



Commentary

Tandem Repeats and Repeatomes: Delving Deeper into the ‘Dark Matter’ of Genomes



Anthony J. Hannan^{a,b,*}

^a Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, Victoria, Australia

^b Department of Anatomy and Neuroscience, University of Melbourne, Parkville, Victoria, Australia

The understanding of human biology and diseases, and the practice of modern medicine, is being transformed by genomics. The sequencing of the human genome (from a handful of healthy individuals) almost two decades ago was only a beginning. Over half of the human genome, approximately 3 billion base pairs of DNA, involves repetitive DNA sequences which constitute the repeatome (Hannan, 2012) and could be considered the ‘dark matter’ of the genome. This is because much of it remains under-studied, poorly annotated and functionally mysterious. For large tracts of repetitive sequences within the repeatome, we remain uncertain as to whether they are ‘junk DNA’ (if in fact any part of our genome deserves such a pejorative description) or whether their evolved functions simply remain opaque at this point in time. A key component of the repeatome involves tandem repeats, which constitute over 3% of the human genome. There are over 1.5 million short tandem repeats (STRs), which consist of repeating units of 1–6 base-pair motifs of DNA, in the human genome and recent evidence supports their roles in a range of molecular and cellular processes (Gymrek et al., 2016; Hannan, 2018). Whilst the evidence for tandem repeat expansions causing particular Mendelian human disorders has accumulated in the past three decades, the potential roles of tandem repeat polymorphisms (TRPs) in modulating complex human traits and disorders remain largely unexplored (Hannan, 2018).

A new article (Lee et al., 2018) in *EBioMedicine* describes a very interesting association between a CAG tandem repeat polymorphism in the huntingtin (HTT) gene and specific measures of cognition (which they refer to as ‘general intelligence’). This work is novel and adds to the rapidly expanding tandem repeat genetics field. The HTT gene first came to prominence 25 years ago, when the CAG repeat, encoding a polyglutamine tract, was found to be expanded in patients with Huntington’s disease (HD). However, like many STRs, it is highly polymorphic in the general population, where it may constitute a functional polymorphism. The present study (Lee et al., 2018) involved a cohort of children (6–18 years) at risk of HD and age-matched healthy controls. The most significant aspect of this study is that in the normal range

(below the threshold for HD) there was an association between CAG repeat length and general intelligence. The major limitation of this study, acknowledged by the authors, is that it needs to be followed up with larger independent cohorts of healthy controls, as well as those gene-positive for the HD mutation. Genome-wide association studies (utilising SNP-based microarrays) for such complex traits (or polygenic disorders) now generally involve very large cohorts, to provide statistical power in the face of extensive polygenicity and heterogeneity. Therefore, the present study ($n = 316$), which was appropriately powered for a gene-positive presymptomatic HD study in children (the primary aim of the study), is not well powered as a candidate gene association study of a complex trait (which presumably was not the original intention of the investigators). Nevertheless, it does provide an important framework for larger replication studies.

The other key point is that for repeat lengths within the higher HD range (also part of this study), even though the children were not predicted to become symptomatic for many years, some of the cognitive changes could reflect very early HD-associated cognitive changes. Alternatively, or additionally, it is possible that adult-onset HD (~95% of cases) is at least partly a developmental disorder. Thus, the HD brain may undergo abnormal development and maturation, setting it up for later neurodegeneration, as well as cognitive, psychiatric and motor symptoms. In fact, the involvement of tandem repeats and TRPs in brain development and associated brain functions has been previously proposed (Nithianantharajah and Hannan, 2007), and is relevant to the findings in the present article. It is possible that the tandem repeat and associated TRP in the HTT gene is a modulator of specific aspects of brain development, which then manifest as cognitive function.

The genetic association between the HTT CAG repeat length polymorphism (below the HD-associated length threshold) and clinical depression (Perlis et al., 2010; Gardiner et al., 2017a), is also highly relevant to this latest study. Furthermore, another recent study (Gardiner et al., 2017b) has linked the CAG/glutamine repeat length in two other genes (independent of HTT) to depression, suggesting a broader role for such TRPs in genetic predisposition for depressive disorders, and possibly broader aspects of affective function.

Another implication of this, and related, work is that HD involves ‘change of function’ of the polyglutamine-expanded huntingtin protein, not just ‘toxic gain of function’ (Hannan, 2018). As we start to think of

DOI of original article: <https://doi.org/10.1016/j.ebiom.2018.03.031>.

* Corresponding author at: Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, Victoria, Australia.

E-mail address: anthony.hannan@florey.edu.au (A.J. Hannan).

<https://doi.org/10.1016/j.ebiom.2018.04.004>

2352-3964/© 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

the function of these tandem repeats, and associated TRPs, in the normal range, we need to incorporate this conceptual framework into our understanding of the diseases (in this case HD) which result when the tandem repeats are expanded to extreme lengths.

The genome-wide association study (GWAS) approach is transforming human genetics, as well as our understanding of human biology, in health and disease. However, these studies have been based on single nucleotide polymorphisms (SNPs) and associated microarrays ('SNP chips'), and one major issue for GWAS on complex traits and polygenic disorders is that they do not account for 'missing heritability', which may have multiple origins (Eichler et al., 2010; Hannan, 2010). One possibility is that tandem repeats and TRPs (such as the HTT CAG repeat) might help explain this 'missing heritability' (Hannan, 2010). There are multiple potential approaches to GWAS that incorporates not only SNPs, but also TRPs (and other repeatome polymorphisms), associated with complex traits and disorders (Hannan, 2018). These approaches include whole-genome sequencing using long sequence reads and PCR-free protocols to optimise repeat-length sequencing (Hannan, 2018). However, new approaches, including those that use GWAS SNP data for imputation of STR lengths, are also possible and should be pursued.

In conclusion, this new article (Lee et al., 2018) provides a further piece in the puzzle of tandem repeat biology and the roles of TRPs in modulating human brain development and cognition. We are only beginning to expand our understanding of tandem repeats, and their associated repeatome, and this exciting field of research will offer many new opportunities to understand human biology, as well as to develop novel preventative and therapeutic approaches for a wide range of human disorders.

Acknowledgements

AJH is supported by a Principal Research Fellowship (GNT1117148) and Project Grants (GNT1126885 and GNT1138321) from the NHMRC (Australian Government), as well as Discovery Project grants from the ARC (Australian Government), and DHB Foundation, Equity Trustees (private philanthropy). None of the funding agencies had any influence on the writing of this article.

References

- Eichler, E.E., Flint, J., Gibson, G., et al., 2010. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* 11, 446–450.
- Gardiner, S.L., van Belzen, M.J., Boogaard, M.W., et al., 2017a. Huntingtin gene repeat size variations affect risk of lifetime depression. *Transl. Psychiatry* 7, 1277.
- Gardiner, S.L., van Belzen, M.J., Boogaard, M.W., et al., 2017b. Large normal-range TBP and ATXN7 CAG repeat lengths are associated with increased lifetime risk of depression. *Transl. Psychiatry* 7, e1143.
- Gymrek, M., Willems, T., Guilmatre, A., et al., 2016. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat. Genet.* 48, 22–29.
- Hannan, A.J., 2010. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. *Trends Genet.* 26, 59–65.
- Hannan, A.J., 2012. Tandem repeat polymorphisms: mediators of genetic plasticity, modulators of biological diversity and dynamic sources of disease susceptibility. *Adv. Exp. Med. Biol.* 769, 1–9.
- Hannan, A.J., 2018. Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.* <https://doi.org/10.1038/nrg.2017.115> [Epub ahead of print].
- Lee, J.K., Conrad, A., Epping, E., et al., 2018. Effect of trinucleotide repeats in the Huntington's gene on intelligence. *EBioMedicine*:2018, <https://doi.org/10.1016/j.ebiom.2018.03.031>.
- Nithianantharajah, J., Hannan, A.J., 2007. Dynamic mutations as digital genetic modulators of brain development, function and dysfunction. *BioEssays* 29, 525–535.
- Perlis, R.H., Smoller, J.W., Mysore, J., et al., 2010. Prevalence of incompletely penetrant Huntington's disease alleles among individuals with major depressive disorder. *Am. J. Psychiatry* 167, 574–579.