Title Page

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

Running Title: Genetic substrates of cerebellar functional organization

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], Bing Liu[6], Kristoffer Hougaard Madsen[1,11,12], Tianzi Jiang[1,2,3,4,5,13], 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] Informatics and Mathematical Modelling, Technical University of Denmark, Kongens Lyngby 2800, Denmark

[12] Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Amager and Hvidovre, Hvidovre 2650, Denmark

[13] School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China

[#] Yaping Wang and Lin Chai contributed equally to this work.

***Corresponding Author**: Lingzhong Fan, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. Email: lingzhong.fan@ia.ac.cn, Phone: 010 - 82544523.

Abstract

Decoding the genetic profiles underlying the cerebellar functional organization is critical for uncovering the essential role of the human cerebellum in various high-order functions and malfunctions in neuropsychiatric disorders. However, no effort has been made to systemically address this. By combining transcriptome data with the intrinsic functional connectivity of the human cerebellum, we not only identified 443 network-specific genes but also discovered that their gene co-expression pattern correlated strongly with intra-cerebellar functional connectivity. Of these genes, 90 were also differentially expressed in the cerebral cortex and linked the cortico-cerebellar cognitive-limbic networks. 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 functional connectivity 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 neurological and neuropsychiatric disorders that are closely linked the cerebellar functional abnormalities. Collectively, our results provide new insight into the genetic substrates behind the functional organization of the human cerebellum with relevance to the possible mechanism of cerebellar contributions to related neurological and psychiatric disorders.

Introduction

Converging evidence from animal and human studies is advancing our understanding of the human cerebellum, which has been shown to be engaged in motor, complex cognitive, and emotional behaviors[1,2]. While such functional diversity of the cerebellum was believed to derive from its extensive afferent and efferent connections to extra-cerebellar structures, rather than being limited to a uniform cerebellar cortical cytoarchitecture[1,3-6]. 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[7-10]. Therefore, unraveling the genetic profiles underlying the cerebellar functional organization could help us understand how the cerebellum organizes different functional subregions that have homogeneous cytoarchitecture into functional networks that support its engagement in various functions[11] as well as increasing our understanding of its relevance in diverse brain diseases[12,13].

However, the genetic mechanism supporting the functional organization of the human cerebellum is largely unknown. Only a few studies have attempted to investigate the genetic expression pattern of the human cerebellum, but they provided inconsistent results in terms of genetic expression variability. For instance, Hawrylycz et al.[14] and Negi and Guda[15] both found that gene expression is highly homogeneous across the anatomical regions of the healthy adult cerebellum. In contrast, Aldinger et al.[16] and Wang and Zoghbi[10] found that cerebellar development and function are

75 governed by the precise regulation of molecular and cellular programs and that the
76 gene expression pattern is heterogeneous across spatial and temporal scales. In
77 addition, differences in gene expression patterns between the cerebellar gyri and
78 sulci[17], and considerable cerebellar regional specializations containing specific cell
79 types, as revealed by high-throughput single-nucleus RNA-seq[18] have been found in
80 the mouse cerebellum. This inconsistency in the genetic variability of the cerebellum
81 needs to be further explored because the relevant studies that showed homogeneity[14,15]
82 only explored the overall cerebellar genetic expression pattern across its gross
83 macro-anatomical boundaries (e.g., cerebellar lobules) and might have failed to fully
84 reflect the functional organization of the human cerebellum[19,20].

85 In the past decade, functional topological maps describing the organization of the
86 human cerebellum using task[21] and task-free functional magnetic resonance imaging
87 (fMRI)[22,23], specifically, separate cerebellar functional networks[22,23] and
88 intra-cerebellar functional gradients[24], have been proposed. In particular, Buckner et
89 al.[22] employed resting-state functional connectivity (rsFC) of the cerebello-cortical
90 circuit as a tool to map the intrinsic functional architecture of the human cerebellum
91 and proposed a possible functional parcellation into 7 networks and 17 networks. It is
92 thus possible to decode the genetic profiles of cerebellar functional organization by
93 investigating the molecular genetic substrates simultaneously linking cerebellar
94 functional heterogeneity and its drivers, i.e., the connections. Whether and how the
95 hypothesized determination of connections in cerebellar functional heterogeneity[6]
96 interact with microscale gene expression is still an open question. To address this, one

promising approach is imaging-transcriptomics analysis[25-27], which allows the brain-wide spatial analysis of microscopic transcriptome data to be combined with macroscopic neuroimaging phenotypes[7].

Thus, our goal was to investigate for the first time the neurobiological genetic mechanism 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[7] was combined with a cerebellar functional parcellation atlas[22] 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). Furthermore, by applying a series of functional annotation tools to these genes (Fig. 1c), we identified two gene sets separately involved in cerebellar neurodevelopment and neurotransmission and obtained interesting genetic evidence supporting the implications of cerebellar functional organization in many neurological and psychiatric disorders. 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 many neurological and neuropsychiatric disorders.

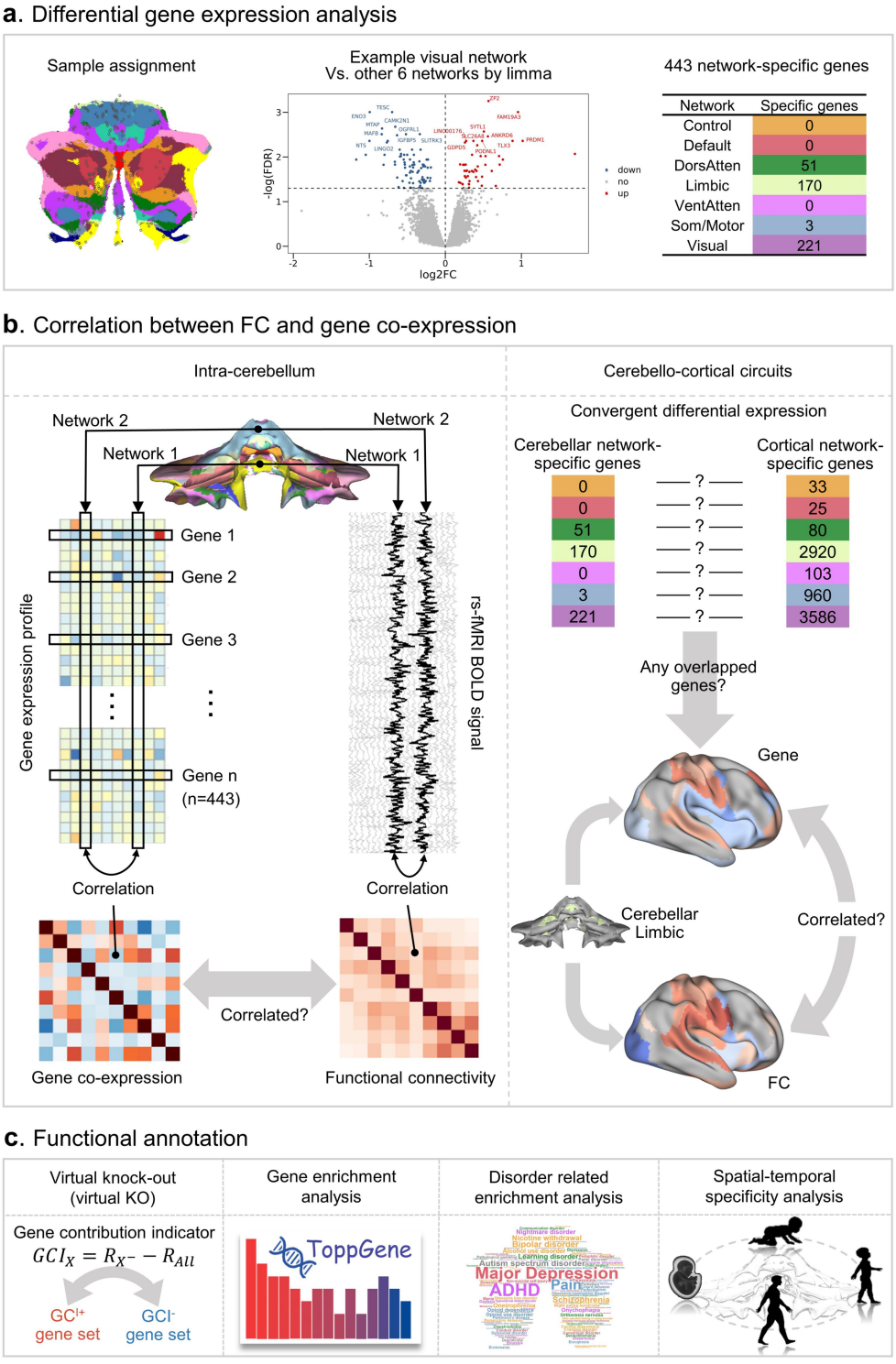


Fig. 1 | Analysis pipeline. a Differential gene expression analysis. We assigned the AHBA cerebellar samples into 7 cerebellar functional networks (left)[22] and averaged each gene's expression within the same network individually. Then we compared the gene expression in each network with all the other networks by limma[28] (middle) with

a fold change > 0 and FDR corrected $p < .05$ as an indicator (Red indicates that the genes we found were significantly positively expressed in the visual network.). Thus, we obtained the network-specific genes for 7 networks (right). **b** Correlations between the gene co-expression and the FC included intra-cerebellar and cerebello-cortical circuits. Intra-cerebellum: for each pair of networks, we calculated the gene expression similarity between them using 443 cerebellar network-specific genes and then constructed the gene co-expression matrix. The FC matrix was constructed by correlating the BOLD signal for all pairs. Then the relationship between the genetic correlation and functional correlation was evaluated. Cerebello-cortical circuits: We first defined the cortical network-specific genes as we had for the cerebellum and tested whether any convergently expressed genes occurred. Then we used the overlapping genes to obtained the cortical genetic correlation for each cerebellar network and evaluated the relationship between the cortical genetic and functional correlation for each cerebellar network. **c** Functional annotation includes virtual knock-out (KO), gene enrichment analysis, disorder-related enrichment analysis, and spatial-temporal specificity analysis.

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 sheet 2, 3) using all samples from 6 donors

148 across 7 networks. The distribution of these network-specific genes is shown in Table
149 1, which shows that these were mainly expressed in the limbic (n = 170), dorsal
150 attention (n = 51), somato/motor (n = 3), and visual (n = 221) networks. We also
151 obtained 6,987 cortical network-specific genes (Supplementary sheet 5, 6, Table 1)
152 using the same strategy and found that the cerebellar and cortical network-specific
153 genes distribution patterns were highly correlated (r = 0.95, p = .00108).

154 Moreover, we found that 90 of these 443 cerebellar network-specific genes (~
155 20%) (Supplementary sheet 7, 8, Table 1) were convergently expressed in the cerebral
156 cortex and that a significant overlap between the cerebellar and cortical
157 network-specific genes of the limbic and somatomotor networks occurred (limbic
158 overlap = 56, hypergeometric ps < .0001; somatomotor overlap = 2, hypergeometric
159 ps < .01). This means that the 56 limbic genes were differentially expressed in the
160 limbic cortex and the limbic cerebellum and that the 2 somatomotor genes were
161 differentially expressed in the somatomotor cortex and somatomotor cerebellum.
162 Overlapped genes were also found in the visual network but failed to pass the
163 hypergeometric test (visual overlap = 33, with hypergeometric ps = .84), and no
164 overlap was found for the other 4 networks (ventral attention, dorsal attention, control,
165 default, Supplementary sheet 7).

	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 network-specific genes for simplicity). Here we defined the cerebellar (n = 443, left column, Supplementary sheets 2, 3) and cortical network-specific genes (n = 6987, middle column, Supplementary sheets 5, 6) across the cerebellar[22] and cortical[29] 7-network strategies. The rightmost column measures the overlap between the cerebellar and cortical network-specific genes for each network (Supplementary sheets 7, 8), * Hypergeometric ps ≤ .01.

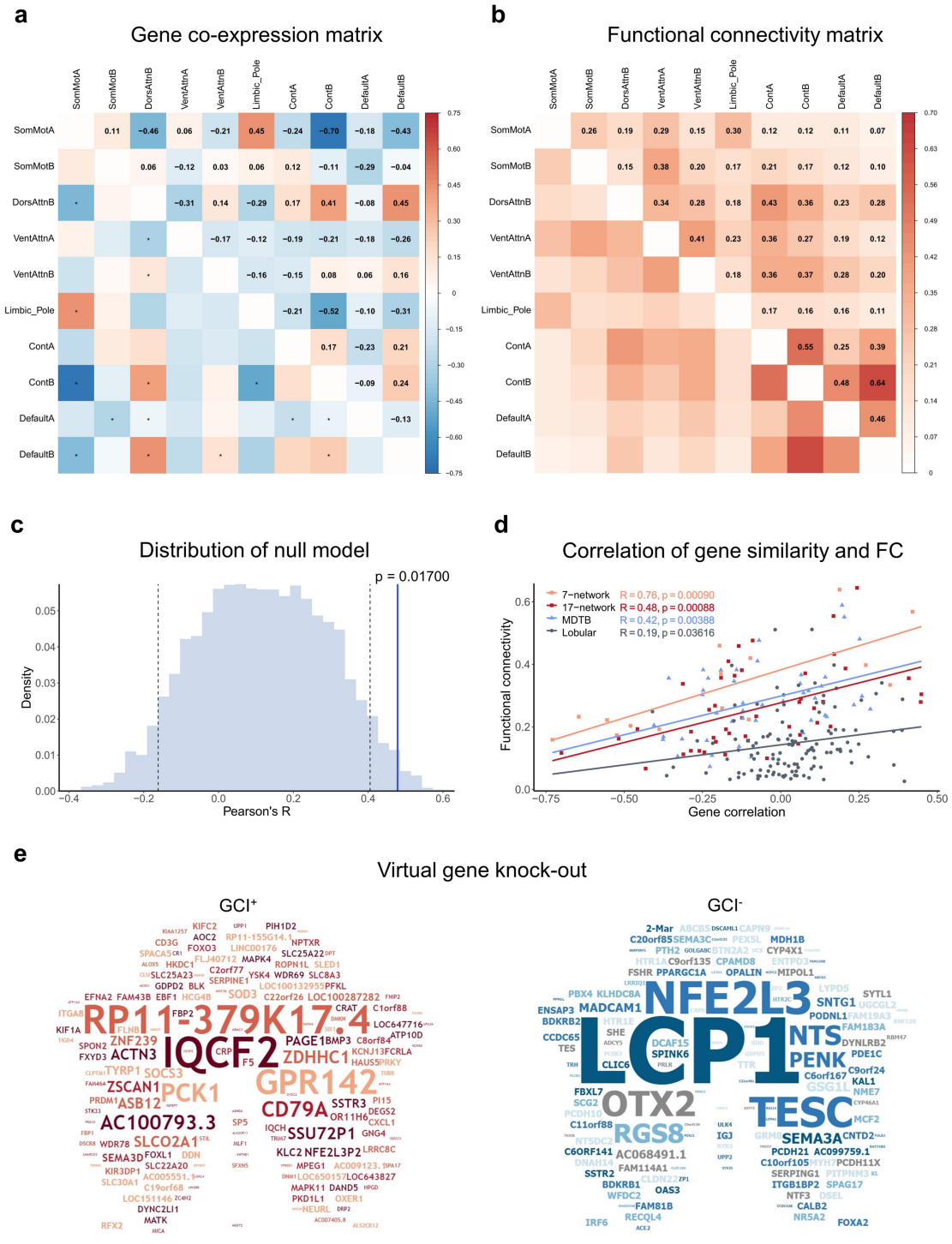
The co-expression of the cerebellar network-specific genes highly correlated with intra-cerebellar FC

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. Across all the available network-network pairs, the genetic co-expression correlates with the FC within the cerebellum (r = 0.48, p = .00088, with permutation test p = .01700, Fig. 2). The genetic correlations were either positive or negative, but the FCs were all positive, and the negative genetic correlation corresponded to a mild functional correlation (Fig. 2d, red). This correlation between gene co-expression and FC was referred to as

Gene-FC correlation throughout present paper for simplicity. To validate the Gene-FC correlation within the cerebellum, we also leveraged the task-free 7-network parcellation, task-based multi-domain task battery (MDTB) functional parcellation[21], and the cerebellar lobular parcellation[30] 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, p = .00090, Fig. 2d, Supplementary Fig. 1), the MDTB functional parcellation (r = 0.42, p = .00388, Fig. 2d, Supplementary Fig. 2), and the cerebellar lobular parcellation (r = 0.19, p = .03616, Fig. 2d, Supplementary Fig. 3). The Gene-FC correlation for the lobular parcellation, however, failed to pass the Bonferroni corrected significance level (p < .05). This is consistent with the observation that, compared with cerebellar morphological boundaries, a functional atlas performs better in terms of functional representativeness[19,20].

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 a different parcellation resolution and independent cerebellar functional atlas although it disappeared in the lobular parcellation. Moreover, the control test exhibited no Gene-FC correlation when the gene co-expression was constructed using non-network-specific genes (Supplementary sheet 28) regardless of whether the test was thresholdless or thresholded. These findings further confirmed that these 443 network-specific genes play a key role in

206 intra-cerebellar functional organization.



207

208 **Fig. 2 | Network-specific gene co-expression correlates with functional**
209 **connectivity (FC) within the cerebellum. a** Genetic correlation was shown by the
210 co-expression matrix (Supplementary sheet 9) constructed for two bi-hemisphere
211 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 (Supplementary Table 3). Genetic correlation revealed both positive (red) and negative (blue) correlations, * Bonferroni corrected $p \leq .05$. **b** The FC matrix (Supplementary sheet 10) shows the functional correlation for the 10 cerebellar networks using 1,018 subjects from the HCP S1200 release[31]. They were positively correlated with each other, and all passed the Bonferroni corrected significant threshold $p \leq .0001$. **c** Distribution of the null model constructed using a permutation test that evaluated whether our Gene-FC correlation was generated by chance. The vertical black dashed lines correspond to the p values of .05 and .95; our observed Gene-FC correlation, shown by the blue vertical line, corresponds to $p = .01700$. **d** The overall intra-cerebellar Gene-FC correlation using different parcellations: task-free 7-network (orange) and 17-network (red) parcellation of the cerebellar functional atlas based on the cerebello-cortical rsFC, task-based MDTB functional parcellation (blue) based on the task activation pattern, and cerebellar lobular parcellation (grey). The Pearson's correlation R and 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.

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 $p = .03026$, Fig. 3a) and control networks ($r = -0.33$, FDR corrected $p = .03449$, Fig. 3b) but was not significant for the somatomotor ($r = -0.15$, FDR

corrected p = .39433), dorsal attention (r = -0.19, FDR corrected p = .28134), ventral attention (r = -0.04, FDR corrected p = .77856), or default (r = 0.10, FDR corrected p = .54382) networks. The high cortical genetic similarity between the limbic system and the adjacent control network (r = -0.90, FDR corrected p < .0001), somatomotor network (r = -0.55, FDR corrected p < .0001), and ventral attention network (r = -0.72, FDR corrected p < .0001) 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 = .31596), 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.

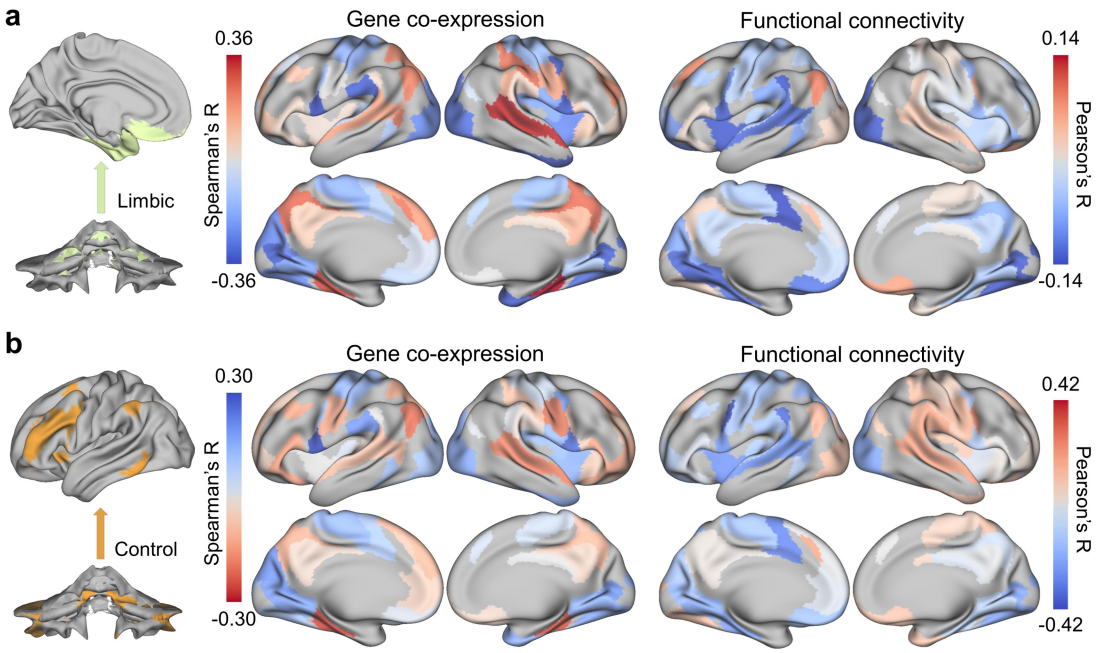


Fig. 3 | Genetic and functional cortical correlation of limbic and control cerebellar networks seeds. Both were calculated for 2 bi-hemisphere donors across 10 cerebellar networks and 59 cortical parcels that contained samples from both bi-hemisphere donors. **a** Limbic: The cortical gene co-expression (Supplementary sheet 11) 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 12). The cortical limbic genetic and functional correlations were correlated with each other (r = 0.36, FDR corrected p = .03026). **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 = .03449). Noted, the color bar of gene co-expression was inverted considering the negative Gene-FC correlation for the control network.

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 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 13) and a 197 GCI negative gene set (GCI$^-$, Fig. 2e right, Supplementary sheet 14). 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, FDR corrected p = .00050), motile cilium (ID: 0031514, FDR corrected p = .00156), and dynein complex (ID:

GO:0030286, FDR corrected p = .00678). 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, FDR corrected p = .03709), regulation of hormone levels (ID: 0010817, FDR corrected p = .04195), response to catecholamine (ID: 0071869, FDR corrected p = .04195), response to monoamine (ID: 0071867, FDR corrected p = .04195), and regulation of neurotransmitter receptor activity (ID: 0099601, FDR corrected p = .04195). This is consistent with their different pathway enrichment results (Supplementary sheets 15,16) in that the GCI^+ was primarily enriched in some basic biological pathways: proximal tubule bicarbonate reclamation (ID: M4361, FDR corrected p = .03197) and glycolysis/gluconeogenesis (ID: M39474, FDR corrected p = .03197), 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, FDR corrected p = .03877).

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 an involvement of GCI^- in various neurological and neuropsychiatric disorders (Fig. 4b, right), including autistic disorder (ID: C0004325, FDR corrected p = .04734), alcoholic intoxication (ID: C0001973, FDR corrected p = .02349), mental depression (ID: C0011570, FDR corrected p = .04167),

312 pain (ID: C0030193, FDR corrected p = .00141), learning disorders (ID: C0023186,
313 FDR corrected p = .02349) and others. Many of these, especially mental depression
314 and autistic disorder, have a close relationship with the human cerebellum, in which
315 patients have shown functional connectivity abnormalities[33,34]. The mental
316 depression- and autistic disorder-associated genes were TRH, PENK, TTR, ADCY5,
317 NRXN1, HTR1A, HTR2C, NTS, PEX5L (n = 9, Supplementary sheet 16) and
318 DLGAP2, TRH, PENK, RYR3, SEMA3A, NRXN1, TESC, ABCG2, PCDH10,
319 CNTN4, HTR1A, CALB2, HTR2C, DNAAF4, FOLR1, NTS, GRM8, UPP2 (n = 18,
320 Supplementary sheet 16), respectively, and the overlapping genes were TRH, PENK,
321 NRXN1, HTR1A, HTR2C, NTS (n = 6).

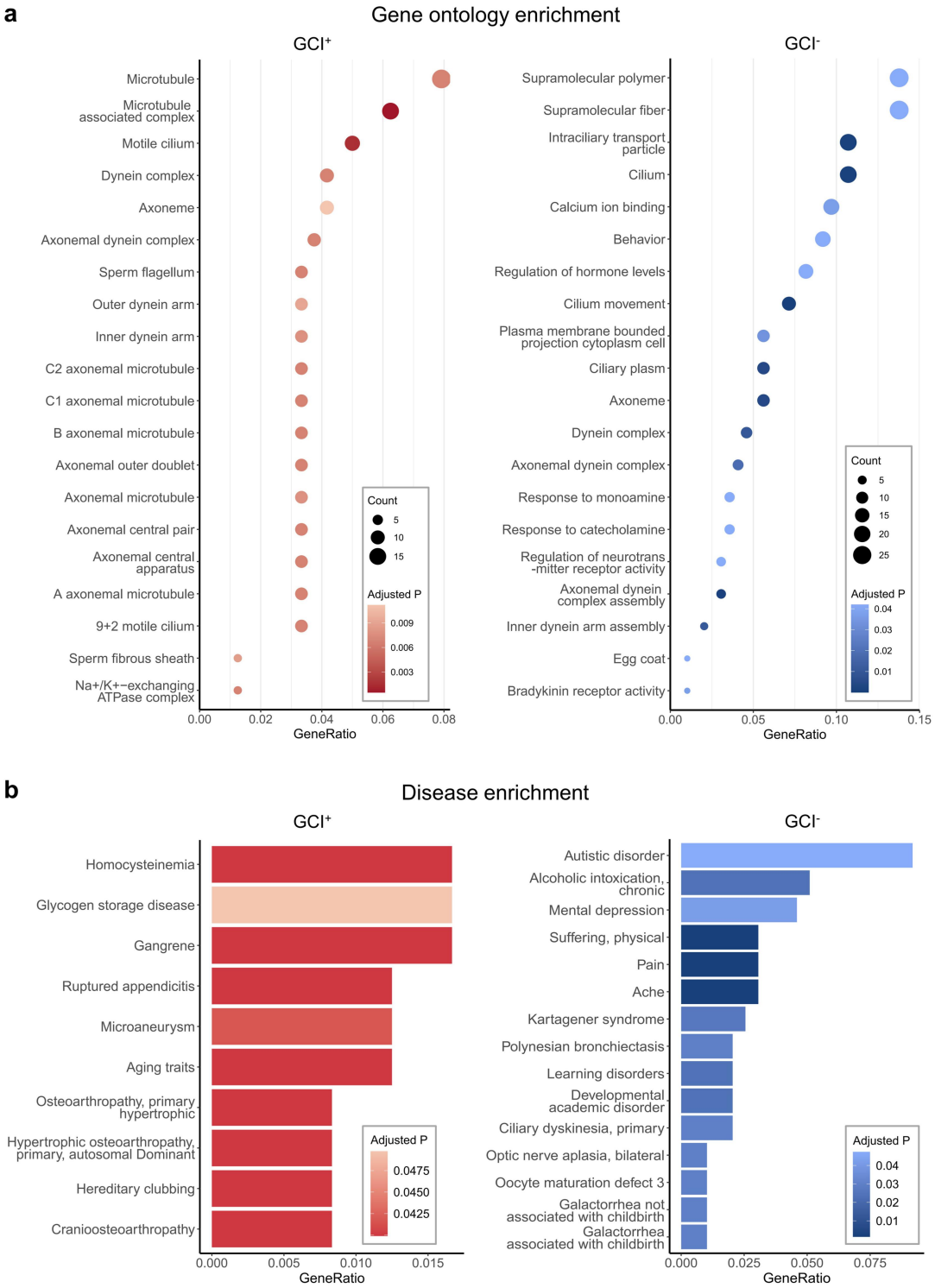


Fig. 4 | The gene ontology (GO) and disease enrichment analysis for GCI$^+$ and GCI$^-$. a Bubble plot shows the GO enrichment top 10 terms for GCI$^+$ (left) and GCI$^-$ (right) (all results are shown in Supplementary sheets 15 and 16, respectively). The biological process (BP), cellular component (CC), 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 color shows 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 color represents the FDR corrected *p* value.

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, and neonatal early infancy 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 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.

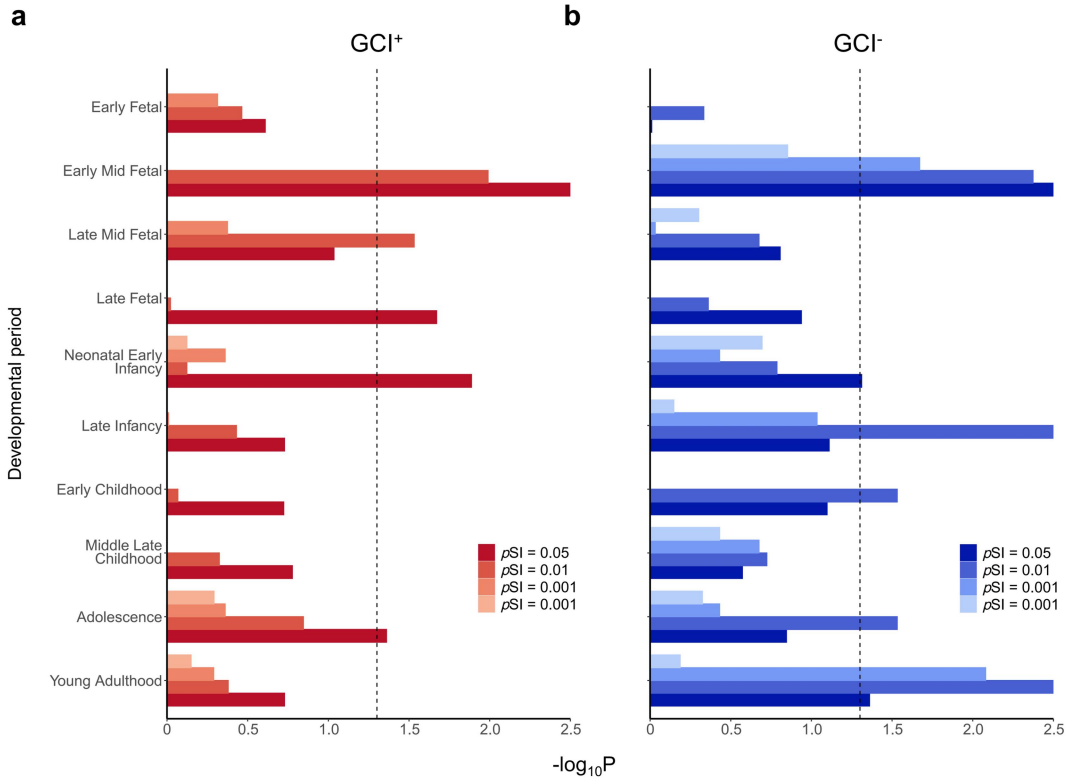


Fig. 5 | Integrative spatial-temporal specificity analysis of GCI$^+$ (a) and GCI$^-$ (b) within the cerebellum. The specificity index probability (*p*SI = .05, .01, .001, and .0001, permutation corrected, shown as different colors) was used to determine how likely a gene was to be expressed in a given time window relative to all other time windows[36]. The x-axis corresponds to the -log10 (FDR corrected *p* value), and for aesthetics if -log10 (FDR corrected *p* value) > 2.5, -log10 (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 17.

Discussion

The current study provided a first 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 neurological and psychiatric disorders, which hints at a possible mechanistic explanation for the cerebellar contributions to related neurological and psychiatric 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 and 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[7] 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[22]. The cerebellar network-specific genes we obtained are not in keeping with the highly homogeneous gene expression within the human cerebellum suggested by its anatomic atlas[7,14]. Even though we selected a definition of differentially expressed genes using an FDR corrected statistical threshold rather than an arbitrary threshold and although lobule-specific genes were identified using our statistical threshold, these genes did not correlate with the FC of the human cerebellum. This indicates that the arbitrary fold change threshold was not appropriate for determining biologically meaningful but subtle differences[42] and that the genes underlying the lobular segregation are not related to the resting-state activity of the human cerebellum. These findings support, from a genetic perspective, the idea that the morphological subdivisions of the cerebellum do not correspond well to its functional representation[19,20].

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[24,43]. These correlated patterns may be related to the way that we defined the cerebellar network, which was by projecting the cerebral cortical networks onto the cerebellum by computing the functional connections between the

two regions[22]. More interestingly, the molecular genetic substrates simultaneously linking functional heterogeneity and integration could only be observed across the functional subdivision, regardless of whether the parcellation was based on the task-free cerebello-cortical rsFC[22] or the intra-cerebellar task-based activation pattern[21], but disappeared in the lobular parcellation. These interpretations are further supported by the widely accepted notion about the human cerebellum that its functional specialization is dominated by its connection with extracerebellar structures rather than within its homogeneous cytoarchitecture[6]. 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[6]. This proposed relationship between extra- and intra-cerebellar connectivity can in turn be expected to affect the resting-state activity between cerebellar regions[24].

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[46]. For instance, the integrated processing by the emotion and cognition areas has been identified solely based on their anatomical connections[47]. 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[48]. 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[49] 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[27]. Considering the large differences between the cerebellum and cortex in terms of their gene expression patterns[14] and structure-function relationships[50] as well as the individual variability of their functional networks[51], 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[52].

Cerebellar neurodevelopment features of GCI$^+$, cerebellar neurotransmission, neurological, 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

446 evidence for GCI$^+$ and GCI$^-$ involvement in cerebellar neurodevelopment and
447 cerebellar neurotransmission, respectively. It is also interesting to speculate that these
448 443 network-specific genes that link both cerebellar functional segregation and
449 integration have a relationship with some brain-related disorders since prior evidence
450 showed that the cerebellar functional organization plays a key role in various
451 neurological[13,53] and neuropsychiatric disorders[12], most of which possess common
452 underlying genetic risks[54]. But a tricky problem emerged in that the genes we are
453 interested in were derived from healthy individuals. This could be tackled to some
454 extent by using the virtual KO method, which can simulate the different expression
455 levels of each gene and thus coarsely corresponds to a fraction of the expression level
456 under normal health and disease situations. This is why we thought that we might be
457 able to see whether the GCI$^+$ and GCI$^-$ are related to a specific disease even though
458 the genes were derived from healthy individuals.

459 The GCI$^+$ is involved in many microtubule-related terms and is overexpressed
460 throughout the protracted development of the cerebellum. The dynamics and
461 flexibility of microtubules were found to be essential throughout cerebellar
462 development because they affect the morphological alterations of Purkinje cells[38]. In
463 addition, some genes of the GCI$^+$, such as GTPBP2[55] and Lin28b[56], were found to
464 play a key role in neurodevelopment; overexpression of the Lin28b gene can induce
465 the development of pathological lobulation in the cerebellum[56]. This converging
466 evidence prompts our speculation that the GCI$^+$ is engaged in cerebellar
467 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[14], 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[57] and neuropsychiatric disorders[58,59]. 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 became potential treatment targets[60]. The time 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[61] and autism[62].

Moreover, we found that the GCI$^-$ was enriched in many neurological and 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[63]; the cerebellum is one of the brain regions most sensitive to the harmful effects of chronic alcohol abuse[64], and the cerebello-cortical FC of patients with alcohol use disorder has been shown to have changes in both flexibility and integration[65]. Therefore, the GCI$^-$ provides a possible micro-macro interacted mechanistic explanation for the engagement of the cerebellum in various neurological and neuropsychiatric disorders; i.e., one of the ways these risk genes play a role in the pathogenesis of corresponding diseases may be through their interactions with the cerebellar FC, which results in pathological manifestations as abnormalities in cerebellar functional connectivity, such as the fluctuation in the correspondence of the Gene-FC relationship found in the present study. The GCI$^-$ also provides a promising genetic resource for investigating the cerebellar engagement 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

512 studies showing that rare and common variants in NRXN1 carried risks for MDD[66],
513 ASD, and schizophrenia[67], and HTR1A, which has a high expression in the
514 cerebellum[15], was found to be involved in pain, mental depression, autistic disorder,
515 alcoholic intoxication, learning disorder, and other conditions.

516 ***Limitations***

517 The interpretation of our findings has several caveats. First, the AHBA dataset itself
518 has many shortcomings, although it provides an unprecedented opportunity to
519 combine brain imaging data with genetic information. The AHBA gene expression
520 data was obtained using microarray technology, which did not include the expression
521 of non-encoding RNA (such as snRNA and microRNA) and lacks cellular level
522 information because it averaged a variety of cell types within a single sample. The
523 overall pattern of gene expression, gene regulation, epigenomics, and improved
524 cellular resolution is helpful for fully understanding the causal relationship between
525 genes and functional organization, which is a greater challenge for neuroscience than
526 just identifying a link between genetic and imaging data. Second, the gene
527 co-expression we constructed only considered one small part of the relationship
528 between the genes and FC thereby it did not fully recapitulate the complexity of the
529 brain transcriptome, such as gene-gene interactions[68]. That is one possible reason why
530 we only found a cerebello-cortical Gene-FC correlation for the cognitive-limbic
531 networks. Last, simple correlation approaches, such as used in this study, are only
532 able to prioritize genes for further investigation and cannot fully explore the

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. Nevertheless, in light of the current limited understanding of the details about how genes contribute to large-scale functional organization, the prioritization of genes and the related functional annotation presented here are still necessary and important[25].

Conclusions

Overall, 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 – disorder" loop as well as of bridging the knowledge gap

554 between the genetic mechanisms driving the cerebellar functional organization and
555 the heritable risks of disorders, especially major depression and autistic disorder. The
556 current study also prioritizes genes for future studies that will focus on the genetic
557 correlates of the cerebellar functional organization, the genetic implications of
558 cerebellar malfunction in the pathogenesis of many neurological and mental disorders,
559 and future genetic treatment targets for the cerebellar functional abnormalities of these
560 disorders.

561 # Materials and Methods

562 ## *AHBA preprocessing*

563 The AHBA[7] is a publicly available transcriptome dataset (http://www.brain-map.org),
564 which provides normalized microarray gene expression data from six adult donors
565 (ages 24, 31, 34, 49, 55, and 57 years; n = 4 left hemisphere only, n = 2 both left and
566 right hemispheres). Supplementary table 1 shows the demographic information.

567 The preprocessing pipeline was referred to in Anderson et al.[27], and included data
568 filtering, probe selection, sample selection, and assignment. We first filtered the
569 probes with the AHBA binary indicator to mitigate the background noise and
570 excluded probes without an Entrez ID. Then for the genes that corresponded to two or
571 more probes, we chose the probe with the maximum summed adjacency to represent
572 the corresponding gene expression; otherwise the probe with the highest mean
573 expression was retained, using the CollapseRows function[69] in R. The first two steps

574 generated 20,738 unique mRNA probes, which provided expression data for 20,738
575 genes. As suggested by Arnatkeviciute et al.[70] and given the known transcriptional
576 differences[14] between the cortical and sub-cortical regions and the cerebellum, we
577 separated the cortical and cerebellar samples a priori based on the slab type and
578 structure name provided by AHBA and processed them separately later. In the end,
579 337 samples were retained for the cerebellar cortex and 1,701 samples for the cortical
580 cortex.

581 Finally, we respectively assigned these 337 cerebellum samples and 1,701
582 cortical samples into the cerebellar functional network atlas[22] and cortical functional
583 networks atlas[29], both of which have 7- and 17-network parcellation strategies. For
584 each cerebellar sample, we first generated a single $1 \times 1 \times 1$ mm^3 region of interest
585 (ROI) at the MNI coordinate for each sample using AFNI[71]. The network label from
586 either region 7 or 17 was assigned, if the ROI fell within a cerebellar network of the
587 Buckner atlas. Considering the uneven and discrete sampling of the AHBA data[7], if
588 the $1 \times 1 \times 1$ mm^3 ROI did not overlap with any network, the associated ROI was
589 expanded to $3 \times 3 \times 3$ mm^3, and if the $3 \times 3 \times 3$ mm^3 ROI overlapped with the
590 functional atlas, the network that had the maximum number of shared voxels with the
591 ROI was assigned. Otherwise, the steps above were repeated for a $5 \times 5 \times 5$ mm^3 ROI.
592 The cerebellar samples were excluded (n = 22) if the $5 \times 5 \times 5$ mm^3 ROI did not
593 overlap with any cerebellar networks. Supplementary tables 2 and 3 show the
594 distributions of the cerebellar sample assignment for the 7-network and 17-network
595 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 Supplementary tables 4 and 5.

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, resulting in 20738 genes × 7 or 17 network matrices for each donor. Then we calculated the differential gene expression across the 7 networks using the R limma package[28] 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[42]. 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 > 0 was used as the key indicator for differentially expressed genes. The residual donor effects were accounted for by using limma's duplicateCorrelation tool[28]. 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)[29] and then averaged within each network.

Cerebellar resting-state functional connectivity (rsFC)

The minimally preprocessed[72,73] Human Connectome Project (HCP) S1200 release dataset[31], which has 1,018 subjects 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[74,75]. 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[22], separately. The rsFC within the cerebellum was computed using the 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

637 fine-grained 17-network parcellation to increase the spatial resolution. The only
638 exception was that the cerebellar 7-network was applied while calculating the FC
639 across the cerebello-cortical circuit to compare each cerebellar network more directly.

640 ***Correlation between gene co-expression with intra-cerebellar rsFC***

641 To fully capture the genetic correlation with the FC within the cerebellum, we
642 leveraged the genetic samples of the 2 bi-hemisphere donors when constructing the
643 gene co-expression matrix because the rsFC of the cerebellum is bilateral. Therefore,
644 the gene co-expression was analyzed for the 2 bi-hemisphere donors using the 443
645 differentially expressed genes derived from all 6 donors across 7 networks, using a
646 finer 17-network parcellation to increase the spatial resolution. Ten networks that
647 contained samples from both bi-hemisphere donors were retained (Table S3). For each
648 bi-hemispheric donor, the log2 gene expression of the cerebellar samples was
649 mean-normalized and then averaged within each network. The cerebellar 10 × 10
650 networks correlation matrix was calculated using the Spearman's correlations
651 individually, then Fisher transformed, and finally averaged to construct the final 10
652 networks gene co-expression matrix. The correlation significance level of the gene
653 co-expression was evaluated using the overlap between the correlation matrix for
654 these two individuals and adjusted by Bonferroni correction. Meanwhile, we
655 transformed the 17 × 17 networks rsFC matrix into a 10 × 10 networks size to be
656 consistent with the gene co-expression matrix. Finally, the relationship between the 10
657 × 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 labels 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. 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., a task-free 7-network parcellation, independent task-based multi-domain task battery (MDTB) functional parcellation[21], and cerebellar lobular parcellation[30]. 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.

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

682 To fully investigate the cerebellar functional organization, we also explored the
683 relationship between the cerebello-cortical FC and the genetic correlation based on
684 the strategy used in Anderson et al.[27]. First, we defined the network-specific genes in
685 the cortex using the same procedure as we had for the cerebellum and examined the
686 genes that overlapped within the same network of the cerebellum and the cortex using
687 a hypergeometric test. Then the gene co-expression matrix was constructed between 6
688 cerebellar networks and 59 cortical parcels from the 2 bi-hemisphere donors, using
689 the 90 unique genes derived from the overlap between the cortical network-specific
690 genes and the cerebellar network-specific genes. Here, the cerebellar 7-network
691 parcellation was selected to compare the different cerebellar networks more directly.
692 The visual network was excluded because it only had two samples that were solely
693 from one of the 2 bi-hemisphere donors. For the cerebral cortex, 59 cortical parcels
694 that contained samples from both bi-hemisphere donors were estimated. The log2
695 mean-normalized expression within each cerebellar network and each cortical parcel
696 was estimated individually and correlated using Spearman's ρ, Fisher-transformed,
697 and averaged. We transformed the 114 cortical parcels × 7 cerebellar networks rsFC
698 matrix into 59 cortical parcels × 6 cerebellar networks size to be consistent with the
699 gene co-expression matrix. Finally, the relationship between the cortical genetic
700 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[76,77] 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). 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

721 portal[78] to conduct a gene ontology (GO), pathway, and disorder enrichment analysis.
722 The Benjamini-Hochberg method for false discovery rate (FDR-BH correction) (q
723 < .05) was chosen to correct for multiple comparisons.

724 *Spatial-Temporal Analysis*

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

735 ## Acknowledgments

736 This work was partially supported by the Natural Science Foundation of China (Grant
737 Nos. 82072099, 91432302, and 31620103905), the Strategic Priority Research
738 Program of Chinese Academy of Sciences (XDB32030200), National Key R&D
739 Program of China (Grant No. 2017YFA0105203), Beijing Municipal Science &
740 Technology Commission (Grant Nos. Z161100000216152, Z171100000117002), the
741 Youth Innovation Promotion Association, and the Beijing Advanced Discipline Fund.

742 The authors appreciate the English language and editing assistance of Rhoda E. and
743 Edmund F. Perozzi, PhDs.

744 Conflict of Interest

745 The authors declare that the research was conducted in the absence of any commercial
746 or financial relationships that could be construed as a potential conflict of interest.

747 Supplementary Material

748 Shown in supplementary figures and supplementary sheets.

749 Data availability

750 R 3.6.1 and custom scripts were used to perform statistical analysis, all R packages
751 were mentioned explicitly in the text where the package was used. The analysis code
752 is freely available (https://github.com/FANLabCASIA/CerebellarGeneFCCorrelation).
753 The ToppGene website (https://toppgene.cchmc.org/) and CSEA tool
754 (http://genetics.wustl.edu/jdlab/csea-tool-2/) which used to do the functional
755 annotation of genes were all freely accessible. A Supplementary Data file provides
756 complete gene lists, the output of differential expression, rs-fMRI, genetic correlation,
757 validation results, and functional annotation results.

References

1 Schmahmann, J. D., Guell, X., Stoodley, C. J. & Halko, M. A. The Theory and Neuroscience of Cerebellar Cognition. *Annu Rev Neurosci* **42**, 337-364, doi:10.1146/annurev-neuro-070918-050258 (2019).

2 De Zeeuw, C. I., Lisberger, S. G. & Raymond, J. L. Diversity and dynamism in the cerebellum. *Nat Neurosci* **24**, 160-167, doi:10.1038/s41593-020-00754-9 (2021).

3 Sathyanesan, A. *et al.* Emerging connections between cerebellar development, behaviour and complex brain disorders. *Nat Rev Neurosci* **20**, 298-313, doi:10.1038/s41583-019-0152-2 (2019).

4 Guell, X., Schmahmann, J. D. & Gabrieli, J. D. E. Functional Specialization is Independent of Microstructural Variation in Cerebellum but Not in Cerebral Cortex. *bioRxiv*, 424176, doi:10.1101/424176 (2018).

5 Voogd, J. & Glickstein, M. The anatomy of the cerebellum. *Trends in Cognitive Sciences* **2**, 307-313, doi:https://doi.org/10.1016/S1364-6613(98)01210-8 (1998).

6 Schmahmann, J. D. The role of the cerebellum in affect and psychosis. *Journal of Neurolinguistics* **13**, 189-214, doi:https://doi.org/10.1016/S0911-6044(00)00011-7 (2000).

7 Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399, doi:10.1038/nature11405 (2012).

8 Johnson, M. B. *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494-509, doi:10.1016/j.neuron.2009.03.027 (2009).

9 Guevara, E. E. *et al.* Comparative analysis reveals distinctive epigenetic features of the human cerebellum. *PLoS Genet* **17**, e1009506, doi:10.1371/journal.pgen.1009506 (2021).

10 Wang, V. Y. & Zoghbi, H. Y. Genetic regulation of cerebellar development. *Nat Rev Neurosci* **2**, 484-491, doi:10.1038/35081558 (2001).

11 Schmahmann, J. D. The cerebellum and cognition. *Neurosci Lett* **688**, 62-75, doi:10.1016/j.neulet.2018.07.005 (2019).

12 Phillips, J. R., Hewedi, D. H., Eissa, A. M. & Moustafa, A. A. The cerebellum and psychiatric disorders. *Front Public Health* **3**, 66, doi:10.3389/fpubh.2015.00066 (2015).

13 Diwakar, S. Cerebellum in Neurological Disorders: A Review on the Role of Inter-Connected Neural Circuits. *Journal of Neurology & Stroke* **6**, doi:10.15406/jnsk.2017.06.00196 (2017).

14 Hawrylycz, M. *et al.* Canonical genetic signatures of the adult human brain. *Nat Neurosci* **18**, 1832-1844, doi:10.1038/nn.4171 (2015).

15 Negi, S. K. & Guda, C. Global gene expression profiling of healthy human

799 brain and its application in studying neurological disorders. *Sci Rep* **7**, 897,
800 doi:10.1038/s41598-017-00952-9 (2017).
801 16 Aldinger, K. A. *et al.* Spatial and cell type transcriptional landscape of human
802 cerebellar development. *Nat Neurosci*, doi:10.1038/s41593-021-00872-y
803 (2021).
804 17 Zeng, T. *et al.* Allen mouse brain atlases reveal different neural connection and
805 gene expression patterns in cerebellum gyri and sulci. *Brain Struct Funct* **220**,
806 2691-2703, doi:10.1007/s00429-014-0821-x (2015).
807 18 Kozareva, V. *et al.* A transcriptomic atlas of the mouse cerebellum reveals
808 regional specializations and novel cell types. *bioRxiv*, 2020.2003.2004.976407,
809 doi:10.1101/2020.03.04.976407 (2020).
810 19 Ren, Y., Guo, L. & Guo, C. C. A connectivity-based parcellation improved
811 functional representation of the human cerebellum. *Sci Rep* **9**, 9115,
812 doi:10.1038/s41598-019-45670-6 (2019).
813 20 Bernard, J. A. *et al.* Resting state cortico-cerebellar functional connectivity
814 networks: a comparison of anatomical and self-organizing map approaches.
815 *Front Neuroanat* **6**, 31, doi:10.3389/fnana.2012.00031 (2012).
816 21 King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B. &
817 Diedrichsen, J. Functional boundaries in the human cerebellum revealed by a
818 multi-domain task battery. *Nat Neurosci* **22**, 1371-1378,
819 doi:10.1038/s41593-019-0436-x (2019).
820 22 Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C. & Yeo, B. T. The
821 organization of the human cerebellum estimated by intrinsic functional
822 connectivity. *J Neurophysiol* **106**, 2322-2345, doi:10.1152/jn.00339.2011
823 (2011).
824 23 Ji, J. L. *et al.* Mapping the human brain's cortical-subcortical functional
825 network organization. *Neuroimage* **185**, 35-57,
826 doi:10.1016/j.neuroimage.2018.10.006 (2019).
827 24 Guell, X., Schmahmann, J. D., Gabrieli, J. & Ghosh, S. S. Functional
828 gradients of the cerebellum. *Elife* **7**, doi:10.7554/eLife.36652 (2018).
829 25 Fornito, A., Arnatkeviciute, A. & Fulcher, B. D. Bridging the Gap between
830 Connectome and Transcriptome. *Trends Cogn Sci* **23**, 34-50,
831 doi:10.1016/j.tics.2018.10.005 (2019).
832 26 Richiardi, J. *et al.* BRAIN NETWORKS. Correlated gene expression supports
833 synchronous activity in brain networks. *Science* **348**, 1241-1244,
834 doi:10.1126/science.1255905 (2015).
835 27 Anderson, K. M. *et al.* Gene expression links functional networks across
836 cortex and striatum. *Nat Commun* **9**, 1428, doi:10.1038/s41467-018-03811-x
837 (2018).
838 28 Ritchie, M. E. *et al.* limma powers differential expression analyses for
839 RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47,
840 doi:10.1093/nar/gkv007 (2015).
841 29 Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by

intrinsic functional connectivity. *J Neurophysiol* **106**, 1125-1165, doi:10.1152/jn.00338.2011 (2011).

30 Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E. & Ramnani, N. A probabilistic MR atlas of the human cerebellum. *Neuroimage* **46**, 39-46, doi:10.1016/j.neuroimage.2009.01.045 (2009).

31 Van Essen, D. C. *et al.* The Human Connectome Project: a data acquisition perspective. *Neuroimage* **62**, 2222-2231, doi:10.1016/j.neuroimage.2012.02.018 (2012).

32 Huntley, R. P. *et al.* The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res* **43**, D1057-1063, doi:10.1093/nar/gku1113 (2015).

33 Zhu, D. M. *et al.* Cerebellar-cerebral dynamic functional connectivity alterations in major depressive disorder. *J Affect Disord* **275**, 319-328, doi:10.1016/j.jad.2020.06.062 (2020).

34 Verly, M. *et al.* Altered functional connectivity of the language network in ASD: role of classical language areas and cerebellum. *Neuroimage Clin* **4**, 374-382, doi:10.1016/j.nicl.2014.01.008 (2014).

35 Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206, doi:10.1038/nature13185 (2014).

36 Dougherty, J. D., Schmidt, E. F., Nakajima, M. & Heintz, N. Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res* **38**, 4218-4230, doi:10.1093/nar/gkq130 (2010).

37 ten Donkelaar, H. J., Lammens, M., Wesseling, P., Thijssen, H. O. & Renier, W. O. Development and developmental disorders of the human cerebellum. *J Neurol* **250**, 1025-1036, doi:10.1007/s00415-003-0199-9 (2003).

38 Munoz-Castaneda, R. *et al.* Cytoskeleton stability is essential for the integrity of the cerebellum and its motor- and affective-related behaviors. *Sci Rep* **8**, 3072, doi:10.1038/s41598-018-21470-2 (2018).

39 Beyreli, I., Karakahya, O. & Cicek, A. E. Deep multitask learning of gene risk for comorbid neurodevelopmental disorders. *bioRxiv*, 2020.2006.2013.150201, doi:10.1101/2020.06.13.150201 (2020).

40 Villarroel, M. A. & Terlizzi, E. P. Symptoms of Depression Among Adults: United States, 2019. *NCHS Data Brief*, 1-8 (2020).

41 Anderson, M. L., Kinnison, J. & Pessoa, L. Describing functional diversity of brain regions and brain networks. *Neuroimage* **73**, 50-58, doi:10.1016/j.neuroimage.2013.01.071 (2013).

42 Dalman, M. R., Deeter, A., Nimishakavi, G. & Duan, Z.-H. Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics* **13**, S11, doi:10.1186/1471-2105-13-S2-S11 (2012).

43 Margulies, D. S. *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, doi:10.1073/pnas.1608282113 (2016).

44 Schmahmann, J. D. & Pandya, D. N. Disconnection syndromes of basal

ganglia, thalamus, and cerebrocerebellar systems. *Cortex* **44**, 1037-1066, doi:10.1016/j.cortex.2008.04.004 (2008).

45 Schmahmann, J. D. From movement to thought: anatomic substrates of the cerebellar contribution to cognitive processing. *Hum Brain Mapp* **4**, 174-198, doi:10.1002/(SICI)1097-0193(1996)4:3<174::AID-HBM3>3.0.CO;2-0 (1996).

46 Storbeck, J. & Clore, G. L. On the interdependence of cognition and emotion. *Cogn Emot* **21**, 1212-1237, doi:10.1080/02699930701438020 (2007).

47 Ghashghaei, H. T. & Barbas, H. Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* **115**, 1261-1279, doi:https://doi.org/10.1016/S0306-4522(02)00446-3 (2002).

48 Riener, C. R., Stefanucci, J. K., Proffitt, D. R. & Clore, G. An effect of mood on the perception of geographical slant. *Cogn Emot* **25**, 174-182, doi:10.1080/02699931003738026 (2011).

49 Mueller, S. *et al.* Individual variability in functional connectivity architecture of the human brain. *Neuron* **77**, 586-595, doi:10.1016/j.neuron.2012.12.028 (2013).

50 Guell, X., Schmahmann, J. D. & Gabrieli, J. D. Functional Specialization is Independent of Microstructural Variation in Cerebellum but Not in Cerebral Cortex. *bioRxiv*, 424176, doi:10.1101/424176 (2018).

51 Marek, S. *et al.* Spatial and Temporal Organization of the Individual Human Cerebellum. *Neuron* **100**, 977-993 e977, doi:10.1016/j.neuron.2018.10.010 (2018).

52 Hensler, J. G. Serotonergic modulation of the limbic system. *Neurosci Biobehav Rev* **30**, 203-214, doi:10.1016/j.neubiorev.2005.06.007 (2006).

53 Schmahmann, J. D. Disorders of the Cerebellum: Ataxia, Dysmetria of Thought, and the Cerebellar Cognitive Affective Syndrome. *The Journal of Neuropsychiatry and Clinical Neurosciences* **16**, 367-378, doi:10.1176/jnp.16.3.367 (2004).

54 Brainstorm, C. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, doi:10.1126/science.aap8757 (2018).

55 Bertoli-Avella, A. M. *et al.* Biallelic inactivating variants in the GTPBP2 gene cause a neurodevelopmental disorder with severe intellectual disability. *Eur J Hum Genet* **26**, 592-598, doi:10.1038/s41431-018-0097-3 (2018).

56 Wefers, A. K., Lindner, S., Schulte, J. H. & Schuller, U. 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, doi:10.1007/s12311-016-0774-0 (2017).

57 Swoboda, K. J. & Hyland, K. Diagnosis and treatment of neurotransmitter-related disorders. *Neurol Clin* **20**, 1143-1161, viii, doi:10.1016/s0733-8619(02)00018-x (2002).

58 Brown, R. P. & Mann, J. J. A clinical perspective on the role of

928 neurotransmitters in mental disorders. *Hosp Community Psychiatry* **36**,
929 141-150, doi:10.1176/ps.36.2.141 (1985).
930 59 Kato, T. A. *et al.* Neurotransmitters, psychotropic drugs and microglia: clinical
931 implications for psychiatry. *Curr Med Chem* **20**, 331-344,
932 doi:10.2174/0929867311320030003 (2013).
933 60 Seo, D., Patrick, C. J. & Kennealy, P. J. Role of Serotonin and Dopamine
934 System Interactions in the Neurobiology of Impulsive Aggression and its
935 Comorbidity with other Clinical Disorders. *Aggress Violent Behav* **13**, 383-395,
936 doi:10.1016/j.avb.2008.06.003 (2008).
937 61 Su, Y., D'Arcy, C. & Meng, X. Research Review: Developmental origins of
938 depression - a systematic review and meta-analysis. *J Child Psychol*
939 *Psychiatry*, doi:10.1111/jcpp.13358 (2020).
940 62 Newschaffer, C. J., Fallin, D. & Lee, N. L. Heritable and nonheritable risk
941 factors for autism spectrum disorders. *Epidemiol Rev* **24**, 137-153,
942 doi:10.1093/epirev/mxf010 (2002).
943 63 Ramos, T. C., Balardin, J. B., Sato, J. R. & Fujita, A. Abnormal
944 Cortico-Cerebellar Functional Connectivity in Autism Spectrum Disorder.
945 *Front Syst Neurosci* **12**, 74, doi:10.3389/fnsys.2018.00074 (2018).
946 64 Baker, K. G., Harding, A. J., Halliday, G. M., Kril, J. J. & Harper, C. G.
947 Neuronal loss in functional zones of the cerebellum of chronic alcoholics with
948 and without Wernicke's encephalopathy. *Neuroscience* **91**, 429-438,
949 doi:10.1016/s0306-4522(98)90664-9 (1999).
950 65 Abdallah, M. *et al.* Altered Cerebro-Cerebellar Dynamic Functional
951 Connectivity in Alcohol Use Disorder: a Resting-State fMRI Study.
952 *Cerebellum*, doi:10.1007/s12311-021-01241-y (2021).
953 66 Freire-Cobo, C. & Wang, J. Dietary phytochemicals modulate
954 experience-dependent changes in Neurexin gene expression and alternative
955 splicing in mice after chronic variable stress exposure. *Eur J Pharmacol* **883**,
956 173362, doi:10.1016/j.ejphar.2020.173362 (2020).
957 67 Hu, Z., Xiao, X., Zhang, Z. & Li, M. Genetic insights and neurobiological
958 implications from NRXN1 in neuropsychiatric disorders. *Molecular*
959 *Psychiatry* **24**, 1400-1414, doi:10.1038/s41380-019-0438-9 (2019).
960 68 Hua, J., Yang, Z., Jiang, T. & Yu, S. Pair-wise interactions in gene expression
961 determine a hierarchical transcription profile of human brain. *Science Bulletin*,
962 doi:https://doi.org/10.1016/j.scib.2021.01.003 (2021).
963 69 Miller, J. A. *et al.* Strategies for aggregating gene expression data: the
964 collapseRows R function. *BMC Bioinformatics* **12**, 322,
965 doi:10.1186/1471-2105-12-322 (2011).
966 70 Arnatkeviciute, A., Fulcher, B. D. & Fornito, A. A practical guide to linking
967 brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353-367,
968 doi:10.1016/j.neuroimage.2019.01.011 (2019).
969 71 Cox, R. W. AFNI: software for analysis and visualization of functional
970 magnetic resonance neuroimages. *Computers and biomedical research, an*

international journal **29**, 162-173, doi:10.1006/cbmr.1996.0014 (1996).

72 Smith, S. M. *et al.* Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144-168, doi:10.1016/j.neuroimage.2013.05.039 (2013).

73 Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105-124, doi:10.1016/j.neuroimage.2013.04.127 (2013).

74 Salimi-Khorshidi, G. *et al.* Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, 449-468, doi:10.1016/j.neuroimage.2013.11.046 (2014).

75 Griffanti, L. *et al.* ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* **95**, 232-247, doi:10.1016/j.neuroimage.2014.03.034 (2014).

76 Kong, X.-Z. *et al.* Gene Expression Correlates of the Cortical Network Underlying Sentence Processing. *Neurobiology of Language* **1**, 77-103, doi:10.1162/nol_a_00004 (2020).

77 Seidlitz, J. *et al.* Morphometric Similarity Networks Detect Microscale Cortical Organization and Predict Inter-Individual Cognitive Variation. *Neuron* **97**, 231-247 e237, doi:10.1016/j.neuron.2017.11.039 (2018).

78 Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305-311, doi:10.1093/nar/gkp427 (2009).