A hierarchical atlas of the human cerebellum for functional precision mapping

Caroline Nettekoven[1,2,*], Da Zhi[1,2], Ladan Shahshahani[1], Ana Luísa Pinho[1,2], Noam Saadon-Grosman[3], and Jörn Diedrichsen[1,2,4,*]

[1]**Western Institute for Neuroscience, Western University, London, Ontario, Canada**
[2]**Department of Computer Science, Western University, London, Ontario, Canada**
[3]**Harvard University, Cambridge, US**
[4]**Department of Statistical and Actuarial Sciences, Western University, London, Ontario, Canada**
[*]**Corresponding authors: Caroline Nettekoven, cr.nettekoven@gmail.com; Jörn Diedrichsen, jdiedric@uwo.ca**

ABSTRACT

The human cerebellum is activated by a wide variety of cognitive and motor tasks. Previous functional atlases have relied on single task-based or resting-state fMRI datasets. Here, we present a functional atlas that integrates information from 7 large-scale datasets, outperforming single dataset parcellations. The new atlas has three further advantages: First, the regions are hierarchically organized across 3 levels, allowing analyses at the appropriate level of granularity. Second, by constraining the boundaries to be the same across hemispheres, we provide a symmetric version of the atlas, which is especially useful in studying functional lateralization. Finally, the atlas allows for precision mapping in individuals: The integration of the probabilistic group atlas with an individual localizer scan results in a marked improvement in prediction of individual boundaries. Overall, the new atlas is an important resource for the study of the interdigitated functional organization of the human cerebellum in health and disease.

Keywords: Cerebellum, Brain Atlassing, Hierarchical Bayesian model, Functional brain parcellation, task-based fMRI, resting-state fMRI

INTRODUCTION

Decades of neuroimaging have shown cerebellar activation in a broad range of tasks, including motor, social, and cognitive tasks - yet its contribution to these different functions remains elusive [1, 2]. A major obstacle to understanding the cerebellar contribution is that the cerebellum consists of a mosaic of functional regions, each of which likely plays a different role [3]. It is still common to use the anatomical subdivision into different lobules [4, 5] to define regions of interest for subsequent analyses. But lobular boundaries do not reflect functional boundaries [3], nor do they align well with developmental gradients in the cerebellum [6].

There are several existing maps based on resting-state or task-based functional Magnetic Resonance Imaging (fMRI) data [7, 8, 3] that parcel the cerebellum into functional regions. These functional atlases clearly outperform anatomical parcellations at predicting functional boundaries, with a task-based parcellation based on a large multi-domain task battery (MDTB) showing an advantage over resting-state parcellations [3]. Nonetheless, task-based parcellations, including the MDTB parcellation have clear weaknesses. First, task-based datasets are limited in the number of tasks, exemplified by the lack of foot movement task in the MDTB dataset. The MDTB parcellation is therefore unable to delineate functional regions in lobules IV, known to host the cerebellar foot region. Second, the number of participants in task-based datasets is generally small due to the substantial efforts required to generate these datasets. To address this shortcoming, we have recently developed a Bayesian Hierarchical method that combines information across datasets into a single parcellation [9]. In this study, we apply this model to seven large task-based datasets to derive a novel cerebellar functional atlas.

Another important limitation of existing group atlases is that they ignore the large inter-individual

variability in functional brain organization [10, 11, 12, 13, 14]. This problem is particularly relevant for the cerebellar cortex, where a large number of functionally heterogeneous regions is packed into a relatively small volume [15, 3]. Numerous groups have therefore proposed to pursue a precision mapping approach, using localizing data to define functional regions at the individual level [11, 12, 13, 16]. To enable such precise and fine-grained analysis, the new atlas is based on a probabilistic framework, which allows the user to use even limited individual data to optimally tailor the atlas to the individual. We evaluated this approach carefully by showing the superiority of the personalized parcellation at predicting functional boundaries in the same individual in different tasks, as compared to both the group atlas, and a parcellation solely based on individual data.

The cerebellum plays a key role in lateralized functions (i.e., language; [17]) and shows lateralized developmental trajectories [6]. The study of lateralization, however, is complicated by existing functional atlases, as they have asymmetric boundaries and no correspondence between left and right regions. We therefore developed the atlas to have symmetric boundaries and have matching hemispheric parcel pairs. Importantly, we did not constrain the functional profiles to be the same across hemispheres, enabling us to study functional lateralization. The comparison to an asymmetric version of the atlas also allowed us to assess whether this symmetry constraint is adequate, or to what degree the spatial organization is truly asymmetric.

Finally, questions about cerebellar function will need to be tested at different levels of granularity. For many anatomical and patient studies, it is often most appropriate to summarize measures in terms of broad functional domains (e.g. motor vs. social-linguistic regions), whereas more detailed functional studies require the definition of finer region distinctions (e.g., separate hand, foot, and tongue regions within the motor domain). We therefore created the atlas with a hierarchical organization of functional regions where the boundaries of the broad domains remain the same at each level of granularity.

RESULTS

State-dependency of functional atlasses

The cognitive state of the brain (rest or specific tasks) will influence how different functional regions work together. Therefore, parcellations based on different datasets may result in different functional boundaries. In deriving a functional atlas, it is thus important to carefully consider which datasets to base the atlas on. As a first step, we therefore sought to characterize similarities between parcellations based on single datasets, using task-based and resting-state data. We trained our probabilistic parcellation model [9] on each dataset in isolation (Fig. 1a) and then compared the resultant parcellations.

This analysis showed some similarities between parcellation maps, but also clear dataset-specific differences. For example, the somatotopic dataset only tested individual body movements, and therefore resulted in a clear somatomotor map, but did not delineate regions of the multi-demand network [18] well, which are involved in working memory and executive functions. In contrast, the Demand dataset delineated working memory regions but had little somatomotor information. Parcellations based on resting-state data (HCP) showed consistent boundaries in regions related to the default-mode network (lobules VII) but appear to delineate other regions (e.g. motor) less finely.

To quantify these similarities, we calculated the adjusted Rand Index (ARI) between parcellations at different levels of parcel granularity (10, 20, 34, 40 and 68 regions). The indices were averaged across granularities and normalized by the within-dataset ARI (see methods). Overall, parcellations of all datasets showed positive ARIs with the other datasets (One-sample t-test of the between-dataset ARIs averaged across granularities $t_{27} = 17.885, p = 1.696 \times 10^{-16}$), indicating that there are common functional boundaries across all different task and resting state datasets [20, 11, 21].

To assess the similarity of the resulting parcellations better, we visualized the reliability-adjusted ARIs using multi-dimensional scaling (Fig. 1b). Unsurprisingly, task-based datasets that are most similar in their task domains (i.e., working memory and multi-demand dataset) resulted in similar parcellations. In contrast, datasets that probed a unique and narrow domain, such as the Somatotopic dataset, differed from the rest. Another pole in the space of parcellations was occupied by the resting state parcellation (HCP), which only showed similarity to the Nishimoto dataset, but was distinct from other task-based parcellations. This is in line with mounting evidence for systematic differences in functional organization during different task-states and during rest [22, 20, 23].

Parcellations based on datasets that included a large range of cognitive tasks (MDTB, MDTB-Highres, and IBC) occupied a middle position, suggesting that such parcellations are more representative of

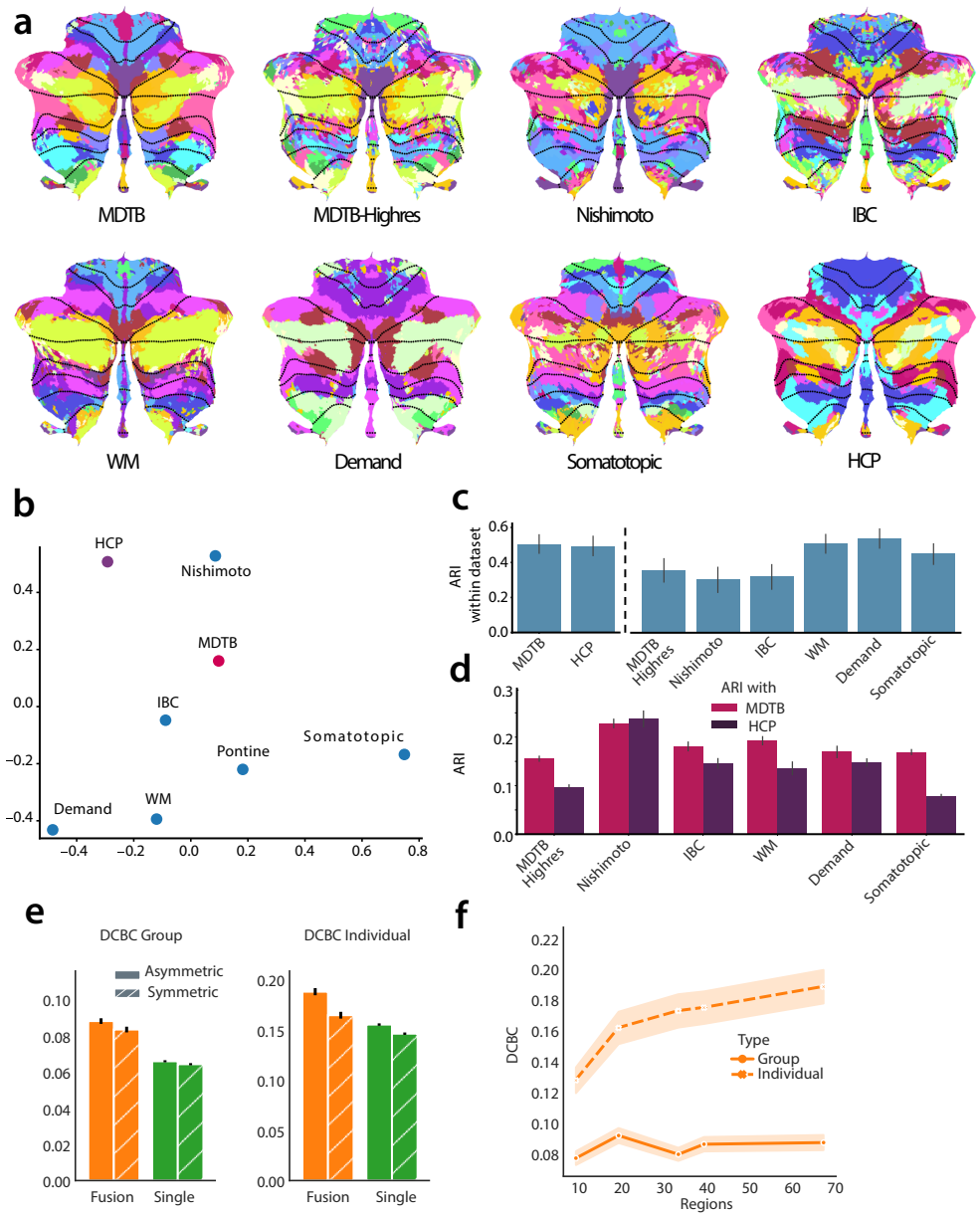

Figure 1. Building a functional atlas of the cerebellum across datasets. **a**, Parcellations (K=68) derived from each single dataset. The probabilistic parcellation is shown as a winner-take-all projection onto a flattened representation of the cerebellum [19]. Dotted lines indicate lobular boundaries. **b**, Projection of the between-dataset adjusted Rand Index (ARI) of single-dataset parcellations into a 2d-space through multi-dimensional scaling (see methods: Single-dataset parcellations and similarity analysis of parcellations). **c**, Within-dataset reliability of parcellation, calculated as the mean ARI across the 5 levels of granularity (10, 20, 34, 40 and 68 regions). Errobars indicate SE of the mean across granularity pairs). **d**, ARI between each single-dataset parcellations and the MDTB (task-based) and HCP (resting parcellation) parcellation. Errorbars indicate standard error of the mean across levels of granularity. **e**, DCBC evaluation of the symmetric and asymmetric atlas at the finest granularity of 68 regions. **f**, DCBC evaluation of the group map and of individual maps derived from the model with 10, 20, 34, 40, and 68 regions.

functional boundaries across tasks than parcellations based only on resting-state data. Indeed, when we compared the normalized ARI for each specific task-based parcellations, we found that they were more similar to the parcellation derived from the MDTB dataset than to one derived from the HCP dataset (paired t-test: $t_{149} = 9.605, p = 2.672 \times 10^{-17}$; Fig. 1d). Testing each task-based dataset separately confirmed that all, except for Nishimoto dataset ($t_{24} = -0.838, p = 0.410$) were significantly more similar to a multi-domain dataset than to rest (MDTB-Highres: $t_{24} = 16.404, p = 1.523 \times 10^{-14}$; IBC: $t_{24} = 3.513, p = .0017$; WM: $t_{24} = 4.727, p = 8.318 \times 10^{-5}$; Demand: $t_{24} = 3.262, p = .0033$; Somatotopic: $t_{24} = 12.538, p = 5.015 \times 10^{-12}$; Bonferroni-corrected significance threshold p = .008)

Fused parcellation outperforms single-dataset parcellations

Our analysis shows that the resting-state parcellation captures some task-based boundaries, but also suggests that it is less representative of task-base boundaries than a parcellation based on a broad task set. Because our main motivation here was to generate an atlas that would identify functional boundaries across task-based paradigms, we decided to combine information from task-based datasets only.

To verify that the fusion of datasets through our framework systematically improved on single-dataset parcellations, we adopted a leave-one-dataset-out approach across different levels of granularity. We trained the fusion parcellation on all task-based datasets except one and tested its ability to predict the functional boundaries within that left-out dataset. This ability was quantified using the Distance-Controlled Boundary Coefficient (DCBC) which compares the correlation between within-parcel voxel-pairs to the correlation between voxels-pairs across a boundary, while controlling for spatial distance [24]. We compared this to the average DCBC value for single-dataset parcellations using a paired t-test. Additionally to the group map (group DCBC), we also evaluated the quality of the individual parcellations derived from this group atlas (see methods). We found that the fused atlas outperformed single dataset parcellations averaged across granularities both for the group ($t_{110} = -4.466, p = 1.936 \times 10^{-5}$; Fig. 1e left) and for the individual parcellations ($t_{110} = -2.564, p = .0171$; Fig. 1e right), confirming our previous results showing that fusion optimally combines information across fMRI datasets [9].

Symmetric vs. asymmetric atlases

To enable the study of hemispheric specialization, we initially constrained our atlas to have spatially symmetric regions across the left and right cerebellar hemispheres. To determine whether this constraint was adequate, we estimated an asymmetric version of the atlas without using the symmetry constraint (see methods, Symmetry constraint).

We then compared the ability of the asymmetric and the symmetric atlas to predict functional boundaries, again adopting a leave-one-dataset-out approach. For the group DCBC, we found a small, but significant difference between the asymmetric and symmetric atlas across levels of granularity (10-68 regions; $t_{110} = -2.344, p = .0201$) (Fig. 4B). This advantage was larger at the individual level ($t_{110} = -5.023, p = 1.981 \times 10^{-6}$). Overall, however, the predictive power of the symmetric atlas was only 5% (group) or 14% (individual) smaller than the asymmetric versions. Given the many practical advantages of the symmetric atlas for controlling for region size and location in lateralization studies, we prepared both symmetric and asymmetric versions of the final atlas.

Spatial granularity of the hierarchical base map

Instead of choosing a fixed number of regions, we designed the atlas to have different levels of resolution linked in a hierarchical scheme. This allows the user to employ different levels of granularity in a consistent fashion. To decide on the "base map" of this hierarchy, we examined the predictive performance of the fusion atlas across the tested levels of granularity at the group and individual levels (Fig. 1f). We found that the performance of the group map saturated early, reaching its best value at 20 regions. However, this peak was not significantly different from the finest granularity of 68 regions ($t_{110} = 2.783, p = .0063$). In contrast, the ability to predict boundaries in the individual increased monotonically, with the finest granularity outperforming the next lower granularity of 40 regions ($t_{110} = 7.584, p = 1.143 \times 10^{-11}$). We therefore based the hierarchical atlas on the map with the finest granularity of 68 functional regions.

Hierarchical atlas organization

We then generated a hierarchically clustered versions of this "base map" according to the functional similarity between regions. These can be visualized using a colormap in which regions with similar functional profiles are also colored similarly. We first clustered the 68 regions (34 per hemisphere) into 32

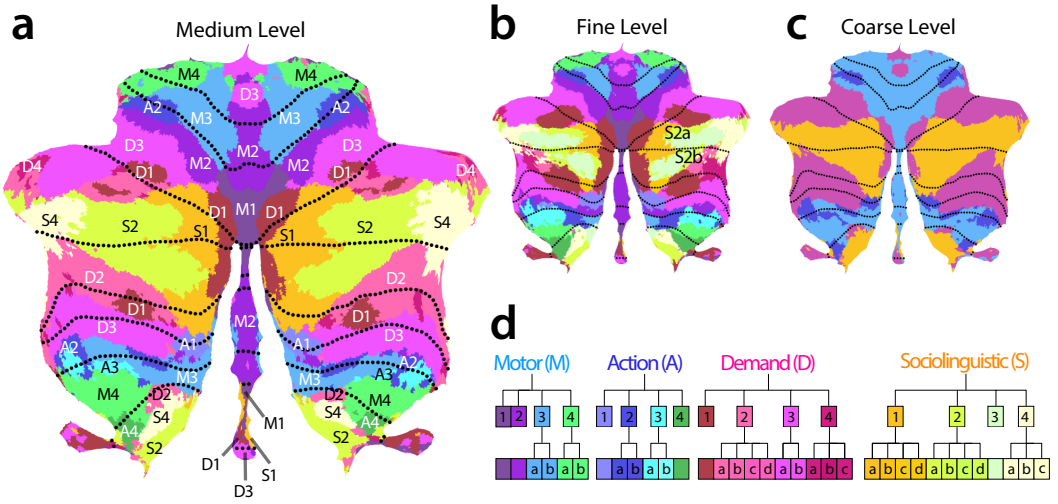

Figure 2. Cerebellar functional atlas at three levels of granularity. **a**, Medium granularity with 32 regions; 16 per hemisphere. **b**, Fine granularity with 68 regions; 34 per hemisphere. **c**, Coarse granularity with 4 functional domains. **d**, Hierarchical organization based on the functional similarity of regions, depicted as a dendrogram. The label of each region indicates the functional domain (M,A,D,S), followed by a region number (1-4), and a lower-case letter for the subregion (a-d).

regions (16 per hemisphere) according to the functional similarity between regions (see methods: parcel similarity and clustering). Finally, we organized these 16 regions into 4 broad functional domains. Based on their functional activation profiles, we denoted these four functional domains as motor (M), action (A), multi-demand (D), and social-linguistic (S) (Fig. 2c). At the medium level, we numbered the four regions within each domain from medial to lateral (Fig. 2d). Finally, the finest level was annotated with a lowercase letter (a-d). In the following description of the regions, we will focus on the medium level, as it provides a good compromise between precision and succinctness.

Characterization of regions

Each functional region is characterized by its response profile across datasets and it's spatial distribution across individuals. In describing the functional profile, we focused on responses estimated from subject-specific regions in the MDTB dataset (see methods: Functional profiles for the MDTB dataset), supplemented by more domain-specific datasets for the motor and demand regions (Somatotopic, Demand, WM).

Motor regions

Regions that exhibited a clear preference for movements of a specific body part were grouped into the motor domain. All regions had a superior (lobules I-VI) and an inferior (lobulus VIII) aspect. We also found a third representation of these body-part-specific regions in the posterior vermis, consistent with recent results at the individual subject level [25].

M1 encompassed the oculomotor vermis, which responded most strongly to saccades (Fig. S1). Even when correcting for the number of saccades, the area was further activated when participants had to read text (Theory-of-Mind), watch a movie (RomanceMovie), or search for visual stimuli (SpatialMap and VisualSearch), likely due to the attentional demands of these tasks. Previous work has shown that this region also has a clear retinotopic organization [26]. M2 comprises a lateral and a vermal part. The lateral section showed strong responses to tongue movements in the somatotopic dataset. In contrast, the vermal component was activated by multiple different bodily movements, but otherwise was functionally most similar to the lateral M2. The M3 regions were clearly and selectively activated by left- and right-hand movements (Supplemental Fig.S1), and therefore can be best described as the cerebellar hand regions with a superior and inferior aspect. Finally, M4 was most activated by movements of the lower body, including flexion and extension of the foot (Highres-MDTB), as well as contraction of the gluteal muscles

(Somatotopic).

Action regions

Directly adjacent to the motor regions lie the action regions, which were activated during action observation and motor imagery tasks. A1 and A2 both comprised spatially separate superior and inferior sections, whereas A3 and A4 primarily occupy the inferior cerebellum (Fig. S2). A1 can be found medially to the hand region in lobule VI and at the border of VIIIa/VIIIb. A2 lies laterally adjacent to the superior hand region M3, and at the border of lobule VIIIa/VIIIb. A3 and A4 only are only present in the inferior cerebellum, with A3 at the border of lobules VIIIa/VIIIb and A4 in lobule IX.

Although both motor and action regions activated during movement execution, only the action regions activated when observing actions without execution: In the MDTB dataset, they showed strong responses to action observation task (VideoAct in Fig. S1). A1 appeared to be particularly involved where spatial simulation is required (strong responses during SpatialMap and MentalRotation tasks). Meanwhile, A2 seems to be a classic action observation region, with little response to tasks that do not involve action observation or execution. In contrast, A3 was also activated during imagined movements (MotorImagery). Finally, A4 appeared to be especially engaged when perceiving or imagining actions with social and emotional content, activating when participants imagined walking through their childhood home (SpatialNavigation) and watched a romantic movie (RomanceMovie).

Multiple-demand regions

Tasks involving executive control consistently activated regions in lobules VI and VII. Based on work by Duncan et al. [18], we labelled these regions the multi-demand domain (D for short). D1 occupied the most medial portion of Crus I and II. Further out in the hemispheres, the demand region formed a ”shell” around the more central social-linguistic domain (Fig. S3b). Here, D3 formed the outermost layer and D2 the innermost, with D1 being interspersed between. The regions (especially D2) also had a repeated representation in lobulus IX (Fig. S2). This is consistent with a 3-fold representation [8]. Intriguingly, we found also a vermal section of D3, both in lobulus IV and IX. D4 was the smallest identified region. Functionally most similar to D1, it occupied the most lateral portion of the demand regions.

Consistent with the characteristics of the cortical multi-demand system [27], all regions showed significant activation during executive tasks (n-back, switch and stop tasks), and increased activity especially with high difficulty. Nonetheless, there was some functional specialization across the regions. In the MDTB dataset, D1 appeared to be involved strongly in spatial tasks, such as the mental rotation, and spatial map task. D1 and D4 were strongly engaged in the n-back task. In contrast, D2 and D3 were specifically activated by the digit span task tested in the WMFS data set - with D2 more active during backwards recall and D3 showing strong increases with working memory load.

Social-linguistic regions

The regions in hemispheric lobules Crus I and Crus II, located laterally to the D1 region, were activated by tasks involving social and linguistic processes. They also showed high activity during rest, consistent with the description of this area as the cerebellar node of the default mode network [8]. We identified four regions, each spanning both sides of the horizontal fissure, with S1 being the most medial and S4 most lateral (Fig. 2). S3 overlapped substantially with S2 and S4 and therefore could only be reliably differentiated from these two regions at the level of the individual (see 5a). In the volume (Supplementary Fig. S3) S1 occupies the depth of the horizontal fissure, and S4 the most lateral tips of Crus I and II. A third representation of S2 and S4 can be found in lobules IX. S1 and S2 also occupy sections in the inferior vermis (VIIIb and IX, Supplementary Fig. S3).

While all social-linguistic regions showed activation for social or language tasks, their precise response profiles differed. The mean evoked responses from the MDTB dataset corrected for motor features (see Supplementary Fig. S1) showed S1 to be primarily involved in linguistic processing, with highest activation during verb generation on the right and verbal working memory tasks on the left. S2 was strongly engaged in social processing, with highest activity on the right during theory-of-mind tasks and during a romantic movie on the left. Both S3 and S4 showed high levels of activity during rest (Supplementary Fig. S4d), but S4 appeared to be particularly involved in imagination and self-projection (Supplementary Fig. S4c), showing the highest activation during the SpatialNavigation task and the motor imagery tasks, which require the participant to imagine themselves walking through their childhood home and playing a game of tennis, respectively. In contrast to S2, S4 showed no activation during the

theory-of-mind or RomanceMovie task (see Supplementary Fig. S4c). This difference in functional profile identifies S4 therefore as the cerebellar node of the recently described default mode network A, and S2 and S3 as part of the default mode network B [28].

Cortical connectivity

Each cerebellar region can be also characterized by the regions in neocortex that it is most functionally connected to. To determine these areas we estimated an effective connectivity model, aiming to explain the data in each cerebellar voxel as a linear combination of cortical areas [29]. For the task-based dataset, we used the condition-averaged profiles, for the resting state data, the preprocessed time-series. We fitted the models individually per subject and dataset. To evaluate the average connectivity model for each dataset, we predicted the cerebellar activity patterns for each test dataset, using only the corresponding cortical activity patterns (see methods: Cortical connectivity).

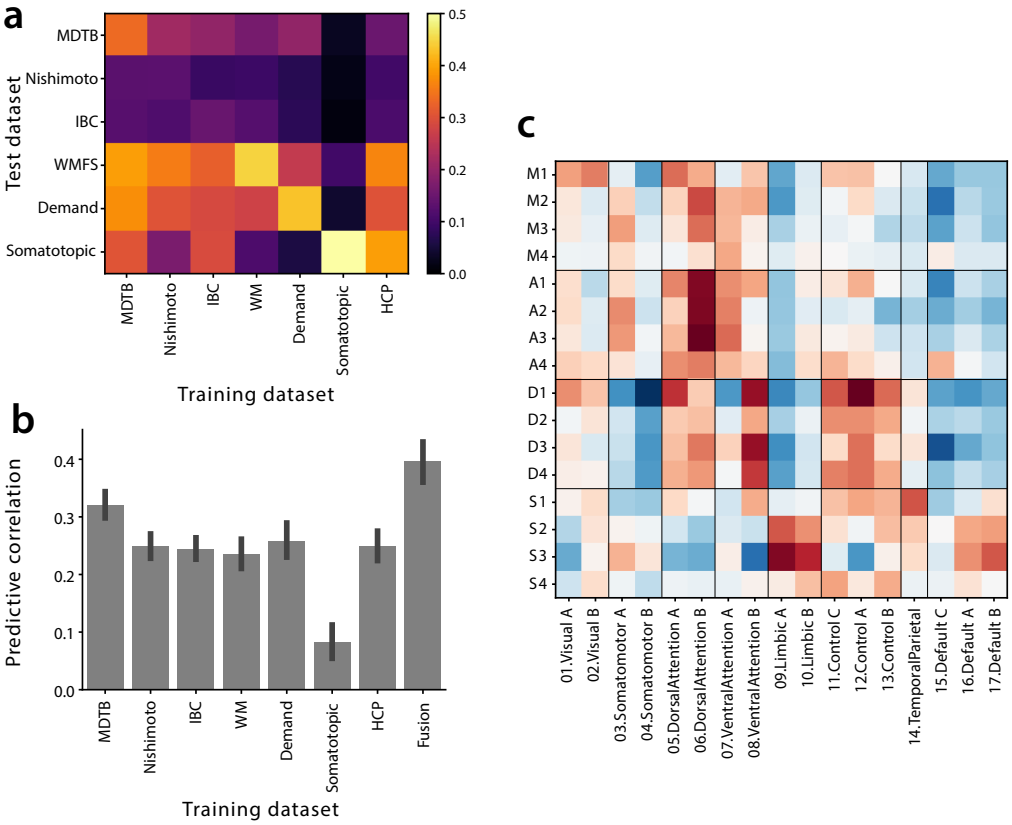

Figure 3. Cortico-cerebellar connectivity models. a, Matrix shows the correlation between observed and predicted cerebellar activity patterns for each test dataset (rows). Connectivity models were trained on each training datasets (columns) separately. Evaluation was cross-validated across subjects when training- and test-dataset were identical. **b**, Correlation between observed and predicted activity patterns, averaged across test-datasets. The Fusion model used the average connectivity weights across all task-based datasets (excluding the HCP resting-state data). **c**, Average connectivity weights for each cerebellar region (row), summed over each of the 17 resting-state networks as described in [30]

.

The average correlation between the predicted and the observed activity patterns (Fig. 3a) were significantly higher than zero for all training / test combinations. One notable exception was the model estimated on the Somatotopic dataset, which generally performed more poorly in predicting the other data sets. Connectivity models generally showed the highest predictive accuracy on the dataset they were trained on, even though this evaluation was cross-validated across subjects.

Averaged over all evaluation datasets (Fig. 3b), the model trained on the MDTB dataset performed best

- with the other models being nearly equivalent in their performance (with the exception of Somatotopic dataset). To fuse across datasets, we simply averaged the connectivity weights across models. We found that average prediction performance was slightly better if it did not include the HCP dataset (.396 vs. .394, $t_{102} = -1.51, p = 0.1349$). The final Fusion model (last bar in Fig. 3b) significantly outperformed the best individual connectivity model (MDTB, $t_{102} = -7.340, p = 5.322 \times 10^{-11}$). The model achieved an average correlation of $r = 0.396$ ($SD = 0.17$). However, taking into account the noise ceiling of this prediction given by the reliability of the cerebellar and the cortical data (see methods: Cortical connectivity), the model achieved a prediction accuracy of $R = 0.6840$, meaning that it predicted on average 47% of the explainable variance.

The weights of these connectivity models for each individual region (Supplemental Figure S5, S6) clearly showed connectivity with the expected cortical regions in the contralateral hemisphere. For example, the left cerebellar hand area showed the highest connectivity with the hand areas of the right primary motor cortex and sensory cortex, and vice versa for the right cerebellar hand area (Fig. S5c).

To summarize these weight maps in terms of standard cortical networks, we averaged the weights within 17 resting-state networks described in [30] (see 3c). This analysis showed the expected results with M1 connected to visual networks, M2-M3 to the Somatomotor network A, D1-D4 to the Ventral Attention network B and control networks, and S1-S4 to limbic and default mode networks. Interestingly, nearly all cerebellar regions showed high connectivity weights with the dorsal and ventral attention networks, consistent with the over-representations of these areas in the cerebellum [8, 15].

Functional lateralization and boundary asymmetry

Our symmetric atlas forced the boundaries between parcels to be the same across hemispheres. Nonetheless, the functional profiles for the left and right parcels were estimated separately (see methods: Symmetry constraint). Therefore, hemispheric differences in functional specialization were captured by the model. To investigate these differences, we correlated the functional profiles of corresponding left and right voxels (Fig. 4c). We observed low functional symmetry in the hand regions (M3). This was mainly caused by task sets that isolated left- vs. right-hand movements. Such task-dependence can be clearly seen in the foot motor region (M4), which appear functionally symmetric in the MDTB Highres dataset, which included bilateral foot movements, and functionally asymmetric in the somatotopic dataset included separate left and right movement conditions S7).

In contrast, the multi-demand regions consistently show high functional symmetry across all datasets, even though the task sets included different executive functions and working memory tasks, using verbal and non-verbal material. While there might be some functional lateralization within this domain, our results suggest that their response profiles are largely symmetric and that it may be difficult to find strongly lateralized tasks in this functional domain. In contrast, the social-linguistic regions showed much lower functional symmetry with substantial differences between left and right response profiles.

Therefore, some functions are clearly lateralized in the cerebellum, reflected in different functional profiles for left and right regions. However, it is also possible that boundaries between functional regions themselves are asymmetric. We therefore estimated an asymmetric version of the atlas with the same functional profiles per region, but without the constraint on symmetry. Overall, the asymmetric atlas was superficially similar to the symmetric atlas (Fig. 4a). However, closer inspection revealed some key differences between the left and right hemispheric parcels of the asymmetric atlas, with the biggest difference observed among the social-linguistic and multiple-demand regions. When we compared the region size between the left and right regions in the asymmetric atlas (see supplementary Fig. S8), S3 and S4 had larger regions on the right, while S2, A2, and D1 were bigger on the left side.

Additionally, we calculated an index of boundary symmetry (see methods: Boundary symmetry) by correlating the parcel probabilities from the asymmetric and symmetric atlas. We found high boundary symmetry in motor and demand regions and low boundary symmetry in social-linguistic regions. Specifically, among the motor regions the oculomotor vermis M1 and the hand region M3 (Fig. 4c) showed high boundary symmetry. All demand regions showed high boundary symmetry with the exception of D2. In the social-linguistic regions, we observed generally low boundary symmetry, indicating that for these regions an asymmetric atlas may be most appropriate.

Individual precision mapping

The fusion atlas reveals several finely inter-digitated regions that have not been well described before and that have only been localized at the single-subject level using large quantities of individual data [16].

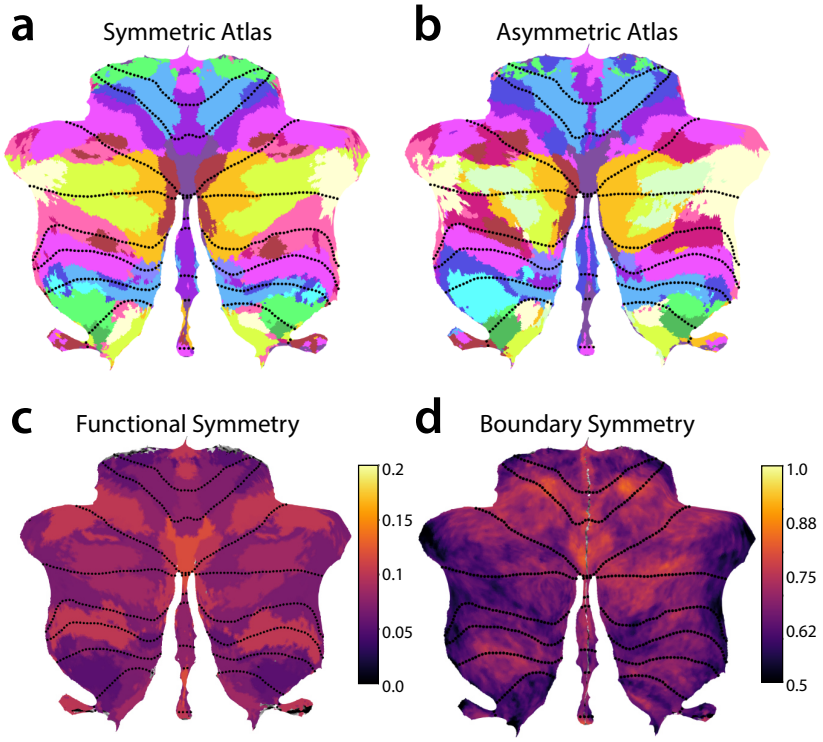

Figure 4. Functional and Boundary asymmetry in the cerebellum. **a**, Symmetric atlas winner-take-all map; **b**, Asymmetric atlas winner-take-all map; **c**, Functional symmetry calculated as the correlations of the functional responses of anatomically corresponding voxel of the left and right hemisphere, averaged across subjects and within each functional region. **d**, Boundary symmetry calculated as the correlations of the probabilistic voxel assignments between the symmetric and asymmetric version of the atlas

However, with the probabilistic framework, the atlas can be used to identify these regions in individual participants even with more limited data. In this section, we will describe the approach of personalizing the atlas to individuals, i.e., using the atlas for precision mapping [11, 12, 13].

We first characterized the spatial pattern of inter-individual variability to understand where in the cerebellum individual localization would offer the greatest utility. Within each dataset, for each voxel, we calculated the Pearson's correlation between the functional profiles of all possible pairs of subjects (methods:Inter-Individual variability). While motor regions showed very consistent functional profiles across subjects (e.g. hand regions M3 and eye regions M1 in Fig. 5b), the social-linguistic regions were much more variable. Only voxels in the core of the S1 region were relatively consistent across individuals; the lateral regions, and especially the boundary to the multi-demand regions demonstrated large inter-individual variability. Consistent with the observed strong inter-individual variability in the social-linguistic regions, our atlas shows considerable overlap in the group probability maps for region S1 and S2 (Fig. 5c). Hence, the study of these regions in Crus I and II and their differentiation from demand regions will benefit most from precision mapping of individuals.

For individual functional localization, a common approach is to acquire functional data from the individual to define individual regional boundaries. However, a substantial amount of functional data is necessary for deriving a parcellation that performs convincingly better than a group map [9, 15, 3]. We quantified this problem here by using 10min-160min of imaging data from the first session of the MDTB data set to derive individual parcellations. We then evaluated these parcellations on how well they predicted the functional boundaries in the same individuals tested on a second task set (Fig. 5d). We found that 20 min of individual data were necessary to be just as good as our new symmetric group atlas, and 30 min to significantly outperform the group map ($t_{23} = 2.283, p = 0.0320$, Fig.5d).

The probabilistic framework, however, allowed us to optimally combine evidence from the individual data with the probabilistic group map (see methods: Individual precision mapping). The final estimate of

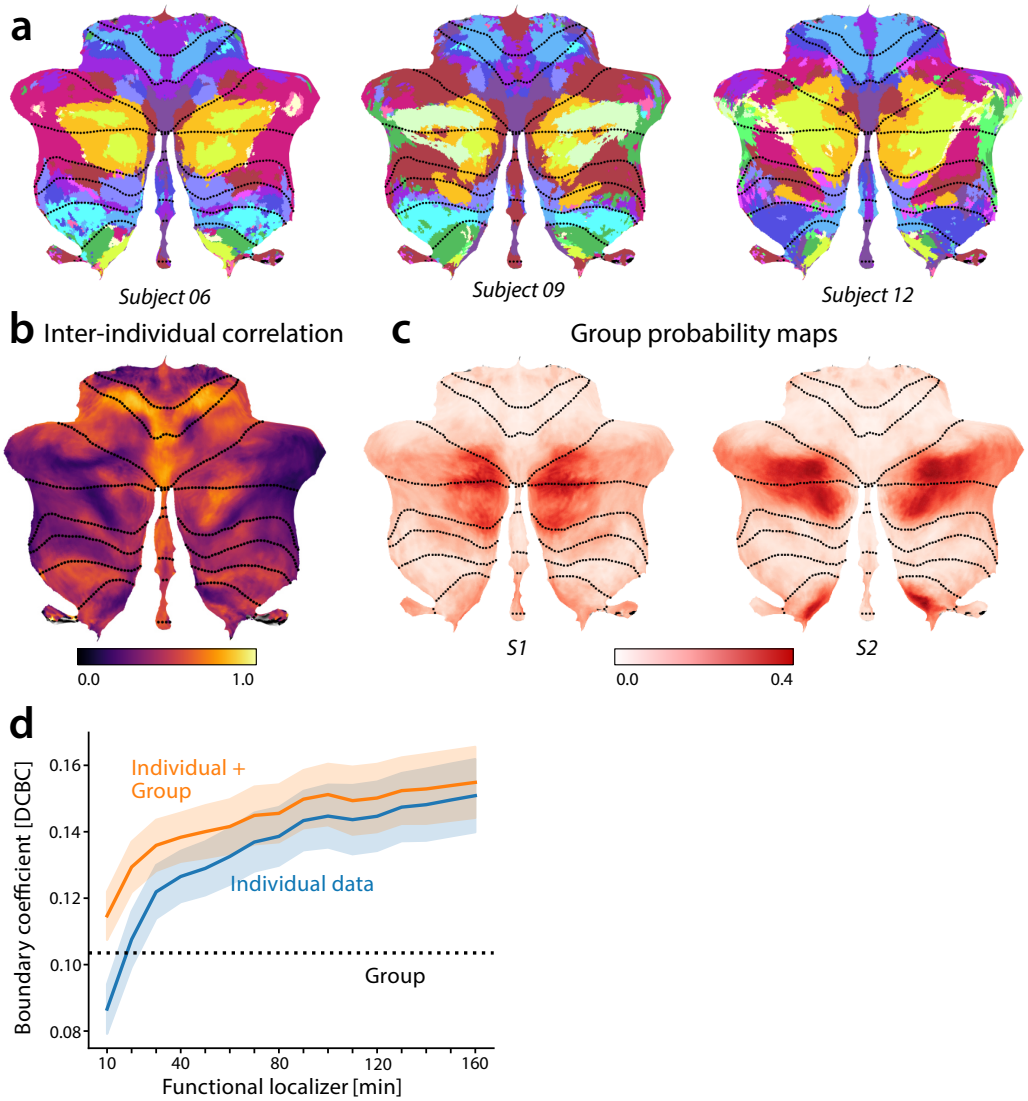

Figure 5. The functional atlas improves individual precision mapping. **a**, Individual parcellations from three participants, using 320min of individual data. **b**, Map of the average inter-subject correlations of functional profiles. Correlations are calculated between any pair of subjects within each dataset, corrected for the reliability of the data (see methods: Inter-individual variability) and then averaged across datasets. **c**, Group probability map for regions S1 and S2 (left and right combined) show the overlap of regions. **d**, DCBC evaluation on individual parcellations (blue line) derived on 10-160min of individual functional localizing data, compared to group parcellation (dashed line) or the combination of group map and individual data (orange line).

the model using only 20min of functional localization data outperformed both the individual data ($t_{23} = 11.468, p = 5.43 \times 10^{-11}$) and the group map ($t_{23} = 3.395, p = 0.0025$). The integrated estimate even improved individual parcellations based on as much as 160 mins of data ($t_{23} = 5.838, p = 5.989 \times 10^{-6}$). Thus the new atlas offers both the advantage of a consistent group map, as well as the possibility to obtain high-precision individualized mapping of brain organization.

DISCUSSION

Summary

In this study, we developed a comprehensive functional atlas of the human cerebellum featuring several important advances: First, using a Hierarchical Bayesian Model, we integrated data across seven large task-based datasets, thereby achieving a more complete coverage of functional domains. Second, by enforcing symmetry, our atlas is particularly suited to study functional lateralization in the cerebellum. Third, the atlas is hierarchically organized, allowing for a consistent description of the cerebellum at different levels of granularity. Finally, the probabilistic group atlas can be combined with a short localizer scan to dramatically improve functional precision mapping of individuals, paving a way to a detailed analysis of small subregions in the future.

Three-fold organization of the human cerebellum

Consistent with previous studies [8, 31, 16], we found overall a three-fold spatial organization of the cerebellum: For most regions, we found a primary representation located between lobulus I and Crus I, a secondary representation between lobulus Crus II and lobulus VIIIb, and a tertiary representation in lobule IX or X. The ordering of the regions was mirrored around the horizontal fissure, such that the demand region formed a shell around the social-linguistic regions, and the action and motor regions again formed a shell around the demand regions. While regions S2-S4 appeared on the flatmap [19] to be spatially contiguous, the volumetric view revealed S3 that these regions too have anatomically distinct primary and secondary representations, separated by the horizontal fissure. This observation exemplifies the importance of considering how regions are distributed on a fully unfolded cerebellar cortical sheet [32] instead of solely relying on the crude approximation that is offered by the flatmap visualization [19].

The group probability maps also show a third representations of cognitive regions in lobule IX. No third motor representation was found in the cerebellar hemispheres. Instead, a third representation of the motor regions in the inferior vermis has recently been described at the individual level using deep phenotyping approaches [25]. Our atlas, which included this data within the training set, now clearly shows this representation both at the group and the individual level S2.

Damage to the primary motor representations leads to more severe deficits that damage to the secondary motor representation [33]. Based on this observation, it has been speculated that there are functional differences between the three representations [31]. So far, however, a definite demonstration of distinct response profiles among the three representations has remained elusive. Two lines of evidence cast doubt on the existence of functional differences between these representations. First, our analysis of functional regions generally grouped the three representations together, implying a significant degree of shared functional profiles across datasets. Second, tracing studies have shown that a single axon from the inferior olive can branch into multiple climbing fibers [34] and innervate different regions in non-contiguous lobules [35]. Because climbing fiber input is believed to organize Purkinje cells into functional modules [36], these findings suggest that all three representations, despite their spatial separation, may receive very similar climbing fiber and mossy fiber inputs, and also project to similar deep cerebellar nucleus targets. Therefore, it is not clear whether the multiple representations of the same functional region can be functionally distinguished. To facilitate further investigations, we provide an atlas version in which each region is subdivided into a superior (lobule I - Crus I inclusive), dorsal inferior (Crus II - VIIIb), ventral inferior (lobule IX - lobule X) and vermal inferior sections (vermis VII - vermis X). However, we recommend inspecting the sectioned probability maps for their spatial contiguity, as hard spatial boundaries imposed at the group level can divide some spatially contiguous parcels into two parts (e.g. A4) .

New functional insights

Although the spatial pattern of most regions adheres to a three-fold organization, our new atlas reveals that several regions deviate from this principle, suggesting a more complex cerebellar functional organization. First, not all functional regions have all three representations, for example A3 and A4 only have an inferior representation, whereas M1 only has a superior representation (see supplementary Fig. S2a). Second, some regions with a primary and secondary representations are spatially connected in the volume. For example, S1 appears to have a distinct primary and secondary representation (supplementary Fig. S2a). However the volumetric view revealed that these representations are connected at the fundus of the horizontal fissure.

Furthermore, while our atlas confirms the well-known functional regions of the cerebellum, it also uncovers regions that have not been reported or only very recently identified. We describe two new action-related regions in lobules VIII and IX, notably A3 which is engaged during spatial simulation and A4 which activates when imagining or observing actions with emotional content. Furthermore, the atlas revealed 5 medial-to-lateral organized regions in Crus I and II. Similar regions have only been recently delineated using several hours of scan time [28, 16]. This work showed that the default mode network can be divided into two parts, one that is associated with episodic memory (network A), the other that is associated with social function (network B). Our atlas captures this distinction, with S4 showing good correspondence with default mode network A, sharing important features such as the spatial location [16], cortical connectivity pattern (Fig. 3c), and its involvement in episodic projection. Meanwhile, S2 and S3 are associated with default mode network B, as these regions are involved in social functions (theory-of-mind and romance movie tasks).

Individual precision mapping

Studying finely inter-digitated regions using a group-level atlas is very difficult. Inter-individual variability is generally high in the cerebellum [37], and our analysis (Fig. 5d) shows that the location and arrangement of the multi-demand and social-linguistic regions is especially variable across individuals. These results stress the importance of using an individualized approach when studying cognitive areas of the cerebellum [38, 39, 40].

The classic approach to individual localization is to run a short localizer scan (often 10 minutes) [41], based on the assumption that these individual-level boundaries reflect the subject's organization better than boundaries defined by a group map. However, our recent analysis suggests that substantial amount of scan data is required to predict individual functional data better than the group map [9]. We confirm this by showing that the probabilistic group map provided by our new atlas is as good as 20 min of individual data (Fig. 5d), rendering individual localization based on only 10 minutes of data suboptimal. Increasing the individual scan time [16] often is not desirable, especially in the clinical context.

Our new atlas offers an alternative, by optimally integrating even limited individual data (10 minutes) with the probabilistic group map. This integration yields a probabilistic map of regions in the individual that is better than both group and individual map.

To apply this approach to a new subject in a new study, one needs to acquire some independent individual localization data (see below). Our framework can then be used to train a new dataset-specific emission model that characterizes - for each cerebellar region - the average group response on the tasks contained in that localizer scan. The final individual parcellations are obtained through a simple integration step (see methods: Group and individual parcellations). This method enables the use of individual functional localization in studies for which the time with each individual is restricted. However, even when a long localizer scan is possible, our approach leads to significant improvement than using the individual data alone.

An important consideration for a precision mapping approach remains the decision of whether to use task-based or resting-state data, and - if using the former - which localizer tasks to include. Generally, it seems advisable to include a set of anchor tasks able to activate each region of interest. We observed that task-based datasets that focused on a narrow functional domain resulted in precise estimates of boundaries for regions of that domain at the expense of region boundaries for other domains (Fig. 1a). In addition to tasks that tap into the domain of interest, it appears also advisable to include task that activate spatially neighboring regions. For example, when aiming to study the language regions of the cerebellum[41], one should add tasks that activate the neighboring multi-demand regions (e.g. spatial map task, Supplementary Fig. S4a and b) to obtain a precise estimate of the functional boundary between social-linguistic and multi-demand regions, which appear especially variable. The development of a principled approach to design optimal task-sets for functional localization remains an important question for future research.

Lateralization

The cerebellum's importance in lateralized higher-order functions, particularly language, has reignited interest in lateralization studies of the cerebellum [6]. Studies of hemispheric specialization are most easily performed using a functional atlas that has regions matched in size and location across hemispheres, while faithfully representing functional boundaries. Prior studies had to rely on anatomical parcellations to examine hemispheric differences in cerebellar development [6] or neurochemistry [42], even though these are not good descriptions of functional subdivisions [3]. Our symmetric atlas addresses this gap. We

show that the symmetry constraint had only a relatively small impact on the ability of the atlas to identify functional subdivisions.

Cortico-cerebellar connectivity

For each of the cerebellar regions, our framework also provides a neocortical connectivity pattern. We showed that a model that integrates data across diverse task-based dataset outperforms our previous model that was only trained on the MDTB dataset [29]. These patterns of neocortical connectivity not only provide an additional description of the identified regions but have two further practical applications.

First, being able to identify a cerebellar region by its neocortical pattern of connectivity allows the use of resting-state data for the functional localization of these regions in single individuals. This enables the extension of the atlas to patient groups and young children and allows users to leverage the broadly available resting-state datasets.

Secondly, the independent identification of the neocortical regions that communicate with each cerebellar region is an important prerequisite for further studies that investigates the relative functional roles of neocortical and cerebellar areas within the same functional module [43]. We therefore believe that the new atlas will provide an important resource for the study of the human cerebellum going forward.

METHODS

Datasets and data organization

We used seven task-based and one resting-state fMRI datasets (see Supplemental Table 1). The first four datasets each comprised a broad battery of tasks tapping into cognitive, motor, perceptual, and social functions: (1) The *Multi-Domain Task Battery* dataset (MDTB, [3]), (2) a high-resolution version of the MDTB (*High-res MDTB*; not yet published), (3) the *Nakai & Nishimoto* dataset [44], and the (4) The *Individual Brain Charting (IBC)* dataset [45, 46]. We also included three further datasets to obtain a better description of the domains of motor movements and executive functions: (5) the working memory (*WM*) dataset [43] which included finger movements and a forward / backwards digit span task; (6) the *Multi-Demand* dataset [27] which included a no-go, N-back, and task-switch task ; and (7) the *Somatotopic* dataset [25] which probed foot, hand, glutes, and tongue movements. Finally, we used the resting-state fMRI dataset *Unrelated 100* subjects, which is made publicly available in the *Human Connectome Project (HCP)* S1200 release [47].

The task-based datasets were preprocessed as described in [9]. For each run and condition, we estimated one contrast image, and divided it by the root-mean-square-error from the first-level GLM to obtain a normalized activation estimate for each condition. These values served as the input data for all subsequent analyses. No smoothing or group normalization was applied at this stage. For the HCP resting-state data, we used minimally preprocessed time series [48]. The preprocessing pipeline included structural registration, correction for spatial distortion, head motion, cortical surface mapping, and functional artifact removal [48, 49]. This resulted in 1200 time points of processed time series per imaging run per cerebellar voxel of the standard MNI152 template [50]. To obtain resting-state functional connectivity (rs-FC) fingerprints of the cerebellar voxels, we used a group Independent Component Analysis (ICA). We applied the group-ICA implemented in FSL's MELODIC [51] with automatic dimensionality estimation to the temporally concatenated functional data of all subjects, sessions and runs, resulting in 1072 group-level components. Sixty-nine signal components were identified from the first 300 ICA components as resting-state networks. We then regressed the 69 group network spatial maps into the subject-and-run-specific cortical time series, resulting in 69 cortical network time courses. The cerebellar rs-FC fingerprints were calculated as Pearson's correlations of the cerebellar voxel time series with each cortical network time course.

Using a unified code framework (available at `github.com/diedrichsenlab/Functional_Fusion`), the data were then extracted in two atlas spaces. For the cerebellum, we computed the non-linear morph into the Symmetric MNI152NLin2009aSym template (`http://nist.mni.mcgill.ca/?p=904`). The functional data was resampled to 2mm isotropic resolution and a cerebellar gray matter mask with 18290 voxels was applied. After masking, a smoothing kernel of 2mm was applied. For the cortical-cerebellar connectivity models, the same data were also extracted in the symmetric freesurfer32LR surface [52].

Hierarchical Bayesian parcellation framework

To integrate different dataset into a unified probabilistic parcellation atlas, we utilized a newly developed Hierarchical Bayesian Framework [for full details 9]. In short, the framework integrates different fMRI datasets, $\mathbf{Y}^{s,n}$, recorded in different sessions (n) from different subjects (s). The model assigns each of the possible brain locations in each individual to one of K functional parcels, with $\mathbf{U}^s_{k,i} = 1$ indicating that the ith voxel is part of the kth parcel. The model estimates the expected value of these parcel assignments, which provides a probabilistic parcellation for that individual.

The model consists of two parts: First, a collection of dataset-specific *emission models* that specify the probability of each observed dataset given the individual brain parcellation, $p(\mathbf{Y}^{s,n}|\mathbf{U}^s)$. Here, we used a van-Mises-Fisher mixture model, in which each parcel had a mean vector $\mathbf{v}^n_k$ for each session, and a separate concentration parameter for each session [κ^n, Model Type 2, see 9]. Each emission model therefore had the parameters $\theta^n_E = \{\mathbf{v}^n_1, ..., \mathbf{v}^n_k, \kappa^n\}$

The second component, the *arrangement model*, specifies the group probability of each brain location belonging to a specific parcel. Here we used a model that treated each voxel independently, with $p(\mathbf{U}^s_{k,i}) = \mathrm{softmax}(\eta_{\mathrm{k,i}})$. The *KxP* arrangement model parameters $\theta_A = \{\eta_{1,1}, ...\}$ could therefore be estimated by averaging across all the individual probability maps.

The parameters of the spatial arrangement models and the emission models were estimated together using an EM-algorithm. We used 5000 different random starting values to avoid local minima. For

computational reasons, the initial fitting and evaluation was done using a 3mm isotropic voxel resolution - the final selected model was upsampled to 2mm and used as a starting value to refit to the higher resolution data.

Symmetry constraint

To achieve a spatially symmetric parcellation, we developed a version of the arrangement model, where parcels $1...K/2$ were restricted to the left hemisphere, and parcel $K/2+1,...,K$ to the right. The assignment of voxels to parcels was symmetric - that is if the left hemisphere voxel was assigned to parcel 1, the corresponding right hemispheric voxel was assigned to parcel $K/2+1$. As a consequence, symmetric brain locations were assigned to corresponding parcels. The mean functional profiles $\mathbf{v}_k^n$, however, were estimated separately for the left and right hemispheric parcels This allowed us to derive a spatially symmetric parcellation of the cerebellum while still capturing the functional specialization of each hemisphere.

We estimated a corresponding asymmetric atlas where the parcel boundaries were unconstrained by hemispheric symmetry. To make the asymmetric atlas comparable to the symmetric version, we retained the constraint that one half of the regions were in the left, the other half in the right hemisphere. Furthermore, we used the fitted emission models (mean functional profiles) from the symmetric model, only refitting the arrangement model without the symmetry constraint. This resulted in an asymmetric version of the atlas in which the regions had the same functional profiles as in the symmetric version.

Group and individual parcellations

After fitting the parameters $\{\theta_A, \theta_E^1, ..., \theta_E^N\}$, the model can be used to derive both a group and individual parcellation maps. The probabilistic group parcellation is based only on the arrangement model, which directly specifies $p_{group} = p(\mathbf{U})$ for each voxel and parcel. Each individual parcellation is based on some *individual training* data, $\mathbf{Y}_s^n$. The data-only parcellation only depends on the corresponding emission model, with $p_{data,s} \propto p(\mathbf{Y}_s^n|\mathbf{U}_s)$. In contrast, the full individual parcellation integrates the probability from both emission and arrangement model $p_{indiv,s} \propto p(\mathbf{Y}_s^n|\mathbf{U}_s)p(\mathbf{U}_s)$, using Bayes rule. For visualization and evaluation of functional boundaries, both group and individual probabilistic parcellation were transformed into hard parcellations by assigning each voxel the parcel with the highest probability.

Individual precision mapping

Our model provides a probabilistic group map (spatial arrangement model) and a probabilistic estimate of parcel membership based on a specific individual data set (using a dataset-specific emission model). By integrating these using Bayes rule, an optimal estimate of brain organization for a new individual can be obtained [9]. For the analysis presented in Fig. 5, we used 1-16 runs of data from the first task set of the MDTB dataset as training. The individual maps were then evaluated on the second task set, which contained 8 overlapping and 9 novel tasks [3].

To apply this approach to new subjects with individual localizing data that is different from the task sets included in our atlas, the user would first estimate a new emission model from the data of all individuals in the study. This new dataset-specific emission model can be used to localize regions in new individuals, given their data.

Single-dataset parcellations and similarity analysis of parcellations

To compare the differences between parcellations derived from different datasets, we trained the model on each dataset separately, estimating parcellation maps with 10, 20, 34, 40 and 68 regions. As an index of parcellation similarity, we calculated the adjusted Rand Index (ARI) between the winner-take-all voxel assignments of the resulting parcellations. The ARI was calculated across all 5 levels of granularity, resulting in a 5x5 matrix of ARIs for each dataset pair. Different datasets are differently reliable which could affect the similarity of two datasets. We therefore estimated the reliability of the parcellation by averaging the ARIs between different levels of granularity within each dataset, with the idea that reliable datasets should result in parcellations that are consistent across granularities. We then divided the ARI (also average across levels of granularity) between two datasets by the geometric mean of the two average within-dataset ARIs. This index served as a reliability corrected measure of correspondence between parcellations.

Finally, we used classic multi-dimensional scaling to visualize the structure of similarities between different parcellations. We calculated the first two eigenvectors of the square matrix of adjusted between-

dataset similarities. The space defined by these two vectors optimally reproduces the overall similarity structure, with the dissimilarity (1-ARI) between two datasets reflected in the Euclidean distance between the two.

DCBC evaluation

To assess how well a given parcellation can predict functional boundaries in the cerebellum, we utilized the Distance-Controlled Boundary Coefficient (DCBC) [24]. This metric compares the correlation between voxel-pairs within a parcel to the correlation between voxel-pairs across a boundary, while accounting for spatial distance. Our evaluation included both the group parcellation (DCBC group) and individual parcellations (DCBC individual) obtained from this group atlas.

Both group and individual DCBC were calculated in a cross-validated fashion, leaving out the test dataset during training of the overall model. The group DCBC was calculated by deriving a winner-take-all parcellation from the group probability map and evaluating the ability of these group-based boundaries to predict functional boundaries in each individual.

To calculate the DCBC for individual parcellations, we used a localizer-like approach for individual precision mapping (see methods: individual precision mapping): One half of the test dataset served as the localizer data. First, we estimated a dataset-specific emission model for the localizer dataset across all subjects. Then, we used the localizer data from one specific subject to estimate the individual boundaries (see methods: group and individual parcellations). Individual boundaries were derived using a winner-take-all approach on the subject's resultant individual probability map. These were then tested for their ability to predict functional boundaries in the second half of the subject's data. We then reversed the role of the two halves of the test set averaged performance across the two within-subject cross-validation folds. To make the evaluation of group-based and individual-based boundaries comparable, we also calculated the group DCBC by splitting each subject's data in half and then averaging the performance across the two halves after individual DCBC calculation.

Parcel similarity and clustering

To develop a hierarchically organized system of maps, we started with the symmetric map with 68 parcels (34 per hemisphere) as our base. For clustering we derived a functional similarity index between parcels. We first averaged the estimated mean response vectors for each parcel and session $\mathbf{v}_k^n$ across the left and right hemisphere, and then calculated the cosine similarity between each pair of parcels. We then took the weighted average of these cosine-similarities across sessions and datasets, with the weight of each session set to product if the dispersion parameter κ^n and number of subjects for that session N^n.

We then iteratively merged the smallest parcels into the functionally most similar parcel, until all parcels had at least one voxel win the winner-take-all assignment, resulting in 32 parcels (16 per hemisphere). When merging parcels, we summed their probability maps to obtain the probability of a voxel to belong to the combined parcel. The emission models for the combined model were then refit to the data, keeping the probabilities in the arrangement model fixed. In a last step, we grouped the 32 parcels (again, based on their functional profiles) into 4 domains. The labels for each parcel then followed the organization of Domain-Region-Hemisphere-Subregion.

The colormap for our functional atlas was based on the weighted cosine similarity of the functional profiles (see above). We used classical multi-dimensional-scaling to represent these similarities in a 3-dimensional space. This arrangement was then projected into RGB space, with 5 anchor points ensuring a consistent coloring across maps. As a result, the similarity of color of different parcels can be directly interpreted as an approximation of their functional similarity across task datasets.

Functional lateralization and Boundary symmetry

To study lateralization, we assessed the symmetry of the functional profiles of left-right voxel pairs. For this, we calculated the cosine similarity of the functional profiles of each voxel pair. Functional profiles were obtained by averaging the estimated mean response vectors for each voxel in each session. The cosine similarities were then weighted by the session weight κ^n and the number of subjects N^n, for session n.

To investigate left-right boundary symmetry in the cerebellum, an asymmetric version of the atlas was estimated (see methods: Symmetry constraint). An index of boundary symmetry was calculated as the correlation between the parcel probability vectors of the asymmetric and the symmetric atlas for each voxel, either for the group map, or for the individual parcellations. For visualization, the correlation

values within all datasets, excluding the Nishimoto and IBC dataset due to the relatively low reliabilities, were averaged across individuals.

Cortical connectivity

Connectivity models were fitted for each individual (and dataset) separately. As described in King et al. [29], we parcellated the neocortex into 1876 parcels using a regular icosahedron. For task-based data we used the normalized activity estimates, for the resting-state data, the preprocessed time series (see methods: Datasets and data organization). These data were averaged across all voxels in each neocortical ROI, forming the NxQ matrix $\mathbf{X}$. The same data was extracted for each cerebellar voxel in atlas space. The connectivity weights were then estimated to form the best predictive model $\mathbf{Y} = \mathbf{XW}$ using Ridge-regression. The ridge coefficient was tuned for each dataset separately to yield the best prediction performance on all the other datasets.

For evaluation, we averaged the connectivity weight across all subjects in each training dataset. For each individual in the evaluation dataset, we used the cortical activity measures and the average connectivity weights to predict the individual cerebellar activity patterns. We then calculated the cosine similarity between the predicted and observed cerebellar activity [29].

When evaluating a connectivity model on the same dataset it was trained on, we adopted a leave-one-subject out approach. For each individual, the connectivity weights were averaged across all other individuals in that dataset, and then applied to make the prediction for that single subject.

Finally, we investigated if an integration across all datasets would increase the predictive power of the connectivity model. For this we simply averaged connectivity models across all task-based datasets, always taking care to leave the particular evaluation subject out of the averaging of the connectivity weights.

Functional profiles for the MDTB dataset

To characterize the functional profile of each cerebellar region, we calculated the mean task response of all parcels in the MDTB dataset. These functional profiles were the normalized activation estimates (see methods: Dataset and Data Organization), averaged across the individualized regions within each individual. To account for activation that can be explained by the motor aspects of each task, we used the number of movements in each condition (left hand presses, right hand presses and saccades per second) as a covariate alongside regressors that coded for each condition separately [3]. The columns of the design matrix and the average functional profiles were z-normalized across conditions. We estimated a linear model using ridge regression (L2 regularization) to arrive at a final estimate for the motor features and task-activations.

Inter-individual variability

To quantify inter-individual variability in the cerebellum, we calculated Pearson's correlation coefficient of each voxel's response profile pairwise between all subjects within each dataset. To account for the measurement noise, we derived two independent estimates for each subject and voxel: one from the first half, the other from the second half of the data. Correlations were computed on the concatenated two profiles and the reliability was calculated by correlating the two independent estimates of the response profile within each subject. The inter-subject correlation was normalized by dividing each value by the square root of the product of the two subject's reliabilities. For purposes of visualization of each voxel's inter-individual variability, we averaged the inter-subject correlation values across subjects and divided it by the reliability averaged across subjects, obtaining a single value per voxel. These voxel values were projected to the flatmap.

REFERENCES

[1] Peter L Strick, Richard P Dum, and Julie A Fiez. Cerebellum and nonmotor function. *Annu. Rev. Neurosci.*, 32:413–434, 2009.

[2] Jeremy D Schmahmann. Disorders of the cerebellum: Ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. *J. Nurse Pract.*, 16(3):367–378, August 2004.

[3] Maedbh King, Carlos R Hernandez-Castillo, Russell A Poldrack, Richard B Ivry, and Jörn Diedrichsen. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nature neuroscience*, 22(8):1371–1378, 2019.

[4] J D Schmahmann, J Doyon, D McDonald, C Holmes, K Lavoie, A S Hurwitz, N Kabani, A Toga, A Evans, and M Petrides. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *Neuroimage*, 10(3 Pt 1):233–260, 1999.

[5] J Diedrichsen, J H H Balsters, J Flavell, E Cussans, and N Ramnani. A probabilistic MR atlas of the human cerebellum. *Neuroimage*, 46(1):39–46, 2009.

[6] Carolin Gaiser, Rick van der Vliet, Augustijn de Boer, Opher Donchin, Pierre Berthet, Gabriel Devenyi, Mallar Chakravarty, Jörn Diedrichsen, Andre Marquand, Maarten Frens, and Ryan Muetzel. Large data on the small brain: Population-wide cerebellar growth models of children and adolescents, 2023.

[7] Jie Lisa Ji, Marjolein Spronk, Kaustubh Kulkarni, Grega Repovš, Alan Anticevic, and Michael W Cole. Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage*, 185:35–57, 2019.

[8] Randy L Buckner, Fenna M Krienen, Angela Castellanos, Julio C Diaz, and BT Thomas Yeo. The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(5):2322–2345, 2011.

[9] Da Zhi, Ladan Shahshahani, Caroline Nettekoven, Ana Luisa Pinho, Danilo Bzdok, and Joern Diedrichsen. A hierarchical bayesian brain parcellation framework for fusion of functional imaging datasets. *bioRxiv*, 2023.

[10] Janine Bijsterbosch, Samuel J Harrison, Saad Jbabdi, Mark Woolrich, Christian Beckmann, Stephen Smith, and Eugene P Duff. Challenges and future directions for representations of functional brain organization. *Nature neuroscience*, 23(12):1484–1495, 2020.

[11] Timothy O Laumann, Evan M Gordon, Babatunde Adeyemo, Abraham Z Snyder, Sung Jun Joo, Mei-Yen Chen, Adrian W Gilmore, Kathleen B McDermott, Steven M Nelson, Nico UF Dosenbach, et al. Functional system and areal organization of a highly sampled individual human brain. *Neuron*, 87(3):657–670, 2015.

[12] Rodrigo M Braga and Randy L Buckner. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron*, 95(2):457–471, 2017.

[13] Evan M Gordon, Timothy O Laumann, Babatunde Adeyemo, Adrian W Gilmore, Steven M Nelson, Nico UF Dosenbach, and Steven E Petersen. Individual-specific features of brain systems identified with resting state functional correlations. *Neuroimage*, 146:918–939, 2017.

[14] Sophia Mueller, Danhong Wang, Michael D Fox, BT Thomas Yeo, Jorge Sepulcre, Mert R Sabuncu, Rebecca Shafee, Jie Lu, and Hesheng Liu. Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3):586–595, 2013.

[15] Scott Marek, Joshua S Siegel, Evan M Gordon, Ryan V Raut, Caterina Gratton, Dillan J Newbold, Mario Ortega, Timothy O Laumann, Babatunde Adeyemo, Derek B Miller, et al. Spatial and temporal organization of the individual human cerebellum. *Neuron*, 100(4):977–993, 2018.

[16] Aihuiping Xue, Ru Kong, Qing Yang, Mark C. Eldaief, Peter Angeli, Lauren M DiNicola, Rodrigo M. Braga, Randy L Buckner, and B.T. Thomas Yeo. The Detailed Organization of the Human Cerebellum Estimated by Intrinsic Functional Connectivity Within the Individual. *J. Neurophysiol.*, 1(September 2020):358–384, 2020.

[17] S. E. Petersen, P. T. Fox, M. I. Posner, M. Mintun, and M. E. Raichle. Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, 1(2):153–170, 1989.

[18] John Duncan, Moataz Assem, and Sneha Shashidhara. Integrated intelligence from distributed brain activity. *Trends in cognitive sciences*, 24:838–852, 10 2020.

[19] Jörn Diedrichsen and Ewa Zotow. Surface-based display of volume-averaged cerebellar imaging data. *PloS one*, 10(7):e0133402, 2015.

[20] Michael W Cole, Danielle S Bassett, Jonathan D Power, Todd S Braver, and Steven E Petersen.

Intrinsic and task-evoked network architectures of the human brain. *Neuron*, 83(1):238–251, 2014.
[21] I Tavor, O Parker Jones, R B Mars, S M Smith, T E Behrens, and S Jbabdi. Task-free MRI predicts individual differences in brain activity during task performance. *Science*, 352(6282):216–220, 2016.
[22] Uri Hasson, Howard C Nusbaum, and Steven L Small. Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences*, 106(26):10841–10846, 2009.
[23] Abigail S Greene, Siyuan Gao, Stephanie Noble, Dustin Scheinost, and R Todd Constable. How tasks change whole-brain functional organization to reveal brain-phenotype relationships. *Cell reports*, 32(8):108066, 2020.
[24] Da Zhi, Maedbh King, Carlos R Hernandez-Castillo, and Jörn Diedrichsen. Evaluating brain parcellations using the distance-controlled boundary coefficient. *Human Brain Mapping*, 43(12):3706–3720, 2022.
[25] Noam Saadon-Grosman, Peter A Angeli, Lauren M DiNicola, and Randy L Buckner. A third somatomotor representation in the human cerebellum. *Journal of Neurophysiology*, 128(4):1051–1073, 2022.
[26] Daniel Marten van Es, Wietske van der Zwaag, and Tomas Knapen. Retinotopic maps of visual space in the human cerebellum. *Current Biology*, 1 2019.
[27] Moataz Assem, Sneha Shashidhara, Matthew F Glasser, and John Duncan. Basis of executive functions in fine-grained architecture of cortical and subcortical human brain networks. *bioRxiv*, 2022.
[28] Lauren M. DiNicola, Rodrigo M. Braga, and Randy L. Buckner. Parallel distributed networks dissociate episodic and social functions within the individual. *Journal of Neurophysiology*, 123(3):1144–1179, 2020.
[29] Maedbh King, Ladan Shahshahani, Richard B Ivry, and Jörn Diedrichsen. A task-general connectivity model reveals variation in convergence of cortical inputs to functional regions of the cerebellum. *eLife*, 12, 4 2023.
[30] BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106:1125–1165, 2011.
[31] Xavier Guell, Jeremy D Schmahmann, John DE Gabrieli, and Satrajit S Ghosh. Functional gradients of the cerebellum. *Elife*, 7:e36652, 2018.
[32] Martin I. Sereno, J. rn Diedrichsen, Mohamed Tachrount, Guilherme Testa-Silva, Helen D Arceuil, and Chris De Zeeuw. The human cerebellum has almost 80neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, 117(32):19538–19543, 2020.
[33] Catherine J. Stoodley, Jason P. MacMore, Nikos Makris, Janet C. Sherman, and Jeremy D. Schmahmann. Location of lesion determines motor vs. cognitive consequences in patients with cerebellar stroke. *NeuroImage: Clinical*, 12:765–775, 2016.
[34] Renée F. Schild. On the inferior olive of the albino rat. *Journal of Comparative Neurology*, 1970.
[35] I. Sugihara, H. S. Wu, and Y. Shinoda. The entire trajectories of single olivocerebellar axons in the cerebellar cortex and their contribution to cerebellar compartmentalization. *Journal of Neuroscience*, 2001.
[36] David J. Herzfeld, Yoshiko Kojima, Robijanto Soetedjo, and Reza Shadmehr. Encoding of action by the Purkinje cells of the cerebellum. *Nature*, 526(7573):439–441, 2015.
[37] Scott Marek, Joshua S. Siegel, Evan M. Gordon, Ryan V. Raut, Caterina Gratton, Dillan J. Newbold, Mario Ortega, Timothy O. Laumann, Babatunde Adeyemo, Derek B. Miller, Annie Zheng, Katherine C. Lopez, Jeffrey J. Berg, Rebecca S. Coalson, Annie L. Nguyen, Donna Dierker, Andrew N. Van, Catherine R. Hoyt, Kathleen B. McDermott, Scott A. Norris, Joshua S. Shimony, Abraham Z. Snyder, Steven M. Nelson, Deanna M. Barch, Bradley L. Schlaggar, Marcus E. Raichle, Steven E. Petersen, Deanna J. Greene, and Nico U.F. Dosenbach. Spatial and Temporal Organization of the Individual Human Cerebellum. *Neuron*, 100(4):977–993.e7, 2018.
[38] Frank Van Overwalle, Min Pu, Qianying Ma, Meijia Li, Naem Haihambo, Kris Baetens, Natacha Deroost, Chris Baeken, and Elien Heleven. The Involvement of the Posterior Cerebellum in Reconstructing and Predicting Social Action Sequences. *Cerebellum*, 21(5):733–741, 2022.
[39] Anila M. D'Mello, Peter E. Turkeltaub, and Catherine J. Stoodley. Cerebellar tdcs modulates neural circuits during semantic prediction: A combined tDCS-fMRI study. *Journal of Neuroscience*,

37(6):1604–1613, 2017.
[40] Anila M. D'Mello, John D.E. Gabrieli, and Derek Evan Nee. Evidence for Hierarchical Cognitive Control in the Human Cerebellum. *Current Biology*, 30(10):1881–1892.e3, 2020.
[41] Evelina Fedorenko, Po Jang Hsieh, Alfonso Nieto-Castañón, Susan Whitfield-Gabrieli, and Nancy Kanwisher. New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, 2010.
[42] Caroline Nettekoven, Leah Mitchell, William T. Clarke, Uzay Emir, Jon Campbell, Heidi Johansen-Berg, Ned Jenkinson, and Charlotte Stagg. Cerebellar GABA Change during Visuomotor Adaptation Relates to Adaptation Performance and Cerebellar Network Connectivity: A Magnetic Resonance Spectroscopic Imaging Study. *The Journal of Neuroscience*, 42(41):JN–RM–0096–22, sep 2022.
[43] Ladan Shahshahani, Maedbh King, Caroline Nettekoven, Richard Ivry, and Jörn Diedrichsen. Selective recruitment: Evidence for task-dependent gating of inputs to the cerebellum. *bioRxiv*, pages 2023–01, 2023.
[44] Tomoya Nakai and Shinji Nishimoto. Quantitative models reveal the organization of diverse cognitive functions in the brain. *Nature communications*, 11(1):1–12, 2020.
[45] Ana Luísa Pinho, Alexis Amadon, Torsten Ruest, Murielle Fabre, Elvis Dohmatob, Isabelle Denghien, Chantal Ginisty, Séverine Becuwe-Desmidt, Séverine Roger, Laurence Laurier, et al. Individual brain charting, a high-resolution fmri dataset for cognitive mapping. *Scientific data*, 5(1):1–15, 2018.
[46] Ana Luísa Pinho, Alexis Amadon, Baptiste Gauthier, Nicolas Clairis, André Knops, Sarah Genon, Elvis Dohmatob, Juan Jesús Torre, Chantal Ginisty, Séverine Becuwe-Desmidt, et al. Individual brain charting dataset extension, second release of high-resolution fmri data for cognitive mapping. *Scientific Data*, 7(1):353, 2020.
[47] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.
[48] Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R. Polimeni, David C. Van Essen, and Mark Jenkinson. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124, 2013.
[49] Stephen M Smith, Christian F Beckmann, Jesper Andersson, Edward J Auerbach, Janine Bijsterbosch, Gwenaëlle Douaud, Eugene Duff, David A Feinberg, Ludovica Griffanti, Michael P Harms, et al. Resting-state fmri in the human connectome project. *Neuroimage*, 80:144–168, 2013.
[50] David C Van Essen, Matthew F Glasser, Donna L Dierker, John Harwell, and Timothy Coalson. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral cortex*, 22(10):2241–2262, 2012.
[51] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.
[52] David C. Van Essen, Matthew F. Glasser, Donna L. Dierker, John Harwell, and Timothy Coalson. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral cortex (New York, N.Y. : 1991)*, 22:2241–2262, 10 2012.

ACKNOWLEDGEMENTS

This study was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2016-04890), and a project grant from the Canadian Institutes of Health Research (CIHR, PJT 159520), both to J.D. Additional funding came from the Canada First Research Excellence Fund (BrainsCAN) to Western University. We would like to acknowledge Suzanne Witt for help in collecting and preprocessing the MDTB-Highres dataset. Special thanks to N. Grosman, R. Buckner, M. Assem, and J. Duncan for sharing their datasets before the official public release.

COMPETING INTERESTS

The authors declare no competing interests.

SUPPLEMENTARY MATERIALS AND FIGURES

Name	Subjects	No. conditions	min/subject	Voxel size (mm)	Description
MDTB	24	62	320	3T, 3mm	Cognitive, motor, perceptual, social
Highres-MDTB	7	9	120	7T, 1.5mm	Cognitive, motor, perceptual, social
Nishimoto	6	103	162	3T, 2mm	Cognitive, motor, perceptual, social
IBC	12	208	822	3T, 1.5mm	Cognitive, motor, perceptual, social
WM	16	17	65	3T, 3mm	Motor and working memory task
Multi-demand	37	12	100	3T, 2mm	Executive Tasks
Somatotopic	8	6	96	1.8/2.4	Motor
HCP-Unrelated 100	100	none	60	3T, 2mm	Resting-state

Table 1. FMRI datasets used for the functional fusion. All datasets but the last are task-based. The last one refers to resting-state data from a subset of the HCP dataset.

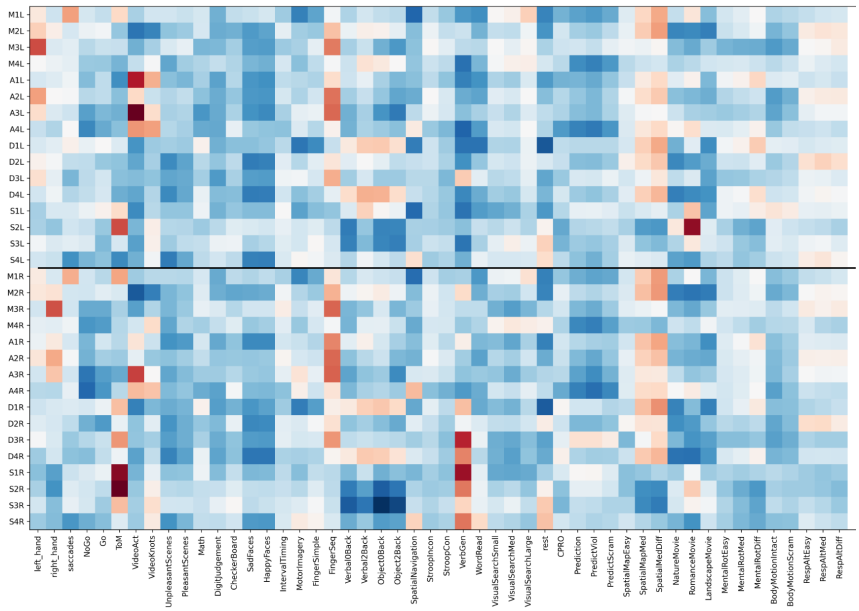

Figure S1. Functional profiles of regions in MDTB dataset. Mean evoked responses in MDTB dataset corrected for motor features. Responses were estimated from subject-specific regions and averaged across subjects for visualization. To account for activation that can be explained by the motor aspects of each task, number of movements were used as covariates alongside regressors that coded for each condition separately. Movements were left hand presses, right hand presses and saccades per second.

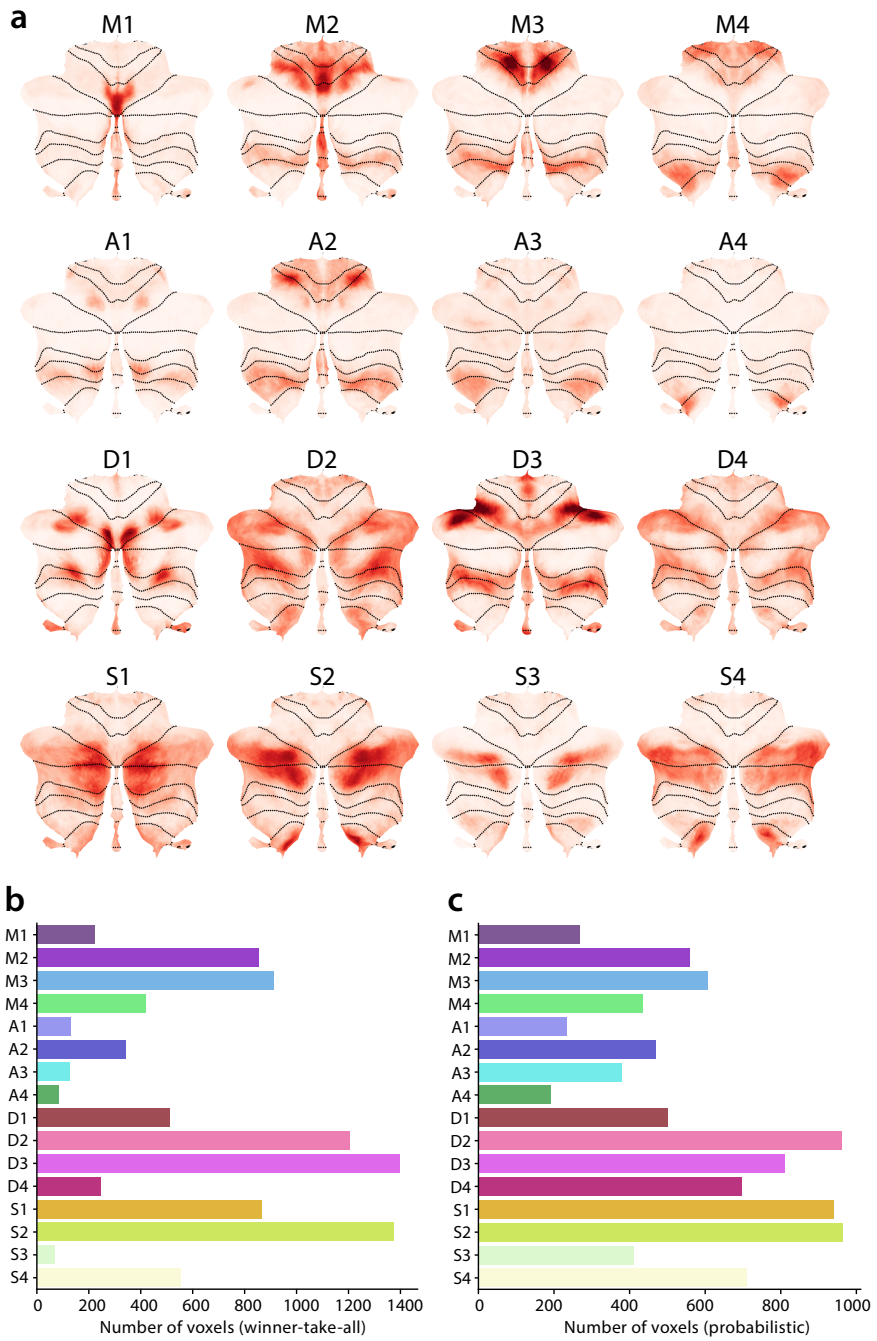

Figure S2. Probability maps and region size. **a**, Probability maps for each region displayed on the flat representation. **b**, Size estimate for each region in terms of the number of voxels ($2mm^3$) using winner-take-all assignment. **c**, Size estimate for each region in terms of the number of voxels ($2mm^3$) using probabilistic assignments.

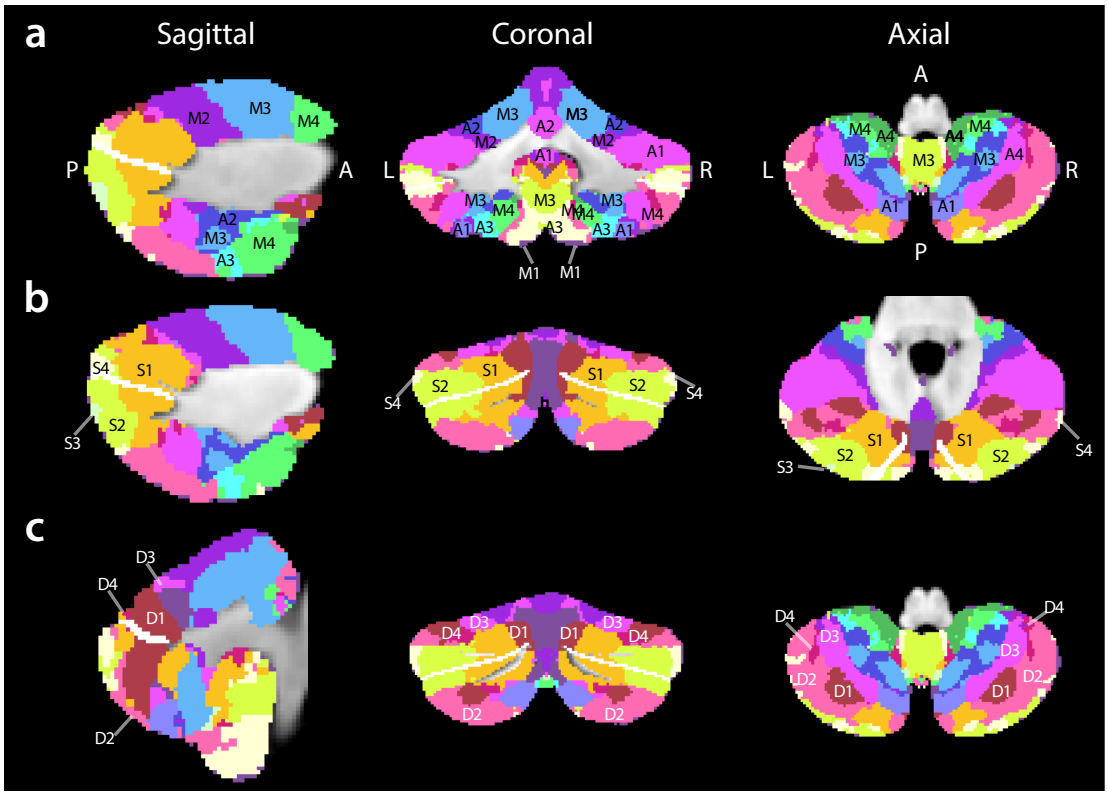

Figure S3. Atlas in volumetric space. Atlas shown at medium granularity (32 regions; 16 per hemisphere). Top row shows motor and action regions, middle row shows multi-demand regions and bottom row shows social-linguistic regions. Horizontal fissure is marked in white.

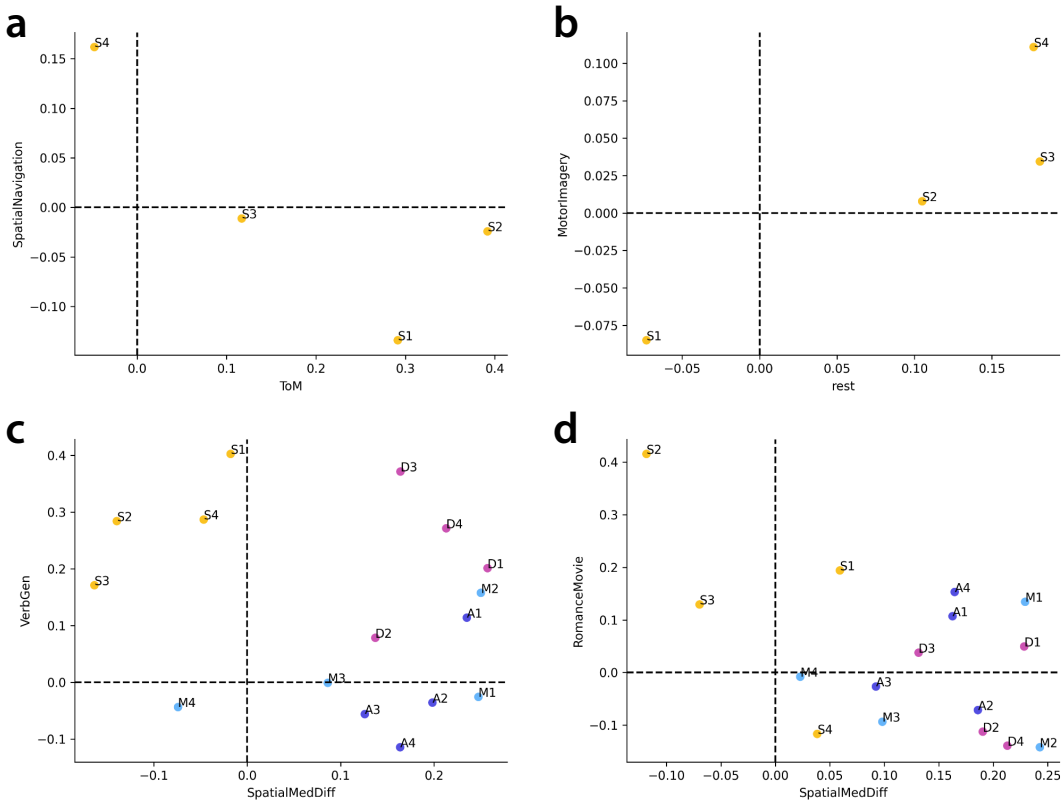

Figure S4. Regional differences in functional responses for selected tasks. **a-b**, Tasks separating social-linguistic (S1-4) regions and **c-d**, social-linguistic regions from other domains. **a**, S4 separated from other S regions. **b**, S1 separated from other S regions. **c**, S regions separated from D regions and **d**, S2 and S3 separated from other regions.

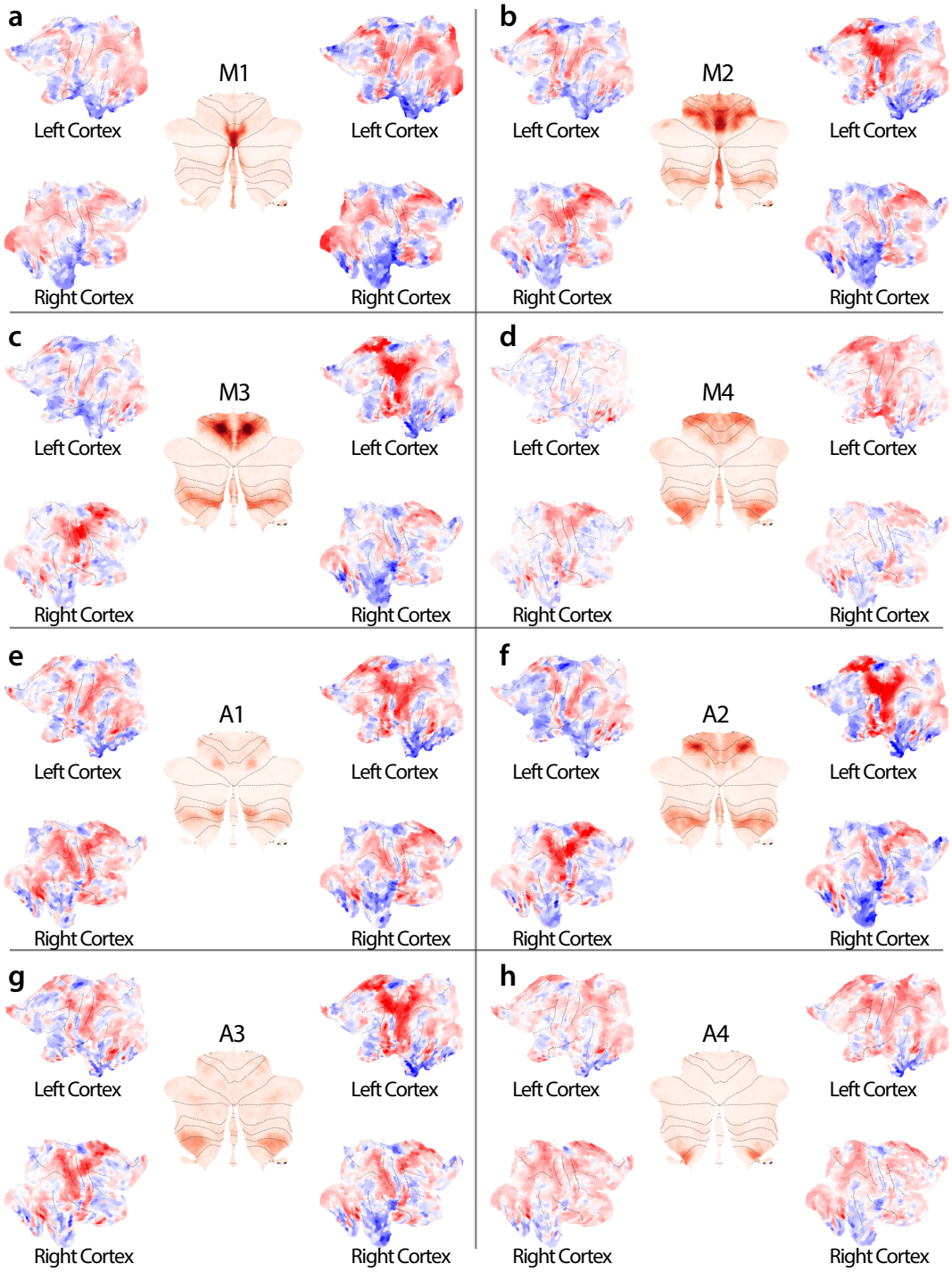

Figure S5. Cortico-cerebellar connectivity weights and probability maps. Parcel probability maps for motor (**a-d**) and action (**e-h**) regions are shown in the middle of each figure inset, surrounded by the cortical input weights for the left and right cerebellar parcel. Weights for the left cerebellar parcel are shown to the left of the probability map and for the right cerebellar parcel to the right of each probability map on the cortical flatmap. Motor regions include oculomotor vermis M1 (**a**), tongue and vermal region M2 (**b**), hand M3 (**c**)) and lower body M4 (**d**)) region. Action regions include spatial simulation regions A1 (**e**), classical action observation A2 (**f**), motor imagery A3 (**g**) and social action region A4 (**h**).

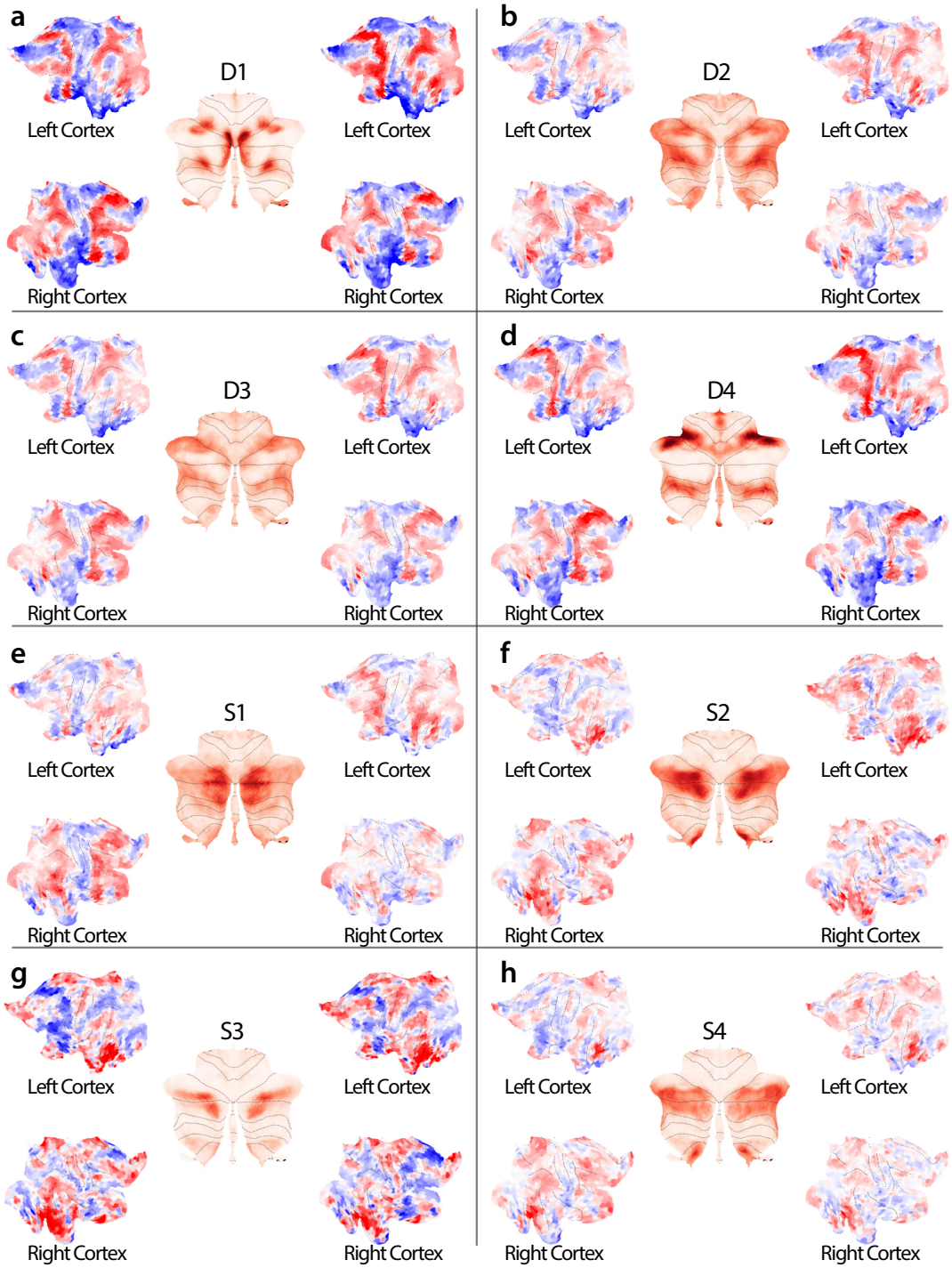

Figure S6. Cortico-cerebellar connectivity weights and probability maps for demand and social-linguistic regions. Parcel probability maps for multiple demand (**a-d**) and social-linguistic (**e-h**) regions are shown in the middle of each figure inset, surrounded by the cortical input weights for the left and right cerebellar parcel. Weights for the left cerebellar parcel are shown to the left of the probability map and for the right cerebellar parcel to the right of each probability map on the cortical flatmap. Demand regions include spatial working memory region (**a**), recall regions (**b**), difficulty-related (**c**)) and N-back region (**d**)) region. Social-linguistic regions include linguistic region S1 (**e**), social region S2 (**f**), rest region S3 (**g**) and self-projection region S4 (**h**)

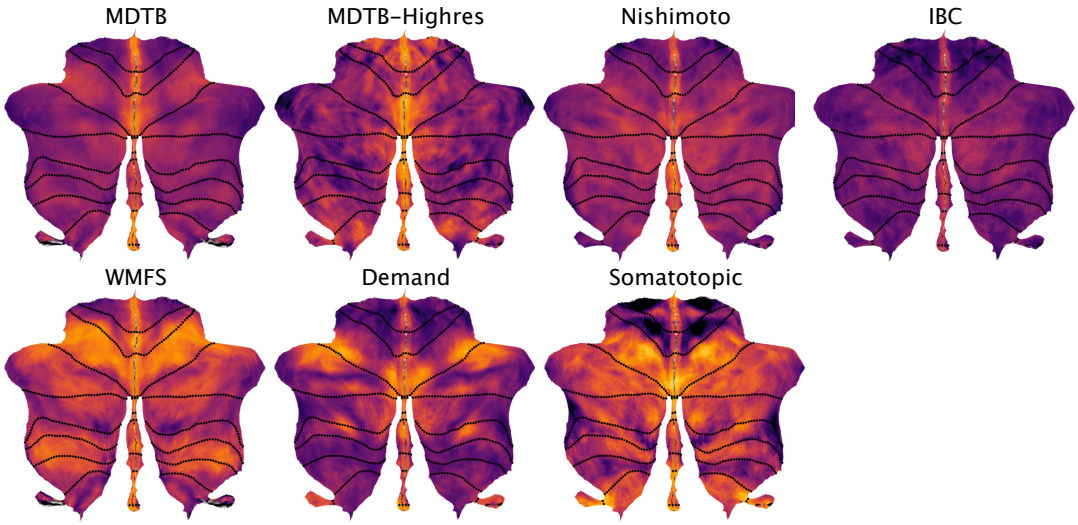

Figure S7. Individual functional symmetry. Functional symmetry across datasets. Functional symmetry calculated as the correlations of the functional responses of anatomically corresponding voxel of the left and right hemisphere. Functional symmetry was averaged across subjects within each dataset.

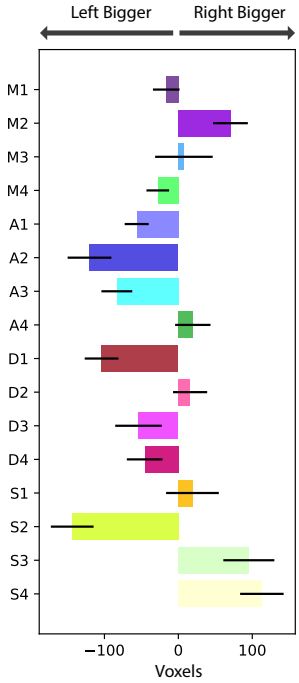

Figure S8. Size difference between left and right region pairs of the asymmetric atlas. Regions were estimated in individual subjects using the asymmetric atlas version. The size difference was calculated as number of voxels ($2mm^3$) in right parcel minus number of voxels in left parcel for each individual. Bars show average size difference across individuals and error bars indicate standard error of the mean.