M3), these mice undergo slight but significant changes in limb duty factor in model M3. *Shank3* and *Del4Aam* are both hypoactive in the open field compared to controls. *Shank3* mice experience a significant decrease in stride speed, whereas *Del4Aam* mice have faster stride speeds (Figures 4A and 4C). All three statistical models show a suggestive or significant decrease in step length in both strains. Using M3, we find that *Shank3* mice have longer step and stride lengths, whereas *Del4Aam* mice have shorter steps and strides. In posture measures, *Shank3* mice have a decrease in nose phase and *Del4Aam* has an increase in tail tip phase. These results indicate that even though both *Shank3* and *Del4Aam* are hypoactive in the open field (Figure 4C), *Shank3* takes slower and longer strides and steps, whereas *Del4Aam* takes faster strides with shorter steps and strides (Figure 4F). Both mutants have some defects in posture. We find each of the ASD models to have a unique set of gait deficits, with *Cntnap2* having the strongest phenotypes. All have some change in stride speed, although the directionality of change and the variance of the phenotypes differ. These results imply that changes in gait and posture are general features of ASD.

Strain survey

After validating that our methods are able to characterize differences in known gait mutants, we sought to understand the range of gait and posture phenotypes in the open field in standard laboratory mouse strains. We surveyed 44 classical inbred laboratory strains, 7 wild-derived inbred strains, and 11 F1 hybrid strains (1,898 animals, 1,740 h of video). All animals were isogenic, and we surveyed both males and females in a 1-h open field assay (see STAR Methods) (Geuther et al., 2019). We then extracted gait metrics from each video and performed an exploratory analysis of the data on a per-animal level (Figures 5A and S5). We analyzed stride data when animals were traveling at a moderate stride speed (20–30 cm/s) and in a straight direction (angular velocity between -20° and $+20^\circ$/s). We could carry out such a selective analysis because of the large amount of data we were able to collect and process in freely moving mice. Since these mice vary considerably in their size (Geuther et al., 2019), we plotted residuals from M1 that adjust for body size (Figures 5A and S5).

We sought to determine whether we could cluster strains based on their open field gait and posture phenotypes. We took a model-free approach and applied the k-means algorithm to cluster the strains. We did not include the circular features in our analyses, as the k-means algorithm requires features that lie in a Euclidean space. We fit a linear model to each linear gait feature with body length and stride speed as covariates and extracted the model residuals. The *Z* score-transformed residuals served as the input features for our analysis (Figure 5B). We initialized the k-means algorithm several times, with random points from the data as means. We picked the initialization that gave the smallest total within-cluster sum of squares. We projected the selected k-means output onto the 2D PC subspace to visualize the clustering structure (Figures 5C–5E). We found three clusters of strains that can be distinguished based on their open field gait behaviors. The number of clusters was chosen based on the gap statistic (Figure S6D). Cluster 1 consisted of primarily classical strains such as A/J, C3H/HeJ, and 129S1/SvImJ; cluster 3 consisted of several classical strains and a large number of wild-derived strains such as MOLF/EiJ and CAST/EiJ. Cluster 2 mainly consisted of C57 and related strains, including the reference C57BL/6J. Next, we visualized the clustering structure in a non-linear embedded space using Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) with two different initializations (Figure S6E). The UMAP dimensions preserved the cluster structure discovered using the k-means algorithm. We further explored the three k-means clusters. PC loading analysis revealed that most of PC1 is highly correlated with posture measures, while PC2 is correlated with open field distance and traditional gait measures (Figure 5C, right). We constructed a consensus stride phase plot of the nose and tail tip for each cluster. Cluster 3 has much higher amplitude, while clusters 1 and 2 have similar amplitude but shifted phase offset (Figure 5D). An examination of the linear gait metrics reveals individual metrics that distinguish the clusters (Figure 5E). For example, cluster 2 has longer stride and step lengths, cluster 3 has higher temporal symmetry, and cluster 2 has low lateral displacement of nose, base, and tip tail. An overall analysis of individual metrics reveals a significant difference in 9 of 10 measures. For comparison, we plotted the output of the k-means analysis of body length-adjusted gait features (Figure S6C). We found strain SEA/GnJ changed cluster membership from cluster 2 to cluster 1 and strains 129X1/SvJ, DBA/2J, LP/J, PWD/PhJ, PWK/PhJ, SWR/J, and WSB/EiJ changed from cluster 1 to cluster 3. This analysis reveals high levels of heritable variation in gait and posture in the laboratory mouse. A combined analysis using multidimensional clustering of these metrics finds three subtypes of gait in the laboratory

(B) Residuals were obtained from a linear model with body length and speed as covariates/features for all gait parameters. We transformed the residuals to obtain *Z* scores and used the scores as inputs to the k-means algorithm of the next step. The heatmap shows the *Z* scores ($|z - \text{score}|>1$; thresholding is applied for easier visualization) along with color-coded cluster memberships (x axis).

(C) We used a k-means algorithm to determine the cluster memberships. We projected the clusters discovered by k-means to a 2D space formed by PC components obtained from the *Z* scores. See Figure S6 for more information on the choice of the number of clusters and the clusters formed when the gait parameters were adjusted for both body length and body length + stride speed.

(D) A consensus view of lateral displacement of nose and tail tip across the clusters. The solid lines represent the mean displacement of stride, while the translucent bands provide a 95% confidence interval for the mean.

(E) Post-clustering analysis: We used a linear model (one-way ANOVA) with cluster membership as a categorical covariate/feature to compare gait parameters between strains across the three clusters. The number of ★s represents the strength of evidence against the null hypothesis of no difference in a gait parameter between strains across three clusters. In contrast, NS represents not sufficient evidence to claim a difference in a gait parameter between strains across 3 clusters.

mouse. Our results also show that the reference mouse strain, C57BL/6J, is distinct from other common mouse strains and wild-derived strains.

We further explored cluster 2, which contains mostly strains from the C57 family. We asked whether our movement phenotypes could distinguish among the C57 family. We used two approaches: a supervised dimension-reduction approach and a classification approach. For the former, we used linear discriminant analysis (LDA) (Machado et al., 2020; Rao, 1948) to quantitatively distinguish between strains C57BL/6J, C57BL/6NJ, C57BLKS/J, C57L/J, C57BR/cdJ, C57BL/10SnJ, and C58/J. C57BL/6J and C57BL/6NJ are considered substrains, while the rest are independent, yet closely related strains (Morse, 2012). For the second approach, we used a multi-class logistic regression (“one versus rest”) model to predict the strain membership for each animal from its gait metrics. We adjusted the gait metrics for both body length (Figures S8A and S8B) and body length + stride speed (Figures S8C and S8D) in our analyses to account for their effect on gait metrics. We found that LDA separated strains when we embedded their adjusted gait metrics in a lower-dimensional 2D space using PCs (Figures S8A and S8C). We plotted absolute PC loadings to understand the gait and posture metrics contributions and found that LD1 consists mainly of base tail lateral displacement and LD2 consists of several gait and posture metrics (Figure S8). In the second approach, we summarized the sensitivity of gait analysis for the multi-class classifier to distinguish between strains using a confusion matrix, which shows the proportion of correctly classified and misclassified animals in each strain (Figures S8B and S8D). Indeed, these combined data indicate that animal movement alone can accurately distinguish genetically similar strains and even substrains.

GWAS

The strain survey demonstrated that the gait and posture features we measure are highly variable, and thus we wanted to understand the heritable components and the genetic architecture of mouse gait in the open field. In human GWAS, both mean and variance of gait traits are highly heritable (Adams et al., 2016). We separated the strides of each animal into four different bins according to the speed at which it was traveling (10–15, 15–20, 20–25, and 25–30 cm/s) and calculated the mean and variance of each trait for each animal to conduct a GWAS to identify quantitative trait loci (QTL) in the mouse genome. We used GEMMA (Zhou and Stephens, 2012) to conduct a GWAS using a linear mixed model. To distinguish between body length-dependent and -independent QTL, we used two models, one taking into account sex and body length as fixed effects and another taking only sex as a fixed effect for comparison. Both models used population structure as a random effect. Since linear mixed models do not handle circular values, we excluded phase gait data from our analysis. The heritability was estimated by determining the proportion of variance of a phenotype that is explained (PVE) by the typed genotypes (Figure 6A, left panel). Heritability of gait measures showed a broad range, and the majority of the phenotypes are moderately to highly heritable. The mean phenotypes with lowest heritability are angular velocity and temporal symmetry, indicating that variance in the symmetrical nature of gait or turning behaviors are not due to genetic variance in the laboratory mouse. In contrast, we find that measures of posture (amplitude measures) and traditional gait measures are moderately to highly heritable. Variance of phenotypes showed moderate heritability, even for traits with low heritability of mean traits (Figure 6A, right panel). For instance, mean angular velocity phenotypes have low heritability (PVE <0.1), whereas the variance angular velocity phenotypes have moderate heritability (PVE between 0.2 and 0.4). These heritability results indicated that the gait and posture traits are appropriate for GWAS of mean and variance traits. When body length was not included in the model, we observed changes in heritability with gait phenotypes that are dependent on animal size (e.g. stride length, step length) (Figure S9). We proceeded with GWAS and excluded traits with low heritability (PVE <0.25).

For the significance threshold, we calculated an empirical p value correction for the association of an SNP with a phenotype by shuffling the values (total distance traveled in the open field) between the individuals 1,000 times. In each permutation, we extracted the lowest p value to find the threshold that represents a corrected p value of 0.05 (1.9×10^{-5}). We took the minimal p value over all mean phenotypes, variance phenotypes, and both classes combined for each SNP to generate combined Manhattan plots (Figures 6B–6D for the model with body length and Figures S9B–S9D for the model without it). Each SNP is colored according to the phenotype associated to the SNP with the lowest p value.

We found 157 QTL for mean traits and 117 QTL for variance traits (Figures 6B and 6C; Table S1). The phenotype with the most associated loci was tail base amplitude (10–15 cm), with 25 loci. Overall, when considering all of the phenotypes together, we found 254 significant genomic regions associated with at least 1 phenotype (Table S1), indicating only 20 QTL were identified for both a mean phenotype and a variance phenotype. Most phenotypes had limited to no overlap between QTL associated with the mean of the feature and its variance. These data argue that the genetic architecture of mean and variance traits in the mouse are largely independent. We compared GWAS models with and without body length as a fixed effect and find 296 QTL that are dependent on animal length. We find 194 and 102 QTL in models without and with body length in model, respectively. We detected 152 QTL that are common in both models (Figure S10A; Table S1). This comparison allows us to assign specific QTL to animal anatomy and sets a framework for detecting feature-dependent genetics.

We extracted the genes residing in the identified QTL and tested for enriched Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, or Mammalian Phenotype (MP) Ontology associated with them using the software INRICH (Lee et al., 2012). Among the most enriched terms are the GO term “positive regulation of synaptic transmission, glutamatergic (BP)” (GO: 0051968) with two QTL for open field variance traits containing genes associated with this term out of 31 genes in the genome, and “proteolysis involved in cellular protein catabolic process (BP)” (GO: 0051603) enriched in the QTL for pose variance (Figure S10B). These results begin to outline the genetic landscape of mouse gait and posture in the open field.

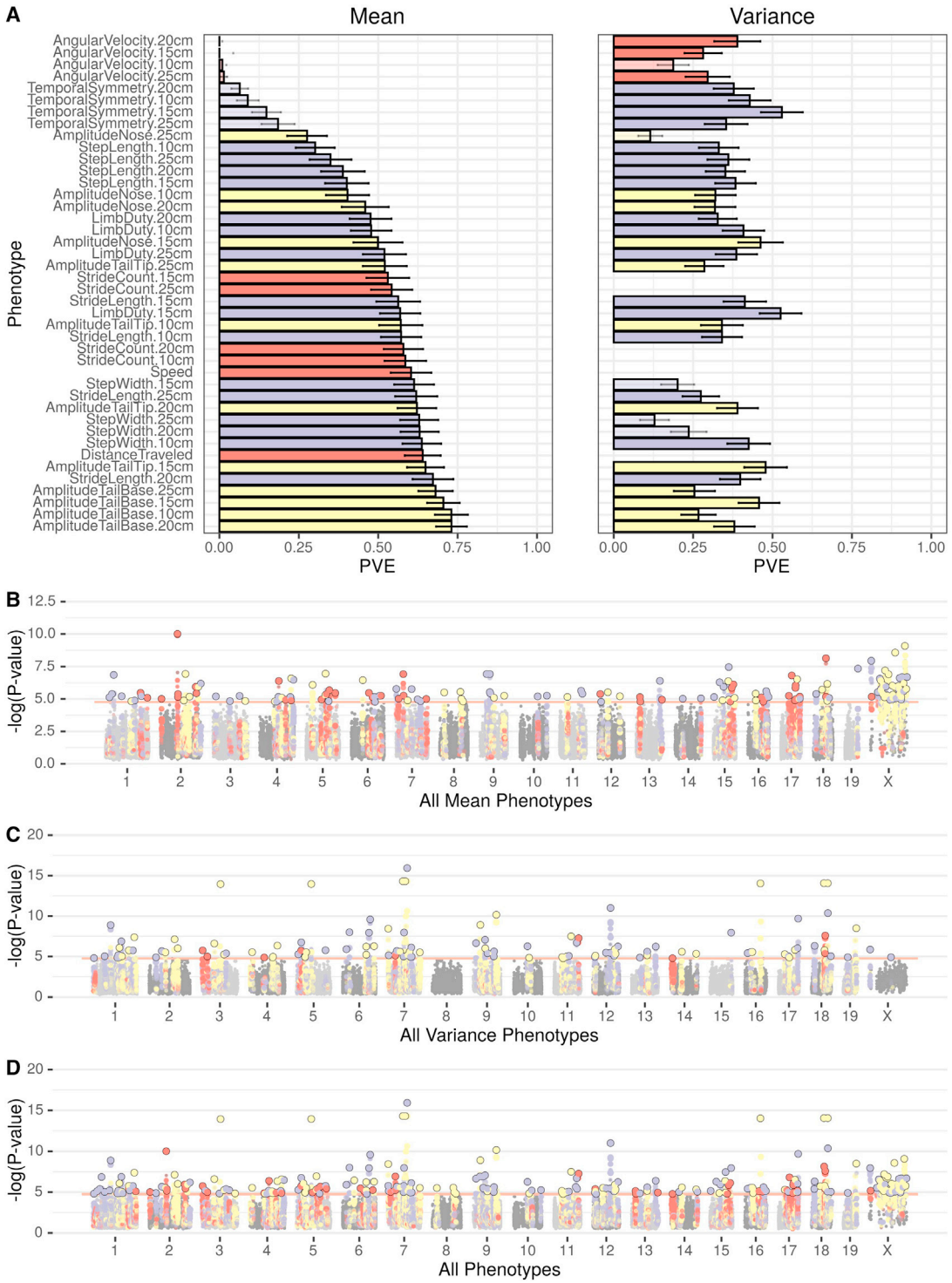


Figure 6. GWAS results for gait phenotypes

(A) Heritability estimates for each phenotype mean (left) and variance (right). Heritability is calculated as PVE (percent variance explained). Colors indicate posture (yellow), gait (blue), and open field (salmon) phenotypes.

(B–D) Manhattan plots of all mean phenotypes (B), variance phenotypes (C), and all of them combined (D); colors correspond to the phenotype, with the lowest p value for the SNP.

DISCUSSION

Gait and posture are important indicators of health, and are perturbed in many neurological, neuromuscular, and neuropsychiatric diseases. The goal of this project was to develop a system that performs pose estimation on mice to extract key gait and posture metrics in an open field from the top-down view. We present a solution that allows researchers to adapt a traditional top-down video imaging system used for open field analysis to extract gait and posture metrics. Our approach has some clear advantages as well as limitations. We are able to process a large amount of data with low effort and cost since the only data that needs to be captured is top-down grayscale video of a mouse in an open field, and all-pose estimation and gait metric extraction are fully automated after that. Top-down videos have routinely been used in behavioral neurogenetics, and this method could be applied to archival video data. We analyzed gait in a strain survey dataset that we partially analyzed for tracking and grooming behaviors (Geuther et al., 2019, 2021). Our method allows the animal to move of its own volition (unforced behavior) in a familiar environment (Jacobs et al., 2014). One limitation of our approach is that we cannot measure kinetic properties of gait because we are limiting ourselves to video (Lakes and Allen, 2016). We also limit ourselves to a 2D representation of pose because of our monocular recording configuration. A 3D representation of pose would allow for the extraction of height metrics from all key points and would likely provide richer gait phenotypes (Wiltschko et al., 2015). The decision to use top-down video also means that forepaw key points are often occluded by the mouse's body. The pose estimation network is robust to some amount of occlusion, as is the case with the hind paws, but the forepaws, which are almost always occluded during locomotion, have pose estimates that are too inaccurate and were excluded from our analysis. Regardless, in all genetic models that we tested, hind paw data are sufficient to detect robust differences in gait and body posture. In addition, we analyze videos at 30 Hz (frames per second) which is standard for video streams. Certain behaviors that occur at high speed such as escape or gallop may be difficult to determine. Kinematic approaches that view the animal from multiple angles and capture data at high frame rates may be more appropriate for certain applications (Machado et al., 2015, 2020). Thus, our methods are not a replacement for kinematic gait analysis that is carried out by certain specialized labs. These labs need higher-resolution approaches with enhanced video and kinetic analysis. Instead, our method offers a high-throughput assessment of gait in a commonly used behavior apparatus, the open field. We hope that our methods will allow these behavior labs to easily access gait and posture for additional biological insight. In addition, the ability to analyze large amounts of data in free-moving animals proves to be highly sensitive, even with very strict heuristic rules around what we consider to be a stride. Future iterations of our method could incorporate data from multiple camera angles and with higher frame rates. Even though we share our training data, code, and trained network weights, the implementation of our methods requires computational expertise. Future development efforts need to focus on turnkey solutions for non-computational labs. This is a problem faced by many of the advanced phenotyping methods that have been developed.

Gait and posture are frequently measured in humans as an endophenotype of psychiatric illness (Sanders and Gillig, 2010; Licari et al., 2020; Flyckt et al., 1999; Walther and Strik, 2012). Our results in mice indicate that gait and posture measures are highly heritable and perturbed in mutants. Specifically, we test neurodegenerative (*Sod1*), neurodevelopmental (Down syndrome, *Mecp2*), and ASD models (*Cntnap2*, *Shank3*, *FMR1*, *Del4Am*) and find altered gait features in all of these mutants. Others have also found similar results with neurodegenerative models (Machado et al., 2015). Of note are the data for Down syndrome. In humans, miscoordination and clumsiness are prominent features of Down syndrome. In mouse models, this miscoordination was previously characterized in inkblot gait assays as a disorganized hind footprint. Here, our analysis revealed perturbed posture differences between control and Tn65Dn mice. Our approach thus enables the quantitation of a previously qualitative trait. We also explored the extent to which differences in the emotionality of the animal could account for differences in gait parameters characterized by its movement in the open field. We split the data for C57BL/6J and C57BL/6NJ into the periphery (high anxiety) and center (low anxiety) strides and analyzed all of the gait and posture measures separately. To our surprise, we found that gait and stride-based posture metrics are identical in center and periphery. This additional analysis, in which we restricted the strides to particular speed (20–30 cm/s) and angular velocity (-20° to 20°/s) bins, found no differences between periphery and center. It indicated that, in C57BL/6J and C57BL/6NJ, gait and posture phenotypes are similar whether or not the animal is anxious. However, the question of independence between emotionality and gait mechanisms may require additional research.

Our analysis of a large number of mouse strains for gait and posture finds three distinct classes of overall movement. We find that the reference C57BL/6J and related strains belong to a distinct cluster separate from other common laboratory and wild-derived strains. We further explored the group containing the reference strain and found that the gait metrics are sensitive enough to distinguish animals between strains belonging to the C57 family (Figure S8). The main difference is seen in the high amplitude of tail and nose movement of the C57BL/6J and related strains. This may be important when analyzing gait and posture in differing genetic backgrounds. The GWAS revealed 254 QTL for gait and posture in the open field for both mean and variance phenotypes. We found that the mean and variance of traits are regulated by distinct genetic loci. We found that most variance phenotypes show moderate heritability, even for mean traits with low heritability. Human GWAS have been conducted for gait and posture, albeit with underpowered samples, which has led to good estimates of heritability, but only a few significantly associated loci (Adams et al., 2016). Similar to rodents, in humans, the variability of traits shows high heritability (Adams et al., 2016). In extended GWAS analysis, we find that a large number of QTL are dependent on body size, emphasizing the need to include this as a covariate in any gait and posture analysis. Enrichment analysis showed a loose set of GO terminologies that are enriched, indicating a wide array of biological

functions that regulates gait and posture. Altered gait and posture could result from QTLs that regulate neuronal or non-neuronal function (e.g., morphometrics). This is challenging to tease apart because specific genes can be expressed broadly and have varying functions during development. One could take a statistical approach to this problem by modeling morphometric features as covariates in GWAS linear mixed models (LMMs). For instance, if the size of the femur is thought to alter gait and posture, modeling its size as a parameter in the LMMs could reveal femur length-specific QTLs. Alternatively, single-gene studies using tissue or cell-type-specific knockouts may elucidate the functional roles of specific pathways to address this question. Our results in the mouse argue that a well-powered study in humans may identify hundreds of genetic factors that regulates gait and posture.

Limitations of the study

We are unable to precisely detect forepaws and limit our analysis to rear paws only. Certain symmetry gait metrics cannot be analyzed. The speed of our video data capture does not allow the detection of fast movements and could benefit from higher frame rate video. Video data do not permit the detection of kinetic events. We also limit our analysis to 2D images, and events that require data in the z direction are challenging.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Modifications to HRNet
  - Neural network training
  - Gait extraction
  - Corner detection
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
  - GWAS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.110231.

ACKNOWLEDGMENTS

We thank members of the Kumar lab for helpful advice and Taneli Helenius for editing. We thank Massimo Daul for the preliminary work on corner detection of our open fields. We thank Dr. Vivek Philip for advice on GWAS analysis. We thank JAX Information Technology team members Edwardo Zaborowski, Shane Sanders, Rich Brey, David McKenzie, and Jason Macklin for infrastructure support. This work was funded by The Jackson Laboratory Directors Innovation Fund, National Institutes of Health DA041668 (NIDA) and DA048634 (NIDA), and the Brain and Behavioral Research Foundation Young Investigator Award (V.K.). All code and training data will be available at Kumarlab.org and the Kumar lab Github upon publication.

AUTHOR CONTRIBUTIONS

V.K., K.S., G.S.S., and J.G. designed the experiments and analyzed the data. A.P. carried out GWAS. M.D. and S.D. generated the animal and training datasets. B.G. provided the animal video datasets. C.M.L. helped plan the gait validation studies. V.K., K.S., G.S.S., and A.P. wrote the manuscript.

DECLARATION OF INTERESTS

The Jackson Laboratory has filed a patent on the machine learning methods described in this paper.

Received: October 27, 2020
Revised: April 29, 2021
Accepted: December 16, 2021
Published: January 11, 2022

SUPPORTING CITATIONS

The following references appear in the supplemental information: Chawla et al., 2002.

REFERENCES

Adams, H.H., Verlinden, V.J., Callisaya, M.L., van Duijn, C.M., Hofman, A., Thomson, R., Uitterlinden, A.G., Vernooij, M.W., van der Geest, J.N., Srikanth, V., et al. (2016). Heritability and genome-wide association analyses of human gait suggest contribution of common variants. J. Gerontol. A Bio. Sci. Med. Sci. *71*, 740–746.

Allan, L.M., Ballard, C.G., Burn, D.J., and Kenny, R.A. (2005). Prevalence and severity of gait disorders in alzheimer's and non-alzheimer's dementias. J. Am. Geriatr. Soc. *53*, 1681–1687.

Amende, I., Kale, A., McCue, S., Glazier, S., Morgan, J.P., and Hampton, T.G. (2005). Gait dynamics in mouse models of Parkinson's disease and huntington's disease. J. Neuroeng. Rehabil. *2*, 20.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in x-linked mecp2, encoding methyl-cpg-binding protein 2. Nat. Genet. *23*, 185–188.

Aristotle. (2004). On the Gait of Animals (Kessinger Publishing), ISBN 9781419138676.

Baker, R. (2007). The history of gait analysis before the advent of modern computers. Gait Posture *26*, 331–342.

Baldaçara, L., Fiorani Borgio, J.G., Tavares de Lacerda, A.L., and Jackowski, A.P. (2008). Cerebellum and psychiatric disorders. Braz. J. Psychiatry *30*, 281–289.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

Batka, R.J., Brown, T.J., Mcmillan, K.P., Meadows, R.M., Jones, K.J., and Haulcomb, M.M. (2014). The need for speed in rodent locomotion analyses. Anat. Rec. *297*, 1839–1864.

Bellardita, C., and Kiehn, O. (2015). Phenotypic characterization of speed-associated gait changes in mice reveals modular organization of locomotor networks. Curr. Biol. *25*, 1426–1436.

Belmonte, M.K., and Bourgeron, T. (2006). Fragile x syndrome and autism at the intersection of genetic and neural networks. Nat. Neurosci. *9*, 1221–1225.

Billauer, E. (2012). peakdet: peak detection using matlab.http://billauer.co.il/peakdet.html.

Borelli, G.A., and Maquet, P. (2012). On the Movement of Animals (Springer Berlin Heidelberg), ISBN 9783642738128.

Braun, M. (1992). Picturing Time: The Work of Etienne-Jules Marey (1830-1904).

Broom, L., Ellison, B.A., Worley, A., Wagenaar, L., Sörberg, E., Ashton, C., Bennett, D.A., Buchman, A.S., Saper, C.B., Shih, L.C., et al. (2017). A translational approach to capture gait signatures of neurological disorders in mice and humans. Sci. Rep. *7*, 1–17.

Brunner, D., Kabitzke, P., He, D., Cox, K., Thiede, L., Hanania, T., Sabath, E., Alexandrov, V., Saxe, M., Peles, E., et al. (2015). Comprehensive analysis of the 16p11. 2 deletion and null cntnap2 mouse models of autism spectrum disorder. PLoS One *10*, e0134572.

Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. (2002). Smote: synthetic minority over-sampling technique. J. Artif. Intell. Res. *16*, 321–357.

Costa, A.C.S., Walsh, K., and Davisson, M.T. (1999). Motor dysfunction in a mouse model for down syndrome. Physiol. Behav. *68*, 211–220.

Crawford, D.C., Acuña, J.M., and Sherman, S.L. (2001). Fmr1 and the fragile x syndrome: human genome epidemiology review. Genet. Med. *3*, 359–371.

Crawley, J.N. (2007). What's wrong with my mouse?. Behavioral Phenotyping of Transgenic and Knockout Mice (John Wiley & Sons).

Dang, Q., Yin, J., Wang, B., and Zheng, W. (2019). Deep learning based 2d human pose estimation: a survey. Tsinghua Sci. Technol. *24*, 663–676.

Dell, A.I., Bender, J.A., Branson, K., Couzin, I.D., de Polavieja, G.G., Noldus, L.P., Pérez-Escudero, A., Perona, P., Straw, A.D., Wikelski, M., et al. (2014). Automated image-based tracking and its application in ecology. Trends Ecol. Evol. *29*, 417–428.

Durand, C.M., Betancur, C., Boeckers, T.M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I.C., Anckarsäter, H., et al. (2007). Mutations in the gene encoding the synaptic scaffolding protein shank3 are associated with autism spectrum disorders. Nat. Genet. *39*, 25–27.

Egnor, S.E.R., and Branson, K. (2016). Computational analysis of behavior. Annu. Rev. Neurosci. *39*, 217–236.

Faizi, M., Bader, P.L., Tun, C., Encarnacion, A., Kleschevnikov, A., Belichenko, P., Saw, N., Priestley, M., Tsien, R.W., Mobley, W.C., et al. (2011). Comprehensive behavioral phenotyping of ts65dn mouse model of down syndrome: activation of β1-adrenergic receptor by xamoterol as a potential cognitive enhancer. Neurobiol. Dis. *43*, 397–413.

Fisher, N.I., and Lee, A.J. (1992). Regression Models for an Angular Response (Biometrics), pp. 665–677.

Flyckt, L., Sydow, O., Bjerkenstedt, L., Edman, G., Rydin, E., and Wiesel, F.-A. (1999). Neurological signs and psychomotor performance in patients with schizophrenia, their relatives and healthy controls. Psychiatry Res. *86*, 113–129.

Gadalla, K.K.E., Ross, P.D., Riddell, J.S., Bailey, M.E.S., and Cobb, S.R. (2014). Gait analysis in a mecp2 knockout mouse model of rett syndrome reveals early-onset and progressive motor deficits. PLoS One *9*, e112889.

Geuther, B.Q., Deats, S.P., Fox, K.J., Murray, S.A., Braun, R.E., White, J.K., Chesler, E.J., Lutz, C.M., and Kumar, V. (2019). Robust mouse tracking in complex environments using neural networks. Commun. Biol. *2*, 1–11.

Geuther, B.Q., Peer, A., He, H., Sabnis, G., Philip, V.M., and Kumar, V. (2021). Action detection using a neural network elucidates the genetics of mouse grooming behavior. Elife *10*, e63207.

Green, D., Charman, T., Pickles, A., Chandler, S., Loucas, T., Simonoff, E., and Baird, G. (2009). Impairment in movement skills of children with autistic spectrum disorders. Dev. Med. Child Neurol. *51*, 311–316.

Greenberg, G., and Haraway, M.M. (1998). Comparative Psychology: A Handbook (Taylor & Francis), ISBN 9781136794506.

Guillot, T.S., Asress, S.A., Richardson, J.R., Glass, J.D., and Miller, G.W. (2008). Treadmill gait analysis does not detect motor deficits in animal models of Parkinson's disease or amyotrophic lateral sclerosis. J. Mot. Behav. *40*, 568–577.

Gurney, M.E., Pu, H., Chiu, A.Y., Dal Canto, M.C., Polchow, C.Y., Alexander, D.D., Caliendo, J., Hentati, A., Kwon, Y.W., Deng, H.-X., et al. (1994). Motor neuron degeneration in mice that express a human cu, zn superoxide dismutase mutation. Science *264*, 1772–1775.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse mecp2-null mutation causes neurological symptoms that mimic rett syndrome. Nat. Genet. *27*, 322–326.

Hall, C.S. (1934). Emotional behavior in the rat. i. defecation and urination as measures of individual differences in emotionality. J. Comp. Psychol. *18*, 385.

Hampton, T.G., and Amende, I. (2009). Treadmill gait analysis characterizes gait alterations in Parkinson's disease and amyotrophic lateral sclerosis mouse models *42*, 1–4.

Hampton, T.G., Stasko, M.R., Kale, A., Amende, I., and Costa, A.C.S. (2004). Gait dynamics in trisomic mice: quantitative neurological traits of down syndrome. Physiol. Behav. *82*, 381–389.

Haslam, R.H.A. (1995). Neurological problems associated with down syndrome. In Down Syndrome: Living and Learning in the Community, Lynn Nadel and Donna Rosenthal, eds. (Wiley-Liss), pp. 107–114.

Hausdorff, J.M., Peng, C.-K., Goldberger, A.L., and Stoll, A.L. (2004). Gait unsteadiness and fall risk in two affective disorders: a preliminary study. BMC Psychiatry *4*, 39.

Herault, Y., Delabar, J.M., Fisher, E.M.C., Tybulewicz, V.L.J., Yu, E., and Brault, V. (2017). Rodent models in down syndrome research: impact and future opportunities. Dis. Models Mech. *10*, 1165–1186.

Hildebrand, M. (1977). Analysis of asymmetrical gaits. J. Mammal. *58*, 131–156.

Hildebrand, M. (1989). The quadrupedal gaits of vertebrates. Bioscience *39*, 766.

Horev, G., Ellegood, J., Lerch, J.P., Son, Y.-E.E., Muthuswamy, L., Vogel, H., Krieger, A.M., Buja, A., Henkelman, R.M., Wigler, M., et al. (2011). Dosage-dependent phenotypes in models of 16p11. 2 lesions found in autism. Proc. Natl. Acad. Sci. U S A *108*, 17076–17081.

Hughes, V. (2011). Motor Problems in Autism Move into Research Focus (Spectrum News). https://www.spectrumnews.org/news/motor-problems-in-autism-move-into-research-focus/.

Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., and Schiele, B. (2016). Deepercut: a deeper, stronger, and faster multi-person pose estimation model. In European Conference on Computer Vision (Springer), pp. 34–50.

Jacobs, B.Y., Kloefkorn, H.E., and Allen, K.D. (2014). Gait analysis methods for rodent models of osteoarthritis. Curr. Pain Headache Rep. *18*, 456.

Jammalamadaka, S.R., and Sengupta, A. (2001). Topics in Circular Statistics, *volume 5* (World Scientific).

Kerr, B., Silva, P.A., Walz, K., and Young, J.I. (2010). Unconventional transcriptional response to environmental enrichment in a mouse model of rett syndrome. PLoS One *5*, e11534.

Kingma, D.P., and Ba, J. (2014). Adam: a method for stochastic optimization. arXiv, arXiv:1412.6980.

Kirtley, C. (2006). Clinical Gait Analysis: Theory and Practice (Elsevier Health Sciences).

Knippenberg, S., Thau, N., Dengler, R., and Petri, S. (2010). Significance of behavioural tests in a transgenic mouse model of amyotrophic lateral sclerosis (ALS). Behav. Brain Res. *213*, 82–87.

Kumar, V., Kim, K., Joseph, C., Thomas, L.C., Hong, H., and Takahashi, J.S. (2011). Second-generation high-throughput forward genetic screen in mice to isolate subtle behavioral mutants. Proc. Natl. Acad. Sci. U S A *108*, 15557–15564.

Laird, N.M., and Ware, J.H. (1982). Random-effects models for longitudinal data. Biometrics, 963–974.

Lakes, E.H., and Allen, K.D. (2016). Gait analysis methods for rodent models of arthritic disorders: reviews and recommendations. Osteoarthr. Cartil. *24*, 1837–1849.

Lanska, D.J. (2016). The dercum-muybridge collaboration and the study of pathologic gaits using sequential photography. J. Hist. Neurosci. *25*, 23–38.

Latash, M.L. (2000). Motor coordination in down syndrome: the role of adaptive changes. In Perceptual-motor Behavior in Down Syndrome (Wolters Kluwer), pp. 199–223.

Lee, P.H., O'Dushlaine, C., Thomas, B., and Purcell, S.M. (2012). INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics *28*, 1797–1799, ISSN 1367-4803. https://doi.org/10.1093/bioinformatics/bts191.

Licari, M.K., Alvares, G.A., Varcin, K., Evans, K.L., Cleary, D., Reid, S.L., Glasson, E.J., Bebbington, K., Reynolds, J.E., Wray, J., et al. (2020). Prevalence of motor difficulties in autism spectrum disorder: analysis of a population-based cohort. Autism Res. *13*, 298–306.

Lund, U., Agostinelli, C., and Agostinelli, M.C. (2017). Package 'Circular' (Repository CRAN).

Machado, A.S., Darmohray, D.M., Fayad, J., Marques, H.G., and Carey, M.R. (2015). A quantitative framework for whole-body coordination reveals specific deficits in freely walking ataxic mice. Elife *4*, e07892.

Machado, A.S., Marques, H.G., Duarte, D.F., Darmohray, D.M., and Carey, M.R. (2020). Shared and specific signatures of locomotor ataxia in mutant mice. Elife *9*, e55356.

Mancuso, R., Oliván, S., Osta, R., and Navarro, X. (2011). Evolution of gait abnormalities in SOD1G93A transgenic mice. Brain Res. *1406*, 65–73.

Manjila, S., Singh, G., Alkhachroum, A.M., and Ramos-Estebanez, C. (2015). Understanding edward muybridge: historical review of behavioral alterations after a 19th-century head injury and their multifactorial influence on human life and culture. Neurosurg. Focus *39*, E4.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). Deeplabcut: Markerless Pose Estimation of User-Defined Body Parts with Deep Learning, Technical report (Nature Publishing Group).

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv, arXiv:1802.03426.

McIntosh, G.C., Brown, S.H., Rice, R.R., and Thaut, M.H. (1997). Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease. J. Neurol. Neurosurg. Psychiatry *62*, 22–26.

Mead, R.J., Bennett, E.J., Kennerley, A.J., Sharp, P., Sunyach, C., Kasher, P., Berwick, J., Pettmann, B., Battaglia, G., Azzouz, M., et al. (2011). Optimised and rapid pre-clinical screening in the SOD1G93A transgenic mouse model of amyotrophic lateral sclerosis (ALS). PLoS One *6*, e23244.

Moeslund, T.B., Hilton, A., and Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Understand. *104*, 90–126.

Morris, A.F., Vaughan, S.E., and Vaccaro, P. (1982). Measurements of neuromuscular tone and strength in down's syndrome children. J. Ment. Defic. Res. *26*, 41–46.

Morse, H.C., III. (2012). Origins of Inbred Mice (Elsevier Science), ISBN 9780323142830. https://books.google.com/books?id=mRyp1vKI3fsC.

Newell, A., Yang, K., and Deng, J. (2016). Stacked hourglass networks for human pose estimation. In European Conference on Computer Vision (Springer), pp. 483–499.

Nutt, J.G., Marsden, C.D., and Thompson, P.D. (1993). Human walking and higher-level gait disorders, particularly in the elderly. Neurology *43*, 268.

Paszke, Adam, et al. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems *32*, 8026–8037.

Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.S.-H., Murthy, M., and Shaevitz, J.W. (2019). Fast animal pose estimation using deep neural networks. Nat. Methods *16*, 117–125.

Pinheiro, J., and Bates, D. (2000). Mixed-effects Models in S and S-PLUS (Springer-Verlag).

Poliak, S., Gollan, L., Martinez, R., Custer, A., Einheber, S., Salzer, J.L., Trimmer, J.S., Shrager, P., and Peles, E. (1999). Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with k+ channels. Neuron *24*, 1037–1047.

Preisig, D.F., Kulic, L., Krüger, M., Wirth, F., McAfoose, J., Späni, C., Gantenbein, P., Derungs, R., Nitsch, R.M., and Welt, T. (2016). High-speed video gait analysis reveals early and characteristic locomotor phenotypes in mouse models of neurodegenerative movement disorders. Behav. Brain Res. *311*, 340–353.

Rao, C.R. (1948). The utilization of multiple measurements in problems of biological classification. J. R. Stat. Soc. Ser. B (Methodological) *10*, 159–193.

Reeves, R.H., Irving, N.G., Moran, T.H., Wohn, A., Kitt, C., Sisodia, S.S., Schmidt, C., Bronson, R.T., and Davisson, M.T. (1995). A mouse model for Down syndrome exhibits learning and behaviour deficits. Nat. Genet. *11*, 177–184.

Robinson, L., Guy, J., McKay, L., Brockett, E., Spike, R.C., Selfridge, J., De Sousa, D., Merusi, C., Riedel, G., Bird, A., et al. (2012). Morphological and functional reversal of phenotypes in a mouse model of rett syndrome. Brain *135*, 2699–2710.

Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., Deng, H.-X., et al. (1993). Mutations in cu/zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature *362*, 59–62.

Sanders, R.D., and Gillig, P.M. (2010). Gait and its assessment in psychiatry. Psychiatry (Edgmont) *7*, 38.

Santos, M., Summavielle, T., Teixeira-Castro, A., Silva-Fernandes, A., Duarte-Silva, S., Marques, F., Martins, L., Dierssen, M., Oliveira, P., Sousa, N., et al. (2010). Monoamine deficits in the brain of methyl-cpg binding protein 2 null mice suggest the involvement of the cerebral cortex in early stages of rett syndrome. Neuroscience *170*, 453–467.

Scherder, E., Eggermont, L., Swaab, D., van Heuvelen, M., Kamsma, Y., de Greef, M., van Wijck, R., and Mulder, T. (2007). Gait in ageing and associated dementias; its relationship with cognition. Neurosci. Biobehav. Rev. *31*, 485–497.

Shumway-Cook, A., and Woollacott, M.H. (1985). Dynamics of postural control in the child with down syndrome. Phys. Ther. *65*, 1315–1322.

Sun, K., Xiao, B., Liu, D., and Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5693–5703.

Takakusaki, K. (2013). Neurophysiology of gait: from the spinal cord to the frontal lobe. Mov. Disord. *28*, 1483–1491.

Takakusaki, K. (2017). Functional neuroanatomy for posture and gait control. J. Mov. Disord. *10*, 1.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). A survey on deep transfer learning. In International Conference on Artificial Neural Networks (Springer), pp. 270–279.

Tesla, R., Wolf, H.P., Xu, P., Drawbridge, J., Estill, S.J., Huntington, P., McDaniel, L., Knobbe, W., Burket, A., Tran, S., et al. (2012). Neuroprotective efficacy of aminopropyl carbazoles in a mouse model of amyotrophic lateral sclerosis. Proc. Natl. Acad. Sci. U S A *109*, 17016–17021.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B (Statistical Methodology) *63*, 411–423.

Toma, C., Pierce, K.D., Shaw, A.D., Heath, A., Mitchell, P.B., Schofield, P.R., and Fullerton, J.M. (2018). Comprehensive cross-disorder analyses of cntnap2 suggest it is unlikely to be a primary risk gene for psychiatric disorders. PLoS Genet. *14*, e1007535.

Verghese, J., Lipton, R.B., Hall, C.B., Kuslansky, G., Katz, M.J., and Buschke, H. (2002). Abnormality of gait as a predictor of non-alzheimer's dementia. N. Engl. J. Med. *347*, 1761–1768.

Vergouts, M., Marinangeli, C., Ingelbrecht, C., Genard, G., Schakman, O., Sternotte, A., Calas, A.-G., and Hermans, E. (2015). Early ALS-type gait abnormalities in AMP-dependent protein kinase-deficient mice suggest a role for this metabolic sensor in early stages of the disease. Metab. Brain Dis. *30*, 1369–1377.

Vimercati, S.L., Galli, M., Stella, G., Caiazzo, G., Ancillao, A., and Albertini, G. (2015). Clumsiness in fine motor tasks: evidence from the quantitative drawing evaluation of children with d own s yndrome. J. Intellect. Disabil. Res. *59*, 248–256.

Walther, S., and Strik, W. (2012). Motor symptoms and schizophrenia. Neuropsychobiology *66*, 77–92.

Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A.R., Green, T., et al. (2008). Association between microdeletion and microduplication at 16p11. 2 and autism. N. Engl. J. Med. *358*, 667–675.

Weiss, K., Khoshgoftaar, T.M., and Wang, D.D. (2016). A survey of transfer learning. J. Big Data *3*, 9.

Wiltschko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abraira, V.E., Adams, R.P., and Robert Datta, S. (2015). Mapping sub-second structure in mouse behavior. Neuron *88*, 1121–1135.

Wooley, C.M., Sher, R.B., Kale, A., Frankel, W.N., Cox, G.A., and Seburn, K.L. (2005). Gait analysis detects early changes in transgenic sod1 (g93a) mice. Muscle Nerve *32*, 43–50.

Zeliadt, N. (2017). Autism in motion: could motor problems trigger social ones. Scientific American, Spectrum, Mental Health.https://www.scientificamerican.com/article/autism-in-motion-could-motor-problems-trigger-social-ones/(uv 11/7/2017).

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. *44*, 821–824.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Mouse: 129P3/J	The Jackson Laboratory	JAX: 000690
Mouse: 129S1/SvImJ	The Jackson Laboratory	JAX: 002448
Mouse: 129X1/SvJ	The Jackson Laboratory	JAX: 000691
Mouse: A/J	The Jackson Laboratory	JAX: 000646
Mouse: AKR/J	The Jackson Laboratory	JAX: 000648
Mouse: B6129PF1/J	The Jackson Laboratory	JAX: 100492
Mouse: B6129SF1/J	The Jackson Laboratory	JAX: 101043
Mouse: B6AF1/J	The Jackson Laboratory	JAX: 100002
Mouse: B6C3F1/J	The Jackson Laboratory	JAX: 100010
Mouse: B6CBAF1/J	The Jackson Laboratory	JAX: 100011
Mouse: B6D2F1/J	The Jackson Laboratory	JAX: 100006
Mouse: B6SJLF1/J	The Jackson Laboratory	JAX: 100012
Mouse: BALB/cByJ	The Jackson Laboratory	JAX: 001026
Mouse: BALB/cJ	The Jackson Laboratory	JAX: 000651
Mouse: BTBR T+ Itpr3tf/J	The Jackson Laboratory	JAX: 002282
Mouse: BUB/BnJ	The Jackson Laboratory	JAX: 000653
Mouse: C3HeB/FeJ	The Jackson Laboratory	JAX: 000658
Mouse: C3H/HeJ	The Jackson Laboratory	JAX: 000659
Mouse: C3H/HeOuJ	The Jackson Laboratory	JAX: 000635
Mouse: C57BL/10SnJ	The Jackson Laboratory	JAX: 000666
Mouse: C57BL/6J	The Jackson Laboratory	JAX: 000664
Mouse: C57BL/6NJ	The Jackson Laboratory	JAX: 005304
Mouse: C57BLKS/J	The Jackson Laboratory	JAX: 000662
Mouse: C57BR/cdJ	The Jackson Laboratory	JAX: 000667
Mouse: C57L/J	The Jackson Laboratory	JAX: 000668
Mouse: C58/J	The Jackson Laboratory	JAX: 000669
Mouse: CAF1/J	The Jackson Laboratory	JAX: 100003
Mouse: CAST/EiJ	The Jackson Laboratory	JAX: 000928
Mouse: CB6F1/J	The Jackson Laboratory	JAX: 100007
Mouse: CBA/CaJ	The Jackson Laboratory	JAX: 000654
Mouse: CBA/J	The Jackson Laboratory	JAX: 000656
Mouse: CByB6F1/J	The Jackson Laboratory	JAX: 100009
Mouse: CZECHII/EiJ	The Jackson Laboratory	JAX: 001144
Mouse: DBA/1J	The Jackson Laboratory	JAX: 000670
Mouse: DBA/2J	The Jackson Laboratory	JAX: 000671
Mouse: FVB/NJ	The Jackson Laboratory	JAX: 001800
Mouse: I/LnJ	The Jackson Laboratory	JAX: 000674
Mouse: KK/HlJ	The Jackson Laboratory	JAX: 002106
Mouse: LG/J	The Jackson Laboratory	JAX: 000675
Mouse: LP/J	The Jackson Laboratory	JAX: 000676
Mouse: MA/MyJ	The Jackson Laboratory	JAX: 000677
Mouse: MOLF/EiJ	The Jackson Laboratory	JAX: 000550
Mouse: MRL/MpJ	The Jackson Laboratory	JAX: 000486

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: MSM/MsJ	The Jackson Laboratory	JAX: 003719
Mouse: NOD/ShiLtJ	The Jackson Laboratory	JAX: 001976
Mouse: NON/ShiLtJ	The Jackson Laboratory	JAX: 002423
Mouse: NOR/LtJ	The Jackson Laboratory	JAX: 002050
Mouse: NU/J	The Jackson Laboratory	JAX: 002019
Mouse: NZB/BINJ	The Jackson Laboratory	JAX: 000684
Mouse: NZBWF1/J	The Jackson Laboratory	JAX: 100008
Mouse: NZO/HlLtJ	The Jackson Laboratory	JAX: 002105
Mouse: NZW/LacJ	The Jackson Laboratory	JAX: 001058
Mouse: PL/J	The Jackson Laboratory	JAX: 000680
Mouse: PWD/PhJ	The Jackson Laboratory	JAX: 004660
Mouse: PWK/PhJ	The Jackson Laboratory	JAX: 003715
Mouse: RIIIS/J	The Jackson Laboratory	JAX: 000683
Mouse: SEA/GnJ	The Jackson Laboratory	JAX: 000644
Mouse: SJL/J	The Jackson Laboratory	JAX: 000686
Mouse: SM/J	The Jackson Laboratory	JAX: 000687
Mouse: SWR/J	The Jackson Laboratory	JAX: 000689
Mouse: TALLYHO/JngJ	The Jackson Laboratory	JAX: 005314
Mouse: WSB/EiJ	The Jackson Laboratory	JAX: 001145
Mouse: B6.129P2(C)-Mecp2tm1.1Bird/J	The Jackson Laboratory	JAX: 0038901
Mouse: B6.Cg-Tg(SOD1*G93A)1Gur/J	The Jackson Laboratory	JAX: 0044351
Mouse: B6EiC3Sn.BLiA-Ts(1716)65Dn/DnJ	The Jackson Laboratory	JAX: 0052521
Mouse: B6EiC3Sn.BLiAF1/J	The Jackson Laboratory	JAX: 0036471
Mouse: B6.129P2-Fmr1tm1Cgr/J	The Jackson Laboratory	JAX: 0030251
Mouse: B6.129S-Del(7Slx1b-Sept1)4Aam/J	The Jackson Laboratory	JAX: 0131281
Mouse: B6.129-Shank3tm2Gfn/J	The Jackson Laboratory	JAX: 0176881
Mouse: B6.129(Cg)-Cntnap2tm1Pele/J	The Jackson Laboratory	JAX: 0286351
Software and algorithms		
Training, validation data, neural network weights	This paper	https://doi.org/10.5281/zenodo.5708437
Gait extraction code version of record	This paper	https://doi.org/10.5281/zenodo.5725641
PyTorch	(Paszke, 2019)	https://pytorch.org/
R	R Core Team	https://www.r-project.org/
Deposited data		
Raw and analyzed data	This paper	MPD: Kumar4. Mouse Phenome Database web resource (RRID:SCR_003212), The Jackson Laboratory, Bar Harbor, Maine USA. https://phenome.jax.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Vivek Kumar (vivek.kumar@jax.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The training and validation data have been deposited within a Zenodo repository at https://doi.org/10.5281/zenodo.5708437 and are publicly available as of the date of publication.

- All original code has been deposited under the Kumar Lab's Github page (https://github.com/KumarLabJax/, https://doi.org/10.5281/zenodo.5725641) within the deep-hrnet-mouse and gaitanalysis repositories using the open source MIT License.
- All phenotype data has been deposited in Mouse Phenome database under "Kumar4" https://phenome.jax.org/ under "Kumar4".
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All behavioral tests were performed in accordance with approved protocols from The Jackson Laboratory Institutional Animal Care and Use Committee guidelines. All animals were obtained from JAX repository or bred in a room adjacent to the behavioral testing room (key resources table). All behavioral protocols have been previously published (Geuther et al., 2019; Kumar et al., 2011). Our open field arena measures 52 cm by 52 cm by 23 cm. The floor is white PVC plastic and the walls are gray PVC plastic. To aid in cleaning maintenance, a white 2.54 cm chamfer was added to all the inner edges. Illumination is provided by an LED ring light (Model: F&V R300). The ring light was calibrated to produce 600 lux of light in each of our 24 arenas. The strain survey data was published before and reanalyzed for gait behavior (Geuther et al., 2019, 2021). We excluded animals that had too few strides which disproportionately affected low activity strains (key resources table). All gait mutants and ASD models were generated in JAX repository and genotyped prior to shipment to Kumar Lab for testing. Animals were acclimated for at least a week prior to any testing. Prior to open field testing, animals were moved to the behavior room and allowed to acclimate for 30-60 minutes. White noise was used for testing to balance the noise between holding and testing rooms. All animals were between 10 and 20 weeks old. The following sexes were used for each strain: 129P3/J (8M, 15F), 129S1/SvlmJ (13M, 4F), 129X1/SvJ (8M, 7F), A/J (4M, 2F), AKR/J (9M, 8F), B6129PF1/J (10M, 20F), B6129SF1/J (15M, 9F), B6AF1/J (17M, 18F), B6C3F1/J (12M, 20F), B6CBAF1/J (9M, 9F), B6D2F1/J (12M, 10F), B6SJLF1/J (5M, 23F), BALB/cByJ (11M, 7F), BALB/cJ (19M, 2F), BTBR T<+>Itpr3<tf>/J (32M, 21F), BUB/BnJ (7M, 8F), C3H/HeJ (11M, 16F), C3H/HeOuJ (6M, 13F), C3HeB/FeJ (5M, 5F), C57BL/10SnJ (10M, 9F), C57BL/6J (298M, 196F), C57BL/6NJ (126M, 167F), C57BLKS/J (19M, 11F), C57BR/cdJ (3M, 12F), C57L/J (10M, 13F), C58/J (7M, 4F), CAF1/J (8M, 6F), CAST/EiJ (10M, 23F), CB6F1/J (18M, 9F), CBA/CaJ (15M, 15F), CBA/J (9M, 5F), CByB6F1/J (14M, 4F), CZECHII/EiJ (4M, 7F), DBA/1J (12M, 15F), DBA/2J (8M, 9F), FVB/NJ (5M, 8F), I/LnJ (6M, 8F), KK/HiJ (5M, 3F), LG/J (3M, 3F), LP/J (15M, 10F), MA/MyJ (7M, 8F), MOLF/EiJ (3M, 6F), MRL/MpJ (4M, 8F), MSM/MsJ (3M, 8F), NOD/ShiLtJ (13M, 12F), NON/ShiLtJ (13M, 14F), NOR/LtJ (7M, 6F), NU/J (5M, 5F), NZB/BINJ (5M, 16F), NZBWF1/J (9M, 8F), NZO/HILtJ (6M, 8F), NZW/LacJ (7M, 4F), PL/J (4M, 8F), PWD/PhJ (7M, 5F), PWK/PhJ (5M, 4F), RIIIS/J (3M, 7F), SEA/GnJ (4M, 3F), SJL/J (4M, 30F), SM/J (8M, 4F), SWR/J (2M, 10F), TALLYHO/JngJ (13M, 9F), WSB/EiJ (8M, 3F).

METHOD DETAILS

Modifications to HRNet

We made some modifications to the HRNet architecture in order to make it work well for our experimental configuration. We used the smaller HRNet-W32 architecture rather than HRNet-W48 because it was shown to provide significant speed and memory improvements for only a small reduction in accuracy (Sun et al., 2019). We added two 5x5 transpose convolutions to the head of the network to match the heatmap output resolution with the resolution of the video input (Figure S1A). Because all of our experiments have a single mouse in an open field, we do not need to rely on object detection for instancing. We thus eliminated this step from our inference algorithm, which also leads to clear runtime performance benefits. Instead of performing pose estimation after object detection, we use the full resolution pose keypoint heatmaps to infer the posture of a single mouse at every frame. This means that for each 480x480 frame of video we generate 12 480x480 heatmaps (one heatmap per keypoint). The maximum value in each heatmap represents the highest confidence location for each respective point. Thus, after taking the argmax of each of the 12 heatmaps we have 12 (x, y) coordinates.

Neural network training

For training our network, we use the same loss function that is used in the original HRNet description (Sun et al., 2019). For each keypoint label, we generate a 2D gaussian distribution centered on the respective keypoint. We then compare the output of the network with our keypoint-centered Gaussian and calculate loss as the mean squared difference between our labeled keypoint Gaussian and the heatmap generated by our network. We train our network over 600 epochs and perform validation at the end of every epoch. The training loss curves (Figure S1) show a fast convergence of the training loss without an overfitting of the validation loss. We used transfer learning (Weiss et al., 2016; Tan et al., 2018) on our network in order to minimize the labeling requirements and improve the generality of our model. We started with the imagenet model provided by the authors of the HRNet paper (hrnet_w32-36af842e.pth) and froze the weights up to the second stage during training. In order to further improve the generality of our network we employed several data augmentation techniques during training including: rotation, flipping, scaling, brightness, contrast and occlusion. We train our network using the ADAM optimization algorithm which is a variant of stochastic gradient descent (Kingma and Ba, 2014). The learning rate is initially set to 5×10^{-4}, then reduced to 5×10^{-5} at the 400*th* epoch and 5×10^{-6} at the 500*th* epoch.

We generated labels that represent a wide diversity of mouse appearances, including variation in coat color, body length and obesity to ensure that the resulting network operates robustly across these differences. We manually labeled 8,910 frames across these diverse strains for training. The resulting network is able to track dozens of mouse strains with varying body size, shape and coat color (Video S1) (Geuther et al., 2019). We calculate the accuracy of our neural network and experimental configuration over two datasets: a set of 1000 images with results for 200 white mice and 200 dark mice broken out (Figure S2B) as well as a set of 120 images containing twenty images each from a set of six visually diverse mouse strains (Figure S2C). We provide these metrics in pixel and centimeter distance units.

Gait extraction

Here we describe our method of extracting gait structure from pose in further detail. The processes of detecting strides begins with first determining intervals of time where the mouse is moving at sufficient speed for strides to take place. We will call these tracks. To determine track intervals we threshold for base of tail speed greater than or equal to 5 cm/sec. Through observation we determined that the base of tail point is highly stable and a good surrogate for overall mouse speed.

The next step is to identify individual steps in both the left hind paw and right hind paw. Initially these steps are determined for each paw respectively without consideration for the other paw. Later in the process we will pair up left and right steps into strides. The process for step detection relies on oscillations in speed as can be seen in (Figure 1D). We calculate individual paw speed and then apply a peak detection algorithm (Billauer, 2012) to identify local maxima in speed. After finding all of the local maxima we use the surrounding local minima to define a step interval with a toe-off event followed by a foot strike event on either side of the step. We then filter out any steps whose peak speed does not exceed 15 cm/sec or the overall animal speed (whichever is greater).

Once we have each set of valid steps from left and right hind paws we need to group pairs of steps together to find strides. We use left hind paw steps to delimit strides. Stride intervals end when the left hind paw step ends and begin at the frame just after the previous stride. There is an additional constraint that stride intervals are not allowed to extend before or after the containing track. After we have defined our stride intervals using left hind paw steps we need to associate right hind paw steps with the stride. If we find a right hind paw step that completes within the given stride interval we then associate that step with the stride. If we cannot find a right hind step that meets this condition the stride is discarded.

Now that we have all of our stride definitions with associated steps we apply additional filtering to improve the quality and consistency of strides. For our application of statistical and genomic analysis of stride metrics we decided to take a fairly aggressive approach at removing strides that have potential to degrade quality or introduce inconsistency. All strides at the start and end of a track are removed. This is done to improve the consistency of gait metrics and avoid introducing variance due to starting and stopping behavior. We also discard strides if the keypoint confidence for Nose, Neck Base, Spine Center, Tail Base, Hind Paw Left, Hind Paw Right, Tail Middle or Tail Tip (Figure 1G) falls below the threshold of 0.3 for any frame in the stride in order to avoid using low quality strides.

Corner detection

In order to perform a comparative analysis of gait metrics between center and periphery of the open field arena (Figure S7) we need to know pixel locations for each of the four corners. Rather than use fixed values which would have been affected by differences in camera placement, we developed a simple corner detector using the same neural network architecture that we developed for pose estimation. The only difference in the network structure is the output; the pose estimation network outputs twelve heatmaps (one per keypoint) whereas the corner detection network just outputs a single heatmap for detecting corner positions. We trained our corner detector using 415 annotated images. We tested accuracy on the 27 validation images which use the same 480x480 resolution that we use for the video in this paper. The Mean Absolute Error (MAE) averaged over the 108 corners from these 27 validation images was 1.63 ±0.51 pixels (0.20 ±0.06 cm). In order to generate corner positions for each of our videos we choose a frame from seven different time points and estimate corner locations at each of these frames. For each corner we then use the median X and Y values from these seven measurements as the final value in order to reduce the impact of inaccurate locations which can result from occlusion or other image noise.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

We have multiple repeated measurements that are collected for each subject (mouse) and each subject has a different number of strides giving rise to imbalanced data. To address this, we used a linear mixed model (LMM) to dissociate within-subject variation from genotype-based variation between subjects (Laird and Ware, 1982; Pinheiro and Bates, 2000). Specifically, in addition to the main effects such as animal size, genotype and age, a random effect that captures the intra-subject variation is included. Finally, we have multiple repeated measurements at two different ages giving rise to a nested hierarchical data structure. The models (M1, M2 M3) follow the standard LMM notation with (Genotype, BodyLength, Speed, TestAge) denoting the fixed effects and (MouseID/TestAge) (test age nested within the animal) denoting the random effect.

$$M1 : Phenotype \sim Genotype + TestAge + BodyLength + (1 \mid MouseID / TestAge)$$

$$M2: Phenotype \sim Genotype + TestAge + Speed + (1 \mid MouseID / TestAge)$$

$$M3: Phenotype \sim Genotype + TestAge + Speed + BodyLength + (1 \mid MouseID / TestAge)$$

In general, we use M1 to detect changes in stride speed and M3 for changes in gait parameters. We include data for M2 in supplement for comparison with previously published data (Figures S3 and S4).

Each gait and ASD mutant were analyzed separately. We analyze *Mecp2* males and females separately. Having fixed a mutant line to analyze, we used dummy variable encoding to encode Genotype, a categorical covariate with two levels - Control (0) and Mutant (1), with the corresponding littermate WTs (or another control strain) serving as the reference level. The numeric covariates, Body length (M1,M2,M3), speed (M2, M3), were normalized using the z-score transformation. We did not include Sex as a covariate in the model; we found it correlated with body length. As described earlier, we treated all gait metrics as response variables except in M2 and M3, where we treated stride speed as a covariate. For the linear gait metrics, we considered the following LMM model for repeated measurements:

$$y_{ij} = x_{ij}^{T}\beta + \gamma_i + \varepsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i$$

$$\gamma_i \sim N\left(0, \sigma_\gamma^2\right), \ \varepsilon_{ij} \sim N\left(0, \sigma^2\right)$$

where n is the total number of animals; y_{ij} is the j^{th} repeat measurement on the i^{th} animal, n_i denotes the number of repeat measurements on animal i; x_{ij} is a $p \times 1$ vector of covariates such as body length, stride speed, genotype, age; β is a $p \times 1$ vector of unknown fixed population-level effects; γ_i is a random intercept which describes subject-specific deviation from the population mean effect; and ε_{ij} is the error term that describes the intrasubject variation of the i^{th} subject that is assumed to be independent of the random effect. We used Type II ANOVA F-test via Satterthwaite's degrees of freedom method to test the null hypothesis of no Genotype-based effect and obtain p-values. We fit our LMM models using the lme4 package in R (Bates et al., 2015).

We modeled the circular gait metrics (phase variables) in Table 1 as a function of linear variables using a circular-linear regression model. To adjust for linear variables such as body length and stride speed, we include them as covariates in the model. Analyzing circular data is not straightforward and statistical models developed for linear data do not apply to circular data (Jammalamadaka and Sengupta, 2001). The circular response variables are assumed to have been drawn from a von-Mises distribution with unknown mean direction μ and concentration parameter κ. The mean direction parameter is related to the variables $\mathbf{X}$ through the equation

$$y_i \sim \text{von Mises}(\mu_i, \kappa), \quad \mu_i = \mu + g\left(x_i^{T}\gamma\right), \quad i = 1, \ldots, n$$

where y_i is the mean circular metric for animal i, $g(u) = 2\tan^{-1}(u)$ is a link function such that for $-\infty < u < \infty$, $-\pi < g(u) < \pi$. The parameters $\mu, \gamma_1, \ldots, \gamma_k$ and κ are estimated via maximum likelihood (Fisher and Lee, 1992). We computed the p-value of no Genotype-based effect in circular phenotypes using a t-test which is based on asymptotic normality of the maximum likelihood estimators of the model parameters. The model is fitted using the circular package in R. (Lund et al., 2017)

We used the q-value (FDR adjusted p-value) to adjust for multiple testing across all gait metrics and control the false positive discovery rate. We reported LOD scores, defined as -log(q-value), in the heatmaps (Figures 3A, 3B, 4A, 4B, S3A, and S4A), and the q-values in Figure S11B.

The gait features for each animal were adjusted for body length (resp. body length + stride speed) by extracting residuals from a linear model with body length (resp. body length and stride speed) covariate/s and the gait metrics as response variables. The residuals were then averaged over animals for each strain to form a representative observation for the strain. The input to the k-means algorithm consisted of a matrix of 62 z-score transformed observations, each corresponding to a strain, on ten gait metrics. We initialized the k-means algorithm several times with random points from the data as means. We picked the initialization that gave the smallest total within-cluster sum of squares. We projected the selected k-means output onto the 2D PC (principal components) subspace to visualize the clustering structure. We performed PCA by applying singular value decomposition (SVD) of the input observations matrix. We obtained the contributions of different gait features to each PC component using the PC loadings, i.e., the corresponding eigenvectors obtained from PCA. The number of clusters was chosen based on the gap statistic (Tibshirani et al., 2001). To see the effect of non-linear embedding, we visualized the clustering structure in a non-linear embedded space using UMAP (McInnes et al., 2018) with two different initializations (SPCA - scaled PCA, Laplace - Laplacian Eigenmap).

GWAS

The R package *mousegwas*, previously described in Geuther et al. (2021), was used to compute genome wide associations. Briefly, the classical mouse strains were used (excluding wild-derived strains), ten individuals from each strain and sex combination were randomly selected and GEMMA was used with its LMM method. The MDA genotypes were obtained from (https://phenome.jax.org/genotypes). Body size and sex were used as covariates in the model. The GWAS can be reproduced using the command:

```
export G=https://raw.githubusercontent.com/TheJacksonLaboratory/mousegwas/
master nextflow run TheJacksonLaboratory/mousegwas -r gait \
-yaml $G/example/gait_nowild_withBL.yaml \
-shufyaml $G/example/gait_shuffle_withBL.yaml \
-input $G/example/gait_paper_strain_survey_2019_11_12.csv \
-outdir gait_output_withBL -clusters 1 \
-addpostp "-colorgroup -meanvariance -set3 -minherit 0.25 \
-loddrop 1.5" -addheatmap "-meanvariance -p 0.1"
nextflow run TheJacksonLaboratory/mousegwas -r gait \
-yaml $G/example/gait_nowild_noBL.yaml \
-shufyaml $G/example/gait_shuffle_noBL.yaml \
-input $G/example/gait_paper_strain_survey_2019_11_12.csv \
-outdir gait_output_noBL -clusters 1 \
-addpostp "-colorgroup -meanvariance -set3 -minherit 0.25 \
-loddrop 1.5" -addheatmap "-meanvariance -p 0.1"
```

Cell Reports, Volume *38*

Supplemental information

Stride-level analysis of mouse open field behavior using deep-learning-based pose estimation

Keith Sheppard, Justin Gardin, Gautam S. Sabnis, Asaf Peer, Megan Darrell, Sean Deats, Brian Geuther, Cathleen M. Lutz, and Vivek Kumar

10 Supplement

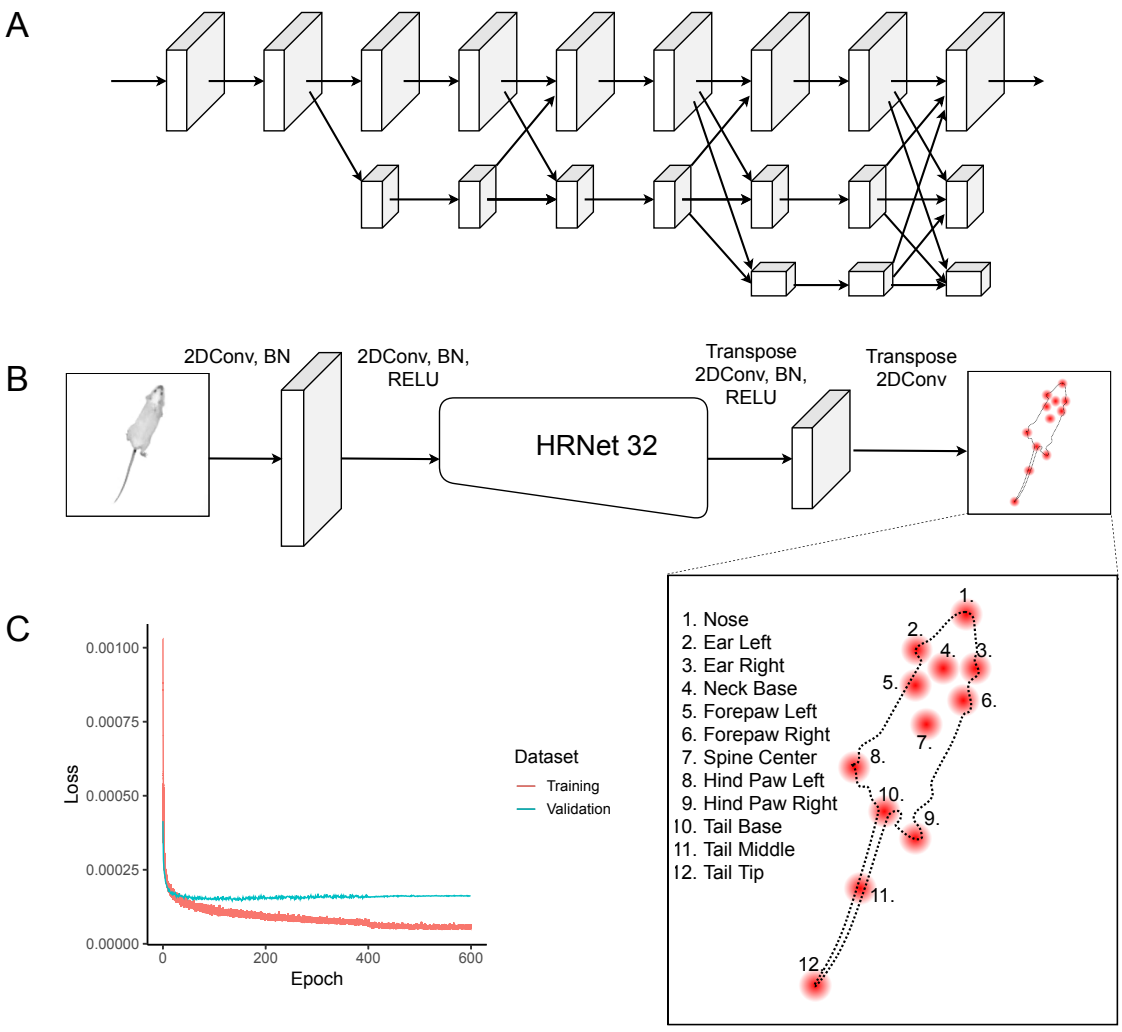
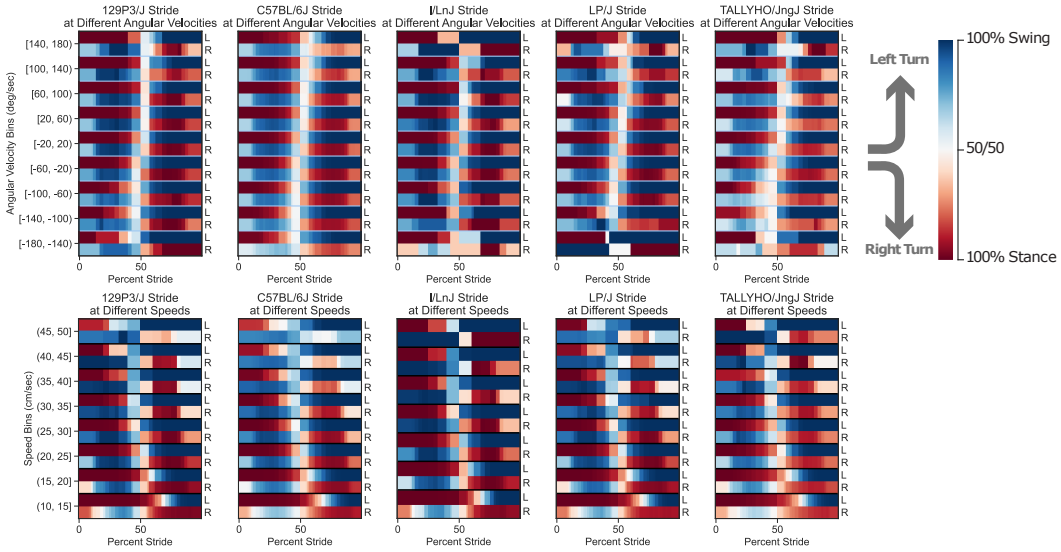


Figure S1: Deep convolutional neural network for pose estimation. Related to Figure 1. (A) the HRNet-W32 neural network architecture for performing pose estimation. (B) The inference pipeline which sends video frames into the HRNet and generates twelve keypoint heatmaps as output. We perform a 2D argmax on the twelve keypoint heatmaps to find the 2D pose coordinates. Also see Video S1. (C) Training loss curves show network convergence without overfitting.

A



B

Keypoint	MAE ±SEM (Pixels)			MAE ±SEM (cm)		
	All 1000	Dark	White	All 1000	Dark	White
All Keypoints	3.27 ±0.05	3.34 ±0.06	3.01 ±0.05	0.41 ±0.01	0.41 ±0.01	0.37 ±0.01
Nose	2.53 ±0.10	2.50 ±0.16	2.23 ±0.11	0.31 ±0.01	0.31 ±0.02	0.28 ±0.01
Ear Left	2.41 ±0.07	2.39 ±0.12	2.12 ±0.10	0.30 ±0.01	0.30 ±0.02	0.26 ±0.01
Ear Right	2.18 ±0.07	1.90 ±0.10	2.08 ±0.11	0.27 ±0.01	0.24 ±0.01	0.26 ±0.01
Neck Base	2.39 ±0.05	2.21 ±0.10	2.24 ±0.09	0.30 ±0.01	0.27 ±0.01	0.28 ±0.01
Forepaw Left	4.14 ±0.09	4.63 ±0.19	4.44 ±0.19	0.51 ±0.01	0.57 ±0.02	0.55 ±0.02
Forepaw Right	4.08 ±0.09	4.79 ±0.20	4.05 ±0.17	0.50 ±0.01	0.59 ±0.02	0.50 ±0.02
Spine Center	3.66 ±0.07	4.41 ±0.16	3.17 ±0.14	0.45 ±0.01	0.55 ±0.02	0.39 ±0.02
Hind Paw Left	5.08 ±0.15	6.38 ±0.36	3.68 ±0.25	0.63 ±0.02	0.79 ±0.04	0.46 ±0.03
Hind Paw Right	4.07 ±0.13	3.79 ±0.30	4.81 ±0.28	0.50 ±0.02	0.47 ±0.04	0.60 ±0.03
Tail Base	2.70 ±0.09	2.90 ±0.23	2.73 ±0.16	0.33 ±0.01	0.36 ±0.03	0.34 ±0.02
Tail Middle	3.24 ±0.12	2.70 ±0.14	2.62 ±0.14	0.40 ±0.01	0.33 ±0.02	0.32 ±0.02
Tail Tip	2.77 ±0.51	1.53 ±0.14	1.91 ±0.14	0.34 ±0.06	0.19 ±0.02	0.24 ±0.02

C

Keypoint	MAE ±SEM (Pixels)						
	All 120	129P3/J	NZO/HILtJ	NU/J	I/LnJ	LP/J	TALLYHO/JngJ
All Keypoints	3.44 ±0.08	2.80 ±0.13	3.80 ±0.24	3.55 ±0.18	4.23 ±0.25	3.33 ±0.22	2.91 ±0.13
Nose	1.97 ±0.14	2.02 ±0.23	1.78 ±0.52	2.47 ±0.43	2.73 ±0.36	1.19 ±0.11	1.65 ±0.17
Ear Left	3.89 ±0.22	2.77 ±0.42	3.52 ±0.47	4.66 ±0.31	5.49 ±0.88	3.77 ±0.35	3.14 ±0.30
Ear Right	4.06 ±0.24	3.01 ±0.31	3.50 ±0.53	4.69 ±0.37	5.86 ±1.10	3.28 ±0.29	4.02 ±0.35
Neck Base	2.18 ±0.15	1.66 ±0.16	1.98 ±0.23	2.40 ±0.28	3.50 ±0.72	1.84 ±0.22	1.67 ±0.19
Forepaw Left	4.06 ±0.27	3.10 ±0.40	4.48 ±0.75	3.93 ±0.68	5.72 ±0.80	4.45 ±0.78	2.66 ±0.28
Forepaw Right	3.90 ±0.31	3.17 ±0.39	4.35 ±0.92	3.73 ±0.81	5.18 ±1.10	3.99 ±0.76	2.97 ±0.26
Spine Center	3.84 ±0.20	3.57 ±0.50	3.63 ±0.47	3.63 ±0.43	4.93 ±0.67	3.26 ±0.44	4.04 ±0.37
Hind Paw Left	3.84 ±0.37	2.87 ±0.37	5.36 ±0.58	3.33 ±0.78	3.51 ±1.31	5.34 ±1.25	2.60 ±0.50
Hind Paw Right	4.11 ±0.40	2.65 ±0.37	5.94 ±1.09	4.07 ±0.70	4.54 ±1.07	5.11 ±1.50	2.35 ±0.51
Tail Base	2.96 ±0.32	2.20 ±0.38	4.04 ±1.39	2.60 ±0.74	3.24 ±0.73	3.03 ±0.59	2.68 ±0.47
Tail Middle	4.24 ±0.29	4.47 ±0.66	4.47 ±1.22	4.74 ±0.73	4.20 ±0.44	3.37 ±0.42	4.15 ±0.56
Tail Tip	2.20 ±0.22	2.13 ±0.64	2.49 ±0.72	2.37 ±0.50	1.91 ±0.26	1.32 ±0.22	2.98 ±0.67
Keypoint	MAE ±SEM (cm)						
	All 120	129P3/J	NZO/HILtJ	NU/J	I/LnJ	LP/J	TALLYHO/JngJ
All Keypoints	0.43 ±0.01	0.35 ±0.02	0.47 ±0.03	0.44 ±0.02	0.52 ±0.03	0.41 ±0.03	0.36 ±0.02
Nose	0.24 ±0.02	0.25 ±0.03	0.22 ±0.06	0.31 ±0.05	0.34 ±0.04	0.15 ±0.01	0.20 ±0.02
Ear Left	0.48 ±0.03	0.34 ±0.05	0.44 ±0.06	0.58 ±0.04	0.68 ±0.11	0.47 ±0.04	0.39 ±0.04
Ear Right	0.50 ±0.03	0.37 ±0.04	0.43 ±0.07	0.58 ±0.05	0.73 ±0.14	0.41 ±0.04	0.50 ±0.04
Neck Base	0.27 ±0.02	0.21 ±0.02	0.24 ±0.03	0.30 ±0.03	0.43 ±0.09	0.23 ±0.03	0.21 ±0.02
Forepaw Left	0.50 ±0.03	0.38 ±0.05	0.55 ±0.09	0.49 ±0.08	0.71 ±0.10	0.55 ±0.10	0.33 ±0.03
Forepaw Right	0.48 ±0.04	0.39 ±0.05	0.54 ±0.11	0.46 ±0.10	0.64 ±0.14	0.49 ±0.09	0.37 ±0.03
Spine Center	0.48 ±0.03	0.44 ±0.06	0.45 ±0.06	0.45 ±0.05	0.61 ±0.08	0.40 ±0.05	0.50 ±0.05
Hind Paw Left	0.48 ±0.05	0.36 ±0.05	0.66 ±0.07	0.41 ±0.10	0.44 ±0.16	0.66 ±0.16	0.32 ±0.06
Hind Paw Right	0.51 ±0.05	0.33 ±0.05	0.74 ±0.13	0.50 ±0.09	0.56 ±0.13	0.63 ±0.19	0.29 ±0.06
Tail Base	0.37 ±0.04	0.27 ±0.05	0.50 ±0.17	0.32 ±0.09	0.40 ±0.09	0.37 ±0.07	0.33 ±0.06
Tail Middle	0.52 ±0.04	0.55 ±0.08	0.55 ±0.15	0.59 ±0.09	0.52 ±0.05	0.42 ±0.05	0.51 ±0.07
Tail Tip	0.27 ±0.03	0.26 ±0.08	0.31 ±0.09	0.29 ±0.06	0.24 ±0.03	0.16 ±0.03	0.37 ±0.08

Figure S2: Validation of pose estimation and gait extraction across strains and key points. Related to Figure S1 and 1.(A) Hildebrand plots for hind paws over five diverse strains organized in columns. Each of these is aggregated over multiple strides from multiple individuals. The top row is separated into different angular velocity bins along the x axis showing how the stride pattern changes with angular velocity and the x axis for the bottom row is separated into stride speed bins showing how the stride pattern changes with speed. (B) We show errors in pixels and centimeters given for all keypoints grouped together and broken out for specific keypoint types. We use 1000 of our validation images to calculate Mean Absolute Error (MAE) with a Standard Error of Mean (SEM). We also calculate MEA and SEM for 200 dark colored mice and 200 white mice to demonstrate that our pose estimation network is robust to visual difference. (C) We show errors in pixels and centimeters given for all keypoints grouped together and broken out for specific keypoint types. We use 120 validation images (20 from each of the six strains shown) to calculate Mean Absolute Error (MAE) with a Standard Error of Mean (SEM). The mice chosen vary significantly in appearance in order to demonstrate the networks robustness to diversity. We include off-white (129P3/J), black obese (NZO/HILtJ), nude (NU/J), piebald (I/LnJ), agouti (LP/J) and moderately obese albino (TALLYHO/JngJ) mice. These are the same strains for which we render pose estimation in our supplementary video: VideoS1_diverse-mouse-pose.mp4.

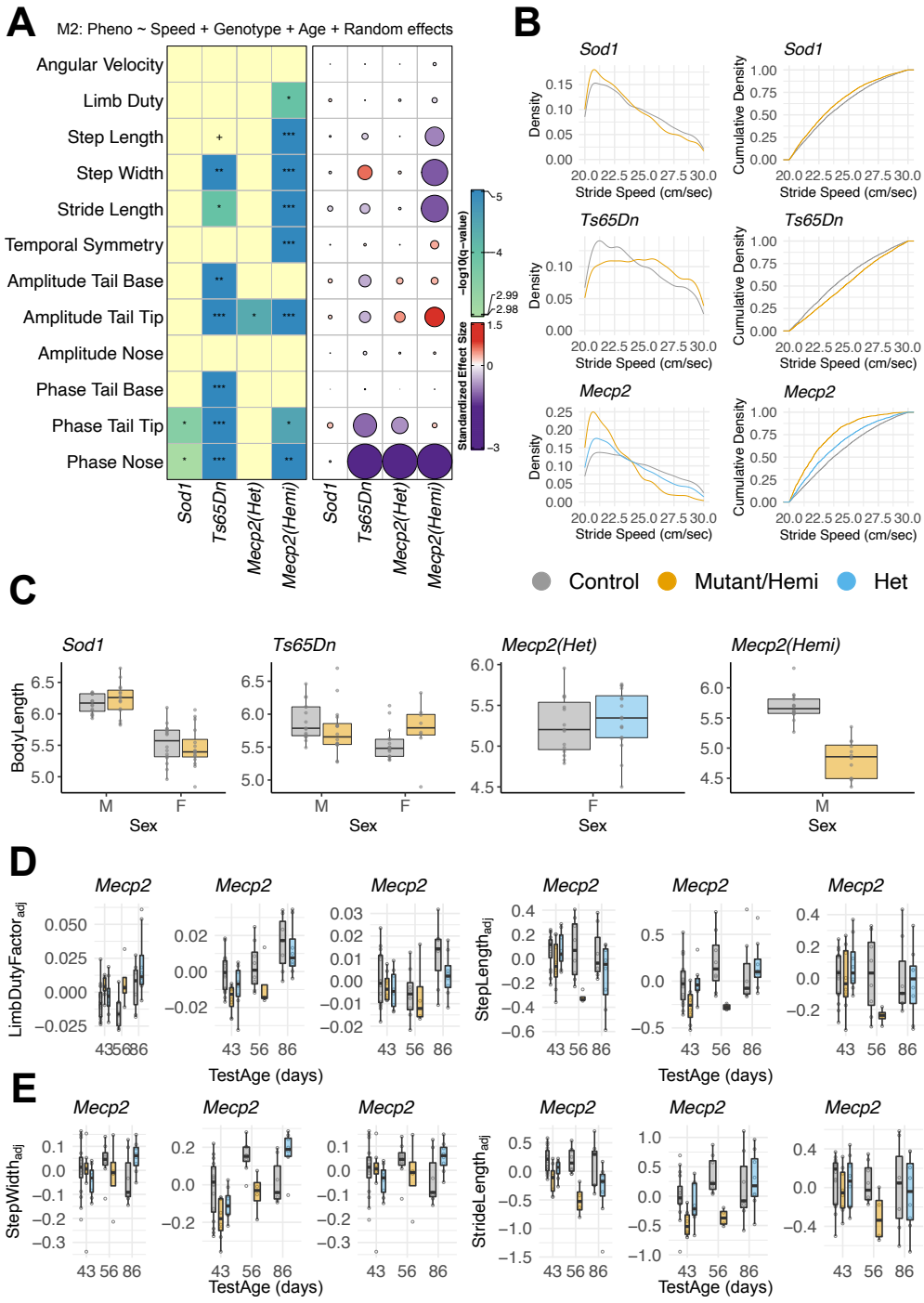


Figure S3: Supplemental analysis of gait mutants described in Figure 3. (A) Heat map summarizing the effect sizes and q-values obtained from model M3: Phenotype ∼ Genotype + TestAge + Speed + BodyLength + (1|MouseID/TestAge). (B) Kernel density (left) and cumulative density (right) curves of stride speed across all strains. (C) A plot showing positive association between body length and sex across different gait mutant strains. (D) Body length (M1), stride speed (M2), body length and stride speed (M3) adjusted residuals for limb duty factor and step length for Mecp2 gait mutant. (E) Body length (M1), stride speed (M2), body length and stride speed (M3) adjusted residuals for step width and stride length for Mecp2 gait mutant. Related to Figure 3.

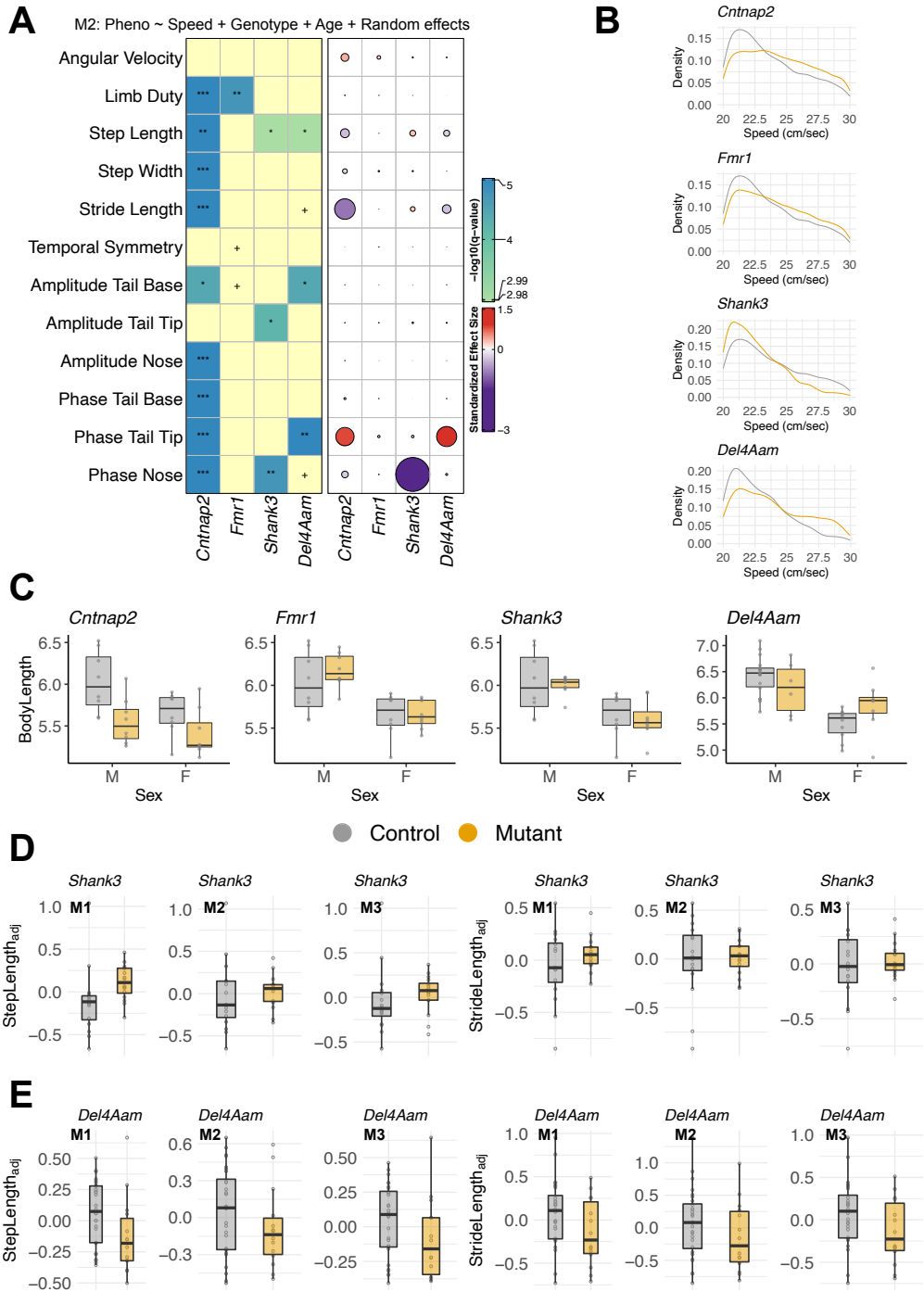


Figure S4: Supplemental analysis of ASD mutants described in Figure 4 (A) Heat map summarizing the effect sizes and q-values obtained from model M2: Phenotype ∼ Genotype + TestAge + Speed + (1|MouseID/TestAge). (B) Kernel density curves (estimates) of stride speed across all strains. (C) A plot showing positive association between body length and sex across different gait mutant strains. (D) Body length (M1), stride speed (M2), body length and stride speed (M3) adjusted residuals for step length and stride length for *Shank3* mutants. (E) Body length (M1), stride speed (M2), body length and stride speed (M3) adjusted residuals for step length and stride length for *Del4Aam*. Related to Figure 4.

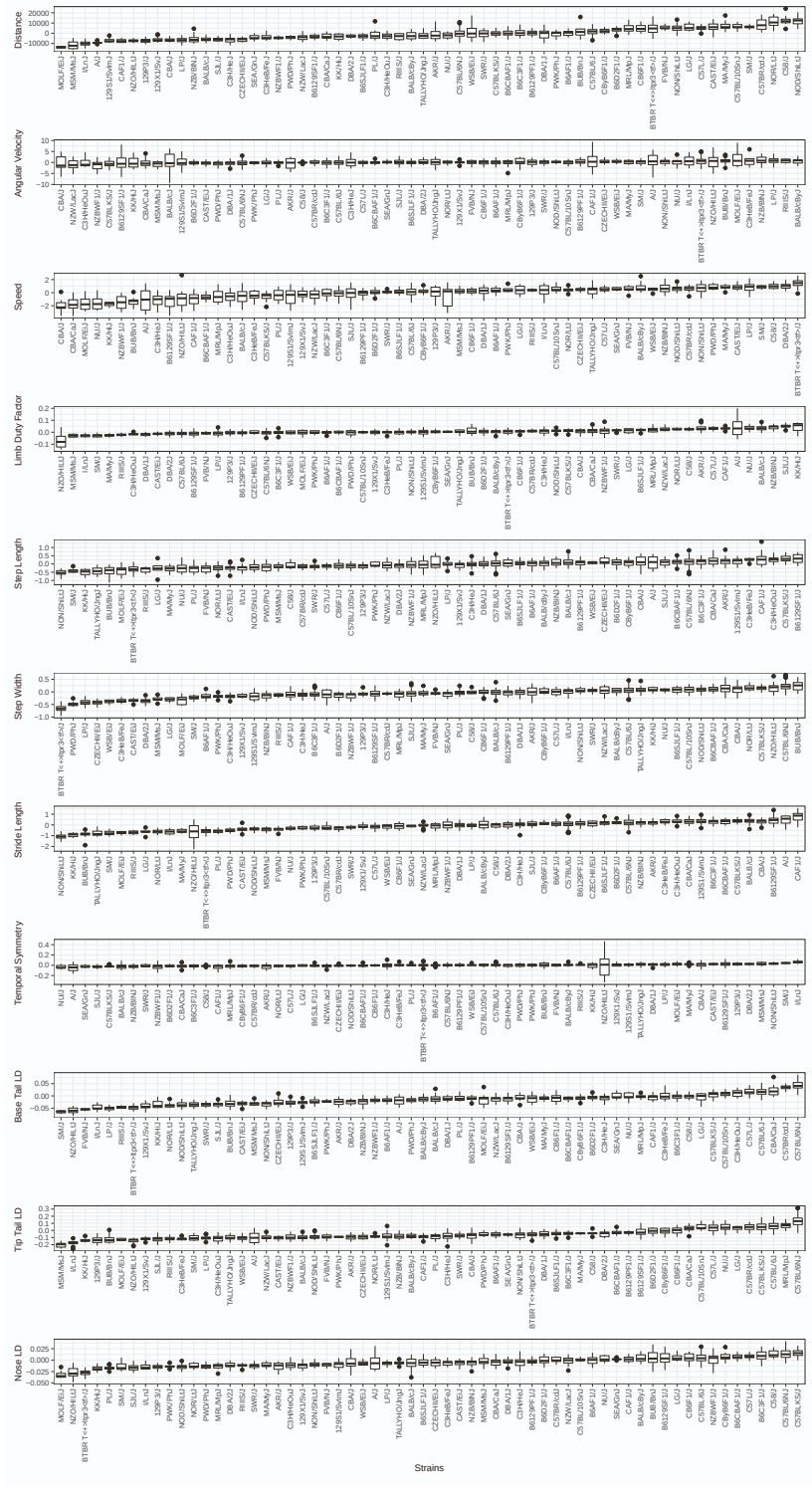


Figure S5: Body length adjusted phenotypes (residuals) are compared across 62 strains in the strain survey. The box plots are displayed in an ascending order with respect to the median measure from left to right. Each row corresponds to a different gait phenotype. Related to Figure 5.

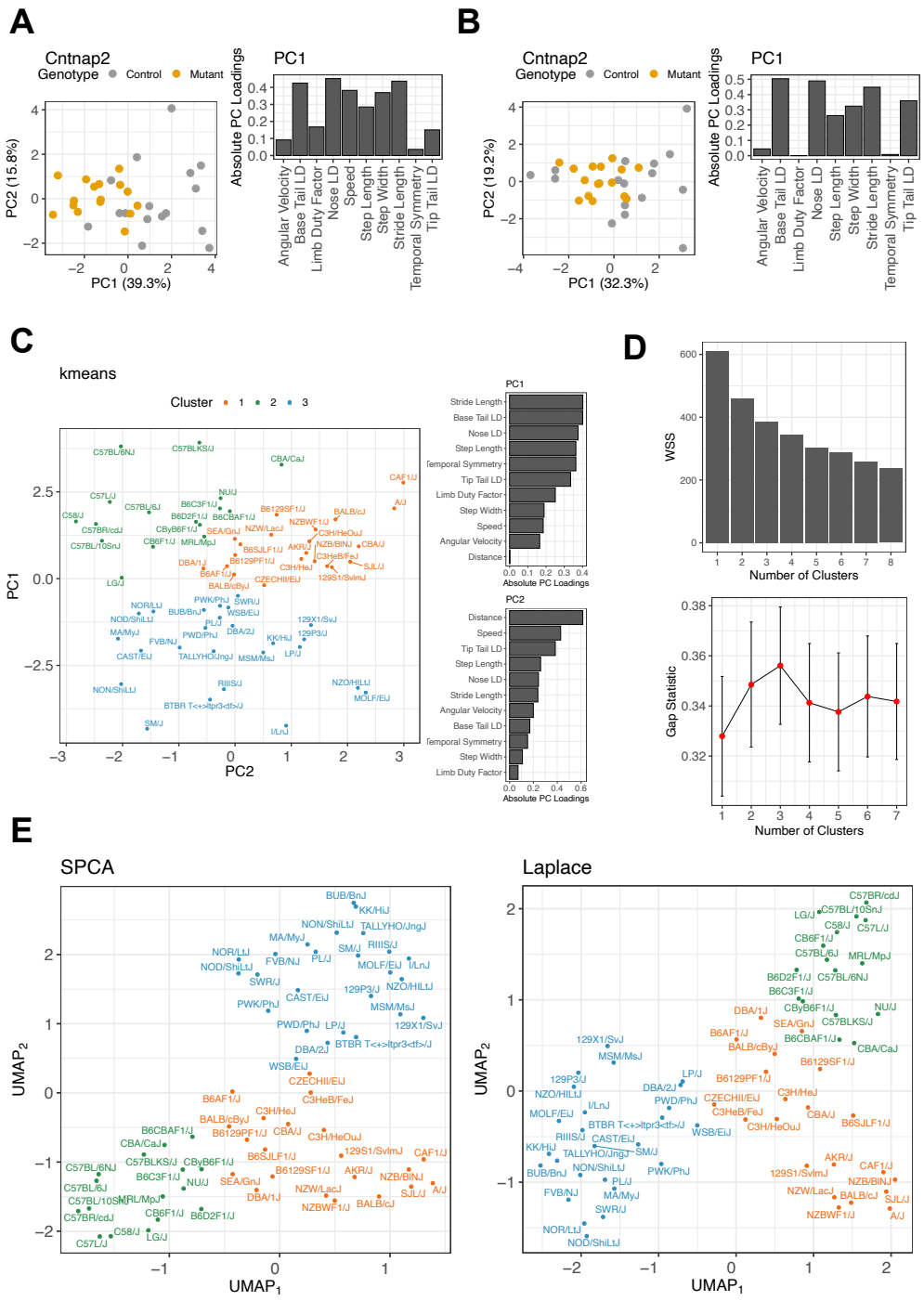


Figure S6: Extended unsupervised clustering analysis of Cntnap2 and Strain survey data. We performed PCA on z-score transformed (A) body-length adjusted gait data, (B) body-length + speed adjusted gait data. Related to Figures 4 and 5. In (A), although there was some overlap between the controls and mutants along PC1, the gait phenotypes combined to separate many mutants from controls along PC1. In (B), the speed-adjusted gait phenotypes were unable to separate mutants from controls since speed was an important contributor to PC1 in (A). Removing any other important contributors to PC1 (stride length, step length, step width, nose, and base tail) by regressing them out from other phenotypes had a similar adverse effect (results not shown) on the ability to separate controls from mutants along PC1. (C) A scatterplot of the strain survey clusters with body length adjusted gait features as input is shown without the shaded areas to present a more unbiased representation of clusters in 2D PCA subspace. (D) For the k-means clusters (Figure 5C), the choice of three clusters was optimal as the gap statistic (top) (Tibshirani et al., 2001) shows a clear peak at three clusters and the within-sum-of-squares (bottom) shows a drop at 3 clusters. (E) To see the effect of non-linear embedding, we visualized the clustering structure in a non-linear embedded space using UMAP with two different initializations (Scaled PCA, Laplace). We projected the strains to the two-dimensional UMAP space. Using the cluster memberships obtained from the k-means algorithm, we found that the UMAP dimensions preserved the separation between the three clusters discovered using the k-means.

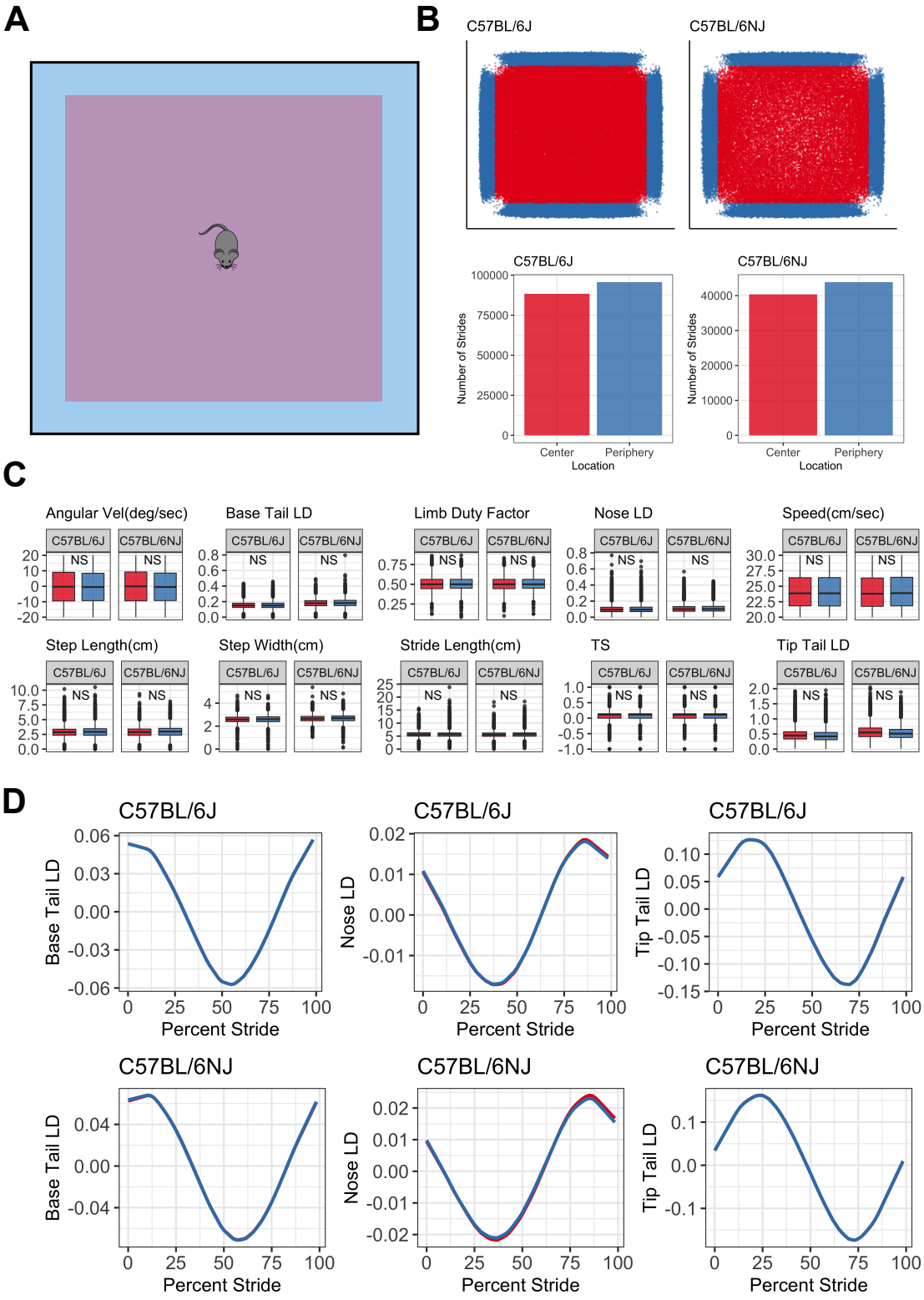


Figure S7: A comparative analysis of gait metrics between center and periphery of the open field arena in strains C57BL/6J and C57BL/6NJ. Related to Figure 2. (A) A schematic diagram of the open field arena marks the center and periphery (outer 10%) as the central 64% area (purple) and outermost 36% area (blue). (B) Top panel shows open field with each dot representing a stride of an animal in the center (red) and periphery (blue). Bottom panel shows the number of strides in the center versus periphery for the two strains. (C) Statistical comparison of linear gait metrics between center and periphery. We fit the linear mixed model Phenotype $\sim$ BodyLength + TestAge + Sex + Location + (1|MouseID) + (1|Location) where BodyLength, TestAge, Sex, Location are fixed effects and MouseID, Location are crossed random effects. To test the the null hypothesis of no effect of Location (no difference between periphery and center) on linear gait metrics, we use the F test with Satterthwaite's approximation method. Our analysis revealed no significant differences across all linear gait metrics. (D) Comparison of circular gait metrics between center and periphery for the two strains showed that the phase measures are almost superimposed for base tail, nose and tip tail.

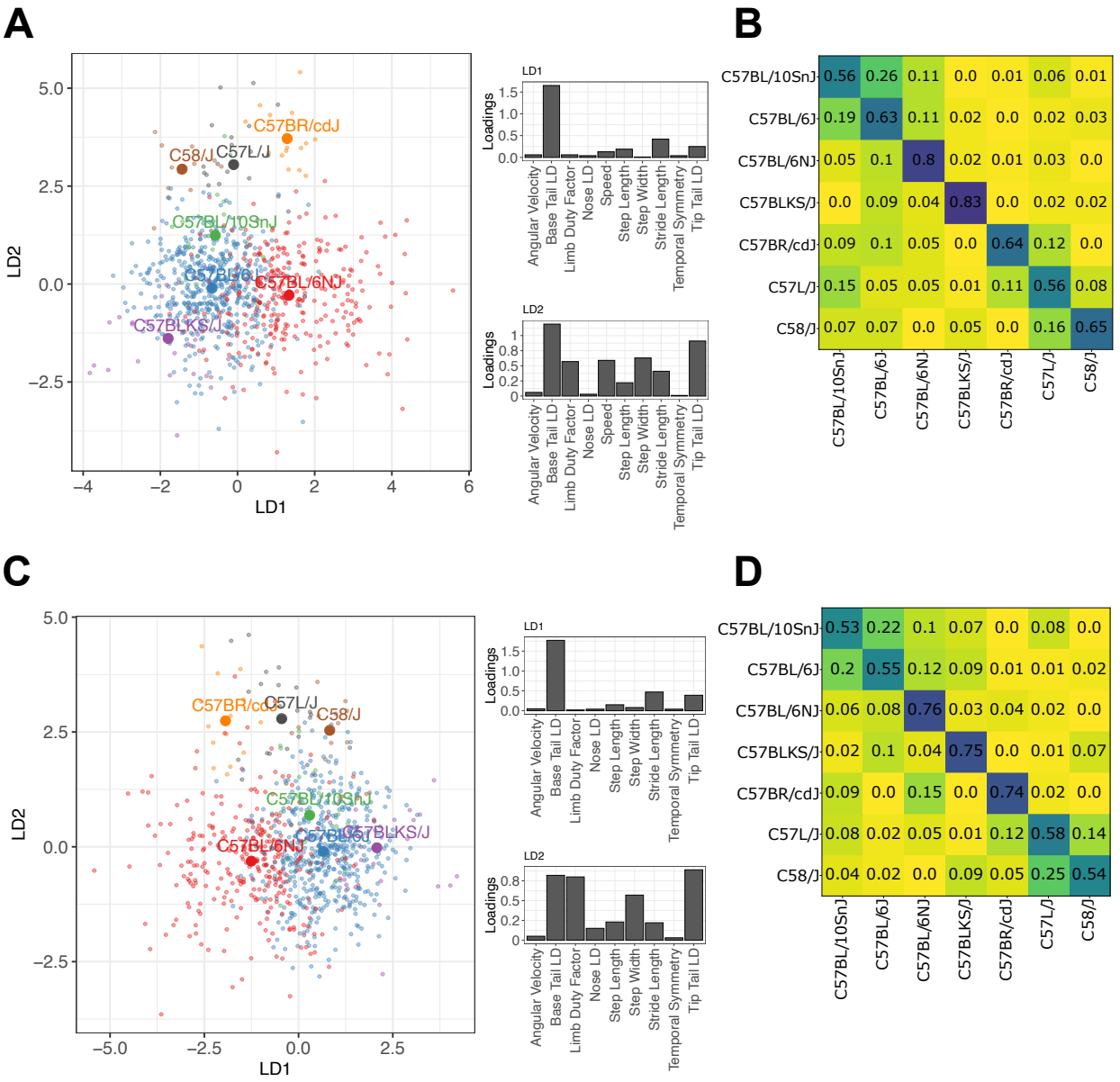


Figure S8: Extended analysis of strains belonging to Cluster 2 in strain survey analysis. Related to Figure 5. We sought to determine the gait metrics' sensitivity to distinguish among similar strains. We took two approaches: a supervised dimension reduction approach and a multi-class classification approach. For the former, we used linear discriminant analysis (Rao, 1948) to quantitatively distinguish between strains C57BL/6J, C57BLKS/J, C57L/J, C57BL/6NJ, C57BR/cdJ, C57BL/10SnJ, and C58/J. First, we adjusted the gait metrics for animals' body length by fitting a linear model with body length as a covariate. Next, we applied principal component analysis (PCA) to the residuals obtained from the linear model to address multicollinearity and prevent overfitting with LDA. We used all principal components (PCs) in the subsequent LDA algorithm to avoid the risk of throwing away critical discriminative dimensions. (A) We found that LDA separated strains when we embedded the PCs in a lower-dimensional 2D space for visualization purposes. Individual dots represent animals and dots (labeled with strain names) represent the mean/average coordinates of all animals belonging to the strain. Next, we multiplied the eigenvectors obtained from PCA with the LDA loadings matrix to identify the gait metrics that contributed to the separation between strains. For example, we found the base tail lateral displacement (Base Tail LD) to be a significant contributor to separating animals between strains C57BL/6N and C57BL/6NJ. We found similar posture differences between these two strains in our exploratory analysis earlier (see Figure 2H,I). We found the features Base Tail LD, Tip Tail LD, step width, stride speed, stride length, limb duty factor to contribute most strongly to LD2, which separated other C57 strains from C57BL/6N and C57BL/6NJ. For the second approach, we used a multi-class logistic regression ('one versus rest') model to predict the strain membership for each animal from its body length-adjusted gait metrics. First, we used stratified sampling to split the data into two parts: train (70%) and test (30%). Next, we used the popular resampling-based SMOTE algorithm (Chawla et al., 2002) to re-balance the number of animals for each strain in the training set. We trained the classifier on the re-balanced training set and tested the performance accuracy on the test set. We performed 100 different splits on the data to allow for a proper assessment of uncertainty in our test set results. (B) We summarized our results using a normalized classification accuracy matrix that shows the proportion of correctly classified (diagonal) and misclassified (off-diagonal) animals in each strain (row). For example, for C57BL/10SnJ (first row), 56% of the test set animals were correctly classified as C57BL/10SnJ. The classifier misclassified 26% (resp. 11%) of the C57BL/10SnJ in the test set as C57BL/6J (resp. C57BL/6NJ). (C,D) We performed similar analyses as in (A,B) except the gait metrics were adjusted for both body length and stride speed of the animals.

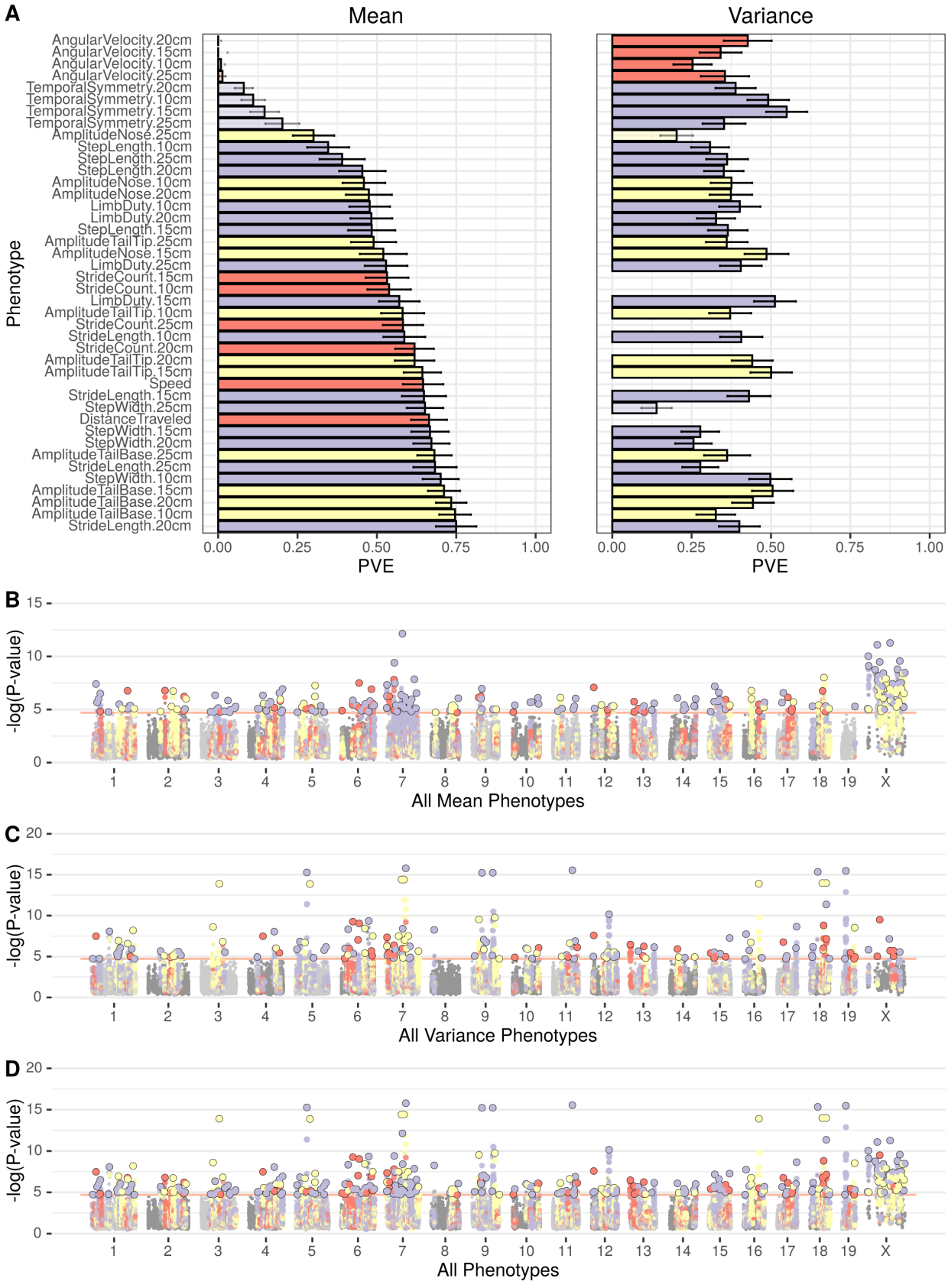


Figure S9: GWAS analysis of pose and gait measures without body length. PVE (A) and Manhattan plots (B-D). Related to Figure 6.

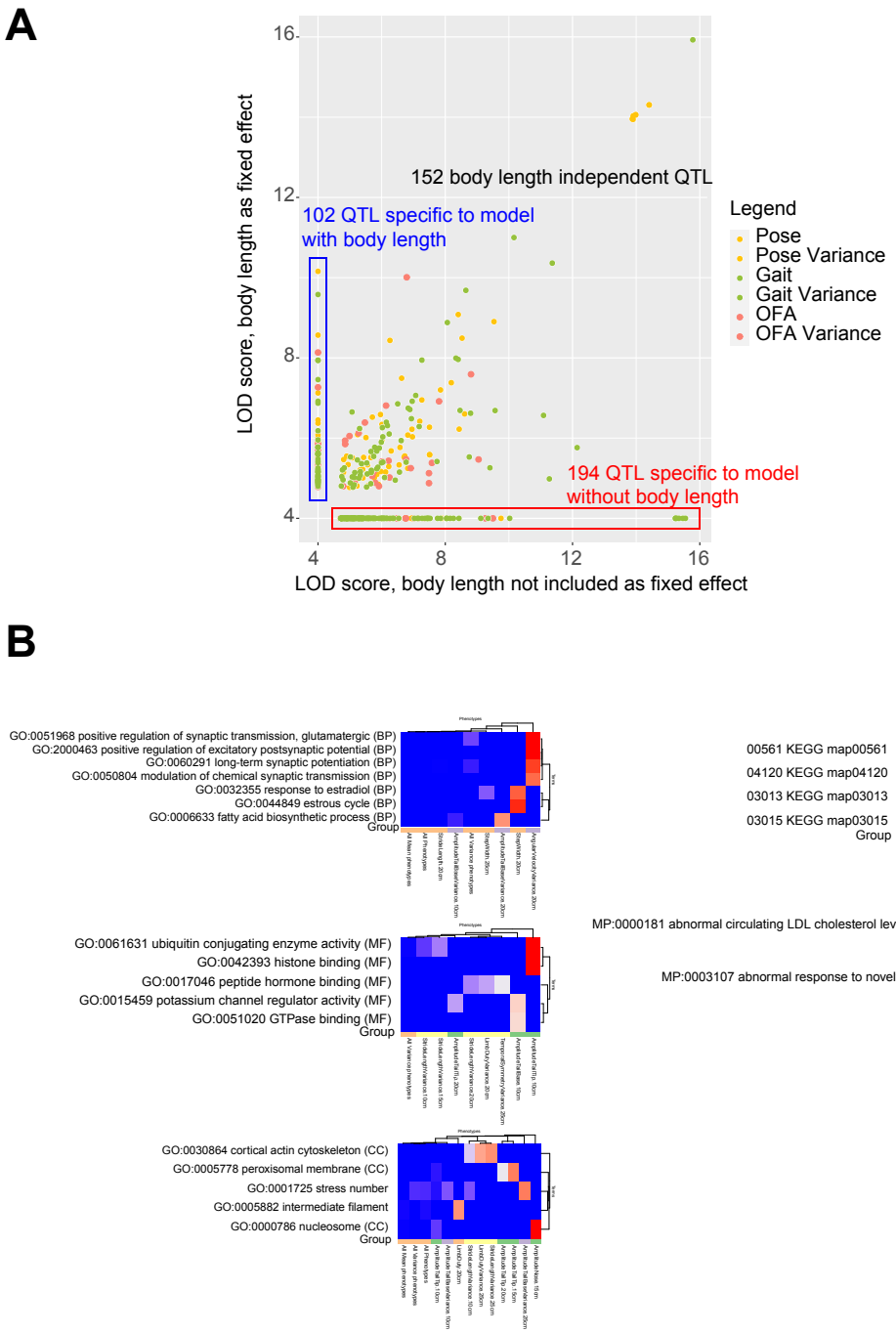


Figure S10: Comparison of QTL models with and without body length in the LMM and candidate INRICH analysis. Related to Figure 6. (A) The scores for the model with body length as a covariate are on the y axis while the scores for the same peaks are on the x axis. Body length-independent QTL were identified using both models and lie on the diagonal, QTL that are body length specific were detected using the model not including body length and were lost when including the body length. A few QTL that were identified only when including body length were supposedly masked by body length differences. (B) INRICH results for GO terms, KEGG pathways and mouse phenotypes. The heatmaps contain terms and phenotypes that passed the 0.1 corrected p-value threshold.

A

Gait Mutants						
Strain	Sex	BodyLength (cm)		Correlation	BodyWeight (g)	
		Control	Mutant	r	Control	Mutant
Sod1	M	6.16 ± .15	6.23 ± .26	0.48	28.90 ± 2.69	26.41 ± 2.25
	F	5.53 ± .31	5.47 ± .32	0.76	21.74 ± 2.38	20.08 ± 1.84
Ts65Dn	M	5.90 ± .29	5.75 ± .37	0.73	30.46 ± 4.16	29.39 ± 6.39
	F	5.55 ± .24	5.77 ± .42	0.73	23.28 ± 2.20	25.51 ± 6.59
Mecp2	M	5.69 ± .23	4.82 ± .31	0.94	23.53 ± 1.77	15.68 ± 1.81
	F	5.25 ± .35	5.33 ± .37	0.91	19.31 ± 3.09	19.86 ± 3.27
Autism Mutants						
Cntnap2	M	6.03 ± .37	5.56 ± .28	0.72	28.28 ± 1.95	23.95 ± 1.71
	F	5.65 ± .25	5.41 ± .29	0.72	22.69 ± 2.05	19.40 ± 0.97
Fmr1	M	6.03 ± .37	6.17 ± .20	0.77	28.28 ± 1.95	29.49 ± 1.37
	F	5.65 ± .25	5.66 ± .17	0.47	22.69 ± 2.05	21.01 ± 1.12
Shank3	M	6.03 ± .37	6.00 ± .11	0.75	28.28 ± 1.95	28.24 ± 1.00
	F	5.65 ± .25	5.60 ± .23	0.70	22.69 ± 2.05	21.60 ± 1.70
Del4Aam	M	6.42 ± .36	6.18 ± .51	0.70	33.13 ± 3.21	23.22 ± 2.43
	F	5.50 ± .29	5.84 ± .49	0.70	21.29 ± 1.49	18.16 ± 2.32

B

M1 : Phenotype Genotype + TestAge + BodyLength + (1 | MouseID/ TestAge)															
Phenotype	Sod1		Ts65Dn		Mecp2 (Het)		Mecp2 (Hemi)		Cntnap2		Fmr1		Shank3		Del4Aam	
	Effect size	p_{adj}	Effect size	p_{adj}	Effect size	p_{adj}	Effect size	p_{adj}	Effect size	p_{adj}	Effect size	p_{adj}	Effect size	p_{adj}	Effect size	p_{adj}
Angular Velocity	0.02	0.69	0.04	0.69	0.02	0.69	-0.25	0.37	0.51	0.79	0.13	0.83	0.13	0.83	-0.17	0.83
Speed	-0.21	0.00	0.15	0.04	-0.38	0.00	-0.53	0.00	1.07	0.00	0.64	0.00	-0.90	0.00	0.77	0.02
Limb Duty	0.23	0.00	-0.10	0.21	0.03	0.74	0.27	0.18	0.02	0.00	0.01	0.10	0.00	0.84	-0.01	0.41
Step Length	-0.12	0.34	-0.24	0.02	-0.10	0.37	-0.38	0.28	-0.19	0.13	0.03	0.77	0.23	0.08	-0.18	0.08
Step Width	-0.09	0.56	0.59	0.00	0.16	0.48	-0.79	0.02	-0.17	0.01	-0.08	0.15	-0.04	0.50	-0.01	0.90
Stride Length	-0.30	0.01	-0.35	0.01	-0.33	0.06	-0.60	0.01	-0.59	0.00	0.04	0.68	0.13	0.25	-0.23	0.08
Temporal Symmetry	-0.09	0.24	0.12	0.15	-0.00	0.96	0.09	0.64	0.00	0.66	-0.01	0.21	-0.01	0.21	-0.00	0.87
Amplitude Tail Base	0.23	0.11	-0.46	0.03	0.34	0.11	0.40	0.09	-0.01	0.01	-0.01	0.08	-0.00	0.45	-0.02	0.01
Amplitude Tail Tip	0.16	0.12	-0.37	0.00	0.50	0.01	0.99	0.00	0.02	0.29	-0.03	0.27	-0.06	0.02	-0.04	0.14
Amplitude Nose	0.02	0.73	-0.16	0.16	0.09	0.53	0.09	0.53	-0.02	0.00	-0.01	0.07	-0.00	0.49	-0.01	0.07
Phase Tail Base	0.02	0.11	-0.05	0.00	0.00	0.39	0.04	0.01	0.07	0.00	0.02	0.19	0.00	0.49	-0.02	0.24
Phase Tail Tip	0.09	0.16	-1.37	0.00	-0.17	0.02	0.50	0.00	0.71	0.00	-0.03	0.39	-0.12	0.15	0.52	0.00
Phase Nose	-0.06	0.01	-0.64	0.00	-0.05	0.06	-0.18	0.00	-0.09	0.00	-0.01	0.28	-0.12	0.00	-0.11	0.04
M2 : Phenotype Genotype + TestAge + Speed + (1 | MouseID/ TestAge)																
Angular Velocity	0.02	0.68	0.03	0.68	0.02	0.68	-0.14	0.28	0.35	0.96	0.16	0.96	0.05	0.96	-0.04	0.96
Limb Duty	0.14	0.12	-0.03	0.71	-0.10	0.36	-0.23	0.02	0.03	0.00	0.02	0.01	-0.01	0.12	0.00	0.85
Step Length	-0.08	0.66	-0.23	0.15	0.03	0.83	-0.77	0.00	-0.40	0.00	0.01	0.91	0.26	0.05	-0.27	0.05
Step Width	-0.12	0.59	0.59	0.00	0.14	0.59	-1.12	0.00	-0.21	0.00	-0.06	0.36	-0.05	0.36	-0.01	0.87
Stride Length	-0.24	0.16	-0.39	0.02	-0.11	0.59	-1.09	0.00	-0.91	0.00	-0.01	0.94	0.21	0.11	-0.38	0.08
Temporal Symmetry	-0.04	0.65	0.08	0.45	0.04	0.65	0.34	0.00	0.00	0.66	-0.02	0.09	-0.01	0.66	-0.01	0.66
Amplitude Tail Base	0.21	0.15	-0.45	0.03	0.28	0.15	0.30	0.15	-0.01	0.01	-0.01	0.10	-0.00	0.36	-0.02	0.01
Amplitude Tail Tip	0.13	0.17	-0.37	0.00	0.46	0.01	0.90	0.00	0.02	0.24	-0.03	0.24	-0.07	0.01	-0.04	0.13
Amplitude Nose	-0.00	0.95	-0.14	0.21	0.05	0.87	0.04	0.87	-0.02	0.00	-0.00	0.14	-0.00	0.23	-0.01	0.14
Phase Tail Base	0.01	0.22	-0.05	0.00	0.01	0.34	0.00	0.08	0.09	0.00	0.02	0.19	-0.00	0.48	-0.02	0.26
Phase Tail Tip	0.00	0.29	-4.55	0.01	-10.50	0.06	0.28	0.00	0.81	0.00	-0.10	0.14	-0.11	0.14	0.89	0.00
Phase Nose	-0.08	0.05	-8.14	0.00	-0.09	0.13	-4.12	0.00	-0.30	0.00	-0.03	0.16	-0.21	0.01	-0.09	0.07
M3 : Phenotype Genotype + TestAge + Speed + BodyLength + (1 | MouseID/ TestAge)																
Angular Velocity	0.02	0.67	0.03	0.67	0.02	0.67	-0.24	0.42	0.54	0.68	0.11	0.98	0.05	0.98	0.02	0.98
Limb Duty	0.14	0.13	-0.03	0.79	-0.12	0.26	0.04	0.79	0.03	0.00	0.02	0.01	-0.01	0.16	0.00	0.78
Step Length	-0.07	0.65	-0.27	0.01	-0.01	0.92	-0.28	0.55	-0.26	0.04	-0.01	0.91	0.28	0.03	-0.24	0.03
Step Width	-0.11	0.49	0.60	0.00	0.12	0.49	-0.81	0.02	-0.17	0.01	-0.08	0.17	-0.04	0.39	0.00	0.98
Stride Length	-0.22	0.07	-0.40	0.00	-0.18	0.32	-0.41	0.07	-0.73	0.00	-0.05	0.66	0.24	0.03	-0.32	0.02
Temporal Symmetry	-0.05	0.41	0.09	0.31	0.06	0.41	0.18	0.31	-0.00	0.59	-0.02	0.11	-0.01	0.59	-0.01	0.59
Amplitude Tail Base	0.21	0.19	-0.45	0.03	0.28	0.19	-0.35	0.23	-0.02	0.00	-0.01	0.11	-0.01	0.34	-0.02	0.00
Amplitude Tail Tip	0.13	0.17	-0.37	0.00	0.46	0.01	0.90	0.00	0.02	0.24	-0.03	0.24	-0.07	0.01	-0.04	0.13
Amplitude Nose	-0.00	0.95	-0.14	0.21	0.05	0.87	0.04	0.87	-0.02	0.00	-0.00	0.14	-0.00	0.23	-0.01	0.14
Phase Tail Base	0.01	0.34	-0.05	0.00	0.01	0.34	0.00	0.34	0.09	0.00	0.02	0.19	0.00	0.47	-0.02	0.26
Phase Tail Tip	0.00	0.03	-2.96	0.00	-0.71	0.22	0.21	0.01	0.31	0.00	-0.29	0.01	0.07	0.35	0.39	0.00
Phase Nose	-0.08	0.05	-5.60	0.00	-275.28	0.22	-8.24	0.00	-0.38	0.00	-0.03	0.16	-225.04	0.02	-0.13	0.08

Figure S11: Body Weight and model results. Related to Figures 3 and 4. (A) Summary data for body length and weight of animals in our experiments. (B) Summary of effect sizes and q-values (FDR-adjusted p-values, p_{adj}) obtained from models M1,M2,M3 for all phenotypes for both gait and autism strains.