scientific reports

Sections of a Stage 1 Registered Report
Scientific Reports

- Title page
- Abstract
- Introduction (no subheadings permitted)
- Methods
  - *Ethics information*
  - *Pilot data (if applicable)*
  - *Design*
  - *Sampling plan*
  - *Analysis plan*
- Data availability statement
- Code availability statement
- References
- Acknowledgements
- Author contributions
- Competing interests
- Figures & Figure captions
- Tables (A Design Table is mandatory)
- Supplementary information

Reproducibility of Cerebellar involvement as quantified by consensus structural MRI biomarkers in Advanced Essential Tremor

Qing Wang[1], Meshal Aljassar[2], Nikhil Bhagwat[1], Yashar Zeighami[3], Alan C Evans[3], Alain Dagher[3], G. Bruce Pike[4], Abbas F. Sadikot[2]*, Jean-Baptiste Poline[1]*

[1] Neuro Data Science - ORIGAMI laboratory, McConnell Brain Imaging Centre, The Neuro (Montreal Neurological Institute-Hospital), Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada
[2] Neurosurgery Clinic, McConnell Brain Imaging Centre (BIC), The Neuro (Montreal Neurological Institute-Hospital), Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada
[3] Ludmer Centre for Neuroinformatics and Mental Health, McConnell Brain Imaging Centre (BIC), The Neuro (Montreal Neurological Institute-Hospital), Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada
[4] Cumming School of Medicine, Hotchkiss Brain Institute (HBI), Department of Radiology, University of Calgary, Calgary, Quebec, Canada

* Corresponding authors: Jean-Baptiste Poline (jean-baptiste.poline@mcgill.ca) and Abbas F. Sadikot (abbas.sadikot@mcgill.ca).

Abstract

Essential Tremor (ET) is the most prevalent movement disorder with poorly understood etiology. Some neuroimaging studies report cerebellar involvement, whereas others find no significant differences between ET and control groups. This discrepancy may stem from the underpowered studies as well as differences in Magnetic Resonance Imaging (MRI) acquisition and processing. To help resolve these differences, we plan to analyze the structural MRI scans from 1) an advanced ET cohort and normal controls (NC) acquired at the Montreal Neurological Institute and 2) additional NC subjects from PPMI and ADNI. We will test the hypothesis that the cerebellar involvement in advanced ET can be detected with multiple neuroimaging biomarkers: 1) cerebellar VBM, 2) cerebellar gray/white matter volumetry, and 3) cerebellar lobular volumetry. We will rigorously evaluate the sensitivity of the hypothesis tests to the underlying methods by varying image processing algorithms and confounder control design choices. Subsequently, we will also report the cortical changes associated with cerebellar "degeneration" in the advanced ET in an exploratory analysis.

Introduction

Essential tremor (ET) is one of the most common chronic neurological movement disorders with an overall prevalence of 0.9 - 4.6%[1]. The International Parkinson and Movement Disorder Society defines ET as a syndrome characterized by isolated bilateral upper limb action tremor with a duration of at least 3 years, with or without tremor in other locations, such as head, voice tremor, or lower limbs. Additional typical neurological signs, such as balance impairment, abnormal posturing of the limbs, or memory loss may emerge as the disease progresses[2].

Although the underlying pathophysiology of ET remains unknown, post-mortem studies reveal changes in the cerebellar cortex, primarily involving Purkinje cell loss in the cerebellar gray matter[3,4]. However, in-vivo neuroimaging studies report inconsistent findings pertaining to cerebellar involvement in ET.

Several Magnetic Resonance (MR) imaging[1] studies suggest structural changes in the cerebellum associated with ET[5–9]. Specifically, earlier work based on voxel-based morphometry (VBM) suggests bilateral cerebellar atrophy in ET subjects[10,11], especially in the vermis[10]. More recently, volumetric studies assisted by high-resolution cerebellar atlases[12] report significant decreases of GM in different cerebellar lobules I-IV, V, VI, VII and VIII in addition to the vermis[13]. Contrary to these findings, other work indicates that there is no significant association between ET symptoms and cerebellar degeneration[14–16]. For example, Rajput et al. found cerebellar Purkinje cell loss could not serve as either a pathological basis for, or the distinctive feature of ET[14]. In addition, a meta-analysis study comprising 16 pooled VBM studies also *fails to find consistent cerebellar abnormalities* and gray matter alterations in the ET population[15].

In another recent voxel-based morphometry (VBM) meta-analysis[17], the authors indicate that *"the cerebellum undergoes certain volumetric changes in essential tremor patients".* The authors also report the significant heterogeneity found in the published studies and conclude that "*the high heterogeneity makes the result less reliable"*. The small median sample size of studies (n=19.5 for ET group, 20 for NC group) considered in their meta-analysis leads to an estimated median power of less than 15% for a conservative estimate of 10 multiple comparisons. This implies the results are likely to suffer from the winner's curse effect and low positive predictive value, in addition to the file drawer effect. Both factors make the meta-analysis results difficult to interpret, such that the authors themselves call for new studies to confirm or infirm their meta-analytic findings. Meta-analysis caveats and winner's curse effects are not specific to the field of neuroimaging[18,19] but pervasive in the low power settings observed here.

Furthermore, beyond the cerebellar involvement hypothesis, cerebello-thalamo-cortical network theory has been proposed as an important pathogenesis in ET[20,21]. However, in neuroimaging studies, the cortical changes in ET and their association with cerebellar degeneration are not well characterized and lacks consensus[10,22–27]. Increase in gray matter in the supplementary motor area of ET patients based on a VBM analysis has also been reported[28]. These inconsistencies motivate further exploration of coincidence of cerebellar and cortical change patterns to improve our understanding of degeneration progression in ET.

The current inconsistencies in MR imaging studies that link varying cerebellar changes to ET can be attributed to various sources. Collection and analysis of disparate cohorts with small sample sizes complicates hypothesis testing and interpretation of findings. Apart from the difficulties of collecting large scale well characterized randomized ET and control subjects, the disagreements between imaging studies may also arise from the complexity and flexibility of the neuroimaging processing pipelines and the statistical models and parameter settings variations[29–31]. These pipelines include VBM, region-of-interest (ROI) volumetry, and cortical thickness estimation which offer quantification of biomarkers at different scales and regional specificities. We refer to the study of the variability in findings resulting from different methodological pipelines as "methods sensitivity analysis". Typically, the methodological variation stems from underlying hypotheses about biomarker's spatial specificity and sensitivity (e.g., voxels vs regions) in identifying case-control differences; as well as the choice of computational pipelines (or toolboxes, e.g., FreeSurfer vs SUIT vs MAGet Brain) used to quantify the phenotype of interest. Most

[1] We note that in the in-vivo MR imaging studies, the quantified structural MRI biomarker can only suggest an underlying biological phenomenon. In studies such as this, the quantified signal loss is often interpreted as "degeneration or atrophy", but confirming such biological mechanisms would require ex-vivo studies. Nevertheless, a consensus in-vivo findings is a critical step towards demonstrating causal biological relationships. This work is an effort towards such an approach that combines multiple voxel- and regional-level neuroimaging phenotypes within a meta-analytic framework and assesses methodological robustness to address the current inconsistencies in the ET imaging studies and offer better insight into mechanisms of tremor in ET.

of the aforementioned studies choose only one among the many available imaging analysis pipelines, such as VBM using SPM, or region-of-interest (ROI) analysis using FreeSurfer. The lack of identical (or similar) pipelines between two studies complicates direct comparison of the results. The next source of variability in the analysis comes from differences in the statistical modeling. The existing literature employs varying approaches towards controlling confounders and covariate selection that can introduce more inconsistencies in the biological findings[5,8,10,11,13]. Problematizing the situation further, at times there are no statistical and neuroimaging reporting standards followed in the literature. For some studies, we were not able to find full details of the statistical analyses , for example, not all the z or t values, effect sizes, and details of multiple comparison corrections[5,10,8,13]. Additionally, studies also perform analysis based on presumed disease subtypes that may in fact exist in a continuum could also dilute statistical power and inflate effect sizes[18,19]in smaller cohort studies. All these complexities, compounded possibly by the file drawer effect, i.e., the publication bias towards reporting of significant findings[32], make the comparison and interpretation of neuroimaging studies difficult, and hinder the translation of research findings to clinical applications[33,34].

To address these methodological issues in the currently reported ET imaging literature results, we plan to carry out multiple neuroimaging analyses at different phenotypic scales and compare against the findings from literature. For these analyses, we will use a local sample of ET patients referred to a specialized neurosurgical movement disorders clinic. The patients present with an advanced stage of ET with disabling upper extremity symptoms. The local sample also comprises a limited number of control subjects however their age and sex are not well-matched with the ET group. We will augment the control sample size by drawing from two publicly available datasets: the Parkinson's progression markers initiative (PPMI)[35] and Alzheimer's Disease Neuroimaging Initiative (ADNI)[36], comprising control subjects with similar age and sex distributions as of local ET sample. A visual summary of the proposed analyses is presented in Fig. 1. With this augmented sample, we aim to investigate group differences between ET (Essential Tremor) and NC (Normal control) groups using structural imaging biomarkers derived from T1 MRIs. Specifically, we aim to answer the following three questions:

1) Can we detect a consensus cerebellar involvement as quantified by derived structural MR imaging features in an advanced Essential Tremor (ET) sample?

2) What is the impact of methodological pipeline selection resulting from the use of different image processing algorithms and statistical models on the above findings? Could these variations explain the literature discrepancies?

3) Are there any covarying structural change patterns between cerebellar and cortical thickness?

To answer question 1, we will test the hypothesis that the ET group will show significant cerebellar changes compared to NC group that are detectable using a consensus of 3 different MRI biomarkers: 1) cerebellar voxel-based morphometry (VBM), 2) cerebellar gray and white matter volumetry, and 3) cerebellar lobular volumetry. The details of the hypothesis testing procedures are summarized in the design table (Tab. 1).

We will answer the second research question of the impact of pipeline selection with a systematic methodological sensitivity analysis that includes: 1) comparisons with alternative segmentation pipelines to estimate cerebellar lobular volumes, 2) alternative confounder control models and intracranial volume

choices. We will investigate the third question by comparing the differences in the correlation patterns between cerebellar and cortical structural features of ET and NC groups in a secondary *exploratory* analysis.

In summary, we propose an innovative and principled approach to analyze a new augmented cohort with high power and methodological rigor. Our novel methodological sensitivity analysis to assess the impact of multiple image processing pipelines on the biological finding will offer a rigorous hypothesis testing framework to characterize and validate cerebellar involvement in ET. Additionally, our complementary exploratory analysis will provide insights into the interaction between cerebellar changes and other brain networks potentially involved in ET, which are not yet described in the literature.

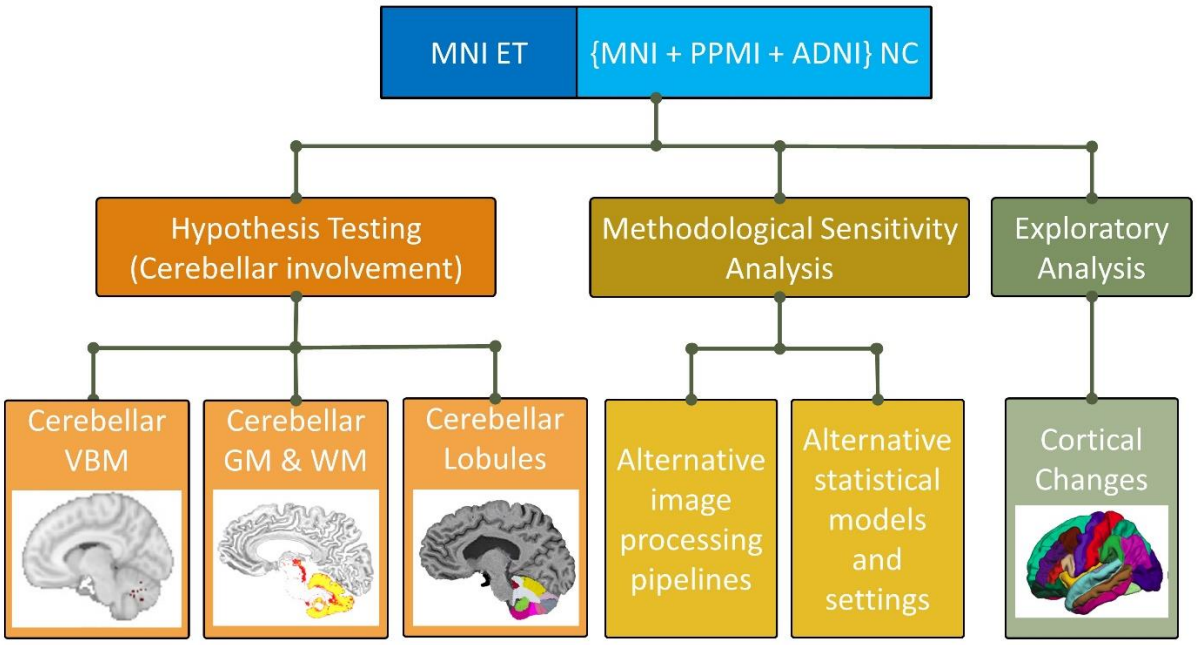

Figure 1. Research summary showing MR image processing tasks and the meta-analytical framework (Brain images are for illustration purposes).

Methods

Pilot data

Not applicable.

Design

Not applicable.

##

Datasets

This study will use 3 datasets which have already been collected. The MNI dataset has 70 subjects including 38 well characterized pre-surgical advanced ET subjects and 32 normal control (NC) subjects.

The PPMI dataset is a subset of the PPMI control cohort with 116 NC subjects. The ADNI2 dataset is a subset of the ADNI control cohort with 312 NC subjects. Note that some subjects will not be included due to image processing failures or not meeting quality control criteria (more details in Quality Control (QC) section). The pooled PPMI and ADNI2 dataset will be sampled according to sample size requirements estimated from the power analysis. The age and sex distribution of these cohorts are illustrated in Fig. 2 and summarized in Tab.2. Cohort membership will be modeled as a linear random effect in the latter analysis.

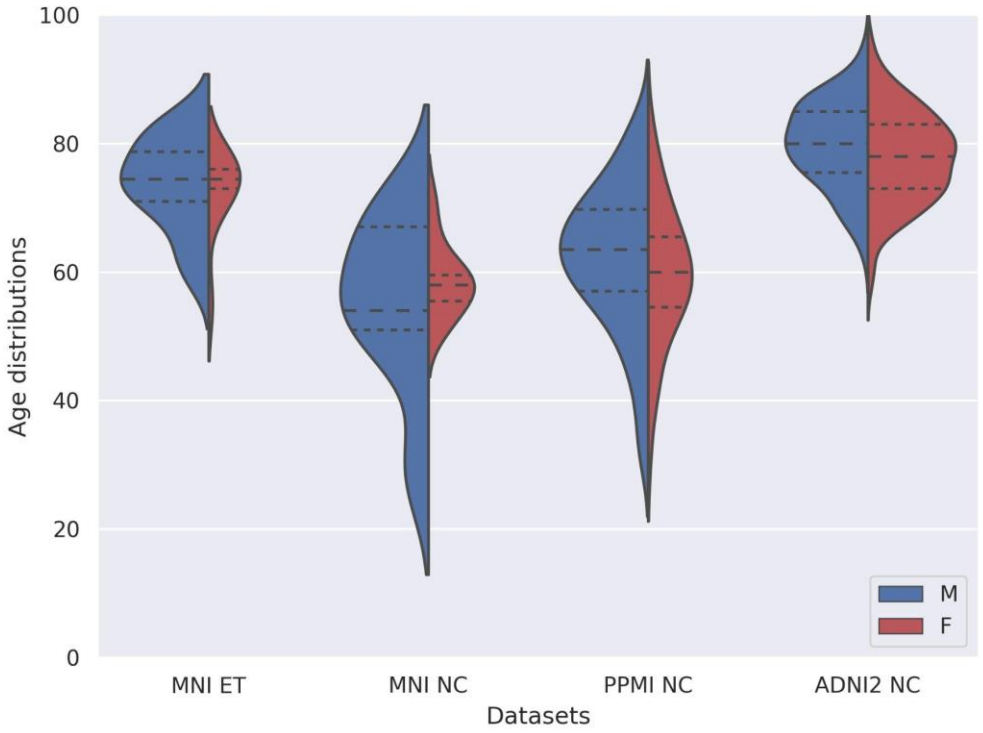

Figure 2. Study cohorts age and sex distributions. Each double sided violin plot shows the distribution for each cohort (red for female, blue for male), and they are MNI ET, MNI NC, PPMI NC and ADNI NC from left to right.

MNI dataset

This dataset has been collected at the MNI (Montreal Neurological Institute) as part of a research protocol, all subjects gave their consent to participating in the imaging acquisition protocol which has been approved by McGill University Research Ethics and Compliance (IRB). The images are acquired with a 3-Tesla Siemens scanner including T1w (T1-weighted), T2w (T2-weighted), diffusion-weighted (DWI) contrasts and resting fMRI acquisitions, we are going to focus on the analysis of T1w structural images in this study. The T1w images are acquired with TR=2300ms, TE=2.96ms and FOV of 256mm, the voxel size is 1mm×1mm×1mm. Since this cohort is being considered for thalamic surgery for tremor, the patients are well characterized for the diagnosis of advanced ET by neurologists and neurosurgeons sub-specializing in movement disorders. Most participants have a greater than 10-year history of

predominantly bilateral hand/arm tremor. Our current local cohort includes multi-modal MRIs from 38 ET patients and 32 healthy controls, representing a large dataset compared to other published imaging studies on ET. The demographics of the subjects are described in Tab. 2 below.

PPMI dataset

PPMI (Parkinson's Progression Markers Initiative, www.ppmi-info.org) provides an open access multicenter longitudinal study designed Parkinson Disease (PD) dataset funded by Michael J Fox Foundation. The PPMI dataset was collected under the approval from a local research ethics committee before study initiation and obtained written informed consent from all subjects participating in the study. This dataset consists of T1w, DWI and resting state fMRI images of the Parkinson Disease (PD) and normal control (NC) subjects. The complete demographic and acquisition protocol details can be found at https://www.ppmi-info.org/. PPMI consists of 116 NC subjects with 3T scan (with similar acquisition parameters as the local MNI dataset) with average age of 61.2±11.1.

ADNI2 dataset

The Alzheimer's Disease Neuroimaging Initiative (ADNI) was launched in 2003 directed by Michael W. Weiner. The ADNI dataset was collected under the approval of a local research ethics committee. This dataset provides T1w structural MRI images of Alzheimer's Disease (AD) and normal control (NC) subjects. The complete demographic and acquisition protocol details can be found at http://adni.loni.usc.edu/. The ADNI2 cohort from the ADNI project consists of 312 NC subjects with 3T T1w scans (with similar acquisition parameters as the local MNI dataset). ADNI2 cohort has an average age of 73.4±6.3 which is closer to our MNI ET cohort (73.4±7.0).

Table2. Characteristics of the ET and control subjects (data are given as mean±standard deviation)

Cohort (n)	Sex(M/F)	Age(years)	Across cohort tests
ET subjects (38)	28/10	73.4±7.0	Age difference between MNI ET/MNI NC ($p=<10^{-5}$, t test) Sex difference between MNI ET/MNI NC (p=0.6375, Chi-square test)
NC subjects (32)	21/11	56.3±11.1	

PPMI NC subjects (116)	77/39	61.2±11.1	Age difference between MNI ET /PPMI NC ($p=<10^{-5}$, t test) Sex difference between MNI ET/PPMI NC (p=0.5435, Chi-square test)
ADNI2 NC subjects (312)	151/161	73.4±6.3	Age difference between MNI ET /ADNI2 NC ($p<10^{-3}$, t test) Sex difference between MNI ET/ADNI2 NC (p=0.0056, Chi-square test)

Sampling plan

Power analysis

We perform our power calculations by setting α = 0.05, power = 0.9, and the median effect size of 0.61 (estimated from the literature in Fig. 3a). We calculated Cohen's d as a measure of the effect size from either the reported z values or the means and standard deviations of each study group. We report a positive effect size when the ET group shows cerebellar atrophy compared with the NC.

The literature review (Fig. 3a) shows that the four ROI volumetry studies report effect sizes in the range of 0.15-0.83. In comparison, the relevant VBM study (with exact z and p values rather than only reporting significant or not) effect sizes are in the range of 0.94-1.76. The higher effect sizes from VBM analysis are likely due to the use of peak z values. We use the median (0.61) as an aggregate estimate of prior effect sizes in our power analysis.

Fig. 3b shows how the statistical power changes with a fixed number of 38 ET subjects and an increasing number of pooled NC subjects from MNI, PPMI and ADNI2 datasets. For the median literature effect size is 0.61, with 38 ET patients, the number of NC subjects needed from MNI, PPMI and ADNI2 is at least *61* for 1-sided test and *116* for 2-sided test. In other words, we will use all of the *38* ET subjects as the ET group, and randomly select *116* age and sex matched NC subjects from the pooled MNI, PPMI and ADNI2 NC subjects as the control group. This means we will use *154* subjects in total for this study. We note that the quality control procedure may result in fewer ET and NC subjects. To achieve the pre-registered power of 0.9, significance level of 0.05 and effect size of 0.61, we will recalculate the number of NC subjects needed based on the number of ET subjects that pass QC before carrying out any further analysis. A reduced number of ET subjects will increase the number of NC subjects needed to achieve the same power.

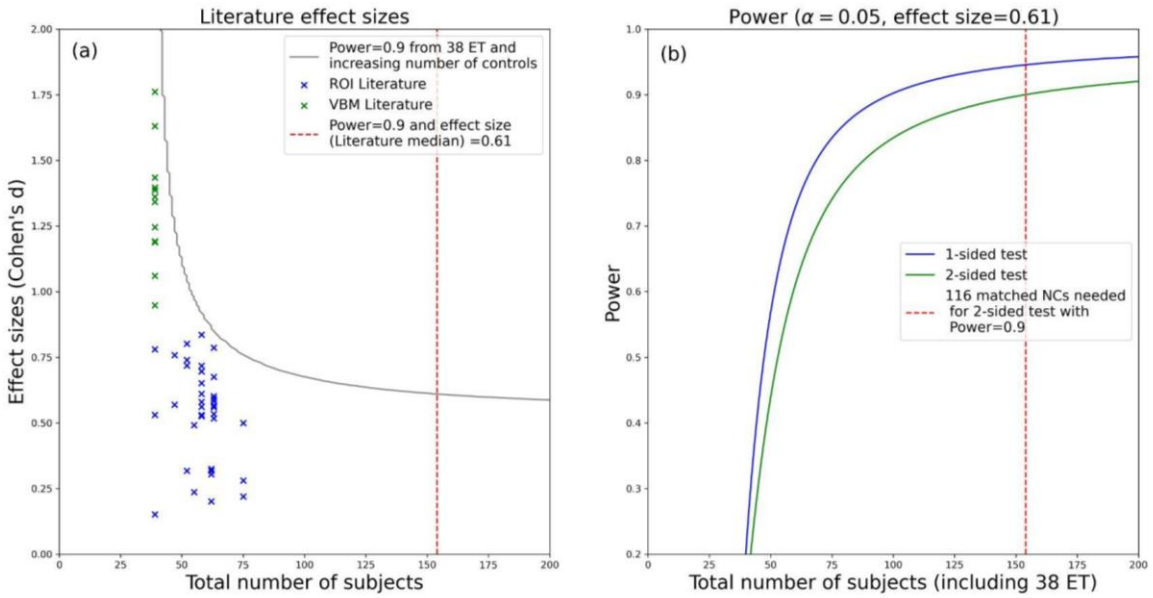

Figure 3. Power analysis. (a) The effect sizes reported from literature of VBM and ROI analyses and the number of subjects used are shown as crosses with blue and green colors, the gray line is the power=0.9 and alpha=0.05 line (ET group fixed at 38 subjects) , the red vertical line is the median literature effect size 0.61; (b) The power obtained with increasing number of control subjects (from MNI, PPMI and ADNI2) for both 1-sided (blue) and 2-sided t tests (green) with effect size set to the median of literature effect size (0.61) and alpha=0.05. The number of NC subjects needed for 1-sides and 2-sided tests are 61 and 116 respectively.

We conducted the above power analysis with the python package statsmodels 0.12.0 and verified by G*Power 3.1.9.7 software. All of the codes and figures are shared here: https://github.com/neurodatascience/ET_biomarker. A copy of the shared code will also be stored as part of the pre-registration.

Control group augmentation

In the primary analysis comprising cerebellar involvement hypothesis testing, we will carry out 2-sided significance tests. We will sample as many subjects as needed (to achieve 0.9 power) from the MNI, PPMI and ADNI2 NC subject pool while matching for their age and sex with the ET group. Similar to Spiel's approach[37], our group matching procedure is based on a L2 distance measure between a sample in the NC subject pool and the MNI ET group's age distribution for each sex. We rank all NC subjects based on this L2 distance, stratified by sex. The first 116 NC subjects with the smallest distance are selected to form the matched NC group. The detailed sampling procedure is described in Procedure 1. We will test the resultant sex and age distribution for the matched NC and ET groups with Chi-square and 2-sided t test, respectively.

Procedure 1: NC group augmentation procedure.

1) Split the MNI ET group into ET male group and ET female group, calculate the ratio of male and female subjects, use this ratio and the number of NCs required (116) based on the power

analysis to calculate the number of male and female NCs needed, denoted by N_M^{Target} and N_F^{Target};

2) For each male NC subject j, calculate the Euclidean distance Mj between NC_j and all the ET male cohort, sort Mj in an ascending (smallest first) order.

$$Mj = \sqrt{\sum_{i=1}^{N_{ETm}} \left(age_i^{ET} - age_j^{NC} \right)^2}$$

3) Repeat step 2 for the female NC subjects, compute the distances F_j and sort in ascending order.

4) Take the top N_M^{Target} male NC subjects with the smallest Mj and take the top N_F^{Target} subjects with the smallest F_j, pool together these selected NC subjects to form the augmented NC group.

Analysis Plan

Proposed analysis pipeline

The proposed computational pipeline consists of MR image preprocessing, a set of image processing methods for quantifying neuroimaging phenotypes, manual quality control procedures, and statistical modelling for hypothesis and exploratory analyses. We will use the default settings for all the imaging preprocessing and processing pipelines for our comparison with literature results.

MR image preprocessing

The original raw (dicom) T1-weighted (T1w) MR images are converted into NIfTI format and further organized according to BIDS standard with HeuDiConv[38]. All the T1 data are preprocessed with the anatomical workflow of fMRIPrep 20.2.0[39,40]. Briefly, the fMRIPrep pipeline performs the following structural image preprocessing tasks: 1) Intensity non-uniformity with N4BiasFieldCorrection[41]; 2); Skull-stripping (i.e. brain extraction) using ANTs workflow[42]; 3) Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) of the brain-extracted T1w using FAST (FSL 5.0.9)[43]; 4) Volume-based spatial normalization to the standard MNI152NLin2009cAsym space[44] through nonlinear registration with ANTs.

Voxel-based morphometry (VBM) processing

We will carry out voxel-based morphometry analysis using SPM 12 (Rev number: 7771) toolbox (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Briefly, VBM is a neuroimaging technique that allows a voxel-wise comparison of regional gray matter 'density' between two groups of subjects[45]. The process involves 1) Spatial normalization involving a nonlinear registration to each T1w image to a common template; 2) Tissue segmentation in gray matter, white matter, and CSF classes; and 3) Spatial smoothing of MR images into a stereotaxic space. Once the voxel-wise correspondence is established, statistical group comparisons via statistical parametric mapping are carried out to detect focal, regional changes in neuroanatomy. Notice that we will apply a cerebellar mask in the hypothesis testing VBM analysis to restrict our analysis for the cerebellar region.

Cerebellar Gray Matter (GM) and White matter (WM) volumetry

We will estimate cerebellar GM and WM volumes using FreeSurfer pipeline (http://surfer.nmr.mgh.harvard.edu/, version 6.0.1) using the default "recon- all" processing part of the fMRIPrep pipeline described earlier. The volume-based stream of recon-all is designed to classify MR voxels into subcortical tissue classes. It involves 1) Affine registration to the MNI305 space[46]; 2) Initial volume labeling and bias field correction; 3) A nonlinear volumetric alignment to the MNI305 atlas; 4) Volume labeling based on the voxel-to-voxel correspondence and probabilistic regional membership. We will quantify gray and white matter volume in the cerebellum using default "DKT atlas+aseg" labels.

Cerebellar segmentation

We will use SUIT pipeline[12] to segment the individual cerebellar volumes into lobules. SUIT is the most commonly used pipeline to segment the cerebellar lobules. It first extracts the cerebellum from the entire brain image, then segments the cerebellar gray and white matter and finally segments the cerebellar gray matter into 34 lobules according to the SUIT atlas.

In contrast to SUIT's single atlas approach, MAGeT Brain[47] pipeline employs a multi-atlas procedure to perform volumetric segmentation of brain structures. The multi-atlas approach combined with an intermediate cohort-specific bootstrapping procedure can better capture the neuroanatomical variability offering more accurate segmentations. We will compare the volume estimates from MAGeT Brain with the SUIT output as part of the methodological sensitivity analysis.

Quality control

The quality of the images and the processed results (registration and segmentation) will be evaluated by two expert neuroanatomists (M.A. and A.F.S.) and an imaging expert (Q.W.). We have 4 levels for the image quality from excellent (4), good (3), acceptable (2) to exclude (1). We will report the final quality assessment results and annotations along with comments for all excluded subjects. We will carry out the quality control procedure on the processed images from both our local ET/NC dataset and the NC subjects from PPMI and ADNI.

We will include all the subjects with acceptable imaging quality (>1) in this dataset, except:

1. Poor imaging quality (ranked as 1): We will exclude the subjects with poor imaging quality, for example heavy acquisition artifacts in cerebellum or severe brain tissue loss.
2. Low preprocessing quality or preprocessing error (ranked as 1), for example low quality of registration or segmentation.
3. We will exclude the subjects with abnormal brain structures (which may not directly relate to brain dysfunction), for example subjects with very large ventricles. The decisions will be made mainly according to the guidelines from our expert neuroanatomists (M.A. and A.F.S.).

Statistics

As described above, we will carry out three sets of analyses. The first set will consist of hypothesis testing pertaining to cerebellar specific differences. Subsequently, we will perform sensitivity analysis to assess the robustness of our findings subject to methodological variation. Finally (not included in Stage 1) we will

perform secondary exploratory analyses contingent on the findings from the hypothesis tests. We detail these analyses below.

Hypothesis testing

To answer the first research question, we will test the hypothesized cerebellar structural differences associated with ET compared to the NC group. We will use the general linear model (GLM) framework for assessing volumetric and morphometric cerebellar differences between ET and NC groups. All three analyses (i.e., VBM, GM-WM volumetry, and lobular volumetry, as described in table 1) will include age, sex, cohort (i.e., MNI, PPMI, ADNI), and estimated intracranial volume as covariates assuming that the individual differences in the head size is confounding the main effect. We will use 2-sided significance tests at the 0.05 level for each analysis. For the VBM analysis we will use False Discovery Rate (FDR) with Benjamini-Hochberg (BH) procedure. For cerebellar gray and white matter volumetry, we will test left and right cerebellar GM and WM separately. Then in the regional analysis we will test vermis VI, VII and VIII and crus I, crus II, dentate nucleus for volumetric differences. In both volumetric analyses, we will use 2-sided significance tests at the 0.05 level with age, sex, cohort (i.e., MNI, PPMI, ADNI) and estimated total intracranial volume (eTIV) as covariates and will correct for the number of ROIs with Bonferroni procedure.

We will confirm the involvement of cerebellum in ET if each of the three hypotheses in Tab. 1 meet the .05 significance level, yielding a conservative statistical threshold.

Methodological sensitivity and robustness analysis

In the methodological sensitivity and robustness analysis, we would like to evaluate the impact of the following image processing methods and statistical models on the results.

1. Cerebellar volumetry and cerebellar segmentation pipeline selection

   Cerebellar volumetry can be sensitive to the choice of segmentation pipelines and anatomical atlas. Therefore, we will compare the lobular volumetric group differences derived from 1) SUIT pipeline with SUIT atlas; 2) MAGeT Brain pipeline with a multi-atlas segmentation method to assess the sensitivity of pipeline selection. Notice that: SUIT atlas and MAGeT Brain atlas have good correspondence for all the hemispheric cerebellar lobules, but only SUIT provides vermis and deep nucleus (dentate nucleus) volumetry estimations.

2. Confounder control methods sensitivity analysis

   We will compare 2 approaches for controlling the effects of known confounders.

   a. Covariate inclusion: Confounders (age, sex, estimated intracranial volume (eTIV), and cohort) are included as covariates in GLM, for example the model can be: Voi = beta0 + beta1*age + beta2*sex + beta3*eTIV + beta4*cohort + beta5*group, where Voi is volume of interest, e.g., the volume of left cerebellar cortex, and group is ET or NC.
   b. Variable transformation: Use log transformed or intracranial volume normalized volumes of interest (proportion adjustment and power proportion adjustment) in the GLM analysis instead of the original variables, for example: Vpa = Voi/eTIV (proportion adjustment, PA); Vppa = Voi/(eTIV^b), log(Vppa)=beta0 + b*log(eTIV) (Power proportion adjustment, PPA), Vpa is the proportion adjusted volume and Vppa is the power proportion adjusted volume,

and the GLM model will be Vpa(Vppa) = beta0 + beta1*age + beta2*sex + beta3*cohort + beta4*group instead.

In addition, we will also test whether using total cerebellar volume instead of eTIV to adjust for global volumetric effects in the above 2 approaches and report the comparisons.

Data availability

We plan to share the data used directly for all the statistical analysis, tables, and figures. Due to the constraints from our research protocol, we are not able to share the raw local clinical imaging dataset directly, however, all derived data will be shared. The PPMI consortium provided open access for their dataset at https://ida.loni.usc.edu/login.jsp?project=PPMI. Access to the ADNI dataset is provided through the ADNI consortium at http://adni.loni.usc.edu/data-samples/access-data/.

Code availability

All the codes and figures are shared via GitHub: https://github.com/neurodatascience/ET_biomarker. (Including the power analysis code with the python package statsmodels 0.12.0.)

Results

Do not include a Results section.

Discussion

Do not include a Discussion section.

References

1. Louis, E. D., Ford, B. & Barnes, L. F. Clinical Subtypes of Essential Tremor. *Arch Neurol* **57**, 1194 (2000).
2. Haubenberger, D. & Hallett, M. Essential Tremor. *New England Journal of Medicine* (2018).
3. Louis, E. D. *et al.* Neuropathological changes in essential tremor: 33 cases compared with 21 controls. *Brain* **130**, 3297–3307 (2007).
4. Louis, E. D. & Faust, P. L. Essential tremor: the most common form of cerebellar degeneration? *Cerebellum & Ataxias* **7**, 12 (2020).
5. Quattrone, A. *et al.* Essential Head Tremor Is Associated with Cerebellar Vermis Atrophy: A Volumetric and Voxel-Based Morphometry MR Imaging Study. *American Journal of Neuroradiology* **29**, 1692–1697 (2008).
6. Pagan, F. L., Butman, J. A., Dambrosia, J. M. & Hallett, M. Evaluation of essential tremor with multi-voxel magnetic resonance spectroscopy. *Neurology* **60**, 1344–1347 (2003).
7. Passamonti, L., Cerasa, A. & Quattrone, A. Neuroimaging of Essential Tremor: What is the Evidence for Cerebellar Involvement? *Tremor Other Hyperkinet Mov (N Y)* **2**, (2012).
8. Shin, H. *et al.* Atrophy of the Cerebellar Vermis in Essential Tremor: Segmental Volumetric MRI Analysis. *Cerebellum* **15**, 174–181 (2016).

9. Han, Q., Hou, Y. & Shang, H. A Voxel-Wise Meta-Analysis of Gray Matter Abnormalities in Essential Tremor. *Front. Neurol.* **9**, (2018).
10. Benito-León, J. *et al.* Brain structural changes in essential tremor: Voxel-based morphometry at 3-Tesla. *Journal of the Neurological Sciences* **287**, 138–142 (2009).
11. Bagepally, B. S. *et al.* Decrease in Cerebral and Cerebellar Gray Matter in Essential Tremor: A Voxel-Based Morphometric Analysis under 3T MRI. *Journal of Neuroimaging* **22**, 275–278 (2012).
12. Diedrichsen, J. A spatially unbiased atlas template of the human cerebellum. *NeuroImage* **33**, 127–138 (2006).
13. Dyke, J. P., Cameron, E., Hernandez, N., Dydak, U. & Louis, E. D. Gray matter density loss in essential tremor: a lobule by lobule analysis of the cerebellum. *Cerebellum & Ataxias* **4**, 10 (2017).
14. Rajput, A. H., Robinson, C. A., Rajput, M. L., Robinson, S. L. & Rajput, A. Essential tremor is not dependent upon cerebellar Purkinje cell loss. *Parkinsonism & Related Disorders* **18**, 626–628 (2012).
15. Luo, R., Pan, P., Xu, Y. & Chen, L. No reliable gray matter changes in essential tremor. *Neurol Sci* **40**, 2051–2063 (2019).
16. Ibrahim, M. F., Beevis, J. C. & Empson, R. M. Essential Tremor – A Cerebellar Driven Disorder? *Neuroscience* **462**, 262–273 (2021).
17. Mavroudis, I. *et al.* A Voxel-Wise Meta-Analysis on the Cerebellum in Essential Tremor. *Medicina* **57**, 264 (2021).
18. Nakaoka, H. & Inoue, I. Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *Journal of Human Genetics* **54**, 615–623 (2009).
19. Button, K. S. *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* **14**, 365–376 (2013).
20. Mazziotta, J. *et al.* A Four-Dimensional Probabilistic Atlas of the Human Brain. *J Am Med Inform Assoc* **8**, 401–430 (2001).
21. Cury, R. G., França, C., Reis Barbosa, E., Jacobsen Teixeira, M. & Ciampi de Andrade, D. Little Brain, Big Expectations. *Brain Sciences* **10**, 944 (2020).
22. Chung, S. J. *et al.* Neuroanatomical Heterogeneity of Essential Tremor According to Propranolol Response. *PLOS ONE* **8**, e84054 (2013).
23. Lin, C.-H. *et al.* VBM Reveals Brain Volume Differences between Parkinson's Disease and Essential Tremor Patients. *Front. Hum. Neurosci.* **7**, (2013).
24. Serrano, J. I. *et al.* A data mining approach using cortical thickness for diagnosis and characterization of essential tremor. *Scientific Reports* **7**, 2190 (2017).
25. Archer, D. B. *et al.* A widespread visually-sensitive functional network relates to symptoms in essential tremor. *Brain* **141**, 472–485 (2018).
26. Pietracupa, S. *et al.* White matter rather than gray matter damage characterizes essential tremor. *Eur Radiol* **29**, 6634–6642 (2019).
27. Nicoletti, V. *et al.* Cerebello-thalamo-cortical network is intrinsically altered in essential tremor: evidence from a resting state functional MRI study. *Scientific Reports* **10**, 16661 (2020).
28. Gallea, C. *et al.* Intrinsic signature of essential tremor in the cerebello-frontal network. *Brain* **138**, 2920–2933 (2015).

29. Khundrakpam, B. S. *et al.* Exploring Individual Brain Variability during Development based on Patterns of Maturational Coupling of Cortical Thickness: A Longitudinal MRI Study. *Cerebral Cortex* **29**, 178–188 (2019).
30. Bhagwat, N. *et al.* Understanding the impact of preprocessing pipelines on neuroimaging cortical surface analyses. *bioRxiv* 2020.05.22.100180 (2020) doi:10.1101/2020.05.22.100180.
31. Botvinik-Nezer, R. *et al.* Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88 (2020).
32. Rosenthal, R. The file drawer problem and tolerance for null results. *Psychological Bulletin* **86**, 638–641 (1979).
33. Cerasa, A. & Quattrone, A. Linking Essential Tremor to the Cerebellum—Neuroimaging Evidence. *Cerebellum* **15**, 263–275 (2016).
34. Scarpazza, C. & Simone, M. S. D. Voxel-based morphometry: current perspectives. *Neuroscience and Neuroeconomics* vol. 5 19–35 https://www.dovepress.com/voxel-based-morphometry-current-perspectives-peer-reviewed-fulltext-article-NAN (2016).
35. Marek, K. *et al.* The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology* **5**, 1460–1477 (2018).
36. Petersen, R. C. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology* **74**, 201–209 (2010).
37. Spiel, C. *et al.* A Euclidean Distance-Based Matching Procedure for Nonrandomized Comparison Studies. *European Psychologist* **13**, 180–187 (2008).
38. Halchenko, Y. *et al. nipy/heudiconv:* (Zenodo, 2021). doi:10.5281/zenodo.5557588.
39. Gorgolewski, K. J. *et al.* BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Comput Biol* **13**, e1005209 (2017).
40. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* **16**, 111–116 (2019).
41. Tustison, N. J. *et al.* N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging* **29**, 1310–1320 (2010).
42. Avants, B. B. *et al.* The Insight ToolKit image registration framework. *Front. Neuroinform.* **8**, (2014).
43. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* **20**, 45–57 (2001).
44. Evans, A. C., Janke, A. L., Collins, D. L. & Baillet, S. Brain templates and atlases. *NeuroImage* **62**, 911–922 (2012).
45. Ashburner, J. & Friston, K. J. Voxel-Based Morphometry—The Methods. *NeuroImage* **11**, 805–821 (2000).
46. Dl, C., P, N., Tm, P. & Ac, E. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* **18**, 192–205 (1994).
47. Park, M. T. M. *et al.* Derivation of high-resolution MRI atlases of the human cerebellum at 3T and segmentation using multiple automatically generated templates. *NeuroImage* **95**, 217–231 (2014).

Acknowledgements

This work was partially funded by the National Institutes of Health (NIH) NIH-NIBIB P41 EB019936 (ReproNim) NIH-NIMH R01 MH083320 (CANDIShare) and NIH RF1 MH120021 (NIDM), the National Institute Of Mental Health under Award Number R01MH096906 (Neurosynth), the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives

initiative and the Brain Canada Foundation with support from Health Canada, Health Canada, through the Canada Brain Research Fund in partnership with the Montreal Neurological Institute. This work was also partially funded by the Brain Canada Foundation with support from the Foundation CERVO and the McGill Initiative in Computational Medicine. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Q.W., J.B. P. and A.F.S. conceptualized this study and wrote this paper together with N.B. and M.A. Y.Z. and A.D. contributed to the data curation and the idea formulation. A.F.S. and M.A. collected the MNI dataset and did the clinical assessments. Q.W. and M.A. curated the MNI dataset. Q.W. and B.N. downloaded the PPMI and ADNI dataset. All these datasets were preprocessed by Q.W., and the overall quality was assessed mainly by M.A. and Q.W. The datasets were analyzed by Q.W., B.N. and M.A. Q.W., N.B. and J.B. P. contributed mainly to the methodology design and evaluation. M.A., A.F.S. and B.P. mainly contributed to the clinical interpretation and clinical implications. J.B. P. and A.C.E. provided the computing resources on Compute Canada to finish this research.

Competing interests

The authors declare no competing financial and/or non-financial interests.

Figures

Figure Legends

Figure 1. Research summary showing MR image processing tasks and the meta-analytical framework (Brain images are for illustration purposes).

Figure 2. Study cohorts age and sex distributions. Each double sided violin plot shows the distribution for each cohort (red for female, blue for male), and they are MNI ET, MNI NC, PPMI NC and ADNI NC from left to right.

Figure 3. Power analysis. (a) The effect sizes reported from literature of VBM and ROI analyses and the number of subjects used are shown as crosses with blue and green colors, the gray line is the power=0.9 and alpha=0.05 line (ET group fixed at 38 subjects) , the red vertical line is the median literature effect size 0.61; (b) The power obtained with increasing number of control subjects (from MNI, PPMI and ADNI2) for both 1-sided (blue) and 2-sided t tests (green) with effect size set to the median of literature effect size (0.61) and alpha=0.05. The number of NC subjects needed for 1-sides and 2-sided tests are 61 and 116 respectively.

Table 1. Design Table

Question	Hypothesis (if applicable)	Sampling plan (e.g., power analysis)	Analysis Plan	Interpretation given to different outcomes

Does ET group show differences in cerebellar regions compared with NC group at the voxel level?	H0: ET group does not show differences in the cerebellar regions compared to the NC group. H1: ET group shows differences in the cerebellar regions compared to the NC group.	Detailed in sampling plan. Power analysis with Python statsmodels version 0.12.0.	VBM with alpha=0.05 using B-H false discovery rate control. Sex, age, eTIV (estimated total intracranial volume) and cohort will be used as covariates.	Reject null hypothesis if p<0.05 (with B-H false discovery rate control).
Does ET group show differences in cerebellar white matter and gray matter volumes compared with NC group?	H0: ET group does not show differences in cerebellar white matter and gray matter volume compared to the NC group. H2: ET group shows differences in cerebellar white matter and gray matter volume compared to the NC group.	Detailed in sampling plan. Power analysis with Python statsmodels version 0.12.0.	General Linear Model with alpha=0.05 using Bonferroni correction. Model: Voi=b0+b1*age+b2*sex+b3*cohort+b4*eTIV+b5*group; among them, Voi is the volume of interest, and it can be left cerebellar white matter and gray matter volume, and eTIV is the estimated intracranial volume.	Reject null hypothesis if p<0.05 (with Bonferroni correction)
Does ET group show differences in the following cerebellar lobules (Vermis VI, Vermis_CrusI, Vermis_CrusII,	H0: ET group does not show any differences in either the following cerebellar lobules compared to NC group: Vermis_VI, Vermis_CrusI,	Detailed in sampling plan. Power analysis with Python statsmodels version 0.12.0.	General Linear Model with alpha=0.05 using Bonferroni correction. Model: Voi=b0+b1*age+b2*sex+b3*cohort+	Reject null hypothesis if p<0.05 (Bonferroni correction).

CrusI, CrusII, Dentate nucleus) than NC group?	Vermis_CrusII, CrusI, CrusII, Dentate nucleus. H3: ET group shows differences in any of the following cerebellar lobules compared to NC group: Vermis_VI, Vermis_CrusI, Vermis_CrusII, CrusI, CrusII, Dentate nucleus.		b4*eTIV+b5*group; among them, Voi could be Vermis_VI, Vermis_CrusI, Vermis_CrusII, CrusI, CrusII, Dentate nucleus, and eTIV is the estimated intracranial volume.	

513

514

515 Supplementary information

516 Not applicable.