Tensor-Based Analysis of Genetic Influences on Brain Integrity Using DTI in 100 Twins

Agatha D. Lee[1], Natasha Leporé[1], Caroline Brun[1], Yi-Yu Chou[1], Marina Barysheva[1], Ming-Chang Chiang[1], Sarah K. Madsen[1], Greig I. de Zubicaray[2], Katie L. McMahon[2], Margaret J. Wright[3], Arthur W. Toga[1], and Paul M. Thompson[1]

[1] Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Los Angeles, CA
[2] Functional MRI Laboratory, Centre for Magnetic Resonance, University of Queensland, Brisbane, Australia
[3] Queensland Institute of Medical Research, Brisbane, Australia

Abstract. Information from the full diffusion tensor (DT) was used to compute voxel-wise genetic contributions to brain fiber microstructure. First, we designed a new multivariate intraclass correlation formula in the log-Euclidean framework [1]. We then analyzed used the full multivariate structure of the tensor in a multivariate version of a voxel-wise maximum-likelihood structural equation model (SEM) that computes the variance contributions in the DTs from genetic (A), common environmental (C) and unique environmental (E) factors. Our algorithm was tested on DT images from 25 identical and 25 fraternal twin pairs. After linear and fluid registration to a mean template, we computed the intraclass correlation and Falconer's heritability statistic for several scalar DT-derived measures and for the full multivariate tensors. Covariance matrices were found from the DTs, and inputted into SEM. Analyzing the full DT enhanced the detection of A and C effects. This approach should empower imaging genetics studies that use DTI.

1 Introduction

Imaging genetics is a rapidly growing field that combines mathematical methods from genetics and imaging to identify factors that contribute to variations in brain structure and function. A more mechanistic understanding of brain structure could be achieved if the contributing genes could be identified. Each gene's contribution is expected to be minor, so researchers have first looked for genetically influenced measures in images in the hope of finding specific features or brain regions in which gene effects are the strongest. Twin studies, in particular, examine statistical differences between pairs of identical and fraternal twins; as these twins share all or half their genes on average, structural equation models may be fitted to understand what proportion of the variance in image-derived measures is attributed to genetic differences.

Genetic factors influence various aspects of brain anatomy including cortical thickness [16], regional gray and white matter volumes [4], and univariate measures of fiber integrity derived from DTI [4,10]. Even so, few studies have analyzed genetic influences on signals that are inherently multidimensional, such as diffusion tensors.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 967–974, 2009.

By reducing the full DT to scalar measures of diffusion anisotropy, potentially relevant information is thrown away. DTI measures the multidirectional profile of water diffusion in brain tissue, revealing information on brain architecture and composition. In most DTI studies, a diffusion tensor (DT) is fitted to model at each voxel; its eigenvalues represent the magnitude of diffusion along three orthogonal principal directions - fiber directions may be inferred from the principal eigenvector. The local fractional anisotropy (FA), computed from the eigenvalues, is widely accepted as an index of fiber integrity and is correlated with intellectual performance [5]. An alternative anisotropy measure, the geodesic anisotropy (GA) [3], measures the geodesic distance between tensors on the symmetric positive-definite tensor manifold. In [9], we found that a multivariate statistical analysis of the full diffusion tensor outperformed derived scalar signals in detecting group differences in the blind; others also suggest that effect sizes in group DTI studies can be greater when retaining the full tensor information [17].

Here we present a new approach that compute differences and correlations between DTs using the "Log-Euclidean" framework. Our new intraclass correlation is derived from a Mahalanobis-like distance in the SPD(3) manifold of symmetric positive definite matrices. DTs that correspond to a physically possible diffusion process must lie in the SPD(3) Lie group, and do not form a vector subspace of the vector space of matrices with the usual matrix addition and scalar multiplication. To account for this, we performed all statistical computations in the Log-Euclidean framework [2], which allows standard Euclidean computations on the DT manifold. We used this manifold distance to compute 3D whole-brain maps of heritability using the classical Falconer method [6], and we compared heritability maps from the full DT to maps based on standard univariate measures (including FA, GA and tGA).

We also generalized the A/C/E model used in quantitative genetics, to assess genetic influences on multidimensional signals such as DTI. This method enables us to assess genetic influences in brain architecture by computing variances within and between members of twin pairs, using distances computed on the diffusion tensor manifold. We compared multivariate and scalar DTI-derived measures to determine the most heritable measures. Honing in on heritable measures is typically the first step in identifying specific genes that affect brain structure and function [7].

2 New Multivariate Statistical Formulae for Twin Study

2.1 Intra-class Correlations on the Tensor Manifold

The intraclass correlation [14] between pairs of observations is defined, for univariate quantities such as FA, as:

$$r_{uni} = \frac{MS_{between} - MS_{within}}{MS_{between} + MS_{within}}, \tag{1}$$

where $MS_{between}$ and MS_{within} are mean-square estimates of the between-pair and within-pair variance, respectively.

For tensor-valued data such as the full 3 x 3 DT, equation (1) cannot be used. Instead we use a distance in the tensor manifold measuring the deviation of subject's diffusion tensor T_i from the mean of the entire sample, weighted by the variances:

$$r_{multi} = \frac{\frac{1}{Np}\sum_{i=1}^{Np}((\log T_1^i - \log \bar{T})^T I(\log T_2^i - \log \bar{T}))}{\sigma_x \cdot \sigma_y}, \tag{2}$$

where $\sigma_x = \sqrt{\frac{1}{Np}\sum_{i=1}^{Np}((\log T_1^i - \log \bar{T})^T I(\log T_1^i - \log \bar{T}))}$ and $\sigma_y = \sqrt{\frac{1}{Np}\sum_{i=1}^{Np}((\log T_2^i - \log \bar{T})^T I(\log T_2^i - \log \bar{T}))}$, $\bar{T}$ is the mean of a set of vectors T^i, i=1,...,m, $\boldsymbol{T_1}^i$ and $\boldsymbol{T_2}^i$ represent the log-transformed tensors for each member of the *i*-th twin pair (for simpler notation, we consider the 6 unique components of the log-transformed tensor as belonging to a 6-component Euclidean vector). Here, we reshaped 3x3 diffusion tensor matrix to 6x1 matrix as diffusion tensor is symmetric giving only 6-variate data. We define $\bar{T} = \exp(\frac{1}{M}\sum_{i=1} m \log T^i)$ where M is the total number of subjects, N_p is number of twin pairs and I is the 6x6 identity tensor. I could be omitted, but is included to show that it could be generalized to an inverse covariance matrix expressing the empirical correlation between the 6 unique DT components at each voxel (we reserve this generalization for future work).

2.2 Structural Equation Modeling (A/C/E model) of Variance in DTI Volume

To understand the relative contribution of additive genetic (A) versus shared (C) and unique (E) environmental effects on the variance in a DTI-derived measure, for example the FA or GA, we compute the measure at each voxel in a set of MZ pairs and DZ pairs and measure the covariance between twin 1 and twin 2 at each voxel. These empirically estimated covariance matrices can be computed for any observed variable (Z), and a structural equation model (SEM) can be fitted to the covariances to infer the proportion of the variance attributable to each of A, C and E. Measurement errors or inter-subject registration errors will both be classified as part of the E component of variance. Z for one twin pair may be modeled as:

$$Z = aA + cC + eE\,. \tag{3}$$

where A/C/E are latent variables and *a, c, e* are the weights of each parameter to be determined.

We used a maximum-likelihood estimate (MLE) [12] to estimate the proportion of the voxel-based intersubject variance that is attributable to each of the 3 free model parameters. The 3 variance components combine to create the total observed inter-individual variance, so that $a^2+c^2+e^2=1$. The weights $\theta = (a, c, e)$ are estimated by comparing the covariance matrix implied by the model, $\Sigma(\theta)$, and the sample covariance matrix of the observed variables, S, using maximum-likelihood fitting:

$$F_{ML,\theta} = \log|\Sigma(\theta)| + trace(\Sigma^{-1}(\theta)\mathbf{S}) - \log|\mathbf{S}| - p, \tag{4}$$

where $p = 2$ is the number of observed variables. Under the null hypothesis that Z is multivariate normal (i.e., each of A, C and E are normally distributed), the MLE model follows a χ^2 distribution with $p(p+1) - t$ degrees of freedom, where t is number of model parameters (3 in our case). We used the Broyden-Fletcher-Goldfarb-Shanno method [13] to obtain the minimum $F_{ML,\theta}$. This algorithm is used for both univariate and multivariate data, the only difference being in the computation of the components of the covariance matrix S.

2.3 Illustration of the Covariance Computations on the Manifold

For the univariate data such as FA, the covariance matrix is defined as:

$$Cov(t_1, t_2) = \frac{1}{Np}\sum_{i=1}^{Np}(t_1^i - \bar{t}_1)(t_2^i - \bar{t}_2) \cdot \tag{5}$$

where $\bar{t}$ is the mean of a set t^i i=1,...,m, $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ represent each subject of the twin pair, $\bar{t} = \exp(\frac{1}{M}\sum_{i=1} m \log t^i)$.

In the case of the 3-component vector whose elements are the eigenvalues, or the full DT, there are either 3 or 6 parameters per voxel, all containing potentially useful information for genetic analysis. The covariance matrices can no longer be computed using the previous general formula. Here we propose a new multivariate generalization of Eq. 5 in the Log-Euclidean formalism (Fig 1). In the Log-Euclidean formalism, distances are computed in the tangent space at the origin of the manifold of positive-definite, symmetric matrices. This plane is reached via a matrix logarithm. Hence, for multivariate measures such as 3x3 full DT, which has been reshaped to 6x1 matrix for the computational purpose, the covariance equation becomes:

$$Cov(T_1, T_2) = \frac{1}{Np}\sum_{i=1}^{Np}((\log T_1^i - \log \bar{T})(\log T_2^i - \log \bar{T})^T) \tag{6}$$

This distance is illustrated in **Fig 1**, where each twin's tensors are represented as points on the curved manifold. This distance may be thought of as a bilinear form that takes two tensors as arguments and returns their discrepancy. As noted earlier, the

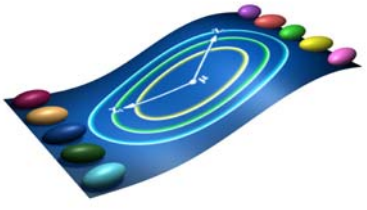

Fig. 1. Here we define a distance between two tensors in the tensor manifold, SPD(3), as $(x_1 - \mu)^T(x_2 - \mu)$, where μ is the mean tensor of all the twins in log-Euclidean space

distances in the manifold could be generalized to take into account the naturally occurring (and perhaps genetically mediated) correlations between the tensor component *i* in twin 1 and tensor component *j* in twin 2. For simplicity, we use the standard metric on Log-Euclidean space, not a statistical metric.

3 Implementation

To test our analysis methods, we acquired 3D structural brain MRI and DTI scans from 100 subjects: 25 identical (MZ) twin pairs (25.1±1.5SD yrs old) and 25 fraternal (DZ) twin pairs (23.1±2.1 yrs) on a 4 Tesla Bruker Medspec MRI scanner with an optimized diffusion tensor sequence. Imaging parameters were: 21 axial slices (5 mm thick), FOV = 23 cm, TR/TE 6090/91.7 ms, 0.5 mm gap, with a 128x100 acquisition matrix. 30 gradient images were collected: three scans with no diffusion sensitization (i.e., T2-weighted images) and 27 diffusion-weighted images for which gradient directions were evenly distributed on the hemisphere [7]. The reconstruction matrix was 128x128, yielding a1.8x1.8 mm^2 in-plane resolution.

3.1 Image Preprocessing and Registration

3D structural MR images were automatically skull-stripped using the Brain Surface Extraction software (BSE) [15] followed by manual editing, and registered via 9-parameter affine transformation to a high-resolution single-subject brain template image, the Colin27 template, using the FLIRT software [8]. 3D structural images were registered to a Mean Deformation Template (MDT) using a 3D fluid registration [11]. Jacobian matrices were obtained from the resulting deformation fields.

DTs were computed from the diffusion-weighted images and smoothed using Log-Euclidean tensor denoising to eliminate singular, negative definite, or rank-deficient tensors, using MedINRIA (http://www.sop.inria.fr/asclepios/software/MedINRIA). To eliminate extracerebral tissues, non-brain tissues were manually deleted from one of the diagonal component images (D_{xx}), yielding a binary brain extraction mask (cerebellum included). Masked images were registered by 12-parameter transformation to the corresponding 3D structural images in the standard template space using *FLIRT* [8].

Transformation parameters from affine and nonlinear registrations were used to rotationally reorient the tensors at each voxel [1] to ensure that the multidimensional tensor orientations remained consistent with the anatomy after image transformation [1,18]. We used two separate algorithms to compute the tensor rotations: the finite strain (FS) and the preservation of principal direction (PPD) algorithms ([1,18])

3.2 Statistical Analysis for Twins

We computed FA, GA and tGA values and the matrix logarithms of the full diffusion tensors for each subject. GA is the manifold equivalent of the FA in the Log-Euclidean framework [2,10]:

$$GA(S) = \sqrt{Trace(\log S - < \log S > I)^2}\,. \tag{7}$$

with $< \log S >= \frac{Trace(\log S)}{3}$.

We used the hyperbolic tangent of GA, *tGA* as in [3], to create maps with a comparable range to the *FA*, i.e., [0,1]. Two sets of voxel-wise covariance matrices for the MZ pairs and DZ pairs were computed for all the univariate and multivariate measures detailed above. For each statistic, we estimated the intraclass correlation, heritability (i.e., variance proportion due to genetic differences among individuals), and computed the best-fitting A/C/E model. For intraclass correlations, we computed permutation-based *p* values to assess their significance.

4 Results and Discussion

To examine the intra-pair variance for each type of twin, Fig 2 shows intraclass correlation (*r*) maps between MZ pairs and DZ pairs for FA (Fig 2a and e), GA (Fig 2b and f), tGA (Fig 2c, g), and the full multivariate DT (Fig 2d, h).

Supporting the validity of these measures, the ventricles, corpus callosum and some anterior temporal regions show higher resemblance among MZs than among DZ twins for all anisotropy measures. Maps based on DT-derived scalars (FA, GA, tGA) are relatively noisy, but maps from the full multivariate tensor show higher SNR than the univariate analyses. Maps of Falconer's heritability estimate for FA, GA, tGA, and the full DT matrix (Fig 2i-l) show that the fiber characteristics of the corpus callosum and some anterior temporal regions are heritable (as confirmed statistically in the p(A/E) maps of Fig 3), consistent with prior studies that only examined FA [5]. Heritability maps confirm the correlation patterns seen in the MZ twins, but adjust for the correlations in DZ twins.

Contributions of factor A and C are shown in Fig 3a-h for the scalar measures (FA, GA, tGA) and the multivariate full DT. Probability maps based on voxel-wise chi-squared statistics confirm the A/C/E model's goodness of fit for all measures. We stress that in A/C/E and other structural equation models, a probability of less than 0.05 indicates that the model is *not* a good fit, so values of p>0.05 are the values of interest (contrary to the usual case in brain maps).

Importantly, maps based on the full diffusion tensor have higher p-values, which means the models are a better fit, and they also appear more spatially coherent. The A and C factors are higher too, suggesting that there is less unmodeled residual variance (which is lumped into the E term).

The p-values from SEM are corrected for the multiple comparisons using FDR. pFDR values for each of the DT-derived measures are pFDR=0.038 for the FA, pFDR=0.0234 for the GA, pFDR=0.038 for the tGA, pFDR=0.0408 for the full DT, indicating that the A/C/E model is a good fit after multiple comparisons correction.

In the *r*-value maps (Fig 2), maps generated using multivariate data (Fig 3d and h) have higher SNR than the univariate data (Fig 3a-c and e-g). Our genetic maps based on the full DT afford better fitting of genetic and environmental models than scalar indices that discard relevant information. These highly heritable phenotypes should facilitate the quest for single-gene polymorphisms that influence the fiber architecture

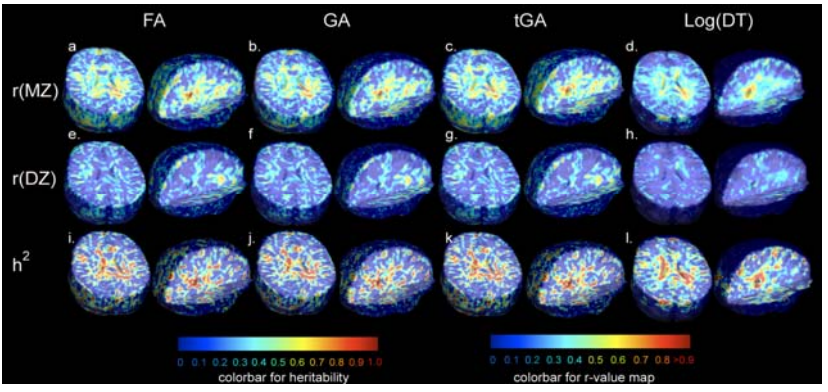

Fig. 2. (**a-h**) show intraclass correlation maps for FA, GA**,** tGA and the full DT

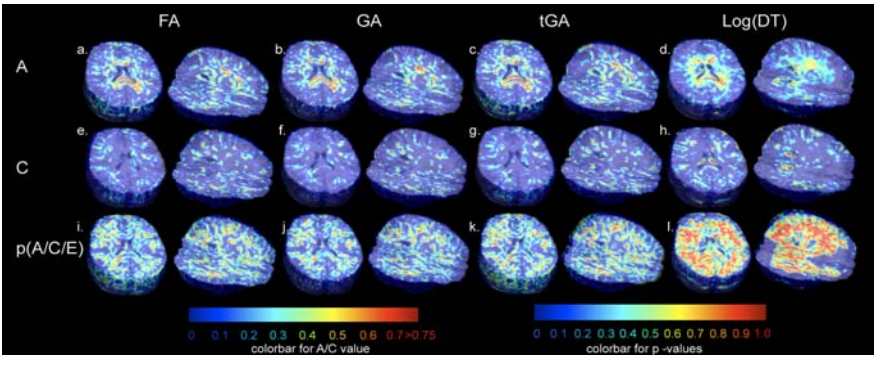

Fig. 3. Maps **(a-d)** show genetic (A) and **(e-h)** shared environmental (C) proportions of variance for various DTI-derived measures: FA, GA**,** tGA and the full DT. The goodness of fit of the A/C/E genetic model is confirmed at voxels where p exceeds 0.05 in panels **(i-l).** In general, the full DT shows best effect sizes for fitting a genetic model; the anisotropy indices show a moderately good fit.

of the living brain, as they hone in on measures and specific brain regions where genetic influences are the most powerfully detected.

The resulting maps show the expected phenotypic patterns of genetic and environmental influences. The ventricles, corpus callosum and anterior temporal regions that develop earlier in life showed strong genetic influences even after multiple comparisons correction. Multivariate analysis also offered improved SNR.

References

1. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. IEEE-TMI 20, 1131–1139 (2001)
2. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Fast and Simple Calculus on Tensors in the Log-Euclidean Framework. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 115–122. Springer, Heidelberg (2005)

3. Batchelor, P., Moakher, M., Atkinson, D., Calamante, F., Connelly, A.: A rigorous framework for diffusion tensor calculus. Magn. Reson. Med. 53, 221–225 (2005)
4. Brun, C.A., Leporé, N., Pennec, X., Chou, Y.-Y., Lee, A.D., Barysheva, M., de Zubicaray, G., Meredith, M., McMahon, K.L., Wright, M.J., Toga, A.W., Thompson, P.M.: A tensor-based morphometry study of genetic influences on brain structure using a new fluid registration method. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 914–921. Springer, Heidelberg (2008)
5. Chiang, M.C., Barysheva, M., Lee, A.D., Madsen, S., Klunder, A.D., Toga, A.W., McMahon, K.L., Zubicaray, G.I., Meredith, M., Wright, M.J., Srivastava, A., Balov, N., Thompson, P.M.: Mapping genetic influences on brain fiber architecture with HARDI. In: ISBI, pp. 871–874 (2008)
6. Falconer, D.S.: Introduction to Quantitative Genetics, 2nd edn. Longman, Harlow (1981)
7. Glahn, D.C., Thompson, P.M., Blangero, J.: Neuroimaging endophenotypes: Strategies for finding genes influencing brain structure and function. Hum. Br. Mapp. 28(6), 488–501 (2007)
8. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. Med. Image. Anal. 5, 143–156 (2001)
9. Lee, A.D., Lepore, N., Lepore, F., Alary, F., Voss, P., Brun, C., Chou, Y.-Y., Barysheva, M., Toga, A.W., Thompson, P.M.: Brain Differences Visualized in the Blind Using Tensor Manifold Statistics and Diffusion Tensor Imaging. In: FBIT 2007, pp. 470–476 (2007)
10. Lee, A.D., Lepore, N., Barysheva, M., Brun, C., Chou, Y.-Y., Madsen, S., McMahon, K.L., Zubicaray, G.I., Wright, M.J., Toga, A.W., Thompson, P.M.: Comparison of fractional and geodesic anisotropy in monozygotic and dizygotic twins from diffusion tensor imaging. In: IEEE ISBI, pp. 943–946 (2008)
11. Lepore, N., Brun, C.A., Chiang, M.-C., Chou, Y.-Y., Dutton, R.A., Hayashi, K.M., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T., Thompson, P.M.: Multivariate statistics of the jacobian matrices in tensor based morphometry and their application to HIV/AIDS. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 191–198. Springer, Heidelberg (2006)
12. Neale, M.C., Boker, S.M., Xie, G., Maes, H.H.: Mx: Statistical modeling (1999)
13. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes in C, Cambridge (1994)
14. Scout, P.E., Fleiss, J.L.: Intraclass correlations: Uses in assessing rater reliability. Psych. Bull. 2, 420–428 (1979)
15. Shattuck, D.W., Leahy, R.M.: BrainSuite: an automated cortical surface identification tool. Medical Image Analysis 8(202), 129–141 (2001)
16. Thompson, P.M., Cannon, T.D., Narr, K.L., et al.: Genetic influences on brain structure. Nat. Neuro. 1253–1258 (2001)
17. Verma, R., Davatzikos, C.: Manifold based analysis of diffusion tensor images using isomaps. Biomedical Imaging: Nano to Macro. 790–793 (2006)
18. Zhang, H., Yushkevich, P.A., Alexander, D.C., Gee, J.C.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. Medical Image Analysis 10, 764–785 (2006)