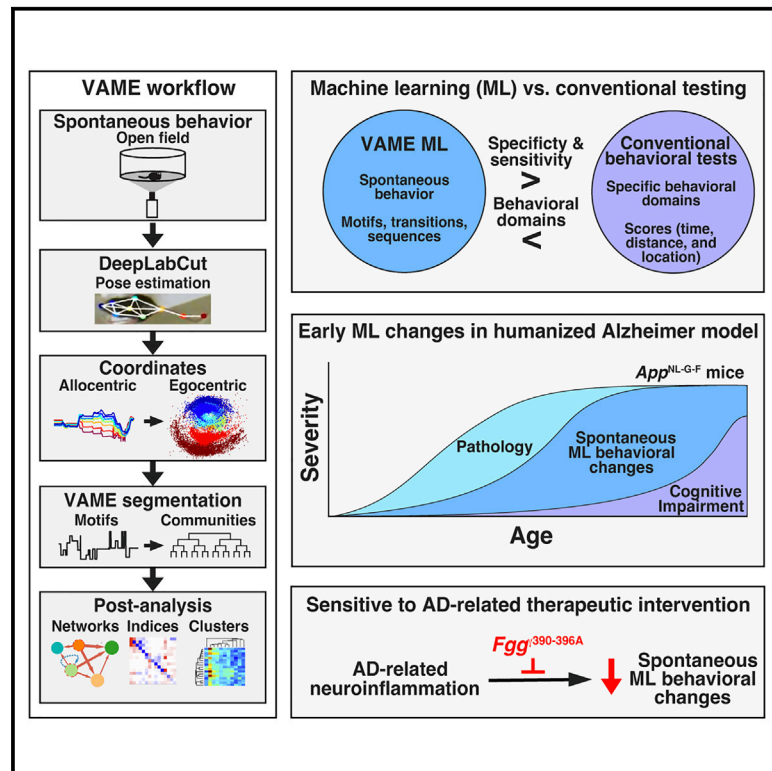


# Machine learning reveals prominent spontaneous behavioral changes and treatment efficacy in humanized and transgenic Alzheimer's disease models

## Graphical abstract



## Authors

Stephanie R. Miller, Kevin Luxem, Kelli Lauderdale, ..., Pavol Bauer, Stefan Remy, Jorge J. Palop

## Correspondence

stephanie.miller@gladstone.ucsf.edu (S.R.M.),

jorge.palop@gladstone.ucsf.edu (J.J.P.)

## In brief

Miller et al. use the VAME machine-learning platform to reveal robust spontaneous behavioral deficits in Alzheimer's disease mouse models. These deficits were ameliorated by a neuroinflammation-related therapeutic intervention and provided higher specificity and sensitivity than conventional approaches. VAME outcomes captured early disease manifestations and treatment efficacy in Alzheimer's models.

## Highlights

- VAME reveals robust alterations in spontaneous behavior in *App<sup>NL-G-F</sup>* and 5xFAD mice
- VAME has higher specificity and sensitivity than conventional behavioral approaches
- Fibrinogen-induced neuroinflammation drives behavioral changes in 5xFAD mice
- VAME outcomes are sensitive to AD-related therapeutic intervention



## Resource

# Machine learning reveals prominent spontaneous behavioral changes and treatment efficacy in humanized and transgenic Alzheimer's disease models

Stephanie R. Miller,<sup>1,2,\*</sup> Kevin Luxem,<sup>3,4,5</sup> Kelli Lauderdale,<sup>1,2</sup> Pranav Nambiar,<sup>1</sup> Patrick S. Honma,<sup>1,2</sup> Katie K. Ly,<sup>1,2</sup> Shreya Banger,<sup>1,2</sup> Mary Bullock,<sup>6</sup> Jia Shin,<sup>1,2</sup> Nick Kaliss,<sup>1</sup> Yuechen Qiu,<sup>1</sup> Catherine Cai,<sup>1</sup> Kevin Shen,<sup>1</sup> K. Dakota Mallen,<sup>1</sup> Zhaoqi Yan,<sup>1,7</sup> Andrew S. Mendiola,<sup>1,7</sup> Takashi Saito,<sup>8</sup> Takaomi C. Saido,<sup>9</sup> Alexander R. Pico,<sup>10</sup> Reuben Thomas,<sup>10</sup> Erik D. Roberson,<sup>6</sup> Katerina Akassoglou,<sup>1,2,7</sup> Pavol Bauer,<sup>3,4,5</sup> Stefan Remy,<sup>3,4,5,11</sup> and Jorge J. Palop<sup>1,2,12,\*</sup>

<sup>1</sup>Gladstone Institute of Neurological Disease, San Francisco, CA 94158, USA

<sup>2</sup>Department of Neurology, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94158, USA

<sup>3</sup>German Center for Neurodegenerative Diseases (DZNE), 39118 Bonn and Magdeburg, Germany

<sup>4</sup>Department of Cellular Neuroscience, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany

<sup>5</sup>Center for Behavioral Brain Sciences (CBBS), 39106 Magdeburg, Germany

<sup>6</sup>Center for Neurodegeneration and Experimental Therapeutics, Alzheimer's Disease Center, Department of Neurology, University of Alabama at Birmingham, Birmingham, AL 35233, USA

<sup>7</sup>Center for Neurovascular Brain Immunology at Gladstone and UCSF, San Francisco, CA 94158, USA

<sup>8</sup>Department of Neurocognitive Science, Nagoya City University Graduate School of Medical Sciences, Nagoya 467-8601, Japan

<sup>9</sup>Laboratory for Proteolytic Neuroscience, RIKEN Center for Brain Science, Wako-shi 351-0198, Japan

<sup>10</sup>Gladstone Institute of Data Science and Biotechnology, San Francisco, CA 94158, USA

<sup>11</sup>German Center for Mental Health (DZPG), 39118 Magdeburg, Germany

<sup>12</sup>Lead contact

\*Correspondence: [stephanie.miller@gladstone.ucsf.edu](mailto:stephanie.miller@gladstone.ucsf.edu) (S.R.M.), [jorge.palop@gladstone.ucsf.edu](mailto:jorge.palop@gladstone.ucsf.edu) (J.J.P.)

<https://doi.org/10.1016/j.celrep.2024.114870>

## SUMMARY

Computer-vision and machine-learning (ML) approaches are being developed to provide scalable, unbiased, and sensitive methods to assess mouse behavior. Here, we used the ML-based variational animal motion embedding (VAME) segmentation platform to assess spontaneous behavior in humanized *App* knockin and transgenic APP models of Alzheimer's disease (AD) and to test the role of AD-related neuroinflammation in these behavioral manifestations. We found marked alterations in spontaneous behavior in *App*<sup>NL-G-F</sup> and 5xFAD mice, including age-dependent changes in motif utilization, disorganized behavioral sequences, increased transitions, and randomness. Notably, blocking fibrinogen-microglia interactions in 5xFAD-*Fgg*<sup>γ390-396A</sup> mice largely prevented spontaneous behavioral alterations, indicating a key role for neuroinflammation. Thus, AD-related spontaneous behavioral alterations are prominent in knockin and transgenic models and sensitive to therapeutic interventions. VAME outcomes had higher specificity and sensitivity than conventional behavioral outcomes. We conclude that spontaneous behavior effectively captures age- and sex-dependent disease manifestations and treatment efficacy in AD models.

## INTRODUCTION

Behavioral alterations, the defining manifestations of neurological disorders, are intricate and multifaceted, posing challenges to accurate definition and measurement. Thus, our knowledge of disease-induced behavioral alterations is incomplete and largely limited to domain-specific and task-oriented behavioral tests. Analyzing complete sequences of spontaneous behavior can offer deep insight into disease-induced behavioral changes and may provide unbiased and scalable measures of brain dysfunction to evaluate disease progression and therapeutic interventions.<sup>1</sup> Until recently, this has not been feasible. However, advances in computer-vision and machine-learning (ML)

techniques for pose estimation (DeepLabCut,<sup>2,3</sup> SLEAP,<sup>4</sup> and LightningPose<sup>5</sup>) and behavioral segmentation (MoSeq<sup>6</sup> and variational animal motion embedding [VAME]<sup>7,8</sup>) make it possible to deconstruct full sequences of spontaneous mouse behavior into brief postural units (motifs or syllables) and reveal their sequence and hierarchical structure. These emerging methods expand the depth, breadth, and sensitivity of behavioral quantifications and may provide insights into disease pathogenesis.<sup>9</sup>

Alzheimer's disease (AD) begins several decades before the clinical onset of dementia with progressive amyloid accumulation preceding significant tau pathology and neurodegeneration.<sup>10,11</sup> During this extended preclinical period prior to the



dementia diagnosis, many patients develop behavioral or neuropsychiatric changes, including agitation, irritability, decreased motivation, loss of empathy, and depression, which increase the risk of transition to mild cognitive impairment and AD.<sup>12–14</sup> Thus, these behavioral changes may be the earliest symptoms of preclinical stages of AD.<sup>13,14</sup> This extended period of incipient and subtle behavioral alterations holds significant promise for investigations of early AD pathogenesis and therapeutic interventions, particularly given the recent development of ML approaches that may effectively harness both large-scale and precise behavioral data.

Humanized *App* knockin (KI) and APP transgenic mouse models mimic certain aspects of AD pathogenesis, including progressive amyloidosis, gliosis, and neuroinflammation, as well as network hyperexcitability and cognitive impairment, albeit to different degrees of severity across models and paradigms.<sup>15–27</sup> However, our understanding of alterations in spontaneous behavior in AD models has been limited by the technical challenges of assessing non-task-oriented behavior. Previously, we described the VAME ML architecture and identified motif usage alterations in APP/PS1 transgenic mice.<sup>7</sup> Here, we advanced the development of our VAME ML pipeline with analytical tools to assess organization of behavioral sequences of spontaneous behavior, disease progression, and sex-dependent effects in humanized *App*<sup>NL-G-F</sup> KI and 5xFAD transgenic mice. We also validated spontaneous behavioral outcomes for assessing therapeutic interventions by testing the role of AD-related neuroinflammation in 5xFAD mice with or without *Fgg*<sup>γ390–396A</sup> expression, which inhibits fibrinogen-microglia interactions and reduces AD-related neuroinflammation.<sup>21,22,28</sup> We conclude that spontaneous behavior effectively captures age- and sex-dependent early disease manifestations and treatment efficacy in AD models with higher sensitivity and specificity than conventional behavioral outcomes.

## RESULTS

### Aged *App*<sup>NL-G-F</sup> mice have mild spatial memory deficits and severe amyloidosis and gliosis

First, we used the Morris water maze test to determine whether aged *App*<sup>NL-G-F</sup> mice develop spatial learning and memory deficits. Consistent with previous findings,<sup>16–19,23</sup> 22-month-old *App*<sup>NL-G-F</sup> mice exhibited mild learning and memory deficits in the Morris water maze relative to their wild-type (WT) littermate controls (Figures 1A and 1B). In the hidden platform (spatial) component, *App*<sup>NL-G-F</sup> mice showed deficits in distance ( $p = 0.025$ ) and latency ( $p = 0.017$ ) (Figure 1A). In the probe trial (platform removed), *App*<sup>NL-G-F</sup> mice had no preference for the target quadrant (Figure 1B), indicating memory impairment. However, their time in the target quadrant was not significantly different from WT littermate controls ( $p = 0.108$ , Figure 1B). As expected,<sup>15</sup> 22-month-old *App*<sup>NL-G-F</sup> mice had robust AD-related pathology, including amyloid deposition and neuroinflammation (Figures 1C and 1D). Unlike APP transgenic mice,<sup>21,25</sup> *App* KI mice seem to develop milder cognitive changes.<sup>16–19,23</sup> As suggested,<sup>15,20</sup> *App* KI mice may model preclinical stages of AD, characterized by progressive amyloidosis

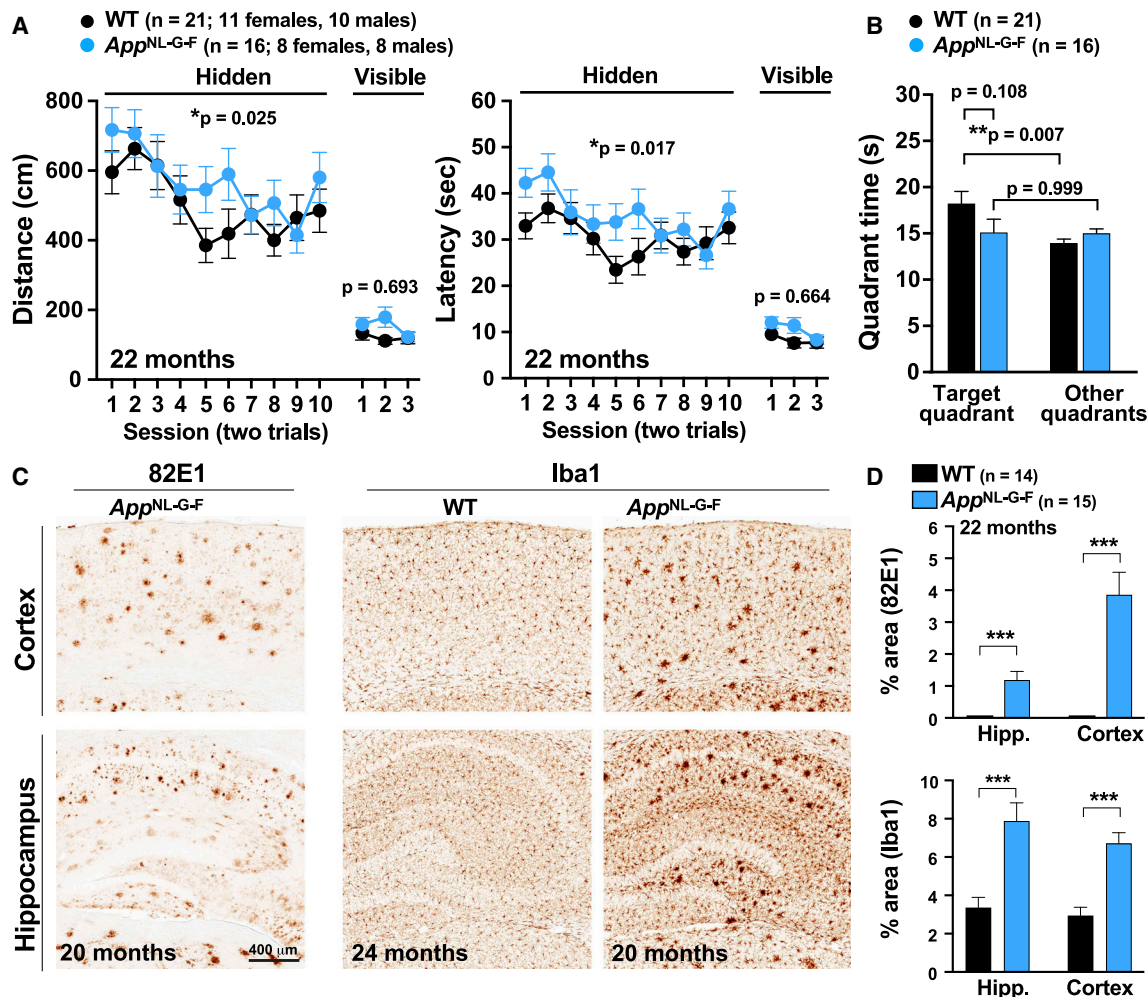
and gliosis without prominent cognitive deterioration, tau pathology, or neurodegeneration.

### VAME deconstructs sequences of mouse behavior into postural motifs

To study spontaneous behavioral manifestations in *App*<sup>NL-G-F</sup> mice, we built upon our ML pipeline and deconstructed complete sequences of spontaneous mouse behavior during exploration of an open arena. To improve upon the original VAME method,<sup>7</sup> we used refined algorithms for egocentric alignment of body parts, clustering of motifs into behavioral communities, and advanced kinematic and network analyses (see STAR Methods). Using these revised approaches, we studied spontaneous behavior for 25 min in an open arena to capture innate spontaneous behaviors that are not prompted by an external behavioral task or reward. Ventral-view videos (25 frames per second) of 13-month-old *App*<sup>NL-G-F</sup> mice ( $n = 18$ ) and littermate WT controls ( $n = 14$ ) (Figure 2A) allowed continuous visibility of limbs and other body parts that define mouse behavior. A DeepLabCut supervised neural network was used for pose estimation of nine body parts: nose, four paws, center, and the base, mid-point, and tip of the tail (Figure 2B). We then generated allocentric (relative to arena) and egocentric (relative to mouse center)  $x$  and  $y$  coordinates and speeds of the defined body parts (Figure 2C).

Egocentric  $x$  and  $y$  coordinates captured full motion of the nine body parts within the mouse left-right and rostral-caudal axes, respectively, relative to the mouse center (Figure 2D). Egocentric coordinates of seven body parts (nose, four paws, center, and tail base) were fed into VAME to train an unsupervised dynamical model, which assessed postural patterns across a sliding window of 16 frames (640-ms window) to identify 30 distinct postural motifs.<sup>7</sup> We compared  $k$ -means and hidden Markov model (HMM) algorithms for clustering the latent space embeddings into motifs and performed a logistic regression (classifier) analysis to assess the probability of either approach to correctly classify *App*<sup>NL-G-F</sup> and WT mice (Figure S1A). The  $k$ -mean algorithm had a higher overall probability of classifying genotype correctly than HMM (93.8% vs. 90.6%) and was used to generate a unique sequence of motifs for each mouse (Figure 2E). Thus, each motif represents a brief sequence of behavior that is consistently identified across mice and over time.

To assess motif specificity across sex and genotypes, we used uniform manifold approximation and projection (UMAP) and kinematic analyses. We found that each motif had a well-delimited UMAP location with smooth transitions between motifs (Figure 2F), indicating specificity of motif features and continuity of behavioral postural motion across motifs. Notably, motifs of each mouse fully overlapped with the corresponding motifs of the other mice regardless of sex and genotype (Figure S1B), indicating that motif identification was consistent throughout the cohort. Kinematics analyses of egocentric coordinates of the 30 motifs also revealed high specificity regardless of sex and genotype (Figures 2G, S1C, and S2), enabling us to precisely study mouse behavior. Thus, VAME can identify postural patterns in motion time series without supervised annotation, permitting reliable and unbiased behavioral segmentation.



**Figure 1. Aged *App*<sup>NL-G-F</sup> mice show robust AD-related pathology and mild impairments in the Morris water maze**

Spatial learning and memory of 22-month-old *App*<sup>NL-G-F</sup> mice (n = 16; eight females and eight males) and WT littermate controls (n = 21; 11 females and 10 males) were tested in the Morris water maze.

(A) Distance swum (left) and latency (right) in the hidden (spatial) and visible (cued) platform components of the Morris water maze test. *App*<sup>NL-G-F</sup> mice had mild deficits in distances ( $p = 0.025$ ) and latencies ( $p = 0.017$ ) in the spatial component of the Morris water maze.  $p$  values were determined by repeated two-way ANOVA.

(B) Time spent in the target (platform removed) and nontarget quadrants during the 24-h probe trial after 10 sessions of hidden training. *App*<sup>NL-G-F</sup> and WT mice did not perform differently at the target quadrant ( $p = 0.108$ ). However, WT ( $p = 0.007$ ), but not *App*<sup>NL-G-F</sup> ( $p = 0.999$ ), mice had a target quadrant preference.  $p$  values were determined by one-way ANOVA and Bonferroni *post hoc* test.

(C) Representative images of hippocampus and cortex stained for 82E1-positive A $\beta$  deposits and Iba1-positive microglia show severe amyloidosis and microgliosis in 22-month-old *App*<sup>NL-G-F</sup> mice.

(D) Quantification of hippocampal and cortical area (%) occupied by 82E1-positive A $\beta$  deposits or by Iba1-positive microglia in 22-month-old *App*<sup>NL-G-F</sup> (n = 14) and WT (n = 15) mice. \*\*\* $p < 0.01$  by Student's *t* test. Values are mean  $\pm$  SEM.

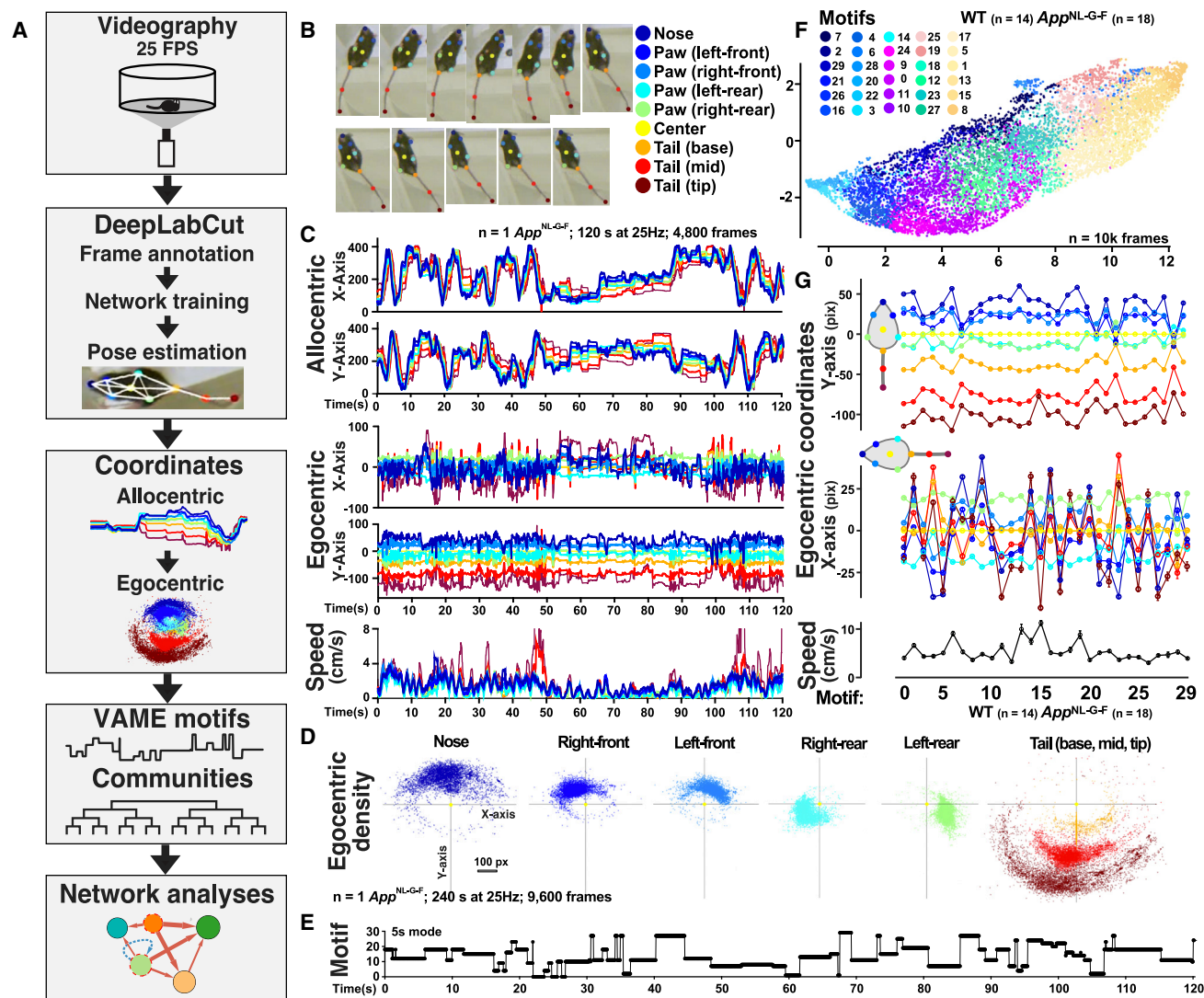
### Male and female *App*<sup>NL-G-F</sup> mice develop prominent age-dependent alterations in spontaneous behavior

Next, we determined whether VAME can detect spontaneous behavior deficits in young (6 months) and middle-aged (13 months) *App*<sup>NL-G-F</sup> mice compared to age-matched littermate controls. While 6-month-old *App*<sup>NL-G-F</sup> mice did not display deficits (Figure S1D), we found that motif usage was markedly altered in 13-month-old *App*<sup>NL-G-F</sup> mice, with significant ( $q < 0.05$ ) alterations in eight of 30 identified motifs, as shown by false discovery rate test with Benjamini-Hochberg correction for multiple compar-

isons (FDR-BH) (Figure 3A). Kinematic and video analyses of the eight motifs affected in *App*<sup>NL-G-F</sup> mice revealed that the motifs belong to the following generic behavioral categories: walk (motif 0), fast explore (motif 5), slow explore (motif 15), rear (motifs 2, 6), run (motifs 12 and 18), and sit up (motif 19) (Figures 3B, 3C, S2, and S3A). Thus, *App*<sup>NL-G-F</sup> mice exhibit robust age-dependent behavioral alterations in spontaneous behavior.

To assess gene and sex interactions, we analyzed the time spent in the eight *App*-affected motifs by two-way ANOVA (sex and genotype). Notably, VAME identified consistent and robust





**Figure 2. VAME segmentation of behavioral motifs during exploration in an open arena**

(A) Workflow. Mice were recorded from below at 25 frames per second (25 Hz) for 25 min (~37,500 frames/mouse) in a circular arena 12 inches in diameter. DeepLabCut (DLC) was used for pose estimation of nine body parts. VAME was used to identify behavioral motifs. Custom scripts were used to define egocentric coordinates, communities, and transition sequences.

(B) DLC pose estimation of nine defined body parts (ventral view).

(C) Allocentric (relative to arena) and egocentric (relative to mouse center) x and y coordinates (pixels) and allocentric speeds of the nine defined body parts (B) of a representative *App<sup>NL-G-F</sup>* mouse recorded over 2 min at 25 Hz.

(D) Egocentric coordinates relative to center (yellow) for the defined body parts (B) over 4 min in a representative *App<sup>NL-G-F</sup>* mouse (1 point per 40 ms; 9,600 frames). Note the left-right and rostral-caudal motion range of all body parts.

(E) Motif sequence (5-s mode) for (C). VAME model trained on egocentric coordinates for nose, paws, center, and tail base.

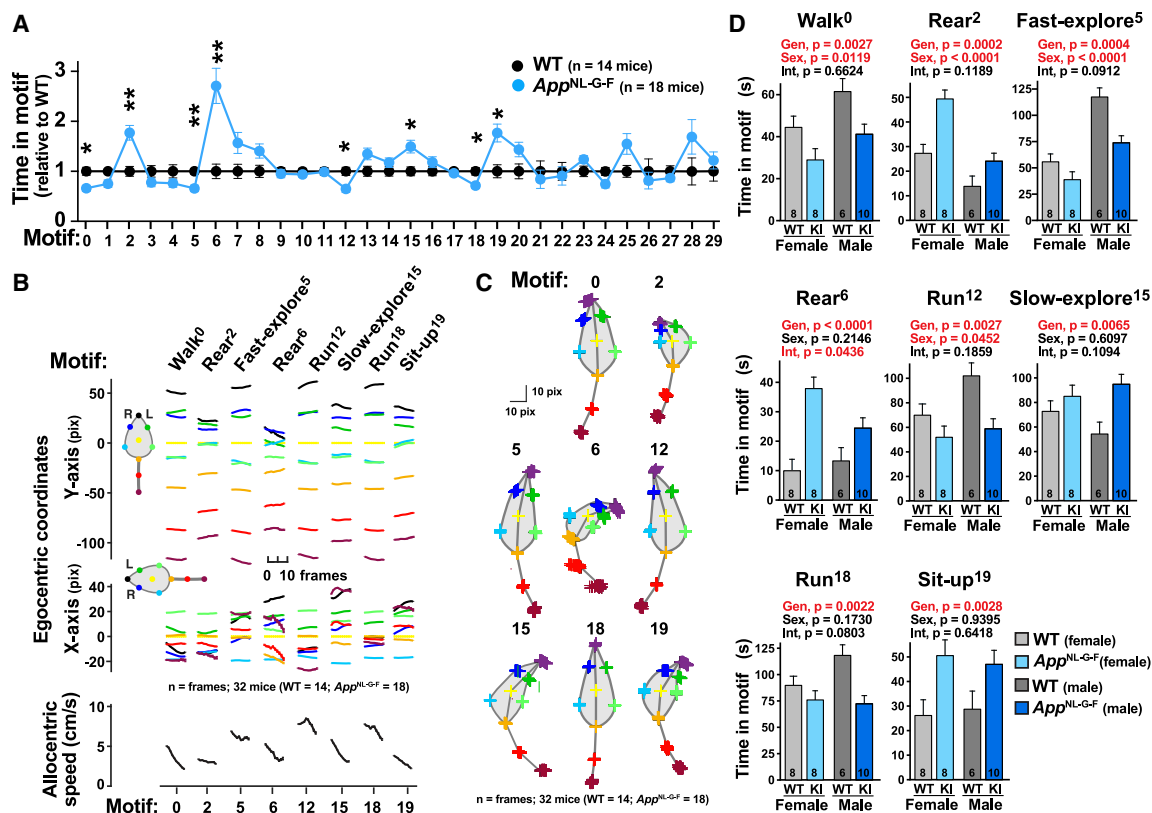
(F) UMAP representation of the 30 motifs identified by VAME for all mice ( $n = 32$ ; 18 *App<sup>NL-G-F</sup>* and 14 WT mice). Motifs occupy distinct locations with smooth transitions as expected for continuous behavioral sequences; 10,000 random frames shown from the entire cohort and duration.

(G) Egocentric coordinates in the rostral-caudal (y axis, top) and left-right (x axis, middle) axes of the 30 identified motifs ( $n = \text{frames}$ ; 32 mice). Points are the average location across all mice for the first 20 frames (800 ms) from motif identification. Allocentric center speed (bottom) across the full cohort ( $n = \text{frames}$ ; 32 mice). Motifs show specific and distinct coordinates and speeds. Values are mean  $\pm$  SEM.

differences in motif use between male and female controls (Figure 3D). However, the effects of *App* genotype and the direction of disease-associated changes in motif use were remarkably consistent across females and males, indicating that VAME can resolve sex and genotype interactions.

### Hierarchical clustering of motifs reveals behavioral communities

To leverage information about motif organization, we used hierarchical clustering of motif-motif transitions to identify higher-order behaviors. These analyses revealed a distinct hierarchical



**Figure 3. Middle-age *App*<sup>NL-G-F</sup> mice display robust behavioral alterations in motif use**

Spontaneous behavior of 13-month-old *App*<sup>NL-G-F</sup> mice (n = 18; eight females and 10 males) and WT littermate controls (n = 14; eight females and six males) was recorded for 25 min in a circular open arena.

(A) Thirty behavioral motifs were identified by VAME. Motif use (relative to sex-matched WT littermates) of 30 identified motifs for all mice. Relative to littermate controls, *App*<sup>NL-G-F</sup> mice displayed usage alterations in eight motifs. \*\*q < 0.01, \*q < 0.05 by false discovery rate with Benjamini-Hochberg correction for multiple comparisons (FDR-BH).

(B) Kinematic analyses of egocentric coordinates of the defined body parts for the first 10 frames (400 ms) of the eight motifs significantly affected in *App*<sup>NL-G-F</sup> mice (A, q < 0.05 by FDR-BH; n = frames; 32 mice).

(C) Schematic of body part positions for the indicated motifs. Data points are the average location for 10 frames after motif identification (n = frames; 32 mice). Motifs show precise and distinct coordinates.

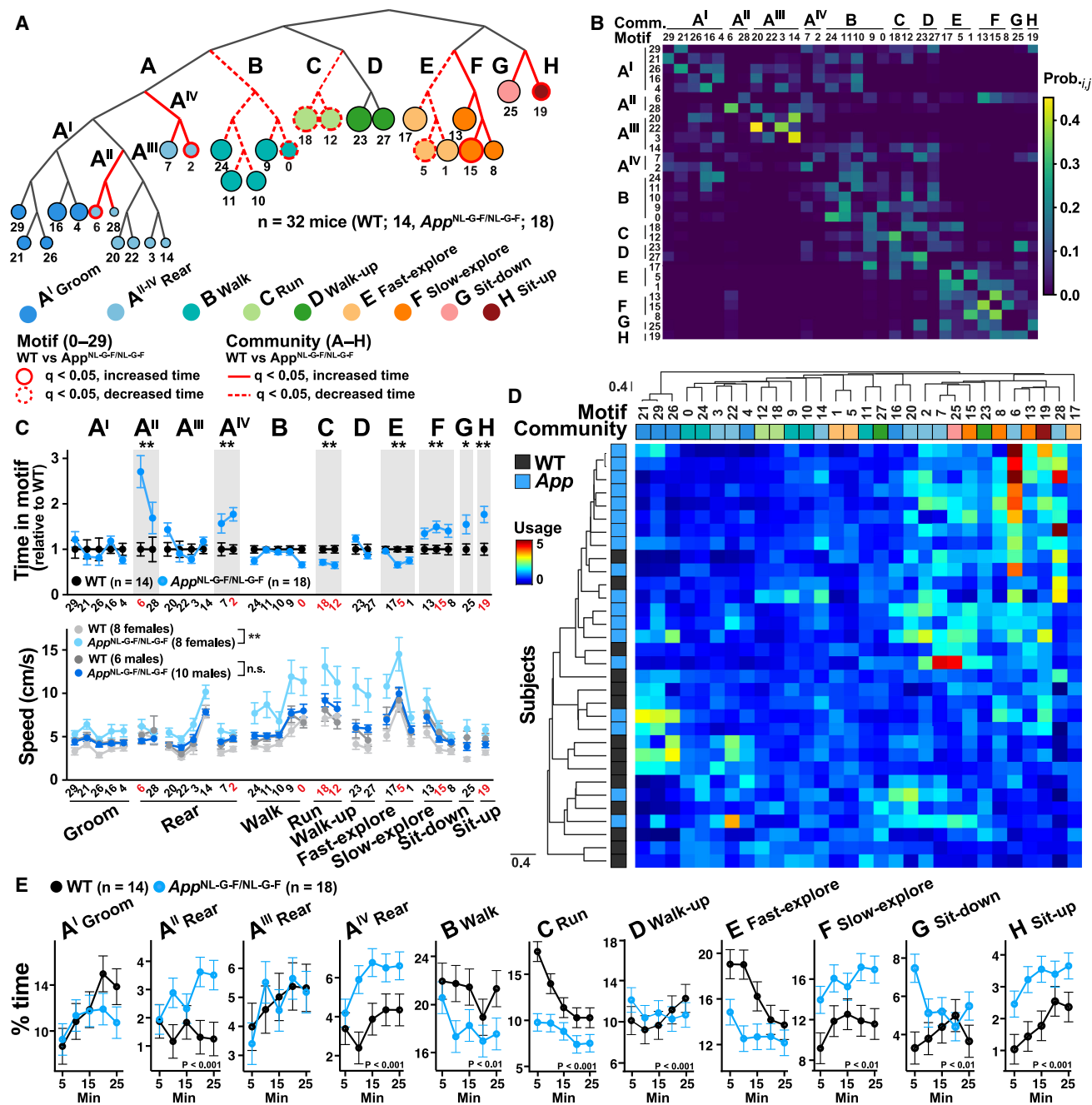
(D) Absolute motif use (in seconds) by genotype and sex for the eight motifs altered in *App*<sup>NL-G-F</sup> mice (A). Males and females differed in their use of motifs 0, 2, 5, and 12, but *App*<sup>NL-G-F</sup> expression affected both sexes similarly. p values were determined by two-way (genotype and sex) ANOVA. Values are mean ± SEM.

organization among the 30 motifs, which was used to cluster the motifs into 11 behavioral communities (Figures 4A and S2) according to total usage and the likelihood of sequential use (as opposed to pose similarity). The probability matrix of motif-motif transitions consistently showed that motifs within each community formed distinct clusters (Figure 4B), demonstrating highly structured behavioral dynamics. Notably, motifs of the same community were similarly affected by *App* (Figure 4C, top), validating the concept that communities represent cohesive behaviors. Communities were generically labeled as groom, rear, walk, run, walk-up, fast- and slow-explore, sit-down, and sit-up (see STAR Methods for definition and dendrogram cost function).

Although the VAME training data included only egocentric coordinates (relative to mouse center) without allocentric speed information, communities displayed distinct allocentric

speeds consistent with the represented behaviors. Female, but not male, *App*<sup>NL-G-F</sup> mice moved faster than sex-matched WT controls, particularly in the communities walk, run, and fast and slow-explore (Figure 4C, bottom). Thus, clustering postural motifs into behavioral communities provides an interpretable and ethologically relevant organization of mouse behavior in the context of disease-associated alterations.

To assess the distribution of genotypes and interindividual variability of mice, we also clustered motifs and mice hierarchically based on motif usage. We detected significant differences in clustering between *App*<sup>NL-G-F</sup> and WT mice (p = 0.0015, Mann-Whitney rank-sum test) (Figure 4D), indicating abnormalities in the overall motif use profile. *App*<sup>NL-G-F</sup> mice also showed a well-defined gradient of behavioral impairment, indicating that spontaneous behavioral changes capture interindividual variability.



**Figure 4. Hierarchical clustering of motifs identifies behavioral communities and reveals experience-dependent alterations in *App*<sup>NL-G-F</sup> mice**

(A) Cost-function-based hierarchical organization of motifs into 11 communities based on motif transitions and motif use at the cohort level (*n* = 32; 18 *App*<sup>NL-G-F</sup> and 14 WT mice). Motifs with higher probability of transitions and closer hierarchical distance were clustered into communities on the dendrogram. Node size is proportional to motif use. Red borders and branches indicate significant differences between *App*<sup>NL-G-F</sup> and WT mice by FDR-BH for 30 motifs and 11 communities, respectively.

(B) Matrix depicting the probability of forward transitions from motif<sub>i</sub> (rows) to motif<sub>j</sub> (columns) at the cohort level (*n* = 32; 18 *App*<sup>NL-G-F</sup> and 14 WT mice). Motifs were organized according to the dendrogram order found in (A).

(C) Motif use (relative to sex-matched WT littermates) (top) and speeds (bottom) organized by community. Top: motifs in the same community had reliably similar *App*<sup>NL-G-F</sup> effects, consistent with the notion that each community reflects a tightly associated set of postural units. \*\**q* < 0.01, \**q* < 0.05 by FDR-BH for indicated communities. Bottom: allocentric speed of mouse center in motifs organized by community for female and male *App*<sup>NL-G-F</sup> mice and WT controls. Relative to sex-matched controls, female, but not male, *App*<sup>NL-G-F</sup> mice performed ambulatory motifs with excessive allocentric speed. Red indicates disease-affected motifs.

(legend continued on next page)

### Experience-dependent modulation of behavioral motifs and communities is altered in *App*<sup>NL-G-F</sup> mice

Next, we investigated whether motif and community use is modulated by time and genotype as mice gain experience during their 25-min exploration of the open arena. We found that motifs and communities were strongly modulated by time and genotype. Interestingly, communities that show habituation (decreased use over time) in WT mice (B-walk, C-run, and E-fast-explore) were underrepresented in *App*<sup>NL-G-F</sup> mice, whereas communities that showed sensitization (increased use over time) in WT mice (A<sup>IV</sup>-rear, F-slow-explore, and H-sit-up) were overrepresented in *App*<sup>NL-G-F</sup> mice (Figure 4E). Similar results were found at the motif level (Figure S3B). Thus, *App*<sup>NL-G-F</sup> mice displayed abnormal habituation and sensitization responses to the environmental context in virtually all identified behaviors. Similar habituation and context-dependent deficits have been observed in APP transgenic and *App* KI mice, as judged by total distance traveled in the OF test.<sup>16,24,27</sup>

### Network analysis of motif transitions reveals increased randomness and altered behavioral sequences in *App*<sup>NL-G-F</sup> mice

Deconstructing the full sequence of motif-motif transitions may ultimately provide a more sensitive measure of brain function than unitary motif analyses (Figures 2, 3, and 4). To address the overall structure of full behavioral sequences, we used the discrete Markov chain model to assess the probability of motif transitions during 25 min of exploration in the open arena (Figure 5A). This model assesses stochastic infinite sequences of events (e.g., motifs) in which the probability of each event depends on previous events.<sup>29,30</sup> To assess the predictability of behavior, we fitted Markov chain models at a range of orders (0–3; 640-ms timesteps) (Figure 5B and Table S1).<sup>30</sup> In *App*<sup>NL-G-F</sup> mice, particularly females, behavioral sequences were less predictable than in WT mice at all orders, indicating increased entropy or randomness in behavioral organization.

To assess the probability of transitions between motifs, we used Cytoscape<sup>31</sup> to generate an unbiased network visualization and graphed the top 20% most probable transitions between motifs at the cohort level ( $n = 32$  mice). As expected from the transition probability matrix (Figure 4B), we found that motifs were well organized into communities (Figures 5C and 5D). Interestingly, ambulatory behaviors were the hub around which the rest of the behaviors were arranged. *App*<sup>NL-G-F</sup> mice had substantial increases in the probability of transitions, particularly between motifs of different communities, resulting in decreased dwell time in ambulatory hub motifs (Figure 5C). Notably, *App*<sup>NL-G-F</sup> mice prematurely transitioned from ambulatory communities (fast exploration or running) to slower exploratory behaviors (walk-up, sit-up, or slow exploration) and were more likely to engage in this behavioral sequence than WT controls (Figure 5D).

Finally, we calculated motif delta indices to score the aggregated motif differences between *App*<sup>NL-G-F</sup> and WT mice with regard to motif transition probability, motif use, and motif speed. We found that *App*<sup>NL-G-F</sup> mice had significant abnormalities in delta indices of motif transitions and motif use, and there was a trend toward impaired motif speed delta index (Figures 5E and 5F). Overall, full behavioral sequence analyses revealed significant alterations in the organization of behavioral structure in *App*<sup>NL-G-F</sup> mice, characterized by increased transitions and randomness, and premature escape of fast ambulatory behaviors to engage in slow and static exploration.

### Spontaneous behavioral alterations in 5xFAD mice are rescued by blocking fibrinogen-microglia interactions

The blood coagulation factor fibrinogen is deposited in the brain of patients with AD and is a key mediator of neurotoxic microglia polarization and cognitive impairment in mouse models of AD.<sup>21,22,32,33</sup> To define spontaneous behavior alterations in a second AD model and determine whether these alterations are sensitive to therapeutic intervention, we studied 8- to 10-month-old 5xFAD mice expressing WT fibrinogen or a mutant fibrinogen (*Fgg*<sup>γ390–396A</sup>) that lacks the inflammatory CD11b receptor-binding site and thus prevents CD11b receptor-mediated interactions with microglia.<sup>22,33</sup> *Fgg*<sup>γ390–396A</sup> reduces AD-related pathogenesis, including amyloid burden, dystrophic neurites, synaptic loss, microgliosis, and neuroinflammation in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice relative to 5xFAD mice.<sup>21,22,34</sup> We found that 5xFAD mice had prominent ML behavioral alterations in 17 of 30 motifs (Figure 6A). Remarkably, nearly all of these 5xFAD-dependent alterations were ameliorated in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice, indicating global restoration of spontaneous behavior, including motif use and speeds (Figures 6A and 6B).

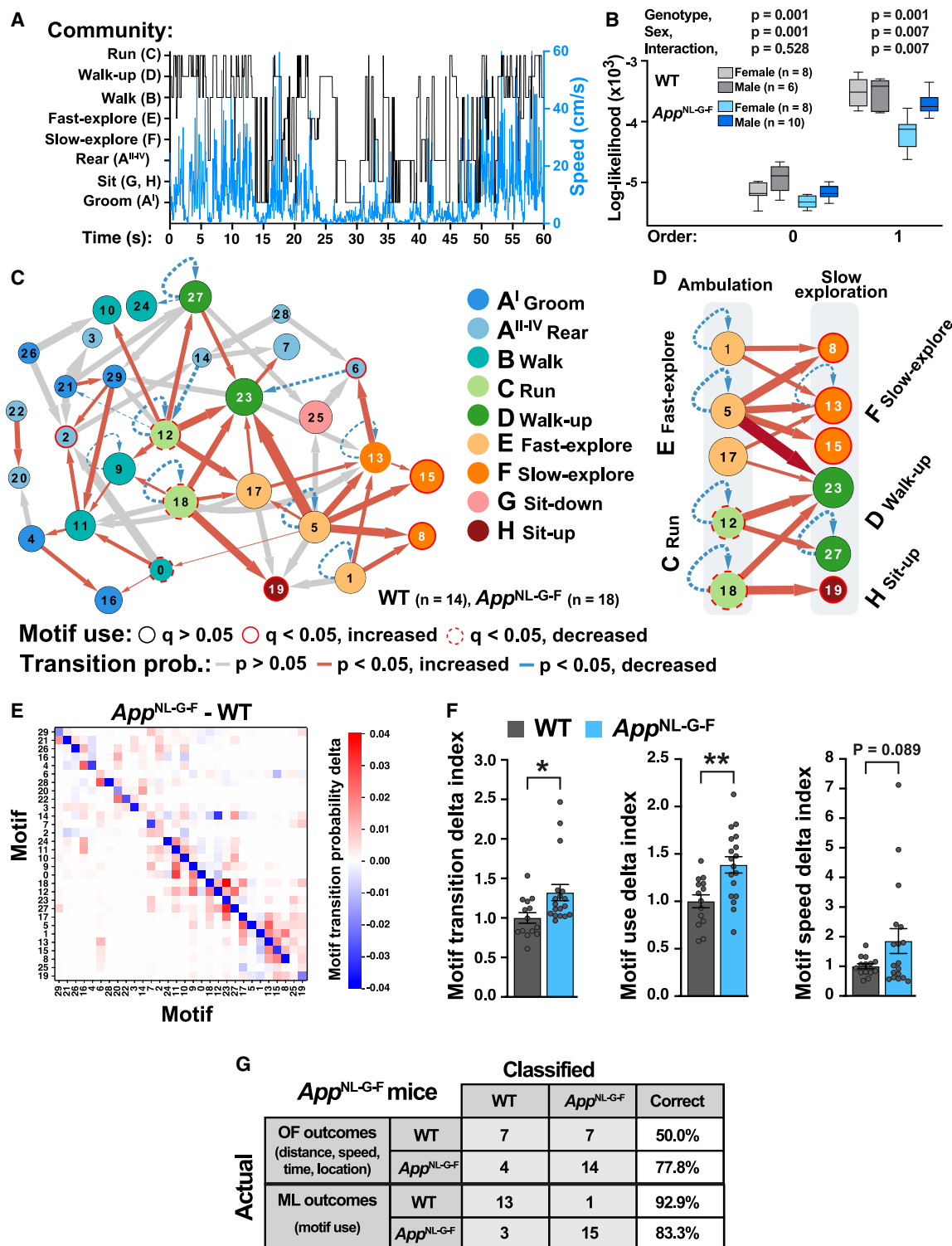
We also calculated delta indices for motif transitions, motif use, and motif speed to score the aggregated differences between 5xFAD, 5xFAD-*Fgg*<sup>γ390–396A</sup>, and control mice. We found that 5xFAD-dependent abnormalities in the motif transition and motif speed delta indices were ameliorated in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice (Figures 6C and 6D). Direct comparisons of the control groups, WT (*Fgg*<sup>+/+</sup>), and *Fgg*<sup>γ390–396A</sup> mice revealed no significant differences in the usage of any of the motifs (Figure S4A). Thus, *Fgg*<sup>γ390–396A</sup> does not change spontaneous behavior in a non-AD background. No sex effects were seen in 5xFAD and 5xFAD-*Fgg*<sup>γ390–396A</sup> mice (Figure S4B). Overall, this global restoration of spontaneous behavior shows that fibrinogen contributes to AD-related behavioral alterations and that ML alterations are susceptible to therapeutic intervention.

Kinematic and video analyses of the motifs that were significantly improved in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice were assigned to the following generic behavioral categories: run (motif 2), walk-up (motifs 6, 11, and 28), groom-up (motif 17), groom-down (motifs 10 and 12), and walk (motif 27) (Figures 6E and S5). These categories were confirmed by the hierarchical behavioral

(D) Hierarchical organization of motifs and subjects based on motif use relative to sex-matched WT littermates ( $n = 32$ ; 18 *App*<sup>NL-G-F</sup> and 14 WT mice). *App*<sup>NL-G-F</sup> and WT mice exhibited different clustering ( $p = 0.0015$  by Mann-Whitney rank-sum test).

(E) Percentage of time spent in identified communities in 5-min bins during 25 min of exploration in an open arena. Community usages were strongly modulated by time (experience), and *App*<sup>NL-G-F</sup> mice exhibited impaired time-dependent responses.  $p$  values were determined by repeated one-way ANOVA. Values are mean  $\pm$  SEM.





**Figure 5. Deconstructing full sequences of spontaneous behavior reveals disorganized behavioral sequences, increased motif transitions, and randomness in *App*<sup>NL-G-F</sup> mice**

Full behavioral sequences of motif transitions in 13-month-old *App*<sup>NL-G-F</sup> mice (n = 18; eight females and 10 males) and WT littermate controls (n = 14; eight females and six males) were annotated at 25 Hz during 25 min of open-arena exploration (~37,500 potential transitions per mouse) and analyzed with the discrete Markov chain model.

(A) Community transition sequences (black) and center speed (blue) of an *App*<sup>NL-G-F</sup> mouse during 60 s of behavioral exploration.

(legend continued on next page)

dendrogram, which showed significant 5xFAD-dependent usage alterations in eight of nine identified behavioral communities (FDR-BH,  $q < 0.05$ ) (Figure 7A, red branches). Furthermore, in a representative selection of 5xFAD-affected motifs (Figure 7B), motif use over time was strongly modulated by genotype, and 5xFAD-*Fgg*<sup>γ390–396A</sup> mice largely showed full or partial recovery to performances seen in control mice ( $p < 0.05$ , two-way ANOVA). Hierarchical clustering of motif usage and mice also revealed prominent disease-associated differences (Figure 7C). Together, these results indicate that fibrinogen is a key mediator of behavioral abnormalities in 5xFAD mice. They also suggest that alterations in spontaneous behavior detected by VAME are responsive to AD-related therapeutic interventions and provide unbiased and sensitive outcomes for assessment of treatments.

### Behavioral transition network analysis reveals therapeutic restoration in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice

To determine whether 5xFAD mice have disease-associated alterations of motif transitions, we generated transition networks at the cohort level for 5xFAD and 5xFAD-*Fgg*<sup>γ390–396A</sup> mice relative to controls (Figure 7D). As in *App*<sup>NL-G-F</sup> mice (Figure 5C), motifs were well organized into communities, and ambulatory behaviors (run, walk) formed the hub of the transition network. 5xFAD mice had marked abnormalities in the probability of most transitions, and the transition network was characterized by increased transitions between motifs and reduced probability of dwelling in the same motif (self-transitions). Thus, both *App*<sup>NL-G-F</sup> and 5xFAD mice exhibit behavioral instability, suggesting that the abnormalities in behavioral structure are an early preclinical feature of AD-related pathogenesis that progresses during clinical stages. To determine whether the motif transition network can be restored by blocking fibrinogen-microglia interactions, we compared 5xFAD-*Fgg*<sup>γ390–396A</sup> mice and controls. We found that most of the 5xFAD-dependent alterations in transitions were rescued or greatly mitigated in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice (Figure 7D), indicating restoration of behavioral sequences.

### Classifier analyses reveal VAME outcomes provide better sensitivity and specificity than conventional behavioral outcomes

To directly compare the specificity and sensitivity of VAME to those of conventional behavioral outcomes, we performed classifier analysis (logistic regression) using VAME outcomes (motif usage) or conventional open field (OF) outcomes (distance, speed, time, and location) obtained from the same videos (Figure S6). The genotype classifier using VAME outcomes had a

specificity (WT classified as WT) of 92.9% and a sensitivity (*App*<sup>NL-G-F</sup> classified as *App*<sup>NL-G-F</sup>) of 83.3%, whereas the genotype classifier using conventional OF outcomes had much lower specificity (50.0%) and sensitivity (77.8%) (Figure 5G). Similarly, for all genotypes, VAME outcomes for the 5xFAD cohort also had better sensitivity and specificity than conventional outcomes, categorizing 100% of mice correctly (Figure 6F). Thus, VAME analysis is more sensitive and specific than conventional outcomes for *App*<sup>NL-G-F</sup> and 5xFAD mice. Notably, only the VAME outcomes differentiated 5xFAD-*Fgg*<sup>γ390–396A</sup> from control mice, suggesting caution in interpreting therapeutic interventions in mouse models, as conventional approaches may not fully capture remaining disease phenotypes in rescued mice. Finally, to compare our VAME approach with keypoint-MoSeq,<sup>35</sup> we performed a side-by-side comparison using the same *App*<sup>NL-G-F</sup> and WT behavioral videos. Although VAME and keypoint-MoSeq used two different algorithms to segment the behavior into motifs or syllables, both approaches identified behavioral alterations in *App*<sup>NL-G-F</sup> mice (Figure S7).

## DISCUSSION

We studied spontaneous behavior in humanized *App*<sup>NL-G-F</sup> and 5xFAD mice using the ML-based VAME segmentation platform (<https://github.com/EthoML/VAME>).<sup>7</sup> We found that *App*<sup>NL-G-F</sup> mice exhibit robust age-dependent impairments in spontaneous behavior, including motif utilization, disorganized behavioral sequences, increased transitions, and randomness, indicating that these behavioral alterations reflect disease progression. Spontaneous behavior was altered in both male and female *App*<sup>NL-G-F</sup> mice, but females were particularly vulnerable to changes in motif usage, motif speed, and randomness of behavioral sequences. 5xFAD mice had more pronounced abnormalities in spontaneous behavior than controls and *App*<sup>NL-G-F</sup> mice, indicating a gene-dose-dependent effect. Interestingly, both *App*<sup>NL-G-F</sup> and 5xFAD mice displayed overlapping abnormalities, including increased motif speed, a higher number of transitions, and reduced motif dwell time. These results suggest that these distinct mouse models share common mechanisms underlying AD-related pathogenesis. These results also indicate that AD-related spontaneous behavioral alterations do not require APP overexpression and are likely linked to the pathologically elevated levels of amyloid in *App*<sup>NL-G-F</sup> mice.<sup>16,19</sup> We also tested the role of neuroinflammation in AD-related spontaneous behavioral alterations by blocking fibrinogen-microglia interactions in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice. Notably, the organization of spontaneous behavior was largely restored in 5xFAD-*Fgg*<sup>γ390–396A</sup>

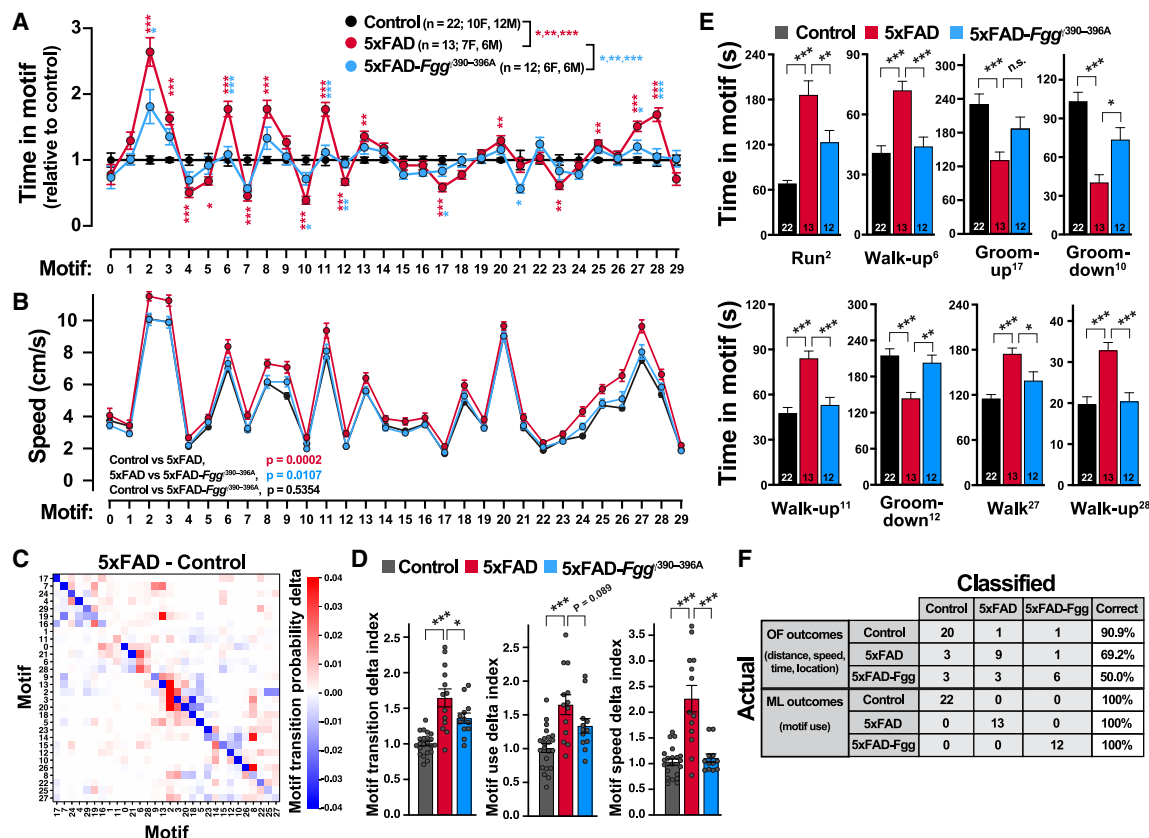
(B) Leave-one-out log likelihood estimates of orders 0 and 1 for the Markov chain model of community transitions. *App*<sup>NL-G-F</sup> mice, particularly females, had less predictable behavior.  $p$  values were determined by two-way (genotype and sex) ANOVA.

(C and D) Cytoscape network visualization of the top 20% most probable motif transitions depicting topological proximity of associated motifs and communities (color code) and significant alterations of motif use (circular borders) and motif transitions (arrows) in *App*<sup>NL-G-F</sup> mice. *App*<sup>NL-G-F</sup> mice had increased transitions (red arrows) and reduced dwell time in motifs (blue dashed arrows) (C). *App*<sup>NL-G-F</sup> mice systematically favored transitions from fast ambulatory to slow exploratory communities (D).

(E) Transition difference matrix depicting the delta probability of transitions between *App*<sup>NL-G-F</sup> and WT mice. Blue and red indicate reduced or increased probability of transition, respectively, in *App*<sup>NL-G-F</sup> mice relative to controls.

(F) Delta indices (*App*<sup>NL-G-F</sup> WT mice) of motif transition probability, motif use, and motif speed. \*\* $p < 0.01$ , \* $p < 0.05$  by  $t$  test.

(G) Classifier analyses (logistic regression) of ML outcomes (motif use) versus conventional open field (OF) outcomes (distance, speed, time, and location) obtained from the same videos in *App*<sup>NL-G-F</sup> mice. ML outcomes were more sensitive and specific than conventional outcomes. Values are mean  $\pm$  SEM.



**Figure 6. Pronounced 5xFAD-dependent alterations in spontaneous behavior are prevented by blocking fibrinogen-microglia interactions in 5xFAD-Fgg<sup>390-396A</sup> mice**

Spontaneous behavior of 8- to 10-month-old 5xFAD mice (n = 13; seven females and six males), 5xFAD-Fgg<sup>390-396A</sup> mice (n = 12; six females and six males), and controls (n = 22; 10 females and 12 males) was recorded for 60 min in a circular open arena. The control group included WT (n = 13; six females and seven males) and Fgg<sup>390-396A</sup> (n = 9; four females and five males) mice, as their behavior did not differ (Figure S4A). See Figure S4B for sex effects.

(A) Motif use of 30 VAME-identified motifs for all mice. Relative to controls, 5xFAD mice displayed motif use alterations in 17 motifs (red asterisks). Relative to 5xFAD mice, 5xFAD-Fgg<sup>390-396A</sup> showed significant improvements in many motifs (blue asterisks). \*\*\*q < 0.001, \*\*q < 0.01, \*q < 0.05 by FDR-BH. F, female; M, male.

(B) 5xFAD mice performed motifs at higher speeds, which was prevented in 5xFAD-Fgg<sup>390-396A</sup> mice. Genotype and motif effects were determined by two-way ANOVA (p values).

(C) Motif transition subtraction matrix depicting the delta probability of transitions between 5xFAD and control mice. Blue and red indicate reduced or increased probability of transition, respectively, in 5xFAD mice relative to controls.

(D) Delta indices of motif transition probability, motif use, and motif speed (n = 47 mice; 13 5xFAD, 12 5xFAD-Fgg<sup>390-396A</sup>, and 22 control mice). \*\*\*p < 0.001, \*p < 0.05 by one-way ANOVA and Bonferroni post hoc test for multiple comparisons.

(E) Time in motif (seconds) by genotype for the eight 5xFAD-affected motifs restored in 5xFAD-Fgg<sup>390-396A</sup> mice (A, q < 0.05 by FDR-HB). \*\*\*p < 0.001, \*\*p < 0.01, \*p < 0.05 by one-way ANOVA and Bonferroni post hoc test for multiple comparisons.

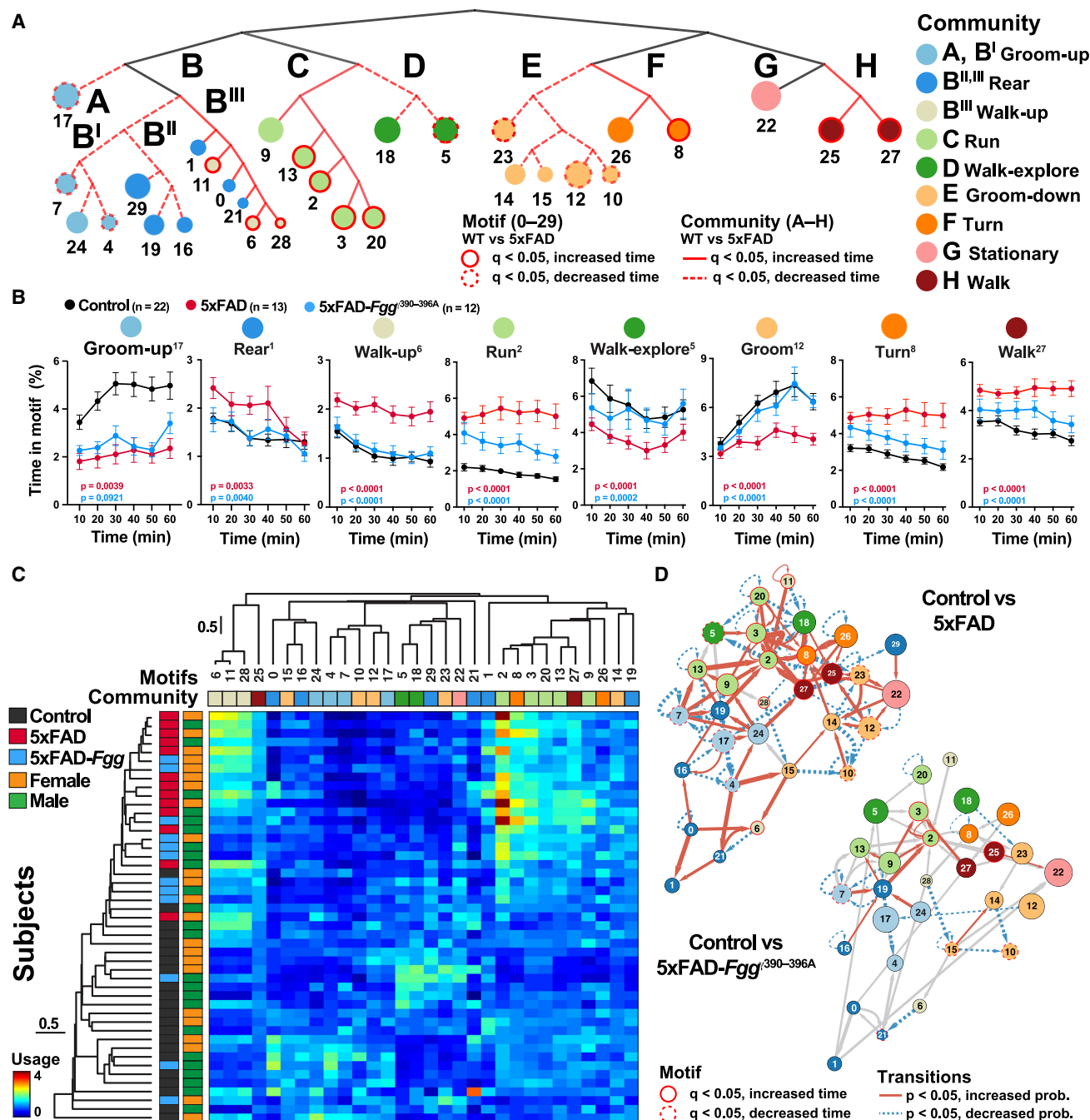
(F) Classifier analyses (logistic regression) of ML outcomes (motif use) versus conventional OF outcomes (distance, speed, time, and location) obtained from the same videos. ML outcomes were more sensitive and specific than conventional outcomes. Values are mean ± SEM.

mice, indicating that AD-related spontaneous behavioral changes are sensitive to therapeutic interventions and driven by neuroinflammation.

### Spontaneous behavioral alterations and AD pathogenesis

A major question is whether behavioral alterations in spontaneous behavior are mechanistically related to AD pathogenesis and are therefore driven by the same biological pathways and mechanisms of cognitive decline. Several observations suggest that this is the case. First, App<sup>NL-G-F</sup> and 5xFAD mice shared

behavioral alterations (e.g., increased motif speed, elevated number of transitions, and reduced dwell time in motifs), suggesting that these deficits are caused by common mechanisms of disease pathogenesis, such as familial AD mutations or abnormally high amyloid levels. It is worth noting that App<sup>NL-G-F</sup> and 5xFAD mice, in addition to differing in being knockin and transgenic models, have distinct sets of familial AD mutations and patterns of amyloid deposition.<sup>16,36,37</sup> Second, spontaneous behavioral abnormalities were age dependent and therefore tracked the progression of AD-related pathological and cognitive alterations.<sup>16,19,23</sup> Third, and more importantly, spontaneous



**Figure 7. Pronounced 5xFAD-dependent alterations in community usage and motif transition networks are prevented in 5xFAD-Fgg<sup>390–396A</sup> mice**

(A) Hierarchical organization of motifs into communities based on motif transitions and motif use at the cohort level (n = 47 mice; 13 5xFAD, 12 5xFAD-Fgg<sup>390–396A</sup>, and 22 control mice). Node size is proportional to motif use. Significant differences (indicated in red) between 5xFAD and control mice were determined by FDR-BH for 30 motifs and nine communities.

(B) Time in motif (%) of a representative motif for each disease-affected community (A) during the 60 min of exploration in an open arena. 5xFAD mice exhibited profound deficits, which were largely prevented in 5xFAD-Fgg<sup>390–396A</sup> mice. p values were determined by repeated one-way ANOVA with Bonferroni *post hoc* test for multiple comparisons. Values are mean ± SEM.

(C) Hierarchical organization of motifs and mice based on motif use relative to sex-matched control littermates.

(D) Cytoscape was used to visualize the top 20% most probable motif transitions to depict topological proximity of associated motifs and communities (color code) and significant alterations of motif use (borders) and motif transitions (arrows) for control versus 5xFAD mice (top) and control versus 5xFAD-Fgg<sup>390–396A</sup> mice (bottom).



behavioral abnormalities in 5xFAD mice were prevented by an AD-related therapeutic intervention, *Fgg*<sup>γ390–396A</sup>, that prevents CD11b receptor-mediated interactions with microglia<sup>22,33</sup> and reduces AD-related pathogenesis, including amyloid burden, dystrophic neurites, synaptic loss, microgliosis, and neuroinflammation.<sup>21,22,34</sup> Thus, our findings suggest that abnormalities in spontaneous behavior are mechanistically related to AD pathogenesis and may offer mechanistically relevant endpoints for assessing AD-related pathways and therapeutic interventions in mouse models.

Although we did not test other neurodegenerative models to assess the disease selectivity of the motif abnormalities, we hypothesize that the structure of spontaneous behavior (e.g., transition networks) will exhibit selective or even pathognomonic elements for different disorders (e.g., AD versus Parkinson's disease models). Overall, unbiased and high-throughput ML approaches such as the one we used to assess spontaneous behavior address a major unmet need in translational research by identifying translationally relevant, functional outcome measures in AD preclinical models. The current approach does not directly assess cognitive functions, but our findings suggest that these outcomes capture key elements of AD pathogenesis, such as inflammation, that could complement traditional task-oriented behavioral approaches.

Behavioral or neuropsychiatric changes, such as depression, agitation, anxiety, apathy, irritability, delusions, or sleep disruptions, are often apparent years before cognitive decline or dementia. These behavioral changes are associated with increased risk of dementia and AD,<sup>13,14</sup> suggesting that they contribute to AD pathogenesis.<sup>13,14</sup> ML approaches have been developed to predict these behavioral abnormalities in AD patients by using data on sleep patterns and physical activity collected by actigraphy from wearable devices.<sup>38</sup> It is important to emphasize that our results and conclusions apply solely to mouse models of AD and that our ML pipeline currently has no clinical application. However, ML approaches to assess spontaneous behavior in humans, as we have done in mice, can provide early functional outcomes for disease diagnosis and the assessment of treatments and have the potential to reveal and quantify AD-related behavioral manifestations that are otherwise difficult to observe.

### **Fibrinogen blockade protects from neurobehavioral abnormalities**

Our findings show that VAME is a high-throughput approach for unbiased assessment of therapeutic interventions. In patients and mouse models of AD, fibrinogen extravasates in the brain at sites of impaired blood-brain barrier integrity and cerebrovascular dysfunction, which are early features of AD pathogenesis.<sup>21,39–42</sup> Fibrinogen is converted to fibrin, which is deposited at sites of microglia activation in the AD brain. Fibrin is present at increased levels in *APOE4* AD patients, correlates with the severity of cerebral amyloid angiopathy, and is detected in amyloid-related imaging abnormalities.<sup>32,43,44</sup> The binding of Aβ to fibrinogen inhibits fibrin degradation, resulting in chronic fibrin deposition in the AD brain.<sup>32</sup> Biomarker studies further support the notion that fibrinogen elevation in plasma and cerebrospinal fluid is an early biomarker of dementia risk and is associated with progression of cognitive decline in AD.<sup>39,45–48</sup> Notably, we found

that genetic blockade of fibrinogen-microglia interactions in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice reduced virtually all the motif usage abnormalities found in 5xFAD mice. These findings, coupled with the restoration of the impaired transition networks in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice, suggest that fibrinogen is a key driver of the disorganization of behavioral sequences in 5xFAD mice. Our results are consistent with the reduced expression of neurodegenerative and oxidative stress genes in microglia and protection against cognitive impairment in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice.<sup>21,22</sup> Accordingly, a fibrin-targeting antibody that selectively inhibits microglia activation without adverse anticoagulant effects protects from neurodegeneration and suppresses AD-related gene networks in 5xFAD mice.<sup>28,34</sup> Fibrin-targeting immunotherapy is currently in phase 1 clinical trials.<sup>49</sup>

### **ML approaches for behavioral segmentation**

ML approaches for behavioral segmentation have the potential to detect subtle or overlooked behavioral abnormalities linked to early brain dysfunction. For example, MoSeq has uncovered previously undetected patterns of spontaneous behavior associated with epileptic activity, providing a behavioral biomarker of brain dysfunction.<sup>9</sup> MoSeq and VAME are based on unsupervised dimensionality reduction methods, which analyze the implicit structure and patterns of postural motion to identify the most relevant features that capture the spatiotemporal variability of the data to define brief stereotypical postural units. MoSeq<sup>5</sup> and keypoint-MoSeq<sup>35</sup> use a statistical framework known as autoregressive hidden Markov model (AR-HMM),<sup>6,50,51</sup> whereas VAME<sup>1,7</sup> uses deep recurrent neural networks and variational autoencoders, providing two well-established complementary approaches.<sup>7,52,53</sup> In a side-by-side comparison of VAME and keypoint-MoSeq, these two approaches for behavioral segmentation differed in motif/syllable distributions but both identified robust behavioral alterations in *App*<sup>NL-G-F</sup> mice.

VAME works with ventral imaging to assess the movement of seven body parts (four paws, nose, center, and tail base) relative to the mouse center (egocentric coordinates). Although VAME segmentation does not incorporate additional variables such as mouse speed, height, or location, motifs with ambulatory kinematics (e.g., run or fast-explore) exhibited higher allocentric speeds than stationary motifs (e.g., groom or sit). This finding suggests that VAME segmentation and egocentric coordinates reflect allocentric speeds. This property might be important when assessing disease models that have locomotor abnormalities, since adding speed into the neural network may bias motif clustering toward differences in speeds among genotypes and dilute other aspects of behavioral features.

### **Communities and transition networks for assessing behavioral organization**

Clustering motifs into communities provides a hierarchical organization of mouse behavior that incorporates the probability of motif-motif transitions, thereby grouping motifs according to the likelihood of sequential use rather than pose similarity. The resulting communities represent cohesive postural sequences of ethologically relevant behavior. Indeed, motifs in the same communities responded to *App*<sup>NL-G-F</sup> and 5xFAD in similar

ways in both direction and magnitude, suggesting that communities cluster disease-relevant behaviors. The same communities were also apparent in the unsupervised layout of transition networks, a method that relies on the overall topological connectivity among motifs and is not biased by prior community assignments. By creating and calculating global delta indices (transitions, usage, and speed indices), we found that transitions were particularly sensitive to *App*<sup>NL-G-F</sup> and 5xFAD, suggesting that the core abnormality in AD models is the disorganization of behavioral sequences rather than the total time spent in each behavior. While sex and genotype interactions are difficult to detect and assess with conventional behavioral tests, VAME captured robust differences in the spontaneous behavior of males and females, making it an ideal platform to assess sex-genotype interactions.

### Scalable and sensitive methods for assessing mouse behavior

Scalable, sensitive, and unbiased methods for behavioral phenotyping are essential for basic and translational neuroscience efforts aimed at phenotyping or assessing therapeutic interventions in mouse models of neurodegenerative disease. The ability to scale ML approaches to large cohorts of mice also facilitates well-powered evaluation of treatment effectiveness across sex and dosages. Moreover, when we assessed the specificity and sensitivity of ML (motif usage) and conventional behavior (distance, speed, and location) outcomes from the same behavioral videos, the ML outcomes classified genotypes and treatments with greater specificity and sensitivity. Remarkably, ML, but not conventional, outcomes distinguished rescued 5xFAD-*Fgg*<sup>γ390-396A</sup> mice from controls, indicating that conventional approaches lack sufficient sensitivity and specificity to capture remaining disease phenotypes in rescued mice, which may result in an overestimation of the beneficial effects of the intervention.

### Limitations of the study

Our findings suggest that ML approaches to assess spontaneous behavior provide a direct, sensitive, and unbiased measure of brain dysfunction induced by AD-related mechanisms in mouse models. A limitation of our study assessing spontaneous behavior is that specific behavioral domains (e.g., cognitive functions) and functional brain topology were not explicitly assessed. Thus, future studies will be needed to mechanistically determine the contribution of distinct behavioral domains, neural systems, or brain regions to the identified spontaneous behavioral changes. For example, our ML approach may be directly applied to study spontaneous behavior during learning and memory tasks. Another limitation of the study is that it does not resolve whether impairments of spontaneous behavior are driven by biological pathways and mechanisms of cognitive decline. It remains to be determined whether some of these outcomes could represent biomarkers of cognitive dysfunction. Despite these unresolved questions, we speculate that unbiased ML approaches will improve translatability of preclinical testing by providing rigorous quantification of disease-induced behavioral alterations to assess the construct and predictive validity of mouse models of neurodegenerative diseases.

### RESOURCE AVAILABILITY

#### Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Jorge J. Palop ([jorge.palop@gladstone.ucsf.edu](mailto:jorge.palop@gladstone.ucsf.edu)).

#### Materials availability

This study did not generate reagents.

#### Data and code availability

VAME code of record has been archived at Zenodo (<https://doi.org/10.5281/zenodo.13787335>). An actively maintained VAME package and repository for behavioral phenotyping can be found at <https://github.com/EthoML/VAME>. Code and algorithms are available at the GitHub repository or are described in the [STAR Methods](#). Any additional information required to reanalyze the data reported in this work paper is available from the [lead contact](#) upon request.

### ACKNOWLEDGMENTS

We thank Takashi Saito and Takaomi C. Saido for the *App*<sup>NL-G-F</sup> mice; the Gladstone Institute Behavioral Core, including Jeffrey Simms and Iris Lo, for behavioral interpretation; Stephen Ordway for editorial review; and Allan Villanueva, Rhodora Gacayan, Pilar Alzamora, and Eilidh MacDonald for technical assistance. The study was supported by US National Institutes of Health (NIH) grants RF1AG062234, P01AG062629, R01AG073082, and AG082147 (J.J.P.), K01-AG083732 (S.R.M.), K99 NS126707 (A.S.M.), R35NS097976 (K.A.), RF1AG06926 (K.A. and J.J.P.), and R01AG061785 and RF1AG059405 (E.D.R.); the European Research Council (K. Luxem., P.B., and S.R.); DFG Collaborative Research Center 1436 (S.R.); and DFG RTG 2413 SynAGE (S.R.).

### AUTHOR CONTRIBUTIONS

Conceptualization, S.R.M., K. Luxem, S.R., and J.J.P.; experimentation, S.R.M., K. Lauderdale, P.N., C.C., P.S.H., J.S., and K.S.; computation, K. Luxem, S.R.M., P.N., N.K., K. Lauderdale, Y.Q., K.D.M., M.B., A.R.P., R.T., S.B., and P.B.; mouse models, Z.Y., A.S.M., T.S., T.C.S., and K.A.; supervision, S.R.M., K.A., E.D.R., and J.J.P.; writing, S.R.M. and J.J.P. with input from all authors.

### DECLARATION OF INTERESTS

K.A. is the scientific founder, advisor, and shareholder of Therini Bio; her conflicts of interests are managed by the Gladstone Institutes. K.L. is currently an employee of REWE Group (Germany). P.B. is currently an employee of Telekom (Germany).

### STAR★METHODS

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

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
  - Mice
  - Mouse cohorts and experimental groups
- [METHOD DETAILS](#)
  - Behavioral experiments
  - Videography for machine-learning analysis
  - Pose estimation
  - Behavioral segmentation and sequencing
  - Uniform manifold approximation (UMAP)
  - Motif and community usage
  - Motif usage correlation cluster analyses
  - Motif transition matrix
  - Hierarchical community dendrogram

- Transition network analyses and visualization
- Evaluation of order of Markov chain models and pseudo Bayes factors
- Delta indices for motif transitions, motif use, and motif speed
- Keypoint-MoSeq
- Community glossary
- Histology and immunohistochemistry
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Statistical analysis

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2024.114870>.

Received: October 9, 2023

Revised: July 18, 2024

Accepted: September 27, 2024

## REFERENCES

1. Luxem, K., Sun, J.J., Bradley, S.P., Krishnan, K., Yttri, E., Zimmermann, J., Pereira, T.D., and Laubach, M. (2023). Open-source tools for behavioral video analysis: Setup, methods, and best practices. *Elife* 12, e79305. <https://doi.org/10.7554/eLife.79305>.
2. 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. *Nat. Neurosci.* 21, 1281–1289. <https://doi.org/10.1038/s41593-018-0209-y>.
3. Nath, T., Mathis, A., Chen, A.C., Patel, A., Bethge, M., and Mathis, M.W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protoc.* 14, 2152–2176. <https://doi.org/10.1038/s41596-019-0176-0>.
4. Pereira, T.D., Tabris, N., Matsliah, A., Turner, D.M., Li, J., Ravindranath, S., Papadopoulos, E.S., Normand, E., Deutsch, D.S., Wang, Z.Y., et al. (2022). SLEAP: A deep learning system for multi-animal pose tracking. *Nat. Methods* 19, 486–495. <https://doi.org/10.1038/s41592-022-01426-1>.
5. Biderman, D., Whiteway, M.R., Hurwitz, C., Greenspan, N., Lee, R.S., Vishnubhotla, A., Warren, R., Pedraja, F., Noone, D., Scharfner, M.M., et al. (2024). Lightning Pose: improved animal pose estimation via semi-supervised learning, Bayesian ensembling and cloud-native open-source tools. *Nat. Methods* 21, 1316–1328. <https://doi.org/10.1038/s41592-024-02319-1>.
6. Wilschko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abaira, V.E., Adams, R.P., and Datta, S.R. (2015). Mapping sub-second structure in mouse behavior. *Neuron* 88, 1121–1135. <https://doi.org/10.1016/j.neuron.2015.11.031>.
7. Luxem, K., Mocellin, P., Fuhrmann, F., Kürsch, J., Miller, S.R., Palop, J.J., Remy, S., and Bauer, P. (2022). Identifying behavioral structure from deep variational embeddings of animal motion. *Commun. Biol.* 5, 1267. <https://doi.org/10.1038/s42003-022-04080-7>.
8. Mocellin, P., Barnstedt, O., Luxem, K., Kaneko, H., Vieweg, S., Henschke, J.U., Dalügge, D., Fuhrmann, F., Karpova, A., Papan, J.M.P., et al. (2024). A septal-ventral tegmental area circuit drives exploratory behavior. *Neuron* 112, 1020–1032.e7. In press. <https://doi.org/10.1016/j.neuron.2023.12.016>.
9. Gschwind, T., Zeine, A., Raikov, I., Markowitz, J.E., Gillis, W.F., Felong, S., Isom, L.L., Datta, S.R., and Soltesz, I. (2023). Hidden behavioral fingerprints in epilepsy. *Neuron* 111, 1440–1452.e5. <https://doi.org/10.1016/j.neuron.2023.02.003>.
10. Hanseeuw, B.J., Betensky, R.A., Jacobs, H.I.L., Schultz, A.P., Sepulcre, J., Becker, J.A., Cosio, D.M.O., Farrell, M., Quiroz, Y.T., Mormino, E.C., et al. (2019). Association of amyloid and tau with cognition in preclinical Alzheimer disease: A longitudinal study. *JAMA Neurol.* 76, 915–924. <https://doi.org/10.1001/jamaneurol.2019.1424>.
11. Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Jr., Kaye, J., Montine, T.J., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 280–292. S1552-5260(11)00099-9 [pii]. <https://doi.org/10.1016/j.jalz.2011.03.003>.
12. McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Jr., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 263–269. <https://doi.org/10.1016/j.jalz.2011.03.005>.
13. Rosenberg, P.B., Mielke, M.M., Appleby, B.S., Oh, E.S., Geda, Y.E., and Lyketsos, C.G. (2013). The association of neuropsychiatric symptoms in MCI with incident dementia and Alzheimer disease. *Am. J. Geriatr. Psychiatr.* 21, 685–695. <https://doi.org/10.1016/j.jagp.2013.01.006>.
14. Peters, M.E., Rosenberg, P.B., Steinberg, M., Norton, M.C., Welsh-Bohmer, K.A., Hayden, K.M., Breitner, J., Tschanz, J.T., and Lyketsos, C.G.; Cache County Investigators (2013). Neuropsychiatric symptoms as risk factors for progression from CIND to dementia: the Cache County Study. *Am. J. Geriatr. Psychiatr.* 21, 1116–1124. <https://doi.org/10.1016/j.jagp.2013.01.049>.
15. Saito, T., Matsuba, Y., Mihira, N., Takano, J., Nilsson, P., Itoharu, S., Iwata, N., and Saido, T.C. (2014). Single App knock-in mouse models of Alzheimer's disease. *Nat. Neurosci.* 17, 661–663. <https://doi.org/10.1038/nn.3697>.
16. Xia, D., Lianoglou, S., Sandmann, T., Calvert, M., Suh, J.H., Thomsen, E., Dugas, J., Pizzo, M.E., DeVos, S.L., Earr, T.K., et al. (2022). Novel App knock-in mouse model shows key features of amyloid pathology and reveals profound metabolic dysregulation of microglia. *Mol. Neurodegener.* 17, 41. <https://doi.org/10.1186/s13024-022-00547-7>.
17. Latif-Hernandez, A., Shah, D., Craessaerts, K., Saido, T., Saito, T., De Strooper, B., Van der Linden, A., and D'Hooge, R. (2019). Subtle behavioral changes and increased prefrontal-hippocampal network synchronicity in APP(NL-G-F) mice before prominent plaque deposition. *Behav. Brain Res.* 364, 431–441. <https://doi.org/10.1016/j.bbr.2017.11.017>.
18. Whyte, L.S., Hemsley, K.M., Lau, A.A., Hassiotis, S., Saito, T., Saido, T.C., Hopwood, J.J., and Sargeant, T.J. (2018). Reduction in open field activity in the absence of memory deficits in the App(NL-G-F) knock-in mouse model of Alzheimer's disease. *Behav. Brain Res.* 336, 177–181. <https://doi.org/10.1016/j.bbr.2017.09.006>.
19. Johnson, E.C.B., Ho, K., Yu, G.Q., Das, M., Sanchez, P.E., Djukic, B., Lopez, I., Yu, X., Gill, M., Zhang, W., et al. (2020). Behavioral and neural network abnormalities in human APP transgenic mice resemble those of App knock-in mice and are modulated by familial Alzheimer's disease mutations but not by inhibition of BACE1. *Mol. Neurodegener.* 15, 53. <https://doi.org/10.1186/s13024-020-00393-5>.
20. Sasaguri, H., Nilsson, P., Hashimoto, S., Nagata, K., Saito, T., De Strooper, B., Hardy, J., Vassar, R., Winblad, B., and Saido, T.C. (2017). APP mouse models for Alzheimer's disease preclinical studies. *EMBO J.* 36, 2473–2487. <https://doi.org/10.15252/embj.201797397>.
21. Merlini, M., Rafalski, V.A., Rios Coronado, P.E., Gill, T.M., Ellisman, M., Muthukumar, G., Subramanian, K.S., Ryu, J.K., Syme, C.A., Davalos, D., et al. (2019). Fibrinogen induces microglia-mediated spine elimination and cognitive impairment in an Alzheimer's disease model. *Neuron* 101, 1099–1108.e6. <https://doi.org/10.1016/j.neuron.2019.01.014>.
22. Mendiola, A.S., Yan, Z., Dixit, K., Johnson, J.R., Bouhaddou, M., Meyer-Franke, A., Shin, M.G., Yong, Y., Agrawal, A., MacDonald, E., et al. (2023). Defining blood-induced microglia functions in neurodegeneration through multiomic profiling. *Nat. Immunol.* 24, 1173–1187. <https://doi.org/10.1038/s41590-023-01522-0>.

23. Wang, S., Ichinomiya, T., Savchenko, P., Devulapalli, S., Wang, D., Beltz, G., Saito, T., Saido, T.C., Wagner, S.L., Patel, H.H., and Head, B.P. (2022). Age-Dependent Behavioral and Metabolic Assessment of App<sup>NL-G-F/NL-G-F</sup> Knock-in (KI) Mice. *Front. Mol. Neurosci.* 15, 909989. <https://doi.org/10.3389/fnmol.2022.909989>.
24. Verret, L., Mann, E.O., Hang, G.B., Barth, A.M.I., Cobos, I., Ho, K., Devidze, N., Masliah, E., Kreitzer, A.C., Mody, I., et al. (2012). Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in Alzheimer model. *Cell* 149, 708–721. <https://doi.org/10.1016/j.cell.2012.02.046>.
25. Palop, J.J., Jones, B., Kekoni, L., Chin, J., Yu, G.Q., Raber, J., Masliah, E., and Mucke, L. (2003). Neuronal depletion of calcium-dependent proteins in the dentate gyrus is tightly linked to Alzheimer's disease-related cognitive deficits. *Proc. Natl. Acad. Sci. USA* 100, 9572–9577. <https://doi.org/10.1073/pnas.113381100>.
26. Palop, J.J., Chin, J., Roberson, E.D., Wang, J., Thwin, M.T., Bien-Ly, N., Yoo, J., Ho, K.O., Yu, G.-Q., Kreitzer, A., et al. (2007). Aberrant excitatory neuronal activity and compensatory remodeling of inhibitory hippocampal circuits in mouse models of Alzheimer's disease. *Neuron* 55, 697–711. <https://doi.org/10.1016/j.neuron.2007.07.025>.
27. Martinez-Losa, M., Tracy, T.E., Ma, K., Verret, L., Clemente-Perez, A., Khan, A.S., Cobos, I., Ho, K., Gan, L., Mucke, L., et al. (2018). Nav1.1-over-expressing interneuron transplants restore brain rhythms and cognition in a mouse model of Alzheimer's disease. *Neuron* 98, 75–89.e5. <https://doi.org/10.1016/j.neuron.2018.02.029>.
28. Ryu, J.K., Yan, Z., Montano, M., Sozmen, E.G., Dixit, K., Suryawanshi, R.K., Matsui, Y., Helmy, E., Kaushal, P., Mankanani, S.K., et al. (2024). Fibrin drives thromboinflammation and neuropathology in COVID-19. *Nature* 633, 905–913. <https://doi.org/10.1038/s41586-024-07873-4>.
29. Pentland, A., and Liu, A. (1999). Modeling and prediction of human behavior. *Neural Comput.* 11, 229–242. <https://doi.org/10.1162/089976699300016890>.
30. McComb, C., Cagan, J., and Kotovsky, K. (2017). Capturing human sequence-learning abilities in configuration design tasks through Markov chains. *J. Mech. Des.* 139, 0911011. <https://doi.org/10.1115/1.4037185>.
31. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>.
32. Cortes-Canteli, M., Paul, J., Norris, E.H., Bronstein, R., Ahn, H.J., Zamo- lodchikov, D., Bhuvanendran, S., Fenz, K.M., and Strickland, S. (2010). Fibrinogen and beta-amyloid association alters thrombosis and fibrinolysis: a possible contributing factor to Alzheimer's disease. *Neuron* 66, 695–709. <https://doi.org/10.1016/j.neuron.2010.05.014>.
33. Petersen, M.A., Ryu, J.K., and Akassoglou, K. (2018). Fibrinogen in neurological diseases: mechanisms, imaging and therapeutics. *Nat. Rev. Neurosci.* 19, 283–301. <https://doi.org/10.1038/nrn.2018.13>.
34. Ryu, J.K., Rafalski, V.A., Meyer-Franke, A., Adams, R.A., Poda, S.B., Rios Coronado, P.E., Pedersen, L.O., Menon, V., Baeten, K.M., Sikorski, S.L., et al. (2018). Fibrin-targeting immunotherapy protects against neuroinflammation and neurodegeneration. *Nat. Immunol.* 19, 1212–1223. <https://doi.org/10.1038/s41590-018-0232-x>.
35. Weinreb, C., Pearl, J.E., Lin, S., Osman, M.A.M., Zhang, L., Annapragada, S., Conlin, E., Hoffmann, R., Makowska, S., Gillis, W.F., et al. (2024). Key-point-MoSeq: parsing behavior by linking point tracking to pose dynamics. *Nat. Methods* 21, 1329–1339. <https://doi.org/10.1038/s41592-024-02318-2>.
36. Sato, K., Watamura, N., Fujioka, R., Mihira, N., Sekiguchi, M., Nagata, K., Ohshima, T., Saito, T., Saido, T.C., and Sasaguri, H. (2021). A third-generation mouse model of Alzheimer's disease shows early and increased cored plaque pathology composed of wild-type human amyloid beta peptide. *J. Biol. Chem.* 297, 101004. <https://doi.org/10.1016/j.jbc.2021.101004>.
37. Oakley, H., Cole, S.L., Logan, S., Maus, E., Shao, P., Craft, J., Guillozet-Bongaarts, A., Ohno, M., Disterhoft, J., Van Eldik, L., et al. (2006). Intra-neuronal  $\beta$ -amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* 26, 10129–10140.
38. Cho, E., Kim, S., Heo, S.J., Shin, J., Hwang, S., Kwon, E., Lee, S., Kim, S., and Kang, B. (2023). Machine learning-based predictive models for the occurrence of behavioral and psychological symptoms of dementia: model development and validation. *Sci. Rep.* 13, 8073. <https://doi.org/10.1038/s41598-023-35194-5>.
39. Iturria-Medina, Y., Sotero, R.C., Toussaint, P.J., Mateos-Pérez, J.M., and Evans, A.C.; Alzheimer's Disease Neuroimaging Initiative (2016). Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. *Nat. Commun.* 7, 11934. <https://doi.org/10.1038/ncomms11934>.
40. Nation, D.A., Sweeney, M.D., Montagne, A., Sagare, A.P., D'Orazio, L.M., Pachicano, M., Seppeband, F., Nelson, A.R., Buennagel, D.P., Harrington, M.G., et al. (2019). Blood-brain barrier breakdown is an early biomarker of human cognitive dysfunction. *Nat. Med.* 25, 270–276. <https://doi.org/10.1038/s41591-018-0297-y>.
41. Iadecola, C. (2013). The pathobiology of vascular dementia. *Neuron* 80, 844–866. <https://doi.org/10.1016/j.neuron.2013.10.008>.
42. Ryu, J.K., and McLarnon, J.G. (2009). A leaky blood-brain barrier, fibrinogen infiltration and microglial reactivity in inflamed Alzheimer's disease brain. *J. Cell Mol. Med.* 13, 2911–2925. <https://doi.org/10.1111/j.1582-4934.2008.00434.x>.
43. Kozberg, M.G., Yi, I., Freeze, W.M., Auger, C.A., Scherlek, A.A., Greenberg, S.M., and van Veluw, S.J. (2022). Blood-brain barrier leakage and perivascular inflammation in cerebral amyloid angiopathy. *Brain Commun.* 4, fcac245. <https://doi.org/10.1093/braincomms/fcac245>.
44. Hultman, K., Strickland, S., and Norris, E.H. (2013). The APOE  $\epsilon$ 4/ $\epsilon$ 4 genotype potentiates vascular fibrin(ogen) deposition in amyloid-laden vessels in the brains of Alzheimer's disease patients. *J. Cerebr. Blood Flow Metabol.* 33, 1251–1258. <https://doi.org/10.1038/jcbfm.2013.76>.
45. Xu, G., Zhang, H., Zhang, S., Fan, X., and Liu, X. (2008). Plasma fibrinogen is associated with cognitive decline and risk for dementia in patients with mild cognitive impairment. *Int. J. Clin. Pract.* 62, 1070–1075. <https://doi.org/10.1111/j.1742-1241.2007.01268.x>.
46. van Oijen, M., Witterman, J.C., Hofman, A., Koudstaal, P.J., and Breteler, M.M.B. (2005). Fibrinogen is associated with an increased risk of Alzheimer disease and vascular dementia. *Stroke* 36, 2637–2641. <https://doi.org/10.1161/01.STR.0000189721.31432.26>.
47. Fan, D.Y., Sun, H.L., Sun, P.Y., Jian, J.M., Li, W.W., Shen, Y.Y., Zeng, F., Wang, Y.J., and Bu, X.L. (2020). The correlations between plasma fibrinogen with amyloid-beta and tau levels in patients with Alzheimer's disease. *Front. Neurosci.* 14, 625844. <https://doi.org/10.3389/fnins.2020.625844>.
48. Yang, H., Lyutvinskiy, Y., Herukka, S.K., Soininen, H., Rutishauser, D., and Zubarev, R.A. (2014). Prognostic polypeptide blood plasma biomarkers of Alzheimer's disease progression. *J. Alzheimers Dis.* 40, 659–666. <https://doi.org/10.3233/JAD-132102>.
49. Kantor, A.B., Akassoglou, K., and Stavenhagen, J.B. (2023). Fibrin-targeting immunotherapy for dementia. *J. Prev. Alzheimers Dis.* 10, 647–660. <https://doi.org/10.14283/jpad.2023.105>.
50. Markowitz, J.E., Gillis, W.F., Jay, M., Wood, J., Harris, R.W., Cieszkowski, R., Scott, R., Brann, D., Koveal, D., Kula, T., et al. (2023). Spontaneous behaviour is structured by reinforcement without explicit reward. *Nature* 614, 108–117. <https://doi.org/10.1038/s41586-022-05611-2>.
51. Markowitz, J.E., Gillis, W.F., Beron, C.C., Neufeld, S.Q., Robertson, K., Bhagat, N.D., Peterson, R.E., Peterson, E., Hyun, M., Linderman, S.W., et al. (2018). The striatum organizes 3D behavior via moment-to-moment action selection. *Cell* 174, 44–58.e17. <https://doi.org/10.1016/j.cell.2018.04.019>.



52. Shi, C., Levy, S., Achvat, S., Abboud, M., Ghanayim, A., Schiller, J., and Mishne, G. (2021). Learning disentangled behavior embeddings. *NeurIPS* 7734, 1–11.
53. Whiteway, M.R., Biderman, D., Friedman, Y., Dipoppa, M., Buchanan, E.K., Wu, A., Zhou, J., Bonacchi, N., Miska, N.J., Noel, J.P., et al. (2021). Partitioning variability in animal behavioral videos using semi-supervised variational autoencoders. *PLoS Comput. Biol.* 17, e1009439. <https://doi.org/10.1371/journal.pcbi.1009439>.
54. Flick, M.J., Du, X., Witte, D.P., Jirousková, M., Soloviev, D.A., Busuttill, S.J., Plow, E.F., and Degen, J.L. (2004). Leukocyte engagement of fibrin(ogen) via the integrin receptor  $\alpha$ 5 $\beta$ 1 is critical for host inflammatory response in vivo. *J. Clin. Invest.* 113, 1596–1606. <https://doi.org/10.1172/JCI20741>.
55. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proc. CVPR* 770–778. <https://doi.org/10.1109/Cvpr.2016.90>.
56. Favuzzi, E., Huang, S., Saldi, G.A., Binan, L., Ibrahim, L.A., Fernández-Otero, M., Cao, Y., Zeine, A., Sefah, A., Zheng, K., et al. (2021). GABA-receptive microglia selectively sculpt developing inhibitory circuits. *Cell* 184, 4048–4063.e32. <https://doi.org/10.1016/j.cell.2021.06.018>.
57. Gelfand, A.E., and Dey, D.K. (1994). Bayesian Model Choice - Asymptotics and Exact Calculations. *J. Roy. Stat. Soc. B* 56, 501–514.
58. Bates, D., Machler, M., Bolker, B.M., and Walker, S.C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Software* 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>.
59. <https://www.r-project.org/about.html>, T.R.F.S.C. (2013).

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Amyloid $\beta$ (N) (82E1) Anti-Human Biotin	IBL	RRID: AB_10705565
Iba1	Wako	RRID: AB_839504
<b>Experimental models: Organisms/strains</b>		
<i>App</i> <sup>NL-G-F</sup> : C57BL/6- <i>App</i> <tm3(NL-G-F)Tcs>	RIKEN; Dr. Saido	RRID:IMSR_RBRC06344 <sup>15</sup>
5xFAD: B6.Cg-Tg (APPSwF1Lon, PSEN1*M146L*L286V)6799Vas/Mmjax	Jackson Laboratory;	RRID:MMRRC_034848-JAX <sup>37</sup>
<i>B6.Fgg</i> <sup><math>\gamma</math>390-396A</sup>	Dr. Jay Degen (University Of Cincinnati, Cincinnati, OH, USA)	Flick et al. <sup>54</sup>

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

#### Mice

*App*<sup>NL-G-F</sup> mice were obtained from Drs. Takaomi Saido and Takashi Saito (RIKEN Brain Science Institute, Japan).<sup>15</sup> We studied male and female homozygous *App*<sup>NL-G-F/NL-G-F</sup> mice expressing humanized A $\beta$  and the familial Alzheimer's disease (FAD) Swedish (K670N, M671L), Arctic (E693G), and Iberian (717, I) mutations on the C57BL/6J background.<sup>15</sup> To obtain wildtype (WT) littermate controls, we crossed heterozygous *App*<sup>NL-G-F/+</sup> mice to produce homozygous *App*<sup>NL-G-F/NL-G-F</sup> mice and *App*<sup>+/+</sup> (WT) littermate controls of both sexes. Homozygous *App*<sup>NL-G-F/NL-G-F</sup> mice are referred to as *App*<sup>NL-G-F</sup> mice. Heterozygous *App*<sup>NL-G-F/+</sup> mice were not studied.

5xFAD mice were from The Jackson Laboratory (Jax 0087030).<sup>37</sup> We studied male and female heterozygous 5xFAD mice overexpressing both human amyloid beta precursor protein (APP) with the Swedish (K670N, M671L), Florida (I716V), and London (V717I) FAD mutations and human presenilin 1 (PS1) with the M146L and L286V FAD mutations on the C57BL/6J background.<sup>37</sup> *Fgg* <sup>$\gamma$ 390-396A</sup> mice were from Dr. Jay Degen (University of Cincinnati, OH, USA) and bred in the Akassoglou Lab.<sup>54</sup> Heterozygous 5xFAD mice were crossed with wildtype (WT) mice to produce heterozygous 5xFAD and WT mice on the C57BL/6J background. Homozygous *Fgg* <sup>$\gamma$ 390-396A/ $\gamma$ 390-396A</sup> mice were crossed with 5xFAD-*Fgg* <sup>$\gamma$ 390-396A/ $\gamma$ 390-396A</sup> mice to produce 5xFAD-*Fgg* <sup>$\gamma$ 390-396A/ $\gamma$ 390-396A</sup> and *Fgg* <sup>$\gamma$ 390-396A/ $\gamma$ 390-396A</sup> mice. Homozygous *Fgg* <sup>$\gamma$ 390-396A/ $\gamma$ 390-396A</sup> mice are referred to as *Fgg* <sup>$\gamma$ 390-396A</sup> mice. Experimental mice of the same sex were group-housed with access to water and food *ad libitum* in a controlled environment with a 12-h light-dark cycle.

#### Mouse cohorts and experimental groups

Young *App*<sup>NL-G-F</sup> cohort: 5–6-month-old *App*<sup>NL-G-F</sup> mice ( $n = 13$ ; 7 females and 6 males; mean age,  $6.17 \pm 0.62$  [SD] months) and WT littermate controls ( $n = 11$ ; 8 females and 3 males;  $5.98 \pm 0.75$  months). Middle-age *App*<sup>NL-G-F</sup> cohort: 11–15-month-old *App*<sup>NL-G-F</sup> mice ( $n = 18$ ; 8 females and 10 males; mean age,  $13.25 \pm 1.59$  [SD] months) and WT littermate controls ( $n = 14$ ; 8 females and 6 males;  $13.27 \pm 1.40$  months). Young and middle-age *App*<sup>NL-G-F</sup> cohorts were behaviorally tested in an open arena to assess spontaneous behavior. To assess spatial learning and memory, advanced-age 19–25-month-old *App*<sup>NL-G-F</sup> mice ( $n = 17$ ; 9 females and 8 males; mean age  $21.88 \pm 2.18$  months) and WT littermate controls ( $n = 22$ ; 12 females and 10 males;  $22.42 \pm 1.98$  months) were tested in the Morris water maze. These three cohorts are referred to as 6-, 13- and 22-month-old *App*<sup>NL-G-F</sup> mice, respectively. To assess spontaneous behaviors in the 5xFAD-*Fgg* <sup>$\gamma$ 390-396A</sup> mice, 8–10-month-old 5xFAD mice ( $n = 13$ ; 7 females and 6 males; mean age,  $8.78 \pm 0.89$  months), 5xFAD-*Fgg* <sup>$\gamma$ 390-396A</sup> mice ( $n = 12$ ; 6 females and 6 males;  $9.06 \pm 0.85$  months), *Fgg* <sup>$\gamma$ 390-396A</sup> mice ( $n = 9$ ; 4 females and 5 males;  $9.32 \pm 0.82$  months), and WT mice ( $n = 13$ ; 6 females and 7 males;  $9.14 \pm 1.01$  months) were behaviorally tested in an open arena. Because there were no significant differences in motif usage between *Fgg* <sup>$\gamma$ 390-396A</sup> and WT mice (Figure S4A), these two groups of mice were combined and treated as control group. This cohort is referred to as the 9-month-old 5xFAD-*Fgg* <sup>$\gamma$ 390-396A</sup> mice.

Behavioral experiments were performed in sex-balanced time blocks (morning and afternoon) by investigators who were unaware of the genotype of the mice. Sex effects are reported in the manuscript. All mice were bred and housed at the Gladstone Institutes vivarium, a UCSF AAALAC-approved facility. All mouse experiments were approved by the Institutional Animal Care and Use Committee of the University of California, San Francisco, and were conducted in accordance with the NIH guidelines for the care and use of laboratory animals.

## METHOD DETAILS

### Behavioral experiments

#### Spontaneous behavior in an open arena

Mice were handled and habituated to the chamber and experimenter before testing. Naturalistic behavior was recorded in an acrylic cylindrical chamber (12 inches in diameter and 14 inches high) for 25 (*App*<sup>NL-G-F</sup> cohort) or 60 (5xFAD-*Fgg*<sup>γ390-396A</sup> cohort) minutes during the day, avoiding the first and last 2 h of the lights-on period. Male or female mice were tested in alternating blocks to ensure balanced experimental groups throughout the day. The chambers were cleaned with 70% ethanol between trials.

#### Morris water maze

Twenty-two-month-old *App*<sup>NL-G-F</sup> mice ( $n = 17$ ; 9 females and 8 males) and WT littermate controls ( $n = 22$ ; 12 females and 10 males) were tested in the Morris water maze. As previously described,<sup>27</sup> the water maze pool (122 cm diameter) contained opaque white water with a 14-cm<sup>2</sup> platform submerged 1 cm below the surface. Distinct visual cues were placed on the four walls surrounding the maze. For the hidden platform training, mice were trained to locate the hidden platform over 10 sessions (two sessions per day, 3–4 h apart), each with two trials (60 s each, 15 min apart). The platform location remained constant, and entry points were changed semi-randomly between trials. A 24-hour-probe trial of 60 s (platform removed) was performed after the last day of hidden platform training. For the visible platform test, mice were trained to locate a platform marked with a visual cue (black and green pole; 2 × 15 cm) on top of the platform over 3 sessions (each consisting of two trials, 15 min apart). The visible platform was in the same location used in hidden platform training. Latency, speed, and distance were monitored with an EthoVisionXT video-tracking system (Noldus Information Technology). Three *App*<sup>NL-G-F</sup> mice were excluded from analysis because they failed to locate the visible platform and had combined latencies and distances for the visible training 3 SD above the mean of WT mice.

#### Videography for machine-learning analysis

Mice exploring an open circular chamber were recorded from below in RGB color (*App*<sup>NL-G-F</sup> cohort) or monochrome (5xFAD-*Fgg*<sup>γ390-396A</sup> cohort) video at 25 frames per second with a 1.3 MP GigE camera (acA1300-60gc; Basler) and a varifocal lens (4.5–12.5 mm; Computar) using EthoVisionXT (Noldus Information Technology). Four adjustable light bricks (Panel Go; Lume Cube) lit the chambers from below for homogeneous illumination of the paws and other body features. Videos were converted from mpg to mp4 format before machine-learning analysis.

#### Pose estimation

To capture mouse postural movement, we used DeepLabCut (DLC, version 2.1.8.2),<sup>2</sup> a supervised deep-learning tool which tracks animal posture by using deep convolutional neural networks pretrained on ImageNet.<sup>55</sup> To create supervised annotations for the *App*<sup>NL-G-F</sup> cohort, 10 frames were extracted from each of the 32 videos, and 9 body parts (nose, paws, center, tail base, mid-tail, tail tip) were manually labeled in each extracted frame. The network (ResNet-50) was trained up to 10<sup>6</sup> iterations until the error converged (train error: 1.01 pixels, 0.75 mm; test error: 5.02 pixels, 3.73 mm) and precise virtual skeletons were estimated for all mice. Supervised annotations for the 47 videos of the 5xFAD-*Fgg*<sup>γ390-396A</sup> cohort were created by using the same approach, with 20 frames from 13 representative videos labeled manually (10<sup>6</sup> iterations; train error: 1.44 pixels, 1.06 mm; test error: 5.30 pixels, 3.94 mm). Google Colaboratory (Colab Pro GPU) was used for DLC network training, video analysis, and labeled video creation. DLC pose-tracking results were exported to CSV files for further analysis.

#### Behavioral segmentation and sequencing

To study behavioral manifestations in AD mouse models, we further developed our unsupervised behavioral segmentation and sequencing tool, Variational Animal Motion Embedding (VAME).<sup>7</sup> We used this tool to identify multidimensional clusters of postural motions (motifs) from egocentric spatiotemporal coordinates derived from DeepLabCut pose estimation, thereby permitting calculation of kinematic variables to precisely identify cohesive ethologically-relevant behaviors. For the *App*<sup>NL-G-F</sup> cohort, custom MATLAB code (version 2021b; MathWorks) was used to align mouse coordinates along the midline for every frame. The midline was defined as a vector from the mouse center to the tail-base and the center of the mouse was defined as the origin (0,0). For the 5xFAD-*Fgg*<sup>γ390-396A</sup> cohort, pose data was egocentrically aligned using the VAME alignment module using the nose-tailbase axis.<sup>7</sup> In both cohorts, mid-tail and tail tip coordinates were not included in the VAME training data, as these body parts reflect a time-lagged version of mouse motion. The remaining seven egocentrically-aligned coordinates and associated DLC confidence values for each cohort were used to train separate VAME networks until the test-training loss had converged. We used the standard RNN network as defined as default configuration in the original publication<sup>7</sup> with the default number of latent features (zdims in the config.yaml = 30), time window size in seconds (time\_window in config.yaml = 16 frames or 0.64 s), and number of motif clusters (n\_clusters in the config.yaml = 30). Latent information (30 z-dimensions) was used to identify 30 *k*-means- or HMM-clustered motifs in the training data, and each frame was assigned a motif within a sliding time window of 16 frames (±320 ms). Because *k*-means demonstrated higher probability of correct genotype classification using logistic regression compared to HMM (93.8% vs. 90.6%) (Figure S1A), *k*-means was chosen to identify motifs. DeepLabCut-labeled motif videos and community videos were generated for visual inspection of VAME-identified postural motions. VAME outputs associated with *App*<sup>NL-G-F</sup> mice were exported to

MATLAB and a Jupyter notebook for post-processing. VAME outputs associated with 5xFAD-*Fgg*<sup>γ390–396A</sup> mice were exported to a separate Google Colab notebook for post-processing (see Code Availability for download link).

An independent VAME model was created for each of the two cohorts, to maximize detection of disease-associated alterations observed within these distinct AD mouse models. Therefore, motifs and communities expressed by the respective control/WT animals differ across the two VAME models.

### Uniform manifold approximation (UMAP)

We used the UMAP dimensionality reduction approach (*umap-learn* Python package) to visualize the topology of latent vectors produced by ethoML for *App*<sup>NL–G–F</sup> mice. A random sample of 10,000 frames was selected from the full cohort and segmented into corresponding genotype and sex groups to generate comparison manifolds (male vs. female and *App*<sup>NL–G–F</sup> vs. WT), as well as a UMAP manifold of the respective 30 motifs identified by ethoML for each cohort.

### Motif and community usage

Motif and community use were defined as the total number of frames or seconds a mouse performed a given motif or community during the naturalistic behavior assay. For *App*<sup>NL–G–F</sup> mice, differences in motif use between genotypes relative to WT were assessed in sex-matched normalized data by false discovery rate with Benjamini-Hochberg test for multiple comparisons as described.<sup>56</sup> Significant *App*-affected motifs ( $q < 0.05$ ) were analyzed by two-way ANOVA to assess genotype and sex interactions. Changes in motif and community use over time were assessed by repeated measures one-way ANOVA in 5-min bins. The 5xFAD-*Fgg*<sup>γ390–396A</sup> cohort was analyzed in the same manner. 5xFAD-affected motifs restored in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice ( $q < 0.05$ ) were analyzed by one-way ANOVA and Bonferroni post hoc test for multiple comparisons.

### Motif usage correlation cluster analyses

Correlation cluster analyses were performed on a matrix of motif usages for all mice in a given cohort in order to group motifs that were used to a similar extent, and to rank mice according to their motif usage patterns. We first calculated the Pearson's correlation between pairs of columns containing each subject's motif usage relative to the average total usage of that motif by sex-matched control mice, then subtracted each element of the resultant correlation matrix from 1 to obtain the dissimilarity matrix. The *linkage()* and *dendrogram()* MATLAB functions were used to obtain and visualize a hierarchical cluster tree that grouped motifs with similar usage patterns across mice. Correlations between mice with similar motif usage patterns were found according to the same method. By applying this clustering approach to both dimensions of the motif usage matrix, both motifs and mice with similar motif usage patterns were grouped and ordered according to the out permutation order of the *dendrogram()* function.

### Motif transition matrix

To create the motif transition matrix at the cohort level, we created a 2D histogram of motif-motif transitions observed over all trials. To focus on transitions from one motif to another, the diagonal was set to zero, and each row was normalized to the total number of times a mouse transitioned out of that motif. The resulting asymmetric transition matrix, *T*, was used to determine the motif transition alteration index (TAI) and was used as an input to the hierarchical community dendrogram function.

### Hierarchical community dendrogram

A hierarchical motif-community dendrogram at the cohort level was generated in a Jupyter notebook with the NetworkX Python library. Using the transition probability matrix and motif usage for the full cohort, we implemented a cost function (Equation 1) to iteratively merge clustered nodes with higher transition probabilities relative to each other.

$$\text{cost} = \min \left( \text{usage}_i + \frac{\text{usage}_j}{\text{prob}_{ij} + \text{prob}_{ji}} \right). \quad (\text{Equation 1})$$

To identify behavioral communities of associated motifs, we inspected kinematic and speed trends for each motif according to dendrogram-order. For the *App*<sup>NL–G–F</sup> cohort, communities (A–H) were identified by segmenting the dendrogram at height = 3. Community A had distinct kinematics and genotype effects within the branched clusters and was therefore further segmented into 4 communities (Figures 4A, 4C and S2). For the 5xFAD-*Fgg*<sup>γ390–396A</sup> cohort, communities (A–H) were similarly identified by segmenting at height = 3. Community B had distinct kinematics and genotype effects within the branched clusters and was therefore further segmented into three communities. Communities A and B<sup>I</sup> had similar kinematics and together constituted the Groom-up community. Within community B<sup>III</sup>, motifs 1, 0, and 21 were associated with rearing, whereas motifs 11, 6, and 28 were associated with rapid ambulatory wall approach and rear initiation (Figure S5). These two sets of motifs were differentially color-coded and respectively grouped into the B<sup>II</sup>/B<sup>III</sup> Rear community and the B<sup>III</sup> Walk-up community.

### Transition network analyses and visualization

Transition probabilities between motifs revealed important information about the organization of behavior in healthy and disease-model mice. We used Cytoscape for network visualization of motif transition trends for *App*<sup>NL–G–F</sup> mice. Node size represents the



non-normalized time in a motif across all mice. Directional edges show the highest 20% of transitions; edge width is proportional to transition probability. Significant differences between genotypes (*App*<sup>NL-G-F</sup> vs. WT) are indicated at the node level with red borders (solid, overexpressed; dashed, underexpressed) and at the transition level with edge color (solid red, overexpressed; dashed blue, underexpressed). To identify disease-associated alterations in sequential motif transitions, we used a discrete Markov Chain model to identify altered “self-edges” (i.e., probability of staying in a given state).

$$\text{edge weight} = \frac{\text{transitions}_{ij}}{\text{usage}_i}$$

This definition satisfies a property of the transition probability matrix such that the sum of transition probabilities out of motif *i* is equal to 1. It also allows for estimates of the probability that the mouse will return to the same motif. For example, once the mouse starts exhibiting a motif that could be called an “absorbing” state, it would continue to do so for the rest of the observation period; in this case, the transition probability of returning to this motif would be exactly 1.

5xFAD and 5xFAD-*Fgg*<sup>γ390–396A</sup> mice were analyzed in a similar manner. We found the union of the set of transitions was significantly affected by 5xFAD-affected motifs and significantly restored in 5xFAD-*Fgg*<sup>γ390–396A</sup> mice. We then used Cytoscape to visualize the highest 20% of transitions in the case of control mice vs. 5xFAD mice (Figure 7D, top), and control mice vs. 5xFAD-*Fgg*<sup>γ390–396A</sup> mice (Figure 7D, bottom).

For the transition network statistics, we used the false discovery rate with Benjamin-Hochberg multiple testing adjustment. To estimate the change in the log odds (or the log odds ratio) between the two genotypes of transitioning from motif *i* to motif *j*, we used the generalized linear mixed effects model (implemented in the *glmer* function, with family argument set to binomial, in the lme4 package in R) with a random effect for each mouse.

## Evaluation of order of Markov chain models and pseudo Bayes factors

### Order of Markov chain models

For *App*<sup>NL-G-F</sup> mice, we defined 11 communities (A<sup>I</sup>, A<sup>II</sup>, A<sup>III</sup>, A<sup>IV</sup>, B, C, D, E, F, G, and H) for each frame over the 25-min period of open arena exploration (see community glossary below) and fit Markov chain models of different orders (0, 1, 2, and 3)<sup>30</sup> to determine whether behavior exhibited in the recent past could predict future behavior. We evaluated these models in the context where time was discretized in consecutive 640 ms intervals.

The models were evaluated using a Leave-One-Out Cross Validation (LOOCV) framework to prevent overfitting. The LOOCV framework was implemented separately for WT and *App*<sup>NL-G-F</sup> mice. The parameters (probability or transition probabilities) of the different models were estimated using the observations from all but one mouse and the log likelihood of the sequence of communities exhibited by the held-out mouse was evaluated by using the fitted parameters.

For the order 0 model, the probability of the *i*th mouse exhibiting community *c* at any observed 640-ms interval is estimated as

$$p_c^i = \frac{N_c^i}{N}$$

where *N* denotes the total number of 640 ms intervals over the period of observation and *N*<sub>*c*</sub><sup>*i*</sup> denotes the number of intervals where the *i*th mouse displayed community *c*. This estimate for all mice but the *i*th mouse, *p*<sub>*c*</sub><sup>*i*</sup>, is given by the sample mean of the probability estimates in these mice:

$$p_c^{i'} = \frac{\sum_{j=1, j \neq i}^M p_c^j}{M-1}$$

where *M* is the number of mice in the group. The leave-one-out log likelihood estimate of the order 0 model for the observed sequence of communities in the *i*th mouse is then given by

$$\mathcal{L}_0^i = \sum_{c=1}^{11} N_c^i \cdot \log(p_c^{i'})$$

Transition probabilities underlie the models of order greater than 0. Let *s* denote a state in these models. The communities themselves would be the states for the order 1 model, ordered pairs of communities would encode states for the order 2 model, and ordered triplets of communities would encode states for the order 3 model. The estimate of the transition probability from state *s* to community *c* for the *i*th mouse, *q*<sub>*sc*</sub><sup>*i*</sup> is given by

$$q_{sc}^i = \frac{N_{sc}^i}{N_s^i}$$

where  $N_s^i$  denotes the number of 640 ms intervals mouse  $i$  spent in state  $s$ .  $N_{sc}^i$  denotes the number of times mouse  $i$  transitions to community  $c$  from state  $s$ . This estimate for all but the  $i$ th mouse,  $q_{sc}^i$ , is similarly estimated by the sample mean. The leave-one-out log likelihood estimate of the order  $k$  ( $k > 0$ ) model for the observed sequence of communities in the  $i$ th mouse is then given by

$$\mathcal{L}_k^i = \sum_{s=1}^{S_k} \sum_{c=1}^{11} N_{sc}^i \cdot \log(q_{sc}^i)$$

**Pseudo Bayes factors.** The estimate of the log of pseudo Bayes factors<sup>57</sup> for the Markov chain model comparison between order  $k$  and order  $l$  is

$$\log PBF_{kl} = \sum_{i=1}^M (\mathcal{L}_k^i - \mathcal{L}_l^i)$$

### Delta indices for motif transitions, motif use, and motif speed

To compare the impact of disease and therapeutic interventions on ML behavioral measures at the individual and group levels, we created a set of three summary measures (delta indices) that quantify the magnitude of behavioral changes. For visual comparison across the three delta index types in Figures 5F and 6D, individual subject index scores were normalized to the control group's mean index value for that index type.

The Transition Delta Index (TDI) is a summary measure that quantifies how the motif transition probability matrix differs as a function of genotype group membership. Mathematically, the TDI is the grand sum of the matrix depicting the absolute difference between the transition probability matrix of the  $i$ th subject of a particular group and the transition probability matrix of the baseline control group average  $\bar{C}$  (Figures 5E and 6C). The TDI is normalized by the inverse of number of possible transitions between  $m$  motifs ( $m^2$ ) to ensure invariance to the number of motifs identified in a given experiment.

The TDI can be formulated by the given expression below,

$$TDI = \sum_m \frac{|T_i - T_{\bar{C}}|}{m^2}$$

where,

$m$  is the number of motifs,

$T_i$  denotes the transition probability matrix of the  $i$ th subject of genotype group, and

$T_{\bar{C}}$  denotes the group-average transition probability matrix of the appropriate baseline control group.

The Use Delta Index (UDI) is a summary measure that quantifies how the motif usages are altered across genotype. Mathematically, it is the grand sum of the matrix depicting absolute difference between an array consisting of the motif usages (in percent) of the  $i$ th subject of a particular genotype and the mean motif usage (in percent) across all motifs for all subjects in the Control group  $C$ . This grand sum is further normalized with respect to the total number of motifs identified in a given experiment.

The UDI can be formulated by the given expression below,

$$UDI = \sum_m \frac{|U_i - U_{\bar{C}}|}{m}$$

where,  $m$  is the number of motifs

$U_i$  denotes the of the  $i$ th subject of the genotype group

$U_{\bar{C}}$  denotes the mean motif usage across all motifs for all subjects in the control group.

The Speed Delta Index (SDI) is a summary measure that quantifies how the motif speeds are altered across genotype. Mathematically, it is the grand sum of arrays depicting the absolute difference between an array consisting of the motif speeds of the  $i$ th subject of a particular genotype and the mean motif speed across all motifs for all subjects in the control group. This grand sum is further normalized with respect to the total number of motifs identified in the given experiment.

The SDI can be formulated by the given expression below,

$$SDI = \sum_m \frac{|S_i - S_{\bar{C}}|}{m}$$

Where

$m$  is the number of motifs

$S_i$  denotes the motif speed array of the  $i$ th subject of the genotype group, and

$S_{\bar{C}}$  denotes the mean motif speed across all motifs for all subjects in the Control group.

### Keypoint-MoSeq

To compare VAME with alternative behavioral sequencing platforms, a separate analysis for the middle-age 13-month-old *App<sup>NL-G-F</sup>* mice was conducted on the same videos at the University of Alabama at Birmingham in the laboratory of Dr. Erik Robeson, employing keypoint-MoSeq for the independent identification of pose trajectories (i.e., syllables). The pose trajectories were derived from egocentric coordinates obtained through the DeepLabCut pose estimation, which was previously examined using VAME, as detailed earlier. Adhering to the established procedures outlined in the keypoint-MoSeq repository (<https://github.com/dattalab/keypoint-moseq/>), we executed the standard workflow to predict syllables for each video frame. The supercomputing cluster at the University of Alabama at Birmingham (A100 GPUs) was used for keypoint-MoSeq training and analysis.

For egocentric alignment, the anterior reference point was determined using the nose, while the posterior reference point was identified as the base of the tail. A hyperparameter, kappa, was configured at a value of 10000 to ensure a comparable syllable length to the output generated by VAME. Notably, during the analysis, the middle of the tail and the tail tip were excluded from both training and subsequent analysis.

To create a model in keypoint-MoSeq, pose trajectories were derived from a Principal Component Analysis (PCA) of the egocentrically-aligned key point time series extracted from DeepLabCut. Subsequently, an autoregressive hidden Markov model (AR-HMM) was fitted to the pose trajectory data over 50 iterations. Finally, a complete keypoint-MoSeq model was created using the provided documentation and trained for 500 iterations. Keypoint-MoSeq outputs including individual trial data, cumulative data frames, and statistical summaries were sent to Gladstone for statistical analysis.

For the side-by-side comparison between VAME and keypoint-MoSeq, we created independent models. Thus, the motifs in [Figure S7](#) are different from those in the rest of the manuscript.

### Community glossary

The VAME probabilistic algorithm assigned a motif to each frame by determining the motif with the dominant likelihood given the eight previous and following frames. We therefore note that VAME community videos contain some flexibility in behavioral expression. Glossary definitions represent the most dominant behavior and were determined by observing community videos and kinematics ([Figure S2](#)), which reveal ethologically-relevant behaviors not obvious to human observers. For *App<sup>NL-G-F</sup>* mice, we defined the 11 identified communities as described below:

**A<sup>I</sup>, Groom:** Grooming behavior. Mouse is stationary, sometimes sitting back on its hind legs, and bends neck down then rapidly moves paws over ears and nose with head movement toward the mouse's right ([Video S1](#)).

**A<sup>II</sup>, Rear:** Rearing behavior. Fully extended rearing posture, supported against the chamber wall with head movement to the mouse's left ([Video S2](#)).

**A<sup>III</sup>, Rear:** Rearing behavior. Supported rearing against the chamber wall with body curvature to the mouse's right ([Video S2](#)).

**A<sup>IV</sup>, Rear:** Rearing behavior. Partially to fully extended rearing posture, not supported by the chamber wall. Motifs observed show the initiation of an unsupported rear and subsequent extension to a full rear ([Video S2](#)).

**B, Walk:** Slow locomotor ambulation behavior. Walks across the center of the chamber or circles the chamber perimeter. Motifs observed demonstrate the initiation of slow movement with the head extending outward in no particularly consistent direction ([Video S3](#)).

**C, Run:** Fast locomotor ambulation behavior. Motifs observed show rapid movement along the arena perimeter ([Video S4](#)).

**D, Walk-up:** Exploratory behavior. Slowly taking a few steps forward and sniffing, often with one or two lifted front paws and a slight raise of the torso and head ([Video S5](#)).

**E, Fast-explore:** Exploratory behavior. Moving forward with side-to-side sweeping of the head, neck angled up. Motifs observed demonstrate return from a raised posture and rapid forward ambulation ([Video S6](#)).

**F, Slow-explore:** Exploratory behavior. Slow forward movement interrupted by pauses, sniffing upwards, and neck extensions ([Video S7](#)).

**G, Sit-down:** Sitting on rump with rounded back, with side-to-side nose movement and forward head extension ([Video S8](#)).

**H, Sit-up:** Sitting back on hindlimbs (hindlimbs vertically aligned with center) and resting without full extension into an unsupported rear ([Video S9](#)).

### Histology and immunohistochemistry

*App<sup>NL-G-F</sup>* mice were anesthetized and transcardially flush-perfused with 0.1 M phosphate buffer (PB), and brains were extracted and drop-fixed in 4% phosphate-buffered paraformaldehyde at 4°C for 48 h. After rinsing with PB saline (PBS), brains were transferred to 30% sucrose in PBS at 4°C for 24 h and coronally sectioned with a sliding microtome. Ten subseries of floating sections (30 μm) were collected per mouse and kept at −20°C in cryoprotectant medium until use. Each subseries contained sections throughout the rostrocaudal extent of the forebrain. Brain sections were washed 3 times for 10 min with PBS to remove cryoprotectant medium and once with 0.5% Triton X-100 (PBTx) to permeabilize the tissue. Endogenous peroxidases were blocked with a 15-min incubation with 3% hydrogen peroxide and 10% methanol in PBS. Sections were subsequently washed three times for 10 min each in PBSTx. Nonspecific binding was blocked with a blocking solution containing 10% normal donkey serum (Jackson ImmunoResearch, 017-000-121) and 0.2% gelatin (Sigma-Aldrich, G2500) in PBSTx for 1 h. Brain sections were incubated overnight with biotinylated 82E1 mouse anti-Aβ antibody (IBL, 10326) at 1:250 dilution for staining or rabbit anti-Iba1 antibody

(Wako, 019–19741) in 3% normal donkey serum 0.2% gelatin and PBSTx at 4°C. For anti-Iba1 staining, after washes, brain sections were incubated with biotinylated donkey anti-mouse antibody (Jackson Labs) at 1:500 dilution for 2.5 h at room temperature. Unbound detection antibodies were removed with three washes of 10 min with 0.5% PBSTx and one wash with PBS for 10 min. Iba1 and 82E1 stainings were developed with an avidin-biotin complex (ABC) kit (Vector Laboratories, PK-6100) according to manufacture protocol and sections were incubated for 1 h at room temperature. Brain sections were subsequently washed three times for 10 min each with PBS at room temperature and incubated with diaminobenzidine (DAB; Vector Labs) for colorimetric development. Sections were then washed three times for 10 min each with PBS and mounted in 1.2% gelatin in H<sub>2</sub>O and allowed to dry. After two 5-min washes in xylene, brain sections were permanently mounted on slides and coverslipped for analysis.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Statistical analysis

Statistical analyses and graphs were done with SPSS (28.0.1.1), MATLAB (2018b), or Prism 10. Statistical tests and *p* or *q* values are indicated in the figure or figure legends. Briefly, false discovery rate (FDR) with Benjamin-Hochberg correction for multiple comparisons (*q* values) was used to assess genotype effect on motif or community usage or speeds among all motifs and communities. two-way ANOVA was used to assess sex, genotype, and sex-genotype interaction effects on motif use and log likelihood estimates. Repeated one-way ANOVA was used to assess genotype effects on motif use changes over time. Pseudo Bayes factors was used to assess Markov chain model comparisons of orders and experimental groups. *N* represents number of mice or frames and is indicated in the figure or figure legends. Values are mean ± SEM. Numbers of males and females per experimental group are indicated in figure legends and text. Null hypotheses were rejected by double-tailed tests with an alpha value of 0.05.

Probability of transitions between motifs for the motif transition networks were calculated by a generalized linear mixed-effects model (GLMM) to model the disease-associated differences in the rate of transitions between every pair of identified motifs in the *App*<sup>NL-G-F</sup> and 5xFAD cohorts. Specifically, two numbers representing the number of frames showing a transition from the first motif to the second motif in the evaluated pair of motifs and the total number of frames spent in the first motif (in the pair) were recorded for each mouse in the two cohorts. For the *App*<sup>NL-G-F</sup> cohort, this pair of numbers was modeled by using a GLMM assuming a binomial probability error distribution with a random effect for each mouse and a fixed effect capturing the log odds ratio representing the differences in log odds of transitioning between the first and second motifs for the *App*<sup>NL-G-F</sup> mice versus the WT mice. The control mice in the 5xFAD cohort included both WT and *Fgg*<sup>γ390-396A</sup> mice. Since their motif usage did not differ (Figure S4A), these two groups were combined into a single control group. The random effects for the GLMM included an animal model (control, *Fgg*<sup>γ390-396A</sup>, and 5xFAD-*Fgg*<sup>γ390-396A</sup>) effect, a separate mouse-specific effect sampled hierarchically within the animal model, and a fixed effect capturing the disease-specific log odds ratio. The GLMM was implemented by using the *glmer* function in the lme4<sup>58</sup> package in R.<sup>59</sup>

We performed a binomial (WT vs. *App*<sup>NL-G-F</sup>) or multinomial (control, *Fgg*<sup>γ390-396A</sup>, and 5xFAD-*Fgg*<sup>γ390-396A</sup>) logistic regression classifier analyses to assess the sensitivity and specificity of VAME outcomes (motif use of the 30 motifs; Figures S6G and S6M) and conventional behavioral outcomes in the open field (23 variables including speed, distance, duration in the different arenas; Figures S6A–S6E and S6H–S6L). We used forward stepwise conditional method for variable entry and removal (<0.05) for all mice (no held out group was tested). For WT vs. *App*<sup>NL-G-F</sup>: ML outcomes the model included constant and motifs 6 and 28 (entry order) (correct classification: WT = 92.9% and *App*<sup>NL-G-F</sup> = 83.3%; AIC = 24.652); conventional behavior outcomes model included middle inactive (correct classification: WT = 50.0% and *App*<sup>NL-G-F</sup> = 77.8%; AIC = 37.052). For control, *Fgg*<sup>γ390-396A</sup>, and 5xFAD-*Fgg*<sup>γ390-396A</sup>: ML outcomes the model included intercept and motifs 2, 28, 22, 4, and 11 (entry order) (correct classification: control = 100%, *Fgg*<sup>γ390-396A</sup> = 100%, and 5xFAD-*Fgg*<sup>γ390-396A</sup> = 100%; AIC = 24.002); conventional behavior outcomes model included intercept and middle distance, wall speed, and center inactive (entry order) (correct classification: control = 90.9%, *Fgg*<sup>γ390-396A</sup> = 69.2%, and 5xFAD-*Fgg*<sup>γ390-396A</sup> = 50.0%; AIC = 70.857). For HMM and *k*-means, HMM model included constant and motifs 9, 25, and 10 (entry order) (correct classification: WT = 92.9% and *App*<sup>NL-G-F</sup> = 88.9%; overall, 90.6%; AIC = 20.029); *k*-means model included constant and motifs 17 and 7 (entry order) (correct classification: WT = 94.4% and *App*<sup>NL-G-F</sup> = 92.8%; overall, 93.8%; AIC = 34.749).