Uncovering the Genetic Profiles Underlying the Intrinsic Organization of the Human Cerebellum

Yaping Wang[1,2,3,4#], Lin Chai[2,3,4#], Deying Li[2,3,4], Chaohong Gao[1,2,3,4], Congying Chu[6], Zhengyi Yang[2,3,4], Yu Zhang[7], Junhai Xu[8], Jens Randel Nyengaard[1,9,10], Simon B. Eickhoff[11,12], Bing Liu[6], Kristoffer Hougaard Madsen[1,13,14], Tianzi Jiang[1,2,3,4,5], Lingzhong Fan[1,2,3,4,5*]

[1]Sino-Danish Center, [2]University of Chinese Academy of Sciences, Beijing 100190, China
[3]Brainnetome Center, [4]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[5]CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[6]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China
[7]Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou 311100, China
[8]School of Computer Science and Technology, Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin 300350, China
[9]Core Centre for Molecular Morphology, Section for Stereology and Microscopy, Department of Clinical Medicine, Aarhus University, Aarhus 8000, Denmark
[10]Department of Pathology, Aarhus University Hospital, Aarhus 8000, Denmark
[11]Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Research Centre Jülich, Jülich, Germany
[12]Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[13]Informatics and Mathematical Modelling, Technical University of Denmark, Kongens Lyngby 2800, Denmark
[14]Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Amager and Hvidovre, Hvidovre 2650, Denmark

[#] Yaping Wang and Lin Chai contributed equally to this work.
[*] Corresponding Author: Lingzhong Fan, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. Phone: 010 - 82544523.
Email: lingzhong.fan@ia.ac.cn

Abstract

The functional diversity of the human cerebellum is hypothesized to be derived more from its extensive connections rather than being limited to its mostly invariant architecture. However, whether and how the hypothesized determination of cerebellar connections in its intrinsic organization interact with microscale gene expression is still unknown. Here we decode the genetic profiles of the cerebellar functional organization by investigating the genetic substrates simultaneously linking cerebellar functional heterogeneity and its drivers, i.e., the connections. We not only identified 443 network-specific genes but also discovered that their co-expression pattern correlated strongly with intra-cerebellar connectivity. Of these genes, 90 were also linked to the cortico-cerebellar cognitive-limbic networks. Then, using a "virtual gene knock-out" analysis, two gene sets involved in different biological processes and diseases were distinguished. Collectively, our results could help to rethink the genetic substrates underlying the cerebellar functional organization, as well as offer possible interpretability for its engagement in neuropsychiatric disorders.

Keywords: human cerebellum, functional connectivity, gene expression, resting-state network,

neuropsychiatric disorder

Teaser

The network-specific gene co-expression links the human cerebellar functional connectivity.

Introduction

Converging evidence from animal and human studies is advancing our understanding of the human cerebellum, which is engaged in motor, complex cognitive, and emotional behaviors (*1-4*). While such functional diversity of the cerebellum was believed to derive more from its extensive afferent and efferent connections to extra-cerebellar structures, rather than being limited to a regular lattice-like anatomic feature of cerebellar cortical cytoarchitecture (*1, 5-8*). It is well known that the macroscale functional organization of the human nervous system is widely accepted as being ultimately regulated by the underlying microscale gene expression (*9-12*). Therefore, unraveling the genetic profiles underlying the cerebellar functional organization could help us understand how the cerebellum organizes different functional subregions that have mostly invariant architecture (*13*) into functional networks that support its engagements in various functions (*14*) as well as increasing our understanding of its relevance in diverse brain diseases (*15, 16*).

However, the genetic mechanism supporting the functional organization of the human cerebellum is largely unknown. Only a few studies have attempted to investigate the gene expression pattern of the human cerebellum, but they provided inconsistent results in terms of gene expression variability. For instance, Hawrylycz et al. (*17*) and Negi and Guda (*18*) both found that gene expression is highly homogeneous across the anatomical regions of the healthy adult cerebellum. In contrast, Aldinger et al. (*19*) and Wang and Zoghbi (*12*) found that cerebellar development and function are governed by the precise regulation of molecular and cellular programs and that the gene expression pattern is heterogeneous across spatial and temporal scales. In addition, differences in gene expression patterns between the cerebellar gyri and sulci (*20*), and considerable cerebellar regional specializations containing specific cell types, as revealed by high-throughput single-nucleus RNA-seq (*21*) have been found in the mouse cerebellum. This inconsistency in the genetic variability of the cerebellum needs to be further explored because the relevant studies that showed homogeneity (*17, 18*) only explored the overall cerebellar genetic expression pattern across its gross macro-anatomical boundaries (e.g., cerebellar lobules) and might have failed to fully reflect the functional organization of the human cerebellum (*22, 23*).

In the past decade, functional topological maps describing the organization of the human cerebellum using task (*24*) and task-free functional magnetic resonance imaging (fMRI) (*25, 26*), specifically, separate cerebellar functional networks (*25, 26*) and intra-cerebellar functional gradients (*27*), have been proposed. In particular, Buckner et al. (*25*) employed resting-state functional connectivity (rsFC) of the cerebello-cortical circuit as a tool to map the intrinsic functional architecture of the human cerebellum and proposed a possible functional parcellation into 7 networks and 17 networks. It is thus possible to decode the genetic profiles of the cerebellar functional organization by investigating the molecular genetic substrates simultaneously linking cerebellar functional heterogeneity and its drivers, i.e., the connections. Whether and how the hypothesized determination of connections in cerebellar functional heterogeneity (*8*) interact with microscale gene expression is still an open question. To address this, one promising approach is imaging-transcriptomics analysis (*28-30*), which allows the brain-wide spatial analysis of microscopic transcriptome data to be combined with macroscopic neuroimaging phenotypes (*9*).

Thus, our goal was to investigate the neurobiological genetic basis underlying the functional organization of the human cerebellum to examine the correlation between the genes linking cerebellar functional heterogeneity and the functional integration of the human cerebellum. The schematic of the experimental design was shown in Fig. 1. Specifically, the Allen Human Brain Atlas (AHBA) transcriptome data (*9*) was combined with a cerebellar functional parcellation atlas (*25*) to identify the cerebellar network-specific genes (Fig. 1A). Then we found that the gene co-expression pattern of the network-specific genes showed a high correlation with the intra-

cerebellar FC (Fig. 1B, left). In addition, we observed coupling between the gene co-expression of ~20% network-specific genes and FC across the cerebello-cortical limbic and control networks (Fig. 1B, right). To further discover the biological functions of these genes, we performed a "virtual gene knock-out" by observing the change in the coupling between gene co-expression and FC and divided the genes into two subsets, i.e., a positive gene contribution indicator (GCI^+) and a negative gene set (GCI^-). GCI^+ is mainly involved in cerebellar neurodevelopment, while GCI^- is related to neurotransmission and is significantly enriched in various neuropsychiatric disorders that are closely linked the cerebellar functional abnormalities. The current exploration can provide a starting point in the effort to understand the molecular basis of cerebellar functional organization and open the door for investigating the pivotal role played by the cerebellum in various related neuropsychiatric disorders.

Results

The cerebellar network-specific genes derived based on the functional segregation within the cerebellum

The genes that were expressed much more in one network than in all the other six networks in the cerebellum and cerebral cortex were identified based on the differential gene expression analysis and are referred to as cerebellar network-specific genes and cortical network-specific genes, respectively. We identified 443 cerebellar network-specific genes (Supplementary sheets 1, 2) using all samples from 6 donors across 7 networks. The distribution of these network-specific genes is shown in Table 1 (Supplementary sheet 3), which shows that these were mainly expressed in the limbic (n = 170), dorsal attention (n = 51), somatomotor (n = 3), and visual (n = 221) networks.

To test the specificity of these 443 network-specific genes in reflecting the difference of gene expression across task-free cerebellar functional atlas, we applied the differential expression analysis with the same procedure in other different cerebellar atlases, includes different resolutions of the same task-free functional atlas (*25*), i.e., 17-, 32-network parcellations (32-network parcellation was derived based on the split of left and right on the 17-network parcellation to increase granularity), independent task-based multi-domain task battery (MDTB) (*24*) functional atlas, and different resolutions, i.e., 11-, 28-lobular parcellations, of the anatomical atlas (*31*). More overlaps between these 443 network-specific genes with the differentially expressed genes were observed in the different resolutions of the same task-free functional atlas, whereas less in the independent task-based functional atlas and different resolutions of the lobular atlas (Supplementary sheet 4). We also test the sensibility of network-specific genes in only 4 left-hemisphere donors with the same procedure and observed 230 cerebellar network-specific genes (Supplementary sheet 5) with preferentially expressed in the limbic (n = 49, overlap = 44, hypergeometric ps < .001) and visual (n = 181, overlap = 97, hypergeometric ps < .001).

Meanwhile, we also obtained 6,987 cortical network-specific genes (Supplementary sheets 6-8, Table 1) using the same strategy and found that the cerebellar and cortical network-specific genes distribution patterns were highly correlated (r = 0.95, p = .001). Moreover, we found that 90 of these 443 cerebellar network-specific genes (~ 20%) (Supplementary sheets 9, 10, Table 1) were convergently expressed in the cerebral cortex and that a significant overlap between the cerebellar and cortical network-specific genes of the limbic and somatomotor networks occurred (limbic overlap = 56, hypergeometric ps < .001; somatomotor overlap = 2, hypergeometric ps = .006). This means that the 56 limbic genes were differentially expressed in the limbic cortex and the limbic cerebellum and that the 2 somatomotor genes were differentially expressed in the somatomotor cortex and somatomotor cerebellum. Overlapped genes were also found in the visual network but failed to pass the hypergeometric test (visual overlap = 33, with hypergeometric ps = .839), and no overlap was found for the other 4 networks (ventral attention, dorsal attention, control, default, Supplementary sheet 9).

The co-expression of the cerebellar network-specific genes highly correlated with intra-cerebellar FC

Using the 443 cerebellar network-specific genes, we constructed the gene co-expression matrix for the 2 bi-hemisphere donors and explored the relationship between gene correlation and FC within the cerebellum. Noted, the 443 network-specific genes were defined based on 7-nework parcellation, but genetic co-expression was constructed using 17-network parcellation to increase the spatial precision. Across all the available network-network pairs, the genetic co-expression correlates with the FC within the cerebellum ($r = 0.48$, permutation test $p = .013$, Fig. 2). This correlation between gene co-expression and FC was referred to as Gene-FC correlation throughout the present paper for simplicity. To validate the Gene-FC correlation within the cerebellum, we also leveraged the different resolutions (7 and 32-network parcellations) of the same task-free functional atlas and the independent task-based multi-domain task battery (MDTB) functional parcellation (*24*) to re-perform the aforementioned steps (Fig. 2D). The gene co-expression and FC within the cerebellum also correlated when analyzed based on the 7-network parcellation ($r = 0.76$, permutation test $p < .001$, Fig. 2D, Fig. S1), 32-network parcellation ($r = 0.41$, permutation test $p = .007$, Fig. 2D, Fig. S2), and the MDTB functional parcellation ($r = 0.42$, permutation test $p = .003$, Fig. 2D, Fig. S3). Moreover, this Gene-FC correlation was also obtained using unrelated participants from HCP S1200 release ($R = 0.49$, permutation test $p = .013$, Supplementary sheet 19), and another independent HCP-Aging neuroimaging dataset ($R = 0.45$, permutation test $p = .015$, Supplementary sheet 20).

Therefore, the 443 cerebellar network-specific genes that we derived based on the functional segregation of the cerebellum also correlated with the functional integration of the cerebellum. This Gene-FC correlation was not generated by chance, so it was consistent using unrelated participants from HCP, independent HCP-Aging dataset, different parcellation resolutions, and independent cerebellar functional atlas. Moreover, the control test exhibited no Gene-FC correlation when the gene co-expression was constructed using non-network-specific genes (Supplementary sheet 21) regardless of whether the test was thresholdless or thresholded. These findings further confirmed that these 443 network-specific genes play a key role in intra-cerebellar functional organization.

Convergently expressed genes among the cerebellar and cortical network-specific genes correlated with the FC across the cerebello-cortical cognitive-limbic networks

Since 90 of the 443 cerebellar network-specific genes were convergently expressed across the cerebello-cortical circuit, we wanted to know whether these ~20% genes correlated with the FC across the cerebello-cortical circuit. A correspondence between the genetic and functional correlations was identified for the limbic ($r = 0.36$, FDR corrected $p = .030$, Fig. 3A) and control networks ($r = -0.33$, FDR corrected $p = .034$, Fig. 3B) but was not significant for the somatomotor ($r = -0.15$, FDR corrected $p = .394$), dorsal attention ($r = -0.19$, FDR corrected $p = .281$), ventral attention ($r = -0.04$, FDR corrected $p = .779$), or default ($r = 0.10$, FDR corrected $p = .544$) networks. The high cortical genetic similarity between the limbic system and the adjacent control network ($r = -0.90$, FDR corrected $p < .001$), somatomotor network ($r = -0.55$, FDR corrected $p < .001$), and ventral attention network ($r = -0.72$, FDR corrected $p < .001$) indicates that the gene co-expression between the cerebellar limbic network and the cortex reflects a gradual genetic gradient rather than genetic dissimilarity between the cerebellar limbic network and the other cerebellar networks. In addition, while controlling the effect of the cortical genetic similarity between the limbic and control networks, the partial correlation showed no cortical Gene-FC correlation for the control network ($r = -0.13$, $p = .316$), which implies that the significant cortical Gene-FC correlation for the control network was induced by the high cortical genetic similarity between the cerebellar limbic and control networks. This is also consistent with the finding that convergently expressed genes were only observed in the limbic network, but not in the control network (Table 1). Overall,

these 443 cerebellar network-specific genes not only correlated with the intra-cerebellar FC, but ~20% of them were also linked with the cerebello-cortical cognitive-limbic networks.

Functional annotation revealed distinct biological properties of GCI$^+$ and GCI$^-$ separated by virtual KO

In addition to the overall correlation between gene co-expression and the functional integration of the human cerebellum, we investigated each gene's importance to the intra-cerebellar Gene-FC correlation by scoring the 443 cerebellar network-specific genes based on the gene contribution indicator (GCI). Using the virtual gene knock-out (KO) procedure, we were able to classify the 443 network-specific genes that linked cerebellar functional segregation and integration into two groups: a 246 GCI positive gene set (GCI$^+$, Fig. 2E left, Supplementary sheet 24) and a 197 GCI negative gene set (GCI$^-$, Fig. 2E right, Supplementary sheet 25). The distinction between the two sets is that the virtual KO of GCI$^+$ genes increased the Gene-FC correlation, whereas the virtual KO of GCI$^-$ genes decreased the Gene-FC correlation. Based on the winner-take-all principle, GCI$^-$ genes may have a critical impact on the functional organization of the cerebellum; an example is that the top genes, LCP1 and TESC, enable GTPase binding and calcium binding, respectively (*32*), which are key functions within signaling transduction. Therefore, we applied a range of bioinformatics tools to further explore the underlying roles of the GCI$^+$ and GCI$^-$.

The gene ontology (GO) enrichment analysis of the GCI$^+$ and GCI$^-$ is shown in Fig. 4A. The GCI$^+$ was mainly enriched in microtubule-related terms, including the microtubule associated complex (ID: 0005875), motile cilium (ID: 0031514), and dynein complex (ID: 0030286). Compared with GCI$^+$, the GCI$^-$ was not only enriched in microtubule-related terms but was also significantly enriched in terms related to neurotransmitter transport, such as calcium ion binding (ID: 0005509), regulation of hormone levels (ID: 0010817), response to catecholamine (ID: 0071869), response to monoamine (ID: 0071867), and regulation of neurotransmitter receptor activity (ID: 0099601). This is consistent with their different pathway enrichment results (Supplementary sheets 26, 27) in that the GCI$^+$ was primarily enriched in some basic biological pathways: proximal tubule bicarbonate reclamation (ID: M4361) and glycolysis/gluconeogenesis (ID: M39474), which provides the energy need during microtube-related processes. In contrast, the GCI$^-$ was primarily involved in signaling transduction, especially in some neurotransmission pathways, such as the neuroactive ligand-receptor interaction (ID: M13380).

Since the GCI$^+$ and GCI$^-$ are involved in different biological processes, we hypothesized that they also play different roles in brain disease or related to different brain diseases. Unexpectedly, we found no link between GCI$^+$ and any brain-related illnesses (Fig. 4B, left) but observed involvement of GCI$^-$ in various neuropsychiatric disorders (Fig. 4B, right), including autistic disorder (ID: C0004325), alcoholic intoxication (ID: C0001973), mental depression (ID: C0011570), pain (ID: C0030193), learning disorders (ID: C0023186) and others. Many of these, especially mental depression and autistic disorder, have a close relationship with the human cerebellum, in which patients have shown functional connectivity abnormalities (*33, 34*). The mental depression- and autistic disorder-associated genes were TRH, PENK, TTR, ADCY5, NRXN1, HTR1A, HTR2C, NTS, PEX5L (n = 9, Supplementary sheet 27) and DLGAP2, TRH, PENK, RYR3, SEMA3A, NRXN1, TESC, ABCG2, PCDH10, CNTN4, HTR1A, CALB2, HTR2C, DNAAF4, FOLR1, NTS, GRM8, UPP2 (n = 18, Supplementary sheet 27), respectively, and the overlapping genes were TRH, PENK, NRXN1, HTR1A, HTR2C, NTS (n = 6).

In light of the distinct properties of GCI$^+$ and GCI$^-$, we wanted to know whether the roles played by these two gene sets showed variable prevalence at different ages. By leveraging the BrainSpan dataset (*35*) and applying the analysis strategy of CSEA tool (*36*), we found that GCI$^+$ showed significant overexpression in early middle fetal, late middle fetal, late fetal, neonatal early infancy, and adolescence compared with GCI$^-$ (Fig. 5A). These stages neatly correspond to the timeline of the protracted development of the human cerebellum (*37*), which extends from the early embryonic period until the end of the first postnatal year. This appears to be consistent with the

observation that the GCI^+ is involved in some fundamental biological processes, especially microtubule-related activity, whose dynamics play a key role in cerebellar neurodevelopment (*38*). In contrast, compared with the GCI^+, the GCI^- was significantly expressed in early mid fetal, neonatal early infancy, late infancy, early childhood, adolescence, and young adulthood (Fig. 5B), which includes the highest neurodevelopmental risk windows for autism spectrum disorder (ASD) (*39*) and major depression disorder (MDD) (*40*), both of which we found in the disease enrichment analysis.

Discussion

The current study provided a tentative exploration of the genetic differential and co-expression linked with the functional organization of the human cerebellum and has the potential for elaborating and rethinking the neurobiological underpinnings of the cerebellar functional organization. Furthermore, we identified two gene sets involved in cerebellar neurodevelopment and neurotransmission and found interesting, indirect genetic evidence supporting the key role played by the cerebellar functional network in many neuropsychiatric disorders, which hints at a micro-macro interacted mechanistic possibility for the cerebellar contributions to related neuropsychiatric disorders.

The genetic profiles underlying cerebellar functional segregation correlate with intra-cerebellar and cerebello-cerebral connections

In this study, we found correlations between the identified cerebellar network-specific genes with the intra-cerebellar connection and cerebello-cerebral FC. These findings could provide possible empirical genetic support for the hypothesized decisive role of cerebellar connectivity in the functional heterogeneity of the cerebellum. First, while obtaining the network-specific genes, we found significant differences in the number of identified genes between the functional specificity (i.e., limbic, visual networks) and functional diversity networks (i.e., the control, default networks); specifically, more differentially expressed genes were in the former and vice versa in the latter (*41*). This was also found in a previous cortical gene expression homogeneity analysis (*9*) that showed that a relatively high differential expression pattern was observed in the primary sensory cortex, area 38, and the primary visual cortex, a finding that closely corresponds with the somatomotor, limbic, and visual networks. But the findings related to the inconsistency in the amount of somatomotor cerebellar (n = 3) and somatomotor cortical network-specific genes (n = 960) were not completely clear. One possible explanation may be that the preferential links between the cerebellar representations of body space and the motor, somatosensory, and premotor cortices are difficult to distinguish (*25*). This finding of cerebellar network-specific genes is consistent with the elaborate regional difference in cerebellar cortical cytoarchitecture (*2, 42*) in addition to the uniform cell types and their connectivity, such as regional variation in the cell size and packing density of Purkinje, granule, and Golgi cells.

Second, the overall distribution patterns of the cerebellar and cortical network-specific genes were highly correlated, a finding that is consistent with a similar macroscale principle that was identified in the cerebellar and cortical functional organization (*27, 43*). These correlated patterns may be related to the way that we defined the cerebellar networks, which were by projecting the cerebral cortical networks onto the cerebellum by computing the functional connections between the two structures (*25*). More interestingly, the molecular genetic substrates simultaneously linking functional heterogeneity and integration could be observed across the functional subdivision, regardless of whether the parcellation was based on the task-free cerebello-cortical rsFC (*25*) or the intra-cerebellar task-based activation pattern (*24*). These interpretations are further supported by the widely accepted notion about the human cerebellum that its functional specialization is dominated more by its connection with extracerebellar structures rather than within its generally homogeneous cytoarchitecture (*8*). Although no intra-cerebellar anatomical fiber connections linking adjacent or distant cerebellar regions with each other have been found (*44, 45*), it is widely accepted that the intra-cerebellar functional map is a consequence of the topological arrangement

of its extra-cerebellar anatomical connections (*8*). This proposed relationship between extra- and intra-cerebellar connectivity can in turn be expected to affect the resting-state activity between cerebellar regions (*27*). Furthermore, for the interaction between gene and FC, they influence one another bidirectional through gene–environment (G–E) interplay (*46*) with the development as an essential element (*47*), instead of a unidirectional determination role of the genes in the brain connection. Here we observed the Gene-FC interaction from the adult perspective (AHBA donors aged from 24 to 57 years old) which lacks exploration in the G-E interplay compared with the early development (*47*).

Third, in addition to the intra-cerebellar Gene-FC correlation, we observed a direct correlation between genes underlying the cerebellar functional specialization and cerebello-cerebral FC with respect to the limbic and control networks. The Gene-FC correlation in the control network was mainly caused by the genetic similarity between these two networks; this interaction between limbic-emotion and control-cognition has been confirmed both anatomically and behaviorally (*48*). For instance, the integrated processing by the emotion and cognition areas has been identified solely based on their anatomical connections (*49*). This relationship can also be observed in that when looking at the top of a hill, a sad mood induces a steeper perception of the hill than a happy one (*50*). One possible reason why we only obtained this correspondence in the limbic network may be the low functional heterogeneity (*41*) and inter-individual functional variability (*51*) of the limbic network compared with others as well as the complexity of gene expression; i.e., the Gene-FC correlation is not fully portrayed by the differentially expressed genes (*30*). Considering the indirect connection between cerebellum and cortex, and large differences between the cerebellum and cortex in terms of their gene expression patterns (*17*) and structure-function relationships (*52*) as well as the individual variability of their functional networks (*53*), identifying 90 convergently expressed genes that linked the cerebello-cortical cognitive-limbic networks is very significant and may hold clues to the molecular underpinnings of the cognitive-emotion roles played by the cerebello-cortical circuit. For example, the HTR1A and HTR2C, which are both preferentially expressed in the cerebellar and cortical limbic network, are pivotal genes in serotonin transmission, play a modulation role in the limbic system, and act as important therapeutic targets in limbic system-related disorders (*54*).

Cerebellar neurodevelopment feature of GCI$^+$, cerebellar neurotransmission, and neuropsychiatric disorders-related features of GCI$^-$

Interestingly, we derived two gene subsets with pronouncedly different characteristics based only on the direction in which each gene influenced the intra-cerebellar Gene-FC correlation by applying a simple virtual KO approach on the 443 cerebellar network-specific genes. By using a series of bioinformatic tools, we found converging evidence for GCI$^+$ and GCI$^-$ involvement in cerebellar neurodevelopment and cerebellar neurotransmission, respectively. It is also interesting to speculate that these 443 network-specific genes that link both cerebellar functional segregation and integration have a relationship with some brain-related disorders since prior evidence showed that the cerebellar functional organization plays a key role in various neurological (*16, 55*) and psychiatric disorders (*15*), most of which possess common underlying genetic risks (*56*). But a tricky problem emerged in that the genes we are interested in were derived from healthy individuals. This could be tackled to some extent by using the virtual KO method, which can simulate the different expression levels of each gene and thus coarsely corresponds to a fraction of the expression level under normal health and disease situations. This is why we thought that we might be able to see whether the GCI$^+$ and GCI$^-$ are related to a specific disease even though the genes were derived from healthy individuals.

The GCI$^+$ is involved in many microtubule-related terms and is overexpressed throughout the protracted development of the cerebellum. The dynamics and flexibility of microtubules were found to be essential throughout cerebellar development because they affect the morphological alterations of Purkinje cells (*38*). In addition, some genes of the GCI$^+$, such as GTPBP2 (*57*) and

Lin28b (*58*), were found to play a key role in neurodevelopment; overexpression of the Lin28b gene can induce the development of pathological lobulation in the cerebellum (*58*). This converging evidence prompts our speculation that the GCI^+ is engaged in cerebellar neurodevelopment. Unexpectedly, the GCI^+ showed no link to brain-related diseases, which appears to be consistent with its primary involvement in many fundamental biological functions. However, this lack of disease linkage is inconsistent with the significant overexpression of GCI^+ genes during the protracted development of the cerebellum, in that many researchers pointed out that this protracted development increased the susceptibility of the cerebellum to many psychiatric disorders (*37*). This likely is complemented by the overexpression of GCI^- in the early middle fetal and neonatal early infancy periods. Other possible explanations include that there are few genetic studies of the cerebellum compared with the cerebral cortex as well as large genetic expression differences between the cerebellum and extra-cerebellar structures (*17*), so the related datasets may lack sufficient information that is specific to the cerebellum. This calls for future studies seeking to provide a more complete explanation by considering multiple perspectives.

The GCI^- was found to be involved in many neurotransmission processes, enriched in various neurological and psychiatric disorders, and significantly overexpressed in late infancy, early childhood, adolescence, and young adulthood compared with GCI^+. These results are mutually supportive. Neurotransmission has long been thought to play a crucial role in various neurological (*59*) and psychiatric disorders (*60, 61*). For example, the abnormal transmission of monamines and catecholamines, such as serotonin and dopamine, has been widely linked with many psychiatric disorders, and these transmitters have thus become potential treatment targets (*62*). The period through which the GCI^- genes are expressed includes the high-risk time windows for GCI^--enriched disorders, such as mental depression (aged 18–29) (*40*) and autistic disorder (from infancy to childhood) (*39*), and the high expression of GCI^- in early middle fetal life might be associated with the prenatal risk factors associated with depression (*63*) and autism (*64*).

Moreover, we found that the GCI^- was enriched in many neuropsychiatric disorders including mental depression, autistic disorder, pain, alcoholic intoxication, learning disorder, and others. These disorders are closely related to alterations of the cerebellar FC. Examples include: the dynamic FC of the cerebello-cortical affective-limbic network is associated with the severity of MDD patients (*33*); ASD patients display decreased FC between the cerebellum and some cortical regions involved in cognitive systems (*65*); the cerebellum is one of the brain regions most sensitive to the harmful effects of chronic alcohol abuse (*66*), and the cerebello-cortical FC of patients with alcohol use disorder has been shown to have changes in both flexibility and integration (*67*). Therefore, the GCI^- provides a possible micro-macro interacted mechanistic explanation for the engagements of the cerebellum in various neuropsychiatric disorders; i.e., presumably, one of the ways these risk genes play a role in the pathogenesis of corresponding diseases maybe through their interactions with the cerebellar FC, which results in dysregulation of cerebellar FC and thus pathological manifested as its functional connection abnormalities, such as the fluctuation in the correspondence of the Gene-FC relationship found in the present study and widespread altered cerebellar FC excavated previously in diverse neuropsychiatric disorders (*15*). The GCI^- also provides a promising genetic resource for investigating the cerebellar involvements in a range of brain diseases. For example, finding that overlapping genes, i.e., NRXN1, are associated with mental depression and autistic disorder supported previous clinical studies showing that rare and common variants in NRXN1 carried risks for MDD (*68*), ASD, and schizophrenia (*69*), and HTR1A, which has a high expression in the cerebellum (*18*), was found to be involved in pain, mental depression, autistic disorder, alcoholic intoxication, learning disorder, and other conditions.

Methodological considerations

The interpretation of our findings is not without limitations. First, we validated our results across independent imaging datasets which cover the ages of AHBA, different resolutions of the task-free

cerebellar atlas, and inter-independent cerebellar functional atlases. But we are unable to test our results in an independent gene dataset as there is no other publicly available dataset with a detailed sampling of subregions of the human cerebellum. Second, the gene co-expression we constructed only considered one small part of the relationship between the genes and FC thereby it did not fully recapitulate the complexity of the brain transcriptome, such as gene-gene interactions (*70*). That is one possible reason why we only found a cerebello-cortical Gene-FC correlation for the cognitive-limbic networks. Third, the simple correlation approach (*29, 41*) such as used in this study and other linear regression models like partial least squares analysis (*71*), are only able to prioritize genes for further investigation and cannot fully explore the causal relationship between genes and functional organization. As a result, further exploration is hindered by the intricacies of genetic and epigenetic regulation. This makes the discussion and explanation of the different directions of this correlation challenging. For example, why the direction of influence on the Gene-FC correlation could separate these 443 genes into two distinct gene sets with definitely different functions remains unclear, so further related exploration is necessary but very challenging. Briefly, more donors, the overall pattern of gene expression, gene regulation, epigenomics, and improved cellular resolution is needed and imperative for developing more appropriate and ingenious approaches to fully understand the causal relationship between genes and functional organization, which is a greater challenge for neuroscience than just identifying a link between genetic and imaging data. Nevertheless, in light of the currently limited understanding of the details about how microscale genes contribute to macroscale brain functional organization, the prioritization of genes and the related functional annotation presented here are still necessary (*28, 30*).

In conclusion, we found that the network-specific genes underlying cerebellar functional heterogeneity correlated with the intra-cerebellar and cerebello-cerebral FC, a finding which indicates that the genetic infrastructure associated with functional segregation coalesces to form a collective system, which has a close relationship with the functional integration of these functional subregions. The current study has thus unveiled part of the neurobiological genetic substrate underlying the cerebellar functional organization. We also identified important indirect genetic markers that support the key role played by the cerebellar functional network in many brain disorders. This hints at the possibility of establishing a "cerebellar functional abnormality – gene – neuropsychiatric disorder" loop as well as of help to bridge the knowledge gap between the genetic mechanisms driving the cerebellar functional organization and the heritable risks of disorders, especially major depression and autistic disorder. The current study also prioritizes genes for future studies that will focus on the genetic correlates of the cerebellar functional organization, the genetic implications of cerebellar malfunction in the pathogenesis of many neuropsychiatric disorders, and future genetic treatment targets for the cerebellar functional abnormalities of related neuropsychiatric disorders.

Materials and Methods

AHBA preprocessing

The AHBA (*9*) is a publicly available transcriptome dataset (http://www.brain-map.org), which provides normalized microarray gene expression data from six adult donors (ages 24, 31, 34, 49, 55, and 57 years; n = 4 left hemisphere only, n = 2 both left and right hemispheres). Table S1 shows the demographic information. Although AHBA provides gene expression from only six adult donors, it still has many unprecedented advantages. While some existing human gene expression atlases just cover multiple brain regions, only the AHBA delivers high-resolution coverage of nearly the entire brain (*72*), includes expression of more than 20,000 genes taken from 3,702 spatially distinct tissue samples (*9*) range from the cerebral cortex to cerebellum.

The preprocessing pipeline was referred to in Anderson et al. (*30*), and included data filtering, probe selection, sample selection, and assignment. We first filtered the probes with the AHBA binary indicator to mitigate the background noise and excluded probes without an Entrez ID. Then

for the genes that corresponded to two or more probes, we chose the probe with the maximum summed adjacency to represent the corresponding gene expression; otherwise, the probe with the highest mean expression was retained, using the CollapseRows function (*73*) in R. The first two steps generated 20,738 unique mRNA probes, which provided expression data for 20,738 genes. As suggested by Arnatkeviciute et al. (*72*) and given the known transcriptional differences (*17*) between the cortical and sub-cortical regions and the cerebellum, we separated the cortical and cerebellar samples a priori based on the slab type and structure name provided by AHBA and processed them separately later. In the end, 337 samples were retained for the cerebellar cortex and 1,701 samples for the cortical cortex.

Finally, we respectively assigned these 337 cerebellum samples and 1,701 cortical samples into the cerebellar functional network atlas (*25*) and cortical functional networks atlas (*74*), both of which have 7- and 17-network parcellation strategies. For each cerebellar sample, we first generated a single $1 \times 1 \times 1$ mm^3 region of interest (ROI) at the MNI coordinate for each sample using AFNI 3dmaskdump -nbox function (*75*). The network label from either region 7 or 17 was assigned if the ROI fell within a cerebellar network of the Buckner atlas. Considering the uneven and discrete sampling of the AHBA data (*9*), if the $1 \times 1 \times 1$ mm^3 ROI did not overlap with any network, the associated ROI was expanded to $3 \times 3 \times 3$ mm^3, and if the $3 \times 3 \times 3$ mm^3 ROI overlapped with the functional atlas, the network that had the maximum number of shared voxels with the ROI was assigned. Otherwise, the steps above were repeated for a $5 \times 5 \times 5$ mm^3 ROI. The cerebellar samples were excluded ($n = 22$) if the $5 \times 5 \times 5$ mm^3 ROI did not overlap with any cerebellar networks. Tables S2 and S3 show the distributions of the cerebellar sample assignment for the 7-network and 17-network atlases. The assignment of the AHBA cortical samples into the cortical functional network atlas was consistent with the method used for the cerebellum, and the cortical sample distributions are shown in Tables S4 and S5.

Differential gene expression analysis across functional networks

The gene expressions of the cerebellar samples within the same network were averaged for each gene across the samples individually, resulting in 20,738 genes × 7 or 17 network matrices for each donor. Here we used as many of the most original overall genes as possible, because some existing methods (*17, 72*) for screening genes do not consider the specificity of gene expression within the human cerebellum. Then we calculated the differential gene expression across the 7 networks using the R limma package (*76*) by comparing the gene expression in one network (e.g., control) with the remaining 6 networks (e.g., default, limbic, visual, etc.). The traditional minimum fold change threshold was not suitable for determining biologically meaningful but subtly different expressions (*77*). Instead, we applied the Benjamini-Hochberg (BH) method to control the false discovery rate (FDR), and the FDR corrected statistical threshold $q \leq .05$ combined with a fold change > 1 was used as the key indicator for differentially expressed genes. The inter-donor effects were accounted for by limma's duplicateCorrelation tool (*76*). For simplicity, the genes that were differentially expressed across cerebellar networks are referred to as cerebellar network-specific genes throughout this paper. The cortical network-specific genes were identified in the same way. The only difference was that the gene expression of the cortical samples was first averaged within each parcel (51 and 114 parcels, which corresponded to the 7- and 17-networks, respectively) (*74*) and then averaged within each network.

To test the specificity of these network-specific genes derived based on task-free 7 functional networks, we also applied the differential gene expression analysis with the same procedure in other different cerebellar atlases, includes different resolutions of the same task-free functional atlas (*25*), i.e., 17-, 32-network parcellations (32-network parcellation was derived based on the split of left and right on the 17-network parcellation to increase granularity), independent task-based multi-domain task battery (MDTB) (*24*) functional atlas and different resolutions, i.e., 11-, 28-lobular parcellations, of the anatomical atlas (*31*). We estimated the overlapped genes between the network-specific genes derived based on the task-free 7 functional networks with other atlases.

Moreover, we also applied the differential gene expression analysis in only 4 left-hemisphere donors with the same procedure to test the sensibility of these network-specific genes.

Cerebellar resting-state functional connectivity (rsFC)

The minimally preprocessed (*78, 79*) Human Connectome Project (HCP) S1200 release dataset (*80*), which has 1,018 subjects (aged from 22 to 37 years old) with both structural MRI and resting-state functional MRI (rs-fMRI, HCP S1200 manual), was used. The preprocessing pipeline includes artifact correction (correction of gradient nonlinearity distortion, realignment for head motion, registration of fMRI data using structural data, reduction of geometric distortions due to B0 field inhomogeneity, etc.) as well as denoising by ICA-FIX (*81, 82*). Time courses were extracted from these CIFTI grayordinate-format preprocessed rs-fMRI images, and the global signal was regressed as well. The resting-state BOLD time series were averaged within each cortical parcel of the 7- or 17-network cortical atlases and within each cerebellar network of the 7- or 17-network cerebellar atlases (*25*), separately. The rsFC within the cerebellum was computed using Pearson's correlation for the averaged time courses for each ROI of interest. Because four runs were performed for each subject, the correlation values were separately calculated for each run, Fisher's z-transformed, and averaged across the runs, resulting in a 17 × 17 networks matrix. The same process was used to calculate the correlations between each functional cerebellar network and each cortical parcel, resulting in a 114 cortical parcels × 7 cerebellar networks functional correlation matrix, which represents the rsFC across the cerebello-cortical circuit. Regardless of whether the FC was within the cerebellum or across the cerebello-cortical circuit, both categories of FC were defined using the more fine-grained 17-network parcellation to increase the spatial resolution. The only exception was that the cerebellar 7-network was applied while calculating the FC across the cerebello-cortical circuit to compare each cerebellar network more directly.

Correlation between gene co-expression with intra-cerebellar rsFC

To fully capture the genetic correlation with the FC within the cerebellum, we leveraged the genetic samples of the 2 bi-hemisphere donors when constructing the gene co-expression matrix because the rsFC of the cerebellum is bilateral. Therefore, the gene co-expression was analyzed for the 2 bi-hemisphere donors using the 443 differentially expressed genes derived from all 6 donors across 7 networks, using a finer 17-network parcellation to increase the spatial resolution. Ten networks that contained samples from both bi-hemisphere donors were retained (Table S3). For each bi-hemispheric donor, the log2 gene expression of the cerebellar samples was mean-normalized and then averaged within each network. The cerebellar 10 × 10 networks correlation matrix was calculated using Spearman's correlations individually, then Fisher transformed, and finally averaged to construct the final 10 networks gene co-expression matrix. The correlation significance level of the gene co-expression was evaluated using the overlap between the correlation matrix for these two individuals and adjusted by Bonferroni correction. Meanwhile, we transformed the 17 × 17 networks rsFC matrix into a 10 × 10 networks size to be consistent with the gene co-expression matrix. Finally, the relationship between the 10 × 10 networks gene co-expression and the 10 × 10 networks rsFC matrix was computed using Pearson's correlation. The correlation between the gene co-expression and FC is referred to as the Gene-FC correlation throughout the present paper for simplicity.

To test whether these Gene-FC relationships were identified by chance, we randomly shuffled the network label of each cerebellar sample 10,000 times, kept the distribution probability of the sample in each network consistent, and then reperformed the previous analyses with the same criteria for each permutation. Since recent efforts to control the auto-correlation such as the spatial permutation (spin) test was not suitable for the cerebellum (*83*) and our analyses were all at the network level, the classical permutation test was used. In addition, to confirm that the verified Gene-FC correlation within the cerebellum is meaningful and to evaluate its robustness, we also

recalculated it using several different parcellations, i.e., task-free 7- and 32-network parcellations, and independent task-based multi-domain task battery (MDTB) functional parcellation (*24*). The criteria for each step were consistent with our main method. Lastly, we employed a control test to learn whether the Gene-FC correlation could be obtained using only the network-specific genes, that is, no Gene-FC correlation while using other genes. We randomly select 443 genes from the full gene set without the network-specific genes and referred to them as non-network-specific genes. Then we calculated the Gene-FC correlation using the non-network-specific genes and ran this step randomly 10,000 times. In addition to these thresholdless non-network-specific genes, we applied a set of thresholds to the averaged original log2 gene expression data to confirm that these non-network-specific genes were expressed in the cerebellum and to test whether the gene co-expression pattern constructed using these threshold non-network-specific genes was correlated with FC.

Moreover, we also verified the robustness of this Gene-FC correlation using the 218 unrelated participants from the HCP S1200 release (*80*). And 296 participants with four runs (aged from 38 to 58 years old) from preprocessed HCP-Aging Lifespan 2.0 Release (*84*) were leveraged to constitute an independent neuroimaging validation dataset, which together with the 1,018 participants from the HCP S1200 release of our main strategy, neatly covered the age ranges of AHBA (extends from 24 to 57 years old). The analysis strategy for these two additional datasets is the same as our main approach, which has been described above.

Correlation between gene co-expression and rsFC across the cerebello-cortical circuit

To fully investigate the cerebellar functional organization, we also explored the relationship between the cerebello-cortical FC and the genetic correlation based on the strategy used in Anderson et al. (*30*). First, we defined the network-specific genes in the cortex using the same procedure as we had for the cerebellum and examined the genes that overlapped within the same network of the cerebellum and the cortex using a hypergeometric test. Then the gene co-expression matrix was constructed between 6 cerebellar networks and 59 cortical parcels from the 2 bi-hemisphere donors, using the 90 unique genes derived from the overlap between the cortical network-specific genes and the cerebellar network-specific genes. Here, the cerebellar 7-network parcellation was selected to compare the different cerebellar networks more directly. The visual network was excluded because it only had two samples that were solely from one of the 2 bi-hemisphere donors. For the cerebral cortex, 59 cortical parcels that contained samples from both bi-hemisphere donors were estimated. The log2 mean-normalized expression within each cerebellar network and each cortical parcel was estimated individually and correlated using Spearman's ρ, Fisher-transformed, and averaged. We transformed the 114 cortical parcels × 7 cerebellar networks rsFC matrix into 59 cortical parcels × 6 cerebellar networks size to be consistent with the gene co-expression matrix. Finally, the relationship between the cortical genetic correlation and the cerebello-cortical rsFC matrix was computed using Pearson's correlation across 6 cerebellar networks and adjusted by the Benjamini-Hochberg method to correct for multiple comparisons.

Gene functional annotation

Virtual Gene Knock-out (KO)

To extend our investigation of the overall relationship between gene co-expression and FC within the cerebellum, we referred to a similar previous approach (*85, 86*) and termed it the "Virtual Gene Knock-out (KO)" to evaluate each gene's contribution to the Gene-FC correlation. In brief, we deleted each of the 443 cerebellar network-specific genes one-by-one to simulate the gene knock-out, then constructed the gene co-expression matrix without that gene, analyzed the correlation between the FC and the gene co-expression, and finally calculated the difference in the correlation coefficient between before and after the simulated deletion, with the result being defined as the gene contribution indicator (GCI) (*85*). Based on the GCI, we identified two different gene sets

that had opposite effects on the Gene-FC correlation: a GCI positive gene set (GCI^+) and a GCI negative gene set (GCI^-). The virtual KO of GCI^+ increased the Gene-FC correlation, and, accordingly, its expression decreased the Gene-FC correlation; in contrast, the virtual KO of GCI^- decreased the correlation, and, accordingly, its expression increased the Gene-FC correlation.

GO, pathway, and disorder enrichment analysis (ToppGene portal)

To characterize the biological role of GCI^+ and GCI^-, we applied the ToppGene portal (*87*) to conduct a gene ontology (GO), pathway, and disorder enrichment analysis. The GO (*88*) enrichment analysis provides ontologies to describe accumulated knowledge of genes in three biological domains: biological process (BP), cellular component (CC), and molecular function (MF). The Benjamini-Hochberg method for false discovery rate (FDR-BH correction) ($q < .05$) was chosen to correct for multiple comparisons.

Integrative temporal specificity Analysis

To investigate the overall temporal expression features of these genes, we applied an online cell type-specific expression analysis (CSEA) tool (*36*) to do the enrichment analysis of the genes within the cerebellum during different lifespan windows. Here, a specificity index probability (pSI = .05, .01, .001, and .0001, permutation corrected) was used to define the probability of a gene being expressed in each time window relative to all other time windows to represent the varying stringencies for enrichment. The significance of the overlap between the interest gene set and those enriched in a specific time window was evaluated by Fisher's exact test, and the Benjamini-Hochberg method for false discovery rate (FDR-BH correction) was chosen to correct for multiple comparisons.

References

1. J. D. Schmahmann, X. Guell, C. J. Stoodley, M. A. Halko, The Theory and Neuroscience of Cerebellar Cognition. *Annu Rev Neurosci* **42**, 337-364 (2019).
2. C. I. De Zeeuw, S. G. Lisberger, J. L. Raymond, Diversity and dynamism in the cerebellum. *Nat Neurosci* **24**, 160-167 (2021).
3. P. L. Strick, R. P. Dum, J. A. Fiez, Cerebellum and nonmotor function. *Annu Rev Neurosci* **32**, 413-434 (2009).
4. R. L. Buckner, The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron* **80**, 807-815 (2013).
5. A. Sathyanesan *et al.*, Emerging connections between cerebellar development, behaviour and complex brain disorders. *Nat Rev Neurosci* **20**, 298-313 (2019).
6. X. Guell, J. D. Schmahmann, J. D. E. Gabrieli, Functional Specialization is Independent of Microstructural Variation in Cerebellum but Not in Cerebral Cortex. *bioRxiv*, 424176 (2018).
7. J. Voogd, M. Glickstein, The anatomy of the cerebellum. *Trends Cognit. Sci.* **2**, 307-313 (1998).
8. J. D. Schmahmann, The role of the cerebellum in affect and psychosis. *J Neurolinguistics* **13**, 189-214 (2000).
9. M. J. Hawrylycz *et al.*, An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399 (2012).
10. M. B. Johnson *et al.*, Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494-509 (2009).
11. E. E. Guevara *et al.*, Comparative analysis reveals distinctive epigenetic features of the human cerebellum. *PLoS Genet* **17**, e1009506 (2021).
12. V. Y. Wang, H. Y. Zoghbi, Genetic regulation of cerebellar development. *Nat Rev Neurosci* **2**, 484-491 (2001).

13. S. Ramón y Cajal, *Histologie du système nerveux de l'homme & des vertébrés*. (Maloine, 1909).
14. J. D. Schmahmann, The cerebellum and cognition. *Neurosci. Lett.* **688**, 62-75 (2019).
15. J. R. Phillips, D. H. Hewedi, A. M. Eissa, A. A. Moustafa, The cerebellum and psychiatric disorders. *Front Public Health* **3**, 66 (2015).
16. S. Diwakar, Cerebellum in Neurological Disorders: A Review on the Role of Inter-Connected Neural Circuits. *J. neurol. stroke* **6**, (2017).
17. M. Hawrylycz *et al.*, Canonical genetic signatures of the adult human brain. *Nat Neurosci* **18**, 1832-1844 (2015).
18. S. K. Negi, C. Guda, Global gene expression profiling of healthy human brain and its application in studying neurological disorders. *Sci Rep* **7**, 897 (2017).
19. K. A. Aldinger *et al.*, Spatial and cell type transcriptional landscape of human cerebellar development. *Nat Neurosci*, (2021).
20. T. Zeng *et al.*, Allen mouse brain atlases reveal different neural connection and gene expression patterns in cerebellum gyri and sulci. *Brain Struct Funct* **220**, 2691-2703 (2015).
21. V. Kozareva *et al.*, A transcriptomic atlas of the mouse cerebellum reveals regional specializations and novel cell types. *bioRxiv*, 2020.2003.2004.976407 (2020).
22. Y. Ren, L. Guo, C. C. Guo, A connectivity-based parcellation improved functional representation of the human cerebellum. *Sci Rep* **9**, 9115 (2019).
23. J. A. Bernard *et al.*, Resting state cortico-cerebellar functional connectivity networks: a comparison of anatomical and self-organizing map approaches. *Front Neuroanat* **6**, 31 (2012).
24. M. King, C. R. Hernandez-Castillo, R. A. Poldrack, R. B. Ivry, J. Diedrichsen, Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nat Neurosci* **22**, 1371-1378 (2019).
25. R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, B. T. Yeo, The organization of the human cerebellum estimated by intrinsic functional connectivity. *J Neurophysiol* **106**, 2322-2345 (2011).
26. J. L. Ji *et al.*, Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage* **185**, 35-57 (2019).
27. X. Guell, J. D. Schmahmann, J. Gabrieli, S. S. Ghosh, Functional gradients of the cerebellum. *Elife* **7**, (2018).
28. A. Fornito, A. Arnatkeviciute, B. D. Fulcher, Bridging the Gap between Connectome and Transcriptome. *Trends Cogn Sci* **23**, 34-50 (2019).
29. J. Richiardi *et al.*, BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241-1244 (2015).
30. K. M. Anderson *et al.*, Gene expression links functional networks across cortex and striatum. *Nat Commun* **9**, 1428 (2018).
31. J. Diedrichsen, J. H. Balsters, J. Flavell, E. Cussans, N. Ramnani, A probabilistic MR atlas of the human cerebellum. *Neuroimage* **46**, 39-46 (2009).
32. R. P. Huntley *et al.*, The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res* **43**, D1057-1063 (2015).
33. D. M. Zhu *et al.*, Cerebellar-cerebral dynamic functional connectivity alterations in major depressive disorder. *J. Affect. Disord.* **275**, 319-328 (2020).
34. M. Verly *et al.*, Altered functional connectivity of the language network in ASD: role of classical language areas and cerebellum. *Neuroimage Clin* **4**, 374-382 (2014).
35. J. A. Miller *et al.*, Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206 (2014).

36. J. D. Dougherty, E. F. Schmidt, M. Nakajima, N. Heintz, Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res* **38**, 4218-4230 (2010).
37. H. J. ten Donkelaar, M. Lammens, P. Wesseling, H. O. Thijssen, W. O. Renier, Development and developmental disorders of the human cerebellum. *J Neurol* **250**, 1025-1036 (2003).
38. R. Munoz-Castaneda *et al.*, Cytoskeleton stability is essential for the integrity of the cerebellum and its motor- and affective-related behaviors. *Sci Rep* **8**, 3072 (2018).
39. I. Beyreli, O. Karakahya, A. E. Cicek, Deep multitask learning of gene risk for comorbid neurodevelopmental disorders. *bioRxiv*, 2020.2006.2013.150201 (2020).
40. M. A. Villarroel, E. P. Terlizzi, Symptoms of Depression Among Adults: United States, 2019. *NCHS Data Brief*, 1-8 (2020).
41. M. L. Anderson, J. Kinnison, L. Pessoa, Describing functional diversity of brain regions and brain networks. *Neuroimage* **73**, 50-58 (2013).
42. N. L. Cerminara, E. J. Lang, R. V. Sillitoe, R. Apps, Redefining the cerebellar cortex as an assembly of non-uniform Purkinje cell microcircuits. *Nat Rev Neurosci* **16**, 79-93 (2015).
43. D. S. Margulies *et al.*, Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc Natl Acad Sci U S A* **113**, 12574-12579 (2016).
44. J. D. Schmahmann, D. N. Pandya, Disconnection syndromes of basal ganglia, thalamus, and cerebrocerebellar systems. *Cortex* **44**, 1037-1066 (2008).
45. J. D. Schmahmann, From movement to thought: anatomic substrates of the cerebellar contribution to cognitive processing. *Hum Brain Mapp* **4**, 174-198 (1996).
46. L. Krubitzer, J. Kaas, The evolution of the neocortex in mammals: how is phenotypic diversity generated? *Curr Opin Neurobiol* **15**, 444-453 (2005).
47. W. T. Boyce, M. B. Sokolowski, G. E. Robinson, Genes and environments, development and time. *Proc Natl Acad Sci U S A* **117**, 23235-23241 (2020).
48. J. Storbeck, G. L. Clore, On the interdependence of cognition and emotion. *Cogn Emot* **21**, 1212-1237 (2007).
49. H. T. Ghashghaei, H. Barbas, Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* **115**, 1261-1279 (2002).
50. C. R. Riener, J. K. Stefanucci, D. R. Proffitt, G. Clore, An effect of mood on the perception of geographical slant. *Cogn Emot* **25**, 174-182 (2011).
51. S. Mueller *et al.*, Individual variability in functional connectivity architecture of the human brain. *Neuron* **77**, 586-595 (2013).
52. X. Guell, J. D. Schmahmann, J. D. Gabrieli, Functional Specialization is Independent of Microstructural Variation in Cerebellum but Not in Cerebral Cortex. *bioRxiv*, 424176 (2018).
53. S. Marek *et al.*, Spatial and Temporal Organization of the Individual Human Cerebellum. *Neuron* **100**, 977-993 e977 (2018).
54. J. G. Hensler, Serotonergic modulation of the limbic system. *Neurosci Biobehav Rev* **30**, 203-214 (2006).
55. J. D. Schmahmann, Disorders of the cerebellum: ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. *J. Neuropsychiatry Clin. Neurosci.* **16**, 367-378 (2004).
56. C. Brainstorm *et al.*, Analysis of shared heritability in common disorders of the brain. *Science* **360**, (2018).
57. A. M. Bertoli-Avella *et al.*, Biallelic inactivating variants in the GTPBP2 gene cause a neurodevelopmental disorder with severe intellectual disability. *Eur J Hum Genet* **26**, 592-598 (2018).

58. A. K. Wefers, S. Lindner, J. H. Schulte, U. Schuller, Overexpression of Lin28b in Neural Stem Cells is Insufficient for Brain Tumor Formation, but Induces Pathological Lobulation of the Developing Cerebellum. *Cerebellum* **16**, 122-131 (2017).
59. K. J. Swoboda, K. Hyland, Diagnosis and treatment of neurotransmitter-related disorders. *Neurol Clin* **20**, 1143-1161, viii (2002).
60. R. P. Brown, J. J. Mann, A clinical perspective on the role of neurotransmitters in mental disorders. *Hosp Community Psychiatry* **36**, 141-150 (1985).
61. T. A. Kato *et al.*, Neurotransmitters, psychotropic drugs and microglia: clinical implications for psychiatry. *Curr Med Chem* **20**, 331-344 (2013).
62. D. Seo, C. J. Patrick, P. J. Kennealy, Role of Serotonin and Dopamine System Interactions in the Neurobiology of Impulsive Aggression and its Comorbidity with other Clinical Disorders. *Aggress Violent Behav* **13**, 383-395 (2008).
63. Y. Su, C. D'Arcy, X. Meng, Research Review: Developmental origins of depression - a systematic review and meta-analysis. *J Child Psychol Psychiatry*, (2020).
64. C. J. Newschaffer, D. Fallin, N. L. Lee, Heritable and nonheritable risk factors for autism spectrum disorders. *Epidemiol Rev* **24**, 137-153 (2002).
65. T. C. Ramos, J. B. Balardin, J. R. Sato, A. Fujita, Abnormal Cortico-Cerebellar Functional Connectivity in Autism Spectrum Disorder. *Front Syst Neurosci* **12**, 74 (2018).
66. K. G. Baker, A. J. Harding, G. M. Halliday, J. J. Kril, C. G. Harper, Neuronal loss in functional zones of the cerebellum of chronic alcoholics with and without Wernicke's encephalopathy. *Neuroscience* **91**, 429-438 (1999).
67. M. Abdallah *et al.*, Altered Cerebro-Cerebellar Dynamic Functional Connectivity in Alcohol Use Disorder: a Resting-State fMRI Study. *Cerebellum*, (2021).
68. C. Freire-Cobo, J. Wang, Dietary phytochemicals modulate experience-dependent changes in Neurexin gene expression and alternative splicing in mice after chronic variable stress exposure. *Eur J Pharmacol* **883**, 173362 (2020).
69. Z. Hu, X. Xiao, Z. Zhang, M. Li, Genetic insights and neurobiological implications from NRXN1 in neuropsychiatric disorders. *Mol. Psychiatry* **24**, 1400-1414 (2019).
70. J. Hua, Z. Yang, T. Jiang, S. Yu, Pairwise interactions in gene expression determine a hierarchical transcriptional profile in the human brain. *Sci. Bull.* **66**, 1437-1447 (2021).
71. G. Ball, J. Seidlitz, R. Beare, M. L. Seal, Cortical remodelling in childhood is associated with genes enriched for neurodevelopmental disorders. *Neuroimage* **215**, 116803 (2020).
72. A. Arnatkeviciute, B. D. Fulcher, A. Fornito, A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353-367 (2019).
73. J. A. Miller *et al.*, Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics* **12**, 322 (2011).
74. B. T. Yeo *et al.*, The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* **106**, 1125-1165 (2011).
75. R. W. Cox, AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162-173 (1996).
76. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
77. M. R. Dalman, A. Deeter, G. Nimishakavi, Z.-H. Duan, Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics* **13**, S11 (2012).
78. S. M. Smith *et al.*, Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144-168 (2013).
79. M. F. Glasser *et al.*, The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105-124 (2013).
80. D. C. Van Essen *et al.*, The Human Connectome Project: a data acquisition perspective. *Neuroimage* **62**, 2222-2231 (2012).

81. G. Salimi-Khorshidi *et al.*, Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, 449-468 (2014).
82. L. Griffanti *et al.*, ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* **95**, 232-247 (2014).
83. A. F. Alexander-Bloch *et al.*, On testing for spatial correspondence between maps of human brain structure and function. *Neuroimage* **178**, 540-551 (2018).
84. S. Y. Bookheimer *et al.*, The Lifespan Human Connectome Project in Aging: An overview. *Neuroimage* **185**, 335-348 (2019).
85. J. Seidlitz *et al.*, Morphometric Similarity Networks Detect Microscale Cortical Organization and Predict Inter-Individual Cognitive Variation. *Neuron* **97**, 231-247 e237 (2018).
86. X.-Z. Kong *et al.*, Gene Expression Correlates of the Cortical Network Underlying Sentence Processing. *Neurobiol Lang* **1**, 77-103 (2020).
87. J. Chen, E. E. Bardes, B. J. Aronow, A. G. Jegga, ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305-311 (2009).
88. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

Acknowledgements
The authors appreciate the English language and editing assistance of Rhoda E. and Edmund F. Perozzi, PhDs. Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen, and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.
Funding: This work was partially supported by
Natural Science Foundation of China (Grant Nos. 82072099, 91432302, and 31620103905)
Strategic Priority Research Program of Chinese Academy of Sciences (XDB32030200)
National Key R&D Program of China (Grant No. 2017YFA0105203)
Beijing Municipal Science & Technology Commission (Grant Nos. Z161100000216152, Z171100000117002)
Youth Innovation Promotion Association
Beijing Advanced Discipline Fund
Author Contributions: Y.W. and L.F. designed the research; Y.W. and C.L. performed the research; Y.W., L.F., T. J., B.L., and Z.Y. contributed new regents/analytic tools; Y.W., C.L., D.L., and C.G. analyzed data; Y.W. and L.F. wrote the paper; and L.F., K.H.M., J.R.N., C.C., Y.Z., Z.Y., J.X., C.G., S.S.E., B.L., and T.J. contributed analytic expertise, theoretical guidance, paper revisions, and informed interpretation of the results.
Competing interest: The authors declare no competing interest.
Data and materials availability: R 3.6.1 and custom scripts were used to perform statistical analysis, all R packages were mentioned explicitly in the text where the package was used. The code is freely available at https://github.com/FANLabCASIA/CerebellarGeneFCCorrelation. The ToppGene (https://toppgene.cchmc.org) and CSEA tool (http://genetics.wustl.edu/jdlab/csea-tool-2/) which used to do the functional annotation of genes were all freely accessible. All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary materials.

Figures and Tables

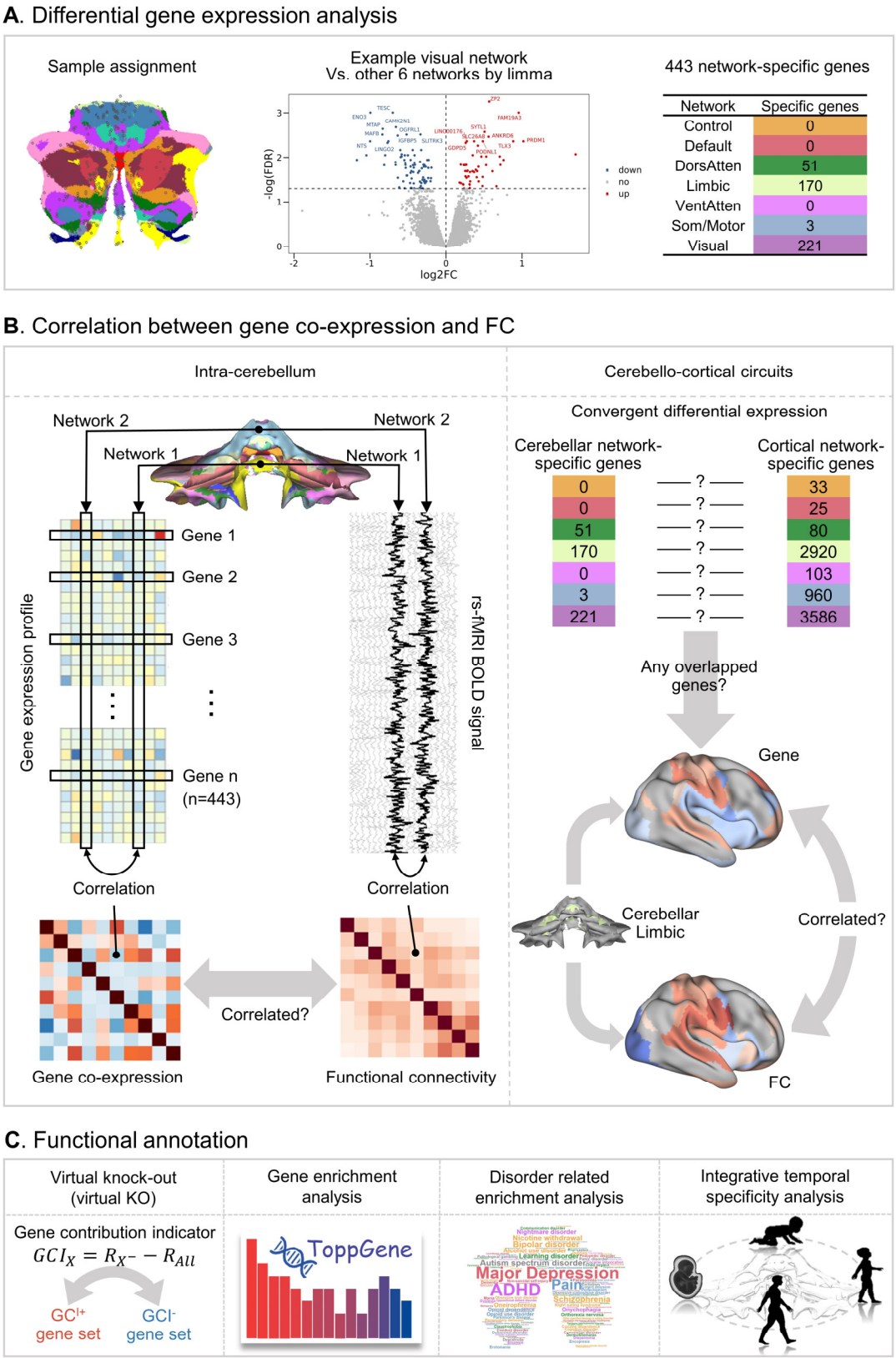

Fig. 1. Analysis pipeline. a Differential gene expression analysis. (A) We assigned the AHBA cerebellar samples into 7 cerebellar functional networks (left) (*25*) and averaged each gene's expression within the same network individually. Then we compared the gene expression in each network with all the other networks by limma (*76*) (middle) with a fold change > 1 and FDR

842 corrected $p < .05$ as an indicator (Red indicates that the genes we found were significantly
843 positively expressed in the visual network.). Thus, we obtained the network-specific genes for 7
844 networks (right). **(B)** Correlations between the gene co-expression and the FC included intra-
845 cerebellar and cerebello-cortical circuits. Intra-cerebellum: for each pair of networks, we
846 calculated the gene expression similarity between them using 443 cerebellar network-specific
847 genes and then constructed the gene co-expression matrix. The FC matrix was constructed by
848 correlating the BOLD signal for all pairs. Then the relationship between the genetic correlation
849 and functional correlation was evaluated. Cerebello-cortical circuits: We first defined the cortical
850 network-specific genes as we had for the cerebellum and tested whether any convergently
851 expressed genes occurred. Then we used the overlapping genes to obtained the cortical genetic
852 correlation for each cerebellar network and evaluated the relationship between the cortical genetic
853 and functional correlation for each cerebellar network. **(C)** Functional annotation includes virtual
854 knock-out (KO), gene enrichment analysis, disorder-related enrichment analysis, and integrative
855 temporal specificity analysis.
856

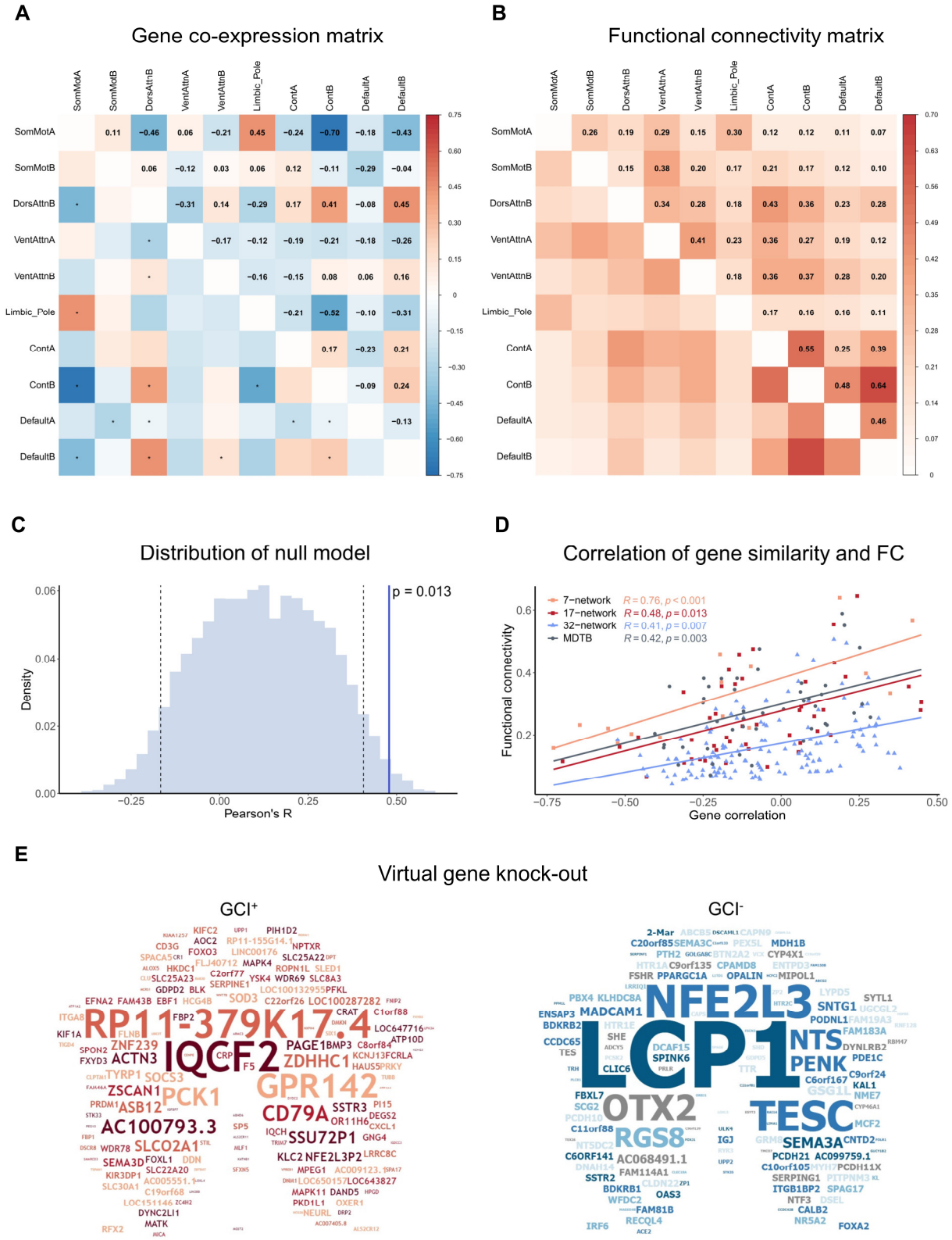

Fig. 2. Network-specific gene co-expression correlates with functional connectivity (FC) within the cerebellum. (A) Genetic correlation was shown by the co-expression matrix (Supplementary sheet 11) constructed for two bi-hemisphere donors across 10 cerebellar networks using 443 cerebellar network-specific genes derived from all six donors. The 10 cerebellar networks corresponded to the networks containing samples from both bi-hemisphere donors (Table S3). * Bonferroni corrected p ≤ .05. **(B)** The FC matrix (Supplementary sheet 12) shows the

functional correlation for the 10 cerebellar networks using 1,018 subjects from the HCP S1200 release (*80*). All passed the Bonferroni corrected significant threshold $p \leq .01$. **(C)** Distribution of the null model constructed using a permutation test that evaluated whether the Gene-FC correlation was generated by chance. The vertical black dashed lines correspond to the p values of .05 and .95; the observed Gene-FC correlation, shown by the blue vertical line, corresponds to $p = .013$. **(D)** The overall intra-cerebellar Gene-FC correlation using different atlases: task-free 7-network (orange), 17-network (red), and 32-network (blue) parcellation of the cerebellar functional atlas based on the cerebello-cortical rsFC, and task-based MDTB functional parcellation (grey) based on the task activation pattern. The Pearson's correlation R and permutation test p values are shown by corresponding colors. **(E)** The GCI^{+} (n = 246, left) and GCI^{-} (n = 197, right) gene list were displayed on flattened shape of the cerebellum.

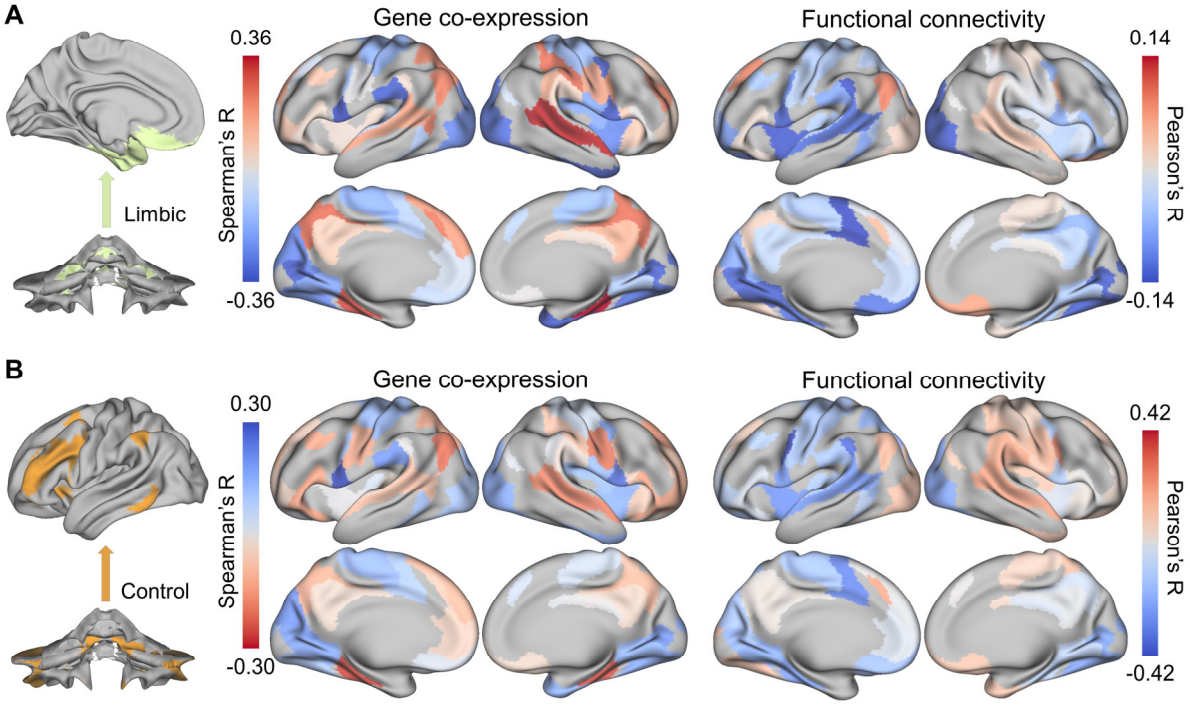

Fig. 3. Cortical genetic and functional correlation of cerebellar limbic and control networks seeds. Both were calculated for 2 bi-hemisphere donors across 6 cerebellar networks and 59 cortical parcels that contained samples from both bi-hemisphere donors. **(A)** Limbic: The cortical gene co-expression (Supplementary sheet 22) was calculated using the 90 overlapping genes between the cerebellar and cortical network-specific genes by Spearman's correlation. The FC across each cerebellar network with each cortical parcel was calculated using Pearson's correlation (Supplementary sheet 23). The cortical limbic genetic and functional correlations were correlated with each other ($r = 0.36$, FDR corrected $p = .030$). **(B)** Control: The cortical gene co-expression and the FC for the control network were correlated with each other ($r = -0.33$, FDR corrected $p = .034$). Noted, the color bar of gene co-expression was inverted considering the negative Gene-FC correlation for the control network.

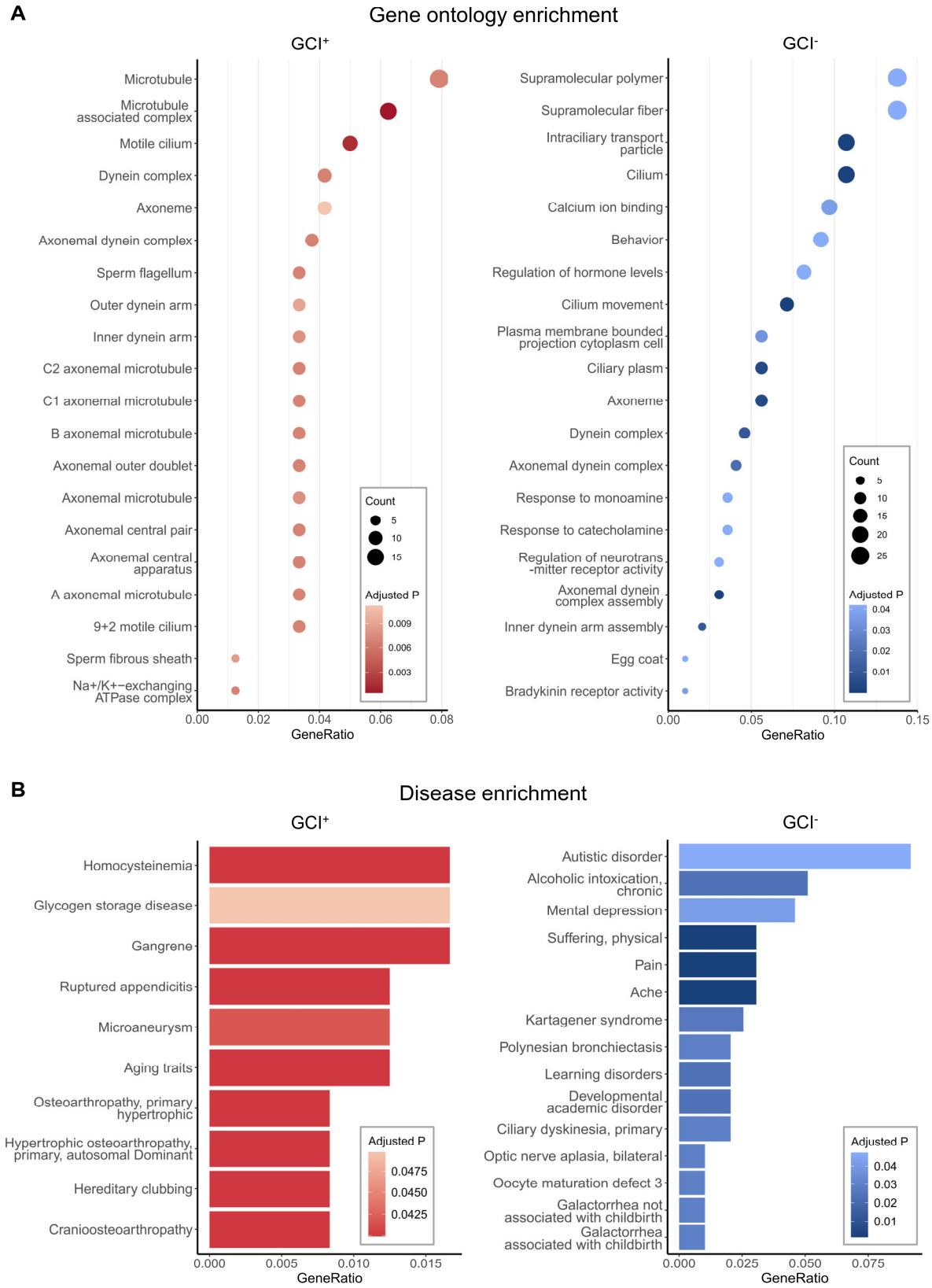

Fig. 4. The gene ontology (GO) and disease enrichment analysis for GCI$^+$ and GCI$^-$. (A) Bubble plot showing the GO enrichment top 20 terms for GCI$^+$ (left) and GCI$^-$ (right) (all results are shown in Supplementary sheets 26 and 27, respectively). The biological process (BP), cellular

component (CC), and molecular function (MF) are displayed together. The dot size (count) represents the number of genes that are within the interest GCI^+ or GCI^- gene panels as well as a specific GO term (y-axis). The different color intensities indicate the FDR corrected p value. **(B)** Gradient barplot showing the disease enrichment for all representative results for GCI^+ and top 15 representative terms for GCI^-. The different color intensities indicate the FDR corrected p value.

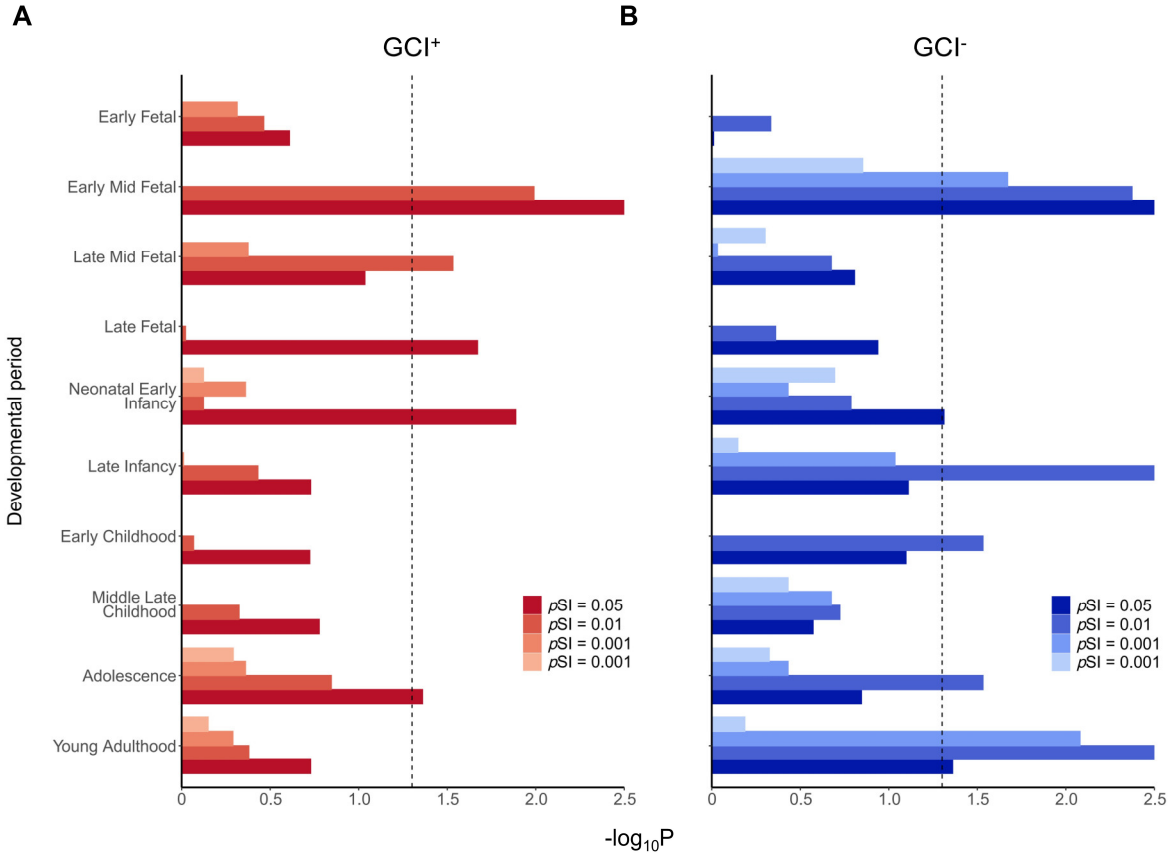

Fig. 5. Integrative temporal specificity analysis of GCI^+ (A) and GCI^- (B) within the cerebellum. The specificity index probability (pSI = .05, .01, .001, and .0001, permutation corrected, shown as different colors) was used to determine how likely a gene was to be expressed in a given time window relative to all other time windows (*36*). The x-axis corresponds to the -$\log_{10}$ (FDR corrected p value), and for aesthetics if -$\log_{10}$ (FDR corrected p value) > 2.5, -$\log_{10}$ (FDR corrected p value) = 2.5; the y axis represents the 10 development windows collected by BrainSpan (*35*). The vertical dark dashed line corresponds to the FDR corrected p = .05. All results are shown in Supplementary sheet 28.

	Cerebellum	Cortex	Overlap genes
Control	0	33	0
Default	0	25	0
Dorsal Attention	51	80	0
Limbic	170	2920	56*
Ventral Attention	0	103	0
SomatoMotor	3	960	2*
Visual	221	3586	33
Total (unique)	443	6987	90

Table 1. Counts of significantly expressed genes within each network compared to other networks, referred to as the network-specific genes. The cerebellar (n = 443, left column, Supplementary sheets 2, 3) and cortical network-specific genes (n = 6987, middle column, Supplementary sheets 7, 8) were defined across the cerebellar (*25*) and cortical (*74*) 7-network strategies. The rightmost column measures the overlap between the cerebellar and cortical network-specific genes for each network (Supplementary sheets 9, 10), *Hypergeometric ps ≤ .01.