



High-throughput neuro-imaging informatics

Michael I. Miller^{1,2,3*}, Andreia V. Faria⁴, Kenichi Oishi⁴ and Susumu Mori⁴

¹ Center for Imaging Science, Johns Hopkins Whiting School of Engineering, The Johns Hopkins University, Baltimore, MD, USA

² Institute for Computational Medicine, Johns Hopkins School of Medicine and Whiting School of Engineering, The Johns Hopkins University, Baltimore, MD, USA

³ Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD, USA

⁴ The Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, MD, USA

Edited by:

Xi Cheng, Lieber Institute for Brain Development, USA

Reviewed by:

Davide Chicco, Politecnico di Milano, Italy

*Correspondence:

Michael I. Miller, Center for Imaging Science, The Johns Hopkins University, Clark 301, 3400 N Charles St., Baltimore, MD 21218, USA
e-mail: mim@cis.jhu.edu

This paper describes neuroinformatics technologies at 1 mm anatomical scale based on high-throughput 3D functional and structural imaging technologies of the human brain. The core is an abstract pipeline for converting functional and structural imagery into their high-dimensional neuroinformatic representation index containing $O(1000-10,000)$ discriminating dimensions. The pipeline is based on advanced image analysis coupled to digital knowledge representations in the form of dense atlases of the human brain at gross anatomical scale. We demonstrate the integration of these high-dimensional representations with machine learning methods, which have become the mainstay of other fields of science including genomics as well as social networks. Such high-throughput facilities have the potential to alter the way medical images are stored and utilized in radiological workflows. The neuroinformatics pipeline is used to examine cross-sectional and personalized analyses of neuropsychiatric illnesses in clinical applications as well as longitudinal studies. We demonstrate the use of high-throughput machine learning methods for supporting (i) cross-sectional image analysis to evaluate the health status of individual subjects with respect to the population data, (ii) integration of image and personal medical record non-image information for diagnosis and prognosis.

Keywords: neuro-imaging, neuroinformatics, computational anatomy, functional imaging

INTRODUCTION

Imaging is one of the most powerful medical tools for monitoring human health. In the era of personalized medicine, periodic checkups via whole body imaging, combined with routine medical screening, genetic information, and comparison with population data is expected to be key information for monitoring health status, pathological condition, and therapeutic effect. High-throughput imaging technologies are becoming ubiquitous, driven by the deployment of whole body high resolution MR, CT, and PET imaging devices. While huge personal MR/CT based data records are routinely being collected for cross-sectional and longitudinal examination of the progression of diseases as manifest via tumor growth or atrophic neurodegeneration, currently while this information is stored in the medical PACS, usually only linguistic diagnostic encoding from the physician is stored in the searchable patient record. Such a lack of direct feature representation of the dense structural and functional phenotype precludes its use for systematic medical analysis such as population statistics or cross-modality correlation. Contrast this to what is emerging in high-throughput genomics.

There are several reasons. Clearly, utilizing the information from dense imagery from a longitudinal study, for example, presents daunting challenges. High-resolution whole body CT scans at 0.5 mm resolution for full body coverage would generate gigabytes of data. Visual inspection by a radiologist is overwhelming at the original resolution. Most often the images are down-sampled or low-resolution images are acquired to accommodate the storage and retrieval challenges. Constructing

a parsimonious encoding of the discriminating information presents a fundamental challenge. In high-dimensional spaces such as that represented by the millions of measurements generated by 3D imagers, parsimonious representation of the measurable structural and functional phenotype is essential.

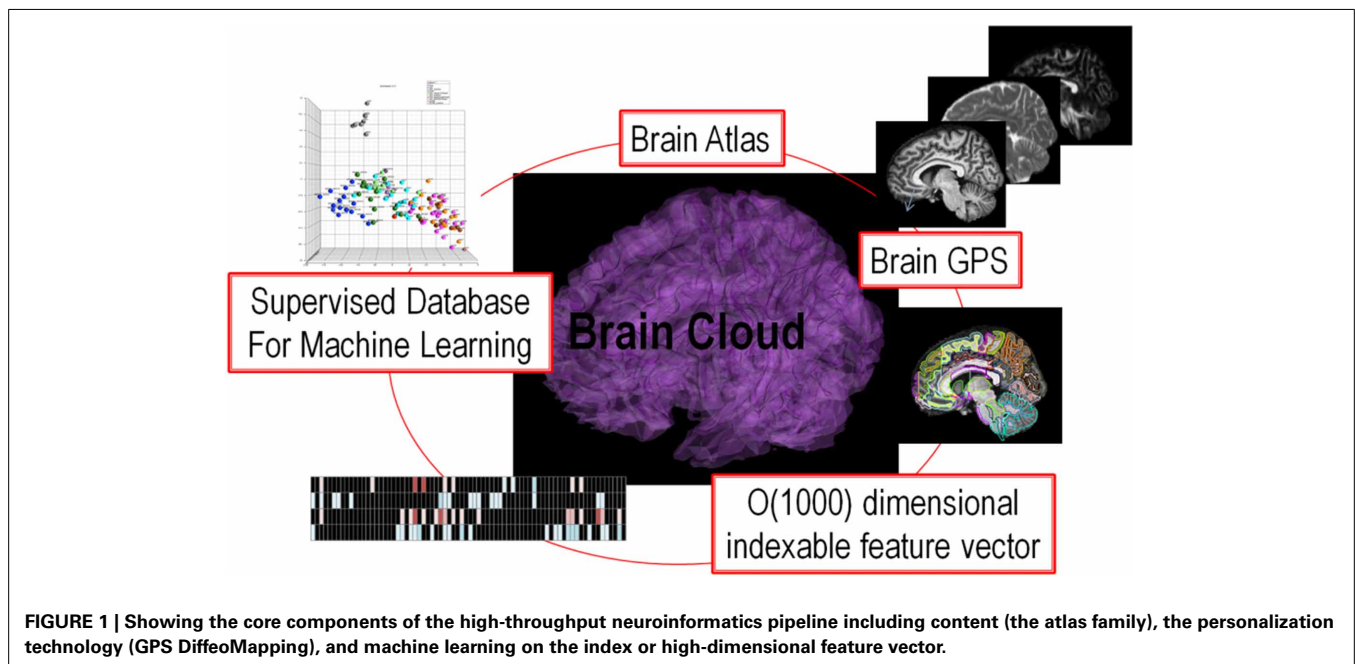
Exploiting the maximum potential of the imagers or the associated scans appears impractical without some form of encoding, or extreme data reduction. Reduction of high-dimensional imagery to symbolic knowledge representations encoded via the informative discriminating dimensions is one of the holy-grails of image analysis, a field which has advanced dramatically in the past several decades. From our own school of Grenander's metric pattern (Grenander, 1993) has emerged the field of computational anatomy (CA) for medical image analysis (Grenander and Miller, 1998, 2007; Toga and Thompson, 2001; Miller et al., 2002; Thompson and Toga, 2002; Ardekani et al., 2009; Ashburner, 2009; Pennec, 2009). The organizing principle in CA is that while there are variations in human structure and function, representation of the evolutionarily stable organization of processing in human beings are to a great extent organized around the structural manifestation of the genotype, throughout what we term the structural or anatomical phenotype. The evolutionary process has been masterful in its conservation of neural processing and its apparent organization around the macroscopic scales of human anatomy. We assume throughout that while functional layout is highly variable and ultimately associated with cellular architecture, it is manifest at the macroscopic scale of the topological organization of human anatomy and is preserved in large

part cross-sectionally. Striking examples include the tonotopic organization of the auditory system for representing the axes of complex spectral representation, the somatosensory and motor homunculus in sensory and motor cortex, and the conformal like representation of visual space in the visual field. In each case the spatial axis encodes the functional axis representation.

The fact that functional topography is supported via dense topologic correspondence to the anatomical coordinates is the basis of our personalization of atlas based neuroinformatics. The personalization step is accomplished via the construction of a positioning system for neuroinformatics termed DiffeoMaps. This is an infinite dimensional positioning system which we term a Geodesic Positioning System (GPS) (Miller et al., 2013b) transferring information between atlas or world coordinate systems and individualized coordinate systems. We term it geodesic positioning since the metric is constructed based on the shortest (geodesic) flow of diffeomorphisms which connect the coordinates (Miller et al., 2013b). Such a transfer of the atlas representation to the coordinates of the individual allows for the organization of the high-throughput medical image record into a high-dimensional “feature vector” or an “index.” Indexing via DiffeoMaps is the essential reduction or parsing of the individual into metadata representations upon which the machine learning phase of high-throughput neuroinformatics may be applied. Shown in **Figure 1** is our overall solution for high-throughput neuroinformatics, which includes atlases, diffeomorphic mapping for position (GPS), reduction to a high-dimensional feature vector or index encoding the anatomical and functional phenotypes, and machine learning via supervised clustering. This paper examines (i) cross-sectional image analysis to evaluate the health status of individual subjects with respect to the population data, (ii) integration of image and non-image information for diagnosis and prognosis.

RELATED WORKS

The high-throughput Neuro-Imaging Informatics introduced in this article is based on three core technologies; deformable multi-modal brain atlases, geodesic positioning of meta-data or semantic labels via diffeomorphic image transformation, and machine learning algorithms, as detailed in the Materials and Methods below. The deformable multi-modal brain atlases have been developed in both the coordinate systems provided by Montreal Neurological Institute (MNI) and the International Consortium of Brain Mapping (ICBM), which is a multicenter effort known for the MRI database and various brain atlases (<http://loni.usc.edu/ICBM/>). The atlases developed through this consortium have been implemented in leading software packages for the functional and anatomical brain analyses, such as Statistical Parametric Mapping (SPM, <http://www.fil.ion.ucl.ac.uk/spm/>), FSL (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>), MRICron (<http://www.mccauslandcenter.sc.edu/mricro/mricron/>), as well as our own MRISTUDIO (<https://www.mristudio.org/>). The primary uses of the MNI or ICBM atlases are to be a reference space for the voxel-based image analysis, in which statistical analyses are performed on each voxel after all images are normalized to the atlas space. This voxel-based approach has been widely used since it enables researchers to statistically analyze the whole brain with very high spatial specificity, and to report their findings on a standardized coordinate system. This approach also allows users to apply various types of anatomical parcellation maps, such as the automatic anatomical labeling atlas (AAL) (Tzourio-Mazoyer et al., 2002) and the LONI Probabilistic Brain Atlas (LPBA) (Shattuck et al., 2008) for quantifying gray matter functions and anatomy. Our deformable multi-modal brain atlases are extensions of these attempting to group voxels based on anatomical or functional units, through which features of each brain are preserved at 1 mm scale. While most of the “atlas-based”



approaches previously have targeted the gray matter areas of single contrast images, the atlas used in our approach is multi-modal, which means that the atlas consists of a set of images with different contrasts [e.g., T1- and T2-weighted images, Diffusion Weighted Imaging (DWI), Diffusion Tensor Imaging (DTI), and Susceptibility Weighted Image (SWI) contrasts] to allow multi-modal image analysis of both gray and white matter structures in the common anatomical framework. The multimodal capability is supported by the Large Deformation Diffeomorphic Metric Mapping (LDDMM) methods that employ single and multi-channel algorithms (Beg et al., 2005; Ceritoglu, 2008; Ceritoglu et al., 2009; Djamanakova et al., 2013), allowing for the incorporation of multiple imaging modalities while performing simultaneous mapping that maximally satisfies registration of the multiple modalities.

Another distinction we have made is to explicitly model both the geometric component of the atlas and associate to that the anatomical phenotype, simultaneously with the contrast component of the atlas which we generally associate to the function. This we do by providing a direct model in which the anatomical geometry carries the function, and demonstrate explicitly how to code via informatics both the anatomical geometry simultaneously with the functional contrasts. This forms the heart of our personalization via DiffeoMaps below. This allows us to directly generate classifiers and perform hypothesis generation about disease groups by both the anatomical phenotype as well as the contrast or function phenotype, and index them to different atlases. This can be contrasted to alternative approaches which use normalization viewing the geometric or anatomical phenotype as a nuisance parameter which is normalized out, like the affine group is removed rather than explicitly modeled.

Since the robustness of the machine learning framework to detect disease related anatomical and functional features of the brain has been demonstrated (Teipel et al., 2007; Hinrichs et al., 2011; Zhang et al., 2011), our approach is the generalizable extension toward high-throughput whole-brain multimodality analysis of heterogeneous brain conditions.

MATERIALS AND METHODS

ATLAS REPRESENTATION OF 1 mm STRUCTURAL—FUNCTIONAL CONTRASTS

The core of our high-throughput neuroinformatics technology is the conversion of the raw images into a structured, quantitative, and searchable high-dimensional feature vector. The basis for reduction to the numerical knowledge representation are the evolutionarily stable categorizations which neuroscientists have defined over the past decade. Our starting point is dense atlases of neuroanatomical structure and function indexed against age and group. We model the individual's imagery as an orbit under transformation of 1 mm scale coordinatized atlas information. **Figure 2** depicts our coordinatized human atlases demonstrating 3D anatomical information at different developmental stages (multi-dimensional) (Oishi et al., 2011c), different MR contrasts (multi-contrast) (Oishi et al., 2009), and varying coordinatized structural and functional definitions (Mori et al., 2013). The coordinate systems support MNI (Mazziotta et al., 1995, 2001) and Talairach (Talairach and Tournoux, 1988) coordinates as well as parcellations into different cortical areas as well as approximately

20 deep gray matter and 100 deep white matter structures all based on anatomical parcellation. The cortical partition includes structures such as parietal gyrus, frontal gyrus, pre-central gyrus, cuneus, lingual and others; the subcortical structures include amygdala, caudate, globus pallidus, hippocampus, putamen, thalamus, red nucleus, substantia nigra, hypothalamus, nucleus accumbens; the white matter structures include corticospinal, internal capsule, thalamic radiation, corona radiata, fornix, longitudinal fasciculus, corpus callosum, and others. Such a modern atlas also includes parcellations based on different anatomical and functional criteria such as cytoarchitecture, vascular territories, and anatomical and functional connectivity. This type of effort to parcellate the brain has been a subject of research based on histology (von Economo and Koskinas, 1925; Sarkisov et al., 1955; Mai et al., 1997; Schleicher et al., 1999; Tzourio-Mazoyer et al., 2002; Zilles et al., 2002) or MRI for the cortex (Lancaster et al., 2000; Mazziotta et al., 2001; Tzourio-Mazoyer et al., 2002; Hammers et al., 2003; Maldjian et al., 2003; Shattuck et al., 2008), white matter (Meyer et al., 1999; Mori et al., 2008; Oishi et al., 2008) and the whole brain (Fischl et al., 2002; Desikan et al., 2006; Oishi et al., 2009, 2011c, 2013).

PERSONALIZATION VIA DIFFEOMAPS AS A GEODESIC POSITIONING SYSTEM

Reduction to a high-dimensional feature vector which can be indexed requires us to model the high-throughput imagery. The underlying assumption of our model is that the meta-data representing the individual's structure and function is carried by the individual's coordinate systems, and there exists a structure preserving mapping which transforms the individual's coordinates into the stereotypical atlases. We term these transformations morphisms, these transformations form a group $\phi \in G$. The structure preserving morphisms provide correspondence between "charts" of the human brain as contained within atlas and the individual's coordinates. In this sense the morphisms provide a positioning system through their algebraic group action. Our group has come to call this the *metamorphism* model (Miller and Younes, 2001; Trouvé and Younes, 2005), organizing the structural and functional informatics, the images $I \in \mathfrak{S}$ into the transformation–image pair $[\phi(x), I(x)]$, $x \in X$ related via the algebraic pairing

$$\bullet : (\phi, I) \mapsto I' \doteq \phi \bullet I \in \mathfrak{S}. \quad (1)$$

In this model the morphisms denoted by $\phi(x)$, $x \in X$ carries the coordinatized contrast metadata imagery denoted by $I(x)$, $x \in X$.

Personalization occurs via smooth transformation of the atlas meta-data $\phi \cdot I_{\text{atlas}}$. For this we define a distance $\inf_{\phi} d(I, \phi \cdot I_{\text{atlas}})$ between the individual's representation and transformed atlas solving a variational problem for the coordinate (Dupuis et al., 1998; Beg et al., 2005; Ceritoglu et al., 2009) transformation. The correspondence between the individual and atlas is termed the "DiffeoMap," which provides an infinite dimensional positioning between atlas and world coordinates. This is in sharp contrast to the 7-dimensional similarity maps used in geographic positioning. To see this, the Eulerian velocities of Equation (2) below, while spatially smooth are a high-dimensional field, implying the Jacobian expressing first order transformation of coordinates in

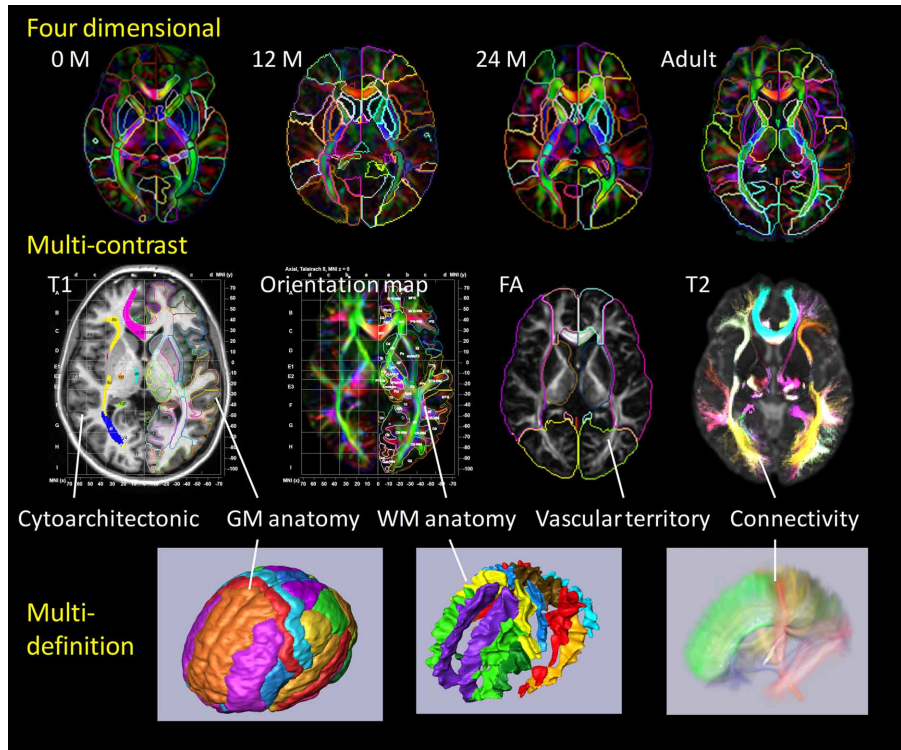


FIGURE 2 | Panels show a current brain atlas, including 3D anatomical information at different developmental stages (multi-dimension), different MR contrasts (multi-contrast), and different structural definitions. The coordinate systems include MNI and Talairach coordinates with the brain depicted as

parcellated into multiple cortical and subcortical areas including deep gray and white matter structures based on anatomical features (anatomical parcellation) as well as functional parcellation based on cytoarchitecture, vascular territories, and anatomical and functional connectivity.

space allows the tissue to locally scale and twist while at the same time preserving relative organization.

Shown in **Figure 3** are instantiations of our structure-function metamorphosis model, including structural contrast imagery T1, orientation vector imagery such as DTI, metabolic contrast as measured via magnetic resonance spectroscopy and functional connectivity via resting-state fMRI (Faria et al., 2012).

Each of the modalities has its own definition of the morphism acting on the meta-data of the contrast imagery explicating the algebra represented by \bullet , specifically (i) for the submanifolds of subcortical structures, gyral curves and cortical surfaces the morphism acts $\phi \cdot x = \phi(x)$, (ii) for scalar imagery such as T1 the morphism acts via the inverse $\phi \cdot I = I \circ \phi^{-1}$, and (iii) for symmetric matrix-valued DTI (color in **Figure 1**) with eigen elements $\{\lambda_i, \phi_i\}$, the morphism acts to preserve the eigenvalues and determinant, rotating the eigenvectors $\phi \cdot I \doteq (\lambda_1 \hat{e}_1 \hat{e}_1^t + \lambda_2 \hat{e}_2 \hat{e}_2^t + \lambda_3 \hat{e}_3 \hat{e}_3^t) \circ \phi^{-1}$ with $\hat{e}_1 = \frac{(d\phi)\hat{e}_1}{\|(d\phi)\hat{e}_1\|}$, $\hat{e}_2 = \frac{(d\phi)\hat{e}_2 - \langle \hat{e}_1, (d\phi)\hat{e}_2 \rangle \hat{e}_1}{\sqrt{\|(d\phi)\hat{e}_2\|^2 - \langle \hat{e}_1, (d\phi)\hat{e}_2 \rangle^2}}$, $\hat{e}_3 = \hat{e}_1 \times \hat{e}_2$, and $d\phi = \begin{pmatrix} \partial\phi_i \\ \partial x_j \end{pmatrix}$ the 3 by 3 Jacobian matrix, with x denoting the vector cross-product.

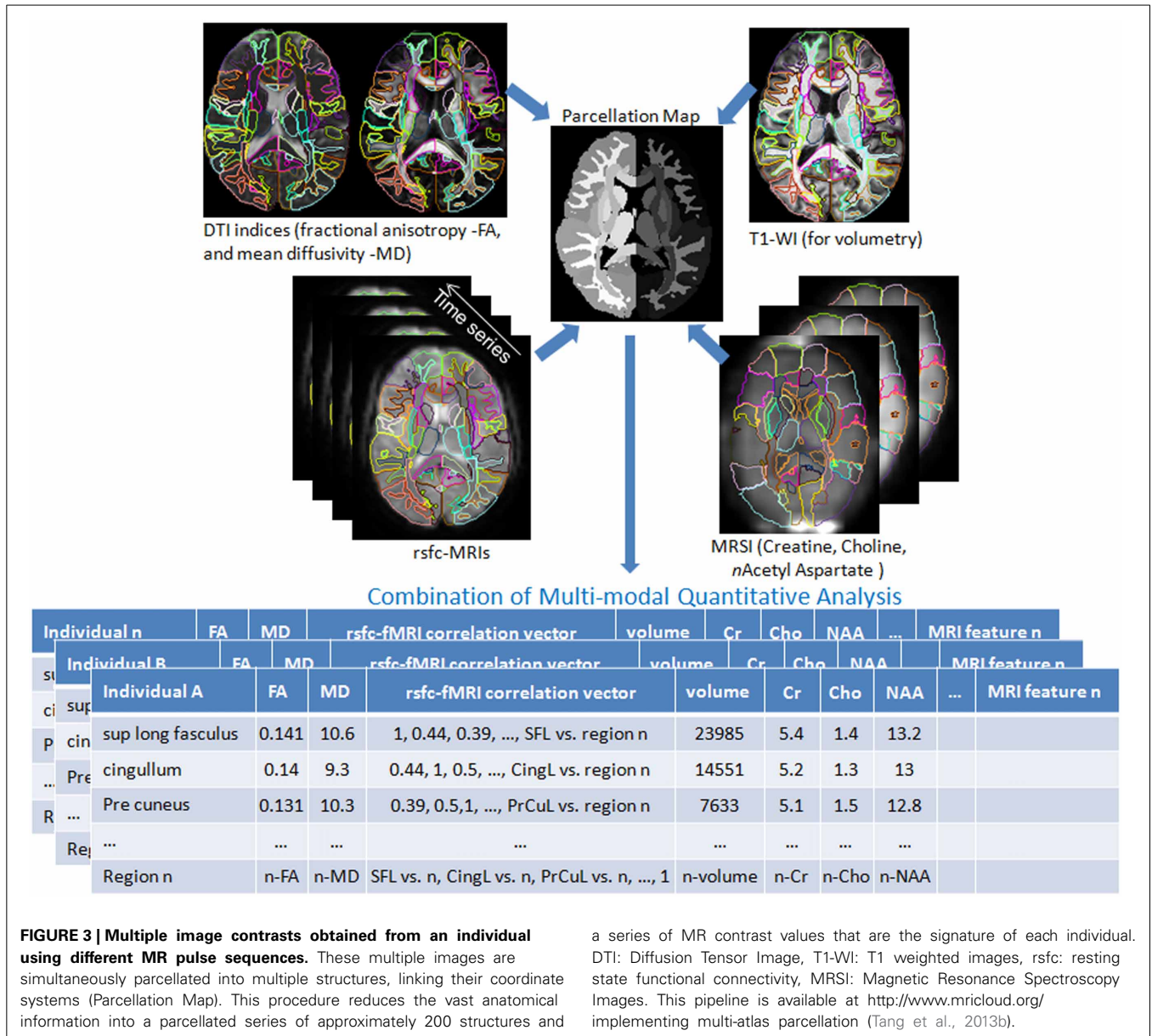
THE HIGH-DIMENSIONAL FEATURE VECTOR AND MACHINE LEARNING

The scanners are the high-throughput devices generating the high-dimensional raw images of $O(10,000,000)$ in complexity,

and the pipeline converts it into a quantitative searchable feature vector $\{f = X_1, X_2, X_3, \dots\}$ representing the individual at $O(1000-10,000)$ complexity. Diffeomorphic GPS (Miller et al., 2013b) provides the basis for data reduction, since the anatomical structure phenotype is encoded by the morphisms and the meta-data of structure-function are encoded by the contrast imagery represented in atlas coordinates. The metamorphism model organizes the structural and functional informatics into the pair $[\phi(x), I(x)], x \in X$.

The GPS correspondences are diffeomorphisms, one-to-one and smooth mappings between coordinate systems $\phi : X \leftrightarrow Y$, $\phi(x), x \in X$ providing correspondences $\phi : I \leftrightarrow I_{atlas}$ between the individuals in the population and the atlas. The correspondences are generated as solutions of the classical Lagrangian flow equations, $\dot{\phi}_t = v_t(\phi_t), t \in [0, 1]$ the time derivative of the flow v_t is termed the Eulerian velocity (Christensen et al., 1996). Constructing the DiffeoMaps occurs via the geodesic connection of one coordinate system to the other (Miller et al., 2006), solving for the geodesic connection between individual I and atlas $\phi \cdot I_{atlas}$ according to

$$\inf_{v_t, t \in [0, 1]: \dot{\phi} = v(\phi), \phi_0 \cdot I_{atlas} = I_{atlas}} \int_0^1 \|v_t\| v dt \text{ subject to } I = \phi_1 \cdot I_{atlas}. \tag{2}$$



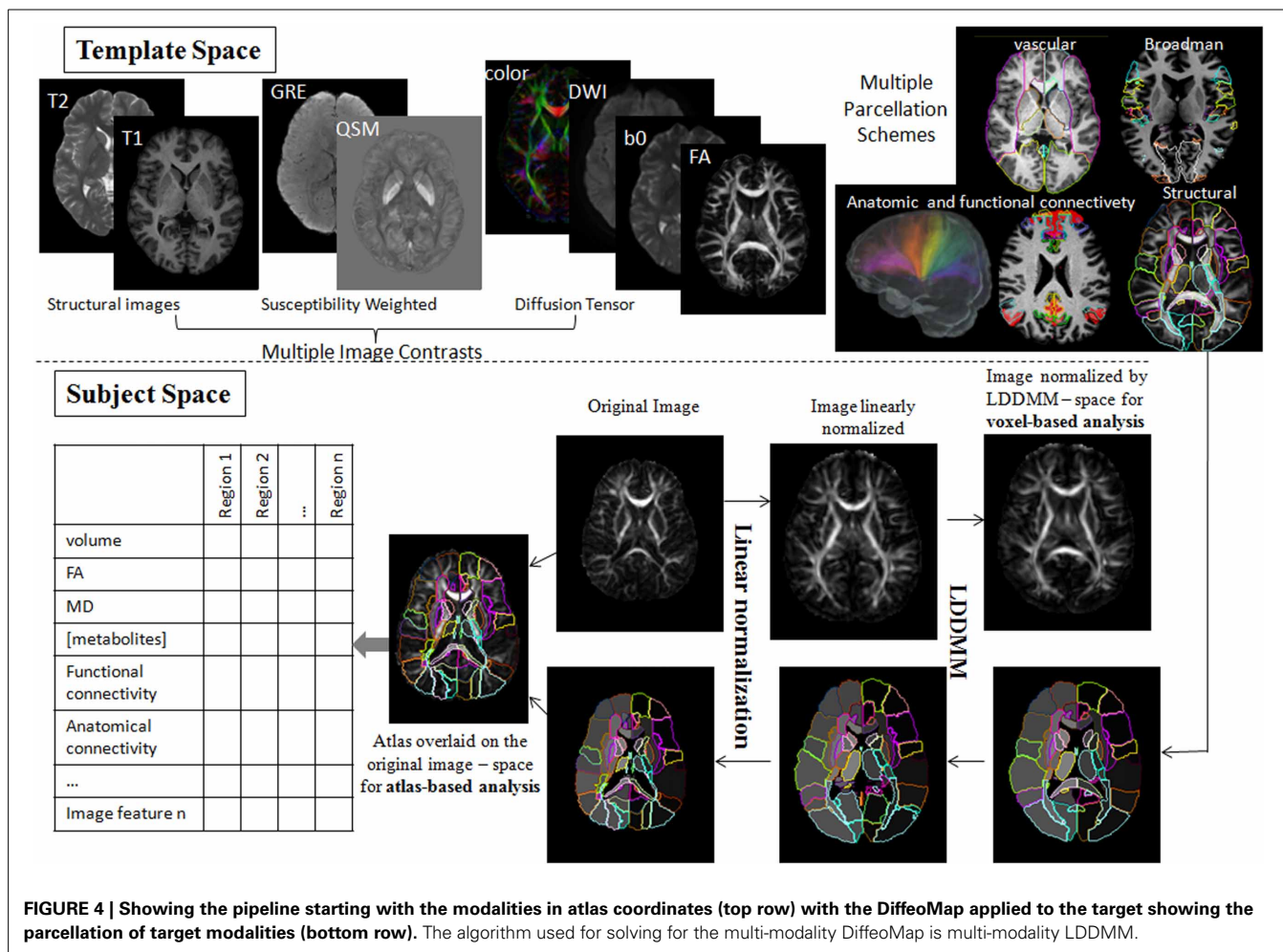
The geodesic connections are encoded via their initial tangent vector at the identity, denoted as $v = v_{t=0} \in V$. This forms the natural coordinate system of our GPS (Miller et al., 2013b). We have reduced the anatomical phenotype to a set of coordinates $v = v_{t=0} \in V$ centered at the atlas.

This is a natural representation of the anatomical or shape phenotype since the norm of the coordinates preserves the metric structure on the space of anatomies using this framework (Miller et al., 2006). The shortest flows connecting the template and individual coordinate systems define the metric in this space, the metric of Equation (2) is given by the integrated norm of the vector fields generating the morphisms. The reduction of the shape phenotype to these diffeomorphic connections we call *diffeomorphometry* (Miller et al., 2013b). At the 1 mm scale of MR imagery the anatomical phenotype is extremely sparse relative to the high-dimension of the initial data. For smooth imagery such as MRI linear functions of the vector fields, termed the shape

momentum, are concentrated to the boundaries of the homogeneous subcomponents of the object (Miller et al., 2006; Qiu and Miller, 2008). At places in the image that are constant the shape is coded as zero. Plainly put, at the 1 mm scale gyral and subcortical regions of the MRI contrasts do not discriminate the cellular architecture.

Shown in **Figure 4** is an instantiation of our pipeline, depicting the personalization phase via DiffeoMap. The generation of the geodesic of Equation (2) for image matching via the solution of a quadratic variation problem on the vector field we call large deformation diffeomorphic metric mapping (LDDMM) (Beg et al., 2005). The modalities are shown in the top row in atlas coordinates with the DiffeoMap applied to the target showing the parcellation of target modalities shown in the bottom row.

Figure 5 shows a depiction of the subcortical neuronanatomy atlas as measured in 1 mm scale MR. The left panel shows



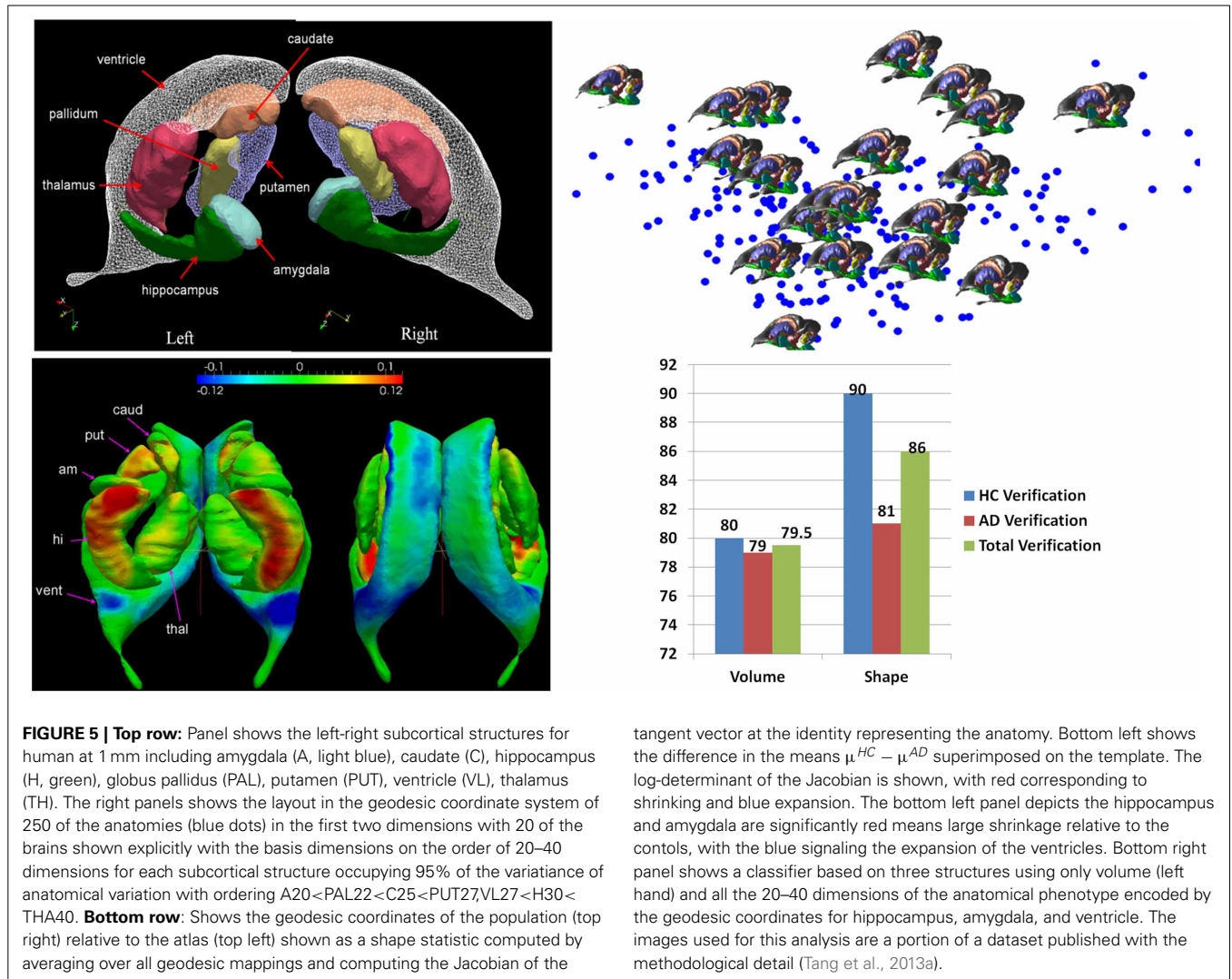
the atlas of 14 subcortical structures, amygdala (A, light blue), caudate (C), hippocampus (H, green), globus pallidus (PAL), putamen (PUT), ventricle (VL), thalamus (TH), each surface in the atlas containing order 1000 vertices. The set of structures correspond to an atlas generated from the population of healthy controls (HC) and Alzheimer’s disease (AD) computed using the surface template estimation algorithm described in Ma et al. (2010). To demonstrate the sparsity of the anatomical phenotype at 1 mm scale, the geodesic correspondence between the atlas and a database of 250 subcortical brains were generated giving a coordinate identification of each element in the population, $I \sim \nu$, where ν is the geodesic coordinate representation of the anatomy to the atlas. To understand the variation over the population, they were expanded via principle component analysis into a basis $\nu(f_1, f_2 \dots) = \sum_i f_i U_i$; the f ’s are reduction of the anatomical phenotypes to the basis of eigenfunctions U . The sparsity of the anatomical phenotype was calculated across the population calculating the dimension required for encompassing 95% of the energy of subcortical variation. Generally each structure requires between 20 and 40 dimensions, with hippocampus and thalamus having the greatest shape variation within

the population in terms of number of dimensions. The 95% variance cutoff as a function of dimensions for each structure is $A20 < PAL22 < C25 < PUT27, VL27 < H30 < THA40$, the sparse subcortical shape phenotype at 1 mm scale is $O(1000)$. The right panel shows the layout in the geodesic coordinate system of 250 of the anatomies (blue dots) in the first two geodesic dimensions with 20 of the brains shown explicitly.

This huge data reduction is noteworthy as it is the direct generalization of the sparsity of rigid body momentum which itself encodes translation and angular momentum to single 3-vectors, even though the inertia is extended over the entire object. Taking the midbrain as roughly 1/3 of the total brain volume of 2–4 Million voxels implies a data reduction of three orders of magnitude to $O(1000)$.

CORTICAL, SUBCORTICAL AND WHITE MATTER PARCELLATION FEATURE VECTOR

The global positioning solution provides registered coordinates for the encoding of the target coordinates system into a parcellation corresponding to the anatomically defined partition of atlas coordinates in the 200 white and gray matter parcels. Denoting the atlas partition p_i , and since there can be as many as 7 MR contrast



values including T1, T2, B0, trace, FA, spectroscopy, gives O(1000) features

$$f_{P_i}^c = \int_{P_i} I^c(x) dx, i = 1, \dots, 200, c = 1, \dots, 7. \quad (3)$$

Shown in **Figure 6** are examples of the neuroinformatics parcellation which is transported via the personalization phase. **Figure 6A** shows the DiffeoMap personalization of the atlas into the coordinates of a spastic cerebral palsy patient with visually appreciable anatomical abnormalities (the color highlights the volume change larger than two standard deviations). The three rows show measurement results for volume, FA, and MD. Each column is an entry for one of the 200 anatomical structures. The top row represents anatomical information of each parcellated structure. In feature space, the neuroinformatics atlas supports both empirical means as well as empirical variances. Only features which demonstrate as outliers are depicted.

The bottom part, **Figure 6B**, shows an example of population data, in which the atlas partition of the anatomical phenotype for the listed structures (the bottom row in **Figure 6A**) are presented for 10 cerebral palsy patients (P3 is the individual shown in **Figure 6A**). All patients shared the same spastic phenotype with varying degree of motor impairment indicated by GMFCS scores. Abnormal parcellation volume values are presented by z-scores. At a glance, even though the patients were selected by similar clinical manifestations, a marked degree of anatomical variability can be recognized implying the importance of clustering on the spectrum of anatomical phenotype.

FUNCTIONAL MRI AND CONNECTIVITY MAPS IN ATLAS COORDINATES

Functional magnetic resonance imagery (fMRI) also provides ideal measurements for studying pairs of interactions in the brains. fMRI connectivity is based on empirical correlations of temporal responses between pairs of elements in the representation. **Figure 7** shows an example of empirical correlation of fMRI at lag-0 using the common atlas coordinate system to parcellate

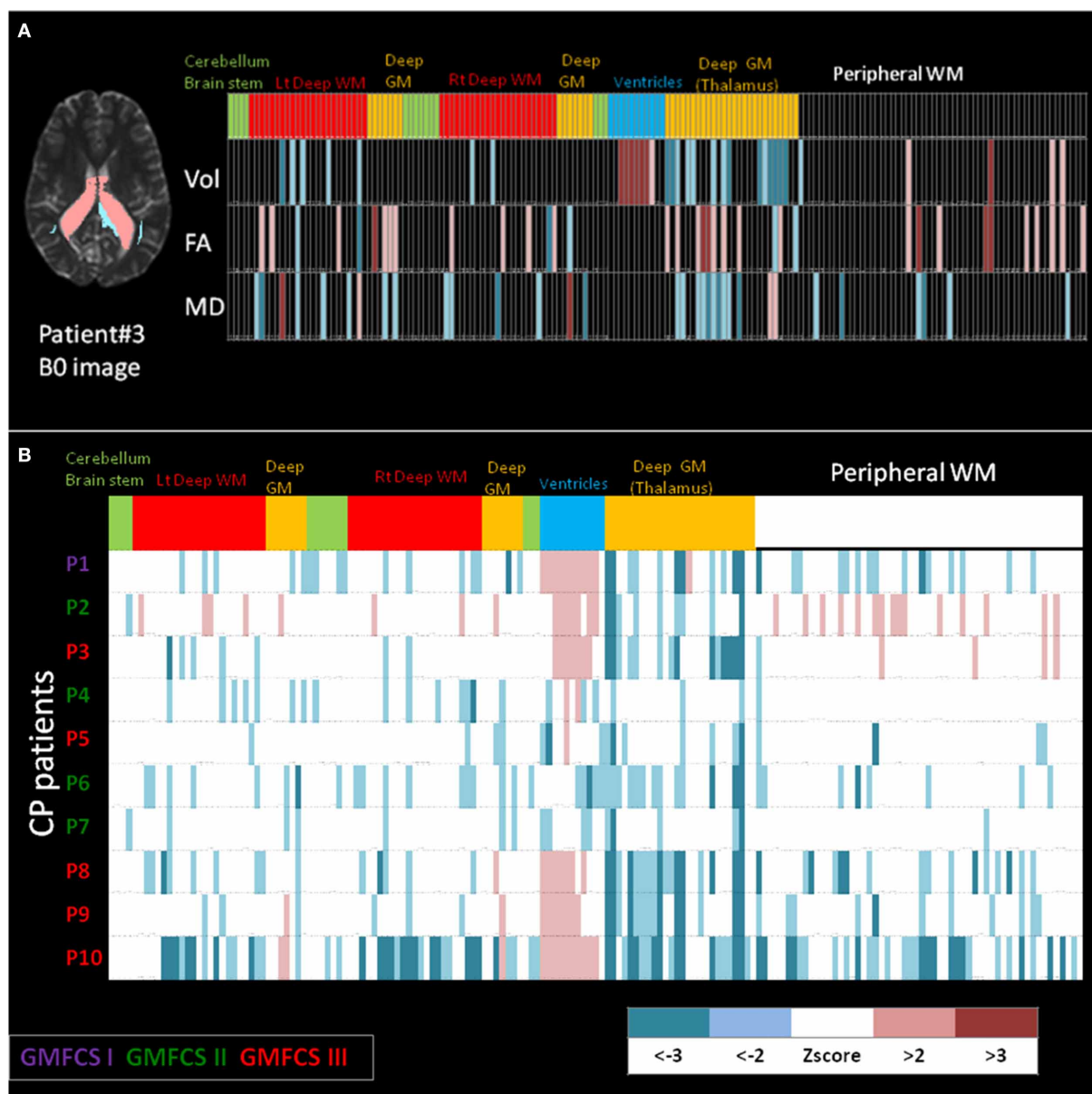


FIGURE 6 | (A) Feature vector from the personalization DiffeoMap correspondence between the atlas and an individual's coordinate system associated with focal disease category. Informatics partition with 200 structures including, volume, FA and MD values into peripheral white and gray matter, and deep white and gray matter structures. The features are color coded according to the statistics to depict color-coded outliers: WM:

white matter, GM: gray matter. **(B)** Example of population data including 10 cerebral palsy patients with different prognoses in their motor disability. Informatics partition with 200 structures of volume for each patient is shown as 10 rows. The features are color coded according to z scores calculated based on normal control population. CP: cerebral palsy, GMFCS: gross motor function classification system.

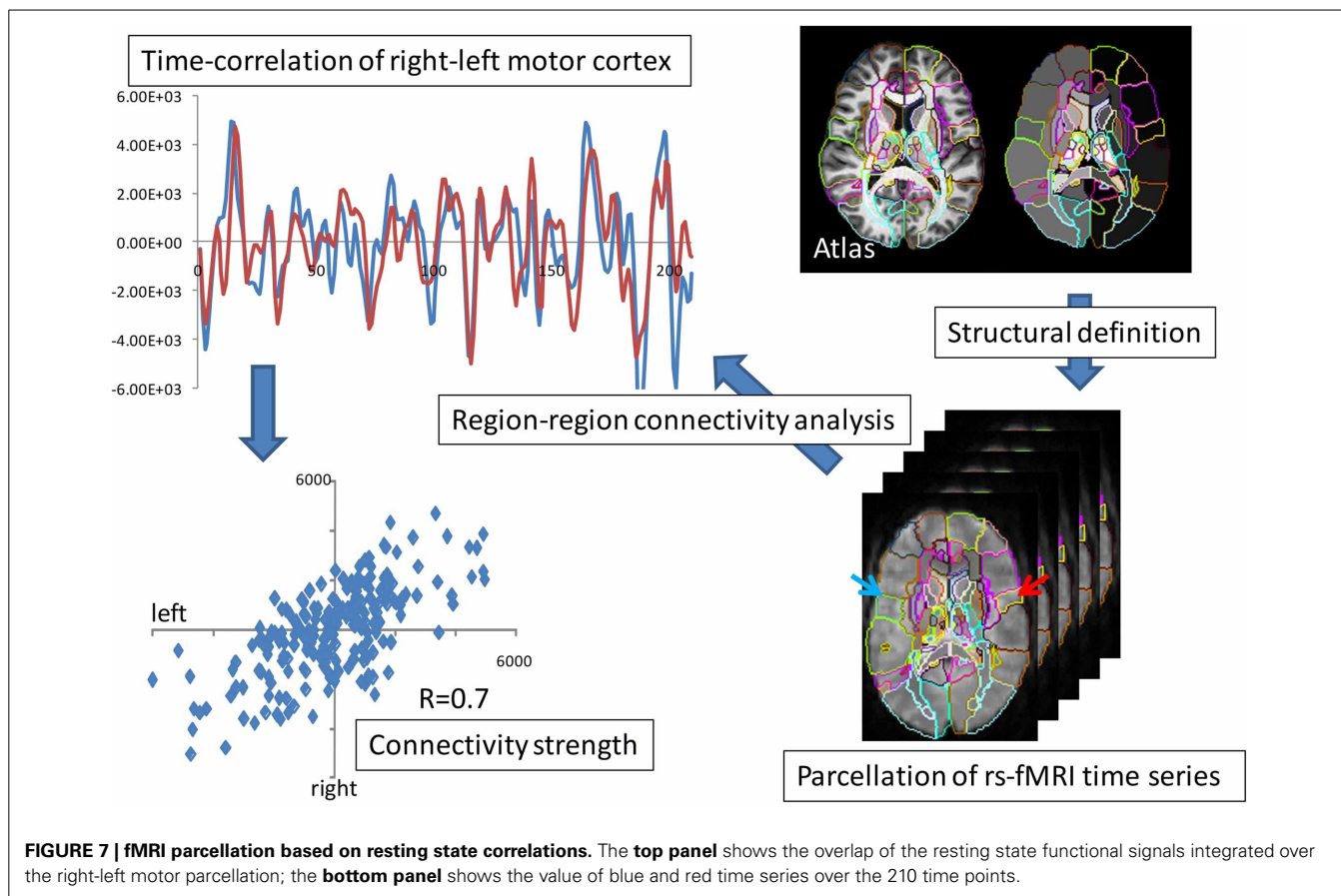
symmetrically associated motor cortex areas plotting the resting-state MRI functions (rs-fMRI) (Tzourio-Mazoyer et al., 2002; Eickhoff et al., 2005; Achard et al., 2006; Hagmann et al., 2008; He et al., 2009; Wang et al., 2009).

Shown is the time series of the fMRI image modality $I^{\text{fMRI}}(x, t)$ integrated over the right and left motor cortex parcels. Notice the strong correlation depicted via the superposition of the red-blue time sequences. These highly correlated patches of tissue has resulted in the widely used

ICA model in which the measured functional signal is the superposition of “networks,”

$$I^{\text{fMRI}}(x, t) = \sum_i f_i^{(t)} U_i^{(x)} \quad (4)$$

the U 's playing the role of the resting-state networks. Working in the registered coordinates of the atlas allows for the construction of these resting state networks in the parcellations of the atlas by



simply replacing the functional MR signal by its parcellated representation $I_{P_i}(t) = \int_{P_i} I_{fMRI}(x, t) dx$, $P_i = 1, \dots, 200$. The f 's are the dimensions of the functional MRI signal representing 10–20 resting state dimensions added to the feature vector.

MACHINE LEARNING INVESTIGATION OF DISEASE-SPECIFIC PHENOTYPES

High spatial resolution is one of the most significant advantages of clinical MRI and its usefulness in studying pathological condition and detecting abnormalities. It seems clear from many studies that because of the noise versus signal tradeoff most detectable pathologies from MRI are signaled via small groups of spatially correlated voxel contrasts. Dimensionality reduction becomes the central methodology for MRI analysis in clinical applications. Combining unsupervised principal component analysis (PCA) along with supervised training, on the supervised group means under the common covariance model gives linear discriminant analysis (LDA).

Given m -length feature vectors, a collection n of them $\{f_j\}$, then PCA calculates the singular value decomposition (SVD) of the $m \times n$ matrix $F = (f_1, f_2, \dots, f_n) = U\Sigma V^t$, where U is an $m \times m$ orthonormal matrix of vectors with Σ diagonal with entries the singular values. The connection to least-squares and covariance modeling of Gaussian processes is that the left singular vectors $U = (U_1, \dots, U_m)$ are the eigenfunctions of the empirical covariance FF^t ; the set of diagonal entries squared of Σ are the

variances in the rotated independent representation of the left singular vectors. LDA then is the supervised version. Given groups of labeled feature vectors $\{f_j^g\}$, $g = 1, \dots$ then each labeled group has a mean and covariance:

$$\mu^g = \sum_{j=1}^{n_g} f_j^g / n_g, K^g = \sum_{j=1}^{n_g} (f_j^g - \mu^g)(f_j^g - \mu^g)^t / n_g. \quad (5)$$

Then LDA is PCA on the group means μ^g using the common covariance $K = \sum_g K^g$. Quadratic discriminant analysis is a particular non-linear discriminant analysis (QDA) relaxing the common across groups covariance assumption. The high-dimensional structural and functional phenotypes are encoded via high-dimensional feature vectors. The classifiers are constructed from the cohorts of neuropsychiatric illnesses collected via the supervised training component. The crucial advantage of this approach is that the anatomical and structural phenotypes are indexed to the coordinates of the template. For the subcortical structures, for example, the anatomical phenotype is immediately reduced from a feature vector of dimension $O(10,000,000)$, to the dimension of the surfaces which is $O(10,000)$. Similarly, the functional feature is indexed over the anatomical substructures. This of course requires the notion of a template coordinate system which is centered in the population. Unlike other methods since we have explicitly modeled the anatomical and functional phenotypes, we can perform classification on both rather than viewing coordinate system transformation as a nuisance variable.

RESULTS

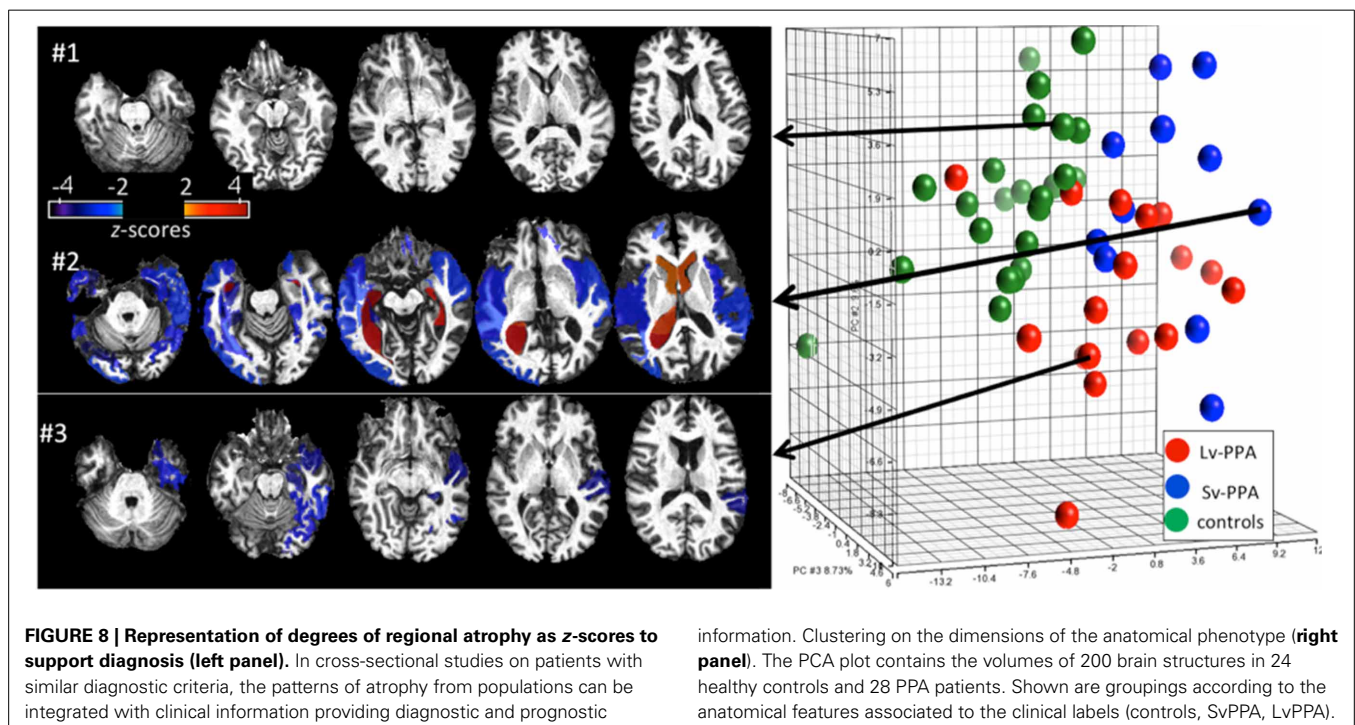
THE ANATOMICAL PHENOTYPE: IMAGE RETRIEVAL AND CLUSTERING

With an estimated 100 million scans every year in radiology, a huge amount of imaging data are generated every day, with these data stored in clinical Picture Archiving and Communication Systems (PACS) and are rarely used to support medical decision-making in cross-sectional examination of patient populations. Similar to the role of genomic and proteomic information for personalized medicine, anatomical phenotypes are fundamentally important for medical decision-making, yet often not systematically utilized in daily medical practice. While text-based patient records for retrieval of disease cohorts is commonly used, to utilize the anatomical phenotype for medical decision-making for individual patients we need to be able to use the patient image as the search key with the diagnostic label being the retrieved information. Using the high-dimensional feature vector without any diagnostic supervised labeling allows us to group and retrieve based on the structural phenotype.

Figure 8 shows an example of retrieval based on the anatomical phenotype, essentially delivering previously supervised cases with clinical information already in the data base. There are two types of information delivered in this analysis. For this we represent the anatomical variance of the population as shown in **Figure 5** for the subcortical structures to represent the coordinates of the anatomical position of the patient with respect to the atlas coordinate system relative to the population. This can be highly illustrative. For example, **Figure 8** shows healthy individuals as controls (green dots in the PCA plot of the structural volumes) and patients with two variants of Primary Progressive Aphasia (PPA), a neurodegenerative

disease characterized by predominant and progressive deterioration in language in the absence of major change in personality, behavior or cognition other than praxis for at least two years. The *z*-score map in patient #3 reveals atrophy at the temporal left side that could be dubious at visual inspection only. In the addition, this subject is closer to other PPA patients than to the controls in the PCA plot, evidencing that the anatomical phenotype identified agrees with the clinical label.

Defining cohorts of similar patients is commonly done based on a host of features, including clinical behavioral and structural and functional phenotypes as measured in the functional and structural imagery. **Figure 8** examines clusters of cohorts based solely on the anatomical phenotype feature vector. The first principal component (PC1) accounts for global cortico-subcortical atrophy and ventricle enlargement, and mainly segregates age-matched controls from the PPA population. The segregation between two PPA variants (Semantic-SvPPA and Logopenic-LvPPA) is driven by severe and global atrophy of deep areas in SvPPA (PC2) and the predominant fronto-parietal atrophy in LvPPA, with relative preservation of temporal areas, particularly left, when compared with SvPPA (PC3). This agrees with past anatomical qualitative description of these populations. The existence of “outliers” corresponding to patients or controls surrounded by subjects of different labels are due to the singular anatomical features of these subjects. This type of analysis provides a platform for hypothesis-free comprehensive characterization of anatomical phenotype. Such quantitative analysis allows the investigation of various anatomy-associating factors, such as disease progression and functional outcomes, in a systematic manner.



DISEASE QUANTIFICATION OF ANATOMICAL PHENOTYPE VIA GEODESIC COORDINATES

The GPS provides geodesic coordinates for representing every element in the population relative to the templates. We have examined machine learning on the subcortical structure coordinates shown in **Figure 5**. In the ADNI (Mueller et al., 2005) project there is extensive diagnostic supervised labeling enabling group based discriminations such as LDA/QDA for cross-sectional study of cohorts in dementia. A total of 385 subjects were segmented into their subcortical structures and lateral ventricle using FreeSurfer (Fischl et al., 2002) based on the analyses published by the Dale group (Fennema-Notestine et al., 2009). There were a total of 210 HC and 175 subjects with AD. To illustrate the average differences between the healthy control and Alzheimer's disease populations a HC-template surface and an AD-template surface was generated representing the center of each of the populations. We compared these two, HC-only and AD-only, template surfaces from the two different populations, and which are represented in geodesic coordinates relative to the overall template representing both HA-AD populations via the two class means μ^{HC} , μ^{AD} . To visualize these as shapes we calculated a scalar field corresponding to the log-determinant of the Jacobian of the map between the two averages, and visualized it on the template surface generated from the HC population generated using the algorithm described in Ma et al. (2008, 2010). This scalar field measures how much expansion/atrophy at each vertex of averaged surface from AD compared to that from HC in the logarithmic scale: i.e., positive value corresponds to surface expansion in the AD averaged surface at a particular location, while negative value denotes surface atrophy. The bottom left panel of **Figure 5** shows the mean differences between the two populations (bottom left), and is a visualization of one of the direction vectors in the Fischer discriminant the difference between the means $\mu^{HC} - \mu^{AD}$, shown as a plot of the Jacobian determinant. The color red represents the determinant being less than one, corresponding to shrinkage. The blue color corresponds to expansion. We see most of the shape change occurring at the ventricle expansion and the hippocampus and amygdala shrinkage.

The bottom right panel of **Figure 5** shows the result of building classifiers via machine learning whose discriminating dimensions are encoded in the picture in the lower left panel. We constructed the LDA and quadratic QDA classifiers using one of the *Leave-One-Out Cross-Validation* resampling method generating 385 LDA classifiers, testing on one of the subjects treating them as the testing data, and constructing the LDA class means μ^{HC} , μ^{AD} from the other 384 subjects (Tang et al., 2013a). For the two class problem the discriminating direction resulting from LDA on the geodesic coordinates is the projection of the differences in the means according to $K^{-1}(\mu^{HC} - \mu^{AD})$ on the common covariance. The Bayes classifier for the two class problem becomes a comparison to a threshold of the inner product of the feature vector on the discriminating direction:

$$f^t K^{-1} (\mu^{HC} - \mu^{AD})_{AD}^{HC} \underset{<}{\overset{>}{\geq}} \theta. \quad (6)$$

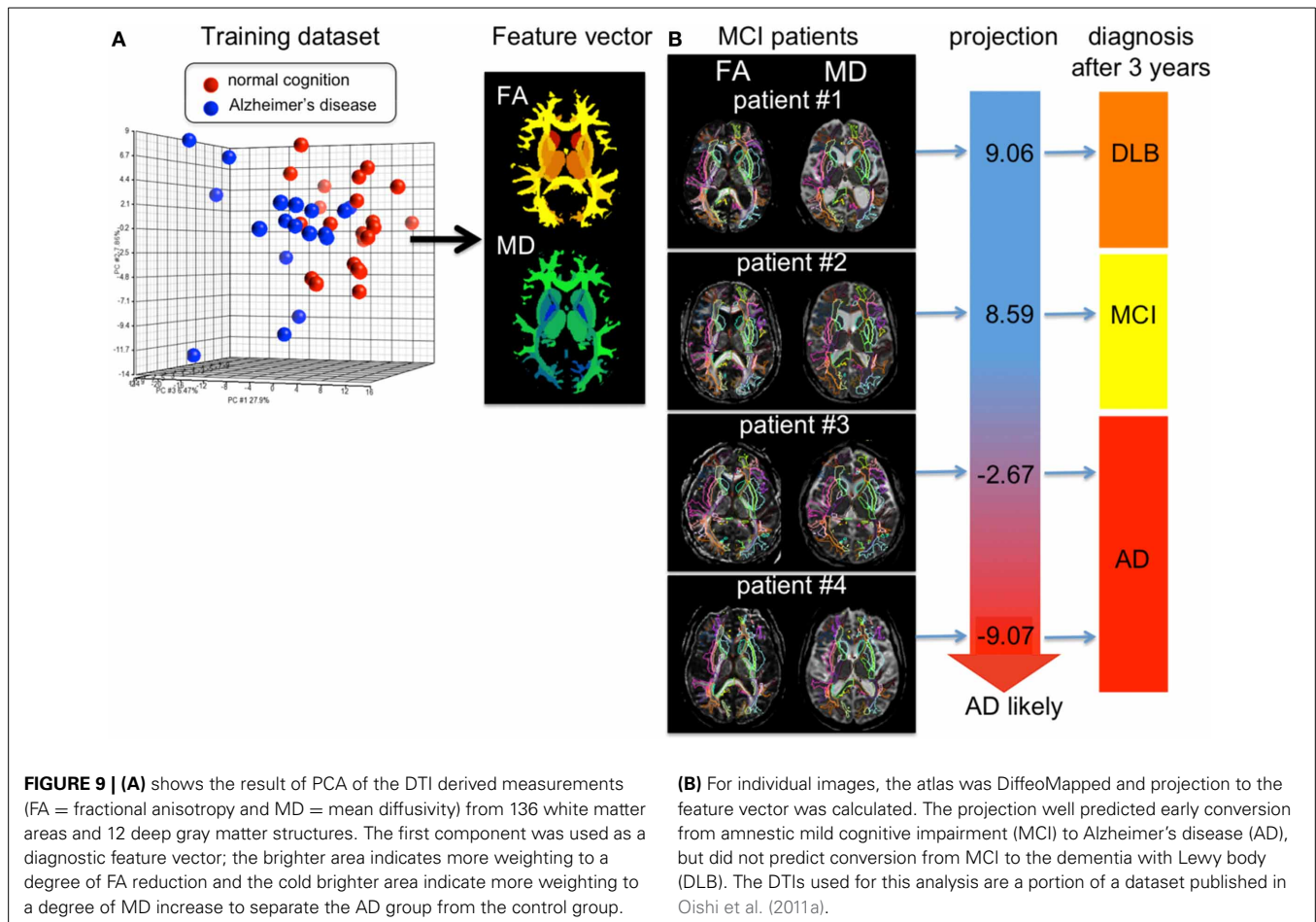
As shown in **Figure 5** we find uniformly, as depicted by the blue bars in the classifier diagram, that the shape dimensions

associated with the subcortical structures are significantly more discriminating than the volumes, generally reducing the errors in discrimination by more than 10%. The significant dimensions in volume and shape are associated with hippocampus and amygdala agreeing with previous results (Qiu et al., 2009). The specificity and sensitivity based on using the PCA shape dimensions in the feature vector for these three subcortical structure phenotypes is 90 and 81%, respectively. This is consistent with recent findings in another preclinical dementia study (Miller et al., 2013a) in which the shape of the temporal lobe subcortical structures is more discriminating than volume measures as well as in a Huntington's disease study tracking caudate, putamen, and globus pallidus (Younes et al., 2012).

PERSONALIZED ANALYSES: PREDICTING FUTURE CONVERSION BASED ON WHITE MATTER STRUCTURAL REPRESENTATIONS

While most research studies are based on cross-sectional population-based analyses, clinical diagnosis is always based on single individuals. This is performed by visual inspection in daily radiological diagnosis, in which images are most likely analyzed in a structure-by-structure basis, not in a voxel-by-voxel basis. Atlas-based neuroinformatic analyses in terms of their aggregate scale of the feature vector is compatible with many current diagnostic practices. Interestingly, histopathological studies indicate that white matter is an excellent target for both the early diagnosis of AD and for monitoring disease progression, motivating the use of DTI for studying patients with AD (Brun and Englund, 1986; Englund et al., 1988; Meier-Ruge et al., 1992; Gunawardena and Goldstein, 2001; Pigino et al., 2003; Sjobeck et al., 2005; Stokin et al., 2005; Chevalier-larsen and Holzbaur, 2006; Oishi et al., 2011b). There are already a large number of cross-sectional group comparison studies reporting significant differences in DTI derived measurements between the patients and controls, suggesting that white matter damage may exist in the pre-symptomatic phase of AD (Rose et al., 2000; Kantarci et al., 2001; Medina et al., 2006; Ringman et al., 2007; Stahl et al., 2007; Zhou et al., 2008; Damoiseaux et al., 2009; Salat et al., 2010; Sexton et al., 2011). One of the important questions after group analyses is whether these findings can be applicable to each individual to predict future conversion from memory impairment without other cognitive deficits (amnestic mild cognitive impairment) to dementia caused by AD. This is important because the amnestic mild cognitive impairment is a clinical category including multiple diseases or conditions with different pathological background, and not all of them develop AD (Albert et al., 2011).

Figure 9 shows results of personalizing the cross-sectional atlas statistics to several patients. A weighted feature vector, which could separate AD from cognitively normal population, was created from training datasets including groups of patients and cognitively normal age-matched individuals using dimensionality reduction applied to the atlas feature vector, and then DiffeoMapped to each individual to calculate the projection onto each patient. Shown in **Figure 9** are examples of the prediction of the conversion to Alzheimer's dementia in successive followup. Notice this projection doesn't predict conversion from amnestic mild cognitive impairment to the dementia with Lewy body, which is another type of neurodegenerative dementia. This type of analysis requires a large database with



longitudinal follow up which is an important current focus of our efforts.

DISCRIMINATING BETWEEN MULTIPLE DISEASES

The concept of group analysis in research studies assumes consistent locations of abnormalities, which does not hold for clinical situations, with heterogeneous patient populations and lack of an age-matched control group. The atlas-based neuroinformatics is compatible with the analysis of multiple diseases with different anatomical features. **Figure 10** shows the applicability of atlas-based neuroinformatics to capture anatomical features of multiple neurodegenerative diseases with known macroscopic anatomical alterations. To appropriately integrate diagnostic information to characterize the anatomical features related to each disease category, PCA and LDA were applied sequentially to a dataset consisting of 102 T1-weighted images from AD, primary progressive aphasia, Huntington's disease, hereditary spinocerebellar ataxia and normal control participants. These were parcellated based on the JHU-atlas [the images used for this analysis are a portion of a dataset published with the methodological detail (Qin et al., 2013)]. The weighted feature vectors efficiently captured known disease-specific anatomical alterations. For example, the medial temporal lobe and the parietal lobe were negatively weighted in the feature vector of AD

to give a higher discriminant score for AD compared with other diseases and the control group. It should be noted that ventricular enlargement was not emphasized in the feature vector, although it was seen in most of the AD patients. Ventricular enlargement has been regarded as one of the disease-related features in past studies based on a cross-sectional comparison between AD and a control group, but seems to contain less information for separating AD from other neurodegenerative diseases.

FUNCTIONAL MRI PHENOTYPES IN ATLAS COORDINATES

Shown in **Figure 11** are results from functional magnetic resonance imaging done in registered atlas coordinates. Given the accompanying structural T1 images the functional responses can be examined in atlas coordinates with hypotheses formed at the scale of the partition of the atlas. Shown is a comparison between 7 patients with stroke at deep gray matter (cortex is preserved) and age-paired HC. The intensity plot shows the average of Fisher-transformed correlations between the rs-fMRI time courses of each pair of 42 cortical regions in controls (bottom) and individuals with stroke (top). In general, correlations between temporal and frontal areas are the main source of differences between the groups.

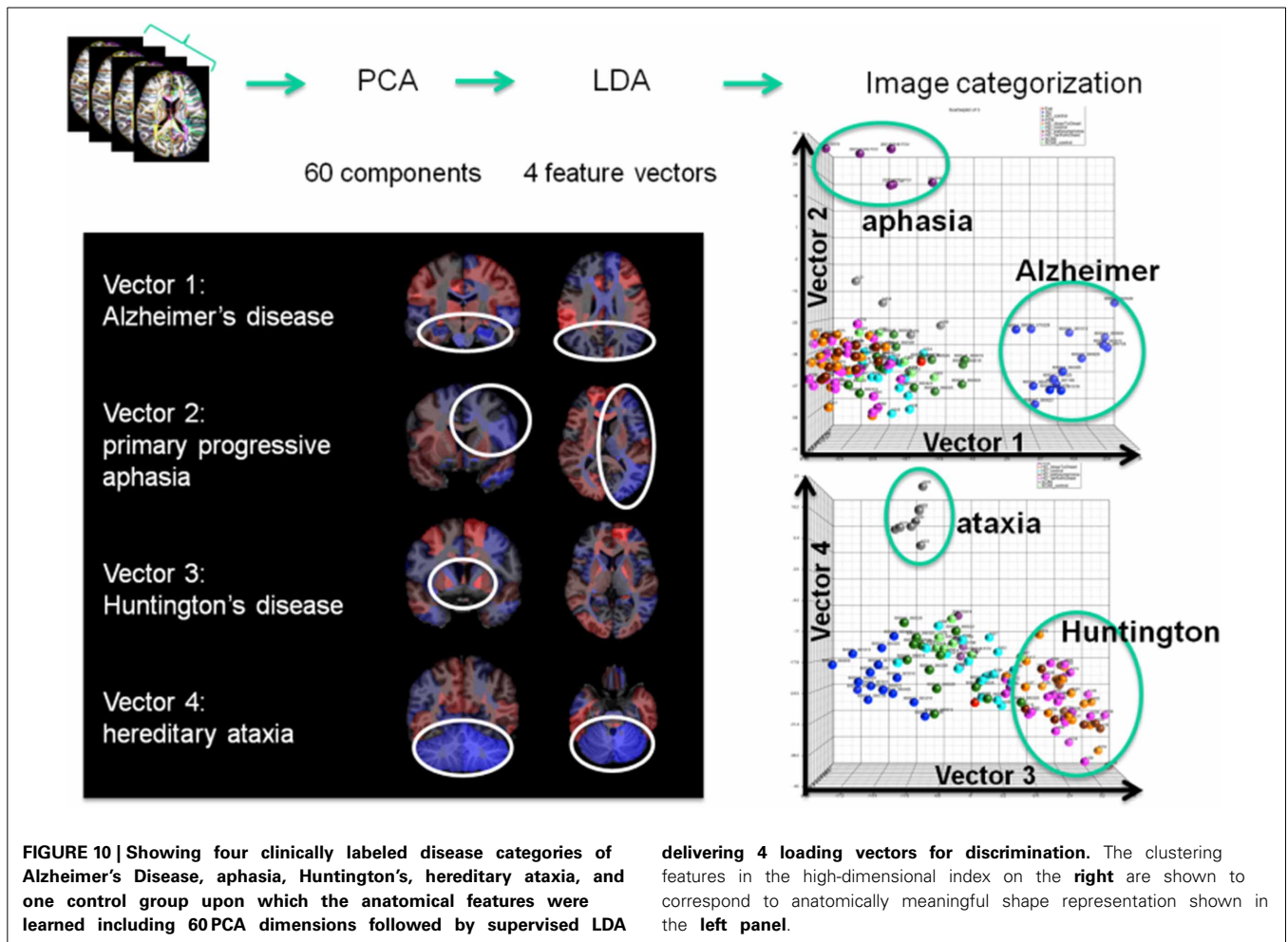


FIGURE 10 | Showing four clinically labeled disease categories of Alzheimer's Disease, aphasia, Huntington's, hereditary ataxia, and one control group upon which the anatomical features were learned including 60 PCA dimensions followed by supervised LDA

delivering 4 loading vectors for discrimination. The clustering features in the high-dimensional index on the right are shown to correspond to anatomically meaningful shape representation shown in the left panel.

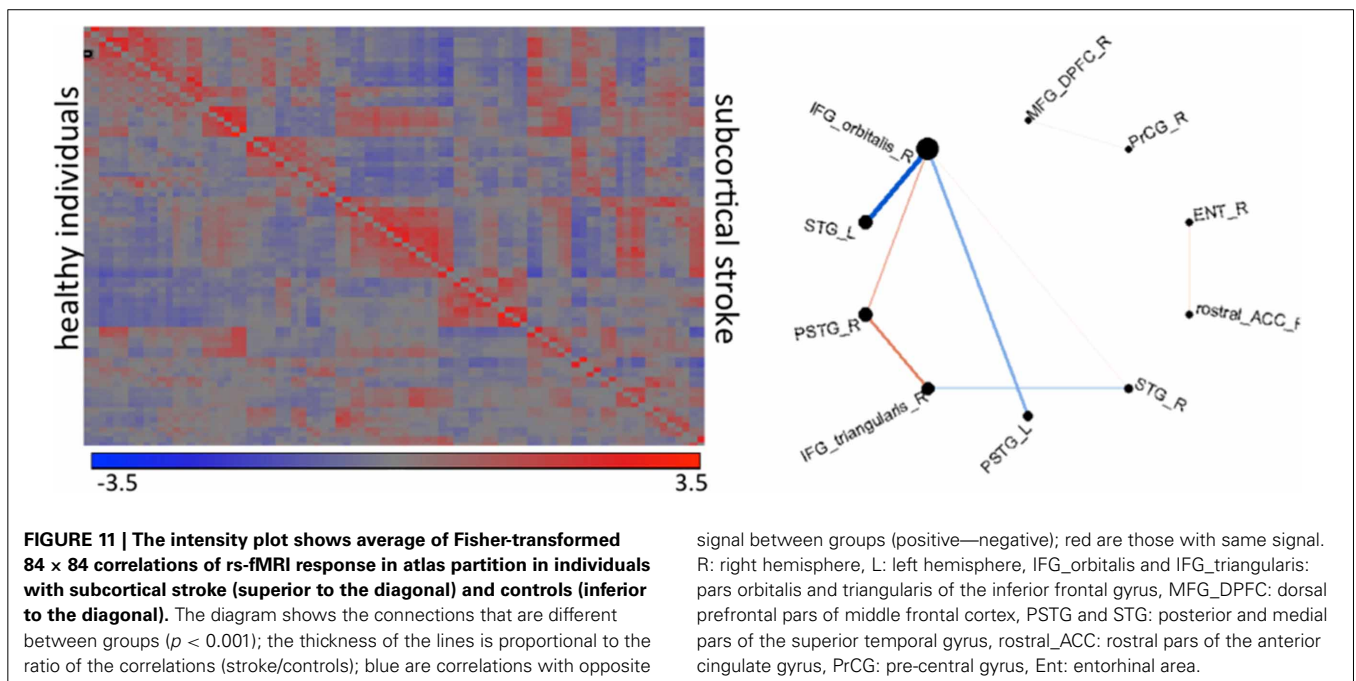


FIGURE 11 | The intensity plot shows average of Fisher-transformed 84 x 84 correlations of rs-fMRI response in atlas partition in individuals with subcortical stroke (superior to the diagonal) and controls (inferior to the diagonal). The diagram shows the connections that are different between groups ($p < 0.001$); the thickness of the lines is proportional to the ratio of the correlations (stroke/controls); blue are correlations with opposite

signal between groups (positive—negative); red are those with same signal. R: right hemisphere, L: left hemisphere, IFG_orbitalis and IFG_triangularis: pars orbitalis and triangularis of the inferior frontal gyrus, MFG_DPF_C: dorsal prefrontal pars of middle frontal cortex, PSTG and STG: posterior and medial parts of the superior temporal gyrus, rostral_ACC_f: rostral parts of the anterior cingulate gyrus, PrCG: pre-central gyrus, Ent: entorhinal area.

CLINICAL INFORMATICS AND BEHAVIOR PHENOTYPES AND FUNCTIONAL PHENOTYPES

Shown in **Figure 12** are results of demonstrating high-throughput informatics used to classify individuals into clinical phenotypes based on functional MRI coupled to clinical behaviors. The 3 clinical variants of PPA (logopenic—Lv, semantic—Sv, and non-fluent—NFv) may differ in terms of disease progression and response to therapeutics. In the early stages of the disease, when some therapeutics are being tested and will hopefully be effective, the clinical tests are not always able to classify all the patients. In addition, although anatomical differences among these variants are reported at group level, the individual classification based on qualitative evaluation is not usually possible. High-throughput imaging informatics can contribute for individual classification. **Figure 12** contains the volumetric data of 120 parcellated areas from 37 PPA patients that were scanned when, in their majority, the variant diagnosis wasn't completely clear, based on clinical information only. Our classification model, created using partial least squares—discriminant analysis (PLS-DA) and volumetric features (120 areas) demonstrated reasonable accuracy on predicting the variant diagnosis with a significant (higher than “by-chance”) *p*-value, both when tested by bootstrapping or by external testing sample. The detection prevalence is low, particularly in the smallest group (NFv) with the sample size needed to be increased.

High-throughput informatics is also an effective tool to scrutinize anatomical-functional/ behavioral correlations. Much of the mapping of brain functions has been via lesion based studies, by relating regions affected by a stroke or trauma, for example, with the functional deficit. Lesion-based studies, however, have significant limitations such as (i) areas most strongly associated with the deficit depend on the vulnerability to ischemia/trauma (ii) determining the part of the lesion which is responsible for the deficit is difficult, or whether it represents a reorganization of cognitive networks that are less efficient, and (iii) the challenge of determining the proportion of changes leading to functional recovery, more than functional loss, (iv) the lack of multiple parameters, local or widespread, that might be concomitantly affected and whose interaction might correlate with the deficit.

Shown in the **Figure 13** is the application of quantitative analysis to assess anatomical-functional correlations in progressive disease models that affect specific functions (such as PPA, that affects primarily language) carried out by investigating the pattern of errors and their relationship to cortical impairment. It shows correlations between regional volumes and PPA patients' performance in a Naming test (Race et al., 2013). This type of anatomical-behavioral analysis provides a better understanding of the relationship between cognitive processes and regions necessary for particular aspects of processing. In more practical terms, we can use this information to monitor the disease progression,

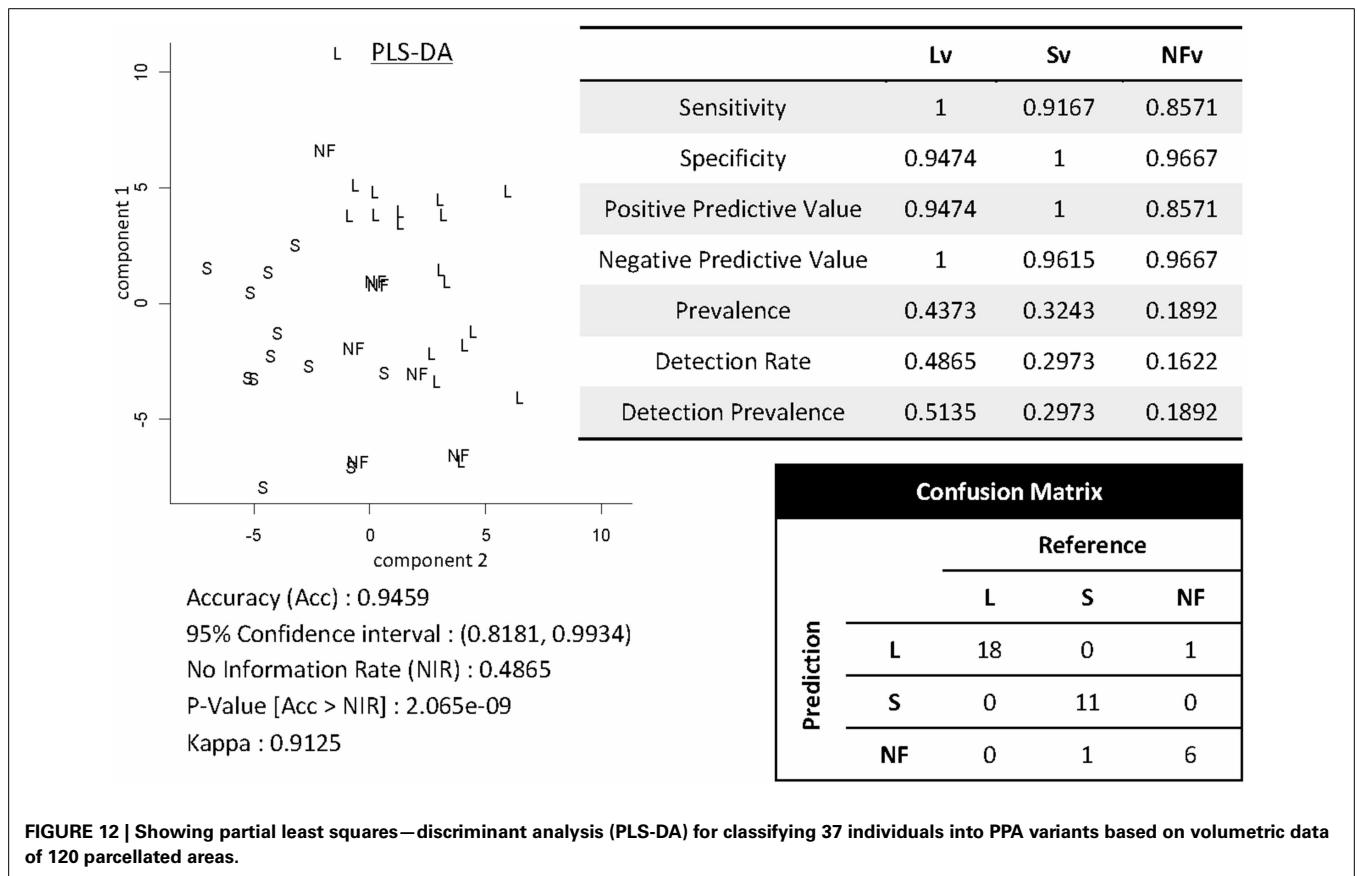
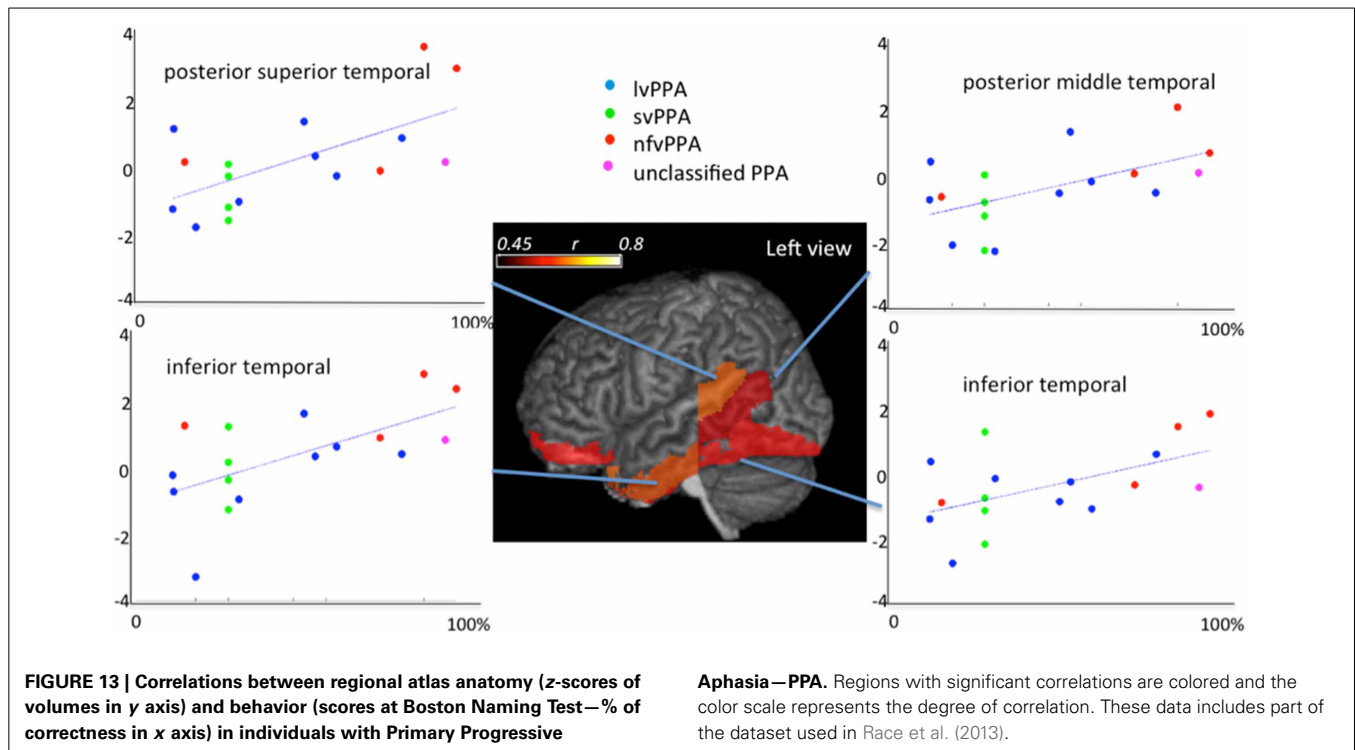


FIGURE 12 | Showing partial least squares—discriminant analysis (PLS-DA) for classifying 37 individuals into PPA variants based on volumetric data of 120 parcellated areas.



or to categorize a clinical entity into more homogeneous groups, which can be meaningful if such subgroups express differences in prognosis or response to various treatments.

DISCUSSION

We have described neuroinformatics technologies at 1 mm anatomical scale based on high-throughput 3D functional and structural imaging technologies of the human brain. The core is the conversion of functional and structural imagery into their high-dimensional neuroinformatic representations index containing $O(1000-10,000)$ discriminating dimensions. The pipeline is based on advanced image analysis coupled to digital knowledge representations in the form of dense atlases of the human brain at gross anatomical scale. We demonstrate the integration of these high-dimensional representations with machine learning methods.

The neuroinformatics pipeline is used to examine cross-sectional and personalized analyses of neuropsychiatric illnesses in clinical applications as well as longitudinal studies. We have demonstrated the use of high-throughput machine learning methods for supporting (i) cross-sectional image analysis to evaluate the health status of individual subjects with respect to the population data, (ii) integration of image and non-image information for diagnosis and prognosis.

ACKNOWLEDGMENTS

We gratefully acknowledge Shoko Yoshida for her efforts in the informatics, and Xioaying Tang for her efforts on the classifiers shown in Figure 5 and Daniel Tward for his work on the geodesic coordinate and variability representation of subcortical structures. This work was supported by grants R01 MH056584, R24 HL085343, R01 EB000975, P50 MH071616, R01 EB001838, P41

EB015909, P41 EB015909, R01 EB008171, R01 MH084803, U01 AG033655, RR025053-01, U01 NS082085-01, R03 EB014357, R21AG033774, R01HD065955, AHA12sdg12080169, and the Yousef Family Research Fund. This effort is being supported by the Computational Anatomy Science Gateway using the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575.

REFERENCES

- Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* 26, 63–72. doi: 10.1523/JNEUROSCI.3874-05.2006
- Albert, M. S., Dekosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 270–279. doi: 10.1016/j.jalz.2011.03.008
- Ardekani, S., Weiss, R. G., Lardo, A. C., George, R. T., Lima, J. A. C., Wu, K. C., et al. (2009). Computational method for identifying and quantifying shape features of human left ventricular remodeling. *Ann. Biomed. Eng.* 37, 1043–1054. doi: 10.1007/s10439-009-9677-2
- Ashburner, J. (2009). Computational anatomy with the SPM software. *Magn. Reson. Imaging* 27, 1163–1174. doi: 10.1016/j.mri.2009.01.006
- Beg, M. F., Miller, M. I., Trounev, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* 61, 139–157. doi: 10.1023/B:VISI.0000043755.93987.a
- Brun, A., and Englund, E. (1986). A white matter disorder in dementia of the Alzheimer type: a pathoanatomical study. *Ann. Neurol.* 19, 253–262. doi: 10.1002/ana.410190306
- Ceritoglu, C. (2008). *Multichannel Large Deformation Diffeomorphic Metric Mapping and Registration of Diffusion Tensor Images*. Ph. D., The Johns Hopkins University.
- Ceritoglu, C., Oishi, K., Li, X., Chou, M. C., Younes, L., Albert, M., et al. (2009). Multi-contrast large deformation diffeomorphic metric

- mapping for diffusion tensor imaging. *Neuroimage* 47, 618–627. doi: 10.1016/j.neuroimage.2009.04.057
- Chevalier-larsen, E., and Holzbaur, E. L. (2006). Axonal transport and neurodegenerative disease. *Biochim. Biophys. Acta* 1762, 1094–1108. doi: 10.1016/j.bbadis.2006.04.002
- Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1996). Deformable templates using large deformation kinematics. *IEEE Trans. Image Process.* 5, 1435–1447. doi: 10.1109/83.536892
- Damoiseaux, J. S., Smith, S. M., Witter, M. P., Sanz-arigita, E. J., Barkhof, F., Scheltens, P., et al. (2009). White matter tract integrity in aging and Alzheimer's disease. *Hum. Brain Mapp.* 30, 1051–1059. doi: 10.1002/hbm.20563
- Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Djavanakova, A., Faria, A. V., Hsu, J., Ceritoglu, C., Oishi, K., Miller, M. I., et al. (2013). Diffeomorphic brain mapping based on T1-weighted images: improvement of registration accuracy by multichannel mapping. *J. Magn. Reson. Imaging*, 37, 76–84. doi: 10.1002/jmri.23790
- Dupuis, P., Grenander, U., and Miller, M. I. (1998). Variational problems on flows of diffeomorphisms for image matching. *Q. Appl. Math.* 56, 587–600.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034
- Englund, E., Brun, A., and Alling, C. (1988). White matter changes in dementia of Alzheimer's type. Biochemical and neuropathological correlates. *Brain* 111(pt 6), 1425–1439. doi: 10.1093/brain/111.6.1425
- Faria, A. V., Joel, S. E., Zhang, Y., Oishi, K., Van zijl, P. C., Miller, M. I., et al. (2012). Atlas-based analysis of resting-state functional connectivity: evaluation for reproducibility and multi-modal anatomy-function correlation studies. *Neuroimage* 61, 613–621. doi: 10.1016/j.neuroimage.2012.03.078
- Fennema-Notestine, C., McEvoy, L. K., Hagler, D. J. Jr., Jacobson, M. W., Dale, A. M., and The Alzheimer's Disease Neuroimaging, I. (2009). Structural neuroimaging in the detection and prognosis of pre-clinical and early AD. *Behav. Neurol.* 21, 3–12. doi: 10.3233/BEN-2009-0230
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–55. doi: 10.1016/S0896-6273(02)00569-X
- Grenander, U. (1993). *General Pattern Theory: A Mathematical Study of Regular Structures*. Oxford, NY, Clarendon: Oxford University Press.
- Grenander, U., and Miller, M. I. (1998). Computational anatomy: an emerging discipline. *Q. Appl. Math.* 56, 617–694.
- Grenander, U., and Miller, M. I. (2007). *Pattern Theory: From Representation to Inference*. Oxford, NY: Oxford University Press.
- Gunawardena, S., and Goldstein, L. S. (2001). Disruption of axonal transport and neuronal viability by amyloid precursor protein mutations in *Drosophila*. *Neuron* 32, 389–401. doi: 10.1016/S0896-6273(01)00496-2
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6:e159. doi: 10.1371/journal.pbio.0060159
- Hammers, A., Allom, R., Koepp, M. J., Free, S. L., Myers, R., Lemieux, L., et al. (2003). Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum. Brain Mapp.* 19, 224–47. doi: 10.1002/hbm.10123
- He, Y., Wang, J., Wang, L., Chen, Z. J., Yan, C., Yang, H., et al. (2009). Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PLoS ONE* 4:e5226. doi: 10.1371/journal.pone.0005226
- Hinrichs, C., Singh, V., Xu, G., and Johnson, S. C. (2011). Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *Neuroimage* 55, 574–589. doi: 10.1016/j.neuroimage.2010.10.081
- Kantarci, K., Jack, C. R. Jr., Xu, Y. C., Campeau, N. G., O'Brien, P. C., Smith, G. E., et al. (2001). Mild cognitive impairment and Alzheimer disease: regional diffusivity of water. *Radiology* 219, 101–107. doi: 10.1148/radiology.219.1.r01ap14101
- Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131. doi: 10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8
- Ma, J., Miller, M. I., Trounev, A., and Younes, L. (2008). Bayesian template estimation in computational anatomy. *Neuroimage* 42, 252–261. doi: 10.1016/j.neuroimage.2008.03.056
- Ma, J., Miller, M. I., and Younes, L. (2010). A bayesian generative model for surface template estimation. *Int. J. Biomed. Imaging* 2010. doi: 10.1155/2010/974957
- Mai, J., Assheuer, J., and Paxinos, G. (1997). *Atlas of the Human Brain*. San Diego, CA: Academic.
- Maldjian, J. A., Laurienti, P. J., Kraft, R. A., and Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239. doi: 10.1016/S1053-8119(03)00169-1
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., et al. (2001). A probabilistic atlas and reference system for the human brain: international Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1293–1322. doi: 10.1098/rstb.2001.0915
- Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., and Lancaster, J. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). *Neuroimage* 2, 89–101. doi: 10.1006/nimg.1995.1012
- Medina, D., Detoleto-Morrell, L., Urresta, F., Gabrieli, J. D., Moseley, M., Fleischman, D., et al. (2006). White matter changes in mild cognitive impairment and AD: a diffusion tensor imaging study. *Neurobiol. Aging* 27, 663–672. doi: 10.1016/j.neurobiolaging.2005.03.026
- Meier-Ruge, W., Ulrich, J., Bruhlmann, M., and Meier, E. (1992). Age-related white matter atrophy in the human brain. *Ann. N.Y. Acad. Sci.* 673, 260–279. doi: 10.1111/j.1749-6632.1992.tb27462.x
- Meyer, J. W., Makris, N., Bates, J. F., Caviness, V. S., and Kennedy, D. N. (1999). MRI-Based topographic parcellation of human cerebral white matter. *Neuroimage* 9, 1–17. doi: 10.1006/nimg.1998.0383
- Miller, M. I., and Younes, L. (2001). Group actions, homeomorphisms, and matching: a general framework. *Int. J. Comput. Vis.* 41, 61–84. doi: 10.1023/A:1011161132514
- Miller, M. I., Trounev, A., and Younes, L. (2002). On the metrics and Euler-Lagrange equations of computational anatomy. *Annu. Rev. Biomed. Eng.* 4, 375–405. doi: 10.1146/annurev.bioeng.4.092101.125733
- Miller, M. I., Trounev, A., and Younes, L. (2006). Geodesic shooting for computational anatomy. *J. Math. Imaging Vis.* 24, 209–228. doi: 10.1007/s10851-005-3624-0
- Miller, M. I., Younes, L., Ratnanather, J. T., Brown, T., Trinh, H., Postell, E., et al. (2013a). The diffeomorphometry of temporal lobe structures in preclinical Alzheimer's disease. *Neuroimage* 3, 352–360. doi: 10.1016/j.nicl.2013.09.001
- Miller, M. I., Younes, L., and Trounev, A. (2013b). Diffeomorphometry and geodesic positioning systems for human anatomy. *Technology* 2. Doi: 10.1142/S2339547814500010
- Mori, S., Oishi, K., Faria, A. V., and Miller, M. I. (2013). Atlas-based neuroinformatics via mri: harnessing information from past clinical cases and quantitative image analysis for patient care. *Annu. Rev. Biomed. Eng.* 15, 71–92. doi: 10.1146/annurev-bioeng-071812-152335
- Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., et al. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage* 40, 570–582. doi: 10.1016/j.neuroimage.2007.12.035
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., et al. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement.* 1, 55–66. doi: 10.1016/j.jalz.2005.06.003
- Oishi, K., Akhter, K., Mielke, M., Ceritoglu, C., Zhang, J., Jiang, H., et al. (2011a). Multi-modal MRI analysis with disease-specific spatial filtering: initial testing to predict mild cognitive impairment patients who convert to Alzheimer's disease. *Front. Neurol.* 2:54. doi: 10.3389/fneur.2011.00054
- Oishi, K., Mielke, M. M., Albert, M., Lyketsos, C. G., and Mori, S. (2011b). DTI analyses and clinical applications in Alzheimer's disease. *J. Alzheimers Dis.* 26(Suppl. 3), 287–296. doi: 10.3233/JAD-2011-0007
- Oishi, K., Mori, S., Donohue, P. K., Ernst, T., Anderson, L., Buchthal, S., et al. (2011c). Multi-contrast human neonatal brain atlas: application to normal neonate development analysis. *Neuroimage* 56, 8–20. doi: 10.1016/j.neuroimage.2011.01.051

- Oishi, K., Faria, A., Jiang, H., Li, X., Akhter, K., Zhang, J., et al. (2009). Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants. *Neuroimage* 46, 486–499. doi: 10.1016/j.neuroimage.2009.01.002
- Oishi, K., Faria, A. V., Yoshida, S., Chang, L., and Mori, S. (2013). Quantitative evaluation of brain development using anatomical MRI and diffusion tensor imaging. *Int. J. Dev. Neurosci.* 31, 512–524. doi: 10.1016/j.ijdevneu.2013.06.004
- Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., et al. (2008). Human brain white matter atlas: identification and assignment of common anatomical structures in superficial white matter. *Neuroimage* 43, 447–457. doi: 10.1016/j.neuroimage.2008.07.009
- Pennec, X. (2009). Statistical computing on manifolds: from Riemannian geometry to computational anatomy. *Emerg. Trends Visual Comput.* 5416, 347–386. doi: 10.1007/978-3-642-00826-9_16
- Pigino, G., Morfini, G., Pelsman, A., Mattson, M. P., Brady, S. T., and Busciglio, J. (2003). Alzheimer's presenilin 1 mutations impair kinesin-based axonal transport. *J. Neurosci.* 23, 4499–4508.
- Qin, Y. Y., Hsu, J. T., Yoshida, S., Faria, A. V., Oishi, K., Unschuld, P. G., et al. (2013). *Gross Feature Recognition of Anatomical Images based on Atlas Grid (GAIA): Using the Degree of Local Atlas-Image Segmentation Disagreement to Capture the Features of Anatomic Brain MRI*. NeuroImage: Clinical, In Press.
- Qiu, A., and Miller, M. I. (2008). Multi-structure network shape analysis via normal surface momentum maps. *Neuroimage* 42, 1430–1438. doi: 10.1016/j.neuroimage.2008.04.257
- Qiu, A. Q., Fennema-Notestine, C., Dale, A. M., Miller, M. I., and Neuroimaging, A. S. D. (2009). Regional shape abnormalities in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 45, 656–661. doi: 10.1016/j.neuroimage.2009.01.013
- Race, D. S., Tsapkini, K., Crinion, J., Newhart, M., Davis, C., Gomez, Y., et al. (2013). An area essential for linking word meanings to word forms: evidence from primary progressive Aphasia. *Brain Lang.* 127, 167–176. doi: 10.1016/j.bandl.2013.09.004
- Ringman, J. M., O'Neill, J., Geschwind, D., Medina, L., Apostolova, L. G., Rodriguez, Y., et al. (2007). Diffusion tensor imaging in preclinical and presymptomatic carriers of familial Alzheimer's disease mutations. *Brain* 130, 1767–1776. doi: 10.1093/brain/awm102
- Rose, S. E., Chen, F., Chalk, J. B., Zelaya, F. O., Strugnell, W. E., Benson, M., et al. (2000). Loss of connectivity in Alzheimer's disease: an evaluation of white matter tract integrity with colour coded MR diffusion tensor imaging. *J. Neurol. Neurosurg. Psychiatry* 69, 528–530. doi: 10.1136/jnnp.69.4.528
- Salat, D. H., Tuch, D. S., Van der Kouwe, A. J., Greve, D. N., Pappu, V., Lee, S. Y., et al. (2010). White matter pathology isolates the hippocampal formation in Alzheimer's disease. *Neurobiol. Aging* 31, 244–256. doi: 10.1016/j.neurobiolaging.2008.03.013
- Sarkisov, S. A., Filimonoff, I. N., Kononowa, E. P., Preobrachenskaja, I. S., and Kukuev, L. A. (1955). *Atlas of the Cytoarchitectonics of the Human Cerebral Cortex*. Moscow: Medgiz.
- Schleicher, A., Amunts, K., Geyer, S., Morosan, P. and Zilles, K. (1999). Observer-independent method for microstructural parcellation of cerebral cortex: a quantitative approach to cytoarchitectonics. *Neuroimage* 9, 165–177. doi: 10.1006/nimg.1998.0385
- Sexton, C. E., Kalu, U. G., Filippini, N., Mackay, C. E., and Ebmeier, K. P. (2011). A meta-analysis of diffusion tensor imaging in mild cognitive impairment and Alzheimer's disease. *Neurobiol. Aging* 32, 2322.e5–2322.e18. doi: 10.1016/j.neurobiolaging.2010.05.019
- Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., et al. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 39, 1064–1080. doi: 10.1016/j.neuroimage.2007.09.031
- Sjoberg, M., Haglund, M., and Englund, E. (2005). Decreasing myelin density reflected increasing white matter pathology in Alzheimer's disease—a neuropathological study. *Int. J. Geriatr. Psychiatry* 20, 919–926. doi: 10.1002/gps.1384
- Stahl, R., Dietrich, O., Teipel, S. J., Hampel, H., Reiser, M. F., and Schoenberg, S. O. (2007). White matter damage in Alzheimer disease and mild cognitive impairment: assessment with diffusion-tensor MR imaging and parallel imaging techniques. *Radiology* 243, 483–492. doi: 10.1148/radiol.2432051714
- Stokin, G. B., Lillo, C., Falzone, T. L., Bruschi, R. G., Rockenstein, E., Mount, S. L., et al. (2005). Axonopathy and transport deficits early in the pathogenesis of Alzheimer's disease. *Science* 307, 1282–1288. doi: 10.1126/science.1105681
- Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: an Approach to Cerebral Imaging*. New York, NY: Thieme.
- Tang, X., Holland, D., Dale, A. M., Younes, L., Miller, M. I., and ADNI (2013a). Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and Alzheimer's disease: detecting, quantifying, and predicting. *Hum. Brain Mapp.*
- Tang, X., Oishi, K., Faria, A. V., Hillis, A. E., Albert, M. S., Mori, S., et al. (2013b). Bayesian Parameter Estimation and Segmentation in the Multi-Atlas Random Orbit Model. *PLoS ONE* 8:e65591. doi: 10.1371/journal.pone.0065591
- Teipel, S. J., Stahl, R., Dietrich, O., Schoenberg, S. O., Perneckzy, R., Bokde, A. L., et al. (2007). Multivariate network analysis of fiber tract integrity in Alzheimer's disease. *Neuroimage* 34, 985–995. doi: 10.1016/j.neuroimage.2006.07.047
- Thompson, P. M., and Toga, A. W. (2002). A framework for computational anatomy. *Comput. Visual Sci.* 5, 13–34. doi: 10.1007/s00791-002-0084-6
- Toga, A. W., and Thompson, P. M. (2001). Maps of the brain. *Anat. Rec.* 265, 37–53. doi: 10.1002/ar.1057
- Trouvé, A., and Younes, L. (2005). Metamorphoses through lie Group Action. *Foundations of Computational Mathematics* 5, 173–198. doi: 10.1007/s10208-004-0128-z
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- von Economo, C., and Koskinas, G. N. (1925). *Die Cytoarchitektonik der Hirnrinde des Erwachsenen Menschen*. Berlin: Springer.
- Wang, J., Wang, L., Zang, Y., Yang, H., Tang, H., Gong, Q., et al. (2009). Parcellation-dependent small-world brain functional networks: a resting-state fMRI study. *Hum. Brain Mapp.* 30, 1511–1523. doi: 10.1002/hbm.20623
- Younes, L., Ratnanather, J. T., Brown, T., Aylward, E., Nopoulos, P., Johnson, H., et al. (2012). Regionally selective atrophy of subcortical structures in prodromal HD as revealed by statistical shape analysis. *Hum. Brain Mapp.* doi: 10.1002/hbm.22214
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867. doi: 10.1016/j.neuroimage.2011.01.008
- Zhou, Y., Dougherty, J. H. Jr., Hubner, K. F., Bai, B., Cannon, R. L., and Hutson, R. K. (2008). Abnormal connectivity in the posterior cingulate and hippocampus in early Alzheimer's disease and mild cognitive impairment. *Alzheimers Dement.* 4, 265–270. doi: 10.1016/j.jalz.2008.04.006
- Zilles, K., Palomero-Gallagher, N., Grefkes, C., Scheperjans, F., Boy, C., Amunts, K., et al. (2002). Architectonics of the human cerebral cortex and transmitter receptor fingerprints: reconciling functional neuroanatomy and neurochemistry. *Eur. Neuropsychopharmacol.* 12, 587–599. doi: 10.1016/S0924-977X(02)00108-6

Conflict of Interest Statement: Susumu Mori and Michael I. Miller own Anatomy Works, with Susumu Mori serving as its CEO. This arrangement is being managed by Johns Hopkins University in accordance with its conflict of interest policies. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 September 2013; accepted: 04 November 2013; published online: 17 December 2013.

Citation: Miller MI, Faria AV, Oishi K and Mori S (2013) High-throughput neuro-imaging informatics. *Front. Neuroinform.* 7:31. doi: 10.3389/fninf.2013.00031

This article was submitted to the journal *Frontiers in Neuroinformatics*.

Copyright © 2013 Miller, Faria, Oishi and Mori. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.