

# Multimodal Hyper-connectivity Networks for MCI Classification

Yang Li<sup>1</sup>, Xinqiang Gao<sup>1</sup>, Biao Jie<sup>2</sup>, Pew-Thian Yap<sup>3</sup>,  
Min-jeong Kim<sup>3</sup>, Chong-Yaw Wee<sup>4</sup>(✉), and Dinggang Shen<sup>3</sup>(✉)

<sup>1</sup> Department of Automation Science and Electrical Engineering,  
Beihang University, Beijing, China

{liyang, gaoxq}@buaa.edu.cn

<sup>2</sup> Department of Computer Science and Technology,  
Anhui Normal University, Wuhu, China

jbiao@nuaa.edu.cn

<sup>3</sup> Department of Radiology and BRIC,  
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

{ptyap, mjkim, dgshen}@med.unc.edu

<sup>4</sup> Department of Biomedical Engineering, National University of Singapore,  
Singapore, Singapore

cywee2000@gmail.com

**Abstract.** Hyper-connectivity network is a network where every edge is connected to more than two nodes, and can be naturally denoted using a hyper-graph. Hyper-connectivity brain network, either based on structural or functional interactions among the brain regions, has been used for brain disease diagnosis. However, the conventional hyper-connectivity network is constructed solely based on single modality data, ignoring potential complementary information conveyed by other modalities. The integration of complementary information from multiple modalities has been shown to provide a more comprehensive representation about the brain disruptions. In this paper, a novel multimodal hyper-network modelling method was proposed for improving the diagnostic accuracy of mild cognitive impairment (MCI). Specifically, we first constructed a multimodal hyper-connectivity network by simultaneously considering information from diffusion tensor imaging and resting-state functional magnetic resonance imaging data. We then extracted different types of network features from the hyper-connectivity network, and further exploited a manifold regularized multi-task feature selection method to jointly select the most discriminative features. Our proposed multimodal hyper-connectivity network demonstrated a better MCI classification performance than the conventional single modality based hyper-connectivity networks.

## 1 Introduction

Hyper-connectivity brain network is a network where each edge is connected to more than two brain regions, which can be naturally denoted using a hyper-graph. Hyper-connectivity network, either based on structural or functional interactions among the brain regions, has been used for brain disease diagnosis [1]. Functional interactions and

structural interactions can be extracted from functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI), respectively [2]. However, the conventional hyper-network, which is constructed solely based on single modality data, ignores the potential complementary information conveyed by other modalities. Integration of complementary information from different modalities has been shown to provide a more comprehensive representation on the brain structural and functional organizations [3, 4]. Inspired by this observation, classification framework based on multimodal brain networks constructed from resting-state fMRI (rs-fMRI) and DTI has been proposed to enhance the classification performance of mild cognitive impairment (MCI) [5].

In this paper, we proposed the first multimodal hyper-connectivity network modelling method that simultaneously considers the information from rs-fMRI and DTI data during the network construction. Specifically, the multimodal hyper-connectivity network was constructed using a star expansion method [6] based on the anatomically weighted functional distance between pairs of brain regions. The anatomically weighted functional distance, which is defined as the strength of the anatomically weighted functional connectivity (awFC) [7], was computed using the complementary information conveyed by the rs-fMRI and DTI data. We then extracted network features from the constructed hyper-connectivity network, and selected the most discriminative features using a manifold regularized multi-task feature selection method (M2TFS) [1]. Finally, we applied a support vector machine (SVM) on the selected features for MCI classification. Promising classification results demonstrated the superiority of the proposed multimodal hyper-connectivity network over the single-modal hyper-connectivity networks which were constructed either from rs-fMRI or DTI data.

## 2 Materials and Methodology

### 2.1 Dataset

Ten MCI patients (5M/5F) and 17 normal controls (8M/9F) were included in this study with informed consent obtained from all participants, and the experimental protocols were approved by the institutional ethics board. The mean age for MCI and control groups are  $74.2 \pm 8.6$  and  $72.1 \pm 8.2$  (years), respectively. All the subjects were scanned using a 3.0-Tesla scanner to acquire the rs-fMRI and DTI data. The acquisition parameters for rs-fMRI were as follows: repetition time (TR) = 2000 ms, echo time (TE) = 32 ms, flip angle =  $77^\circ$ , acquisition matrix =  $64 \times 64$ , voxel size = 4 mm. One hundred fifty fMRI volumes were acquired. During the scanning, all subjects were instructed to keep their eyes open and stare at a fixation cross in the middle of the screen, which lasted for 5 min. The acquisition parameters for DTI were as follows:  $b = 0$  and  $1000 \text{ s/mm}^2$ , flip angle =  $90^\circ$ , TR/TE = 17000/78 ms, imaging matrix =  $128 \times 128$ , FOV =  $256 \times 256 \text{ mm}^2$ , voxel thickness = 2 mm, and 72 continuous slices.

## 2.2 Data Preprocessing

Resting-state fMRI images were preprocessed using Statistical Parametric Mapping software package (SPM8). Specifically, the first 10 fMRI volumes were removed before parcellating the brain space into 116 regions-of-interest (ROIs) based on the automated anatomical labeling (AAL) [8] template. We averaged the fMRI time series over all voxels in each ROI to compute the mean fMRI time series. Prior to constructing the hyper-connectivity network, a temporal band-pass filtering with frequency interval ( $0.025 \leq f \leq 0.100$  Hz) was applied to the mean time series of each individual ROI to reduce the effects of physiological and measurement noises. Following previous study, global signal regression was not performed due to its controversy in the rs-fMRI preprocessing procedure [9].

Similar to the fMRI preprocessing, DTI images were aligned to the AAL template space using a deformable DTI registration algorithm (F-TIMER) [10] before the parcellating the brain space into 116 ROIs. A whole-brain streamline fiber tractography was then applied on each image using ExploreDTI [11] with the minimal seed point fractional anisotropy (FA) of 0.45, stopping FA of 0.25, minimal fiber length of 20 mm, and maximal fiber length of 400 mm.

## 2.3 Methods

**Anatomically Weighted Functional Distance.** We proposed a novel multimodal hyper-connectivity network modelling method that simultaneously utilizes the information from rs-fMRI and DTI data. Our method is based on the anatomically weighted functional distance which reflects the evidence for the underlying DTI data to supplement the fMRI data as defined below [7]

$$awFD_{ij} = \left(1 - \frac{\pi_{ij}}{\lambda}\right)FD_{ij} \quad (1)$$

where  $\pi_{ij} \in [0,1)$  is the strength of DTI-based structural connectivity between the brain regions  $i$  and  $j$ ,  $\lambda \in [1,\infty) = \Omega$  is an unknown parameter that potentially attenuates the anatomically weighting, and  $FD_{ij}$  is the functional distance between the fMRI profiles. Equation (1) explicitly incorporates the brain anatomy for guiding a more accurate inference of the functional connectivity between two brain regions. Following the premise that structural connection is neither a sufficient nor necessary condition for the functional connection [7], a parameter  $\lambda$  was imposed in Eq. (1) to regulate the contribution of the structural connection especially for the case where no fibers connect two regions. The functional distance between the fMRI profiles of ROIs  $i$  and  $j$  at lag- $o$  is defined as [7]

$$FD_{ij} = 1 - \min_{o \in O} FD_{ij}(o) = 1 - \min_{o \in O} \left\{ \frac{\sum_{t=1}^{T-o} [x_i(t+o) - \bar{x}_i][x_j(t) - \bar{x}_j]}{\hat{\sigma}_i \hat{\sigma}_j} \right\} \quad (2)$$

where  $x_i(t)$  denotes the fMRI time series of the ROI  $i$  at time  $t$ ,  $T$  is the total number of rs-fMRI volumes,  $\hat{\sigma}_i$  and  $\hat{\sigma}_j$  denote the standard deviations of samples  $x_i$  and  $x_j$ ,  $\bar{x}_i$  and  $\bar{x}_j$

indicate the sample means of  $x_i$  and  $x_j$ , respectively. For the ease of explanation, we considered only the positive correlation. In view of the potential differences in the hemodynamic responses of resting-state neuronal activity between different brain regions, we estimated the functional distance with a few lagging  $o$  in  $O = [-3, 3]$  and obtained the minimum lag- $o$  distance [7].

The structural distance, which represents the strength of the DTI-based structural connectivity between pairs of ROIs, is defined as [7]

$$SD_{ij} = (1 - \frac{\pi_{ij}}{\lambda}) \quad (3)$$

where  $\pi_{ij}$ , which is the average on-fiber FA, denotes the strength of structural connection between ROIs  $i$  and  $j$ , and  $\lambda$  denotes an unknown parameter that potentially reduces the effect of structural data. The indirect structural connections were allowed by defining  $\pi_{ij} = \max[\pi_{ij}, \max_l(\pi_{il}, \pi_{lj})]$  [7]. The optimal  $\lambda$  was determined empirically through minimizing the impact of false positive structural connectivity [7].

**Hyper-graph Construction.** We employed a multimodal hyper-graph construction method to estimate the anatomically weighted functional distance. Let  $V$  be the vertex set and  $E$  the hyper-edge set of a hyper-graph  $G$ . For the  $n$ -th subject with  $P$  ROIs, a hyper-graph  $G_n = (V_n, E_n)$  with  $P$  vertices can be constructed with each of its vertices representing an ROI. We employed a star expansion method [6] to generate hyper-edges among vertices. Specifically, for each distance matrix, a vertex was first selected as the centroid vertex and a hyper-edge was then constructed by linking the centroid vertex to its nearest neighbors within  $\varphi \bar{d}$  distance [6]. Here,  $\bar{d}$  is the average anatomically weighted distance between regions and  $\varphi$ , which was set to 0.78 via grid search on training data, is a hyper-parameter controlling the sparsity of the hyper-network. It is noteworthy that the constructed hyper-edges were non-weighted edges.

**Feature Extraction and Selection.** Topological properties derived from a hyper-connectivity network provide quantitative measures to effectively study the differences in terms of brain organization between MCI subjects and normal controls (NC). In this study, we extracted three different types of clustering coefficients from the constructed multimodal hyper-connectivity network. Given a multimodal hyper-network  $G = (V, E)$ , let  $M(v)$  be the hyper-edges adjacent to the vertex  $v$ , i.e.,  $M(v) = \{e \in E : v \in e\}$ , and  $N(v)$  the neighboring vertices to  $v$ , i.e.,  $N(v) = \{u \in V : \exists e \in E, u, v \in e\}$ . Then, three different types of clustering coefficients [1] can be computed on the vertex  $v$  as

$$HCC^1(v) = \frac{2 \sum_{u, q \in N(v)} I(u, q, \neg v)}{|N(v)|(|N(v)| - 1)} \quad (4)$$

$$HCC^2(v) = \frac{2 \sum_{u, q \in N(v)} I'(u, q, v)}{|N(v)|(|N(v)| - 1)} \quad (5)$$

$$\text{HCC}^3(v) = \frac{2 \sum_{e \in M(v)} (|e| - 1) - |N(v)|}{|N(v)|(|M(v)| - 1)} \quad (6)$$

where  $u, q, v \in V$  and  $e \in E$ ,  $I(u, q, -v) = 1$  if there exists  $e \in E$  such that  $u, q \in e$  but  $v \notin e$ , and 0 otherwise.  $I'(u, q, v) = 1$  if there exists  $e \in E$  such that  $u, q, v \in e$ , and 0 otherwise. Three types of clustering coefficient features represent the topological properties of the multimodal hyper-connectivity network from three different perspectives. Specifically, the  $\text{HCC}^1$  denotes the number of neighboring nodes that have connections not facilitated by node  $v$ . In contrast, the  $\text{HCC}^2$  denotes the number of neighboring nodes with connections facilitated by node  $v$ , giving that these nodes may share some brain functions with each other and node  $v$ . The  $\text{HCC}^3$  denotes the amount of overlap among adjacent hyper-edges of node  $v$ . We jointly selected features from these three types of clustering coefficients using a manifold regularized multi-task feature selection method (M2TFS) defined as [1]

$$\min_W \frac{1}{2} \sum_{c=1}^C \|Y - Z^c w^c\|_2^2 + \beta \sum_{c=1}^C (Z^c w^c)^T L^c (Z^c w^c) + \gamma \|W\|_{2,1} \quad (7)$$

$$S^c(n, m) = \exp\left(-\|z_n^c - z_m^c\|^2/h\right) \quad (8)$$

where  $Z^c = [z_1^c, \dots, z_n^c, \dots, z_N^c]^T \in R^{N \times P}$  denotes a set of features from a total of  $N$  training subjects, each with  $P$  regions, and  $z_n^c = [\text{HCC}^c(v_i)]_{i=1:P} \in R^P$  is the vector of clustering coefficients from the  $n$ -th training subject on task  $c$  (in our case, a task represents selecting features from one type of clustering coefficients),  $Y = [y_1, \dots, y_n, \dots, y_N]^T \in R^N$  is the response vector for those  $N$  training subjects, where  $y_n$  is the class label for the  $n$ -th training subject.  $L^c = D^c - S^c$  is the combinatorial Laplacian matrix on task  $c$ .  $S^c$  is a matrix that describes the similarity on the  $c$ -th task across training subjects, where  $D^c$  is a diagonal matrix defined as  $D^c(n, n) = \sum_{m=1}^N S^c(n, m)$ .  $W = [w^1, w^2, \dots, w^C] \in R^{P \times C}$  is a weight matrix with  $C$  being the total number of tasks (i.e.,  $C = 3$ ), and  $\|W\|_{2,1} = \sum_{i=1}^P \|w_i\|_2$  is the group sparsity regularizer that encourages features from different tasks to be jointly selected. Here,  $w_i$  is the  $i$ -th row vector of  $W$ .  $\beta$  and  $\gamma$  are the corresponding regularization coefficients.  $h$  is a free parameter to be tuned empirically. The values of  $h$ ,  $\beta$  and  $\gamma$  can be determined via inner cross-validation on the training subjects.

**Classification.** We employed a multi-kernel SVM to fuse three types of clustering coefficient features for MCI classification. Specifically, let  $f_n^c$  be the selected features from the  $c$ -th task of the  $n$ -th subject. We computed a linear kernel based on the features selected by the M2TFS method for each type of clustering coefficients and then fused them via a multi-kernel technique given as follows:

$$k(f_n, f_m) = \sum_{c=1}^C \mu^c k^c(f_n^c, f_m^c) \quad (9)$$

where  $k^c(f_n^c, f_m^c)$  denotes the linear kernel function between the  $n$ -th and  $m$ -th subjects for the  $c$ -th set of selected clustering coefficients, and  $\mu^c$  is a non-negative weight coefficient with  $\sum_{c=1}^C \mu^c = 1$ . A coarse-grid search was used to optimize  $\mu^c$  through a nested cross-validation on the training subjects.

### 3 Experiment Results

Due to the limited sample size, we employed a nested leave-one-out cross-validation (LOOCV) scheme to evaluate the performance and generalization power of our proposed method. In the inner LOOCV loop, the training data was used to optimize the parameters  $h$ ,  $\beta$  and  $\gamma$  that identify a set of the most discriminative features for classification. To determine the weights  $\mu^c$  for integrating multiple kernels, we used a grid search with the range [0, 1] at a step size of 0.1.

The proposed method was compared to three single-modal models, i.e., hyper-networks derived either from fMRI or DTI data individually and also hyper-networks constructed from fMRI using sparse representation (fMRI-SR) [1]. Multiple values of the regularization parameter that determines the sparsity level of hyper-networks in the fMRI-SR model were set to [0.1, 0.2, ..., 0.9]. As shown in Table 1, the proposed method yielded an accuracy of 96.3%, which is 7.4% better than the second best performed DTI-based hyper-network model. The fMRI-based hyper-network model performed the worst with an accuracy of 74.1%. The area under receiver operating characteristic curve (AUC) was used to evaluate the generalization performance and the proposed method achieved an AUC of 0.98, indicating an excellent generalization performance.

**Table 1.** Classification performance for four comparison approaches.

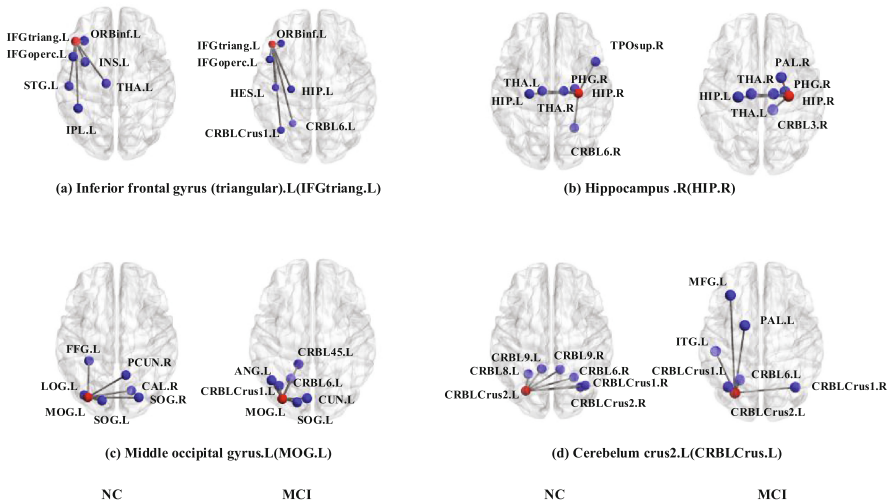
Approach	Accuracy (%)	AUC	SEN (%)	SPE (%)	BAC (%)
fMRI-SR [1]	85.2	0.92	82.4	90.0	86.2
fMRI	74.1	0.87	76.5	70.0	73.2
DTI	88.9	0.94	88.2	90.0	89.1
Proposed	96.3	0.98	100.0	94.1	97.1

As shown in Table 2, there were 11 most discriminative features that were always selected in each LOOCV fold. These brain regions included regions located in the frontal lobes (e.g., left inferior frontal gyrus (triangular) [12] and left rectus gyrus [13]), the temporal lobes (e.g. left temporal pole and middle temporal gyrus [14]), cerebellum, and other regions including hippocampus [14] and occipital gyrus [14]. Our findings are consistent with previous findings that (1) atrophies of regions in the temporal lobe and frontal lobe were found at the early AD [15], and (2) gial

accumulation of redox-active iron in the cerebellum was found significant in preclinical Alzheimer’s disease patients [16]. Figure 1 graphically illustrates the significant differences in terms of hypergraph structure between MCI and NC [1]. For example, in Fig. 1(b), the right hippocampus (HIP.R) was connected to the left hippocampus (HIP.L), left thalamus (THA.L), right thalamus (THA.R), right parahippocampal gyrus (PHG.R), right lenticular nucleus (pallidum) (PAL.R) and right cerebellum 3 (CRBL3.R) in MCI, while it was connected to the left hippocampus (HIP.L), left thalamus (THA.L), right thalamus (THA.R), right parahippocampal gyrus (PHG.R), right temporal pole (superior) (TPOsup.R) and right cerebellum 6 (CRBL6.R) in NC. As the hippocampus is highly associated with the memory performance, this pattern of alteration in functional connectivity involving the hippocampus may provide clues on the underpinnings of cognitive deficit in MCI.

**Table 2.** The most discriminative ROIs that were selected during MCI classification.

No.	ROI	No.	ROI	No.	ROI
1	IFGtraing.L	5	TPOmid.L	9	Vermis1
2	REC.L	6	CRBLCrus1.R	10	Vermis6
3	HIP.L	7	CRBLCrus2.L	11	Vermis8
4	MOG.L	8	CRBL9.L		



**Fig. 1.** The average degree of hyper-edges for NC and MCI for the 4 brain regions listed in Table 2. Each sub-figure represents a hyper-edge between the corresponding brain region (indicated by the red node) and other nodes. The average degree of hyper-edges for a node is computed from the top  $d$  ROIs with the highest occurrence number among all subjects.

## 4 Conclusion

In this paper, we proposed a novel multimodal hyper-network modelling method for improving the diagnostic accuracy of MCI. The proposed hyper-connectivity network encodes complementary information from multiple modalities to provide a more comprehensive representation on the brain structural and functional organizations. We demonstrated the superiority of our proposed method via MCI classification. Compared to the single-modal method, our proposed method achieved a higher classification accuracy and a better generalization performance. In the future, we will evaluate the performance of the proposed method on larger datasets.

## References

1. Jie, B., Wee, C.-Y., Shen, D., Zhang, D.: Hyper-connectivity of functional networks for brain disease diagnosis. *Med. Image Anal.* **32**, 84–100 (2016)
2. Zhu, D., Zhang, T., Jiang, X., Hu, X., Chen, H., Yang, N., Lv, J., Han, J., Guo, L., Liu, T.: Fusing DTI and fMRI data: a survey of methods and applications. *Neuroimage* **102**, 184–191 (2014)
3. Greicius, M.D., Supekar, K., Menon, V., Dougherty, R.F.: Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb. Cortex* **19**, 72–78 (2009)
4. Van den Heuvel, M.P., Mandl, R.C., Kahn, R.S., Pol, H., Hilleke, E.: Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Hum. Brain Mapp.* **30**, 3127–3141 (2009)
5. Wee, C.-Y., Yap, P.-T., Zhang, D., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D.: Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage* **59**, 2045–2056 (2012)
6. Gao, Y., Wee, C.-Y., Kim, M., Giannakopoulos, P., Montandon, M.-L., Haller, S., Shen, D.: MCI identification by joint learning on multiple MRI Data. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9350, pp. 78–85. Springer, Cham (2015). doi:[10.1007/978-3-319-24571-3\\_10](https://doi.org/10.1007/978-3-319-24571-3_10)
7. Bowman, F.D., Zhang, L., Derado, G., Chen, S.: Determining functional connectivity using fMRI data with diffusion-based anatomical weighting. *Neuroimage* **62**, 1769–1779 (2012)
8. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002)
9. Murphy, K., Birn, R.M., Handwerker, D.A., Jones, T.B., Bandettini, P.A.: The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage* **44**, 893–905 (2009)
10. Yap, P.-T., Wu, G., Zhu, H., Lin, W., Shen, D.: F-TIMER: fast tensor image morphing for elastic registration. *IEEE Trans. Med. Imaging* **29**, 1192–1203 (2010)
11. Leemans, A., Jeurissen, B., Sijbers, J., Jones, D.: ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In: *ISMRM*, p. 3537 (2009)
12. Bell-McGinty, S., Lopez, O.L., Meltzer, C.C., Scanlon, J.M., Whyte, E.M., DeKosky, S.T., Becker, J.T.: Differential cortical atrophy in subgroups of mild cognitive impairment. *Arch. Neurol.* **62**, 1393–1397 (2005)



13. Fleisher, A.S., Sherzai, A., Taylor, C., Langbaum, J.B., Chen, K., Buxton, R.B.: Resting-state BOLD networks versus task-associated functional MRI for distinguishing Alzheimer's disease risk groups. *Neuroimage* **47**, 1678–1690 (2009)
14. Salvatore, C., Cerasa, A., Battista, P., Gilardi, M.C., Quattrone, A., Castiglioni, I.: Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Front. Neurosci.* **9**, 307 (2015)
15. Möller, C., Vrenken, H., Jiskoot, L., Versteeg, A., Barkhof, F., Scheltens, P., van der Flier, W.M.: Different patterns of gray matter atrophy in early- and late-onset Alzheimer's disease. *Neurobiol. Aging* **34**, 2014–2022 (2013)
16. Smith, M.A., Zhu, X., Tabaton, M., Liu, G., McKeel Jr., D.W., Cohen, M.L., Wang, X., Siedlak, S.L., Dwyer, B.E., Hayashi, T.: Increased iron and free radical generation in preclinical Alzheimer disease and mild cognitive impairment. *J. Alzheimers Dis.* **19**, 363–372 (2010)