CLUSTERING OF CONTRAST ESTIMATE PATTERNS OF FMRI TO UNTANGLE GENOTYPIC EFFECTS ON WHOLE BRAIN NETWORKS

Kayako Matsuo[1], Chih-Min Liu[2,3], Shen-Hsing Annabel Chen[4], Aleksandr Aleksandrovich Simak[5,6], Chen-Chung Liu[2], Tzung-Jeng Hwang[2,3], Ming-Hsien Hsieh[2], Yi Ling Chien[2], Hai-Gwo Hwu[2,7], and Wen-Yih Isaac Tseng[1,3]

[1]*Advanced Biomedical MRI Lab, National Taiwan University College of Medicine, Taipei, Taipei, Taiwan,* [2]*Department of Psychiatry, National Taiwan University Hospital & College of Medicine, Taipei, Taiwan,* [3]*Graduate Institute of Brain and Mind Sciences, National Taiwan University College of Medicine, Taipei, Taiwan,* [4]*Division of Psychology, School of Humanities and Social Sciences, Nanyang Technological University, Singapore, Singapore,* [5]*Functional Neuroimaging Group, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan,* [6]*Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan,* [7]*Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan*

Target audience: researchers who seek to untangle genotypic effects on whole brain functions.

Purpose: As we lack an efficient methodology to obtain veritable genotypic interaction patterns on whole brain activation, we developed a clustering method of contrast estimate values during fMRI of verbal working memory (VWM) in a data-driven manner and observed genetic/epigenetic interactions related to a new gene, NRG1-P3 (rs56228203), which was recently discovered in Taiwan.

Background: A dysfunction in working memory (WM) is a core cognitive impairment in schizophrenia [1-3]. NRG1 has been identified as a candidate susceptibility gene for schizophrenia [4-5] and has been associated with WM [6]. In statistics, we cannot model "true" patterns of genotypic effects because of the lack of information in advance (Fig. 1).

Methods: A total of 104 people with schizophrenia (SCH) and healthy controls (CON) (52 each) were studied after obtaining written informed consent. The protocol was approved by the National Taiwan University Hospital institutional review board in adherence to the Declaration of Helsinki. Four groups were classified by disease state and genotype: SCH-C (27 people), SCH-TT (25), CON-C (27) and CON-TT (25), where C indicates C-allele carrier (TC/CC) whereas TT indicates TT genotype carrier. Age, handedness score, year of education, reaction time (RT) and accuracy during the VWM performance in the scanner and gender distribution were matched and indicated no significant differences. Age of onset, duration of illness (DOI), chlorpromazine equivalent dose (CPZ) [7] and positive, negative and general scores of PANSS [8] also demonstrated no significant differences between SCH-C and SCH-TT. Scanning was performed using a 3T MR scanner with a 32- or 16-channel phased array coil (Trio Tim, Siemens, Erlangen, Germany). A GRE-EPI sequence was employed for fMRI using the following parameters: TR/TE = 2500 ms/24 ms, flip angle = 90 deg, 43 slices, 3.5 mm thick with no gap interleaved, FOV 240 mm, matrix size 64 x 64, voxel size of 3.75 x 3.75 x 3.5 mm3, and 191 volumes per run (about 8 min). Two runs were administered to each subject. A block-design paradigm of Sternberg VWM was employed with blocks of high (6 letters) or low (1 letter and 5 '#' signs) WM load in alternation for eight cycles [9]. SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) was used for image analyses. After preprocessing, contrast estimates of high>low loads were computed for each subject. A random-effects group statistics was performed including a signal-to-fluctuation-noise ratio (SFNR) as a covariate because of the employment of two types of coils [10]. As a result, we obtained 4 volumes of averaged contrast estimates ('betas') corresponding to each subject group. The 4-value patterns from same coordinates of the 4 groups (180840 voxels) were clustered using a k-means clustering algorithm with correlation as the distance measure. The obtained cluster means were normalized to zero with standard deviation of 1. A cluster-specific SPM was conducted entering the cluster means as contrast definition values. A challenge of the current method was how to define the cluster division number for the main analysis. We performed k-means clustering of 2 to 15 cluster divisions and estimated the results by (1) silhouette coefficient, (2) overlapped voxel number between cluster-specific SPMs ($p<0.005$, $k=20$; [11]) and (3) inner product between cluster means. We decided to employ cluster division 10 in this analysis (Fig. 2). To examine the characteristics of the obtained clusters, we computed a Pearson correlation coefficient between individual summation of contrast estimates within masks of the cluster-specific SPM and the subjects' parameters including age, accuracy and PANSS scores.

Results: We extracted clusters of disease effect (CON>SCH, SCH>CON), a cluster of TT>C and a cluster of additive interaction of CON>SCH and C>TT (Fig. 3). Four clusters (Cl4, 5, 3, 9) demonstrated correlations with age ($p<0.01$); some of them also showed milder correlations with positive and/or general scores in patients ($p<0.05$). Other 2 clusters (Cl6, 8) demonstrated those with accuracy ($p<0.01$).

Discussion: Among the obtained SPMs, those derived from clusters 10, 1 and 2 represented relatively genuine genotypic effects with less influences of age, accuracy and disease state. Specifically, SPM derived from cluster means of the cluster 1 demonstrated a clear TT>C pattern (Fig. 3, 3rd row); the highest peaks were found in the medial prefrontal cortex (mPFC) and the posterior cingulate cortex (PCC), which are core structures of the default mode network (DMN). Genotypic effects are thought to manifest themselves in a network-specific manner in the whole brain but the exact patterns are unknown in advance. The patterns would be influenced by individual parameters such as age and task accuracy (possibly affected by attention ability); these were successfully demonstrated in the cluster-specific SPM results.

Conclusions: The clustering of contrast estimate patterns is a promising data-driven method to investigate genotypic effects on brain networks that are interacted with disease state, age, performance accuracy and other epigenetic factors.

References:

[1] Honey, G.D. et al. Neuroscience 139, 59-71 (2006). [2] Manoach, D.S. Schizophr Res 60, 285-98 (2003). [3] Conklin, H.M. et al. Am J Psychiatry 157, 275-7 (2000). [4] Stefansson, H. et al. Am J Hum Genet 71, 877-92 (2002). [5] Liu, C.M. et al. Am J Med Genet B Neuropsychiatr Genet 134B, 79-83 (2005). [6] Krug, A. et al. Neuroimage 42, 1569-76 (2008). [7] Andreasen, N.C. et al. Biol Psychiatry 67, 255-62 (2010). [8] Kay et al. Schizophr Bull 13, 261-276 (1987). [9] Matsuo, K. et al. Schizophr Res (in press). [10] Friedman et al. Neuroimage 33(2), 471-481 (2006). [11] Lieberman, M.D. et al. Soc Cogn Affect Neurosci 4, 423-8 (2009).

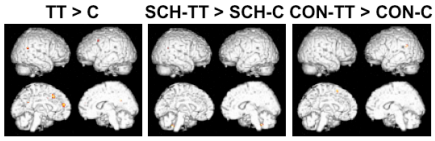

Fig. 1. 3D rendered maps by conventional SPM contrast definitions. N=104. P<0.005, k=20. Upper row, lateral view. Lower row, medial view. Genotypic differences are small and the interactive relationships between disease and genotype are difficult to discern.

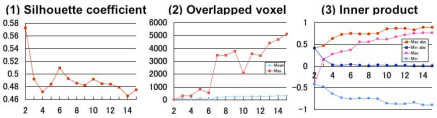

Fig. 2. Examination of cluster division number. Cluster division 10 was chosen because of (1) local maximum of silhouette coefficient, (2) local minimum of overlapped voxel number and (3) sufficiently large absolute inner product between clusters of SCH>CON and CON>SCH (corresponding to Cl9 and Cl7 in Fig. 3).

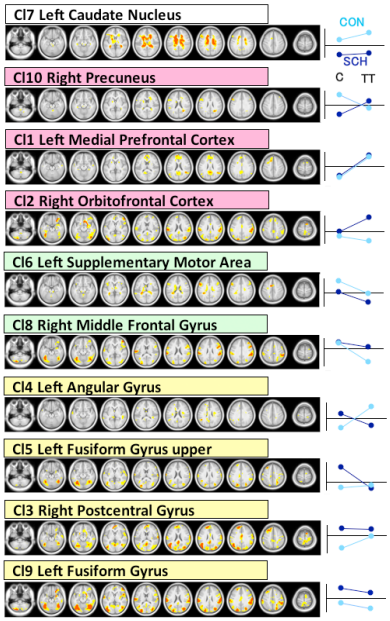

Fig. 3. Cluster-specific SPMs. N=104. In the left side sections, red indicates maps of $p<0.005$, $k=20$ whereas yellow, $p<0.05$. The right side line graphs represent cluster means that were used as contrast definition values; pale blue indicates controls whereas dark blue, patients, and left side indicates C whereas right, TT. Labels above sections indicate peak coordinate anatomy labeling. Yellow labels (Cl4, 5, 3, 9) indicate that contrast estimates were correlated to age whereas pale green labels (Cl6, 8) indicate the correlation to accuracy ($p<0.01$). Cl7 represents a genuine disease effect. SPMs with pink labels (Cl10, 1, 2) were less affected by age, accuracy and disease state.