We recently demonstrated that genetic variation in and around exon 1 of MSH3 is associated with somatic expansion of the CAG•CTG repeat and disease severity in Huntington’s disease (HD) and myotonic dystrophy type 1 (DM1) (Flower et al., 2019). Specifically, we revealed that in our well characterised European cohorts (HD, n = 218; DM1, n = 247), the MSH3 9 bp repeat allele, 3a, was associated with reduced MSH3 expression in blood and brain, with less somatic expansion, and with later age at onset and slower progression. Three non-coding single nucleotide polymorphisms (SNPs) in MSH3 were in near complete linkage disequilibrium (LD) with the 3a allele, and therefore similarly associated with each phenotype. In addition, after controlling for MSH3 repeat length, two other SNPs were independently associated with onset. These same MSH3 exon 1 haplotype associations with both age at onset and somatic instability have recently been independently replicated in HD patients of European ancestry by the latest Genetic Modifiers of HD genome wide association study (n = 9,064) (Lee et al., 2019) and by Ciosi et al. (2019a) (n = 734). These data thus firmly establish that MSH3 variants are associated with both somatic expansion and disease severity in HD.

An obvious question arising, is whether MSH3 variants are also associated with disease severity in other simple sequence repeat expansion disorders? In Bettencourt et al. (2016), we showed that DNA repair variants that influence HD onset are also associated with onset in other CAG repeat expansion polyglutamine diseases, though this was primarily driven by spinocerebellar ataxia type 2 (SCA2) and SCA6.

In their letter, Yau et al. investigated whether the MSH3 3a allele was associated with disease severity in 132 spinocerebellar ataxia type 3 (SCA3) and 136 Friedreich’s ataxia (FRDA) patients. They found no significant associations. Of course, such a finding may indicate that somatic expansion is not a critical feature of these disorders and/or that MSH3 does not drive somatic expansion of the CAG repeat in SCA3 or the GAA repeat in FRDA. However, it is interesting to note, that their pattern of age at onset variation with 3a genotype (Figure 1) is not dissimilar to our data (Figure 2, (Flower et al., 2019)), in that two copies of 3a appear to delay onset. Indeed, as the authors acknowledge, it remains highly plausible that the failure of Yau et al. to detect a statistically significant association represents a false negative type II error. In our analysis, the number of 3a alleles accounted for around 5% of variability in onset. With such a small sample, Yau et al.
would have had only around 70% power to detect a similar effect at a significance level of 0.05. Our HD progression score integrated longitudinal motor, cognitive and imaging data and has proved to be a highly sensitive measure of HD severity (Hensman Moss et al., 2017; Ciosi et al., 2019b). In contrast, Yau et al. investigated the effect on disease progression based on a single measure, the SARA score, which was only available for a very small sub-sample of FRDA patients (n = 57). SARA is a relatively insensitive semi-quantitative measure of ataxia (Burk et al., 2013), requiring a sample size over 200 to detect a 50% change in annual progression (Jacobi et al., 2015). Additionally, whilst SCA3 onset correlates with repeat length, SARA score does not (Huang et al., 2017), potentially because of its insensitivity, or because mechanisms underlying onset and progression are distinct.

In addition to the small sample size and insensitivity of the clinical progression measures, technical constraints in the assay and possible population-specific effects may also have limited the ability of Yau et al. to detect an effect for MSH3 variants. Whereas we unambiguously genotyped all MSH3 repeat alleles using amplicon sequencing, Yau et al. identified 3a alleles by a combination of fragment length analysis, and by deconvoluting superimposed Sanger sequencing traces in heterozygotes. Notably, Yau et al., revealed a 3a allele frequency in the combined SCA3 and FRDA cohort of 0.35. This was significantly higher than the 0.25 and 0.27 in our European HD and DM1 subjects (Chi-squared = 15.48, p = 0.0008). This is likely due to the heterogeneous ancestry of their cohort, which included African, Caribbean, Asian and South American subjects, regional differences in MSH3 allele frequencies (Nakajima et al., 1995), and/or the inability to resolve 3a alleles from other similar alleles, especially in populations likely to contain additional diversity from that detected in Europe. In addition, whilst our data implicates 3a (Flower et al., 2019), we cannot exclude an effect from other MSH3 variants in LD with 3a. It is very possible that MSH3 SNP and repeat LD patterns differ in non-European populations, potentially further confounding the ability of Yau et al., to define an association with 3a in their heterogenous cohort.

In our analyses, we used Illumina sequencing in HD (Ciosi et al.), and, small pool PCR and restriction digest analysis in DM1 (Cumming et al., 2019), to define both the sequence and inherited progenitor allele length of disease-associated CAG•CTG alleles. In contrast, Yau et al., measured the length of the SCA3 and FRDA disease-associated repeats by fragment analysis and took no account of the potential confounding effects of age-dependent somatic expansions. The gross confounding effects of overestimating allele length due to age-dependent somatic expansion on establishing genotype to phenotype correlations in DM1 is well established (Morales et al., 2012). Although the levels of somatic expansion of small HD disease-causing alleles (40 to 50 repeats) in blood DNA are relatively low, permitting the unambiguous identification of the inherited progenitor allele, larger HD alleles (>50 CAG repeats) can show much higher levels of somatic expansion, obviating the unambiguous identification of the progenitor allele (Ciosi et al., 2019b). Given the
much longer average length of SCA3 (mean CAG = 70 repeats) and FRDA (mean GAA ~ 1000) alleles, it seems very likely that determination of the progenitor allele length has been compromised by somatic expansion. Such an effect would introduce a bias in which the length of more somatically unstable alleles are systematically even more overestimated, leading to an underestimation of the degree of variation due to allele length, and thus potentially directly masking the disease accelerating effects of MSH3 variants that promote somatic expansion. Failure to take account of the effects of somatic expansion in driving measured allele length may also explain the lower proportion of variability in onset observed in SCA3 and FRDA (46% and 36% respectively), relative to the ~60-70% typically observed in HD (Andresen et al., 2007) and DM1 (Morales et al., 2012). Moreover, fragment length analysis sheds no light on the presence or absence of variant repeats that have been shown to have major disease modifying effects in HD (Ciosi et al., 2019b; Lee et al., 2019), DM1 (Braida et al., 2010; Cumming et al., 2019) and FRDA (Cossee et al., 1997; McDaniel et al., 2001; Sakamoto et al., 2001; Pollard et al., 2004), and are known to exist in some SCA3 alleles (Limprasert et al., 1996).

Although it is clear that the expanded SCA3 repeat is somatically unstable (Watanabe et al., 1996; Hashida et al., 1997; Cancel et al., 1998), the degree of somatic instability is lower than in HD and SCA1 (Maciel et al., 1995; Cancel et al., 1998). This is despite the fact that expanded SCA3 alleles are typically much larger (70 to 80 CAG repeats) than in HD and SCA1 (40 to 50 CAG repeats). Indeed, it has been suggested that inherited SCA3 alleles need to be so large to cause disease because they are somatically more stable and need to start closer to the pathogenic threshold in neurons to illicit disease during the lifetime of an individual (Nestor and Monckton, 2011). Likewise, the inherited disease-causing GAA alleles in FRDA are already very large, and although they are somatically unstable (De Biase et al., 2007; Long et al., 2017), the extensive somatic expansions observed in HD (Kennedy et al., 2003) and DM1 (Anvret et al., 1993; Thornton et al., 1994) affected tissues have not been reported. These observations do not preclude a role for further somatic expansion in driving SCA3 and FRDA, but suggest that these two disorders may be less susceptible to somatic expansion than HD and DM1, and effects of MSH3 variants may therefore be more subtle.

In conclusion, Yau et al. did not challenge our associations between MSH3 variants and symptomatic variability and somatic expansion in HD and DM1, but rather their extension to SCA3 and FRDA. Their study was small and underpowered to detect changes in onset and progression, and does not preclude the involvement of MSH3 in other polyglutamine diseases. Indeed, we think it premature to assert that “repeat variant in MSH3 is not a genetic modifier for spinocerebellar ataxia 3 and Friedreich ataxia”, but agree with the authors that a larger study of repeat disorders will be important to identify common genetic factors.
References


