Neuroinformatics manuscript No.
(will be inserted by the editor)

Multi-Table Differential Correlation Analysis of Neuroanatomical and Cognitive Interactions in Turner Syndrome

Christof Seiler* · **Tamar Green*** · **David Hong** · **Lindsay Chromik** · **Lynne Huffman** · **Susan Holmes**** · **Allan L. Reiss****

Received: date / Accepted: date

Abstract Girls and women with Turner syndrome (TS) have a completely or partially missing X chromosome. Extensive studies on the impact of TS on neuroanatomy and cognition have been conducted. The integration of neuroanatomical and cognitive information into one consistent analysis through multi-table methods is difficult and most standard tests are underpowered. We propose a new two-sample testing procedure that compares associations between two tables in two groups. The procedure combines multi-table methods with permutation tests. In particular, we construct cluster size test statistics that incorporate spatial dependencies. We apply our new procedure to a newly collected dataset comprising of structural brain scans and cognitive test scores from girls with TS and healthy control participants (age and sex matched). We measure neuroanatomy with Tensor-Based Morphometry (TBM) and cognitive function with Wechsler IQ and NEuroPSYchological tests (NEPSY-II). We compare our multi-table testing procedure to a single-table analysis. Our new procedure reports differential correlations between two voxel clusters and a wide range of cognitive tests whereas the single-table analysis reports no differences. Our findings are consistent with the hypothesis that girls with TS have a different brain-cognition association structure than healthy controls.[1]

Keywords Permutation Tests · Multi-Table Analysis · Sparse Canonical Correlation Analysis · Turner Syndrome · Tensor-Based Morphometry · Cognitive Abilities

*joint first authors and **joint senior authors

Christof Seiler · Susan Holmes
Department of Statistics, Stanford University

Tamar Green
Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine; Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

David Hong · Lindsay Chromik
Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine

Lynne Huffman
Department of Pediatrics, Stanford University School of Medicine

Allan L. Reiss
Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine; Departments of Radiology, Psychiatry and Behavioral Sciences, Stanford University School of Medicine

1 Introduction

Turner syndrome (TS) is a genetic disorder caused by a complete or partial absence of one X-chromosome in females. TS occurs in about 1 in 2000 newborn girls (Sybert and McCauley, 2004; Gravholt, 2005) and has been linked to complex aberrant neuroanatomy (e.g. Hong et al. (2014)) and selective impairments in cognitive function (e.g. Hong et al. (2009)). Associations between neuroanatomical and cognitive profiles are rarely elucidated with few notable exceptions (e.g. Brown et al. (2004)).

We will refer to the joint statistical analysis of two separate data tables or matrices as multi-table analysis. One reason for the paucity of multi-table analysis between neuroanatomy and cognition is the heterogeneity of the different data sources. Brain images and cognitive subtest scores are measurements that live in different

[1] This is a post-peer-review, pre-copyedit version of an article published in Neuroinformatics. The final authenticated version is available online at: `http://dx.doi.org/10.1007/s12021-017-9351-z`.

mathematical spaces and are interpreted by different experts with significantly different scientific training.

One approach to study brain-cognition associations is through pairwise testing between brain regions and cognitive subtests. This approach ignores the dependencies between regions and between cognitive subtests. It assumes that brain regions are known a priori. However, this is not the case for regions in the presence of brain disorders and disease. This approach is also subject to multiple testing, as statistical tests are performed for each brain region and cognitive subtest pair, and one has to control for false discoveries by adjusting p-values using the Family-Wise Error Rate (FWER) or False Discovery Rate (FDR). Pairwise approaches are also unable to capture multivariate associations between brain regions and cognitive subtest variables.

In contrast, a multi-table approach to this problem can go beyond pairwise associations by considering linear combinations of the columns in both data tables. Many methods are available; Canonical Correlation Analysis (Hotelling, 1936), Partial Least Squares (Tucker, 1958; Wold, 1966; Fornell and Bookstein, 1982), Reduced-Rank Regression (Izenman, 1975) and their penalized versions for the high dimensional setting. The increased power to detect more than pairwise associations comes at the price of interpretability. It is often challenging to link results back to medically actionable information, partly because most multi-table methods are exploratory and do not assign significance levels.

In this article, we introduce a new two-sample testing procedure for multi-table method based on sparse Canonical Correlation Analysis (sCCA) (Parkhomenko et al., 2007; Waaijenborg et al., 2008; Parkhomenko et al., 2009; Lê Cao et al., 2009; Witten et al., 2009). Sparse CCA integrates two data sources observed on the same set of participants. It finds sparse, maximally correlated linear combinations of variables in both datasets simultaneously. Our procedure assigns significance levels using permutation tests and incorporates spatial dependency structure in the image domain using a newly constructed cluster size test statistic. Our procedure comprises the following steps: First, add column with group labels to the non-imaging data table. Second, compute sCCA between the non-imaging and imaging data tables. Third, compute test statistics from sCCA's canonical variables. Fourth, randomize group labels by permutation and recompute steps two and three many times to obtain null distribution of test statistics. Fifth, assign significance levels by comparing observed test statistics to the null distribution.

Our new method provides two advantages over pairwise association testing: First, pairwise tests miss weak associations due to testing many statistical hypotheses, which requires us to adjust for multiple comparisons and thus reduces the power to discover weaker associations. Second, pairwise tests ignore dependencies between neuroanatomy and cognition: multiple brain regions will most likely impact multiple cognitive functions. Such effects may not be detectable when considering brain regions individually.

The primary neuroscience objective of the present article is to elucidate complex interactions between subsets of neuroanatomical features and subsets of cognitive features. We measure neuroanatomy using Tensor-Based Morphometry, where local volume differences are represented by Jacobian determinant maps, and cognitive function using test scores from the NEuroPSYchological (NEPSY) battery and the Wechsler IQ test.

After describing our Turner syndrome dataset comprised of brain images, IQ and cognitive subtests, Tanner stages to track puberty, and other information about our participants, we will describe our preprocessing workflow including image registration and handling of missing values. Next, we will introduce our new two-sample multi-table testing procedure and compare it to a single-table analysis. We will apply our procedure to the Turner syndrome dataset and finish with a report of the potential discoveries of interest from a neuroscience viewpoint.

2 Materials and Methods

2.1 Participants

This study is part of an ongoing longitudinal investigation into gene, brain, and behavior in girls with TS. We included 54 girls with monosomic TS (mean age 10.2 ± 2.5, range $5.5 - 15.9$ years) and 48 healthy control participants (mean age 10 ± 2.1, range $5.2 - 14.2$ years). We recruited participants with TS through the national Turner Syndrome Society and the Turner Syndrome Foundation, and online advertisements at Stanford University School of Medicine. We recruited control participants through local print media and parent networks. The local Institutional Review Board at the Stanford University School of Medicine approved this study. We obtained informed written consent from legal guardians for all participants, and written assent from participants over 7 years of age.

2.2 MR Imaging Acquisition

We utilized a mock MRI scanner prior to the actual scan to desensitize participants to the sights and sounds of

an actual MRI environment. The participants underwent behavioral training to help reduce motion-related artifacts. We acquired all imaging data at the Stanford University Lucas Center for Medical Imaging. Magnetic resonance images of the young cohort were collected between 2012 and 2015 on a GE Healthcare Discovery 750 MRI whole-body MR system (GE Medical Systems, Milwaukee, WI) using a standard birdcage head coil. We employed a fast spoiled gradient recalled (FSPGR) echo pulse sequence to obtain a high-resolution T1 anatomical brain image of each subject sagittal slices, repetition time = 8.2 ms, echo time = 3.2 ms, inversion time = 450 ms, flip angle = 12°, number of excitations = 1, field of view 240×192 mm; matrix 256×256; 176 slices; voxel size = $1.0 \times 1.0 \times 1.0$ mm thickness, acquisition time = 4 min 29 seconds.

2.3 Image Processing

We applied the following preprocessing steps to all scans: alignment to the plane defined by the anterior and posterior commissures (ACPC), voxel resampling isotropically at 1mm, correction of bias field with N4 (Smith, 2002), and removal of voxels comprising skull and scalp (Tustison et al., 2010).

We then employed a standard Tensor-Based Morphometry (TBM) approach (Davatzikos et al., 1996; Freeborough and Fox, 1998; Gee and Bajcsy, 1998; Chung et al., 2001) that permits localization of brain structures that are different in size and shape but do not necessarily conform to exact sulcal-gyral locations. The key step in TBM is the computation of deformation fields (and their inverses) that map from a template image to participant images. For this article, we built a customized brain template from participant images and associated deformations using the Advanced Normalization Tools (ANTS) template construction script (Avants et al., 2011) (details can be found on the first author's GitHub repository). This, in turn, generates a spatial gradient of deformation fields that yields a Jacobian matrix at every voxel position. We focused on the determinant of each Jacobian matrix encoding local volume changes of tissue expansion (bigger than one) and contraction (smaller than one) with respect to the template. We follow convention established by (Leow et al., 2007) and log-transformed the Jacobian determinant maps to symmetrize volume changes around zero to be consistent with previous analyses.

Our statistical analysis focuses on gray matter regions. We use the segmentation algorithm FAST (Zhang et al., 2001) to define the gray matter mask in our custom built template brain. We apply this mask to each log-Jacobian determinant map to extract volume changes of gray matter voxels for each subject.

2.4 Cognitive Assessment

We assessed the cognitive status of our participants using the NEPSY-II (Brooks et al., 2009). The NEPSY-II classifies cognitive functions into six domains: Attention and Executive Functions, Language, Memory and Learning, Sensorimotor, Social Perception, and Visuospatial Processing. Test administrators followed standard procedures as outlined in the published product manuals; all cognitive/neuropsychological variables consisted of age-normed scaled scores. Overall 34 NEPSY-II subtests were used in our analysis.

Additionally, we assessed participants cognitive abilities using the Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III; Wechsler (2002)) for girls aged between 4 and 5 years and the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler (2003)) for girls aged between 6 and 16 years.

2.5 Puberty Stages

In addition to the age of each participant we collected information about puberty status in the form of Tanner stages (Marshall and Tanner, 1969) as derived from examination of the participant by the study physicians (LH, TG). Tanner stages in girls classify puberty into five discrete stages, ranging from prepubertal to mature stage of breast development and pubic hair. Considering this type of information is crucial to account for different developmental trajectories of our participants.

2.6 Missing Values

We excluded all subtests and participants that exhibit more than 20% of missing entries. This reduces the original sample size by four participants to 98. We imputed the remaining 3.7% of missing values using predictive mean matching implemented in the `R` package `mice`. We obtained an imputed dataset by averaging over 20 imputed datasets for subsequent statistical analysis. For visualizations of the missing values pattern and the imputation variability see `Interactions.html` in the supplementary material. To confirm that our imputation is unbiased, we repeated the analysis five times with different random seeds.

2.7 Multi-Table Analysis: Sparse Canonical Correlation Analyses

We collected Jacobian determinant maps in a matrix $\mathbf{X}$ with dimensions 98 (number of participants) $\times$ 710,320 (number of gray matter voxels). The second matrix $\mathbf{Y}$ contained cognitive subtest scores and has dimensions 98 (number of participants) $\times$ 27 (number of NEPSY subtests). It is not possible to apply standard CCA directly to this problem because we have more voxels than observations resulting in an underdetermined system of equations. Sparse CCA reduces the solution space to sparse solutions meaning that many noisy coefficients will be set to zero. In addition to making the CCA problem solvable, sCCA also provides us with isolated clusters of coefficients without having to define an additional thresholding parameter. Spatially contiguous clusters occur because the log-Jacobian maps are spatially smooth.

Successful employment of sCCA in other recent neuroimaging studies have been reported in the literature, Avants et al. (2010); Chi et al. (2013); Duda et al. (2013); Avants et al. (2014). The related PLS method is also widely and successfully used in neuroimaging studies, Streissguth et al. (1993); Bookstein (1994); McIntosh et al. (1996); McIntosh and Lobaugh (2004); Krishnan et al. (2011); Lorenzi et al. (2016a,b).

Sparse CCA provides a linear combination of voxels in Jacobian determinant maps that are maximally correlated to a linear combination of cognitive subtests. We call these linear combinations, canonical variables v and w. To find maximally correlating canonical variables, we solve

$$\underset{v,w}{\text{maximize}} \left\{ \text{Corr}(\mathbf{X}v, \mathbf{Y}w) \right\}$$

subject to a constraint on the l_2 norm $\|v\|_2^2 \leq 1$, $\|w\|_2^2 \leq 1$, and a constraint on the l_1 norm $\|v\|_1 \leq c_1$, $\|w\|_1 \leq c_2$ of the canonical variables. We maximize this objective function (Witten et al., 2009) using the `R` package `PMA` and find an optimal sparsity regularization parameter pair c_1 and c_2 through permutation tests (explained in next section).

This optimization problem has an intuitive interpretation; we project $\mathbf{X}$ onto the candidate canonical variable v, and $\mathbf{Y}$ onto w resulting in two new vectors that are elements in a n-dimensional Euclidean space (this dimension is given by the number of observations). In this common n-dimensional space we optimize correlation by reprojecting $\mathbf{X}$ and $\mathbf{Y}$ onto slightly modified canonical variable candidates until we find the maximum.

The signs of canonical variables v and w are not identifiable. More precisely, the following solutions (v, w) and $(-v, -w)$ are equivalent. We account for this by constructing score functions and test statistics that are invariant to sign flips.

2.7.1 Group Differences

The main goal of this study is to find differences between TS and control girls in terms of brain-cognition associations. To accomplish this goal, we designed a new nonparametric permutation-based test procedure.

We have three different sources of information: First, morphometry measurements in the form of log-Jacobian maps stored as images. Second, the cognitive tests stored in a data table. Third, the group label factor with two levels encoding whether a participant belongs to the TS or healthy control group. To make the data amenable for analysis with sCCA, we split the data into an imaging and non-imaging table by joining group labels and cognitive tests, denoted by $\tilde{\mathbf{Y}}$. In this form, we can pass it to sCCA and compute maximally correlated canonical variables. Cognitive tests will have large coefficients in the canonical variables if they correlate with morphometry measurements. In addition, large group label coefficients provide evidence of differential interactions between morphometry measurements and cognitive tests.

2.7.2 Computing Canonical Variables

In sCCA, we need to define one regularization parameter per data table c_1 and c_2. We find optimal regularization parameters for the TS group by the permutation-approach implemented in `R` package `PMA`, which computes the null distribution of $\text{Corr}(\mathbf{X}v, \mathbf{Y}w)$ and selects the most significant model according to the highest z-statistic. We choose not to regularize the cognitive subtest canonical variables because we have more observations than subtests and thus all cognitive coefficients can be estimated. The regularization strength for the morphometry measurements will set the detectable cluster size. A strong regularization will produce small clusters, whereas weak regularization large clusters. In case of strong correlations this optimization will choose a weak regularization and produce large voxel clusters, whereas with weaker correlations it will choose a strong regularization and produce small voxel clusters.

Before going into details of the randomization test with permutations, we define permutations π to be in the group of permutations $\pi \in G$, an element from that group describes a reordering of the rows of $\tilde{\mathbf{Y}}$ by sampling rows without replacement. We further define that the first permutation π_1 is the identity permutation representing the unpermuted case, and the other $1-B$ per-

mutations are actual permutations $\pi_2, \ldots, \pi_b, \ldots, \pi_B$ to create the null distribution.

Our procedure starts by computing the unpermuted observed case $(\mathbf{X}, \pi_1(\tilde{\mathbf{Y}}))$ by first optimizing regularization parameters c_1 and c_2 and then solving

$$\left(v^{(\pi_1)}, w^{(\pi_1)}\right) = \underset{v,w}{\text{maximize}} \left\{ \text{Corr}(\mathbf{X}v, \pi_1(\tilde{\mathbf{Y}})w) \right\}.$$

Next, we compute sCCA's on the permuted datasets $(\mathbf{X}, \pi_b(\tilde{\mathbf{Y}}))$ by optimizing regularization parameters c_1 and c_2 and then solving

$$\left(v^{(\pi_b)}, w^{(\pi_b)}\right) = \underset{v,w}{\text{maximize}} \left\{ \text{Corr}(\mathbf{X}v, \pi_b(\tilde{\mathbf{Y}})w) \right\}.$$

To assign significance levels to voxels, cognitive tests, and group labels, we need to define test statistics. A good test statistic is problem-specific. In our case, we have to design a separate test statistic for morphometry measurements $\mathbf{X}$, cognition tests, and group labels $\tilde{\mathbf{Y}}$ incorporating the different nature of the data. For $\mathbf{X}$, we include the spatial nature of log-Jacobian maps by defining cluster size test statistics, and for $\tilde{\mathbf{Y}}$, we consider the absolute values of the canonical variables directly.

2.7.3 Test Statistic for Morphometry Measurements

On the imaging side, we design a test statistic that captures the spatial dependency structure of log-Jacobian maps. We begin by mapping the vector $v^{(\pi_b)}$ to its corresponding spatial image. We then threshold voxels that have non-zero coefficients resulting in a binary image, and group neighboring non-zero voxels into clusters. This will result in a list of voxel clusters ordered from largest to smallest $\Omega_1, \ldots, \Omega_k$.

To increase power to detect larger clusters, we only consider clusters of size at least half the smallest region in the Harvard-Oxford cortical atlas (half region because this atlas combines left and right regions into one label). In total we have K test statistics per image. We define the kth cluster size test statistics to be

$$T_{\mathbf{X},k}(v) = \sum_{i \in \Omega_k} |v_i|.$$

We can now compute p-values by counting the number of cluster sizes that are at least as large as the observed size divided by the total number of permutations

$$p\text{-value}_k = \frac{1}{B} \sum_{b=1,\ldots,B} I\left\{ T_{\mathbf{X},k}\left(v^{(\pi_1)}\right) \leq T_{\mathbf{X},k}\left(v^{(\pi_b)}\right) \right\}$$

for all $k = 1, \ldots, K$ clusters.

Finally, to account for multiple testing of the K cluster sizes, we adjust p-values using the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995).

The size of the clusters depend on the location within the gray matter. Some regions are anatomically larger. This leads to larger cluster size test statistics in these regions. To account for this location dependency in our test statistic, we compare the cluster sizes according to their rank, e.g. the unpermuted largest cluster with the permuted largest clusters.

2.7.4 Test Statistic for Group Labels and Cognitive Tests

We define the test statistic for the cognitive canonical variables to be the magnitude of studentized absolute values of the coefficients. We denote $\mu_{|w_l|}$ as the sample mean and $\sigma_{|w_l|}$ as the sample standard deviation of the lth element computed across all permutations. We denote the statistic of the lth element of the coefficient vector as

$$T_{\tilde{\mathbf{Y}},l}(w) = \left| \frac{|w_l| - \mu_{|w_l|}}{\sigma_{|w_l|}} \right|.$$

This test statistic is invariant to sign flips to account for non-identifiability inherent in sCCA.

By counting the number of times that the permuted test statistic is at least as large as the unpermuted statistic

$$p\text{-value}_l = \frac{1}{B} \sum_{b=1,\ldots,B} I\left\{ T_{\tilde{\mathbf{Y}},l}\left(w^{(\pi_1)}\right) \leq T_{\tilde{\mathbf{Y}},l}\left(w^{(\pi_b)}\right) \right\}$$

we obtain the p-values for all subtests $l = 1, \ldots, 27$ on the same $b = 1, \ldots, B$ permutations. Since we obtain 27 p-values we again adjust using the BH procedure.

2.8 Relationship to Suprathreshold Cluster Size Test

We use cluster sizes as the test statistics to incorporate spatial dependencies between voxels. This is inspired by the suprathreshold cluster size permutation test commonly used in neuroimaging studies (Poline and Mazoyer, 1993; Roland et al., 1993; Nichols and Holmes, 2002). The usual procedure is to perform separate two-sample tests at each voxel position using the t-test or a rank-based test such as the Wilcoxon two-sample rank test, and then to adjust for multiple testing using the suprathreshold test. The simplest version of the suprathreshold test comprises of two steps: first threshold the voxelwise p-value image by a fixed primary value, e.g. $\alpha_I = 0.001$, and second, identify

the significantly connected spatial contiguous clusters at this threshold using the secondary significant level, e.g. $\alpha_{II} = 0.05$. By regulating the primary threshold one can choose to detect few large clusters (large α_I) or many small clusters (small α_I). The choice of the primary threshold defines the "scale" of the analysis.

In our procedure, the first stage corresponds to finding connected components in the zero/non-zero coefficient binary image, and the second stage corresponds to counting cluster sizes in the usual way. Our procedure thus differs in the way we handle the first stage, adapting it to multi-table methods.

2.9 Single-Table Analysis: Using Two Separate Principal Component Analyses

We compare our proposed permutation-based multi-table analysis to single-table analyses. First, we perform two separate Principal Component Analysis (PCA) on the morphometry matrix $\mathbf{X}$ and the cognitive test matrix $\mathbf{Y}$. Then, we reduce the dimensionality by identifying the PCs that explain a large amount of variability and separate TS from healthy control participants. Ideally, the major variability in the data is due to the group difference, and we will only need to keep one PC per matrix. We then represent morphometry and cognition with their respective PCs. To test whether there is an interaction between morphometry and cognition, we can use a linear model with an interaction term. We define the cognition PC as the response variable, the morphometry PC as an explanatory variable (including an intercept), and add and interaction term diagnosis $\times$ morphometry PC. To assess if the interaction terms explains additional variability, we can use ANOVA to compare the model "with interaction term" to the model "without the interaction term", or we can test if the coefficient representing the interaction term is zero.

3 Results

After data removal and imputation of missing values in the cognitive subtests, our sample size was 53 TS and 45 control participants. The two groups are age matched. There is no evidence that the two age distributions are shifted (p-value from two-sided Wilcoxon rank sum test is $p = 0.9$), and no evidence for a difference in scale (Mood Two-Sample Test of Scale is $p = 0.3$ and Ansari-Bradley Test is $p = 0.4$). Furthermore, a χ^2 test of homogeneity for Tanner stages for the first three stages (not enough observations in stage four and five, so we merged stages three, four, and five) yielded a p-value of 0.4, thus favoring the null hypothesis of equal distribution of participants for the two groups. The overall Full Scale IQ for TS girls is 94 ± 14 and for control girls is 113 ± 12, clearly different between the two groups.

A sampling bias might exist because of our recruitment methods (conferences, online advertisements, local print media, and parent networks). Comparable differences in Full Scale IQ between children with TS and healthy controls have been reported in the literature: Rovet (1993) reported a mean difference of 13.4 in a study comparing 67 children with TS and 27 healthy controls, and Mazzocco (1998) reported a mean difference of 14.9 in a study comparing 29 children with TS and 16 sibling healthy controls.

All reported p-values are adjusted using the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995) to control the False Discovery Rate (FDR). If not explicitly stated, we declare significance below an FDR of 5%.

3.1 Multi-Table Analysis

3.1.1 Neuroanatomical Canonical Variable Plots

The permutation test for the morphometry measurements yielded two significant clusters (Figure 1 top). The optimal regularization parameter was 0.1. The clusters are larger in the TS than in the control group (Figure 1 bottom). The clusters partially overlap with the intracalcarine, precuneous, cuneal, and supracalcarine cortex in the left and right occipital lobe.

3.1.2 Cognitive Canonical Variable Plots

The permutation test for the cognitive scores found 22 tests and the diagnosis as different (Figure 2). Tests come from all five NEPSY domains. All marginal distributions of each test score are lower or similar in TS compared to the healthy control group (Figure 3).

3.1.3 Joint Neuroanatomy-Cognition Plots

Our new procedure provides a filtering tool to select differential correlations between a priori unknown brain regions and cognitive tests. A non-zero effect of the diagnosis reflects an interaction between brain regions and cognitive test scores. We can explore this correlation structure with scatter plots between cluster sizes and cognitive tests (Figures 4 and 5). We compute cluster sizes for each participant by summing over each participant's Jacobian determinant map within a cluster.

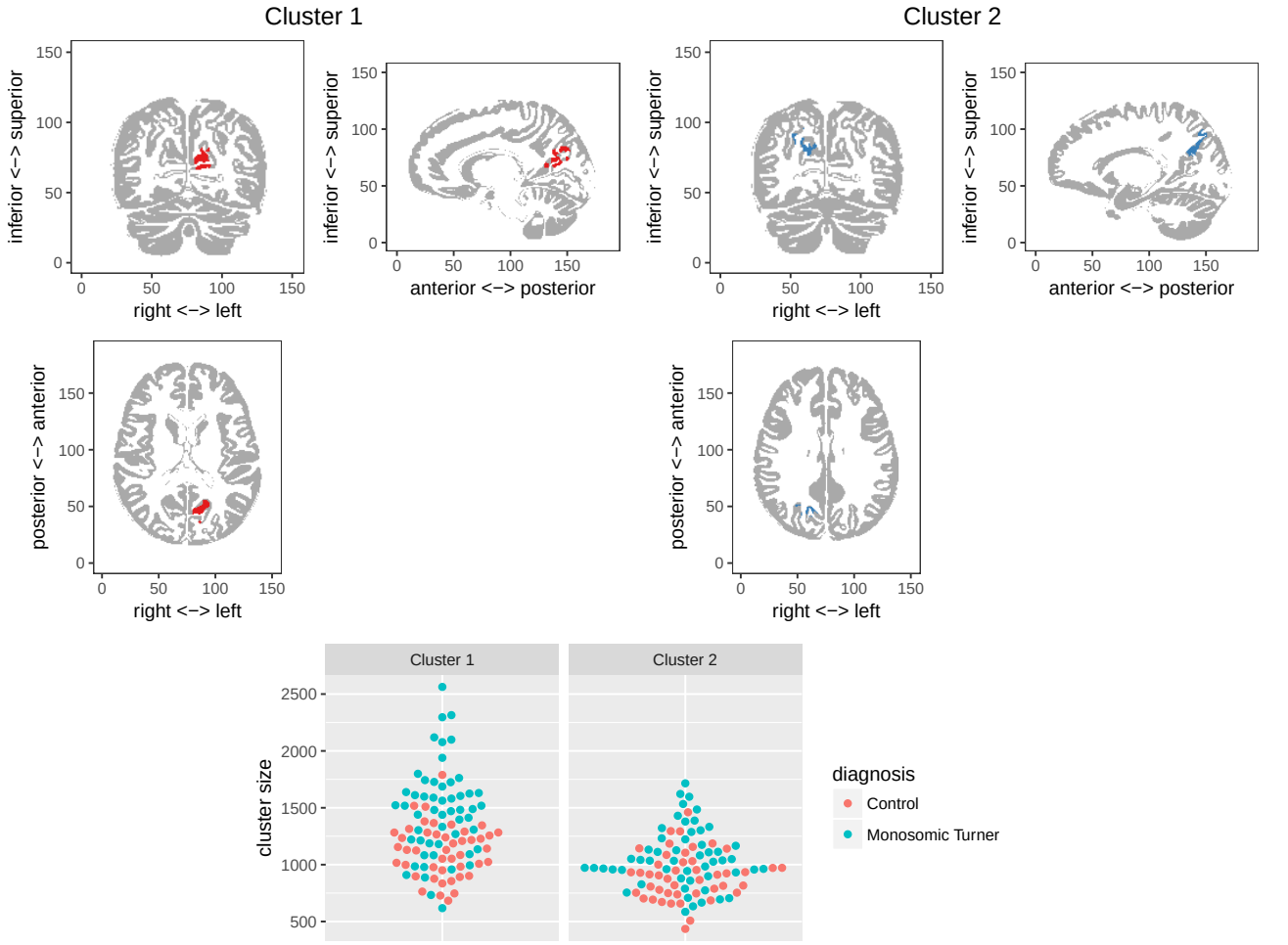

Fig. 1 Top: Two significant voxel clusters from the multi-table differential correlation analysis. Bottom: Actual observed cluster sizes for each participant in mm^3. One point represents one participant. The points are arrange to avoid overlaps. If overlap occurs, points are horizontally shifted by a small amount.

In both Cluster 1 and Cluster 2, we observe that TS participants have lower scores compared to their healthy controls. This confirms our findings from the marginal distribution plots in Figure 3. Visually, we can compare the two slopes per facet. For example, in Cluster 1, Auditory Attention and Auditory Attention Correction show a decline in test score with increasing cluster size in the healthy control group and an increase in test scores in the TS group. In Cluster 2, Arrows and Picture Puzzles show an increase in test score with increasing cluster size in the healthy control group, whereas a decline in test score in the TS group.

We avoid assigning significance levels to the slopes in the scatter plots because the cluster selection and cognitive test selection was done on the same data. We think of these plots as an exploratory way to interpret the multivariate results.

3.2 Single-Table Analysis

In this section, we compare our multi-table analysis with two single-table analyses. When doing PCA on the cognition test and morphometry tables separately, we notice that the first PC separates between control and TS groups (Figures 6 and 7). We take advantage of this by reducing our original data to two vectors. To test for interaction, we compare two linear models. Model one, has as the response variable morphometry PC1, and explanatory variables, cognition PC1 and an intercept. Model two, has the same terms plus an additional cognition PC1 $\times$ diagnosis (TS or control) interaction term. We found no evidence that the interaction term explains more variance as the ANOVA comparison of the two model fits yielded a p-value of 0.6. We also tested the interaction term directly by testing the null hypothesis that its coefficient is zero. This test yielded

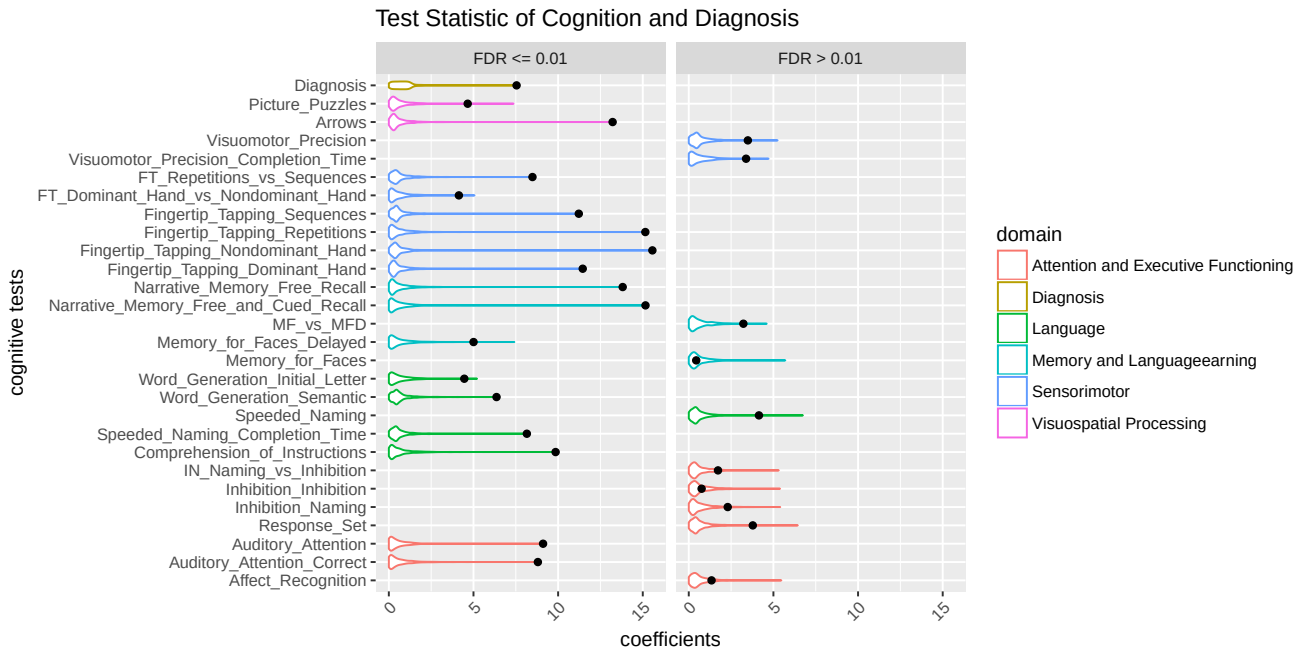

Fig. 2 Violin plots represent the null distribution of the test statistic. Black circles represent the observed test statistic.

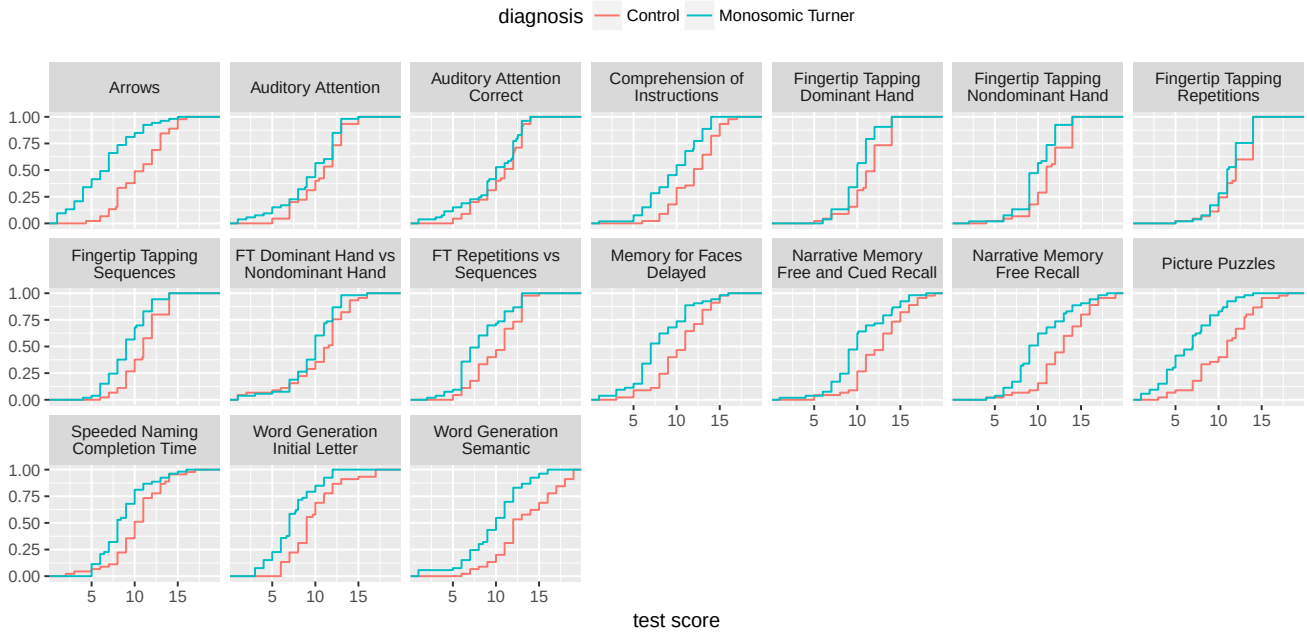

Fig. 3 Empirical cumulative distribution function of selected tests.

a p-value of 0.7 confirming the previous test of no evidence for a global brain-cognition interaction.

4 Discussion

In this study we sought to investigate brain-behavior associations of gray matter volume and cognitive abilities in a large cohort of females with Turner syndrome. Overall we found two voxel clusters and a wide range of cognitive subtests that show an aberrant association in TS compared to the control group.

4.1 Joint Neuroanatomy-Cognition Interpretation

Most functional imaging studies on TS report aberrant activation in frontal and parietal lobes (Molko et al.,

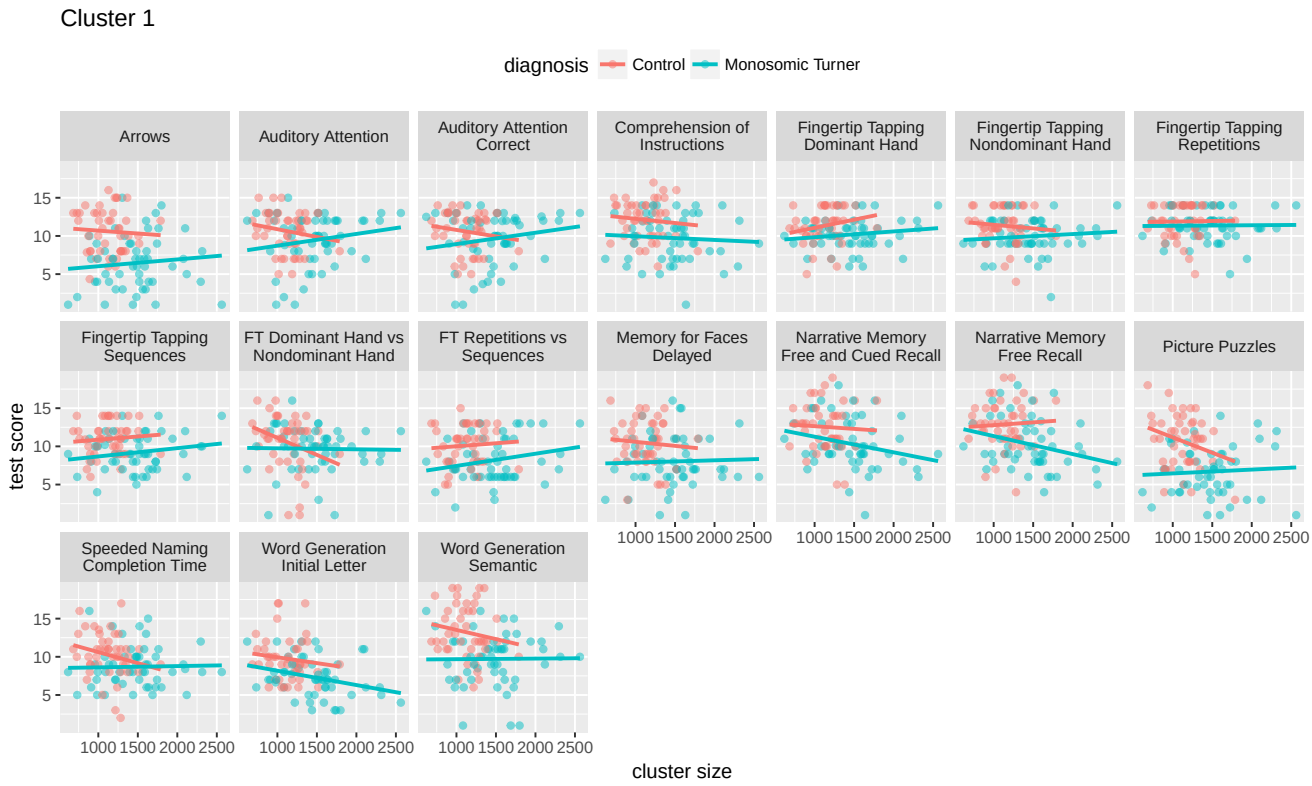

Fig. 4 Scatter plot of cluster size of Cluster 1 versus cognitive test scores. Points are participants. Lines are linear model fits.

2003; Kesler et al., 2004; Hart et al., 2006; Bray et al., 2011). We observe putative, disorder-relevant findings in the occipital cortex. We did not find aberrations in frontal and parietal cortices. Discrepancies between our and previous findings are expected considering our focus on brain-cognition as opposed to brain-only aberrations. Note that these results do not contradict but extend previous findings.

The heterogeneous cognitive profile in TS is believed to span over many cognitive domains (Kesler, 2007; Hong et al., 2009). In particular, the arrow subtest showed significant between-group differences, a finding which is consistent with Green et al. (2014) who found decreased arrows subtest scores in TS. The others subtests are to the best of our knowledge potential new discoveries and have not been investigated in TS.

4.2 Comparison of Multi-Table and Single-Table Analysis

Our single-table analysis is a global analysis describing cognition and morphometry with one variable each, whereas our multi-table analysis is a local analysis on individual voxel clusters and test scores. One can think of the two analyses in terms of a global test, such as testing for equal mean across multiple groups without identifying which group drives the difference, and a local test, such as two-sample pairwise tests. Analog to hypothesis testing, not rejecting the global null hypothesis does not imply that the local tests will not be rejected. It is more helpful to consider it as providing two levels of analysis. Best practice would be to perform both analyses and to form an opinion considering both results.

5 Conclusion

In this article, we present a combined statistical analysis of neuroanatomy and cognition in girls with TS and healthy controls. The estimated canonical variables show aberrant associations located in the occipital lobe and a wide range of cognitive tests.

Sparse CCA in combination with a cluster size test statistic yields a meaningful dimension reduction from thousands of voxels to a few voxel clusters. We showed that multi-table analysis can be used in combination with permutation tests to assign p-values to sets of meaningful coefficients. This is a very promising path for complex neuropsychiatric disease research given that pairwise association tests cannot capture the multivariate nature of the data. For neuropsychiatric diseases we expect to find multi-node pathways and thus our

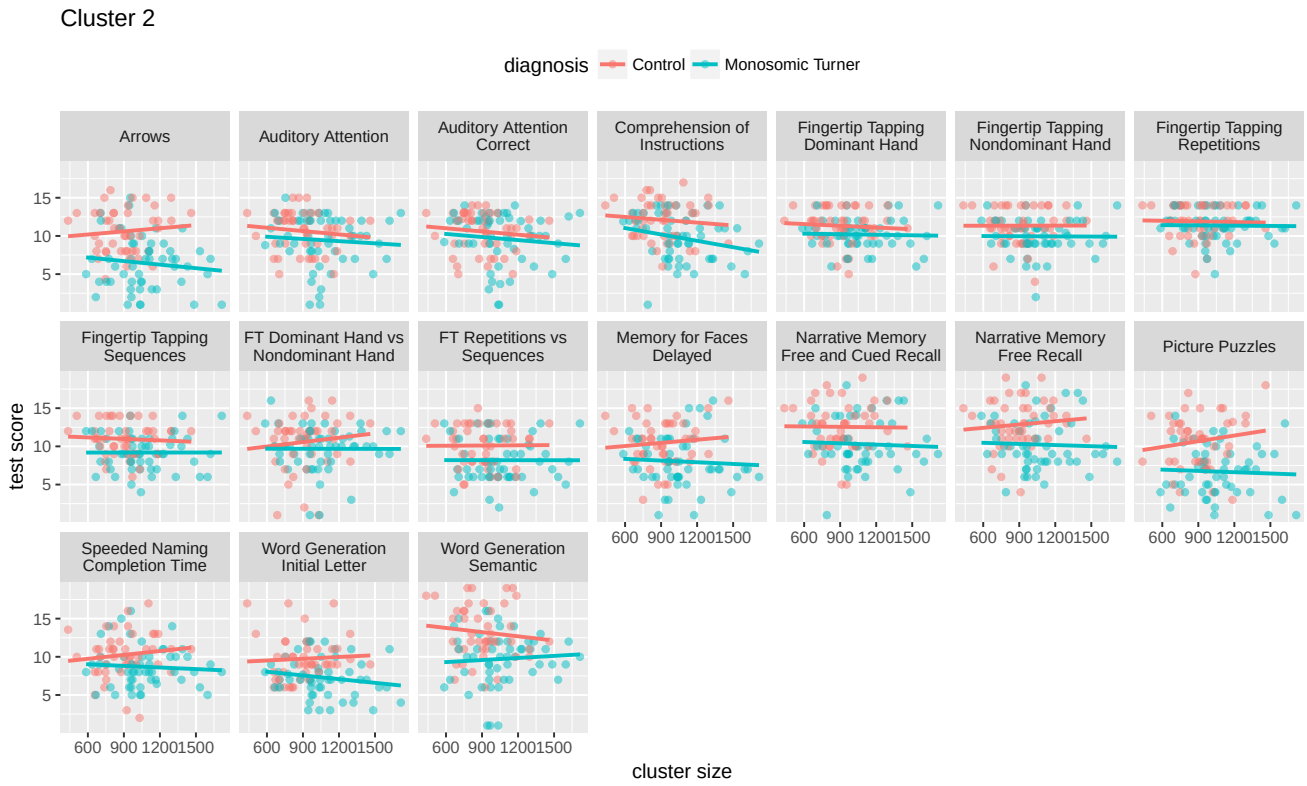

Fig. 5 Scatter plot of cluster size of Cluster 2 versus cognitive test scores. Points are participants. Lines are linear model fits.

presented multi-table method is a valid candidate approach. However, there is no "free lunch", and the price paid for increase in power is greater ambiguity in the interpretation of the results.

As an alternative to sCCA, one can preprocess the data matrices $\mathbf{X}$ and $\mathbf{Y}$ using PCA and reduce dimensionality so that we have fewer components than observations and apply standard CCA (Smith et al., 2015). This approach exploits the strong spatial correlation structure in the data and allows one to reduce the dimensionality. However it involves two choices: first, one has to choose the right number of components to retain, and second, one has to select the top most important features and ignore small but still nonzero features. By assuming sparsity we can avoid both subjective choices and select appropriate regularizations using data driven permutation tests.

Another option would be to split the data into two pairs of data tables, compute two separate sCCA's, and construct a test statistic that contrasts canonical variables. Our experiments with this approach showed that it is challenging to match canonical variables from separate sCCA optimizations into one consistent analysis. However, if such a matching could be found it would provide a more direct separation between groups.

We imputed missing values in the cognitive subtests and computed the average dataset over all imputed datasets. As it is well known, this underestimates the variability. To test how sensitive our results are to imputation, we repeated the analysis five times with different random seeds. The repeated analyses selected the same clusters and mostly the same cognitive tests. In the manuscript, we show the most conservative result. In the supplementary materials, we report all five analyses.

The construction of Jacobian determinant maps involves several processing steps (ACPC alignment, resampling, bias field correction, skull stripping, and registration). Each step can impact our final statistical conclusions. In future work we aim to better quantify uncertainty in neuroimaging pipelines and how uncertainty can be propagated to the analysis level.

Our goal here was not predictive, we have not assigned a response status to either the behavioral tests or the anatomical measurements; we have only looked for TS-aberrant associations.

Interactions between neuroanatomy and cognition can change with age and throughout developmental stages. Such changes can be unrelated to TS. We are currently collecting longitudinal data to investigate how associations change over time.

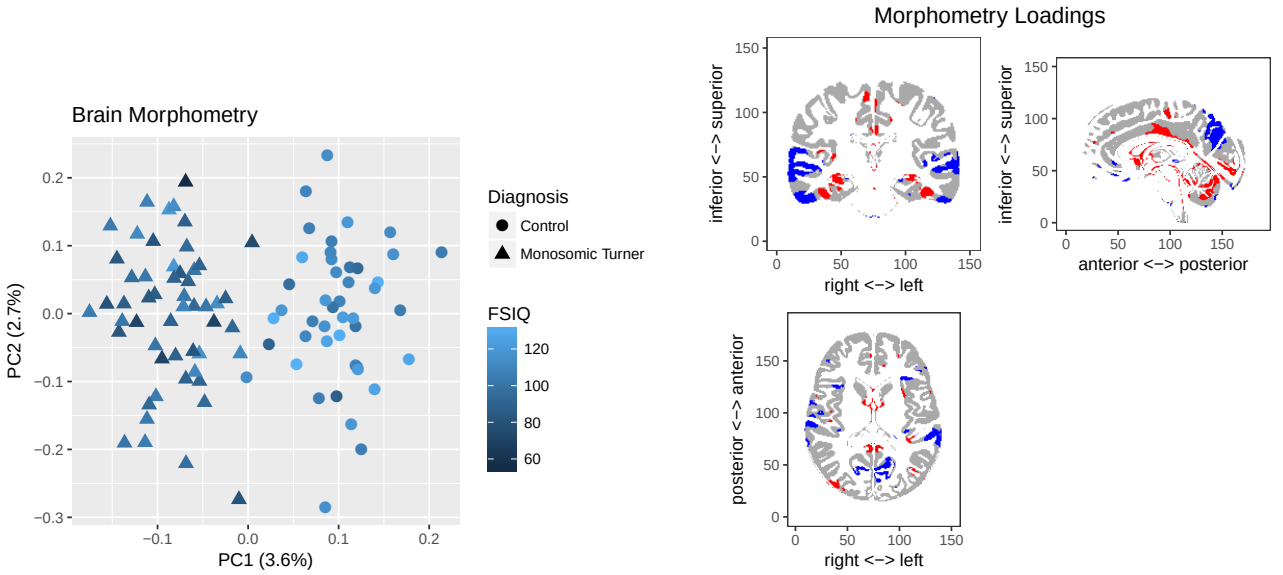

Fig. 6 PCA of brain morphometry measurements. Left: Each shapes is one participant projected onto PC1 and PC2. Right: In red, the 10% largest (positive) coefficients in the PC1 loadings. In blue, the 10% smallest (negative) coefficients in the PC1 loadings. Blue brain regions are larger in the TS group and smaller in the healthy control group. Red brain regions are larger in the healthy control group and smaller in the TS group.

It is critical to extend our understanding of brain-cognition associations to advance the field of clinical neuropsychiatry with the hope of designing new targeted interventions for disorders and diseases such as TS.

Information Sharing Statement

We implemented the `R` package `braincog` available on GitHub.[2] Our complete analysis workflow is in one `Rmd` file called `Interactions.Rmd` also available on GitHub.[3] All results and plots can be completely reproduced by running:

```
R -e "rmarkdown::render('Interactions.Rmd')"
```

This command will produce an `Interactions.html` report. The computation time for 1000 permutations on a regular laptop with two CPU cores is about 5 days. We recommend using a computing cluster reducing computation time to few hours. Our `R` pacakge `braincog` is `slurm` cluster compatible.

The image registration steps including template and Jacobian maps construction are available as batch scripts and need to be run prior to the `Rmd` files.

Processed data are available upon request.

[2] `https://github.com/ChristofSeiler/braincog`

[3] `https://github.com/ChristofSeiler/braincog_manuscript`

Acknowledgments

The Turner Syndrome Society and the Turner Syndrome Foundation made this work possible. The authors would like to sincerely thank all of the families who kindly volunteered to participate.

Christof Seiler was supported by two postdoctoral fellowships from the Swiss National Science Foundation and a travel grant from the France-Stanford Center for Interdisciplinary Studies. Tamar Green was supported by a grant from the Gazit-Globe Post-Doctoral Fellowship Award. Allan L. Reiss is supported by grants from the NICHD (HD049653), NIMH (MH099630), and the Sharon Levine Foundation. Dr. Reiss is an unpaid medical advisor for the Turner Syndrome Society and Turner Syndrome Foundation. The funding sources mentioned above had no role in the study design; in the collection, analysis and interpretation of the data. Susan Holmes is supported by NICHD (HD049653).

We would like to thank two anonymous reviewers for their helpful input that greatly contributed to improved clarity and quality of this manuscript.

References

V. P. Sybert, E. McCauley, Turner's Syndrome, New England Journal of Medicine 351 (12) (2004) 1227–1238.

C. H. Gravholt, Clinical practice in Turner syndrome, Nature Reviews Endocrinology 1 (1) (2005) 41–52.

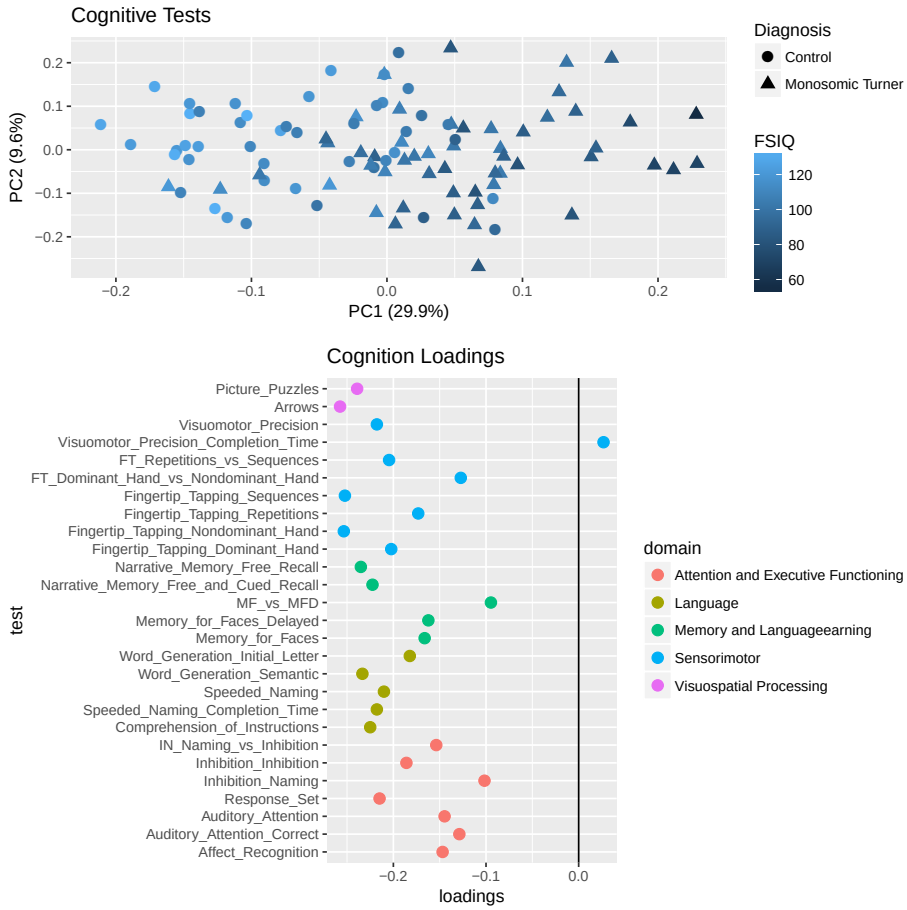

Fig. 7 PCA of cognition tests. Top: Each shapes is one participant projected onto PC1 and PC2. Bottom: Loadings of PC1. All cognitive test scores except Visuomotor Precision Completion Time are higher in the healthy control group.

D. S. Hong, F. Hoeft, M. J. Marzelli, J.-F. Lepage, D. Roeltgen, J. Ross, A. L. Reiss, Influence of the X-Chromosome on Neuroanatomy: Evidence from Turner and Klinefelter Syndromes, The Journal of Neuroscience 34 (10) (2014) 3509–3516.

D. Hong, J. Scaletta Kent, S. Kesler, Cognitive profile of Turner syndrome, Developmental Disabilities Research Reviews 15 (4) (2009) 270–278.

W. E. Brown, S. R. Kesler, S. Eliez, I. S. Warsofsky, M. Haberecht, A. L. Reiss, A volumetric study of parietal lobe subregions in Turner syndrome, Developmental Medicine & Child Neurology 46 (9) (2004) 607–609.

H. Hotelling, Relations between two sets of variates, Biometrika 28 (3/4) (1936) 321–377.

L. R. Tucker, An inter-battery method of factor analysis, Psychometrika 23 (1958) 111–136, ISSN 0033-3123.

H. Wold, Estimation of Principal Components and Related Models by Iterative Least squares, Academic Press, 391–420, 1966.

C. Fornell, F. L. Bookstein, Two structural equation models: LISREL and PLS applied to consumer exit-voice theory, Journal of Marketing research (1982) 440–452.

A. J. Izenman, Reduced-rank regression for the multivariate linear model, Journal of Multivariate Analysis 5 (2) (1975) 248–264.

E. Parkhomenko, D. Tritchler, J. Beyene, Genome-Wide Sparse Canonical Correlation of Gene Expression with Genotypes, in: BMC Proceedings, vol. 1, S119, 2007.

S. Waaijenborg, P. C. Verselewel de Witt Hamer, A. H. Zwinderman, Quantifying the Association between Gene Expressions and DNA-markers by Penalized Canonical Correlation Analysis, Statistical Applications in Genetics and Molecular Biology 7 (1).

E. Parkhomenko, D. Tritchler, J. Beyene, Sparse Canonical Correlation Analysis with Application to Genomic Data Integration, Statistical Applications in Genetics and Molecular Biology 8 (1) (2009) 1–34.

K.-A. Lê Cao, P. G. Martin, C. Robert-Granié, P. Besse, Sparse Canonical Methods for Biological Data In-

tegration: Application to a Cross-Platform Study, BMC Bioinformatics 10 (1) (2009) 34.

D. M. Witten, R. Tibshirani, T. Hastie, A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis, Biostatistics 10 (3) (2009) 515–534.

S. M. Smith, Fast robust automated brain extraction, Human Brain Mapping 17 (3) (2002) 143–155.

N. Tustison, B. Avants, P. Cook, Y. Zheng, A. Egan, P. Yushkevich, J. Gee, N4ITK: Improved N3 Bias Correction, IEEE Transactions on Medical Imaging 29 (6) (2010) 1310–1320.

C. Davatzikos, M. Vaillant, S. M. Resnick, J. L. Prince, S. Letovsky, R. N. Bryan, A computerized approach for morphological analysis of the corpus callosum, Journal of Computer Assisted Tomography 20 (1) (1996) 88–97.

P. A. Freeborough, N. C. Fox, Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images, Journal of Computer Assisted Tomography 22 (5) (1998) 838–843.

J. C. Gee, R. K. Bajcsy, Elastic matching: Continuum mechanical and probabilistic analysis, in: A. W. Toga (Ed.), Brain Warping, Academic Press, 1998.

M. Chung, K. Worsley, T. Paus, C. Cherif, D. Collins, J. Giedd, J. Rapoport, A. Evans, A Unified Statistical Approach to Deformation-Based Morphometry, NeuroImage 14 (3) (2001) 595–606.

B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, J. C. Gee, A reproducible evaluation of ANTs similarity metric performance in brain image registration, NeuroImage 54 (3) (2011) 2033–2044.

A. Leow, I. Yanovsky, M.-C. Chiang, A. Lee, A. Klunder, A. Lu, J. Becker, S. Davis, A. Toga, P. Thompson, Statistical Properties of Jacobian Maps and the Realization of Unbiased Large-Deformation Nonlinear Image Registration, IEEE Transactions on Medical Imaging 26 (6) (2007) 822–832.

Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm, IEEE Transactions on Medical Imaging 20 (1) (2001) 45–57.

B. L. Brooks, E. M. Sherman, E. Strauss, NEPSY-II: A Developmental Neuropsychological Assessment, Second Edition, Child Neuropsychology 16 (1) (2009) 80–101.

D. Wechsler, Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III), San Antonio, TX: The Psychological Corporation, 2002.

D. Wechsler, Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV), San Antonio, TX: The Psychological Corporation, 2003.

W. A. Marshall, J. M. Tanner, Variations in pattern of pubertal changes in girls, Archives of Disease in Childhood 44 (235) (1969) 291.

B. B. Avants, P. A. Cook, L. Ungar, J. C. Gee, M. Grossman, Dementia induces correlated reductions in white matter integrity and cortical thickness: A multivariate neuroimaging study with sparse canonical correlation analysis, NeuroImage 50 (3) (2010) 1004–1016.

E. Chi, G. Allen, H. Zhou, O. Kohannim, K. Lange, P. Thompson, Imaging genetics via sparse canonical correlation analysis, in: International Symposium on Biomedical Imaging – ISBI, 740–743, 2013.

J. T. Duda, J. A. Detre, J. Kim, J. C. Gee, B. B. Avants, Fusing Functional Signals by Sparse Canonical Correlation Analysis Improves Network Reproducibility, in: K. Mori, I. Sakuma, Y. Sato, C. Barillot, N. Navab (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI, vol. 8151 of *Lecture Notes in Computer Science*, Springer, 635–642, 2013.

B. B. Avants, D. J. Libon, K. Rascovsky, A. Boller, C. T. McMillan, L. Massimo, H. B. Coslett, A. Chatterjee, R. G. Gross, M. Grossman, Sparse canonical correlation analysis relates network-level atrophy to multivariate cognitive measures in a neurodegenerative population, NeuroImage 84 (2014) 698–711.

A. P. Streissguth, F. L. Bookstein, P. D. Sampson, H. M. Barr, The enduring effects of prenatal alcohol exposure on child development: Birth through seven years, a partial least squares solution, The University of Michigan Press, 1993.

F. L. Bookstein, Partial least squares: a dose–response model for measurement in the behavioral and brain sciences, Psycoloquy 5 (23) (1994) 1.

A. McIntosh, F. Bookstein, J. V. Haxby, C. Grady, Spatial pattern analysis of functional brain images using partial least squares, Neuroimage 3 (3) (1996) 143–157.

A. R. McIntosh, N. J. Lobaugh, Partial least squares analysis of neuroimaging data: applications and advances, NeuroImage 23 (2004) S250–S263.

A. Krishnan, L. J. Williams, A. R. McIntosh, H. Abdi, Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review, NeuroImage 56 (2) (2011) 455–475.

M. Lorenzi, B. Gutman, D. P. Hibar, A. Altmann, N. Jahanshad, P. M. Thompson, S. Ourselin, Partial least squares modelling for imaging-genetics in Alzheimer's disease: Plausibility and generalization, in: 13th International Symposium on Biomedical Imaging (ISBI), IEEE, 838–841, 2016a.

M. Lorenzi, I. J. Simpson, A. F. Mendelson, S. B. Vos, M. J. Cardoso, M. Modat, J. M. Schott, S. Ourselin,

Multimodal Image Analysis in Alzheimer's Disease via Statistical Modelling of Non-local Intensity Correlations, Scientific Reports 6 (2016b) 22161.

Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Statist. Soc. Ser. B 57 (1) (1995) 289–300.

J.-B. Poline, B. M. Mazoyer, Analysis of individual positron emission tomography activation maps by detection of high signal-to-noise-ratio pixel clusters, Journal of Cerebral Blood Flow & Metabolism 13 (3) (1993) 425–437.

P. Roland, B. Levin, R. Kawashima, S. Åkerman, Three-dimensional analysis of clustered voxels in 15O-butanol brain activation images, Human Brain Mapping 1 (1) (1993) 3–19.

T. E. Nichols, A. P. Holmes, Nonparametric permutation tests for functional neuroimaging: a primer with examples, Human brain mapping 15 (1) (2002) 1–25.

J. F. Rovet, The psychoeducational characteristics of children with Turner syndrome, Journal of Learning Disabilities 26 (5) (1993) 333–341.

M. M. Mazzocco, A process approach to describing mathematics difficulties in girls with Turner syndrome, Pediatrics 102 (Supplement 3) (1998) 492–496.

N. Molko, A. Cachia, D. Rivière, J.-F. Mangin, M. Bruandet, D. Le Bihan, L. Cohen, S. Dehaene, Functional and Structural Alterations of the Intraparietal Sulcus in a Developmental Dyscalculia of Genetic Origin, Neuron 40 (4) (2003) 847–858.

S. R. Kesler, M. F. Haberecht, V. Menon, I. S. Warsofsky, J. Dyer-Friedman, E. K. Neely, A. L. Reiss, Functional Neuroanatomy of Spatial Orientation Processing in Turner Syndrome, Cerebral Cortex 14 (2) (2004) 174–180.

S. J. Hart, M. L. Davenport, S. R. Hooper, A. Belger, Visuospatial executive function in Turner syndrome: functional MRI and neurocognitive findings, Brain 129 (5) (2006) 1125–1136.

S. Bray, B. Dunkin, D. S. Hong, A. L. Reiss, Reduced Functional Connectivity during Working Memory in Turner Syndrome, Cerebral Cortex 21 (11) (2011) 2471–2481.

S. R. Kesler, Turner Syndrome, Child and Adolescent Psychiatric Clinics of North America 16 (3) (2007) 709–722.

T. Green, L. C. Chromik, P. K. Mazaika, K. Fierro, M. M. Raman, L. C. Lazzeroni, D. S. Hong, A. L. Reiss, Aberrant parietal cortex developmental trajectories in girls with Turner syndrome and related visualspatial cognitive development: A preliminary study, American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 165 (6) (2014) 531–540.

S. M. Smith, T. E. Nichols, D. Vidaurre, A. M. Winkler, T. E. Behrens, M. F. Glasser, K. Ugurbil, D. M. Barch, D. C. Van Essen, K. L. Miller, A positive-negative mode of population covariation links brain connectivity, demographics and behavior, Nature Neuroscience 18 (11) (2015) 1565–1567.