Support vector machines detect Huntington's gene effects in mouse brain images with >98% accuracy

Stephen J Sawiak[1,2], A Jennifer Morton[3], and T Adrian Carpenter[1]

[1]Wolfson Brain Imaging Centre, University of Cambridge, Cambridge, England, United Kingdom, [2]Behavioural and Cognitive Neuroscience Institute, University of Cambridge, Cambridge, England, United Kingdom, [3]Department of Pharmacology, University of Cambridge, Cambridge, England, United Kingdom

Target audience

Users of animal models for translational medicine, image analysts, clinicians of neurodegenerative disease

Introduction

Huntington's disease (HD) is an inherited, fatal neurological disorder with no known cure [1]. The R6/2 transgenic mouse is the most common animal model of the disease used for studying its pathology and potential therapeutic interventions [2]. The disease results from an unstable expansion of polyglutamine (CAG) repeats in the gene coding for huntingtin, (*htt*). Support vector machines (SVMs) are a binary classifier that finds a hyperplane offering the best separation of points in multidimensional space [3]. In this study, we trained an SVM classifier with *ex vivo* mouse brain scans using grey matter (GM) segmented portions from SPMMouse [4].

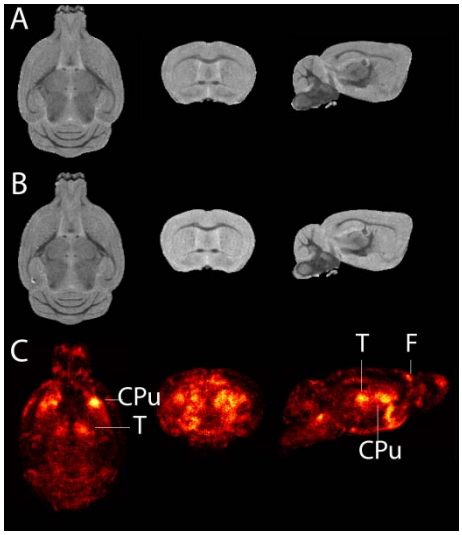

Figure 1 Typical images of WT (A) and R6/2 CAG100 brains (B) in horizontal, coronal and sagittal views. (C) shows the weight vector as described in the text illustrative of regions important for the classification of brains.

Methods

Data were obtained from the Cambridge HD mouse brain public archive [5]. SVMs were trained with Matlab 7.5 bioinformatics toolbox (Mathworks Inc, MA, USA) using masked GM segmented data calculated with SPMMouse. An initial classifier was trained with 29 WT against 29 R6/2 (CAG100) brain images all aged 12 weeks. Classifier performance at generalisation was assessed with leave-one-out cross-validation (LOOCV). The trained classifier was used to investigate three further pools of data: older WT brains (n = 21; age mean±SD 61±9 weeks), R6/2 CAG200 (n = 28; age 12 weeks), R6/2 CAG300 (n = 11; age 12 weeks) and a third model, the YAC128 mouse [6] (n = 21; age 48±3 weeks).

Results

LOOCV accuracy for segmented GM images is 98.3%. Figure 1 shows the weight vector remapped into brain images. For display, the image was thresholded at 20% of the maximum weight and the images shows summed projections on three axes. Brains with lower GM scores in these areas are more like R6/2 brains. The most prominent region seen is the caudate putamen (CPu), with the thalamus (T) and frontal cortical (F) regions also conspicuous, in line with differences seen in similar models with voxel-based morphometry [7].

Applying this classifier to the alternative models, R6/2 (CAG200), R6/2 (CAG300) and YAC128 identified 67.4%, 47.6% and 23.8% of the brains as transgenic accordingly. Of the older WT brains tested with the classifier, 3 of 21 (14.3%) were classified as transgenic according to the original classifier.

Discussion

Classifier performance was excellent and the low error achieved with LOOCV suggests that it will generalise well to populations. Implementation of the method is relatively easy with Matlab. Currently we are exploring if an *ex vivo* derived classifier performs as well on *in vivo* acquisitions. The regions used by the classifier for its detection are well known to be implicated in disease. The differential detection rates for R6/2 mouse strains is further evidence of distinct pathology in these mice [8]. In addition, that older WT mice were generally not classified as transgenic shows that the differences involved in the HD model are distinct from those of normal aging.

References

[1] Walker FO (2007) Lancet 369: 218-228 [2] Mangiarini L *et al.* (1996) Cell 87:493-500 [3] Cortes and Vapnik (1995) Machine learning 20:273-297. [4] Sawiak SJ *et al.* Proc ISMRM (2009) Hawaii 1086 [5] http://www.dspace.cam.ac.uk/handle/1810/243361 [6] Slow EJ *et al.*(2003) Hum Mol Genet 12 :1555-1567 [7] Sawiak *et al.* (2009) Neurobiol. Dis. 33:12-20 [8] Morton AJ *et al.* (2009) Neurobiol Dis 33: 331-341