Review of: "A Variant in Genes of the NPY System as Modifier Factor of Machado-Joseph Disease in the Chinese Population"

Jorge Diogo Da Silva[1], Patricia Maciel[1]

1 Universidade do Minho

Potential competing interests: The author(s) declared that no potential competing interests exist.

The published manuscript "A Variant in Genes of the NPY System as Modifier Factor of Machado-Joseph Disease in the Chinese Population" proposes to assess potential genetic modifiers of age-of-onset (AO) of Machado-Joseph disease/spinocerebellar ataxia type 3 (MJD/SCA3), in addition to the well described AO determinant number of CAG repeats, in the Chinese population. The candidate genetic factors that were chosen for this work were single nucleotide polymorphisms (SNPs) in genes encoding the neuropeptide Y (*NPY*) and NPY receptors 2 and 5 (*NPY2R* and *NPY5R*, respectively). The choice of these candidate genes was based on previous reports that NPY is likely protective for neurons in conditions such as Parkinson's disease, Huntington's disease and even MJD itself. Regarding the working hypothesis, methodology and interpretation of the results, we have the following comments:

1. The hypothesis is well-formulated and based on well-established observations in the literature, as NPY treatment and overexpression has been shown to alleviate the phenotype of an MJD mouse model. Furthermore, the choice of the two NPY receptors for assessment was also based on the fact that they are also well-established as mediators of the neuroprotective effect of this peptide. Finally, while not exhaustive, the choice of SNPs to assess was based on prior association studies with positive effects in other conditions, which made them good candidates for a potential modifying effect in MJD as well.

2. Regarding the study design and methodology, the sample size was excellent given the rarity of this neurodegenerative disease, and the existence of two patient cohorts for validation was also very good. Little information is provided regarding selection of controls (except for sex-matching). The statistical tests and models chosen for comparison of genotype frequency, allelic frequency and effect of the SNPs in the age-of-onset were adequate. The assessment of dominant, genotypic and recessive models was also suitable. However, the reported statistical data on regression models is somewhat incomplete, with only the R^2 and p-value being shown. Furthermore, the regression models could have been enriched with additional variables (such as the sex of the subject). The labeling of the tables, specifically of Tables 3 and 4, is also rather incomplete, which stops the reader from having an immediate understanding of the data. The labelling of the SNPs is also little user-friendly, as it is never clearly stated which SNP reference corresponds to which gene.

3. The effect of the number of CAG repeats in the age-of-onset is in line with previous reports, which is important for

validation of this cohort. The authors opt to compare a linear and quadratic model of prediction, and choose the quadratic model for their assessments as it has a higher percentage of explained variance (R^2). Nevertheless, as prior reports preferably use linear models, it would have been interesting to assess for both models.

4. When it comes to comparing the allele and genotype frequency of each SNP between controls and MJD patients, the authors observe a difference in the distribution of such genotypes/alleles regarding the *NPY5R* SNP rs11100494. The authors then state that, as the CC genotype was present more often in MJD patients, that this was "associated with increased susceptibility of MJD". This is an incorrect and confusing statement, as the susceptibility to MJD is only defined by the presence/absence of a CAG expansion in the *ATXN3* gene. Indeed, it is a bit confusing why the authors decided to include controls in this specific study: since MJD is a monogenic disorder with complete penetrance, it does not make sense to look for additional "susceptibility/risk factors" for the development of disease (versus not developing it). The search for modifiers of disease expression, in contrast, is of relevance, given the marked clinical heterogeneity observed in this disorder. Importantly, other explanations can be put forward for the observation of an increased frequency of the CC genotype of this *NPY5R* SNP in the MJD patient population, namely poor case-control matching and population stratification. Causation could never be inferred from this study methodology, i.e., it cannot be stated that this *NPY5R* SNP is a modifier factor of MJD. Given that the title of the manuscript is based solely on this finding, in our opinion the title itself ends up being incorrect and misleading.

5. In the analysis of the candidate SNPs as modifiers of the clinical expressivity of the disease, the *NPY5R* SNP rs11100494 was then shown not to be a predictor of age-of-onset in MJD. This goes against the main conclusion and title of the manuscript, as this observation favors the hypothesis that this SNP is of small importance for MJD. While it could still be relevant as a modifier of specific symptoms and/or rate of disease progression, this was not assessed in the current study. Therefore, the conclusions are once again misleading, as the working hypothesis of the study - that these SNPs were additional factors (to the number of CAG repeats) that modify the age-of-onset of the disease – was not confirmed.