# Simple sequence repeats: Genetic modulators of brain function and behavior

John W. Fondon III, Elizabeth A. D. Hammock, Anthony J. Hannan and David G. King

**Abstract**

Simple sequence repeats (SSRs), sometimes described as genetic "stutters", are DNA tracts in which a short base-pair motif is repeated several to many times in tandem (e.g., CAGCAGCAG). These sequences experience frequent mutations that alter the number of repeats. Because SSRs are commonly located in promoters, untranslated regions and even coding sequences, such mutations can directly influence almost any aspect of gene function. Mutational expansion of certain triplet repeats is responsible for several hereditary neurodegenerative disorders, but SSR alleles can also contribute to normal variation in brain and behavioral traits. Here we review studies implicating SSRs not just in disease but also in circadian rhythmicity, socio-sexual interaction, aggression, cognition and personality. SSRs can affect neuronal differentiation, brain development and even behavioral evolution.

## Introduction

In the early 1990s, the neuroscience community was surprised by reports that Huntington's disease, fragile X syndrome and several other hereditary neurological disorders all shared a common cause in the form of mutational expansion of triplet repeat DNA sequences within genes[1, 2]. As one researcher was quoted in *Science*[3], "No one expected that DNA sequences could be so unstable or behave as these do."

Nevertheless, DNA researchers have long known that simple sequence repeats (SSRs, also called microsatellites and minisatellites) have a propensity for "slippage mutations" which increase or decrease the number of repeats without otherwise altering the sequence[4, 5]. SSRs based on various motifs are extremely numerous in eukaryotic genomes; many human genes possess multiple SSRs. The abundant polymorphism which results from repeat slippage has become the basis for DNA fingerprinting, lineage analysis and gene mapping. Yet this very mutability, together with the seeming lack of information content in such repetitive genetic "stutters", once appeared to preclude any possibility of critical function for SSRs. Indeed, much of the literature on SSR polymorphism has assumed (and continues to assume [e.g., 6]) that SSRs are genetic "junk", supplying only "neutral" variation with no appreciable effect on phenotype.

However, as we review here, abundant evidence warrants some skepticism toward any presumption of neutrality for SSR alleles[7-15]. Repeat expansion diseases have highlighted the occurrence of repetitive sequences at sites where slippage mutations can have dramatic consequences. But many other examples of repeat number effects have come to light since Hamada and colleagues first established that altering the number of repeating dinucleotides could affect gene activity[7]. By now, SSR sequences are widely recognized for a remarkable set of characteristics [reviewed in 10-15]:

- Slippage mutations, occurring at rates that can be as high as $10^{-2}$ per cell division at a single SSR locus, yield abundant repeat-number variation at innumerable SSR sites.
- SSR slippage mutations are readily reversible, unlike single nucleotide substitutions.
- SSR mutability is a function of locus-specific properties including motif length, total number of repeats, inclusion of variant motifs, and flanking sequences.
- Repeat number variation can affect diverse aspects of gene function including transcription rates and transcript stability; rates of protein folding and turnover; and protein-protein interactions, aggregation and sub-cellular location.
- Repeat number mutations commonly exert incremental quantitative effects, not unlike adjusting a "tuning knob"[11]. They can also act reversibly to switch genes on or off.

These characteristics have prompted speculation that the mutability of SSRs could play an important and potentially beneficial role in evolution[5, 8-11, 13-17]. Thus the triplet repeat expansion diseases represent only the pathological extreme of a much more general mutational process, one which also contributes to normal brain function and development.

## Repeat-expansion diseases

At least twenty different neurological disorders are caused by expanded SSRs (Table 1, reviewed in [1, 2]). Such diseases are frequently characterized by "genetic anticipation," a hereditary tendency toward further expansion of pathological repeat alleles in each generation, leading to earlier onset and accelerated disease progression in subsequent generations.

Huntington's disease (HD) exemplifies neurodegenerative repeat expansion diseases caused by an expanded protein-coding repeat. The huntingtin protein normally contains a variable run of 6 to 35 glutamines encoded by CAG repeats. Expansion of this repeat to lengths greater than 39 glutamines causes abnormalities of protein folding, cleavage, inter-

**JWF**: McDermott Center for Human Growth and Development and Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

**EADH**: Vanderbilt Kennedy Center for Research on Human Development, Vanderbilt University, Nashville, TN 37232, USA

**AJH**: Howard Florey Institute and Department of Anatomy and Cell Biology, University of Melbourne, VIC 3010, Australia

**DGK**: Department of Anatomy and Department of Zoology,Southern Illinois University Carbondale, Carbondale, IL 62901, USA.

*Corresponding author*: dgking@siu.edu

actions, trafficking and degradation, as well as gene transcription, synaptic function and cellular plasticity (reviewed in [18, 19]). These defects in gene expression, protein regulation and function give rise to late-onset progressive, selective neural cell dysfunction and death associated with a movement disorder, psychiatric symptoms and cognitive deficits culminating in dementia. At least eight additional fatal neurological disorders are also caused by expanded polyglutamine tracts. Like HD, these are all believed to involve a gain of toxic protein function, with subtle differences in gene expression, intracellular localization and protein context for the polyglutamine tracts leading to distinct neuropathological profiles. (A potential role for RNA toxicity associated with the transcribed CAG repeat in these diseases has not yet been ruled out and is worthy of investigation.) Nevertheless, apart from the presence of polymorphic CAG repeats encoding polyglutamine tracts, most of the genes involved share no substantial similarities other than expression in the central nervous system.

In contrast, fragile-X mental retardation syndrome (FRAXA) develops when a greatly-expanded CGG repeat in the 5′ untranslated region (UTR) of the *FMR1* gene leads to decreased levels of the RNA-binding protein FMRP, implicated in dendritic and synaptic function (reviewed in [20, 21]). Normal *FMR1* alleles vary between 6 and 53 repeats, while alleles of greater than 200 repeats abolish or drastically reduce gene transcription. Intermediate lengths (55 to 200 repeats), sometimes called "premutation alleles" because of their high probability of pathological expansion, are associated with fragile X tremor/ataxia syndrome (FXTAS) and/or autism spectrum disorder (reviewed in [22, 23]). The favored hypothesis for the repeat's role in FXTAS etiology is a molecular gain of function whereby mRNAs containing expanded CGG repeats introduce novel functional changes, such as protein misfolding and aggregation, leading to depletion of other proteins, progressive neural cell dysfunction and eventual death.

Several additional repeat-expansion neurodegenerative diseases display a wide variety of motif sequences, lengths, and modes of action (Table 1). While the most common and best-studied repeat-expansion diseases involve triplet repeats encoding expanded amino acid tracts, non-coding repeat expansions can also disrupt gene transcription, RNA processing, protein translation, and other cellular processes.

**Table 1. Many neurological diseases are caused by repeat expansion.**

| Disease | SSR type / motif | Repeat length | | Affected gene product | Proposed class of pathogenic mechanism | Reference |
|---|---|---|---|---|---|---|
| | | Normal | Pathology | | | |
| Huntington's disease (HD) | exon / CAG | 6-35 | 36-250 | huntingtin | polyglutamine toxicity, molecular gain of function | 18, 19 |
| Dentatorubral pallidoluysian atrophy (DRPLA) | exon / CAG | 3-36 | 49-88 | atrophin | polyglutamine toxicity, molecular gain of function | 1, 2 |
| Spinocerebellar ataxia | exon / CAG | | | | polyglutamine toxicity, molecular gain of function | 1, 2 |
| type 1 (SCA1) | | 6-39 | 40-88 | ataxin-1 | | |
| type 2 (SCA2) | | 14-32 | 33-200 | ataxin-2 | | |
| type 3 (SCA3, Machado-Joseph) | | 12-40 | 55-86 | ataxin-3 | | |
| type 6 (SCA6) | | 4-19 | 21-33 | CACNA1A | | |
| type 7 (SCA7) | | 4-35 | 37-306 | ataxin-7 | | |
| type 17 (SCA17) | | 25-42 | 43-63 | TBP | | |
| Spinal and bulbar muscular atrophy (SBMA) | exon / CAG | 9-36 | 38-65 | androgen receptor | polyglutamine toxicity, molecular gain of function | 1, 2 |
| Fragile X mental retardation syndrome (FRAXA) | 5'-UTR / CGG | 6-53 | >200 | FMR1/FMRP | decreased RNA/protein levels, molecular loss of function | 20, 21 |
| Fragile X tremor/ataxia syndrome (FXTAS) | 5'-UTR / CGG | 6-53 | 55-200 | FMR1/FMRP | RNA toxicity, molecular gain of function | 22, 23 |
| Fragile XE syndrome (FRAXE) | 5'-UTR / CCG | 4-39 | >200 | FMR2 | decreased RNA/protein levels, molecular loss of function | 1, 2 |
| Spinocerebellar ataxia type 8 (SCA8) | 3'-UTR & exon / bi-directional CTG/CAG (sense/antisense) | 16-37 | >74 | SCA8 and KLHL1 (via antisense) | RNA and/or polyglutamine toxicity? | 54 |
| Spinocerebellar ataxia type 10 (SCA10) | intron / ATTCT | 10-22 | >280 | ataxin-10 | RNA toxicity, molecular gain of function? | 55 |
| Spinocerebellar ataxia type 12 (SCA12) | 5'-UTR / CAG | 7-45 | 55-78 | PPP2R2B | decreased RNA/protein levels, molecular loss of function? | 56 |
| Huntington disease-like 2 (HDL2) | exon / CTG variably spliced | 6-28 | 51-78 | junctophilin-3 | RNA toxicity, molecular gain of function? | 57 |
| Friedreich ataxia (FRDA) | intron / GAA | 6-34 | >100 | FXN/frataxin | decreased RNA/protein levels, molecular loss of function | 58 |
| Myotonic dystrophy DM1 | 3'-UTR / CTG | 5-37 | >50 | DMPK | RNA toxicity, molecular gain of function | 59 |
| Myotonic dystrophy DM2 | intron / CCTG | 10-38 | >50 | ZNF9 | RNA toxicity, molecular gain of function | 59 |
| Myoclonic epilepsy EPM1 (Unverricht–Lundborg disease) | 5'-UTR / CCCCGCCCCGCG | 2-17 | >30 | cystatin B | decreased RNA/protein levels, molecular loss of function | 60 |

We regret that space limitations preclude comprehensive citation. Please consult the cited reports and references therein for access to this literature.

Their frequency and functional diversity hint at how deeply SSRs are woven into functional genomic sites that can influence neurological function.

**Effects of SSR variation on behavior**

Although extreme repeat expansion is deleterious, studies of many different organisms provide evidence that non-pathogenic SSR polymorphism contributes to normal quantitative genetic variation in traits ranging from yeast cell adhesion to dog skeletal morphology[10, 14, 15]. Among the best-studied examples of relationships among phenotype, molecular function and SSR polymorphism are three genes influencing animal behavior. One of these, the *period* (*per*) gene of *Drosophila melanogaster*, participates in biological time-keeping including circadian rhythmicity. Another, involved in social recognition memory, aggression and socio-sexual behaviors in vertebrates, is the *avpr1a* gene of voles (*Microtus sp*., "field mice"). The third, a serotonin transporter gene (*SLC6A4*), is implicated in several aspects of primate behavior.

The *Drosophila* PER protein contains an embedded array of threonine-glycine and serine-glycine dipeptide repeats encoded by a hexanucleotide SSR in exon 5. In laboratory studies, flies carrying a 20-repeat allele show reduced sensitivity of circadian rhythm to temperature fluctuations compared to those with a shorter 17-repeat allele[24]. In natural populations sampled across Europe, northern Africa and Australia, the 20-repeat allele was more frequent at higher latitudes. This latitudinal cline suggests that natural selection discriminates among alleles of different lengths based on their contribution to temperature compensation in regions with greater temperature fluctuations (reviewed in [25]). Complementary data have been independently reported from *Drosophila* populations showing microclimatic divergence on north- and south-facing slopes of Evolution Canyon in Israel[26]. A functional role for the *per* SSR is also supported by interspecies comparison, including gene transplantation experiments that suggest coevolutionary compensation between the effects of SSR and flanking sequence variants[27]. (Although data are less extensive, blue tit populations (*Cyanistes caeruleus*) also show a latitudinal cline in frequencies of polyglutamine repeat number alleles in a different clock gene, again suggesting environmental selection based on functional effects of allele length[28].)

Establishing a causal relationship between genetic variants and phenotype poses considerable difficulties (Box 1). Even in *Drosophila* (above) where experimental manipulation of genes is feasible, evidence that SSR variants actually matter in natural environments remains primarily correlative. In mammals, and especially in humans, the differing functional effects of SSR repeat-number variants is generally supported by circumstantial evidence of varying quality rather than by conclusive proof. Among the more compelling cases is one in which several converging lines of evidence identify functional SSR variation in the 5' regulatory region of the vasopressin 1a receptor gene *avpr1a* as a significant contributor to social behaviors such as selective partner preference and offspring care[29].

Comparison of individual prairie voles (*Microtus ochrogaster*) whose parents carried longer-than-average SSR alleles with those whose parents had shorter alleles revealed differences in V1aR protein distribution in the brain as well as alterations in vasopressin-dependent social behavior[29]. Transgenic introduction of *avpr1a* from a prairie vole into a mouse resulted in mouse brain distribution of V1aR protein and affiliative behavior in response to vasopressin injection that were more similar to those of a prairie vole than to those of a wild-type mouse littermate[30]. Causal efficacy of the SSR difference was demonstrated through *in vitro* transcription assays. Insertion into cultured cells of distinct SSR domains from two common alleles, with other *avpr1a* regions held constant, revealed that variation in SSR length could alter

---

**Box 1. Establishing SSR functionality is challenging.**

Functional SSR polymorphisms typically exert only small, quantitative effects on phenotype, and behavioral traits are typically multigenic, heterogeneous, weakly penetrant, and environmentally influenced[09-15]. These factors combine to make identification of SSRs influencing behavior exceedingly difficult. The power of genetic association studies to identify functional alleles is limited by the magnitude of their effects; even the largest studies likely lack the power to detect most functional SSRs. Linkage disequilibrium mapping is further undermined by the tendency of high SSR mutation rates to rapidly degrade linkage. In addition, the preferred genotyping methods employed for large studies (e.g. DNA microarrays) are incapable of assaying repeats, so SSRs are often simply ignored. However, prospects are improved somewhat since SSRs offer a couple of unique advantages that when exploited can greatly improve chances for success. The first advantage is that their locations are finite and known, so researchers know exactly where to look. The second is statistical: since SSRs usually exert length-dependent incremental effects on phenotype, this feature can be incorporated in statistical tests, increasing power significantly (e.g., [71]).

As with any other type of genetic variant, unequivocal "gold standard" proof of SSR phenotypic effects requires introducing the alleles into a controlled genetic background through genetic manipulation of an appropriate model organism -- an expensive and time-consuming process that has yet to be extensively implemented with SSR variation.

The difficulties inherent in establishing phenotypic effects of repeat variation are illustrated by the example of dopamine receptor D4. Exon three of human DRD4 has a 48 basepair motif SSR, with allele lengths ranging from two to ten repeats. The seven-repeat allele was originally associated with novelty seeking in 1996[64, 65]. Since that time, DRD4 repeat alleles have been associated with several personality and neuropsychiatric traits, including ADHD, alcoholism, schizophrenia, adaptability, drug sensitivity, reward dependence and temperament. Most of these associations have alternately been refuted and supported by subsequent studies, and despite more than a decade of intense study by numerous groups there remains considerable uncertainty about the effects of SSR variation at this locus . As with all putatively functional SSRs, validation of DRD4 SSRs require experimental models. However, this SSR appears to have been under strong directional selection in human DRD4 and is not present in the mouse ortholog. Until an alternative model is developed, the behavioral effects of this SSR seem destined to remain uncertain and contentious.

**Table 2. Several aspects of human behavior are associated with repeat number variation**
(see main text for non-human examples)**.**

| Gene | SSR type and motif | Affected trait | Evidence | References |
|---|---|---|---|---|
| serotonin transporter SERT (SLC6A4), 5-HTTLPR | noncoding, 5' promoter; 44 basepair motif | anxiety-related traits | *in vitro* and *in vivo* assays; inconsistent association studies; moderated by life stress; similar results in Rhesus macaque | 32-39, 61 |
| serotonin transporter SERT (SLC6A4) | noncoding, intron; 17-basepair motif | susceptibility to bipolar disorder, response of SERT to lithium. | association; *in vitro* assays. | 61, 62, 63 |
| dopamine receptor DRD4 | coding; 48 basepair motifs | novelty-seeking behaviors | association | 61, 64, 65 |
| dopamine transporter DAT1 (SLC6A3) | noncoding, 3' UTR; 40 basepair motif | attention deficit hyperactivity disorder; episodic memory formation | *in vitro* and *in vivo* assays; association | 61, 66, 67 |
| DNA binding protein Jarid2 | coding; tetranucleotide | schizophrenia | association | 68 |
| androgen receptor | coding; CAG repeat | cognitive function | association | 69 |
| α2b-adrenoceptor ADRA2B | coding; imperfect glutamic acid repeat | emotional memory | association; *in vitro* assays | 70 |
| arginine vasopressin receptor, AVPR1a | noncoding, upstream, dinucleotide repeats | altruism, other social behaviors | association, *post mortem* gene expression | 71 |

We regret that space limitations preclude a full reference list. Please consult the cited papers for access to literature.

levels of luciferase reporter activity[29]. Nevertheless, while SSR length correlates with brain and behavioral traits among individual prairie voles, phylogenetic analysis of other vole species[31], which exhibit considerable interspecies diversity in the *avpr1a* SSR as well as in social structure and brain distribution of V1aR, has established that this SSR is not by itself a reliable marker for complex social behavior which depends on multiple brain circuits, environmental context and many genes[29].

Several candidate-gene sites where SSR polymorphisms have been associated with human behavior are listed in Table 2. Evidence for most of these remains less than ideal (see Box 1). Perhaps the best human example involves a polymorphic SSR (designated 5-HTTLPR) in the upstream promoter region of the serotonin transporter gene, *SLC6A4*. In a classic study, the shorter of two common alleles, which yields lower levels of gene expression in transcription reporter assays, was correlated with increased neuroticism, tension and harm avoidance[32]. However, many subsequent studies have failed to support those particular associations or have otherwise offered conflicting findings; one meta-analysis of gene association studies indicated that 5-HTTLPR was weakly associated with harm avoidance scores but not significantly with neuroticism[33], while another meta-analysis reversed these relationships[34]. Provocatively, individuals carrying two long alleles of this SSR appear to be somewhat protected from the negative effects of significant life stress[35]. Such gene × environment interactions are an important consideration in human association studies and may be used to clarify conflicting findings. Furthermore, the putative modulation of anxiety-like traits by an SSR polymorphism in the serotonin transporter is also supported by investigation of a similar, but not identical, polymorphism (rh5-HTTLPR) in the rhesus macaque (*Macaca mulatta*) serotonin transporter gene. As in humans, the shorter allele of this polymorphism yields lower gene expression *in vitro*[36]. The shorter allele has been associated with a variety of related measures including reduced levels of the serotonin metabolite 5-HIAA[36], increased distress during assessment[37] and greater inhibition by anxiety-provoking novel environments[38]. Notably, these relationships are most apparent when rearing conditions are taken into account, indicating a gene × environment interaction like that of the human serotonin transporter SSR polymorphism. Curiously, early dispersal of male offspring from their natal group has also been associated with the shorter allele[39].

**Potential significance of mitotic SSR mutation for brain development**

Somatic mosaicism is a common feature of several of the neurodegenerative repeat expansion diseases[40-42, 72]. For example, as seen in spinocerebellar ataxia 2 (a classic triplet repeat disease caused by expanded CAG-encoded polyglutamine tracts), pathologically expanded repeats can expand further during development to create a mosaic of cells with differing repeat lengths. Awareness of this role for mitotic repeat expansion was prompted by the discovery of individuals who possessed expanded alleles but who never developed symptoms of this otherwise fully penetrant disease. Closer examination of these anomalously long but benign alleles revealed the presence of silent, slippage-suppressing CAA interruptions within the CAG repeat sequence[43]. This discovery, combined with observations of hyper-expanded alleles in diseased brains, suggested that the source of the disease was not the length of the inherited allele *per se* but rather the longer allele's propensity for rampant hyper-expansion during ontogeny. The resultant pathogenicity is based on susceptibility of particular SSRs to somatic mutation and can be reduced by the stabilizing influence of repeat imperfections. Additional examples of the stabilization of otherwise pathological repeat alleles by repeat interruptions have been reported[44-46] and may be a general feature of repeat-expansion disorders.

Several observations indirectly suggest that mitotic slippage mutations could also produce somatic variation during normal brain development. First, non-pathological somatic SSR mutations occur in other developing systems, with benign or even potentially adaptive effects. For

**Outstanding Questions**

**To what extent are individual differences in behavioral traits causally associated with variation supplied by SSRs?** Since SSRs are an abundant source for quantitative genetic variation, SSR alleles may be contributing factors in any hereditary trait displaying common quantitative variation, including personality or specific cognitive ability. Intriguing candidates abound. There are many polymorphic SSRs in genes involved in neural function, but experimentally testing them is difficult. Numerous factors, such as environmental effects, confound genetic studies of behavior. SSR's subtle effects on phenotype make them more challenging to link to behavioral traits than other types of mutations (see Box 1).

**To what extent does SSR variation facilitate behavioral adaptation during nervous system evolution?** Putatively non-functional (neutral) SSR allelic frequency differences between populations are common and widely used as measures for population divergence. Since such differences can arise simply through genetic drift, the hypothesis that any SSR divergence reflects a response to selection needs additional support through experimental demonstration of functional effects on phenotype and fitness.

**To what extent might somatic variation contribute to functional differentiation among individual nerve cells?** A necessary first step is determining how much genetic variation is created by somatic mutation of particular SSRs during normal nervous system development. If such mutation is observed, especially in candidate genes with known neural function, then the functional relevance of various alleles must be identified.

example, slippage mutations in a short poly-C tract in the coding region of the melanocortin receptor gene of domestic red pigs cause frame shifts that reversibly disable or enable protein production. Beginning with an inactive germ line allele, subsequent function-restoring mutations occur at a fairly high frequency during skin development, creating a pattern of black spots in these animals[47]. Second, although somatic SSR mutation rates in normal nerve cells have not yet been reported, somatic mosaicism for pathogenic repeat allele length is commonly observed in neurodegenerative diseases[40-42, 72] and at least in non-neural tissues appears to be regulated by other genetic factors[48, 49]. Third, a precedent for effective exploitation of mitotic mutational mechanisms exists in somatic hypermutation of antibody genes[50]. Finally, other sources for genetic diversity within nerve cell populations are already recognized, including constitutional aneuploidy[51] and LINE-1 retrotransposable elements[52].

With thousands of repeat loci in the genome, some mutation of normal SSR alleles appears not only probable but inevitable during the many billions of mitoses occurring during brain development. That SSR mutation can exert a direct impact on gene function, through any of several distinct molecular mechanisms, is also well-known[7-15]. The enrichment of SSRs in genes involved in neuronal differentiation and function[53] thus raises the intriguing possibility that somatic mutation of SSRs might be a normal, perhaps even essential, component of normal brain development[1]. Subtle but abundant variation in the expression and activity of ion channels and transmitter receptors as well as transcription factors and other developmental regulators, caused by somatic SSR mutation, would undoubtedly enhance computational complexity of the human brain. Evolutionary selection for such complexity may explain extensive human-specific increases in the quantity and mutability of SSRs in genes involved in nervous system development and function.

## Conclusion

SSR mutability provides an abundant source of genetic variation. Most genes, including many involved with brain activity, are associated with one and often several polymorphic SSRs located in sites where repeat-number variation could affect gene function[1, 9-15]. A decade ago David Comings[12] wrote, "Our initial assumption was that like the neutral or silent single base pair polymorphisms [SNPs], the micro/minisatellite alleles would be in linkage disequilibrium with other 'critical' mutations that affect gene function. After working with these polymorphisms [as genetic markers] for several years we began to suspect that the micro/minisatellites themselves might be the 'critical' mutations." Indeed, when SSR polymorphisms within a candidate gene are closely linked with a trait of interest, the hypothesis that SSR alleles may be responsible for the trait should be considered just as carefully as the more conventional SNP alternative. While pathological effects of repeat expansion have dominated most neuroscientists' perceptions of SSRs, such pathologies probably represent only a tiny fraction of the functional effects of SSR variation.

Here we have reviewed several cases implicating SSR alleles in normal brain and behavioral function as well as susceptibility to certain neurodegenerative diseases. Addressing the questions raised by such data (see Outstanding Questions) should begin to reveal whether these examples are unusual or whether variation generated by SSRs plays a major role in neurobiological function, development, and evolution.

## Acknowledgments

JWF is supported by the Sara and Frank McKnight Fellowship in Biochemistry. EADH would like to acknowledge recent support from NIH F31MH67397, NIH T32MH65215 and NIH T32MH075883. AJH is supported by the NHMRC (Australia), a Pfizer Australia Research Fellowship and the Lord Mayor's Charitable Fund (Eldon and Anne Foote Trust).

## References

1 Nithianantharajah, J. and Hannan, A.J. (2007) Dynamic mutations as digital genetic modulators of brain development, function and dysfunction. *Bioessays* 29, 525-35

2 Orr, H.T. and Zoghbi, H.Y. (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* 30, 575-621

3 Morell, V. (1993) The puzzle of the triplet repeats. *Science* 260, 1422-1423

4 Richards, R.I. and Sutherland, G. R. (1994) Simple repeat DNA is not replicated simply. *Nat. Genet.* 6, 114-116

5 Kovtun, I.V. and McMurray, C.T. (2008) Features of trinucleotide repeat instability in vivo. *Cell Res.* 18, 198-213

6 Ellegren, H. (2004) Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.* 5, 435-445

7 Hamada, H. *et al*. (1984) Enhanced gene expression by the poly(dT-dG) · poly(dC-dA) sequence. *Mol. Cell. Biol.* 4, 2622-2630

8 Trifonov, E.N. (1989) The multiple codes of nucleotide sequences. *B. Math. Biol.* 51, 417-432

9 Gerber, H.-P. *et al*. (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 263, 808-811

10 Kashi, Y. *et al*. (1997) Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* 13, 74-78

11 King, D.G. *et al*. (1997) Evolutionary tuning knobs. *Endeavour* 21, 36-40.

12 Comings, D.E. (1998) Polygenic inheritance and micro/minisatellites. *Mol. Psychiatr*. 3, 21-31

13 King, D.G. and Soller, M. (1999) Variation and fidelity: The evolution of simple sequence repeats as functional elements in adjustable genes. In *Evolutionary Theory and Processes: Modern Perspectives* (Wasser, S.P., ed.), pp. 65–82, Kluwer Academic Publishers

14 Fondon, J.W. III and Garner, H.R. (2004) Molecular origins of rapid and continuous morphological evolution. *P. Natl. Acad. Sci. USA* 101, 18058–18063

15 Kashi, Y. and King, D.G. (2006) Simple sequence repeats as advantageous mutators in evolution. *Trends Genet*. 22, 253-259

16 Caporale, L.H. (2003) Natural selection and the emergence of a mutation phenotype: an update of the evolutionary synthesis considering mechanisms that affect genome variation. *Annu. Rev. Microbiol.* 57, 467–485

17 King, D.G. and Kashi, Y. (2007) Mutability and evolvability: Indirect selection for mutability. *Heredity* 99, 123-124.

18 Gusella, J.F. and MacDonald, M.E. (2006). Huntington's disease: seeing the pathogenic process through a genetic lens. *Trends Biochem. Sci*. 31, 533-540

19 Spires, T.L. and Hannan, A.J. (2007) Molecular mechanisms mediating pathological plasticity in Huntington's disease and Alzheimer's disease. *J. Neurochem*. 100, 874-882

20 Bear, M.F. *et al*. (2004) The mGluR theory of fragile X mental retardation. *Trends Neurosci*. 27, 370-377

21 Garber, K. *et al*. (2006) Transcription, translation and fragile X syndrome. *Curr. Opin. Genet. Dev.* 16, 270-275

22 Loesch, D.Z. *et al*. (2007) Molecular and cognitive predictors of the continuum of autistic behaviours in fragile X. *Neurosci. Biobehav. R.* 31, 315-326

23 Hagerman, R,J, *et al*. (2005) Recent advances in fragile X: a model for autism and neurodegeneration. *Curr. Opin. Psychiatr*. 18, 490-496

24 Sawyer, L.A. *et al*. (1997). Natural variation in a Drosophila clock gene and temperature compensation. *Science* 278, 2117-2120

25 Kyriacou, C.P. *et al*. (2008) Clines in clock genes: fine-tuning circadian rhythms to the environment. *Trends Genet.* 24, 124-132

26 Zamorzaeva, I. *et al*. (2005) Sequence polymorphism of candidate behavioural genes in *Drosophila melanogaster* flies from 'Evolution Canyon'. *Mol. Ecol.* **14**, 3235–3245

27 Peixoto, A.A. *et al*. (1998). Molecular coevolution within a Drosophila clock gene. *P. Natl. Acad. Sci. U S A* 95, 4475-4480

28 Johnsen, A. *et al*. (2007) Avian *Clock* gene polymorphism: evidence for a latitudinal cline in allele frequencies. *Molec. Ecol.* 16, 4867–4880

29 Hammock, E.A.D. and Young, L.J. (2005) Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* 308, 1630-1634

30 Young, L.J. *et al*. (1999) Increased affiliative response to vasopressin in mice expressing the $V_{1a}$ receptor from a monogamous vole. *Nature* 400, 766-768.

31 Fink, S. *et al*. (2006) Mammalian monogamy is not controlled by a single gene. *P. Natl. Acad. Sci. USA* 103, 10956-10960

32 Lesch, K.-P. *et al*. (1996) Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274, 1527-1531

33 Munafò, M.R. *et al*. (2005) Does measurement instrument moderate the association between the serotonin transporter gene and anxiety-related personality traits? A meta-analysis. *Mol. Psychiatr*. 10, 415–419

34 Sen, S. *et al*. (2004) Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *Am. J. Med. Genet. B* 127B, 85-89

35 Caspi, A. *et al*. (2003) Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science* 301, 386-389

36 Bennett, A. J. *et al*. (2002). Early experience and serotonin transporter gene variation interact to influence primate CNS function. *Mol. Psychiatr*. 7, 118-122

37 Champoux, M., *et al*. (2002). Serotonin transporter gene polymorphism, differential early rearing, and behavior in rhesus monkey neonates. *Mol Psychiatr.* 7, 1058-1063

38 Bethea, C. L., *et al*. (2004). Anxious behavior and fenfluramine-induced prolactin secretion in young rhesus macaques with different alleles of the serotonin reuptake transporter polymorphism (5HTTLPR). *Behav. Genet.* 34, 295-307

39 Trefilov, A. *et al*. (2000) Natal dispersal in rhesus macaques is related to serotonin transporter gene promoter variation. *Behav. Genet*. 30, 295-301

40 Tanaka, F. *et al*. (1996) Differential pattern in tissue-specific somatic mosaicism of expanded CAG trinucleotide repeat in dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and X-linked recessive spinal and bulbar muscular atrophy. *J. Neurol. Sci.* 135, 43-50

41 Telenius, H. *et al*. (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat. Genet.* 6, 409-414 (with erratum *Nat. Genet.* 7, 113)

42 Ueno, S. *et al*. (1995) Somatic mosaicism of CAG repeat in dentatorubral-pallidoluysian atrophy (DRPLA). *Hum. Mol. Genet.* 4, 663-6

43 Choudhry, S., *et al*. (2001) CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. *Hum. Mol. Genet.* 10**,** 2437-46

44 Matsuura, T. *et al*. (2006) Interruptions in the expanded ATTCT repeat of spinocerebellar ataxia type 10: repeat purity as a disease modifier? *Am. J. Hum. Genet.* 78, 125-9

45 Mulvihill, D.J. *et al*. (2005) Effect of CAT or AGG interruptions and CpG methylation on nucleosome assembly upon trinucleotide repeats on spinocerebellar ataxia, type 1 and fragile X syndrome. *J. Biol. Chem.* 280, 4498-503

46 Sobczak, K. and Krzyzosiak, W.J. (2004) Patterns of CAG repeat interruptions in *SCA1* and *SCA2* genes in relation to repeat instability. *Hum Mutat* 24, 236-47

47 Kijas, J.M.H. *et al*. (2001) A frameshift mutation in *MC1R* and a high frequency of somatic reversions cause black spotting in pigs. *Genetics* 158, 779-785

48 Tanaka, F. *et al*. (1999) Tissue-specific somatic mosaicism in spinal and bulbar muscular atrophy is dependent on CAG-repeat length and androgen receptor-gene expression level. *Am. J. Hum. Genet.* 65, 966-73

49 Manley, K. *et al*. (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.* 23, 471-3

50 Beale, R. and Iber, D. (2006) Somatic evolution of antibody genes. In *The Implicit Genome* (Caporale, L.H., ed.) pp. 177-190, Oxford Univ. Press

51 Mosch, B. *et al*. (2007) Aneuploidy and DNA replication in the normal human brain and Alzhiemer's disease. *J. Neurosci.* 27, 6859-6867

52 Muotri, A.R. *et al*. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retroposition. *Nature* 435, 903-910

53 Karlin, S. et al. (2002) Amino acid runs in eukaryotic proteomes and disease associations. *P. Natl. Acad. Sci. USA*

54 Koob, M.D *et al*. (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat. Genet*. 21, 379-384

55 Matsuura, T. *et al*. (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat. Genet*. 26, 191-194

56 Holmes, S.E. *et al*. (1999) Expansion of a novel CAG trinucleotide repeat in the 5' region of *PPP2R2B* is associated with SCA12. *Nat. Genet*. 23, 391-392

57 Holmes, S.E. *et al*. (2001) A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nat. Genet*. 29, 377-378

58 Voncken, M. *et al*. (2004) Friedreich ataxia-update on pathogenesis and possible therapies. *Neurogenetics* 5, 1-8

59 Cho, D.H. and Tapscott, S.J. (2007) Myotonic dystrophy: Emerging mechanisms for DM1 and DM2. *Biochim. Biophys. Acta* 1772, 195-204

60 Lafrenière, R.G. *et al*. (1997) Unstable insertion in the 5' flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. *Nat. Genet*. 15, 298-302

61 D'Souza, U.M. and Craig, I.W. (2006) Functional polymorphisms in dopamine and serotonin pathway genes. *Hum. Mutat.* 27, 1-13

62 Lasky-Su, J.A. *et al*. (2005) Meta-analysis of the association between two polymorphisms in the serotonin transporter gene and affective disorders. *Am. J. Med. Genet. B* 133B, 110 –115

63 Roberts, J. *et al*. (2007) Differential regulation of the serotonin transporter gene by lithium is mediated by transcription factors, CCCTC binding protein and Y-Box binding protein 1, through the polymorphic intron 2 variable number tandem repeat. *J. Neurosci.* 27, 2793-2801

64 Ebstein, R.P. *et al*. (1996) Dopamine D4 receptor (*D4DR*) exon III polymorphism associated with the human personality trait of novelty seeking**.** *Nat. Genet.* 12, 78-80

65 Benjamin, J. *et al*. (1996) Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nat. Genet.* 12, 81−84

66 Cornish, K.M. *et al*. (2005) Association of the dopamine transporter (DAT1) 10/10-repeat genotype with ADHD symptoms and response inhibition in a general population sample *Mol. Psychiatr.* 10**,** 686–698

67 Schott, B.H. *et al*. (2006) The dopaminergic midbrain participates in human episodic memory formation: Evidence from genetic imaging. *J. Neurosci.* 26, 1407-1417

68 Pedrosa, E. *et al*. (2007) Positive association of schizophrenia to *JARID2* gene. *Am. J. Med. Genet. B* 144B, 45–51

69 Manning, J.T. (2007) The androgen receptor gene: A major modifier of speed of neuronal transmission and intelligence? *Med. Hypotheses* 68, 802-804

70 de Quervain, D.J.-F. *et al*. (2007) A deletion variant of the α2b-adrenoceptor is related to emotional memory in Europeans and Africans. *Nat. Neurosci*. 10, 1137 - 1139

71 Knafo, A. *et al*. (2008) Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor RS3 promotor region and correlation between RS3 length and hippocampal mRNA. *Genes Brain Behav*. 7, 266-275.

72 Gonitel, R. *et al*. (2008) DNA instability in postmitotic neurons. *P. Natl. Acad. Sci. USA* 105, 3467–3472

**Glossary**

**Allele** -- a distinct form of DNA sequence at a particular chromosomal locus, differing in base-pair sequence from other alleles at that locus.

**Coding sequences** -- Those portions of genes (exons) that are translated into proteins. In conventional usage, all other sequences, whether or not they have a function, are 'non-coding'.

**Gene** -- a tract of DNA consisting not only of coding sequences but also introns and associated non-coding upstream and downstream regulatory elements.

**Linkage disequilibrium** -- the statistical association between alleles at two (or more) genetic loci, due to proximity along the same chromosome. Linked alleles are inherited as a unit until they become separated by recombination (e.g., by crossing-over between homologous alleles during meiosis).

**Linkage analysis** -- localization of genetic regions responsible for a trait of interest by identifying markers that are associated by linkage with inheritance of the trait. Since SNPs and polymorphic SSRs are widely distributed throughout the genome, most traits will be closely linked with one or more such markers. A marker associated by linkage with a trait of interest may be described as a "genetic risk factor" for that trait, but by itself such correlation does not imply any causal role.

**Microsatellite** -- an SSR whose repeating motif is between one and six basepairs in length (the upper limit is arbitrary and may be defined as low as five or as high as ten).

**Minisatellite** -- an SSR whose motif-length is greater than the upper limit for microsatellites (i.e., longer than 5-10 basepairs).

**Motif** -- a short sequence of DNA basepairs, repeated in tandem to form an SSR. With increasing motif length, the number of possible motif sequences increases. Thus there are two distinct mononucleotide motifs (A/T and C/G), six distinct dinucleotides, ten distinct trinucleotides, etc. (Motifs that can be matched by shifting along the sequence or by reading in either direction are considered equivalent; e.g., CAG = AGC = GCA = CTG = TGC = GCT.)

**Polymorphism** -- a genetic locus (specific location along a chromosome) at which two or more alleles each occur at appreciable frequencies within a population.

**Slippage mutation** -- an increase or decrease in the number of tandem motifs in an SSR; although SSR mutational processes are complex[4,05], basically slippage mutations occur during DNA replication when one strand realigns with the complementary strand after shifting ("slipping") by one or more motif-lengths.

**SSR** -- Simple Sequence Repeat (alternatively STR, Short Tandem Repeat); a DNA tract consisting of a relatively short basepair motif that is repeated several to many times in tandem. Polymorphic SSRs may be called VNTRs (variable number tandem repeats).

**SNP** -- Single Nucleotide Polymorphism; an allelic difference at a single basepair. Although some SNPs may be causally responsible for some phenotypic differences (traits), SNPs may also serve in linkage analysis as markers ("genetic risk factors") for traits of interest without any implication of causal role.