bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Title: Linking lysosomal enzyme targeting genes and energy metabolism with altered gray matter volume in children with persistent stuttering

Abbreviated title: Genetics and anomalous neuroanatomy in stuttering

Authors: Ho Ming Chow[1,2,3]*, Emily O. Garnett[3], Hua Li[2], Andrew Etchell[3], Jorge Sepulcre[4,5], Dennis Drayna[6], Diane C. Chugani[1], Soo-Eun Chang[3,7,8]

Affiliations: [1] Communication Sciences and Disorders Program, University of Delaware, Newark, DE 19713; [2] Katzin Diagnostic & Research PET/MRI Center, Nemours/Alfred I. duPont Hospital for Children, Wilmington, DE 19803; [3] Department of Psychiatry, University of Michigan, Ann Arbor, MI 48109; [4] Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114; [5] Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; [6] National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, MD 20892; [7] Cognitive Imaging Research Center, Department of Radiology, Michigan State University, East Lansing, MI 48824; [8] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI 48824

Corresponding author: Ho Ming Chow, Communication Sciences and Disorders Program, University of Delaware, 100 Discovery Blvd, Newark, DE 19713. Tel: 302-831-7456. Email address: homingchow@gmail.com

Acknowledgements

The authors wish to thank all the children and parents who participated in this study. We also thank Kristin Hicks for her assistance in participant recruitment, behavioral testing, and help with MRI data collection, Scarlett Doyle for her assistance in MRI data acquisition, and Ashley Diener for her assistance in speech data analyses.

Declaration of conflicting interests: All other authors declare that they have no conflict of interest.

Funding: This work was supported by Award Numbers R21DC015853 (HMC), Z1A-000046 (DD), R01DC011277 (SC), and R21DC015312 (SC) from the National Institute on Deafness and Other Communication Disorders (NIDCD), and the Matthew K. Smith Stuttering Research Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDCD or the National Institutes of Health.

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Abstract

Developmental stuttering is a childhood onset neurodevelopmental disorder with an unclear etiology. Subtle changes in brain structure and function are present in both children and adults who stutter. It is a highly heritable disorder, and up to 12-20% of stuttering cases may carry a mutation in one of four genes involved in mannose-6-phosphate mediated protein intracellular trafficking. To better understand the relationship between genetic factors and brain structural changes, we used gene expression data from the Allen Institute for Brain Science (AIBS) and voxel-based morphometry (VBM) to investigate the spatial correspondence between gene expression patterns and differences in gray matter volume (GMV) between children with persistent stuttering (n=26, 87 scans) and their fluent peers (n=44, 139 scans). We found that expression patterns of two stuttering-related genes (*GNPTG* and *NAGPA*) in the brain exhibit a strong positive spatial correlation with the magnitude of GMV differences between groups. Further gene set enrichment analyses revealed that genes whose expression was highly correlated with the GMV differences were enriched for glycolysis and oxidative metabolism in mitochondria. Although the results are correlational and cannot inform us about underlying casual mechanisms, our results suggest a possibility that regions with high expression level of genes associated with stuttering may be particularly vulnerable to the effect of alterations in these genes. This effect may be further exacerbated by the relatively high energy utilization in those brain during the period of a sharp increase in brain energy utilization, which coincides with a period of rapid language development and the onset of stuttering during childhood.

Keywords: longitudinal, lysosome, metabolism, stuttering, VBM

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Introduction

Fluid, effortless speech production forms the basis for communication and is considered a fundamental human ability. Stuttering significantly disrupts fluent speech production, often leading to negative psychosocial and economic consequences throughout life (Blumgart, Tran, & Craig, 2010; Craig, Blumgart, & Tran, 2009; Yaruss, 2010). Developmental stuttering typically has an onset in early childhood, affecting more than 5% of preschool-age children and persisting in about 1% of adults (Craig, Hancock, Tran, Craig, & Peters, 2002; Månsson, 2000; Yairi & Ambrose, 2013). Persistent stuttering is highly heritable, with estimations of genetic contribution exceeding 80% in some studies (Dworzynski, Remington, Rijsdijk, Howell, & Plomin, 2007; Fagnani, Fibiger, Skytthe, & Hjelmborg, 2011; Ooki, 2005; Rautakoski, Hannus, Simberg, Sandnabba, & Santtila, 2012; van Beijsterveldt, Felsenfeld, & Boomsma, 2010). Genes causative of persistent stuttering have begun to be identified (C. Kang et al., 2010; Raza et al., 2015). To date, four such genes, designated *GNPTG*, *GNPTAB* and *NAGPA* and *AP4E1* have been found, and they may cumulatively account for up to 12-20% of unrelated individuals with persistent stuttering (see Frigerio-Domingues & Drayna, 2017 for a comprehensive review). This group of genes is known to play a role in lysosomal enzyme trafficking. *GNPTG*, *GNPTAB* and *NAGPA* are involved in marking lysosomal hydrolases and several nonlysosomal proteins with a mannose-6-phosphate (M6P) tag that is important for intracellular trafficking (Barnes et al., 2011). Homozygous mutations in *GNPTG* and *GNPTAB* genes are known to cause the rare inherited lysosomal storage disorders Mucolipidosis Types II and III (Kornfeld, 2001), which affect many parts of the body including the brain. However, in most of the cases, people who stutter only carry heterozygous mutations in these genes, and do not have the signs or symptoms typically seen in Mucolipidosis types II and III. *AP4E1* is a member of a family of adaptor proteins that are involved in vesicle formation and sorting member proteins for transporting lysosomal enzymes from the trans Golgi network to late endosomes and lysosomes. Mutations in *AP4E1* have been associated with hereditary spastic paraplegia and cerebral palsy (Abou Jamra et al., 2011; Kong et al., 2013). Why

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

mutations in these genes specifically affect the ability to produce fluent speech but not other cognitive or neurologic functions remains unknown. However, neuroimaging studies have shown that persistent stuttering is associated with subtle functional and anatomical anomalies (Chow & Chang, 2017; Garnett et al., 2018; Neef, Anwander, & Friederici, 2015). How genetic factors relate to brain anomalies is not yet clear. To pursue this question, we examined the differences in spatial patterns of gray matter volume (GMV) in children with persistent stuttering (pCWS) with the regional expression of the four genes thus far associated with stuttering using data provided by the Allen Institute for Brain Science (AIBS; http://www.brain-map.org/). This approach has been used to reveal gene-brain relationships in a number of recent studies. A seminal study was published in 2015, in which the authors used gene expression data and resting-state functional magnetic resonance imaging (fMRI) to identify 136 genes associated with intrinsic functional networks in the brain (Richiardi et al., 2015). In another study, Ortiz-Terán et al. used a similar method to demonstrate that the neural reorganization in blind children measured by resting-state functional connectivity is associated with a set of known neuroplasticity-related genes (Ortiz-Terán et al., 2017). Moreover, this approach was also employed to study neuropsychiatric disorders. For example, McColgan et al. identified genes associated with Huntington's disease by comparing regional white matter loss and gene expression in patients with the disorder (McColgan et al., 2017).

Although we do not know the mutation status of our participants with stuttering, high heritability of the disorder suggests that they are likely to carry a known or yet to be discovered gene mutation. While the genetic causes of stuttering are likely to be heterogeneous, it is probable that their effects at the neuroanatomical level could be similar because the disorder affects speech production specifically. Moreover, in support of the previous argument, neuroimaging studies of people who stutter with unknown genetic status have shown some consistent results. In particular, different groups of researchers have independently demonstrated that the anisotropic diffusivity in the corpus callosum and the superior longitudinal fasciculus are decreased in people who stutter compared with matched

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

controls (Neef et al., 2015). Therefore, a core presumption of this study is that the expression patterns of the known genes associated with stuttering, to a certain extent, reflect the magnitude of anatomical anomalies of the disorder. A similar presumption was made in a previous study showing that the expression of risk genes for schizophrenia were positively correlated with the anatomical disconnectivity defined by diffusion tensor imaging (DTI) tractography in patients with schizophrenia, but not in patients with bipolar disorder (Romme, de Reus, Ophoff, Kahn, & van den Heuvel, 2017).

In this study, we hypothesized that the expression of the four stuttering-related genes would be associated with the GMV differences in pCWS. Moreover, genes whose expression is highly associated with the GMV differences were used to explore the potential the biological processes and pathways involved in stuttering, using gene set enrichment analysis.

Materials and Methods

Participants. Participants were primarily recruited from the East Lansing, Michigan area and surrounding 50-mile radius, as part of an on-going longitudinal study conducted at Michigan State University (MSU). Recruitment activities included contacting and sending study flyers to area physicians' offices, childcare facilities, public schools, and outpatient speech clinics, and advertising in parent magazines and in social media. Relevant to the current investigation, a total of 226 children were contacted and screened. Of those, 128 were eligible and willing to participate in the study. Thirty-three of 128 did not participate in the MRI for variable reasons (e.g., uneasiness in the MRI setting and anxiety detected during mock scanner training), while 50 children who stutter (CWS; 20 girls and 30 boys) and 45 controls (23 girls and 22 boys) were scanned. Each subject participated in 1 to 4 longitudinal visits, with an inter-visit interval of approximately 12 months. The mean ages of CWS and controls at the first visit were 5.55 (SD=2.02) and 5.99 (SD=2.00) years, ranging between 3 to 10 years. Participants were monolingual English speakers. Children in the two groups did not differ in

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

chronological age, sex, handedness, or socioeconomic status (SES) based on mother's education level. All children exhibited normal speech and language development except for the presence of stuttering in the stuttering cohort as confirmed through a battery of standardized assessments, including Wechsler Preschool and Primary Scale of Intelligence Third Edition for children 2:6-7:3 (Wechsler, 2002), Wechsler Abbreviated Scale of Intelligence for children 7 and up (Wechsler, 1999), Peabody Picture Vocabulary Test (PPVT-3) for receptive vocabulary ability (Dunn & Dunn, 2007), Expressive Vocabulary Test (EVT-2; Williams, 2007) and Goldman-Fristoe Test of Articulation (GFTA-2; Goldman, 2000). The results of these tests are listed in Table 1. For study inclusion the participants had to score above -2 standard deviations of the norm on all standardized tests. None of the subjects had any concomitant developmental disorders (e.g., dyslexia, ADHD), with the exception of stuttering in CWS, and none were taking any medication affecting the central nervous system. CWS with assessments from 2 visits or more were further categorized as recovered or persistent based on their stuttering severity rating. The Stuttering Severity Instrument (SSI-4) was used to examine frequency and duration of disfluencies occurring in the speech sample, as well as any physical concomitants associated with stuttering. These were incorporated into a composite stuttering severity rating (Riley & Bakker, 2009). A child was considered recovered if the SSI-4 score was 10 or below at the second visit or thereafter. A child was categorized persistent if the composite SSI-4 score was higher than 10 at the second visit or thereafter. Both clinician and parent reports were required to be consistent with stuttering severity assessments in determining whether a child had recovered or was persistent. Three CWS were excluded because they were assessed only once and, therefore their final diagnoses could not be determined. Another CWS was excluded due to incidental findings in the structural scan. Nine scans from seven subjects were excluded due to excessive head movements. The final analysis included 87 scans from 26 children with persistent stuttering (pCWS; 8 girls and 18 boys; mean age at the first visit=6.5 years; SD=1.9), 61 scans from 17 children recovered from stuttering (rCWS; 8 girls and 9 boys; mean age at the first visit= 5.4 years; SD=1.9), and 139 scans from 44 controls (23 girls and 21 boys; mean age at the first visit=6.5 years; SD=2.0). Both persistent and recovered groups did

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

not differ from controls in chronological age, sex, handedness, or SES. Because *GNPTAG*, *NAGPA*, *GNPTG* and AP4E1 are associated with persistent stuttering, only scans collected from children with persistent stuttering and controls were included in the primary analysis. All research procedures were approved by the Michigan State University Institutional Review Board, which follows the ethical standards described in the Belmont Report and complies with the requirements of the Federalwide Assurance for the Protection of Human Subjects regulated by the United States Department of Health and Human Services. Written consents were obtained from all parents of the participating children, and assents were obtained from all children in verbal or written format depending on reading level. Children were paid a nominal remuneration, and were given small prizes (e.g., stickers) for their participation.

Voxel-based morphometry (VBM). Anatomical images were acquired on a GE 3T Signa scanner with an 8-channel head coil at Michigan State University. In each scan session, a whole brain 3D inversion recovery fast spoiled gradient-recalled T1-weighted images with CSF suppressed was obtained using the following parameters: time of echo = 3.8 ms, time of repetition of acquisition = 8.6 ms, time of inversion = 831 ms, repetition time of inversion = 2,332 ms, flip angle = 8◦, and receiver bandwidth = 620.8 kHz. For VBM analysis, we used the optimized procedure proposed by (Good et al., 2001). In summary, anatomical images were first segmented into different tissue partitions (J. Ashburner & Friston, 2005). Gray and white-matter images were nonlinearly registered to a MNI template using diffeomorphic image registration algorithm (DARTEL) (J. Ashburner, 2007). To accommodate for brain size differences, registrations were performed iteratively in a coarse-to-fine manner. Volumetric changes of each voxel were obtained by multiplying (or modulating) voxel values in the gray matter image by the deformation field derived from the registration procedure. Individual, modulated images were resampled to 1.5 mm isotropic voxels and spatially smoothed with a 6-mm FWHM kernel. To account for the dependence of participants' multiple scans in this study, GMV images were analyzed using Sandwich Estimator method, which was designed for analyzing longitudinal and repeated measures data (Guillaume, Hua, Thompson, Waldorp, & Nichols, 2014). The model included group

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

(pCWS and controls) and group by age interaction as well as quadratic age, sex, IQ, brain size, socioeconomic status and stuttering severity as covariates to control potential sources of variation. Although there was a significant difference between CWS and controls in IQ and both language measures, PPVT and EVT (Table 1), only IQ was included in the model because both measures were highly correlated with IQ (PPVT-IQ r=0.70, EVT-IQ r=0.69). The overall means of each covariate, except stuttering severity, were removed to capture the variation and potential differences between groups associated with the covariates. Since stuttering severity is only relevant to CWS, we considered stuttering severity was zero for controls and the mean for CWS was removed from the measure so that it would remove the variation associated with stuttering severity without affecting the group estimates. Voxel-wise t-statistics of the group difference were calculated. For comparing our VBM results with the findings in the literature, we also apply a threshold to visualize the significant GMV differences. However, it is not the primary goal of this study. Voxel-wise height threshold $p<0.005$ and cluster-size threshold $k>316$ voxels were used to control for false positives. This set of threshold corresponds to a corrected $p<0.05$. The cluster-size threshold was determined by AFNI 3dClustSim (version 17.2.13). Specifically, we first generated a non-gaussian noise model according to the spatial smoothness of the residual images using the AFNI 3dFWHMx autocorrelation function (-acf option). Then, we used Monte Carlo simulations implemented in AFNI 3dClustSim to estimate the false positive rate from the noise model (Cox, Chen, Glen, Reynolds, & Taylor, 2017).

Gray matter volume and gene expression correlation. Microarray-based gene expression data were obtained from the AIBS, which provides normalized expression of 29,131 genes using a total of 58,692 probes in each of 3,702 brain samples obtained from six adult donors (5 males, 1 female; age: 24-57 years; see http://www.brain-map.org/ for details) (Hawrylycz et al., 2012). We excluded genes whose symbols could not be identified in the database of HUGO Gene Nomenclature Committee, resulting in a total of 19,174 unique genes. The T1-weighted magnetic resonance images of the donors were segmented into different tissue partitions and normalized to the MNI template using the same

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

procedure used for analyzing the structural images acquired from our pediatric subjects. Using the deformation field generated by DARTEL, the locations of brain samples in native space were transformed into the MNI space. The samples' locations were mapped to 90 cortical and subcortical regions and the cerebellum based on a standard atlas with automated anatomical labeling (AAL Atlas) (Tzourio-Mazoyer et al., 2002). Because samples in the right hemispheres were taken from only two of the six donors, only supratentorial regions in the left hemisphere were included in our GMV-gene expression analysis. Since the right cerebellum has strong anatomical connections with the left cerebral hemisphere, and cerebellar anomalies have been associated with stuttering, the right cerebellum was included in our analyses as a single region. In total, 46 regions were included in the primary GMV-gene expression analysis. For each donor, expression of the same gene at each sample location from different probes was first averaged. Gene expression for each region was represented by the median of all the samples in the region. This step generated a parcellated expression map for each of the 19,174 genes. The GMV difference of each region was calculated by taking the mean of voxel-wise absolute *t*-statistics of between-group GMV differences (|*t*-stat|) within the region. Absolute GMV difference was used because the effect of genetic mutations on GMV is not known. Many neurological disorders such as ADHD and childhood onset schizophrenia are associated with the reduction of GMV (Gogtay et al., 2004; Nakao, Radua, Rubia, & Mataix-Cols, 2011). Like these neurological disorders, the effect of genetics in stuttering could directly impact the function of a brain region, leading to the reduction of GMV in stuttering. Second, the effect of genetics may also delay the cortical developmental trajectories of gray matter in children who stutter. The normal developmental trajectories of GMV differ across different brain areas, with most areas showing general decreases with age (Ducharme et al., 2015; Lange, 2012). Thus, delays in GMV decreases during development may appear as increased GMV when compared with age-matched controls. We do not know how the effect of genetics will lead to decreases or increases of GMV in stuttering, and how the mechanisms underlying the increases and decreases interplay in different brain regions. Moreover, neurological disorders with a strong genetic underpinning can be associated with increased or decreased GMV in different brain regions. For

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

example, autism spectrum disorder (ASD) is associated with a GMV decrease in the bilateral amygdala-hippocampus complex and an increase in the middle-inferior frontal gyrus (Via, Radua, Cardoner, Happé, & Mataix-Cols, 2011). Moreover, a small previous study on GMV in pCWS shows that stuttering is associated with GMV decreases in the inferior frontal gyrus and the putamen as well as increases in other speech motor areas (Beal, Gracco, Brettschneider, Kroll, & De Nil, 2013). Therefore, the effect of genetic variations on GMV could be in different directions, and we chose to use the absolute GMV differences. To minimize the potential adverse effect of outliers, Spearman's rank correlation was used to assess the relationship between gene expression and between-group GMV difference instead of Pearson's correlation as it was used in previous analyses of this kind (Ortiz-Terán et al., 2017; Richiardi et al., 2015). Spearman's correlation coefficients (ρ) between GMV differences and each of the 19,174 genes expressed across the 46 regions were calculated. This procedure established a distribution of correlation coefficients. Statistical threshold was set at $q<0.05$ (adjusted $p<0.05$), corrected for multiple testing by controlling the false discovery rate (FDR) (Benjamini & Hochberg, 1995).

Gene set enrichment analysis. Since genes other than *GNPTG* and *NAGPA* that are expressed in concordance with between-group GMV differences might also be associated with persistent stuttering, we carried out a gene set enrichment analysis to identify biological processes, molecular functions, cellular components or KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways for which the top 2.5% of the genes that were most positively correlated with GMV differences are enriched. The 19,174 genes were used as the input of the background set for the enrichment analyses. We used PANTHER (http://geneontology.org/) to identify enrichment for biological processes, molecular functions or cellular components and DAVID (david.ncifcrf.gov) for identifying enrichment for KEGG pathways (M. Ashburner et al., 2000; Huang, Sherman, & Lempicki, 2009;
The Gene Ontology Consortium, 2017). The redundancy of the resulting gene ontology terms were removed by using REViGO (Supek, Bošnjak, Škunca, & Šmuc, 2011). Fisher's Exact test was used to

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

determine statistical significance of enrichment factors. Statistical threshold was set at $q<0.05$, corrected for multiple testing by controlling the FDR (Benjamini & Hochberg, 1995). Although the regional expression of our targeted four genes were only positively correlated with the GMV differences, for exploratory purposes, we carried out the same enrichment analysis using the 2.5% of the genes that were most negatively correlated with GMV differences.

Results

We used 87 longitudinally-acquired structural scans from 26 pCWS and 140 scans from 44 controls using a well-established neuroimaging technique, voxel-based morphometry (VBM) (J. Ashburner, 2007; Good et al., 2001), to estimate voxel-wise GMV differences across the whole brain. Controlling for sex, age, quadratic age, cranial brain volume, IQ, socioeconomic status and stuttering severity, the longitudinal analysis model showed that in the pCWS group, GMV in the left somatosensory, left anterior prefrontal and the right motor areas was significantly larger than in the control group (Fig. 1A).

We examined the spatial correspondence between the expression of the *GNPTG, NAGPA, GNPTAB* and *AP4E1* genes and between-group differences in GMV across the 46 regions using the Spearman rank correlation (Fig. 1B). The correlation coefficients (ρ) associated with *GNPTG, NAGPA, GNPTAB* and *AP4E1* genes were 0.57 ($q<0.01$), 0.49 ($q<0.05$), -0.08 ($q=0.78$) and 0.08 ($q=0.78$), respectively. As illustrated in the frequency distribution of ρ of all 19,174 genes in our analysis (Fig. 1C), the ρ associated with *GNPTG* and *NAGPA* genes were significantly higher than the 97.5 percentile. Since there was a tendency for a between-group difference in sex ratio [$\chi^2(1, N=70)=3.09$, $p=0.079$], we conducted a follow up analysis including only male pCWS and controls to rule out the possibility that the observed gene-brain relationship was driven by potential sex differences. The results of the male-only analysis were very similar to the original results (i.e., *GNPTG:* $\rho=0.47$, $q<0.05$; *NAGPA:* $\rho=0.57$, $q<0.01$; *GNPTAB*: $\rho=0.03$, $q=0.93$ and *AP4E1* $\rho=0.01$, $q=0.98$), indicating that the results were not

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

driven by a higher proportion of female subjects in the control group. The scatter plots in Fig. 1D show the relationship between gene expression and between-group differences in GMV across regions in the original analysis. We observed that the expression level of the *GNPTG* and *NAGPA* genes in the cerebellum and some subcortical regions such as the pallidum deviated from the relationship seen for the other regions. Repeating the analysis excluding the cerebellum, basal ganglia regions and thalamus, we obtained similar results (i.e., *GNPTG*: ρ=0.57, q<0.05; *NAGPA*: ρ=0.42, q<0.10; *GNPTAB*: ρ=0.02, q=0.58 and *AP4E1*: ρ=0.08, q=0.50), although the correlation for *NAGPA* only showed a tendency toward significance (Fig. S1 in the Appendix). Fig. 1D further illustrates that, in our original analysis, the monotonic relationship between the expression of *GNPTG* and *NAGPA* and between-group differences in GMV in different regions. To further explore whether the relationship between gene expression and between-group differences in GMV was specific to persistent stuttering, we performed the same analysis on 17 children who had recovered from stuttering (rCWS). For rCWS, the relationship was not significant for any of the four genes (i.e., *GNPTG*: ρ=-0.08, q=0.56; *NAGPA*: ρ=0.20, q=0.38: *GNPTAB*: ρ=-0.24, q=0.34; *AP4E1*: ρ=0.15, q=0.45).

Gene set enrichment analysis of the top 2.5% of the genes whose expression was most positively correlated with the between-group differences in GMV (Table S1 in the Appendix) showed that this set of genes was highly enriched with genes involved in energy metabolism and mitochondrial functions (Table S2-S3 in the Appendix). Genes in several KEGG pathways were also over-represented in the gene set. Forty-nine out of 479 (10.2%) genes analyzed were involved in metabolic pathways, including over representation of a subset of genes involved in citrate cycle (has 00020) and oxidative phosphorylation (hsa00190) (Table S4 in the Appendix). Genes involved in oxidative phosphorylation were also linked to a number of neurological disorders, including Parkinson's disease, Alzheimer's disease, Huntington's disease and amyotrophic lateral sclerosis (Table S5 in the Appendix). Using the same method, the top 2.5% genes negatively correlated with GMV differences were significantly enriched for the DNA packaging complex and the nucleosome (cellular components) as well as three

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

KEGG pathways: Alcoholism (hsa05034), Systemic lupus erythematosus (hsa05322) and Viral carcinogenesis (has05203) (Table S6 and S7 in the Appendix).

Discussion

In our VBM analysis, we observed significant GMV increases in the left somatosensory areas, the left middle frontal gyrus and the right motor cortex in pCWS. Moreover, the magnitude of absolute GMV differences across brain regions in pCWS was positively correlated with the expression levels of two stuttering-related genes (*GNPTG* and *NAGPA*). This association suggests that mutations in these genes (and potentially related yet to be identified stuttering-related genes) are associated with the changes in GMV in pCWS and the manifestation of persistent stuttering.

To date, two VBM studies in children who stutter have been published in peer reviewed journals (Beal et al., 2013; Chang, Erickson, Ambrose, Hasegawa-Johnson, & Ludlow, 2008). However, the sample sizes of both studies are small (<12 pCWS) and relatively lenient statistic thresholds were used. Nevertheless, both studies showed that smaller GMV in the bilateral inferior frontal gyrus (IFG) is associated with pCWS. In the current study, decreased GMV in the IFG was observed only at an uncorrected threshold (Fig 1A). On the other hand, our data showed that pCWS were associated with significantly increased GMV in the left prefrontal and bilateral sensorimotor areas, which have been shown to have increased blood flow during speech production in people who stutter in previous PET studies (Braun et al., 1997; Fox et al., 1996). This finding is partially consistent with Beal et al. (2013), where pCWS exhibited increased GMV in right motor regions. From previous VBM studies in typically developing children, it is known that a majority of brain regions undergo GMV decreases during childhood, which reflects refinements of neural circuits via synaptic and dendritic pruning (Gennatas et al., 2017). The increased GMV in pCWS relative to controls may thus reflect a delay of development in those areas. Similarly, developmental delays in the structure of white matter tracts connecting speech-

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

motor areas have been observed in a previous diffusion tensor imaging study using the same group of subjects (Chow & Chang, 2017). However, the connection between white and gray matter anomalies is unclear and warrant further analysis.

While our results are correlational and do not provide direct evidence on the underlying casual mechanisms, here we discuss biological mechanisms that could explain to the observation of our results. Our analysis with regional gene expression showed that the magnitudes of GMV regional differences between pCWS and controls are correlated with the expression patterns of *GNPTG* and *NAGPA* in the left hemisphere and the right cerebellum. This finding supports our hypothesis that the altered GMV in pCWS is related to known genes associated with stuttering. *GNPTAB*, *GNPTG* and *NAGPA* are involved in the formation of the M6P tag that allows the binding of the lysosomal enzymes and other M6P-glycoproteins to the M6P receptor. While the main function of the M6P receptor is to transport the M6P-hydrolases from the trans Golgi network (TGN) to the lysosomes, M6P receptors also bind other M6P-proteins and non-M6P proteins such as the insulin growth factor 2 (IGF2), a hormone that regulates cell metabolism and growth at the cell surface (Barnes et al., 2011; Gary-Bobo, Nirdé, Jeanjean, Morère, & Garcia, 2007; Han, D'Ercole, & Lund, 1987). The cation-independent M6P receptor plays an important role in the regulation of IGF2 levels by mediating its internalization and degradation (Oka, Kawasaki, & Yamashina, 1985). While IGF2 binds at a different site on the receptor than M6P tagged proteins, M6P lysosomal enzymes have been shown to alter the binding of IGF2 to the receptor (De Leon, Terry, Asmerom, & Nissley, 1996; Kiess et al., 1989). Thus, altered binding of these enzymes due to mutations associated with stuttering might alter IGF2 mediated growth leading to altered GMV in regions where these enzymes are highly expressed. IGF2 has been linked to brain growth and differentiation as well as psychiatric and neurodegenerative disorders such as anxiety disorders and Parkinson's disease (Fernandez & Torres-Alemán, 2012; Matrone et al., 2016; Pardo et al., 2018). While speculative, the connection between stuttering-related genes and IGF2 would be an interesting future research direction.

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

The gene set enrichment analysis showed that genes expressed in concordance with the between-group differences in GMV were significantly enriched for metabolic processes in mitochondria (Table S2-4 in the appendix). Moreover, the gene set also enriched for the KEGG pathways of a number of neurological disorders, including Parkinson's disease, Alzheimer's disease, Huntington's disease and amyotrophic lateral sclerosis (Table S5 in the appendix). The genes involved in these disease pathways largely overlapped with genes involved in oxidative phosphorylation, suggesting that metabolic dysfunction similar to these diseases may play a role in stuttering. The positive correlation between GMV differences and expression of metabolic genes indicates that the regions exhibiting large GMV differences have higher energy metabolism rates (Goyal, Hawrylycz, Miller, Snyder, & Raichle, 2014). This result suggested that there may be a link between energy metabolism and the development of anomalous GMV in persistent stuttering. This link could be mediated through mutations in metabolic genes as candidates in persistent stuttering. However, it is also possible that the GMV differences between pCWS and controls were exacerbated in brain regions with relatively high energy consumption related to disturbance in lysosomal function (McKenna, Schuck, & Ferreira, 2018).

How might metabolism and gene mutations that affect lysosomal enzymes trafficking lead to the neurological anomalies associated with stuttering? Recent studies have shown that lysosomes and mitochondria interact physically and functionally, and these interactions play an important role in modulating metabolic functions of the two organelles (Todkar, Ilamathi, & Germain, 2017; Wong, Ysselstein, & Krainc, 2018). The known gene mutations related to stuttering are involved in the trafficking of lysosomal enzymes for the breakdown of macromolecules and organelles, including damaged mitochondria (Plotegher & Duchen, 2017). If the lysosomal function cannot keep pace with the sharp increase of energy metabolism, damaged mitochondria may accumulate, leading to increased oxidative stress, which may in turn negatively impact neurological development (de la Mata et al., 2016; Kiselyov, Yamaguchi, Lyons, & Muallem, 2010). The effect of this vulnerability may be amplified in brain regions with relatively high energy consumption in children between 2 to 5 years of

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

age, a period of time in which brain metabolism sharply increases (Chugani, Phelps, & Mazziotta, 1987; Goyal et al., 2014). This age range also coincides with the typical onset age of stuttering (Bloodstein & Ratner, 1995), as well as a period of rapid development of language and other cognitive functions. Consistent with this hypothesis, significant increases in mitochondrial fragmentation were observed in several lysosomal storage diseases, including Mucolipidosis Types II and III that are caused by homozygous mutations in the same genes linked to stuttering (*GNPTAB* and *GNPTG*) (Jennings et al., 2006). Although mutations in people who stutter are usually heterozygous and in different locations and forms (C. Kang et al., 2010) and do not lead to the detrimental symptoms of Mucolipidosis, a previous study showed that lysosomal enzyme activity is partially compromised by mutations associated with persistent stuttering in *NAGPA* (Lee, Kang, Drayna, & Kornfeld, 2011). However, it is unclear to what extent cellular function is affected by a partial deficiency of lysosomal enzymes, especially related to mitophagy and autophagy.

While our study showed a significant spatial relationship between expression patterns of *GNPTG*/*NAGPA* and GMV regional differences between pCWS and controls , a few methodological caveats should be acknowledged. First, although the sample size of the current study is one of the largest neuroimaging data sets in developmental stuttering and included repeated scans from each subject, it is small in the context of neuroanatomical studies examining correlates in complex heterogeneous traits. Small sample sizes could lead to finding spurious group differences (Fusar-Poli et al., 2014) as well as lower power that limit detection of subtle differences that might otherwise be possible with larger samples. Second, , although gene expression data from AIBS were obtained from six adult donors, there is a high degree of similarity in the regional expression among donors (Hawrylycz et al., 2012). Moreover, the microarray gene expression data provided by AIBS were normalized within and across donors' brains. The details of the current normalization method can be found in a technical paper from Allen Brain Atlas (http://help.brain-map.org/download/attachments/2818165/Normalization_WhitePaper.pdf). This normalization procedure

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

reduces the variability across donors due to technical biases and allows comparison of expression data across two or more brains. However, a certain degree of variability between donors is inevitable. To further explore the individual donor variability, we examined the spatial relationship between the GMV differences with the express of the four gene in each of the six donors. The correlation coefficients are presented in (Table S8 in the Appendix). In summary, we found that the correlation with *GNPTG* and *NAGPA* was positive for all six donors and at least one of them in each donor was larger than 0.30 (p<0.05) in four of the six donors, whereas the correlation with *AP4E1* and *GNPTAB* ranged from -0.23 to 0.20 (p>01). While variability of individual donors in gene expression is inevitable, the results of this individual donor analysis point to the same direction of our results using the gene expression aggregated from the six donors. For readers who are interested in further examining the individual donor variability, the regional expression patterns of the four targeted genes in each of the six donors are presented in Fig. S2 and S3 in the Appendix. To date, the expression data from AIBS is the only source of human gene expression patterns with high spatial resolution. As mentioned in the introduction, AIBS gene expression data and the approach of the current study have been used to reveal relationships between genes and intrinsic brain network, association between neuroplasticity-related genes and altered functional connectivity in blind children, and to confirm the relationship between known risk genes and their associated neurological disorders including Alzheimer's disease, Huntington's disease and schizophrenia (Grothe et al., 2018; McColgan et al., 2017; Ortiz-Terán et al., 2017; Richiardi et al., 2015; Romme et al., 2017). Third, we cannot completely rule out the possibility that expression levels and patterns of some genes in children and adults are different. If this is the case for the genes associated with stuttering, we would expect that the spatial correlation between gene expression in adults and the GMV patterns in children would be weak, whereas a strong relationship was observed in our study. Moreover, previous studies have suggested that the changes of gene expression occur predominately during prenatal and infant development and become relatively stable by around 6 years of age (H. J. Kang et al., 2011). Kang et al. (2011) estimated that only 9.1% of genes exhibit temporally differential expression in the first 20 years of life, and the portion of differentially

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

expressed genes should be even less in our participant's age group. Future studies using gene expression profiles in children should be pursued to refine our understanding of the relationship between brain anomalies and gene expression, when such dataset becomes available.

Conclusions

In conclusion, we showed that relative to controls, pCWS exhibited larger GMV in the left somatosensory, left anterior prefrontal and the right motor areas. The spatial pattern of this GMV difference was positively correlated with the expression of lysosomal targeting genes *GNPTG* and *NAGPA* as well as genes involved in energy metabolism. More research is warranted to further investigate possible roles of M6P mediated intracellular and extracellular trafficking as well as metabolic functions play a role in the development of brain structural anomalies associated with stuttering.

References

Abou Jamra, R., Philippe, O., Raas-Rothschild, A., Eck, S. H., Graf, E., Buchert, R., … Colleaux, L. (2011). Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *American Journal of Human Genetics*, *88*(6), 788–795. https://doi.org/10.1016/j.ajhg.2011.04.019

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). `. *Nature Genetics*, Vol. 25, pp. 25–29. https://doi.org/10.1038/75556

Barnes, J., Lim, J. M., Godard, A., Blanchard, F., Wells, L., & Steet, R. (2011). Extensive mannose

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

phosphorylation on leukemia inhibitory factor (LIF) controls its extracellular levels by multiple mechanisms. *Journal of Biological Chemistry*, *286*(28), 24855–24864. https://doi.org/10.1074/jbc.M111.221432

Beal, D. S., Gracco, V. L., Brettschneider, J., Kroll, R. M., & De Nil, L. F. (2013). A voxel-based morphometry (VBM) analysis of regional grey and white matter volume abnormalities within the speech production network of children who stutter. *Cortex*, *49*(8), 2151–2161. https://doi.org/10.1016/j.cortex.2012.08.013

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Bloodstein, O., & Ratner, N. (1995). A handbook on stuttering Singular. *San Diego, CA*.

Blumgart, E., Tran, Y., & Craig, A. (2010). An investigation into the personal financial costs associated with stuttering. *Journal of Fluency Disorders*, *35*(3), 203–215.

Braun, A. R., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J. M., … Ludlow, C. L. (1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering. *Brain*, *120*(5), 761–784. https://doi.org/10.1093/brain/120.5.761

Chang, S. E., Erickson, K. I., Ambrose, N. G., Hasegawa-Johnson, M. A., & Ludlow, C. L. (2008). Brain anatomy differences in childhood stuttering. *NeuroImage*, *39*(3), 1333–1344. https://doi.org/10.1016/j.neuroimage.2007.09.067

Chow, H. M., & Chang, S. E. (2017). White matter developmental trajectories associated with persistence and recovery of childhood stuttering. *Human Brain Mapping*. https://doi.org/10.1002/hbm.23590

Chugani, H. T., Phelps, M. E., & Mazziotta, J. C. (1987). Positron emission tomography study of human

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

brain functional development. *Annals of Neurology*, *22*(4), 487–497. https://doi.org/10.1002/ana.410220408

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). fMRI clustering and false-positive rates. *Proceedings of the National Academy of Sciences*, 201614961.

Craig, A., Blumgart, E., & Tran, Y. (2009). The impact of stuttering on the quality of life in adults who stutter. *Journal of Fluency Disorders*, *34*(2), 61–71.

Craig, A., Hancock, K., Tran, Y., Craig, M., & Peters, K. (2002). Epidemiology of stuttering in the community across the entire life span. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2002/088)

de la Mata, M., Cotán, D., Villanueva-Paz, M., de Lavera, I., Álvarez-Córdoba, M., Luzón-Hidalgo, R., … Oropesa-Ávila, M. (2016). Mitochondrial Dysfunction in Lysosomal Storage Disorders. *Diseases*, *4*(4), 31. https://doi.org/10.3390/diseases4040031

De Leon, D. D., Terry, C., Asmerom, Y., & Nissley, P. (1996). Insulin-Like Growth Factor II Modulates the Routing of Cathepsin D in MCF-7 Breast Cancer Cells. *Endocrinology*, *137*(5), 1851–1859. https://doi.org/10.1210/endo.137.5.8612524

Ducharme, S., Albaugh, M. D., Nguyen, T. V., Hudziak, J. J., Mateos-Pérez, J. M., Labbe, A., … O'Neill, J. (2015). Trajectories of cortical surface area and cortical volume maturation in normal brain development. *Data in Brief*. https://doi.org/10.1016/j.dib.2015.10.044

Dunn, L. M., & Dunn, D. M. (2007). PPVT-4□: Peabody picture vocabulary test. In *Peabody picture vocabulary test*.

Dworzynski, K., Remington, A., Rijsdijk, F., Howell, P., & Plomin, R. (2007). Genetic etiology in cases of recovered and persistent stuttering in an unselected, longitudinal sample of young twins. *American Journal of Speech-Language Pathology*, *16*(2), 161–178. https://doi.org/10.1044/1058-

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

0360(2007/021)

Fagnani, C., Fibiger, S., Skytthe, A., & Hjelmborg, J. V. B. (2011). Heritability and environmental effects for self-reported periods with stuttering: A twin study from Denmark. *Logopedics Phoniatrics Vocology*, Vol. 36, pp. 114–120. https://doi.org/10.3109/14015439.2010.534503

Fernandez, A. M., & Torres-Alemán, I. (2012). The many faces of insulin-like peptide signalling in the brain. *Nature Reviews Neuroscience*, Vol. 13, pp. 225–239. https://doi.org/10.1038/nrn3209

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., … Lancaster, J. L. (1996). A PET study of the neural systems of stuttering. *Nature*, *382*(6587), 158–162. https://doi.org/10.1038/382158a0

Frigerio-Domingues, C., & Drayna, D. (2017). Genetic contributions to stuttering: the current evidence. *Molecular Genetics and Genomic Medicine*, Vol. 5, pp. 95–102. https://doi.org/10.1002/mgg3.276

Fusar-Poli, P., Radua, J., Frascarelli, M., Mechelli, A., Borgwardt, S., Di Fabio, F., … David, S. P. (2014). Evidence of reporting biases in voxel-based morphometry (VBM) studies of psychiatric and neurological disorders. *Human Brain Mapping*. https://doi.org/10.1002/hbm.22384

Garnett, E. O., Chow, H. M., Nieto-Castañón, A., Tourville, J. A., Guenther, F. H., & Chang, S.-E. (2018). Anomalous morphology in left hemisphere motor and premotor cortex of children who stutter. *Brain*. https://doi.org/10.1093/brain/awy199

Gary-Bobo, M., Nirdé, P., Jeanjean, A., Morère, A., & Garcia, M. (2007). Mannose 6-phosphate receptor targeting and its applications in human diseases. *Current Medicinal Chemistry*, *14*(28), 2945–2953. https://doi.org/10.2174/092986707782794005

Gennatas, E. D., Avants, B. B., Wolf, D. H., Satterthwaite, T. D., Ruparel, K., Ciric, R., … Gur, R. C. (2017). Age-Related Effects and Sex Differences in Gray Matter Density, Volume, Mass, and Cortical Thickness from Childhood to Young Adulthood. *The Journal of Neuroscience*, *37*(20),

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

5065–5073. https://doi.org/10.1523/JNEUROSCI.3550-16.2017

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, a C., … Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8174–8179. https://doi.org/10.1073/pnas.0402680101

Goldman, R. (2000). Goldman Fristoe 2: Test of Articulation. In *Journal of Chemical Information and Modeling*. https://doi.org/10.1017/CBO9781107415324.004

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N., Friston, K. J., & Frackowiak, R. S. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, *14*(1 Pt 1), 21–36. https://doi.org/10.1006/nimg.2001.0786

Goyal, M. S., Hawrylycz, M., Miller, J. A., Snyder, A. Z., & Raichle, M. E. (2014). Aerobic glycolysis in the human brain is associated with development and neotenous gene expression. *Cell Metabolism*, *19*(1), 49–57. https://doi.org/10.1016/j.cmet.2013.11.020

Grothe, M. J., Sepulcre, J., Gonzalez-Escamilla, G., Jelistratova, I., Schöll, M., Hansson, O., & Teipel, S. J. (2018). Molecular properties underlying regional vulnerability to Alzheimer's disease pathology. *Brain*. https://doi.org/10.1093/brain/awy189

Guillaume, B., Hua, X., Thompson, P. M., Waldorp, L., & Nichols, T. E. (2014). Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. *NeuroImage*, *94*, 287–302. https://doi.org/10.1016/j.neuroimage.2014.03.029

Han, V. K. M., D'Ercole, A. J., & Lund, P. K. (1987). Cellular localizaton of somatomedin (insulin-like growth factor) messenger RNA in the human fetus. *Science*, *236*(4798), 193–197. https://doi.org/10.1126/science.3563497

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., … Jones, A.

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, *489*(7416), 391–399. https://doi.org/10.1038/nature11405

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. https://doi.org/10.1038/nprot.2008.211

Jennings, J. J., Zhu, J. H., Rbaibi, Y., Luo, X., Chu, C. T., & Kiselyov, K. (2006). Mitochondrial aberrations in mucolipidosis type IV. *Journal of Biological Chemistry*, *281*(51), 39041–39050. https://doi.org/10.1074/jbc.M607982200

Kang, C., Riazuddin, S., Mundorff, J., Krasnewich, D., Friedman, P., Mullikin, J. C., & Drayna, D. (2010). Mutations in the Lysosomal Enzyme–Targeting Pathway and Persistent Stuttering. *New England Journal of Medicine*, *362*(8), 677–685. https://doi.org/10.1056/NEJMoa0902630

Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., … Šestan, N. (2011). Spatio-temporal transcriptome of the human brain. *Nature*, *478*(7370), 483–489. https://doi.org/10.1038/nature10523

Kiess, W., Thomas, C. L., Greenstein, L. A., Lee, L., Sklar, M. M., Rechler, M. M., … Nissley, S. P. (1989). Insulin-like growth factor-II (IGF-II) inhibits both the cellular uptake of β-galactosidase and the binding of β-galactosidase to purified IGF-II/mannose 6-phosphate receptor. *Journal of Biological Chemistry*, *264*(8), 4710–4714.

Kiselyov, K., Yamaguchi, S., Lyons, C. W., & Muallem, S. (2010). Aberrant Ca2+ handling in lysosomal storage disorders. *Cell Calcium*, *47*(2), 103–111. https://doi.org/10.1016/j.ceca.2009.12.007

Kong, X. F., Bousfiha, A., Rouissi, A., Itan, Y., Abhyankar, A., Bryant, V., … Boisson-Dupuis, S. (2013). A Novel Homozygous p.R1105X Mutation of the AP4E1 Gene in Twins with Hereditary Spastic Paraplegia and Mycobacterial Disease. *PLoS ONE*, *8*(3).

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

https://doi.org/10.1371/journal.pone.0058286

Kornfeld, S. (2001). I-cell disease and pseudo-Hurler polydystrophy: disorders of lysosomal enzyme phosphorylation and localization. *The Metabolic and Molecular Bases of Inherited Disease*, 3469–3482.

Lange, N. (2012). Total and regional brain volumes in a population-based normative sample from 4 to 18 years: The NIH MRI study of normal brain development. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhr018

Lee, W. S., Kang, C., Drayna, D., & Kornfeld, S. (2011). Analysis of mannose 6-phosphate uncovering enzyme mutations associated with persistent Stuttering. *Journal of Biological Chemistry*, *286*(46), 39786–39793. https://doi.org/10.1074/jbc.M111.295899

Månsson, H. (2000). Childhood stuttering: Incidence and development. *Journal of Fluency Disorders*, *25*(1), 47–57.

Matrone, C., Dzamko, N., Madsen, P., Nyegaard, M., Pohlmann, R., Søndergaard, R. V., … Nielsen, M. S. (2016). Mannose 6-phosphate receptor is reduced in -synuclein overexpressing models of parkinsons disease. *PLoS ONE*, *11*(8). https://doi.org/10.1371/journal.pone.0160501

McColgan, P., Gregory, S., Seunarine, K. K., Razi, A., Papoutsi, M., Johnson, E., … Tabrizi, S. J. (2017). Brain Regions Showing White Matter Loss in Huntington's Disease Are Enriched for Synaptic and Metabolic Genes. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2017.10.019

McKenna, M. C., Schuck, P. F., & Ferreira, G. C. (2018). Fundamentals of CNS energy metabolism and alterations in lysosomal storage diseases. *Journal of Neurochemistry*. https://doi.org/10.1111/jnc.14577

Nakao, T., Radua, J., Rubia, K., & Mataix-Cols, D. (2011). Gray matter volume abnormalities in ADHD:

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Voxel-based meta-analysis exploring the effects of age and stimulant medication. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.2011.11020281

Neef, N. E., Anwander, A., & Friederici, A. D. (2015). The Neurobiological Grounding of Persistent Stuttering: from Structure to Function. *Current Neurology and Neuroscience Reports*, Vol. 15. https://doi.org/10.1007/s11910-015-0579-4

Oka, S., Kawasaki, T., & Yamashina, I. (1985). Isolation and characterization of mannan-binding proteins from chicken liver. *Archives of Biochemistry and Biophysics*, *241*(1), 95–105. https://doi.org/10.1016/0003-9861(85)90366-2

Ooki, S. (2005). Genetic and environmental influences on stuttering and tics in Japanese twin children. *Twin Research and Human Genetics*, *8*(1), 69–75. https://doi.org/10.1375/1832427053435409

Ortiz-Terán, L., Diez, I., Ortiz, T., Perez, D. L., Aragón, J. I., Costumero, V., … Sepulcre, J. (2017). Brain circuit–gene expression relationships and neuroplasticity of multisensory cortices in blind children. *Proceedings of the National Academy of Sciences*, 201619121. https://doi.org/10.1073/pnas.1619121114

Pardo, M., Cheng, Y., Sitbon, Y. H., Lowell, J. A., Grieco, S. F., Worthen, R. J., … Barreda-Diaz, A. (2018). Insulin growth factor 2 (IGF2) as an emergent target in psychiatric and neurological disorders. Review. *Neuroscience Research*. https://doi.org/10.1016/j.neures.2018.10.012

Plotegher, N., & Duchen, M. R. (2017). Mitochondrial Dysfunction and Neurodegeneration in Lysosomal Storage Disorders. *Trends in Molecular Medicine*, Vol. 23, pp. 116–134. https://doi.org/10.1016/j.molmed.2016.12.003

Rautakoski, P., Hannus, T., Simberg, S., Sandnabba, N., & Santtila, P. (2012). Genetic and environmental effects on stuttering: A twin study from Finland. *Journal of Fluency Disorders*, *37*(3), 202–210. https://doi.org/http://dx.doi.org/10.1016/j.jfludis.2011.12.003

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Raza, M. H., Mattera, R., Morell, R., Sainz, E., Rahn, R., Gutierrez, J., … Drayna, D. (2015). Association between rare variants in AP4E1, a component of intracellular trafficking, and persistent stuttering. *American Journal of Human Genetics*, *97*(5), 715–725. https://doi.org/10.1016/j.ajhg.2015.10.007

Richiardi, J., Altmann, A., Milazzo, A. C., Chang, C., Chakravarty, M. M., Banaschewski, T., … Greicius, M. D. (2015). Correlated gene expression supports synchronous activity in brain networks. *Science*, *348*(6240), 1241–1244. https://doi.org/10.1126/science.1255905

Riley, G., & Bakker, K. (2009). *SSI-4: Stuttering severity instrument*. Pro-Ed.

Romme, I. A. C., de Reus, M. A., Ophoff, R. A., Kahn, R. S., & van den Heuvel, M. P. (2017). Connectome Disconnectivity and Cortical Gene Expression in Patients With Schizophrenia. *Biological Psychiatry*, *81*(6), 495–502. https://doi.org/10.1016/j.biopsych.2016.07.012

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0021800

The Gene Ontology Consortium. (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Research*, *45*(D1), D331–D338. https://doi.org/10.1093/nar/gkw1108

Todkar, K., Ilamathi, H. S., & Germain, M. (2017). Mitochondria and Lysosomes: Discovering Bonds. *Frontiers in Cell and Developmental Biology*, *5*. https://doi.org/10.3389/fcell.2017.00106

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

van Beijsterveldt, C. E. M., Felsenfeld, S., & Boomsma, D. I. (2010). Bivariate Genetic Analyses of Stuttering and Nonfluency in a Large Sample of 5-Year-Old Twins. *Journal of Speech Language*

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

and Hearing Research, *53*(3), 609. https://doi.org/10.1044/1092-4388(2009/08-0202)

Via, E., Radua, J., Cardoner, N., Happé, F., & Mataix-Cols, D. (2011). Meta-analysis of gray matter abnormalities in autism spectrum disorder: Should asperger disorder be subsumed under a broader umbrella of autistic spectrum disorder? *Archives of General Psychiatry*. https://doi.org/10.1001/archgenpsychiatry.2011.27

Wechsler, D. (1999). Manual for the Wechsler abbreviated intelligence scale (WASI). *San Antonio, TX: The Psychological Corporation*.

Wechsler, D. (2002). Wechsler primary and preschool scale of intelligence. *San Antonio, TX: The Psychological Corporation*.

Williams, K. (2007). EVT-2: Expressive vocabulary test. *Pearson Assessments*.

Wong, Y. C., Ysselstein, D., & Krainc, D. (2018). Mitochondria–lysosome contacts regulate mitochondrial fission via RAB7 GTP hydrolysis. *Nature*, *554*(7692), 382–386. https://doi.org/10.1038/nature25486

Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*. https://doi.org/10.1016/j.jfludis.2012.11.002

Yaruss, J. S. (2010). Assessing quality of life in stuttering treatment outcomes research. *Journal of Fluency Disorders*. https://doi.org/10.1016/j.jfludis.2010.05.010

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Appendix

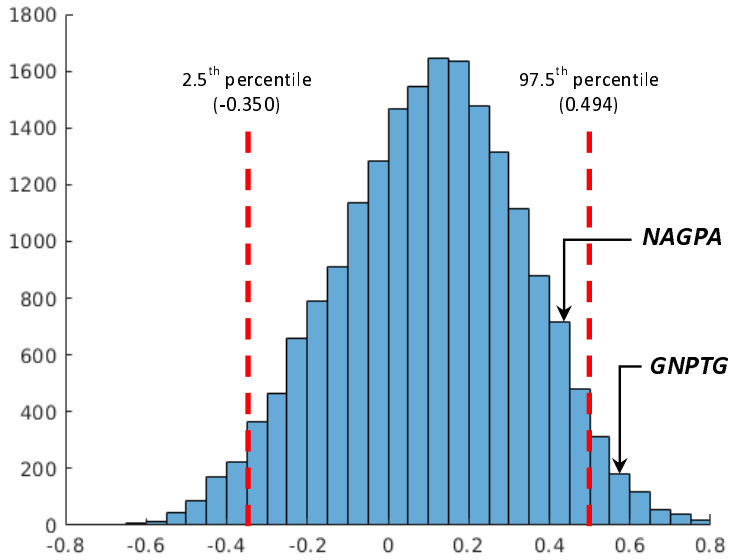


Fig. S1. Frequency plot of Spearman's correlation coefficients between GMV group differences and each of the 19,174 genes expressed across the regions when cerebellum, basal ganglia regions and thalamus were excluded. The red dash lines indicate the levels of correlation at the 2.5th and 97.5th percentiles. To obtain a correlation >0.494 or <-0.350 is less than 5% chance if a gene is randomly selected.

Table S1. Top 2.5% (479) genes positively correlated with regional GMV difference between pCWS and controls

#	Gene Name	*ρ*	*p*	*q*	#	Gene Name	*ρ*	*p*	*q*	#	Gene Name	*ρ*	*p*	*q*
1	ZNF513	0.71	0.000	0.000	161	SV2A	0.55	0.000	0.005	321	RNF208	0.51	0.000	0.009
2	TRPM4	0.69	0.000	0.000	162	KIAA1826	0.55	0.000	0.005	322	SORL1	0.51	0.000	0.009
3	MED12	0.69	0.000	0.000	163	ASAP1	0.55	0.000	0.005	323	ACO2	0.51	0.000	0.009
4	MYO5A	0.69	0.000	0.000	164	PLEKHA6	0.55	0.000	0.005	324	MFHAS1	0.51	0.000	0.009
5	SPHK2	0.68	0.000	0.000	165	SLC25A5	0.55	0.000	0.005	325	FLT3	0.51	0.000	0.010
6	NUDT22	0.68	0.000	0.000	166	TMEM14C	0.55	0.000	0.005	326	ATXN7	0.51	0.000	0.010
7	VPS8	0.67	0.000	0.000	167	ARRDC2	0.55	0.000	0.005	327	DLSTP	0.51	0.000	0.010
8	ZNF358	0.67	0.000	0.000	168	PITPNA	0.55	0.000	0.005	328	HACL1	0.51	0.000	0.010
9	ZBTB45	0.67	0.000	0.000	169	G6PD	0.55	0.000	0.005	329	NEFM	0.51	0.000	0.010
10	ANKRD9	0.66	0.000	0.001	170	KLF12	0.55	0.000	0.005	330	ESAM	0.51	0.000	0.010
11	PHLDB2	0.65	0.000	0.001	171	KCTD20	0.55	0.000	0.005	331	BHLHE40	0.51	0.000	0.010
12	KIAA0427	0.64	0.000	0.001	172	SAP30L	0.55	0.000	0.005	332	SC5DL	0.51	0.000	0.010
13	LONP1	0.64	0.000	0.001	173	UQCRC2	0.55	0.000	0.005	333	SCRT1	0.51	0.000	0.010
14	KIAA0430	0.64	0.000	0.001	174	PDK3	0.55	0.000	0.005	334	GIPC1	0.51	0.000	0.010
15	PEPD	0.64	0.000	0.001	175	COPS5	0.55	0.000	0.005	335	RANBP9	0.51	0.000	0.010
16	NDUFS1	0.64	0.000	0.001	176	OSBPL1A	0.55	0.000	0.005	336	ALDOC	0.51	0.000	0.010
17	KEAP1	0.64	0.000	0.001	177	PITPNM1	0.55	0.000	0.005	337	PDZD7	0.51	0.000	0.010
18	ENPP5	0.63	0.000	0.001	178	C12orf70	0.54	0.000	0.005	338	QARS	0.51	0.000	0.010

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

19	FAM108A5P	0.63	0.000	0.002	179	SLC7A5	0.54	0.000	0.005	339	CRBN	0.51	0.000	0.010
20	RERGL	0.63	0.000	0.002	180	ABCA9	0.54	0.000	0.006	340	HOOK3	0.51	0.000	0.010
21	ATF4	0.62	0.000	0.002	181	BBS7	0.54	0.000	0.006	341	SLC25A5P1	0.51	0.000	0.010
22	RIMKLA	0.62	0.000	0.002	182	PCCA	0.54	0.000	0.006	342	CLEC14A	0.51	0.000	0.010
23	ACSL6	0.62	0.000	0.002	183	CALML3	0.54	0.000	0.006	343	ST3GAL1	0.51	0.000	0.010
24	ZNF385D	0.62	0.000	0.002	184	FGD5	0.54	0.000	0.006	344	ZNF365	0.51	0.000	0.010
25	PTPRD	0.62	0.000	0.002	185	MREG	0.54	0.000	0.006	345	REPIN1	0.51	0.000	0.010
26	RNF220	0.62	0.000	0.002	186	OR3A2	0.54	0.000	0.006	346	NEDD9	0.51	0.000	0.010
27	AIFM3	0.61	0.000	0.002	187	MAGED2	0.54	0.000	0.006	347	TMEM191A	0.51	0.000	0.010
28	RCAN2	0.61	0.000	0.002	188	SURF2	0.54	0.000	0.006	348	CPEB3	0.51	0.000	0.010
29	GPR161	0.61	0.000	0.002	189	CIT	0.54	0.000	0.006	349	SEC62	0.51	0.000	0.010
30	CSRP2	0.61	0.000	0.002	190	KRT31	0.54	0.000	0.006	350	SLC6A8	0.51	0.000	0.010
31	SPARCL1	0.61	0.000	0.002	191	MAST4	0.54	0.000	0.006	351	RRM2B	0.51	0.000	0.010
32	AVEN	0.61	0.000	0.002	192	PCDHB10	0.54	0.000	0.006	352	GNL3	0.50	0.000	0.010
33	PPP2R5B	0.61	0.000	0.002	193	PPP3CC	0.54	0.000	0.006	353	RANBP2	0.50	0.000	0.010
34	AKAP6	0.61	0.000	0.002	194	ZMAT4	0.54	0.000	0.006	354	HAPLN4	0.50	0.000	0.010
35	ZNF629	0.61	0.000	0.002	195	INPPL1	0.54	0.000	0.006	355	PRKAB1	0.50	0.000	0.010
36	PPM1L	0.61	0.000	0.002	196	DEF8	0.54	0.000	0.006	356	GFRA2	0.50	0.000	0.010
37	C12orf10	0.61	0.000	0.002	197	RNF144B	0.54	0.000	0.006	357	MRPL12	0.50	0.000	0.010
38	VWC2	0.61	0.000	0.002	198	RPL13P5	0.54	0.000	0.006	358	MAP4K2	0.50	0.000	0.010
39	HDAC5	0.61	0.000	0.002	199	NDUFA12	0.54	0.000	0.006	359	CORO2A	0.50	0.000	0.010
40	ELOVL6	0.60	0.000	0.002	200	UVRAG	0.54	0.000	0.006	360	GOLGA6L5	0.50	0.000	0.010
41	NEFL	0.60	0.000	0.002	201	KITLG	0.54	0.000	0.006	361	MRPL18	0.50	0.000	0.010
42	CD99L2	0.60	0.000	0.002	202	NXPH3	0.54	0.000	0.006	362	SAMM50	0.50	0.000	0.010
43	RELL2	0.60	0.000	0.002	203	WDR7	0.54	0.000	0.006	363	MTDH	0.50	0.000	0.010
44	STK38L	0.60	0.000	0.002	204	MKRN2	0.54	0.000	0.006	364	CCM2	0.50	0.000	0.010
45	DENND4B	0.60	0.000	0.002	205	C12orf24	0.54	0.000	0.006	365	YIPF3	0.50	0.000	0.010
46	C6orf136	0.60	0.000	0.002	206	TMEM163	0.54	0.000	0.006	366	AP1B1	0.50	0.000	0.011
47	RGMA	0.60	0.000	0.002	207	HSPA12A	0.54	0.000	0.006	367	TTC39B	0.50	0.000	0.011
48	ATP13A2	0.60	0.000	0.002	208	NDUFA4	0.54	0.000	0.006	368	ULK3	0.50	0.000	0.011
49	PTH1R	0.60	0.000	0.002	209	ARL4C	0.54	0.000	0.006	369	RRM2	0.50	0.000	0.011
50	WAC	0.60	0.000	0.002	210	ETV6	0.54	0.000	0.006	370	SLC24A2	0.50	0.000	0.011
51	FBXW5	0.60	0.000	0.002	211	SLC43A2	0.54	0.000	0.006	371	KBTBD11	0.50	0.000	0.011
52	CHCHD1	0.60	0.000	0.002	212	SPHAR	0.54	0.000	0.006	372	FMN1	0.50	0.000	0.011
53	MAP2K2	0.60	0.000	0.002	213	GLT25D2	0.54	0.000	0.006	373	ARHGEF4	0.50	0.000	0.011
54	NDUFB9	0.60	0.000	0.002	214	RALB	0.54	0.000	0.006	374	KCNS3	0.50	0.000	0.011
55	POLDIP2	0.60	0.000	0.002	215	MRPL16	0.54	0.000	0.006	375	AGFG1	0.50	0.000	0.011
56	AHNAK2	0.60	0.000	0.002	216	HPS1	0.53	0.000	0.006	376	BCL3	0.50	0.000	0.011
57	SNX21	0.59	0.000	0.002	217	ZC3H7B	0.53	0.000	0.006	377	ENTPD4	0.50	0.000	0.011
58	PCSK1	0.59	0.000	0.002	218	OSBPL3	0.53	0.000	0.006	378	RCCD1	0.50	0.000	0.011
59	RBPMS2	0.59	0.000	0.002	219	RPH3A	0.53	0.000	0.006	379	ZFY	0.50	0.000	0.011
60	GMPS	0.59	0.000	0.002	220	WDR64	0.53	0.000	0.006	380	HDHD2	0.50	0.000	0.011
61	TAX1BP1	0.59	0.000	0.002	221	HECW1	0.53	0.000	0.006	381	NSMCE2	0.50	0.000	0.011
62	RAB11FIP5	0.59	0.000	0.002	222	NR3C1	0.53	0.000	0.006	382	IL7R	0.50	0.000	0.011
63	KIAA1671	0.59	0.000	0.003	223	DDHD2	0.53	0.000	0.006	383	MCF2L-AS1	0.50	0.000	0.011
64	MRPS36	0.59	0.000	0.003	224	RAB11FIP3	0.53	0.000	0.006	384	MGAT3	0.50	0.000	0.011

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

65	GLRX2	0.59	0.000	0.003	225	UQCRC1	0.53	0.000	0.006	385	GOLGA7B	0.50	0.000	0.011
66	BRSK1	0.59	0.000	0.003	226	MPP1	0.53	0.000	0.006	386	COL6A1	0.50	0.000	0.011
67	NANOS3	0.59	0.000	0.003	227	RUSC2	0.53	0.000	0.007	387	KCNQ5	0.50	0.000	0.011
68	UGP2	0.59	0.000	0.003	228	GAS2	0.53	0.000	0.007	388	SF4	0.50	0.000	0.011
69	PLAGL1	0.59	0.000	0.003	229	NDUFV2	0.53	0.000	0.007	389	SLC39A14	0.50	0.000	0.011
70	SRA1	0.59	0.000	0.003	230	HIST1H3A	0.53	0.000	0.007	390	HMBS	0.50	0.000	0.012
71	SLC7A8	0.59	0.000	0.003	231	PDE4A	0.53	0.000	0.007	391	ZNF839	0.50	0.000	0.012
72	LIMK1	0.58	0.000	0.003	232	TFCP2	0.53	0.000	0.007	392	TMEM184C	0.50	0.000	0.012
73	DPY19L1	0.58	0.000	0.003	233	EIF5A2	0.53	0.000	0.007	393	MGST2	0.50	0.000	0.012
74	SIPA1L2	0.58	0.000	0.003	234	FAM167B	0.53	0.000	0.007	394	DMXL1	0.50	0.000	0.012
75	SLC35E1	0.58	0.000	0.003	235	ZBTB7A	0.53	0.000	0.007	395	ACSF3	0.50	0.000	0.012
76	C16ORF52	0.58	0.000	0.003	236	COX15	0.53	0.000	0.007	396	CCDC104	0.50	0.000	0.012
77	ATP6V0A1	0.58	0.000	0.003	237	NAT8L	0.53	0.000	0.007	397	FECH	0.50	0.000	0.012
78	PPFIBP1	0.58	0.000	0.003	238	ADSS	0.53	0.000	0.007	398	TAF12	0.50	0.000	0.012
79	MRPL27	0.58	0.000	0.003	239	ARHGEF1	0.53	0.000	0.007	399	SUSD5	0.50	0.000	0.012
80	FBXO33	0.58	0.000	0.003	240	A26C1B	0.53	0.000	0.007	400	CIDEA	0.49	0.000	0.012
81	PPP1R28	0.58	0.000	0.003	241	RAGE	0.53	0.000	0.007	401	TSPAN9	0.49	0.000	0.012
82	NR1H2	0.58	0.000	0.003	242	SERGEF	0.53	0.000	0.007	402	ATP1B3	0.49	0.000	0.012
83	SATB1	0.58	0.000	0.003	243	MRPS21	0.53	0.000	0.007	403	MTRF1L	0.49	0.000	0.012
84	TTLL12	0.58	0.000	0.003	244	HERPUD1	0.53	0.000	0.007	404	IDI1	0.49	0.000	0.012
85	NOX4	0.58	0.000	0.003	245	KCNA3	0.53	0.000	0.007	405	C4orf22	0.49	0.000	0.012
86	SLC39A13	0.58	0.000	0.003	246	BTN3A3	0.53	0.000	0.007	406	DCLK1	0.49	0.000	0.012
87	ARID5B	0.58	0.000	0.003	247	EXTL2	0.53	0.000	0.007	407	DPP8	0.49	0.000	0.012
88	CEND1	0.58	0.000	0.003	248	ZDHHC3	0.53	0.000	0.007	408	KCNAB2	0.49	0.000	0.012
89	KCNIP3	0.58	0.000	0.003	249	RASSF5	0.53	0.000	0.007	409	ZFYVE9	0.49	0.000	0.012
90	PSMD12	0.58	0.000	0.003	250	RNF157	0.53	0.000	0.007	410	ABCF2	0.49	0.000	0.012
91	SNX24	0.58	0.000	0.003	251	CHURC1	0.53	0.000	0.007	411	ATG4B	0.49	0.000	0.012
92	GMPR2	0.58	0.000	0.003	252	CDADC1	0.53	0.000	0.007	412	IDH3G	0.49	0.000	0.012
93	RHOBTB1	0.58	0.000	0.003	253	DAZAP1	0.53	0.000	0.007	413	PHLDA3	0.49	0.000	0.012
94	SUCLA2	0.57	0.000	0.003	254	HISPPD1	0.53	0.000	0.007	414	CAMTA2	0.49	0.000	0.012
95	PCDHB9	0.57	0.000	0.003	255	FNDC4	0.53	0.000	0.007	415	DNTTIP1	0.49	0.000	0.012
96	MAP3K6	0.57	0.000	0.003	256	KCNC3	0.52	0.000	0.007	416	SLU7	0.49	0.000	0.012
97	ZNF415	0.57	0.000	0.003	257	PNKD	0.52	0.000	0.008	417	TESK1	0.49	0.000	0.012
98	STARD13	0.57	0.000	0.003	258	SDSL	0.52	0.000	0.008	418	CRTAC1	0.49	0.001	0.012
99	CAB39L	0.57	0.000	0.003	259	ASGR1	0.52	0.000	0.008	419	TRNP1	0.49	0.001	0.012
100	TOMM40	0.57	0.000	0.004	260	CYCS	0.52	0.000	0.008	420	ERG	0.49	0.001	0.012
101	GNPTG	0.57	0.000	0.004	261	CAMK2G	0.52	0.000	0.008	421	KRBA1	0.49	0.001	0.012
102	MRPL15	0.57	0.000	0.004	262	TRIM44	0.52	0.000	0.008	422	GSTT2	0.49	0.001	0.012
103	RHBDL3	0.57	0.000	0.004	263	ANXA6	0.52	0.000	0.008	423	FBXO41	0.49	0.001	0.012
104	RIMS3	0.57	0.000	0.004	264	POU6F1	0.52	0.000	0.008	424	SCRN3	0.49	0.001	0.012
105	MRPL47	0.57	0.000	0.004	265	CDT1	0.52	0.000	0.008	425	SLC25A36	0.49	0.001	0.013
106	PFKM	0.57	0.000	0.004	266	PTPRK	0.52	0.000	0.008	426	ETS1	0.49	0.001	0.013
107	MIR22HG	0.57	0.000	0.004	267	IER5L	0.52	0.000	0.008	427	OSBPL6	0.49	0.001	0.013
108	CDS1	0.57	0.000	0.004	268	EPHB6	0.52	0.000	0.008	428	TTLL1	0.49	0.001	0.013
109	GAPVD1	0.57	0.000	0.004	269	PLEKHM3	0.52	0.000	0.008	429	ARCN1	0.49	0.001	0.013
110	CTSA	0.57	0.000	0.004	270	BTBD11	0.52	0.000	0.008	430	MTIF2	0.49	0.001	0.013

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

111	NPM2	0.57	0.000	0.004	271	SESTD1	0.52	0.000	0.008	431	LSG1	0.49	0.001	0.013
112	C13orf27	0.56	0.000	0.004	272	ARHGAP25	0.52	0.000	0.008	432	NDUFB8	0.49	0.001	0.013
113	MTMR2	0.56	0.000	0.004	273	BCAS3	0.52	0.000	0.008	433	GTPBP1	0.49	0.001	0.013
114	SHD	0.56	0.000	0.004	274	FAM108A4	0.52	0.000	0.008	434	OXR1	0.49	0.001	0.013
115	KIAA0232	0.56	0.000	0.004	275	PLXDC1	0.52	0.000	0.008	435	ASNS	0.49	0.001	0.013
116	CLEC2L	0.56	0.000	0.004	276	IMPA1	0.52	0.000	0.008	436	C12orf29	0.49	0.001	0.013
117	PLEKHH3	0.56	0.000	0.004	277	TMEM183A	0.52	0.000	0.008	437	NEFH	0.49	0.001	0.013
118	RAN	0.56	0.000	0.004	278	CS	0.52	0.000	0.008	438	SEMA7A	0.49	0.001	0.013
119	GRK4	0.56	0.000	0.004	279	PKM2	0.52	0.000	0.008	439	MTR	0.49	0.001	0.013
120	EME1	0.56	0.000	0.004	280	PRPSAP1	0.52	0.000	0.008	440	SGTB	0.49	0.001	0.013
121	PPARGC1A	0.56	0.000	0.004	281	ANKH	0.52	0.000	0.008	441	TUBAP2	0.49	0.001	0.013
122	VPS11	0.56	0.000	0.004	282	GNGT2	0.52	0.000	0.008	442	CASP8AP2	0.49	0.001	0.013
123	C5orf13	0.56	0.000	0.004	283	VWA3A	0.52	0.000	0.008	443	PREP	0.49	0.001	0.013
124	ESRRA	0.56	0.000	0.004	284	GSK3B	0.52	0.000	0.008	444	JHDM1D	0.49	0.001	0.013
125	RHOBTB2	0.56	0.000	0.004	285	NFIX	0.52	0.000	0.008	445	BCL2L2	0.49	0.001	0.013
126	SCN1A	0.56	0.000	0.004	286	ITGA11	0.52	0.000	0.008	446	CTXN3	0.49	0.001	0.013
127	AGSK1	0.56	0.000	0.005	287	PCDHGA5	0.52	0.000	0.008	447	CYTSA	0.49	0.001	0.013
128	RNF168	0.56	0.000	0.005	288	ADAMTS5	0.52	0.000	0.009	448	GSTK1	0.49	0.001	0.013
129	TGS1	0.56	0.000	0.005	289	C1orf174	0.52	0.000	0.009	449	COL5A2	0.49	0.001	0.013
130	ZNF641	0.56	0.000	0.005	290	FNDC5	0.52	0.000	0.009	450	NAGPA	0.49	0.001	0.013
131	CHAF1A	0.56	0.000	0.005	291	RTN2	0.52	0.000	0.009	451	PCDHB1	0.49	0.001	0.013
132	PDCL3	0.56	0.000	0.005	292	PCDH9	0.52	0.000	0.009	452	MNS1	0.49	0.001	0.013
133	EPN3	0.56	0.000	0.005	293	ANKS1A	0.52	0.000	0.009	453	BEND5	0.49	0.001	0.013
134	IQSEC1	0.56	0.000	0.005	294	RNF26	0.52	0.000	0.009	454	SYNGR1	0.49	0.001	0.013
135	HEXIM1	0.56	0.000	0.005	295	MGLL	0.52	0.000	0.009	455	CKMT1A	0.49	0.001	0.013
136	RHOB	0.56	0.000	0.005	296	FAM108A4P	0.52	0.000	0.009	456	C1orf201	0.49	0.001	0.013
137	OSBP	0.55	0.000	0.005	297	ZNF804B	0.52	0.000	0.009	457	FAM119A	0.49	0.001	0.013
138	ANKRD37	0.55	0.000	0.005	298	CLEC16A	0.52	0.000	0.009	458	SUPT5H	0.49	0.001	0.013
139	NAT8	0.55	0.000	0.005	299	STRBP	0.51	0.000	0.009	459	C19orf47	0.49	0.001	0.013
140	SNRK	0.55	0.000	0.005	300	STS	0.51	0.000	0.009	460	TMLHE	0.49	0.001	0.013
141	LAPTM4B	0.55	0.000	0.005	301	SDHB	0.51	0.000	0.009	461	PDXK	0.49	0.001	0.013
142	GPER	0.55	0.000	0.005	302	HES6	0.51	0.000	0.009	462	AHSA1	0.49	0.001	0.013
143	JDP2	0.55	0.000	0.005	303	CCDC25	0.51	0.000	0.009	463	FGF1	0.49	0.001	0.013
144	STK39	0.55	0.000	0.005	304	SDCCAG3P2	0.51	0.000	0.009	464	PPAP2A	0.49	0.001	0.013
145	SPATA2	0.55	0.000	0.005	305	USP31	0.51	0.000	0.009	465	WDR45	0.49	0.001	0.013
146	TRPM2	0.55	0.000	0.005	306	SEC23A	0.51	0.000	0.009	466	MATK	0.49	0.001	0.013
147	ACSL3	0.55	0.000	0.005	307	PLS1	0.51	0.000	0.009	467	ATP1A4	0.49	0.001	0.013
148	MAP1A	0.55	0.000	0.005	308	C15orf59	0.51	0.000	0.009	468	TFRC	0.49	0.001	0.013
149	EEF2	0.55	0.000	0.005	309	FAM20A	0.51	0.000	0.009	469	ENO3	0.49	0.001	0.013
150	ARHGEF11	0.55	0.000	0.005	310	LUZP1	0.51	0.000	0.009	470	FAXC	0.49	0.001	0.013
151	OIP5-AS1	0.55	0.000	0.005	311	EIF2AK1	0.51	0.000	0.009	471	ITGA1	0.49	0.001	0.013
152	HECA	0.55	0.000	0.005	312	RABGGTA	0.51	0.000	0.009	472	BAT3	0.48	0.001	0.013
153	CIRH1A	0.55	0.000	0.005	313	CORO6	0.51	0.000	0.009	473	CHGA	0.48	0.001	0.013
154	UBE2D3	0.55	0.000	0.005	314	ITSN1	0.51	0.000	0.009	474	NDUFA9	0.48	0.001	0.013
155	RAB37	0.55	0.000	0.005	315	ZYX	0.51	0.000	0.009	475	REEP2	0.48	0.001	0.013
156	EML2	0.55	0.000	0.005	316	PCDHB14	0.51	0.000	0.009	476	RFX5	0.48	0.001	0.013

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

157	LRRC66	0.55	0.000	0.005	317	STAC2	0.51	0.000	0.009	477	TFB1M	0.48	0.001	0.013
158	VWA2	0.55	0.000	0.005	318	TMX2	0.51	0.000	0.009	478	C17orf75	0.48	0.001	0.013
159	MFN2	0.55	0.000	0.005	319	ECM1	0.51	0.000	0.009	479	PEX5	0.48	0.001	0.013
160	C6orf106	0.55	0.000	0.005	320	FAM81A	0.51	0.000	0.009					

Table S2. Gene ontology (GO) terms for biological processes associated with in the top 2.5% of the genes whose expression was positively correlated with the between group gray matter volume differences.

GO Term [a]	GO biological process	*p* Value	FDR *q* Value	Enrichment [b]	B	b
GO:0045333	cellular respiration	1.02E-06	1.62E-02	4.43	147	17
GO:0006414	translational elongation	8.37E-06	2.21E-02	4.46	120	14
GO:0006091	generation of precursor metabolites and energy	2.31E-05	3.05E-02	2.57	387	26
GO:0032543	mitochondrial translation	1.69E-05	2.67E-02	4.48	111	13
GO:0006415	translational termination	2.17E-05	3.11E-02	4.73	97	12

[a] GO terms are arranged hierarchically with the most specific subclass first. GO terms associated with more than 1,000 genes are excluded.
[b] Enrichment = [Number of genes associated with a specific GO term in the target list (b) / Total number of genes in the target list] / [Total number of genes associated with a specific GO term (B) / Total number of background genes].

Table S3. Gene ontology (GO) terms for cellular components associated with the top 2.5% of the genes whose expression was positively correlated with the between group gray matter volume differences.

GO Term [a]	GO cellular component	p Value	FDR q Value	Enrichment [b]	B	b
GO:0070469	respirasome	4.16E-06	8.42E-04	5.67	81	12
GO:1990204	oxidoreductase complex	4.50E-06	8.28E-04	5.13	97	13
GO:0000313	organellar ribosome	4.36E-05	5.19E-03	4.78	88	11
GO:0044429	mitochondrial part	9.69E-09	1.96E-05	2.30	966	58
GO:0030964	NADH dehydrogenase complex	1.98E-04	1.82E-02	6.53	41	7
GO:0005746	mitochondrial respirasome	9.04E-06	1.41E-03	5.77	73	11
GO:0005761	mitochondrial ribosome	4.36E-05	4.90E-03	4.78	88	11

[a] GO terms are arranged hierarchically with the most specific subclass first. GO terms associated with more than 1,000 genes are excluded.
[b] Enrichment = [Number of genes associated with a specific GO term in the target list (b) / Total number of genes in the target list] / [Total number of genes associated with a specific GO term (B) / Total number of background genes].

Table S4. KEGG Pathways associated with the top 2.5% of the genes whose expression was positively correlated with the between group gray matter volume differences.

KEGG ID	KEGG Pathway	p Value	FDR q Value	Enrichment [a]	B	b
hsa01200	Carbon metabolism	4.10E-05	1.72E-03	4.3	112	13
hsa05012	Parkinson's disease	5.30E-04	3.82E-03	3.5	126	12
hsa04932	Non-alcoholic fatty liver disease (NAFLD)	1.40E-04	5.03E-03	3.6	146	14

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

hsa05010	Alzheimer's disease	3.60E-04	3.82E-03	3.2	161	14
hsa01230	Biosynthesis of amino acids	5.20E-04	3.55E-03	4.8	70	9
hsa00190	Oxidative phosphorylation	3.20E-04	4.68E-03	3.7	119	12
hsa00020	Citrate cycle (TCA cycle)	1.10E-03	3.82E-03	7.4	30	6
hsa05016	Huntington's disease	3.60E-03	2.48E-02	2.6	183	13
hsa00480	Glutathione metabolism	1.00E-02	3.65E-02	4.4	50	6
hsa05014	Amyotrophic lateral sclerosis (ALS)	1.00E-02	3.34E-02	4.4	50	6

[a] Enrichment = [Number of genes associated with a specific KEGG pathway in the target gene set (b) / Total number of genes in the target set] / [Total number of genes associated with a specific KEGG pathway (B) / Total number of background genes].
KEGG; Kyoto Encyclopedia of Genes and Genomes

Table S5. Genes associated with KEGG pathways and between group gray matter volume differences. The statistics of the enrichment analysis that identified the KEGG pathways is listed in Table S4.

Gene name	Non-alcoholic fatty liver disease (NAFLD)	Alzheimer's disease	Oxidative phosphorylation	Carbon metabolism	Parkinson's disease	Biosynthesis of amino acids	Huntington's disease	Amyotrophic lateral sclerosis (ALS)	Glutathione metabolism	Citrate cycle (TCA cycle)
NEFL								X		
NEFH								X		
NEFM								X		
TOMM40								X		
PPP3CC		X						X		
GSK3B	X	X								
ATF4	X									
PRKAB1	X									
CALML3		X								
NDUFA4	X	X	X		X		X			
NDUFA12	X	X	X		X		X			
NDUFA9	X	X	X		X		X			
NDUFB8	X	X	X		X		X			
NDUFB9	X	X	X		X		X			
NDUFS1	X	X	X		X		X			
NDUFV2	X	X	X		X		X			
UQCRC1	X	X	X		X		X			
UQCRC2	X	X	X		X		X			
SDHB	X	X	X	X	X		X			X

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

CYCS	X	X			X		X	X		
SLC25A5					X		X			
PPARGC1A							X			
COX15			X							
ATP6V0A1			X							
MGST2									X	
GSTT2									X	
GSTK1									X	
RRM2B									X	
RRM2									X	
G6PD				X					X	
PCCA				X						
SUCLA2				X						X
CS				X		X				X
ACO2				X		X				X
IDH3G				X		X				X
PKM				X		X				X
ENO3				X		X				
ALDOC				X		X				
SDSL				X		X				
PFKM				X		X				
DLST				X						
MTR						X				

KEGG; Kyoto Encyclopedia of Genes and Genomes

Table S6. Gene ontology (GO) terms for cellular components associated with the top 2.5% of the genes whose expression was negatively correlated with the between group gray matter volume differences.

GO Term [a]	GO cellular component	p Value	FDR q Value	Enrichment [b]	B	b
GO:0044815	DNA packaging complex	2.48E-07	5.02E-04	5.67	75	13
GO:0000786	nucleosome	3.35E-06	3.39E-03	4.78	67	11

[a] GO terms are arranged hierarchically with the most specific subclass first. GO terms associated with more than 1,000 genes are excluded.
[b] Enrichment = [Number of genes associated with a specific GO term in the target list (b) / Total number of genes in the target list (479)] / [Total number of genes associated with a specific GO term (B) / Total number of background genes (19174)].

Table S7. KEGG Pathways associated with the top 2.5% of the genes whose expression was negatively correlated with the between group gray matter volume differences.

KEGG ID	KEGG Pathway	p Value	FDR q Value	Enrichment [a]	B	b
hsa05034	Alcoholism	4.30E-07	9.96E-06	4.7	170	17
hsa05322	Systemic lupus erythematosus	1.10E-05	1.26E-04	4.9	126	13

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

hsa05203	Viral carcinogenesis	3.10E-03	4.00E-02	2.8	200	12

[a] Enrichment = [Number of genes associated with a specific KEGG pathway in the target gene set (b) / Total number of genes in the target set] / [Total number of genes associated with a specific KEGG pathway (B) / Total number of background genes].
KEGG; Kyoto Encyclopedia of Genes and Genomes

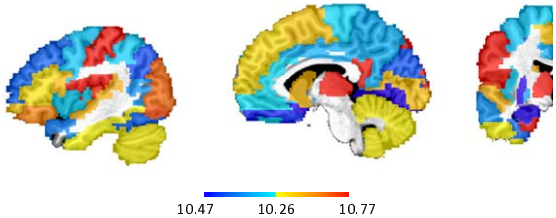
Table S8. Spearman's correlation between regional GMV differences and the expression of the four targeted genes in the six donors.

Donor ID	*GNPTG*	*NAGPA*	*AP4E1*	*GNPTAB*
9861	0.00	0.12	0.20	-0.23
10021	0.36	0.32	0.03	0.03
12876	0.18	0.23	-0.01	-0.02
14380	0.48	0.37	0.02	-0.18
15496	0.02	0.33	-0.08	0.03
15697	0.15	0.32	-0.02	0.11

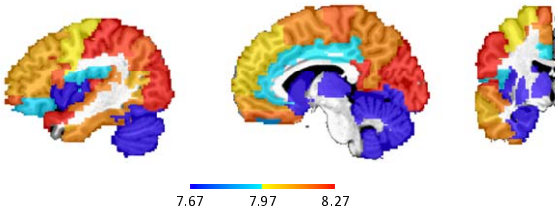
bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Fig. S2. Expression of the *GNPTG* and *NAGPA* in each of the six donors

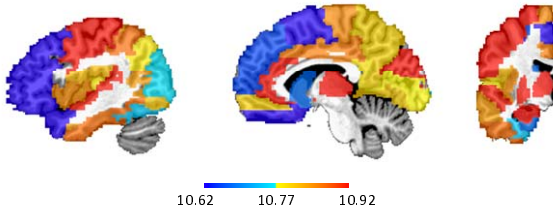
Donor 10021 *GNPTG* expression



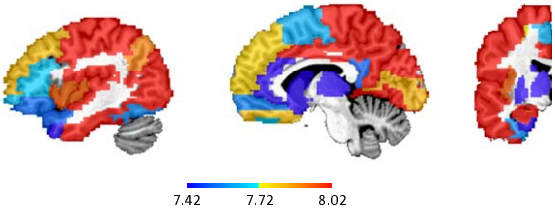
Donor 10021 *NAGPA* expression



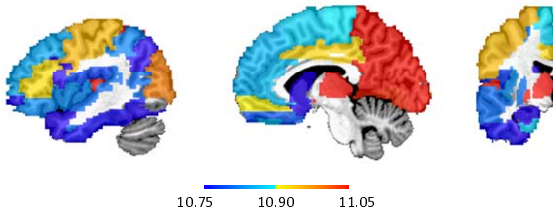
Donor 12876 *GNPTG* expression



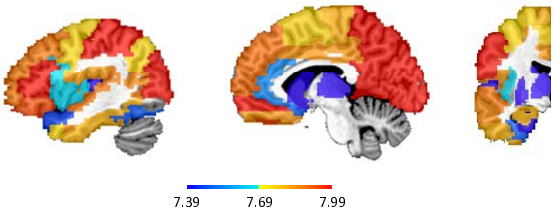
Donor 12876 *NAGPA* expression



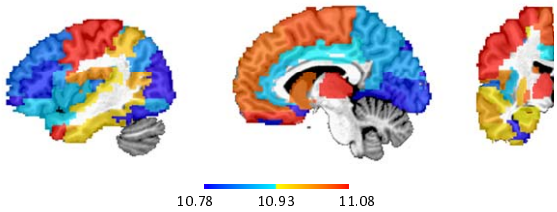
Donor 14380 *GNPTG* expression



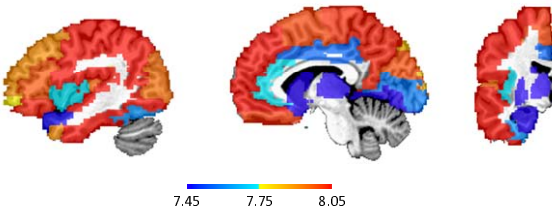
Donor 14380 *NAGPA* expression



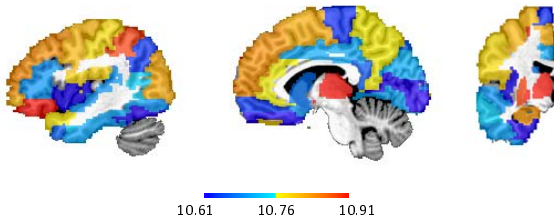
Donor 15496 *GNPTG* expression



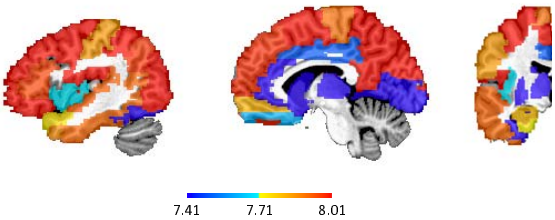
Donor 15496 *NAGPA* expression



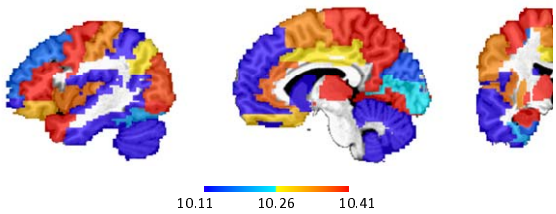
Donor 15697 *GNPTG* expression



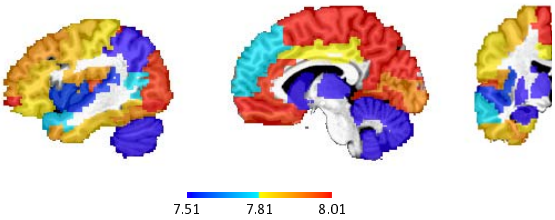
Donor 15697 *NAGPA* expression



Donor 9861 *GNPTG* expression



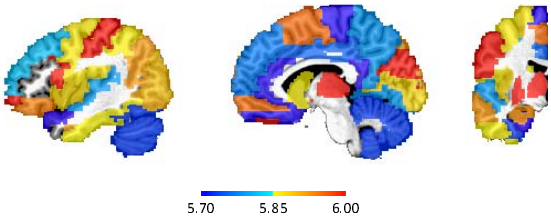
Donor 9861 *NAGPA* expression



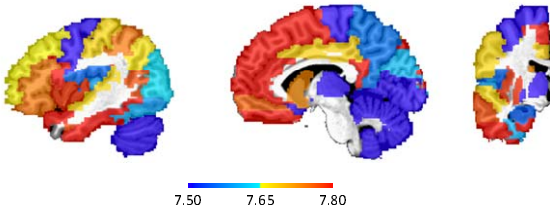
bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Fig. S3. Expression of the *AP4E1* and *GNPTAB* in each of the six donors

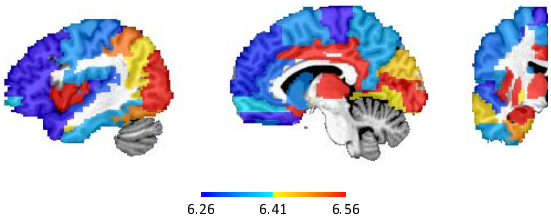
Donor 10021 *AP4E1* expression



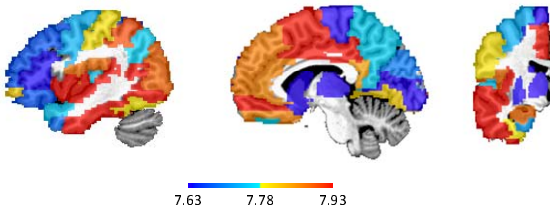
Donor 10021 *GNPTAB* expression



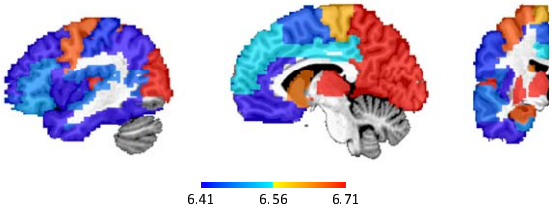
Donor 12876 *AP4E1* expression



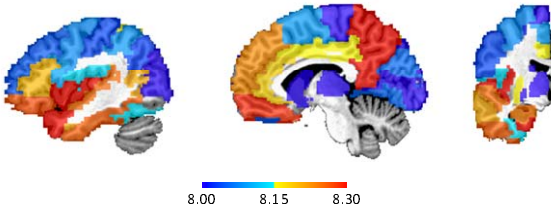
Donor 12876 *GNPTAB* expression



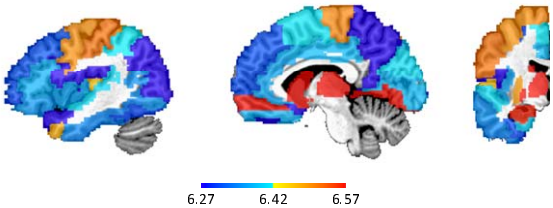
Donor 14380 *AP4E1* expression



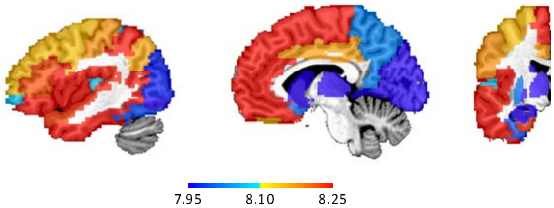
Donor 14380 *GNPTAB* expression



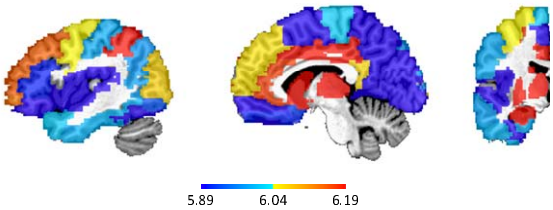
Donor 15496 *AP4E1* expression



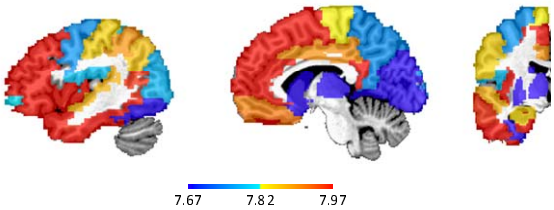
Donor 15496 *GNPTAB* expression



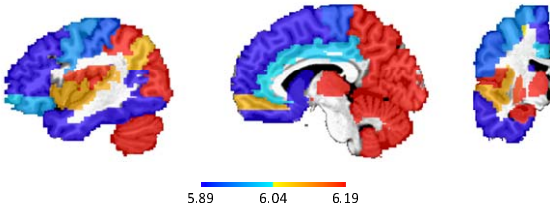
Donor 15697 *AP4E1* expression



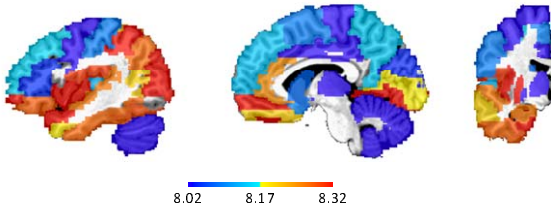
Donor 15697 *GNPTAB* expression



Donor 9861 *AP4E1* expression



Donor 9861 *GNPTAB* expression



bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Titles and legends to figures

Figure 1. Spatial relationship between gene expression and between-groups differences in gray matter volume (GMV). (A) Voxel-wise differences between children with persistent stuttering (pCWS) and controls in GMV. Color-coded *t* values of group differences are overlaid on an anatomical image. Areas exhibited a significant between-group difference at multiple comparisons corrected $p<0.05$ are outlined by black lines. The other colored areas are subthreshold (uncorrected $p<0.1$). (B) Parcellated gene expression of *GNPTG* and *NAGPA* and absolute GMV differences in *t*-statistics (|*t*-stat|) in 45 left hemispheric regions and the right cerebellum (which anatomically connects to the left hemisphere) were overlaid on a single-subject anatomical image. The parcellation of the brain was based on a standard atlas with automated anatomical labeling (AAL). Gene expression and |*t*-stat| of the right cerebellum were displayed in the left cerebellum to save space. (C) Frequency plot of Spearman's correlation coefficients between GMV group differences and each of the 19,174 genes expressed across the regions. The red dash lines indicate the levels of correlation at the 2.5th and 97.5th percentiles. To obtain a correlation >0.474 or <-0.360 is less than 5% chance if a gene is randomly selected. (D) Scatter plots between gene expression and group differences in GMV in the sensorimotor areas (red dots), the parietal lobe (purple dots), the cingulate cortex (orange dots), the middle frontal gyrus (green dots) and the rest of the regions (blue dots). Regions in which gene expression is 2.5 standard deviations above or below mean were labelled.

Table 1. Demographics, intelligent quotient (IQ) and language tests scores averaged across longitudinal visits for each participant

	Controls, n = 44 (21 boys)		Persistent, n = 26 (18 boys)		Recovered, n = 17 (9 boys)	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Age at the first scan (years)	6.5 (2.0)	3.3 – 10.8	6.5 (1.9)	3.6 – 10.3	5.4 (1.9)	3.1 – 9.4
IQ [a]	114 (14.1)	84 – 144	106 (15.5)	81 – 138	106 (13.1)	88 – 128
PPVT [b]	119 (12.7)	95 – 141	110 (13.5)	86 – 146	114 (10.3)	93 –131
EVT [c]	115 (11.8)	93 – 142	106 (12.2)	86 – 138	109 (9.2)	89 – 129
GFTA [d]	104 (6.6)	84 – 115	102 (4.2)	92 – 110	106 (7.3)	91 – 115
SSI [e]	-	-	21 (8.3)	12 – 48	13 (2.9)	7 – 19

[a] intelligence quotient (IQ) No significant difference between any two groups (t-tests, $p>0.05$)

[b] The Peabody Picture Vocabulary Test (PPVT). Scores significantly higher in control than persistent groups (two-sample t-tests, $p<0.05$).

[c] The Expressive Vocabulary Test (EVT). Scores significantly higher in control than both persistent and recovered groups (two-sample *t*-tests, $p<0.05$).

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

[d] The Goldman-Fristoe Test of Articulation (GFTA). Scores significantly higher in control than persistent groups (two-sample t-tests, $p<0.05$).

[e] Scores significantly higher in persistent and recovered groups (two-sample t-tests, $p<0.05$).

bioRxiv preprint doi: https://doi.org/10.1101/848796. this version posted November 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

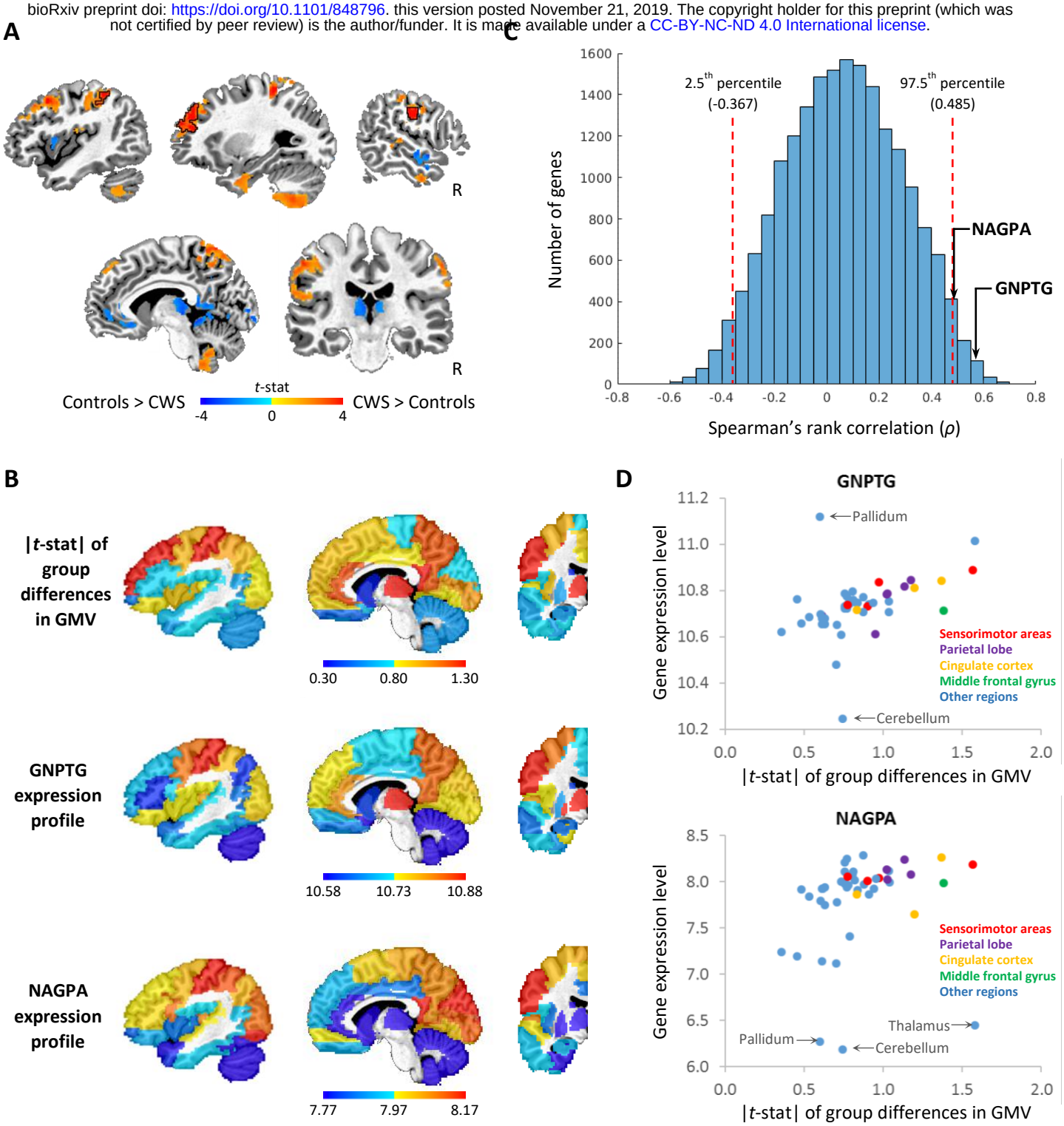


Figure 1. Relationship between gene expression and between-groups differences in gray matter volume (GMV). (A) Voxel-wise differences between pCWS and controls in GMV. Color-coded *t* values of the group differences are overlaid on an anatomical image. Areas exhibited a significant between-group difference at multiple comparisons corrected $p<0.05$ are outlined by black lines. The other colored areas are subthreshold (uncorrected $p<0.1$). (B) Parcellated gene expression of GNPTG and NAGPA and absolute GMV differences in *t*-statistics (|*t*-stat|) in 45 left hemispheric regions and the right cerebellum (which anatomically connects to the left hemisphere) were overlaid on a single-subject anatomical image. The parcellation of the brain was based on a standard atlas with automated anatomical labeling (AAL). Gene expression and |*t*-stat| of the right cerebellum were displayed in the left cerebellum to save space. (C) Frequency plot of Spearman's correlation coefficients between GMV group differences and each of the 19,174 genes expressed across the regions. The red dash lines indicate the levels of correlation at the 2.5th and 97.5th percentiles. To obtain a correlation >0.474 or <-0.360 is less than 5% chance if a gene is randomly selected. (D) Scatter plots between gene expression and group differences in GMV in the sensorimotor areas (red dots), the parietal lobe (purple dots), the cingulate cortex (orange dots), the middle frontal gyrus (green dots) and the rest of the regions (blue dots). Regions in which gene expression is 2.5 standard deviations above or below mean were labelled.