



## FULL LENGTH ARTICLE

# Integrated analysis on transcriptome and behaviors defines HTT repeat-dependent network modules in Huntington's disease

Lulin Huang<sup>a,b,\*</sup>, Li Fang<sup>b</sup>, Qian Liu<sup>b</sup>,  
Abolfazl Dostparast Torshizi<sup>b</sup>, Kai Wang<sup>b,\*\*</sup>

<sup>a</sup> *The Key Laboratory for Human Disease Gene Study of Sichuan Province, Department of Clinical Laboratory, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, PR China*

<sup>b</sup> *Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA*

Received 21 January 2021; received in revised form 13 April 2021; accepted 12 May 2021

Available online 9 June 2021

## KEYWORDS

Behaviors;  
CAG repeat;  
Huntington's disease;  
Mice;  
Single-cell RNA sequencing;  
Small chemicals;  
Striatum;  
Transcriptome

**Abstract** Huntington's disease (HD) is caused by a CAG repeat expansion in the huntingtin (HTT) gene. Knock-in mice carrying a CAG repeat-expanded *Htt* will develop HD phenotypes. Previous studies suggested dysregulated molecular networks in a CAG length genotype- and the age-dependent manner in brain tissues from knock-in mice carrying expanded *Htt* CAG repeats. Furthermore, a large-scale phenome analysis defined a behavioral signature for HD genotype in knock-in mice carrying expanded *Htt* CAG repeats. However, an integrated analysis correlating phenotype features with genotypes (CAG repeat expansions) was not conducted previously. In this study, we revealed the landscape of the behavioral features and gene expression correlations based on 445 mRNA samples and 445 microRNA samples, together with behavioral features (396 PhenoCube behaviors and 111 NeuroCube behaviors) in *Htt* CAG-knock-in mice. We identified 37 behavioral features that were significantly associated with CAG repeat length including the number of steps and hind limb stand duration. The behavioral features were associated with several gene coexpression groups involved in neuronal dysfunctions, which were also supported by the single-cell RNA sequencing data in the striatum and the spatial gene expression in the brain. We also identified 15 chemicals with significant responses for genes with enriched behavioral features, most of them are agonist or antagonist for dopamine receptors and serotonin receptors used for neurology/psychiatry. Our study

\* Corresponding author. The Key Laboratory for Human Disease Gene Study of Sichuan Province, Department of Clinical Laboratory, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, PR China.

\*\* Corresponding author.

E-mail addresses: [huangluling@yeah.net](mailto:huangluling@yeah.net) (L. Huang), [kaichop@gmail.com](mailto:kaichop@gmail.com) (K. Wang).

Peer review under responsibility of Chongqing Medical University.

provides further evidence that abnormal neuronal signal transduction in the striatum plays an important role in causing HD-related phenotypic behaviors and provided rich information for the further pharmacotherapeutic intervention possibility for HD.

Copyright © 2021, Chongqing Medical University. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

Huntington's disease (HD) was first described by George Huntington in 1872.<sup>1</sup> It is an autosomal dominant condition that typically presents in middle age with a combination of movement, cognitive, and psychiatric problems, and it leads to death 20–25 years later.<sup>2</sup> Currently, HD has no effective treatment. The disease-causing gene mutation consists of an expanded CAG tract (>35 repeats) within the first exon of the huntingtin (*HTT*) gene, which is translated into a corresponding poly-glutamine stretch (polyQ). Proteins with the expanded repeats make the medium spiny neurons in the striatum particularly vulnerable to cell death and cause the dysfunction and death of neurons in other brain regions. HTT protein is essential for embryogenesis and is ubiquitously expressed in moderate amounts in and outside the nervous system; the mutant HTT protein is toxic and initiates the disease.<sup>3</sup> Although multiple pathologic mechanisms have been proposed, the exact mechanism by which mutant HTT causes neuronal dysfunction is still unknown. Recent studies demonstrated the mRNA transcriptional dysregulation as a central mechanism.<sup>4</sup>

Knock-in mice carrying a CAG repeat-expanded *Htt* have been shown to develop neurological and neurodegenerative phenotypes similar to HD.<sup>5</sup> Using this model, a previous study provided a large-scale, comprehensive transcriptomic characterization of the molecular pathogenic effects of CAG repeat expansion, identifying a consistent set of genes and networks that were dysregulated in HD mouse brains.<sup>6</sup> Patients with early HD do not have impairments in decision making, but those with later HD have decreased attention, perseveration, construction, conceptualization, and memory in comparison with normal people.<sup>7</sup> Another released behavioral feature dataset used a machine learning approach and analyzed the behavioral effects of CAG repeat length in the *Htt* CAG-knock-in mouse model,<sup>8</sup> and it revealed sufficient discriminatory power to accurately predict the genotype-required combined analysis of more than 200 behavioral features. These previous works raised an open question about the gene network associated with the behaviors in HD conditions. Moreover, the shared data with the research community provides support for a focused understanding of the mechanism that links behavioral consequences to a genetic mutation in HD.<sup>8</sup>

Understanding the behavioral features of HD patients requires clinical knowledge within the larger context of a person's life. The interaction between disease mechanisms and behaviors may be complex. In the CAG repeat knock-in mouse model, previous studies only examined the gene expression profiles with the CAG repeat length genotype or the behavioral features with the CAG repeat length

genotype, respectively, thus providing a limited view of understanding the disease mechanism. In this study, we present a combinatorial analysis of brain transcriptomes and behavioral features by using the *Htt* CAG-knock-in mouse model to establish the network of HD behavioral features and gene expressions linked to the CAG repeat length genotype to find the biomarkers and behavior markers for HD.

## Methods

### mRNA sequencing datasets of brain tissues from knock-in mice

As the main affected tissue of HD is the brain, we used the transcriptomic study (i.e., mRNA and microRNA) of the four brain regions (i.e., striatum, cortex, hippocampus, and cerebellum) with an allelic series of CAG repeat lengths HD knock-in mice (Q7, Q20, Q80, Q92, Q111, Q140, and Q175) at three ages (2, 6, and 10 months), from the previous research.<sup>6</sup> In total, the transcriptomes of 445 mRNA samples and 455 microRNA samples were combined and analyzed in this study. These datasets included GSE65770, GSE65774, GSE65775, GSE73468, GSE73503, GSE65769, GSE65773, GSE73505, and GSE73507 from the Gene Expression Omnibus repository (<https://www.ncbi.nlm.nih.gov/gds>). In these datasets, mRNA was extracted and prepared using the Illumina TruSeq RNA sample prep kit and sequenced on an Illumina HiSeq2000 sequencer using strand-specific, paired-end, and 50-mer sequencing protocols to a minimum read depth of 40 million reads per sample.<sup>6</sup> The sequencing was performed in two separate batches (6-month samples in batch 1 and 2, and 10-month samples in batch 2).<sup>6</sup> Raw data were downloaded and aligned to the mouse genome mm9 using the STAR aligner, and read counts for the individual genes were obtained using HTSeq.<sup>6</sup>

### PGI high-throughput behavioral datasets

PGI's comprehensive high-throughput systems (i.e., PhenoCube, NeuroCube, and SmartCube systems) capture the different domains of behavior, namely, cognitive, motor, circadian, social, anxiety-like, and gait, among others, using custom-built computer vision software and machine learning algorithms.<sup>8,9</sup> PhenoCube is a high-throughput platform that assesses the circadian, cognitive, social, and motor behaviors exhibited by group-housed mice.<sup>8</sup> NeuroCube uses computer vision to detect the stance characteristics, gait geometry, and dynamics in rodent

models of disorders.<sup>8</sup> These two platforms detected the behaviors of HD knock-in mice (Q7, Q20, Q80, Q92, Q111, Q140, and Q175) at three ages (2, 6, and 10 months) were used in this study.<sup>8</sup> The behavior datasets were downloaded from the Open Source Science for the HD Research Community (HDinHD: <https://www.hdinhd.org/>). The behavioral features used in this study are listed in Table S1.

## Data preprocessing

We first conducted a quick check of the samples in each tissue using a sample correlation heat map to determine the sample cluster and delete the outgroup sample. This step resulted in 143 samples in the striatum, 104 samples in the cortex, 104 samples in the hippocampus, and 103 samples in the cerebellum for further analysis. The low count was filtered out from the data to keep only the genes that were informative during the differential expression analysis and network construction. We then performed a series of differential expression contrasts using the samples in each tissue to remove the non-differentially expressed genes using the limma R package, which uses raw counts<sup>10</sup> and using the results to further filter out the genes that did not have a significant amount of variance. The properly normalized and preprocessed data were necessary for the downstream analyses of WGCNA. Thus, 17,548 genes in the striatum, 9,291 genes in the cortex, 755 genes in the hippocampus, and 5,802 genes in the cerebellum were obtained for further analysis. The average values of a group of each mouse line of PhenoCube 396 behaviors and 111 NeuroCube behaviors were used.

## Gene coexpression module detection

The WGCNA<sup>11</sup> starts by constructing a matrix of pairwise correlations between all pairs of genes across the measured samples in a dataset.<sup>6</sup> Constructing a weighted gene network entails the choice of the soft thresholding power  $\beta$  to which coexpression similarity is raised to calculate adjacency.<sup>12</sup> We chose the candidate powers for which the scale-free topology index reaches at least 0.80 to perform the analysis of network topology. We calculated the adjacencies using the soft thresholding power of each tissue and time point. To minimize the effects of noise and spurious associations, we transformed the adjacency into a topological overlap matrix and calculated the corresponding dissimilarity. To quantify the coexpression similarity of all the modules, we calculated their eigengenes and clustered them according to their correlation. As we already had a summary profile (eigengenes) for each module, we simply correlated the eigengenes with behavioral features and looked for the most significant associations. We quantified the associations of individual genes with our behavior of interest by defining the gene significance (GS) as the correlation between the gene and the behavior. For each module, we also defined a quantitative measure of module membership (MM) as the correlation of the module eigengene and the gene expression profile. This step enabled us to quantify the similarity of all genes on the array to every module. Using the GS and MM measures, we can identify the genes with a high significance for the significant behavior

and the high module membership in the modules of interest. The main code used in the WGCNA analysis was shown in supplementary file 1.

## Gene functional annotation clustering and visualization

The DAVID functional annotation clustering<sup>13</sup> uses a similar fuzzy clustering concept as the functional classification by measuring the relationships among the annotation terms based on the degree of their co-association with the genes within the users' list to cluster somewhat heterogeneous, yet highly similar, annotation into functional annotations.<sup>14</sup> This process reduces the burden of associating different terms related to similar biological processes, thus enabling the biological interpretation to be more focused on the "biological module" level.<sup>14</sup> This type of grouping of functional annotation can give a more insightful view of the relationships between annotation categories and terms than the traditional linear list of enriched terms. The enriched functional annotation terms associated with the users' gene list were identified and listed according to their enrichment  $P$  value by DAVID. The DAVID-enriched functional annotation tables (FDR < 0.05) served as the input data for the network visualization by the Enrichment Map app<sup>15</sup> of Cytoscape 3.3.<sup>16</sup>

## Gene functional group network analysis and visualization

GeneMANIA searches many large, publicly available biological datasets to find related genes, including protein–protein, protein–DNA and genetic interactions, pathways, reactions, gene and protein expression data, protein domains, and phenotypic screening profiles.<sup>17–19</sup> The gene lists in the category of the DAVID functional annotation table were used for network analysis by GeneMANIA. The results of the visualization were performed by the GeneMANIA app in Cytoscape 3.3.<sup>19</sup>

## Single-cell RNA sequence data cluster analysis

GSE82187 data was downloaded.<sup>20</sup> The genes were extracted from the significant GO terms in the GO network of Fig. 4B, 6C, 7B and 8B. The gene expression values of each cell type were calculated by the mean values in each cell's populations. The gene expression cluster was down by cluster 3.0 by using hierarchical methods.

## Gene small chemical targets

We aimed to identify the potential small chemical targets by searching for behavior-enriched genes for HD in Connectivity Map<sup>21</sup> (CMap dataset, <https://clue.io/>). The CMap dataset of cellular signatures catalogs the transcriptional responses of human cells to chemical and genetic perturbations (1.3 M L1000 profiles and the tools for their analysis). A total of 27,927 perturbations have been profiled to produce 476,251 expression signatures.<sup>22</sup> We searched the chemicals by the gene for interaction with the

existing CMap database. Fisher's test was then used to identify the significant small chemicals. The *P* values of the Fisher's test were adjusted by Bonferroni correction. An FDR of less than 0.05 was considered significant.

## Results

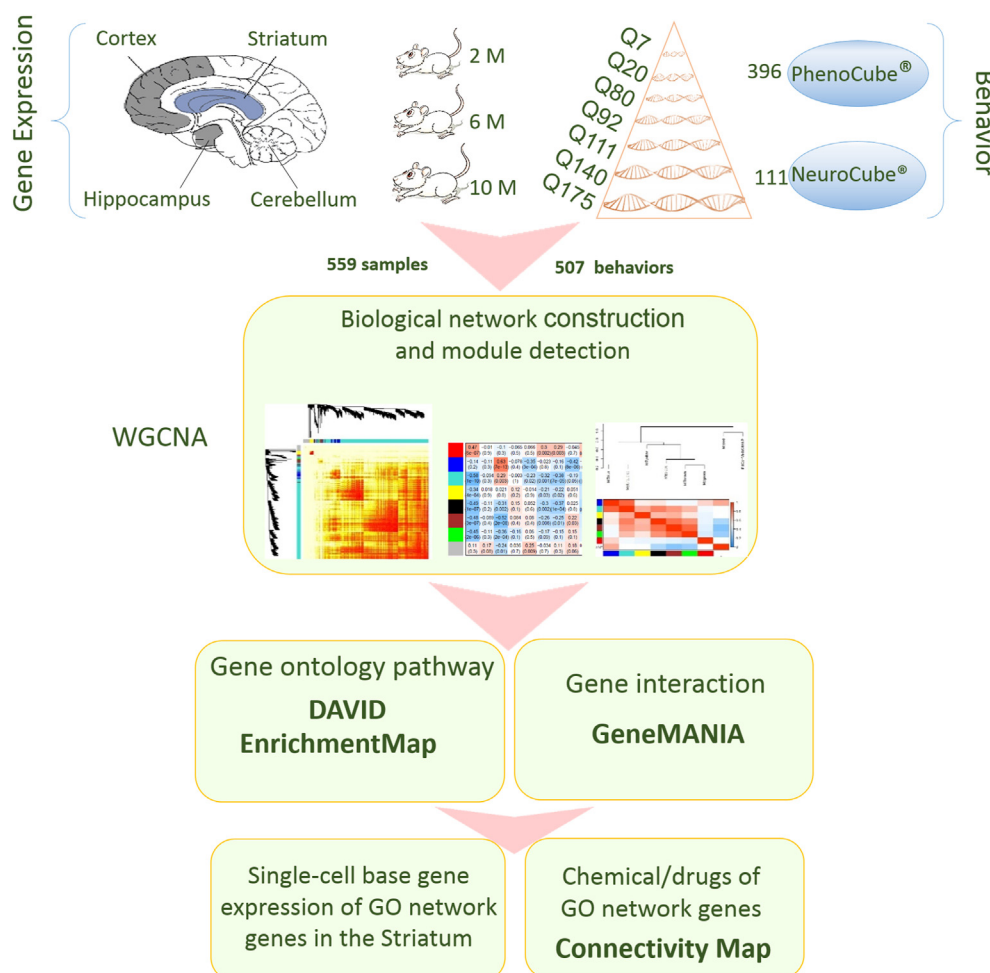
### Study design

As the main affected tissue in HD is the brain, we used the published transcriptomic data of CAG repeat knock-in HD mice (mRNA and microRNA) of the four brain regions, namely, striatum, cortex, hippocampus, and cerebellum, for data analysis. For each tissue, the data contain an allelic series of CAG repeat lengths (Q7, Q20, Q80, Q92, Q111, Q140 and Q175) at three ages (2, 6 and 10 months). For the behavior data, we used the PhenoCube and NeuroCube platform-generated data from mouse lines matched to the transcriptomic data. PhenoCube is a high-throughput platform that assesses circadian, cognitive, and motor behaviors, whereas NeuroCube<sup>®</sup> is a platform that uses computer vision to detect stance characteristics, gait

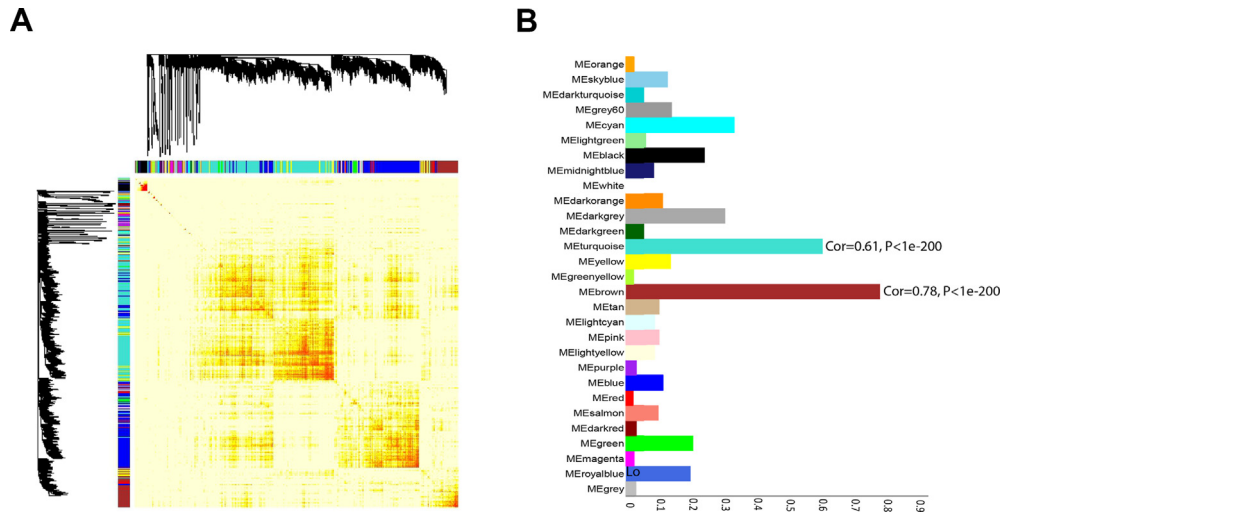
geometry, and dynamics in the HD mouse model. The transcriptome data of the 445 mRNA samples, 445 microRNA samples, 396 PhenoCube behaviors, and 111 NeuroCube behaviors were combined and analyzed in this study (Fig. 1). These samples are expected to provide crucial systems-level insight into CAG length-dependent networks linked to behaviors in HD. We used a consensus weighted gene coexpression network analysis (WGCNA)<sup>23</sup> with the behavioral features to define the modules of the genes related to the CAG length genotype (Fig. 1). Afterward, a gene ontology network analysis was performed to address the disease mechanisms (Fig. 1). Besides, single-cell-based expression patterns of ontology network genes in the most affected tissue striatum were analyzed to reveal the most affected cell type in the striatum. Finally, we predicted small molecules targeted to behavior-enriched genes.

### The landscape of the behavioral features and gene expression correlations

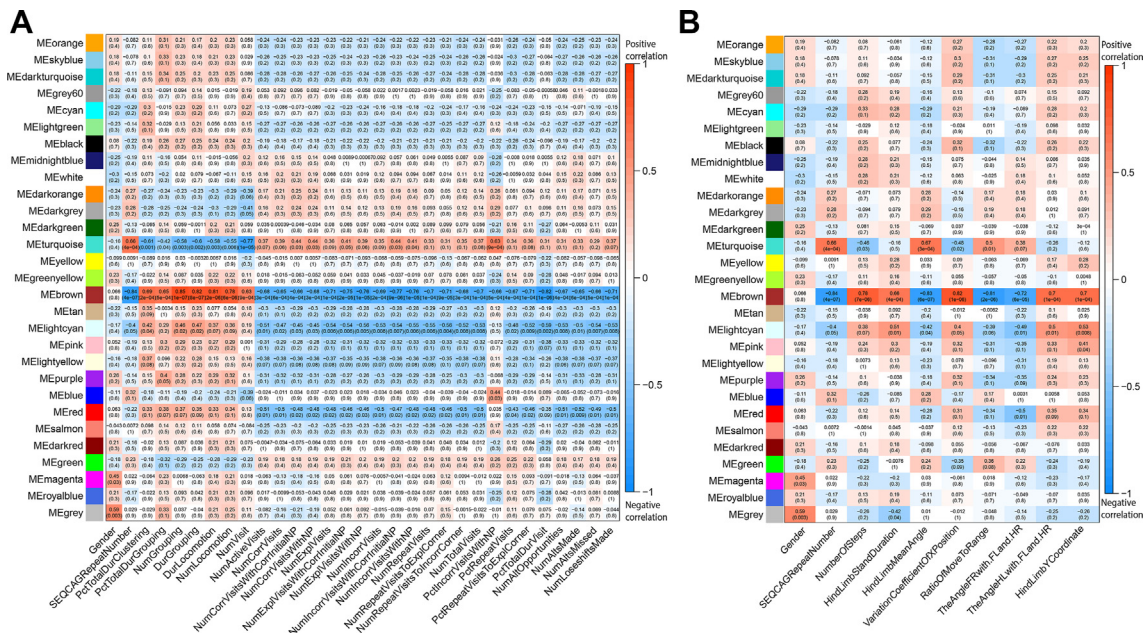
We first used WGCNA to analyze all the datasets to give an overview of our data. The results demonstrated that the



**Figure 1** Overview of the experimental design and data analysis strategy. An allelic series of CAG repeat lengths of the transcriptome data of 445 mRNA samples, 445 microRNA samples, 396 PhenoCube behaviors, 111 NeuroCube behaviors were analyzed by consensus weighted gene coexpression network analysis (WGCNA), followed by gene ontology network analysis, to address the disease mechanisms in the WGCNA detected modules.



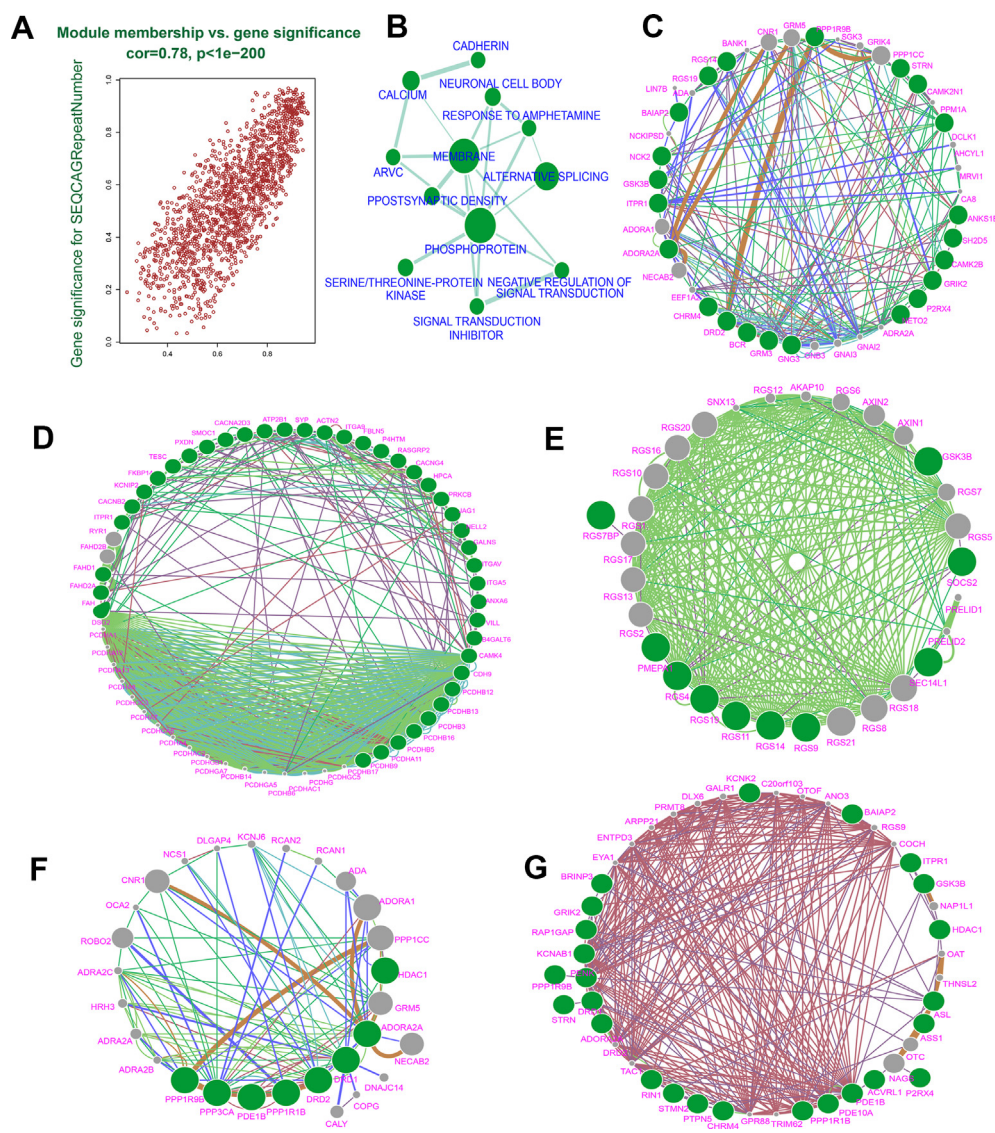
**Figure 2** Integrated effect of the gene expression and behavioral features in the striatum at 10 months. **(A)** The gene co-expression network and **(B)** bar plot of the module significance defined as the mean gene significance across all genes in the module. MEbrown and METurquoise are the two modules with a correlation of more than 0.6 detected in the striatum at 10 months.



**Figure 3** Module-behavioral feature associations in the striatum at 10 months of the PhenoCube behavior **(A)** and NeuroCube behavior **(B)**. Each row corresponds to a module eigengene and each column to a behavioral feature. Each cell contains the corresponding correlation and *P*-value. The full names of the behaviors are listed in Table S1.

batch and tissue effects played the main role and that some behavior sets have a similar pattern (Fig. S1). We then separated the batch (tissues) and combined the similar behavior phenotypes (see methods) for the module detection. The results suggested that the batch and age effect played a larger effect than the CAG repeat length (Fig. S2–S5). To detect the real network linked to the CAG repeat length and behavior, we separated the dataset in each tissue and each time point for behavioral analysis. The separate analysis provided a reasonable pattern for the

CAG repeat length-associated behaviors. In general, the striatum, which is HD’s most affected tissue, showed the strongest correlation of gene expression changes and behavioral changes in both PhenoCube and NeuroCube. For the time course, the correlation at 10 months was the most significant time point. The WGNA module analyses provided the strength and significance of the associations between the modules and the behavioral features according to the CAG repeat length. Among the four brain tissues, namely, the cortex, hippocampus, cerebellum, and

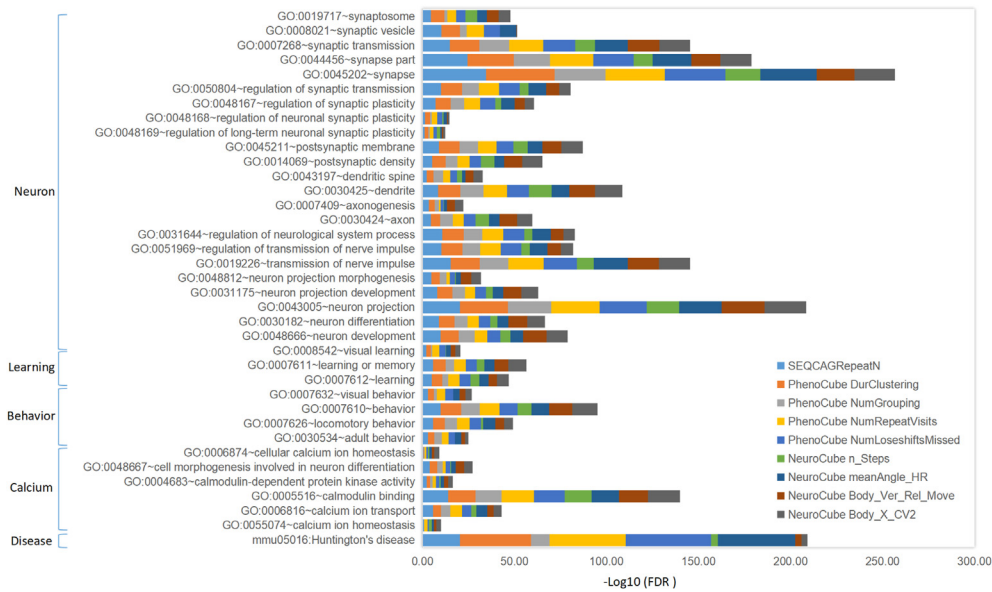


**Figure 4** Networks shared by the CAG repeat length and behaviors. **(A)** A scatter plot of the gene significance for CAG repeat length vs. module membership in the brown module. **(B)** Gene ontology networks for the CAG repeat length. **(C–G)** Gene networks of the significant ontologies for the CAG repeat length. Blue lines, physical interactions. Brown lines, co-localization. Green lines, pathway. Dark pink lines, coexpression. Light green lines, shared protein domains. **(C)** Postsynaptic density (GO: 0014069), **(D)** Calcium, **(E)** Signal transduction inhibitor function, **(F)** Amphetamine ontology, and **(G)** Neuronal cell body ontology (GO: 0043025).

striatum, we detected the strongest association of gene expression profile and behavioral features in the striatum, consistent with the previous studies.<sup>6</sup> At 10 months, two modules were detected in the striatum tissue associated with the CAG repeat length and behaviors, with a correlation score of more than 0.6 (Fig. S6). We also detected a weak association between the CAG repeat length and behavior phenotype in the cerebellum (Fig. S7). No significant behaviors were found in each of the four tissues during the seven CAG repeat lengths at the 2-month and 6-month time points. However, at 6 months, the significant behaviors driven by gender difference were detected in the cortex, striatum, and cerebellum. Especially in the cortex, most of the PhenoCube and NeuroCube behaviors were significantly correlated with gender differences (Fig. S8, S9).

### Behavioral features and CAG repeat length-dependent gene coexpression network in the striatum

Our results suggested several special behavioral changes in HD condition linked to the CAG repeat length controlling gene expression profiles in the striatum at 10 months. The most significant module is the brown module, which was shared by the CAG repeat length and behavioral features (Fig. 2). In the striatum at the 10-month time point, we detected 29 significant ( $P < 7.25 \times 10^{-4}$ , 0.05/69 behavioral features for Bonferroni correction) PhenoCube behaviors associated with CAG length (Fig. 3A, Table S1). The grouping and clustering of the behaviors were negatively correlated with the CAG repeat length. This result suggests that these behavioral features decreased according to the



**Figure 5** The graphical representation of the DAVID FDR of the enriched functional terms for gene modules that are associated with representative PhenoCube and NeuroCube behaviors.

CAG repeat length increase in the HD condition (Fig. 3A). On the contrary, most of the visit activities showed the same direction of correlation with the CAG repeat length, thus indicating that these behaviors increased according to the CAG repeat length expansion in the HD condition (Fig. 3A). We also found eight significant ( $P < 4.63 \times 10^{-4}$ , 0.05/108 behavioral features) NeuroCube behaviors associated with the CAG repeat length in one or more modules. These behaviors presented the stance characteristics, gait geometry, and dynamics (Fig. 3B, Table S1). The behavioral features, including the number of steps and hind limb stand duration, among others, were negatively related to the CAG repeat length, whereas the other behavioral features, including the hind limb, mean angle, among others, were positively related to the CAG repeat length (Fig. 3B).

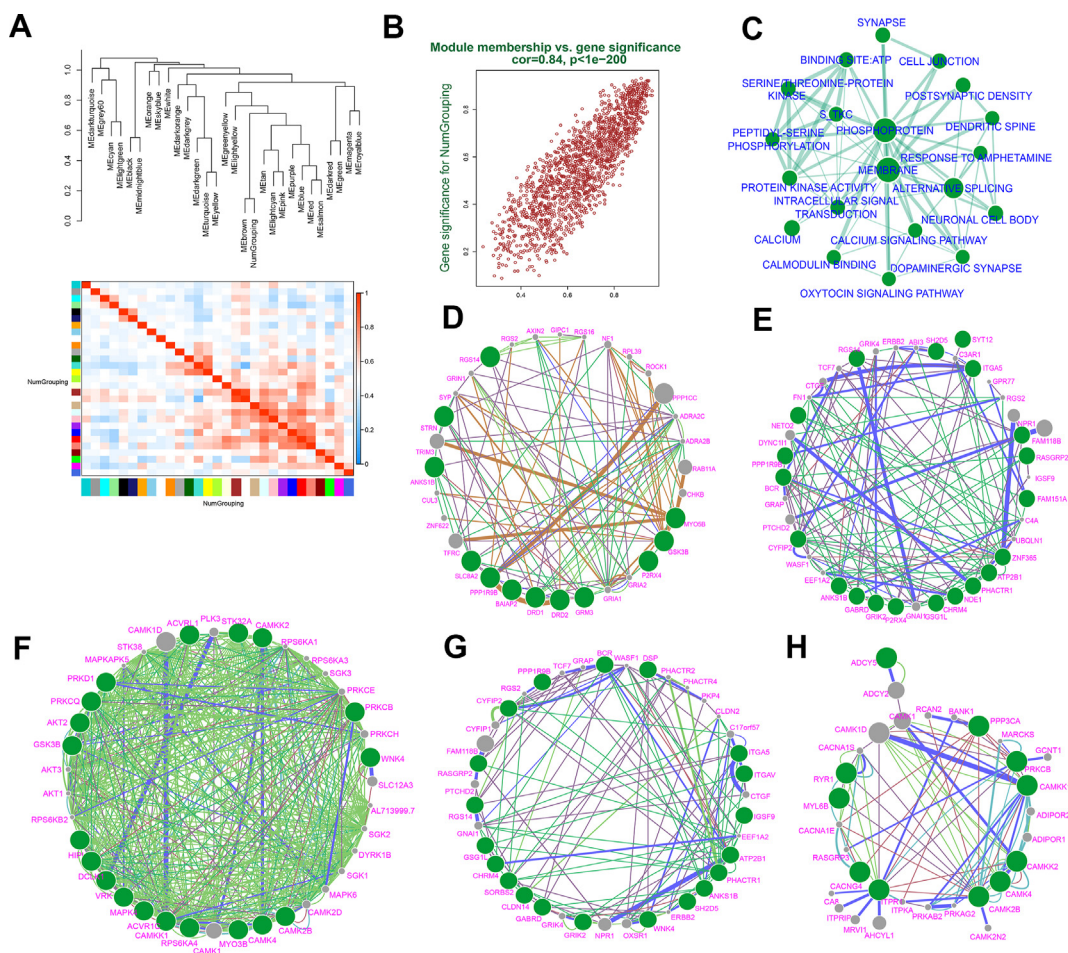
**Highlighted networks shared by the CAG repeat length and behaviors**

As most of the significant genes were shared by the CAG repeat length and significant behavioral features, we initially focused on the CAG repeat length-modified genes. In total, 1111 genes were significantly altered according to the CAG repeat length (false discovery rate,  $FDR < 0.05$ ). A highly significant correlation was found between the gene significance for the CAG repeat length versus the brown module membership detected by WGCNA (Fig. 4A). This gene set highlighted 12 significant functional groups by the DAVID analysis ( $FDR < 0.05$ ) (Fig. 4B).<sup>14</sup> The basic striatal functional proteins involved in phosphoprotein, membrane, and alternative splicing were the largest gene groups enriched according to the CAG repeat lengths and significant behaviors (Fig. 4B). Postsynaptic density (GO: 0014069) was the most significant ontology enriched

( $FDR = 8.33 \times 10^{-7}$ ) by the DAVID annotation. This result is consistent with those of previous studies in which postsynaptic density was reduced in the HD condition (Fig. 4C).<sup>24–28</sup> The next significant ontology is the calcium term ( $FDR = 4.08 \times 10^{-6}$ ). This result is also consistent with the previous finding that neuronal calcium signaling is abnormal in HD.<sup>29,30</sup> A set of 43 genes was enriched in this ontology (Fig. 4D). This ontology was followed by the signal transduction inhibitor function ( $FDR = 4.80 \times 10^{-5}$ ) with 10 genes enriched (Fig. 4E). As expected, the response to amphetamine ontology (GO: 0001975) was also enriched ( $FDR = 9.12 \times 10^{-3}$ ) (Fig. 4F). The neuronal cell body ontology (GO: 0043025) gene set enriched 28 genes ( $FDR = 1.62 \times 10^{-2}$ ), and most of them were co-expressed as markers of the neuronal cell body (Fig. 4G).<sup>31</sup>

**Highlighted networks in the motor and visit behaviors in the PhenoCube platform**

The brown module showed that the PhenoCube-detected motor behavioral features, namely, grouping, clustering, and locomotion, were positively correlated with gene expression. The most significant behavior was the number of grouping (number of short bouts together,  $P = 1 \times 10^{-7}$ ). A total of 725 genes were significantly altered in the PhenoCube feature ( $FDR < 0.05$ ). The summary of the DAVID FDR of the enriched functional terms for gene modules that are associated with representative PhenoCube behaviors was shown in Fig. 5 and Table S2. Fig. 6A illustrates the visualization of the eigengene network representing the relationships between the modules and the number of grouping. Fig. 6B presents the scatter plot of the gene significance for the number of grouping versus the module



**Figure 6** Highlighted networks in the grouping behaviors in the PhenoCube platform. **(A)** Visualization of the eigengene network representing the relationships between the modules and the number of groupings. The upper panel shows a hierarchical clustering dendrogram of the eigengenes in which the dissimilarity of eigengenes  $E_i$  and  $E_j$  is given by  $1 - \text{cor}(E_i; E_j)$ . The heat map in the lower panel shows the eigengene adjacency  $A_{ij} = (1 + \text{cor}(E_i; E_j))/2$ . **(B)** A scatter plot of the gene significance for number of groupings vs. module membership in the brown module. **(C)** Gene ontology networks for number of groupings. **(D–H)** Gene networks of the significant ontologies for the number of grouping. Blue lines, physical interactions. Brown lines, co-localization. Green lines, pathway. Dark pink lines, coexpression. Light green lines, shared protein domains. **(D)** Dendritic spine, **(E)** Synapse, **(F)** S\_TKc domain, **(G)** Cell junction, and **(H)** Oxytocin signaling pathway.

membership in the brown module. This gene set highlighted 20 significant functional groups through the DAVID analysis ( $FDR < 0.05$ ) (Fig. 6C). Besides the functional ontologies detected in the CAG repeat previously, the other five functional groups were enriched in this behavior. The dendritic spine functional group ( $FDR = 4.03 \times 10^{-4}$ ) contained 13 genes, as shown in Fig. 6D. This result is consistent with those of previous reports in which the dendritic spine was found to be unstable and reduced in the HD condition.<sup>32–34</sup> The synapse functional group ( $FDR = 6.31 \times 10^{-3}$ ), which included 21 genes, was also enriched for motor behavior (Fig. 6E). This result is consistent with those of previous reports in which HD was found to be partially caused by abnormal synaptic transmission.<sup>35,36</sup> The S\_TKc domain functional group with 20 genes was also enriched ( $FDR = 0.017$ ) (Fig. 6F). The cell junction function group ( $FDR = 0.02$ ), which contained 22

genes, was also enriched (Fig. 6G). This result is consistent with those of previous reports in which the signal transmission was reduced in the HD condition.<sup>37</sup> Eleven oxytocin signaling pathway genes were enriched in this study ( $FDR = 0.03$ ) (Fig. 6H). This result is consistent with those of previous reports in which oxytocin was found to affect HD patients.<sup>37,38</sup>

A set of visit behaviors, such as the NumExplVitsWithCorrInitialNP behavior (number of visits to non-reinforced corners with a first pseudocorrect nose poke), was also associated with gene expression ( $P = 9 \times 10^{-6}$ ) (Fig. 3A). A total of 263 genes contributing to 13 functional terms were significantly regulated in the NumExplVitsWithCorrInitialNP behavior. The significant gene functional annotation terms were similar to the number of grouping behavior, but the correlation pattern was the opposite. The significant ontologies were postsynaptic



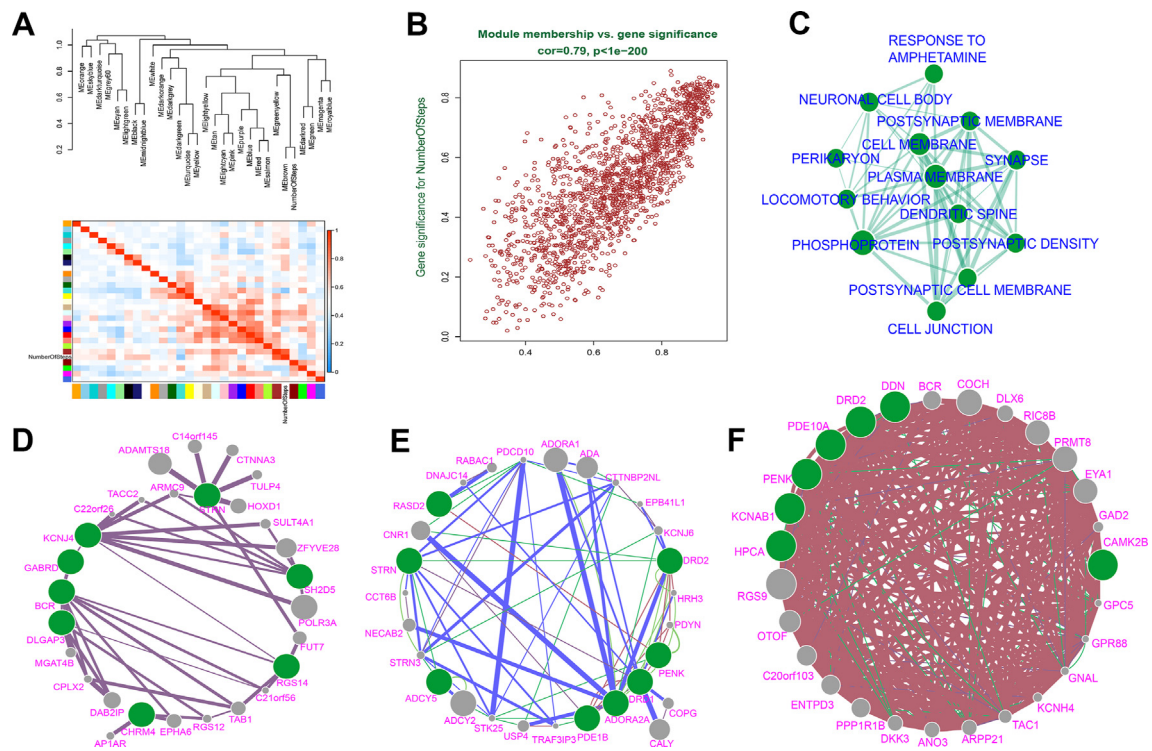
density (FDR =  $1.49 \times 10^{-8}$ ), calcium (FDR =  $4.05 \times 10^{-5}$ ), and dendritic spine (FDR =  $3.08 \times 10^{-4}$ ), among others.

### Highlighted networks in the gait characteristics in the NeuroCube platform

The gait characteristics are known to be abnormal in HD. <sup>39,40</sup> In the NeuroCube platform, we detected eight gait characteristics that were significantly changed in the brown module (Fig. 3B). The summary of the DAVID FDR of the enriched functional terms for gene modules that are associated with representative NeuroCube behaviors was shown in Fig. 5. The expression of the gene set was positively correlated with the number of steps feature (FDR =  $7 \times 10^{-6}$ ); 371 genes were significantly altered in this behavior (FDR < 0.05). Along with the CAG repeat length, the number of steps should be lower in the HD condition than in the normal one. Fig. 7A shows the visualization of the eigengene network, which represents the relationships between the modules and the number of steps. Fig. 7B illustrates the scatter plot of the gene significance for the number of steps versus the module membership in the brown module. Twelve functional ontologies were significantly enriched in the number of steps behavior (Fig. 7C). Except for the ontologies previously mentioned, namely, the response to amphetamine

(FDR =  $3.97 \times 10^{-5}$ ), postsynaptic density (FDR =  $5.52 \times 10^{-5}$ ), phosphoprotein (FDR =  $7.26 \times 10^{-4}$ ), synapse (FDR =  $2.55 \times 10^{-3}$ ), plasma membrane (FDR =  $2.79 \times 10^{-3}$ ), neuronal cell body (FDR =  $4.69 \times 10^{-3}$ ), cell junction (FDR =  $1.41 \times 10^{-2}$ ), and dendritic spine (FDR =  $1.45 \times 10^{-2}$ ), the other three ontologies were specifically highlighted in this feature. The postsynaptic membrane was enriched in the number of steps feature (FDR =  $2.82 \times 10^{-3}$ ) (Fig. 7D). This result is consistent with the previous results that the postsynaptic membrane is a dysfunction in the HD condition. <sup>41-43</sup> A set of genes with the locomotor behavioral feature was also enriched (FDR =  $3.03 \times 10^{-3}$ ) (Fig. 7E). Most of these genes are connected by physical interaction. The remaining gene ontology is perikaryon, with eight genes assigned in this ontology (FDR = 0.016). These genes tightly interacted mainly by co-localization (Fig. 7F).

The other behavioral characteristic gene sets, including the imaging features, body motion, and rhythmicity, were also significantly associated with the HD behavior state. Another significant behavioral feature is the HindLimbMeanAngle (hind limb mean angle of orientation of the maximal diameter to the direction of the run [X]) (FDR =  $6 \times 10^{-7}$ , Fig. 3B). Fig. 8A shows the scatter plot of the gene signature of the number of hind limb angles versus the module membership in the brown module. A total of



**Figure 7** Highlighted networks in gait characteristics and number of steps in the NeuroCube platform. (A) Visualization of the eigengene network representing the relationships between the modules and the number of steps. The upper panel shows a hierarchical clustering dendrogram of the eigengenes in which the dissimilarity of eigengenes EI and EJ is given by  $1 - \text{cor}(EI; EJ)$ . The heat map in the lower panel shows the eigengene adjacency  $A_{IJ} = (1 + \text{cor}(EI; EJ))/2$ . (B) A scatter plot of the gene significance for number of steps vs. module membership in the brown module. (C) Gene ontology networks for number of steps. (D–F) Gene networks of the significant ontologies for number of steps. Blue lines, physical interactions. Brown lines, co-localization. Green lines, pathway. Dark pink lines, coexpression. Light green lines, shared protein domains. (D) Postsynaptic membrane, (E) Locomotor, and (F) Perikaryon.

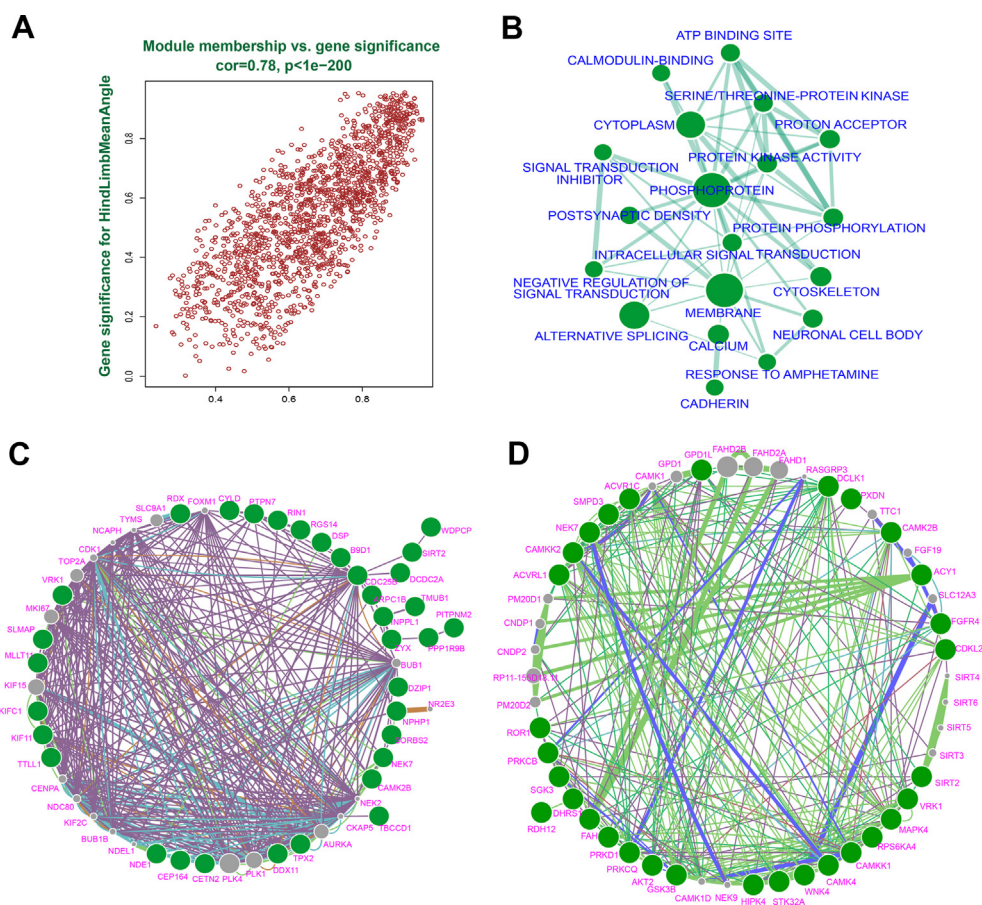
1087 genes, consisting of 19 gene ontologies (Fig. 8B), were significantly altered in this behavioral feature (FDR < 0.05). This behavior shared the most ontologies with other behaviors, such as phosphoprotein (FDR =  $8.47 \times 10^{-8}$ ), postsynaptic density (FDR =  $1.83 \times 10^{-6}$ ), and calcium (FDR =  $2.80 \times 10^{-6}$ ), among others (Fig. 8B). Two specific ontologies were enriched in this behavior. The first one is cytoskeleton ontology (FDR =  $3.12 \times 10^{-3}$ ) (Fig. 8C). This result is consistent with those of previous reports in which cytoskeleton abnormalities were found in the HD condition.<sup>44,45</sup> The second one is the proton acceptor ontology (FDR = 0.03), in which 32 genes were detected (Fig. 8D).

### Single cell-based RNA expression patterns and spatial expression patterns of the HTT repeat-dependent behavior genes

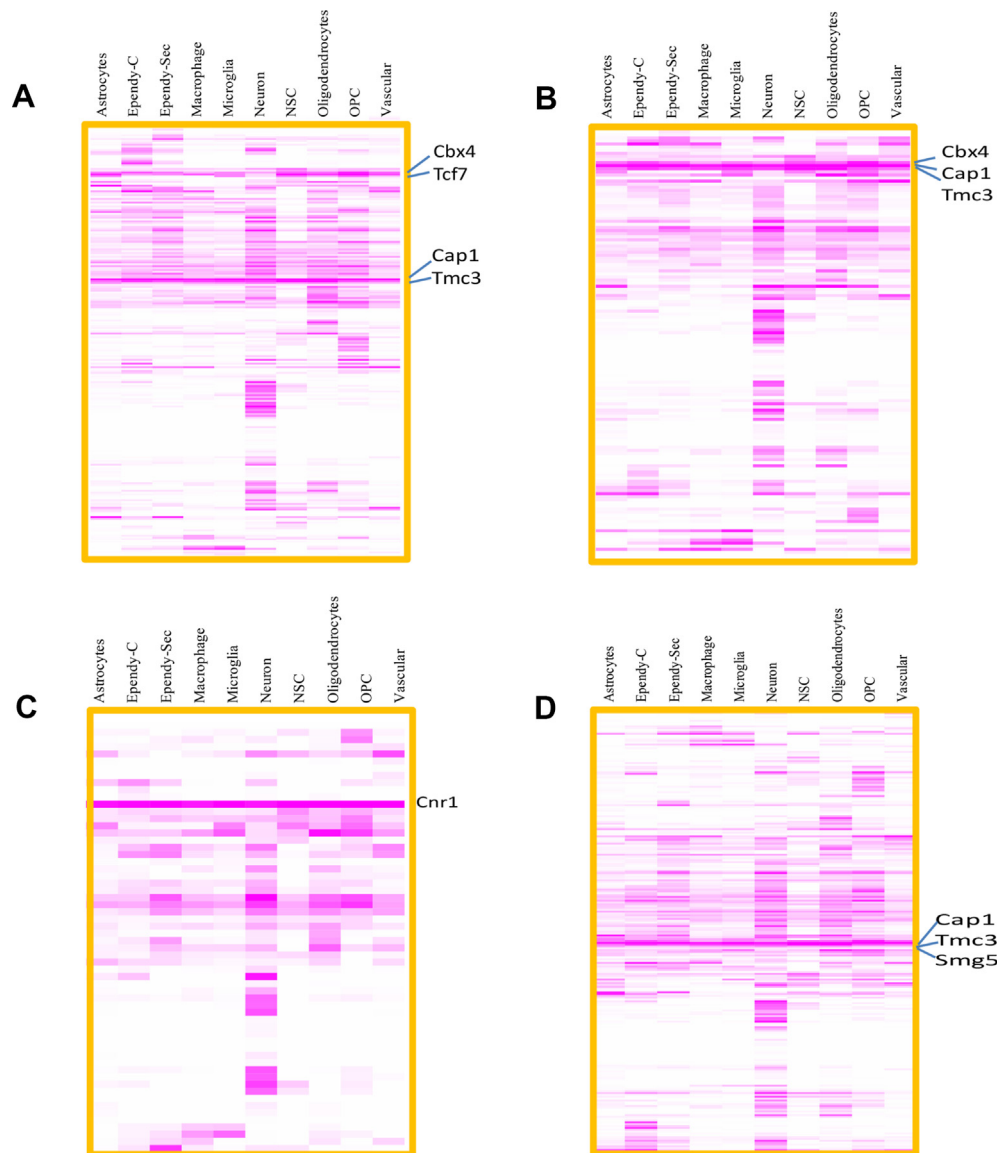
To investigate the expression patterns of the HTT repeat-dependent behavior genes in the striatum cell populations, we downloaded the mouse striatum single-Cell RNA-Seq data.<sup>20</sup> We searched the cell population expression patterns of genes in the GO network of CAG repeat length (Fig. 9A), number of grouping (Fig. 9B), number of

steps (Fig. 9C), and hind limb angle (Fig. 9D). All the ten cell populations in the striatum can detect some gene expression, with the strongest expression in neuron cells (Fig. 9). These results suggested that HD patients' neurons are highly impacted in the striatum. Other cell types were also impacted to a different degree. *Cbx4*, *Tcf7*, *Cap1*, and *Tmc3* were highly expressed in all the ten cell populations in GO network of CAG repeat length genes (Fig. 9A). *Cbx4*, *Cap1*, and *Tmc3* were also involved in the behavior number of grouping (Fig. 9B). *Cap1* and *Tmc3* were also involved in behavior hind limb angle (Fig. 9D). In the behavior number of steps, only *Cnr1* was highly expressed in all the ten cell populations (Fig. 9C).

To validate the gene expression patterns in brain space, we selected some of the enriched functional terms for gene modules that are associated with representative PhenoCube and NeuroCube behaviors. We looked up the genes' spatial expression patterns involved in behavior, locomotor behavior, learning or memory, dendritic spine, postsynaptic density in C57BL/6 mice and *App*<sup>NL-G-F</sup> mice (a model for Alzheimer's diseases, AD) (Fig. S10) (<https://alzmap.org/>). These genes showed dynamic expression in corresponding brain regions with age (3 months vs. 18 months). Comparing to the wild type, these genes generally down-



**Figure 8** Highlighted networks in the gait characteristics of the hind limb angle (hind limb mean angle of orientation of the maximal diameter to the direction of the run [X]) in the NeuroCube platform. (A) A scatter plot of the gene significance for number of hind limb angles vs. module membership in the brown module. (B) Gene ontology networks for the hind limb angle. (C, D) Gene networks of the significant ontologies for the hind limb angle. Blue lines, physical interactions. Brown lines, co-localization. Green lines, pathway. Dark pink lines, coexpression. Light green lines, shared protein domains. (C) Cytoskeleton and (D) Proton acceptor.



**Figure 9** Single-cell-based RNA expression patterns of the HTT repeat-dependent behavior genes in the Striatum cell population. The cell population expression patterns of genes in the network of CAG repeat length (A), number of grouping (B), number of steps (C), and hind limb angle (D). Ependy-C: Ependy cilia; NSCs: neuronal stem cells; OPCs: oligodendrocyte precursor cells.

regulated their expression in the other typical neurodegenerative disease—AD, which is consistent with the expression patterns observed in the HD in our previous analysis.

### Enriched small molecule response to behaviors defines the HTT repeat-dependent genes

We aimed to identify the interactions between genes and small molecules (using the Connectivity Map<sup>21</sup>; <https://clue.io/repurposing-app>) in HD based on the behavior-enriched genes to provide the potential targets for pharmacotherapeutic intervention. The 15 significant small molecules are listed in Table 1, most of them are dopamine receptor agonist or antagonist, serotonin receptor agonist or antagonist. Staurosporine, an inhibitor for CDK, CHK, and

PKC, was reported to be an inducer of apoptosis for HD cells.<sup>46</sup> At the present stage, Chlorprothixene is used as a dopamine receptor antagonist for schizophrenia and bipolar disorder. Naltrexone is used as an opioid receptor antagonist for abstinence from alcohol. Levomepromazine is used as a dopamine receptor antagonist for psychosis, schizophrenia, bipolar disorder, nausea, and insomnia. Pramipexole is a clinically effective non-ergot dopamine agonist and interacts with dopamine D2 subfamily receptors, namely, the D2, D3, and D4 receptor subtypes.<sup>47</sup> It has a positive effect on HD.<sup>47</sup> Aripiprazole was well tolerated and remarkably improved some of the motor and behavioral symptoms in patients affected by HD.<sup>48</sup> Ziprasidone, as a dopamine receptor antagonist and serotonin receptor antagonist, was improved several categories of the motor function of HD.<sup>49</sup> Nortriptyline is a tricyclic antidepressant, is used for depression. Dantrolene was neuroprotective in

**Table 1** Fifteen significant chemicals for HD based on genes with enriched behavior features.

Name	P	FDR	ATC code	Target	Mechanism of action <sup>a</sup>	Disease area <sup>a</sup>
Staurosporine	$1.46 \times 10^{-8}$	$3.44 \times 10^{-6}$		CDK2, GSK3B, CAMK2B, CDK1, CDK5, CHEK1, CHRM1, CHRM2, CHRM4, CSK, DAPK1, GPR 35, IKBKB, ITK, LCK, LRRK2, MAP2K4, MAP2K6, MAPKAPK2, PAK2, PDPK1, PHKG2, PIK3CG, PIM1, PKN1, PRKACB, PRKCI, PRKCQ, RPS6KA1, STK3, SYK, TNIK, ZAP70	CDK inhibitor, CHK inhibitor, PKC inhibitor	undefined
Quinpirole	$4.58 \times 10^{-5}$	$4.95 \times 10^{-3}$		DRD2, DRD3, DRD4, DRD1, HTR1A, HTR2A, HTR2B, HTR2C	dopamine receptor agonist	undefined
Chlorprothixene	$7.92 \times 10^{-5}$	$5.33 \times 10^{-3}$	N05AF03	DRD2, CHRM1, CHRM2, CHRM3, CHRM4, CHRM5, DRD1, DRD3, HRH1, HTR2A, HTR2B, HTR2C	dopamine receptor antagonist	schizophrenia (neurology/psychiatry) bipolar disorder (neurology/psychiatry)
Boldine	$6.3 \times 10^{-5}$	$4.95 \times 10^{-3}$		CHRNA4, CHRN2, DRD1, DRD2		
Naltrexone	$6.3 \times 10^{-5}$	$4.95 \times 10^{-3}$	N07BB04	OPRK1, OPRM1, OPRD1, SIGMAR1	opioid receptor antagonist	abstinence from alcohol (neurology/psychiatry)
Levomepromazine	$2.44 \times 10^{-4}$	$6.16 \times 10^{-3}$	N05AA02	ADRA1A, ADRA1B, ADRA1D, ADRA2A, ADRA2B, ADRA2C, CHRM1, CHRM2, CHRM3, CHRM4, CHRM5, DRD1, DRD2, DRD3, DRD4, DRD5, HRH1, HTR2A, HTR2C	dopamine receptor antagonist	psychosis (neurology/psychiatry) schizophrenia (neurology/psychiatry) bipolar disorder (neurology/psychiatry) nausea (gastroenterology) insomnia (neurology/psychiatry)
Minaprine	$2.59 \times 10^{-4}$	$6.16 \times 10^{-3}$	N06AX07	HTR2B, SLC6A4, ACHE, CHRM1, DRD1, DRD2, HTR2A, HTR2C, MAOA	serotonin reuptake inhibitor	undefined
Pramipexole	$3 \times 10^{-4}$	$6.16 \times 10^{-3}$	N04BC05	DRD3, DRD2, ADRA2A, ADRA2B, ADRA2C, DRD1, DRD4, DRD5, HTR1A, HTR1B, HTR1D, HTR2A, HTR2B, HTR2C	dopamine receptor agonist	Parkinson's Disease (neurology/psychiatry)
Ropinirole	$3 \times 10^{-4}$	$6.16 \times 10^{-3}$	N04BC04	DRD2, DRD3, ADRA2A, ADRA2B, ADRA2C, DRD1, DRD4, DRD5, HTR1A, HTR1B, HTR1D, HTR2A, HTR2B, HTR2C	dopamine receptor agonist	Parkinson's Disease (neurology/psychiatry) restless leg syndrome (neurology/psychiatry)
SCH- 23390	$2.61 \times 10^{-4}$	$6.16 \times 10^{-3}$		DRD1, DRD5, HTR2C, KCNJ4, KCNJ6	dopamine receptor antagonist	undefined
Aripiprazole	$3.79 \times 10^{-4}$	$7.14 \times 10^{-3}$	N05AX12, other antipsychotics, antipsychotics	DRD2, HTR1A, HTR2A, HRH1, HTR1B, HTR1D, HTR2C, ADRA1A, ADRA1B, ADRA2A, ADRA 2B, ADRA2C, CHRM1, CHRM2, CHRM3, CHRM4, CHRM5, DRD1, DRD3, DRD4, DRD5, HTR1E, HTR3A, HTR6, HTR7	serotonin receptor agonist serotonin receptor antagonist	depression (neurology/psychiatry) schizophrenia (neurology/psychiatry) bipolar disorder

Ziprasidone	$3.79 \times 10^{-4}$	$7.14 \times 10^{-3}$	N05AE04, indole derivatives antipsychotics	DRD2, HTR2A, HTR1A, HTR1D, HRH1, HTR1B, HTR1E, HTR2C, HTR7, ADRA1A, ADRA1B, ADRA2A, ADRA2B, ADRA2C, CHRM1, CHRM2, CHRM3, CHRM4, CHRM5, DRD1, DRD3, DRD4, DRD5, HTR3A, HTR6, SLC6A4	dopamine receptor antagonist serotonin receptor antagonist NA	(neurology/psychiatry) schizophrenia (neurology/psychiatry) bipolar disorder (neurology/psychiatry)
Romazine	$5.29 \times 10^{-4}$	$9.51 \times 10^{-3}$	N05AA03	CHRM5, DRD2, ADRA1A, ADRA1B, ADRA1D, CHRM1, CHRM2, CHRM3, CHRM4, DRD1, DRD3, DRD4, HRH1, HTR2A, HTR2C	undefined	
Nortriptyline	$5.45 \times 10^{-4}$	$9.51 \times 10^{-3}$	N06AA10, non-selective monoamine reuptake inhibitors, antidepressants	KCNJ10, SLC6A2, SLC6A4, ADRA1A, ADRA1B, ADRA1D, ADRA2A, ADRA2B, ADRA2C, ADRA2E, ADRA2F, ADRA2G, ADRA2H, ADRA2I, ADRA2J, ADRA2K, ADRA2L, ADRA2M, ADRA2N, ADRA2O, ADRA2P, ADRA2Q, ADRA2R, ADRA2S, ADRA2T, ADRA2U, ADRA2V, ADRA2W, ADRA2X, ADRA2Y, ADRA2Z, ADRA2AA, ADRA2AB, ADRA2AC, ADRA2AD, ADRA2AE, ADRA2AF, ADRA2AG, ADRA2AH, ADRA2AI, ADRA2AJ, ADRA2AK, ADRA2AL, ADRA2AM, ADRA2AN, ADRA2AO, ADRA2AP, ADRA2AQ, ADRA2AR, ADRA2AS, ADRA2AT, ADRA2AU, ADRA2AV, ADRA2AW, ADRA2AX, ADRA2AY, ADRA2AZ, ADRA2BA, ADRA2BB, ADRA2BC, ADRA2BD, ADRA2BE, ADRA2BF, ADRA2BG, ADRA2BH, ADRA2BI, ADRA2BJ, ADRA2BK, ADRA2BL, ADRA2BM, ADRA2BN, ADRA2BO, ADRA2BP, ADRA2BQ, ADRA2BR, ADRA2BS, ADRA2BT, ADRA2BU, ADRA2BV, ADRA2BW, ADRA2BX, ADRA2BY, ADRA2BZ, ADRA2CA, ADRA2CB, ADRA2CC, ADRA2CD, ADRA2CE, ADRA2CF, ADRA2CG, ADRA2CH, ADRA2CI, ADRA2CJ, ADRA2CK, ADRA2CL, ADRA2CM, ADRA2CN, ADRA2CO, ADRA2CP, ADRA2CQ, ADRA2CR, ADRA2CS, ADRA2CT, ADRA2CU, ADRA2CV, ADRA2CW, ADRA2CX, ADRA2CY, ADRA2CZ, ADRA2DA, ADRA2DB, ADRA2DC, ADRA2DD, ADRA2DE, ADRA2DF, ADRA2DG, ADRA2DH, ADRA2DI, ADRA2DJ, ADRA2DK, ADRA2DL, ADRA2DM, ADRA2DN, ADRA2DO, ADRA2DP, ADRA2DQ, ADRA2DR, ADRA2DS, ADRA2DT, ADRA2DU, ADRA2DV, ADRA2DW, ADRA2DX, ADRA2DY, ADRA2DZ, ADRA2EA, ADRA2EB, ADRA2EC, ADRA2ED, ADRA2EE, ADRA2EF, ADRA2EG, ADRA2EH, ADRA2EI, ADRA2EJ, ADRA2EK, ADRA2EL, ADRA2EM, ADRA2EN, ADRA2EO, ADRA2EP, ADRA2EQ, ADRA2ER, ADRA2ES, ADRA2ET, ADRA2EU, ADRA2EV, ADRA2EW, ADRA2EX, ADRA2EY, ADRA2EZ, ADRA2FA, ADRA2FB, ADRA2FC, ADRA2FD, ADRA2FE, ADRA2FF, ADRA2FG, ADRA2FH, ADRA2FI, ADRA2FJ, ADRA2FK, ADRA2FL, ADRA2FM, ADRA2FN, ADRA2FO, ADRA2FP, ADRA2FQ, ADRA2FR, ADRA2FS, ADRA2FT, ADRA2FU, ADRA2FV, ADRA2FW, ADRA2FX, ADRA2FY, ADRA2FZ, ADRA2GA, ADRA2GB, ADRA2GC, ADRA2GD, ADRA2GE, ADRA2GF, ADRA2GG, ADRA2GH, ADRA2GI, ADRA2GJ, ADRA2GK, ADRA2GL, ADRA2GM, ADRA2GN, ADRA2GO, ADRA2GP, ADRA2GQ, ADRA2GR, ADRA2GS, ADRA2GT, ADRA2GU, ADRA2GV, ADRA2GW, ADRA2GX, ADRA2GY, ADRA2GZ, ADRA2HA, ADRA2HB, ADRA2HC, ADRA2HD, ADRA2HE, ADRA2HF, ADRA2HG, ADRA2HH, ADRA2HI, ADRA2HJ, ADRA2HK, ADRA2HL, ADRA2HM, ADRA2HN, ADRA2HO, ADRA2HP, ADRA2HQ, ADRA2HR, ADRA2HS, ADRA2HT, ADRA2HU, ADRA2HV, ADRA2HW, ADRA2HX, ADRA2HY, ADRA2HZ, ADRA2IA, ADRA2IB, ADRA2IC, ADRA2ID, ADRA2IE, ADRA2IF, ADRA2IG, ADRA2IH, ADRA2II, ADRA2IJ, ADRA2IK, ADRA2IL, ADRA2IM, ADRA2IN, ADRA2IO, ADRA2IP, ADRA2IQ, ADRA2IR, ADRA2IS, ADRA2IT, ADRA2IU, ADRA2IV, ADRA2IW, ADRA2IX, ADRA2IY, ADRA2IZ, ADRA2JA, ADRA2JB, ADRA2JC, ADRA2JD, ADRA2JE, ADRA2JF, ADRA2JG, ADRA2JH, ADRA2JI, ADRA2JJ, ADRA2JK, ADRA2JL, ADRA2JM, ADRA2JN, ADRA2JO, ADRA2JP, ADRA2JQ, ADRA2JR, ADRA2JS, ADRA2JT, ADRA2JU, ADRA2JV, ADRA2JW, ADRA2JX, ADRA2JY, ADRA2JZ, ADRA2KA, ADRA2KB, ADRA2KC, ADRA2KD, ADRA2KE, ADRA2KF, ADRA2KG, ADRA2KH, ADRA2KI, ADRA2KJ, ADRA2KK, ADRA2KL, ADRA2KM, ADRA2KN, ADRA2KO, ADRA2KP, ADRA2KQ, ADRA2KR, ADRA2KS, ADRA2KT, ADRA2KU, ADRA2KV, ADRA2KW, ADRA2KX, ADRA2KY, ADRA2KZ, ADRA2LA, ADRA2LB, ADRA2LC, ADRA2LD, ADRA2LE, ADRA2LF, ADRA2LG, ADRA2LH, ADRA2LI, ADRA2LJ, ADRA2LK, ADRA2LL, ADRA2LM, ADRA2LN, ADRA2LO, ADRA2LP, ADRA2LQ, ADRA2LR, ADRA2LS, ADRA2LT, ADRA2LU, ADRA2LV, ADRA2LW, ADRA2LX, ADRA2LY, ADRA2LZ, ADRA2MA, ADRA2MB, ADRA2MC, ADRA2MD, ADRA2ME, ADRA2MF, ADRA2MG, ADRA2MH, ADRA2MI, ADRA2MJ, ADRA2MK, ADRA2ML, ADRA2MM, ADRA2MN, ADRA2MO, ADRA2MP, ADRA2MQ, ADRA2MR, ADRA2MS, ADRA2MT, ADRA2MU, ADRA2MV, ADRA2MW, ADRA2MX, ADRA2MY, ADRA2MZ, ADRA2NA, ADRA2NB, ADRA2NC, ADRA2ND, ADRA2NE, ADRA2NF, ADRA2NG, ADRA2NH, ADRA2NI, ADRA2NJ, ADRA2NK, ADRA2NL, ADRA2NM, ADRA2NN, ADRA2NO, ADRA2NP, ADRA2NQ, ADRA2NR, ADRA2NS, ADRA2NT, ADRA2NU, ADRA2NV, ADRA2NW, ADRA2NX, ADRA2NY, ADRA2NZ, ADRA2OA, ADRA2OB, ADRA2OC, ADRA2OD, ADRA2OE, ADRA2OF, ADRA2OG, ADRA2OH, ADRA2OI, ADRA2OJ, ADRA2OK, ADRA2OL, ADRA2OM, ADRA2ON, ADRA2OO, ADRA2OP, ADRA2OQ, ADRA2OR, ADRA2OS, ADRA2OT, ADRA2OU, ADRA2OV, ADRA2OW, ADRA2OX, ADRA2OY, ADRA2OZ, ADRA2PA, ADRA2PB, ADRA2PC, ADRA2PD, ADRA2PE, ADRA2PF, ADRA2PG, ADRA2PH, ADRA2PI, ADRA2PJ, ADRA2PK, ADRA2PL, ADRA2PM, ADRA2PN, ADRA2PO, ADRA2PP, ADRA2PQ, ADRA2PR, ADRA2PS, ADRA2PT, ADRA2PU, ADRA2PV, ADRA2PW, ADRA2PX, ADRA2PY, ADRA2PZ, ADRA2QA, ADRA2QB, ADRA2QC, ADRA2QD, ADRA2QE, ADRA2QF, ADRA2QG, ADRA2QH, ADRA2QI, ADRA2QJ, ADRA2QK, ADRA2QL, ADRA2QM, ADRA2QN, ADRA2QO, ADRA2QP, ADRA2QQ, ADRA2QR, ADRA2QS, ADRA2QT, ADRA2QU, ADRA2QV, ADRA2QW, ADRA2QX, ADRA2QY, ADRA2QZ, ADRA2RA, ADRA2RB, ADRA2RC, ADRA2RD, ADRA2RE, ADRA2RF, ADRA2RG, ADRA2RH, ADRA2RI, ADRA2RJ, ADRA2RK, ADRA2RL, ADRA2RM, ADRA2RN, ADRA2RO, ADRA2RP, ADRA2RQ, ADRA2RR, ADRA2RS, ADRA2RT, ADRA2RU, ADRA2RV, ADRA2RW, ADRA2RX, ADRA2RY, ADRA2RZ, ADRA2SA, ADRA2SB, ADRA2SC, ADRA2SD, ADRA2SE, ADRA2SF, ADRA2SG, ADRA2SH, ADRA2SI, ADRA2SJ, ADRA2SK, ADRA2SL, ADRA2SM, ADRA2SN, ADRA2SO, ADRA2SP, ADRA2SQ, ADRA2SR, ADRA2SS, ADRA2ST, ADRA2SU, ADRA2SV, ADRA2SW, ADRA2SX, ADRA2SY, ADRA2SZ, ADRA2TA, ADRA2TB, ADRA2TC, ADRA2TD, ADRA2TE, ADRA2TF, ADRA2TG, ADRA2TH, ADRA2TI, ADRA2TJ, ADRA2TK, ADRA2TL, ADRA2TM, ADRA2TN, ADRA2TO, ADRA2TP, ADRA2TQ, ADRA2TR, ADRA2TS, ADRA2TT, ADRA2TU, ADRA2TV, ADRA2TW, ADRA2TX, ADRA2TY, ADRA2TZ, ADRA2UA, ADRA2UB, ADRA2UC, ADRA2UD, ADRA2UE, ADRA2UF, ADRA2UG, ADRA2UH, ADRA2UI, ADRA2UJ, ADRA2UK, ADRA2UL, ADRA2UM, ADRA2UN, ADRA2UO, ADRA2UP, ADRA2UQ, ADRA2UR, ADRA2US, ADRA2UT, ADRA2UU, ADRA2UV, ADRA2UW, ADRA2UX, ADRA2UY, ADRA2UZ, ADRA2VA, ADRA2VB, ADRA2VC, ADRA2VD, ADRA2VE, ADRA2VF, ADRA2VG, ADRA2VH, ADRA2VI, ADRA2VJ, ADRA2VK, ADRA2VL, ADRA2VM, ADRA2VN, ADRA2VO, ADRA2VP, ADRA2VQ, ADRA2VR, ADRA2VS, ADRA2VT, ADRA2VU, ADRA2VV, ADRA2VW, ADRA2VX, ADRA2VY, ADRA2VZ, ADRA2WA, ADRA2WB, ADRA2WC, ADRA2WD, ADRA2WE, ADRA2WF, ADRA2WG, ADRA2WH, ADRA2WI, ADRA2WJ, ADRA2WK, ADRA2WL, ADRA2WM, ADRA2WN, ADRA2WO, ADRA2WP, ADRA2WQ, ADRA2WR, ADRA2WS, ADRA2WT, ADRA2WU, ADRA2WV, ADRA2WW, ADRA2WX, ADRA2WY, ADRA2WZ, ADRA2XA, ADRA2XB, ADRA2XC, ADRA2XD, ADRA2XE, ADRA2XF, ADRA2XG, ADRA2XH, ADRA2XI, ADRA2XJ, ADRA2XK, ADRA2XL, ADRA2XM, ADRA2XN, ADRA2XO, ADRA2XP, ADRA2XQ, ADRA2XR, ADRA2XS, ADRA2XT, ADRA2XU, ADRA2XV, ADRA2XW, ADRA2XX, ADRA2XY, ADRA2XZ, ADRA2YA, ADRA2YB, ADRA2YC, ADRA2YD, ADRA2YE, ADRA2YF, ADRA2YG, ADRA2YH, ADRA2YI, ADRA2YJ, ADRA2YK, ADRA2YL, ADRA2YM, ADRA2YN, ADRA2YO, ADRA2YP, ADRA2YQ, ADRA2YR, ADRA2YS, ADRA2YT, ADRA2YU, ADRA2YV, ADRA2YW, ADRA2YX, ADRA2YY, ADRA2YZ, ADRA2ZA, ADRA2ZB, ADRA2ZC, ADRA2ZD, ADRA2ZE, ADRA2ZF, ADRA2ZG, ADRA2ZH, ADRA2ZI, ADRA2ZJ, ADRA2ZK, ADRA2ZL, ADRA2ZM, ADRA2ZN, ADRA2ZO, ADRA2ZP, ADRA2ZQ, ADRA2ZR, ADRA2ZS, ADRA2ZT, ADRA2ZU, ADRA2ZV, ADRA2ZW, ADRA2ZX, ADRA2ZY, ADRA2ZZ	tricyclic antidepressant	depression (neurology/psychiatry)
Dantrolene	$3.01 \times 10^{-4}$	$6.16 \times 10^{-3}$	M03CA01, dantrolene and derivatives, muscle relaxants, directly acting agents	DRD2, HRH1, HTR1A, HTR2A, HTR2C, HTR6, PGRMC1, PIK3CD, SIGMAR1, RYR1, RYR3	calcium channel blocker	spasms (neurology/psychiatry)malignant hyperthermia (MH) (endocrinology)

<sup>a</sup> <https://clue.io/repurposing-app>.

the HD transgenic mouse model.<sup>50</sup> The interaction analysis of the small molecules and behavior-enriched genes of HD provided information for the further pharmacotherapeutic intervention possibility for HD.

## Discussion

To the best of our knowledge, this analysis is the first systematic and integrated one on a gene coexpression network corresponding to HD behaviors in mice models. In this study, we applied an extensive association analysis of genome-wide transcriptomic characterization with behaviors using an allelic series of the *Htt* CAG-knock-in mouse model to elucidate the CAG length-dependent molecular networks in disease-relevant behaviors. In this study, we used a consensus WGCNA on 445 mRNA samples and 445 microRNA samples along with behavioral features (396 PhenoCube behaviors and 111 NeuroCube behaviors) to define the gene modules that correlate with CAG repeat-dependent behaviors in *Htt* CAG-knock-in mice. We identified 37 behavioral features that were significantly associated with the CAG repeat length, and these behavioral features were associated with several gene coexpression groups involved in neuronal dysfunctions. We also identified the significant responses of 15 chemicals in behavior-enriched HD genes. Our study provides further evidence that the neuronal signal transduction abnormal in the striatum plays an important role in causing HD-related phenotypic behaviors.

The neuronal dysfunction module is considered the molecular system most strongly associated with the pathophysiology of HD.<sup>51</sup> The gene expression in the post-synaptic density, calmodulin-binding, neuronal cell body, synapse, perikaryon, and dendrite functional groups was modulated and linked to behaviors detected by the PhenoCube and NeuroCube platforms. These results dominated the dysfunction of neurons in the HD condition, consistent with the results of previous reports that neuronal death is the hallmark of HD.<sup>52,53</sup> These results were also supported by the single-cell RNA sequencing data in the striatum which suggested many genes in the network were highly expressed in the neurons. Moreover, although most of the functional ontologies were reported in HD, about half of the genes in our networks were newly identified to be associated with HD.

The prominent behavioral changes in HD patients are learning and memory, perception, executive functions, apathy, organization, impulsivity, frustration, irritability, anger, strategies, denial and unawareness, perseveration, depression, anxiety, psychosis, and sleep disturbances, among others.<sup>54,55</sup> In the mouse HD model, we found that grouping, clustering, and locomotion behaviors were significantly reduced. Many functional ontology classes were found to correspond to these behaviors. The postural and gait patterns in HD patients were abnormal, and the walking speed was reduced in HD patients in comparison with normal individuals.<sup>56–58</sup> Previous findings showed compensation production in the lower limbs in HD patients.<sup>40–59</sup> We identified eight HD mouse behaviors similar to the postural and gait patterns of the lower limb compensation production of human HD patients. Our results provided the dominant behaviors to test drug compounds

and narrowed down the behavioral features from thousands to about 40 by using the CAG-expansion HD mouse model for drug screening.

In this study, we also analyzed the miRNA data, but no miRNA markers were found to be significantly associated with the CAG repeat length and behaviors. We also examined the association of transcriptome data and 62 Smart-Cube platform behaviors<sup>60</sup> of 2 months (no 6- and 10-month behavior data were available) and found no associated behaviors in this dataset mainly because of the early age.

In summary, our study provided a large-scale, comprehensive transcriptomic and behavioral characterization of the molecular pathogenic effects of the *Htt* CAG-knock-in mouse model. We provided integrative genomic evidence to show that converging molecular networks linked to behavioral features are perturbed in HD mice. Taken together, our integrative findings capturing the social and activity behaviors with dysregulated genes and ontologies provide rich information for the understanding of HD.

## Author contributions

K. W and L.H. designed the study. L.H., L.F., Q.L., and A.D. performed the statistical analysis. L.H. and K.W wrote the manuscript. All authors critically revised and provided the final approval for this manuscript.

## Conflict of interests

The authors declare no competing financial interests related to this paper.

## Funding

This research was supported by the National Natural Science Foundation of China (No. 81970839 and 81670895 to L.H.) and the Department of Science and Technology of Sichuan Province, China (No. 2021YFS0033 to L.H.). We would like to thank the Wang lab members for their assistance in the R package codes and their helpful comments.

## Acknowledgements

We greatly appreciate the authors of the original studies for making the transcriptome data and the phenotype data on knock-in mice publicly available, thus making it feasible for us to perform the current study.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gendis.2021.05.004>.

## References

- Bates GP, Dorsey R, Gusella JF, et al. Huntington disease. *Nat Rev Dis Primers*. 2015;1:15005.
- Barker RA, Mason SL. Neurodegenerative disease: mapping the natural history of Huntington disease. *Nat Rev Neurol*. 2014;10(1):12–13.
- Cattaneo E, Zuccato C, Tartari M. Normal huntingtin function: an alternative approach to Huntington's disease. *Nat Rev Neurosci*. 2005;6(12):919–930.
- Sadri-Vakili G, Cha JH. Mechanisms of disease: histone modifications in Huntington's disease. *Nat Clin Pract Neurol*. 2006;2(6):330–338.
- Pouladi MA, Morton AJ, Hayden MR. Choosing an animal model for the study of Huntington's disease. *Nat Rev Neurosci*. 2013;14(10):708–721.
- Langfelder P, Cattle JP, Chatzopoulou D, et al. Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nat Neurosci*. 2016;19(4):623–633.
- Gleichgerricht E, Ibáñez A, Roca M, Torralva T, Manes F. Decision-making cognition in neurodegenerative diseases. *Nat Rev Neurol*. 2010;6(11):611–623.
- Alexandrov V, Brunner D, Menalled LB, et al. Large-scale phenome analysis defines a behavioral signature for Huntington's disease genotype in mice. *Nat Biotechnol*. 2016;34(8):838–844.
- Alexandrov V, Brunner D, Hanania T, Leahy E. High-throughput analysis of behavior for drug discovery. *Eur J Pharmacol*. 2015;750:82–89.
- Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:e559.
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17.
- Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*. 2007;8(9):R183.
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
- Isserlin R, Merico D, Voisin V, Bader GD. Enrichment Map - a Cytoscape app to visualize and explore OMICS pathway enrichment results. *F1000Res*. 2014;3:141.
- Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504.
- Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol*. 2008;9(Suppl 1):S4.
- Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010;38(Web Server issue):W214–W220.
- Montojo J, Zuberi K, Rodriguez H, Bader GD, Morris Q. GeneMANIA: fast gene network construction and function prediction for Cytoscape. *F1000Res*. 2014;3:153.
- Gokce O, Stanley GM, Treutlein B, et al. Cellular taxonomy of the mouse striatum as revealed by single-cell RNA-Seq. *Cell Rep*. 2016;16(4):1126–1137.
- Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006;313(5795):1929–1935.
- Subramanian A, Narayan R, Corsello SM, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171(6):1437–1452.

23. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol.* 2007;1:54.
24. Sepers MD, Smith-Dijak A, LeDue J, Kolodziejczyk K, Mackie K, Raymond LA. Endocannabinoid-specific impairment in synaptic plasticity in striatum of Huntington's disease mouse model. *J Neurosci.* 2018;38(3):544–554.
25. Rothe T, Deliano M, Wójtowicz AM, et al. Pathological gamma oscillations, impaired dopamine release, synapse loss and reduced dynamic range of unitary glutamatergic synaptic transmission in the striatum of hypokinetic Q175 Huntington mice. *Neuroscience.* 2015;311:519–538.
26. Fan J, Cowan CM, Zhang LY, Hayden MR, Raymond LA. Interaction of postsynaptic density protein-95 with NMDA receptors influences excitotoxicity in the yeast artificial chromosome mouse model of Huntington's disease. *J Neurosci.* 2009;29(35):10928–10938.
27. Calabresi P, Centonze D, Pisani A, et al. Striatal spiny neurons and cholinergic interneurons express differential ionotropic glutamatergic responses and vulnerability: implications for ischemia and Huntington's disease. *Ann Neurol.* 1998;43(5):586–597.
28. Gil JM, Rego AC. Mechanisms of neurodegeneration in Huntington's disease. *Eur J Neurosci.* 2008;27(11):2803–2820.
29. Thomas EA. Striatal specificity of gene expression dysregulation in Huntington's disease. *J Neurosci Res.* 2006;84(6):1151–1164.
30. Pchitskaya E, Popugaeva E, Bezprozvanny I. Calcium signaling and molecular mechanisms underlying neurodegenerative diseases. *Cell Calcium.* 2018;70:87–94.
31. Sharma S, Moon CS, Khogali A, et al. Biomarkers in Parkinson's disease (recent update). *Neurochem Int.* 2013;63(3):201–229.
32. Murmu RP, Li W, Holtmaat A, Li JY. Dendritic spine instability leads to progressive neocortical spine loss in a mouse model of Huntington's disease. *J Neurosci.* 2013;33(32):12997–13009.
33. Bulley SJ, Drew CJ, Morton AJ. Direct visualisation of abnormal dendritic spine morphology in the hippocampus of the R6/2 transgenic mouse model of Huntington's disease. *J Huntingtons Dis.* 2012;1(2):267–273.
34. Herms J, Dorostkar MM. Dendritic spine pathology in neurodegenerative diseases. *Annu Rev Pathol.* 2016;11:221–250.
35. Smith R, Brundin P, Li JY. Synaptic dysfunction in Huntington's disease: a new perspective. *Cell Mol Life Sci.* 2005;62(17):1901–1912.
36. Carroll JB, Bates GP, Steffan J, Saft C, Tabrizi SJ. Treating the whole body in Huntington's disease. *Lancet Neurol.* 2015;14(11):1135–1142.
37. Khedraki A, Reed EJ, Romer SH, et al. Depressed synaptic transmission and reduced vesicle release sites in Huntington's disease neuromuscular junctions. *J Neurosci.* 2017;37(34):8077–8091.
38. Labuschagne I, Poudel G, Kordsachia C, et al. Oxytocin selectively modulates brain processing of disgust in Huntington's disease gene carriers. *Prog Neuropsychopharmacol Biol Psychiatry.* 2018;81:11–16.
39. Koller WC, Trimble J. The gait abnormality of Huntington's disease. *Neurology.* 1985;35(10):1450–1454.
40. Mirek E, Filip M, Chwala W, et al. Three-dimensional trunk and lower limbs characteristics during gait in patients with Huntington's disease. *Front Neurosci.* 2017;11:566.
41. Indersmitten T, Tran CH, Cepeda C, Levine MS. Altered excitatory and inhibitory inputs to striatal medium-sized spiny neurons and cortical pyramidal neurons in the Q175 mouse model of Huntington's disease. *J Neurophysiol.* 2015;113(7):2953–2966.
42. Fourie C, Kim E, Waldvogel H, et al. Differential changes in postsynaptic density proteins in postmortem Huntington's disease and Parkinson's disease human brains. *J Neurodegener Dis.* 2014;2014:938530.
43. Holley SM, Joshi PR, Parievsky A, et al. Enhanced GABAergic inputs contribute to functional alterations of cholinergic interneurons in the R6/2 mouse model of Huntington's disease. *eNeuro.* 2015;2(1):ENEURO.0008-14.2015.
44. Fernández-Nogales M, Santos-Galindo M, Hernández IH, Cabrera JR, Lucas JJ. Faulty splicing and cytoskeleton abnormalities in Huntington's disease. *Brain Pathol.* 2016;26(6):772–778.
45. Eira J, Silva CS, Sousa MM, Liz MA. The cytoskeleton as a novel therapeutic target for old neurodegenerative disorders. *Prog Neurobiol.* 2016;141:61–82.
46. Sawa A, Wiegand GW, Cooper J, et al. Increased apoptosis of Huntington disease lymphoblasts associated with repeat length-dependent mitochondrial depolarization. *Nat Med.* 1999;5(10):1194–1198.
47. Bonelli RM, Niederwieser G, Diez J, Gruber A, Kölringer P. Pramipexole ameliorates neurologic and psychiatric symptoms in a Westphal variant of Huntington's disease. *Clin Neuropharmacol.* 2002;25(1):58–60.
48. Ciammola A, Sassone J, Colciago C, et al. Aripiprazole in the treatment of Huntington's disease: a case series. *Neuropsychiatric Dis Treat.* 2009;5:1–4.
49. Bonelli RM, Mayr BM, Niederwieser G, Reisecker F, Kapfhammer HP. Ziprasidone in Huntington's disease: the first case reports. *J Psychopharmacol.* 2003;17(4):459–460.
50. Chen X, Wu J, Lvovskaya S, Herndon E, Supnet C, Bezprozvanny I. Dantrolene is neuroprotective in Huntington's disease transgenic mouse model. *Mol Neurodegener.* 2011;6:81.
51. Nikolaus S, Antke C, Kley K, et al. Investigating the dopaminergic synapse in vivo. I. Molecular imaging studies in humans. *Rev Neurosci.* 2007;18(6):439–472.
52. Ross CA, Tabrizi SJ. Huntington's disease: from molecular pathogenesis to clinical treatment. *Lancet Neurol.* 2011;10(1):83–98.
53. Ross CA, Aylward EH, Wild EJ, et al. Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nat Rev Neurol.* 2014;10(4):204–216.
54. Duff K, Paulsen JS, Beglinger LJ, et al. "Frontal" behaviors before the diagnosis of Huntington's disease and their relationship to markers of disease progression: evidence of early lack of awareness. *J Neuropsychiatry Clin Neurosci.* 2010;22(2):196–207.
55. Dumas EM, van den Bogaard SJ, Middelkoop HA, Roos RA. A review of cognition in Huntington's disease. *Front Biosci (Schol Ed).* 2013;5:1–18.
56. Snowden J, Craufurd D, Griffiths H, Thompson J, Neary D. Longitudinal evaluation of cognitive disorder in Huntington's disease. *J Int Neuropsychol Soc.* 2001;7(1):33–44.
57. Bilney B, Morris ME, Churchyard A, Chiu E, Georgiou-Karistianis N. Evidence for a disorder of locomotor timing in Huntington's disease. *Mov Disord.* 2005;20(1):51–57.
58. Grimbergen YAM, Knol MJ, Bloem BR, Kremer BPH, Roos RAC, Munneke M. Falls and gait disturbances in Huntington's disease. *Mov Disord.* 2008;23(7):970–976.
59. Delval A, Krystkowiak P, Blatt JL, et al. Role of hypokinesia and bradykinesia in gait disturbances in Huntington's disease: a biomechanical study. *J Neurol.* 2006;253(1):73–80.
60. Alexandrov V, Brunner D, Hanania T, Leahy E. Reprint of: highthroughput analysis of behavior for drug discovery. *Eur J Pharmacol.* 2015;753:127–134.