Opinion

Cell PRESS

Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'

Anthony J. Hannan[1,2]

[1] Howard Florey Institute, Florey Neuroscience Institutes, University of Melbourne, VIC 3010, Australia
[2] Department of Anatomy and Cell Biology, University of Melbourne, VIC 3010, Australia

A problem of 'missing heritability' has been identified following recent genome-wide association (GWA) studies of single nucleotide polymorphisms (SNPs) associated with complex diseases. Current GWA studies fail to detect key sources of genetic variation, particularly tandem-repeat polymorphisms (TRPs), which provide a unique source of genetic variability by modulating a range of biological processes. Expanded tandem repeats cause various monogenic disorders, including Huntington's disease and various ataxias. However, there is emerging evidence suggesting that TRPs have a role in polygenic diseases. For example, candidate gene studies have found associations between specific TRPs and various brain disorders. Future GWA studies that include all TRPs as genetic variables will reveal the full extent of their association with complex diseases. TRPs might provide substantial genetic variability contributing to complex polygenic diseases and could be an important source of the missing heritability evident in SNP-based GWA studies.

The function of TRPs: biology, disease and evolution

DNA sequences that are repeated in tandem are common throughout the genomes of a wide range of species, including humans, and are often highly conserved, suggesting that they have an important function [1–3]. Most research on tandem repeats has focused on those in the human genome: as microsatellites used as genetic markers and as repeat expansion mutations causing dominant or recessive disorders with Mendelian inheritance patterns. The term 'dynamic mutations' has been used to describe the variation in tandem repeat length that has been found in many human genes to cause a range of repeat-expansion disorders, particularly those affecting the nervous system including Huntington's and other polyglutamine diseases, Friedreich's ataxia and fragile X syndrome (for reviews, see Refs [2,4,5]). It should be noted that one of the genes, *FMR1*, which contains a trinucleotide expansion implicated in fragile X syndrome, has further associations of premutation repeat lengths (below the disease threshold necessary for fragile X syndrome) with more complex disorders, including tremor/ataxia, Parkinsonism, neuropsychiatric symptoms and premature ovarian failure [1,6,7].

A much larger array of tandem repeats are present in, and between, genes that are not known to be involved directly in diseases of Mendelian inheritance. Tandem repeats are also referred to more specifically as simple sequence repeats (SSRs), microsatellites or minisatellites (Box 1). These repetitive sequences can be located in exons, introns or intergenic regions (Figure 1), providing opportunities for the modulation of gene expression, as well as the structure and function of RNAs and proteins (e.g. codon repeats translated into amino acid runs).

Tandem repeats contribute ~3% of the human genome, which is a larger proportion of the genome than the entire sum of protein coding sequences [8]. Most tandem repeats examined thus far are polymorphic in nature and some have formed the basis of microsatellite mapping, although the extent of polymorphism across a broad cross-section of the human population has yet to be investigated systematically. DNA is a digital information encoding system with the nucleotide as the smallest indivisible unit. Tandem repeat lengths can have an extended digital (multiallelic) distribution, as opposed to the usual binary (biallelic) nature of single nucleotide polymorphisms (SNPs), thus providing a wide and variable range of possible genotypes at a given locus [5]. Although short tandem repeat motifs, such as mono-, di-, tri- and tetranucleotides, are common in the human genome, longer motifs are frequently observed, leading to a wide range of polymorphic possibilities.

Tandem repeats can affect a variety of biological processes, including evolution, development, brain function and behaviour [1,3,5,9–12]. They can serve as 'tuning knobs' in evolution [3,9], which might occur via impacts on genetic programs regulating development and other key biological processes. Recent data suggesting that tandem repeats can increase the 'evolvability' of promoter sequences [10] provides further evidence that they have functional consequences and they provide a rich source of genetic diversity that can facilitate evolutionary processes, via actions on development, biological function and behaviour.

Corresponding author: Hannan, A.J. (anthony.hannan@florey.edu.au).

0168-9525/$ – see front matter doi:10.1016/j.tig.2009.11.008 Available online 28 December 2009

Box 1. Different types of repetitive DNA sequences in the human genome

Bioinformatics analysis indicates that there is some bias regarding the distribution of the hundreds of thousands of unique tandem repeats throughout the human genome [4,68,69]. A large proportion of tandem repeats are located in introns and intergenic regions [70]. There is also some specificity with respect to different types of repeat motif and their genomic distribution patterns. For example, trinucleotide and hexanucleotide repeats are more likely to be found in exons [70], and coding regions in particular, where their expansion or contraction will lead to altered lengths of amino acid runs (e.g. polyhistidine tracts [71]), but not catastrophic frameshifts.

Tandem repeats exhibit extended digital (multiallelic) polymorphic distributions; for example, $(XY)^{n+z}$, where X and Y are the bases in a dinucleotide motif, n is the minimum repeat length ($n > 1$), and $z = 0, 1, 2, 3, 4, 5$ etc., until $n + z =$ maximum repeat length. This is in contrast to the binary (biallelic) nature of the vast majority of single nucleotide polymorphisms (SNPs), where either nucleotide X or Y may be present.

Here, I use the term tandem repeats to describe tandemly repeated DNA sequences also known as SSRs, or satellite DNA (which includes both microsatellites and minisatellites). In addition to the dinucleotide example mentioned above, tandem repeats can involve mononucleotides, trinucleotides (triplets), tetranucleotides etc. The definition of microsatellites ranges from 1–5 to 1–10 base pairs in length. Thus, tandem repeats with a motif longer than the upper limit for microsatellites (i.e. >10 base pairs in length) are generally defined as minisatellites. For a more in-depth discussion of the nomenclature and the extent of polymorphism of tandem repeats, see recent articles [1,26]. TRPs, or SSR polymorphisms, have also been referred to as variable-number tandem repeats (VNTRs) or repeat-length polymorphisms (RLPs).

One other common form of genomic polymorphism involving differential repetition of DNA sequences is the copy number variants (CNV). Typical definitions of CNVs generally involve variations in copy numbers of chromosomal segments ranging in length from ~1 kb to multiple Mb of DNA. Recently, CNVs have been implicated in a number of polygenic disorders, including schizophrenia and autism [72,73]. In fact, polymorphic long tandem repeats might be considered to be a form of copy number variation, in their broadest definition [74]. Conversely, a broad definition of CNVs could include variations in tandem repeat lengths [72].

A striking example of a role for tandem repeat polymorphisms (TRPs) in basic biological processes comes from a genetic study of dog breeds, identifying TRPs in the coding regions of genes that appear to control transcriptional programs regulating morphological development [11]. Studies of the vasopressin receptor in a monogamous species of voles have implicated tandem repeats in the regulation of brain function and behaviour [12]. Furthermore, correlative data suggest that specific TRPs can modulate serotonergic, dopaminergic, adrenergic, vasopressin and oxytocin systems, as well as other neurodevelopmental and signalling systems, thus affecting human brain development, behaviour and cognition [1,13–21]. A caveat associated with any candidate gene study, is that not all associations will stand the test of replication in larger cohorts and subsequent meta-analyses [22,23]. Nevertheless, one implication of the evidence for TRPs modulating normal development and function is that they might be involved in dysfunction and pathogenesis, beyond monogenic diseases.

A role for tandem repeat polymorphisms in common polygenic disorders?

Emerging evidence suggests that TRPs might play a key role in modulating disease susceptibility for a range of common polygenic disorders. The characteristics of TRPs are consistent with possible roles in predisposition to biological dysfunction associated with polygenic disorders. TRPs are widely distributed throughout the genome, with a recent bioinformatics analysis suggesting that short tandem repeats occur more frequently in disease-associated genes than those not known to be associated with disease [25]. The presence of TRPs in exons, introns and intergenic regions provides scope for modulation of a variety of molecular processes (Figure 1). This, in turn, could lead to functionally significant variability at cellular and systems levels. A specific role for polymorphisms in satellite DNA as regulators of gene expression, and thus potential contributors to quantitative traits associated with complex disorders, has been proposed [16]. Accumulating data in the past decade are consistent with these ideas, and extend the impact of TRPs to a broader range of molecular processes, and have provided candidate gene associations with specific diseases.

Recent findings have begun to identify TRPs as potential genetic susceptibility loci for a range of disorders [26]. Given the evidence that TRPs function in development, nervous system physiology and behaviour, it is not surprising that many of the associated disorders involve CNS dysfunction. Brain disorders that have been found to show associations with specific TRPs include depression [27–30], bipolar disorder [31–34], schizophrenia [35–40], attention-deficit hyperactivity disorder (ADHD) [17,41,42], Alzheimer's disease and associated behavioural and/or psychological symptoms [43–48], and stroke [49,50]. Examples of these associations of TRPs with various neurological and psychiatric disorders are summarised in Table 1.

There is some evidence for TRP associations with diseases not primarily affecting the nervous system. For example, specific tandem repeat loci have been associated with particular forms of cardiovascular disease [51,52], diabetes [53,54] and cancer [55–58]. One of the first non-neurological diseases to be associated with a specific TRP was a familial form of colorectal cancer [55]. Since then tandem repeat instability has been found to be associated with a range of different sporadic cancers [58].

Many of these studies involve small sample sizes and require validation in large independent cohorts. Some of those reported associations, involving relatively modest cohort sizes, could constitute false positives and might not translate to larger populations. Conversely, only a few TRPs and a limited number of disorders have been examined thus far, so it is likely that genome-wide approaches could lead to a much larger collection of TRP–disease associations. Furthermore, to demonstrate that any specific tandem repeat polymorphism is causatively involved in these various disorders is a more challenging task that needs to be addressed

How might TRPs alter biological processes associated with disease susceptibility?

TRPs can modulate gene expression, but when they are found as trinucleotide repeats in coding regions can have additional roles in modulating protein function (Figure 1). TRPs in non-transcribed genomic regions, including promoters and intergenic DNA, could influence epigenetics

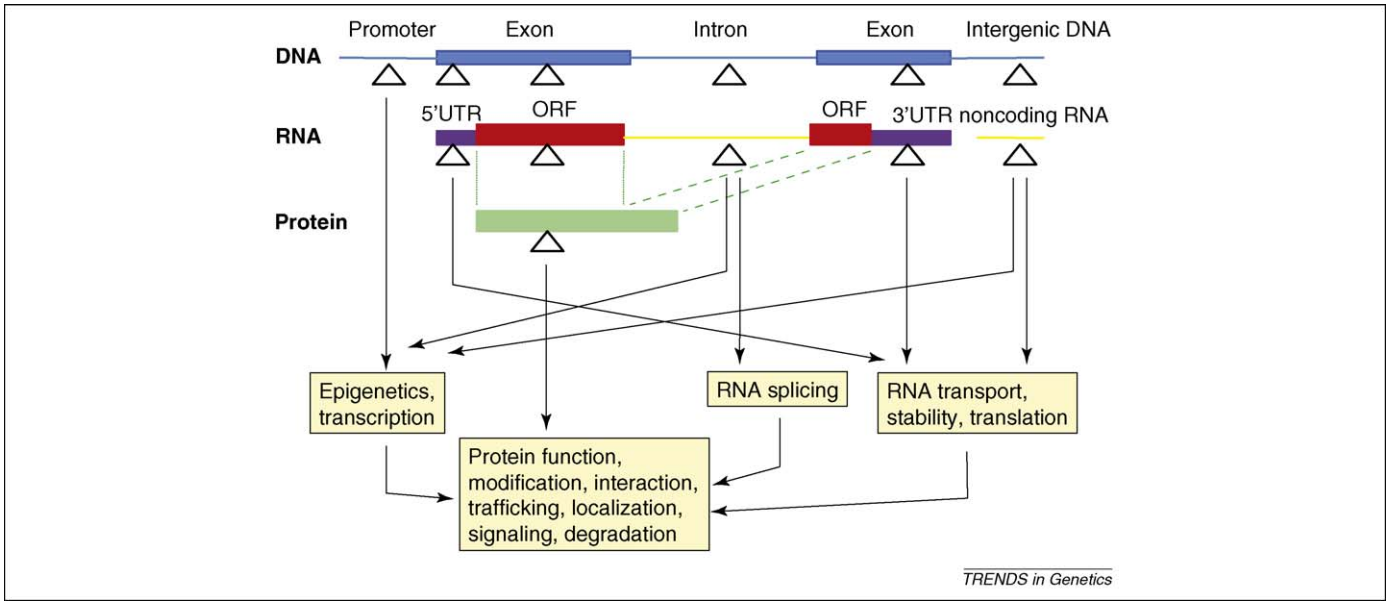

Figure 1. The positions of tandem repeat polymorphisms (TRPs) in the genome and their potential effects. TRPs are indicated by triangles and can be found within promoters, exons and introns of genes, and in intergenic regions. For transcribed regions of the genome, this leads to TRPs in a variety of different RNAs. For those TRPs consisting of in-frame trinucleotides (or less commonly of hexanucleotides and other motifs that are multiples of 3) this leads to polymorphic amino acid runs in the translated proteins. Thus, a wide range of molecular processes can be disrupted due to the effects of TRPs at the DNA, RNA and protein levels. Abbreviations: ORF, open reading frame; UTR, untranslated region.

and transcriptional regulation, and striking examples of this have been shown in yeast [10,59]. Untranslated regions (UTRs) of transcripts have key roles in molecular processes, including RNA transport, stability and translation, all of which could be modulated by TRPs. Furthermore, those TRPs (generally consisting of trinucleotide motifs, but occasionally hexanucleotide motifs and longer multiples of three) encoding amino acid runs in proteins [60], have the potential to influence various aspects of protein processing and function (Figure 1).

Another possible functional impact of TRPs, which might be associated with age-dependent disease susceptibility, is that particular repeat sequences could lead to 'molecular misreading', and thus pathological alterations in repetitive RNA and amino acid sequences. There is evidence that at least some forms of molecular misreading can contribute to increased vulnerability to selective neurodegenerative processes with aging [44,61].

Some unique functions of TRPs are associated with specific sequence repeats. For example, it has been shown recently that specific trinucleotide repeats in transcripts can serve as a target for the ribonuclease Dicer to produce short CUG repeats that can participate in RNA interference mechanisms mediating selective silencing of gene

Table 1. Tandem repeat polymorphisms associated with increased susceptibility to polygenic brain disorders[a,b]

Disorder	Gene	Motif size	TRP type	Functional consequence?	Refs
Depression and anxiety disorders	*5-HTT*	17-44	Noncoding	5-HTT levels	[27,29,30]
	MAOA	30	Noncoding	MAOA levels	[28]
Bipolar disorder	*5-HTT*	17-44	Noncoding	5-HTT levels	[31,33]
	BDNF	2	Noncoding	BDNF levels	[32]
	Per3	18	Coding	Per3 function	[34]
Schizophrenia	*BDNF*	2	Noncoding	BDNF levels	[35]
	GCLC	3	Noncoding	GCLC levels	[39]
	JARID2	4	Coding	JARID2 function	[38]
	NOS1	2	Noncoding	NOS levels	[36]
	NUMBL	3	Coding (CAG)	NUMBL function	[37]
	TBP	3	Coding (CAG)	AR function	[40]
ADHD	*DAT1*	40	Noncoding	DAT levels	[41]
	DRD4	48	Coding	DRD4 function	[42]
Alzheimer's disease	*NOS1*	2	Noncoding	NOS1 levels	[47]
	AR	3	Coding (CAG)	AR function	[43,48]
AD (BPSD)	*5-HTT*	17-44	Noncoding	5-HTT levels	[45,46]
Stroke	*IL1RN*	86	Noncoding	IL1RN levels	[50,51]
	GPIbα	39	Coding	GPIbα function	[49]

[a]This table is not exhaustive, but rather provides representative examples. Many of these studies involve small samples and require validation in large independent cohorts. Studies that have only used microsatellites as mapping tools have been excluded. Furthermore, the table excludes monogenic tandem repeat expansion disorders, with Mendelian inheritance patterns (including those caused by trinucleotide repeat expansions in the *FMR1* gene), which were reviewed recently [1,2]. Space constraints preclude citing all relevant work in the literature, and therefore the earliest valid reports are cited.

[b]Abbreviations: 5-HTT, serotonin transporter; AD (BPSD), Alzheimer's disease (behavioural and psychological symptoms in dementia); ADHD, attention-deficit hyperactivity disorder; AR, androgen receptor; BDNF, brain-derived neurotrophic factor; DAT1, dopamine transporter 1; DRD, dopamine receptor; GPIbα, glycoprotein Ibalpha; GCLC, glutamate cysteine ligase catalytic subunit; IL1RN, interleukin-1 receptor antagonist; MAOA, monoamine oxidase A; NOS, nitric oxide synthase; TBP, TATA-binding protein; TRP, tandem repeat polymorphism.

expression [62]. The impact of TRPs on these and other non-coding RNAs, and associated epigenetic processes [63], could be important for various aspects of development, physiology and behaviour.

Meiotic, mitotic and postmitotic instability of TRPs

The discussion so far has been in the context of a tacit assumption that as an individual develops essentially all cells in the body will have the same repeat number at a given TRP locus. However, evidence for mitotic and postmitotic instability of TRPs raises an additional possibility that this instability in somatic cells of developing and adult individuals is associated with specific diseases. Although this hypothesis is difficult to test, because it requires comparison of repeat lengths across different somatic tissues, this potential for mitotic instability is one characteristic by which TRPs can differ from other kinds of polymorphisms, and is therefore worthy of future investigation.

TRPs differ substantially from SNPs in their susceptibility to meiotic, mitotic and postmitotic instability. SNPs are highly stable, with single nucleotide substitutions that are not corrected by DNA repair being rare genetic events, and are not generally expected to alter during meiosis or mitosis in an individual. However, the evidence for instability of tandem repeats [3,64] suggests that particular sequence motifs and repeat lengths (e.g. the longest alleles) at a given locus are more susceptible to instability during meiosis, and in mitotic cells of developing and adult tissues. One possibility, which was proposed earlier, is that tandem repeat mutation during mitotic cell divisions could give rise to a whole new layer of cellular diversity during development [5]. Furthermore, some evidence for repeat instability in postmitotic neurons, with TRPs consisting of long CAG trinucleotide repeats, has recently been provided [65]. However, further investigation is required to determine whether this phenomenon is common, functionally significant and extends to other types of tandem repeats.

Thus, meiotic instability of TRPs could lead to an increase (or decrease) in genetic susceptibility for a given disorder between generations. It is possible that mitotic and postmitotic instability of tandem repeats also contribute to disease susceptibility, although this is more difficult to prove in humans because it requires tissue and cell-specific analysis of repeat lengths to demonstrate mitotic instability and its functional consequences. There is some evidence that an age-dependent decrease in tandem repeat stability during mitosis could be associated with an increase in incidence of specific cancers [55,58]. One other implication of this proposal would be that tandem repeat instability during development could give rise to particular groups of cells and tissues with altered repeat numbers for a given gene. Such a mechanism would be expected to show variability between individuals, and the resultant epidemiological noise would make it more difficult to pick up such effects based solely on genotyping from readily accessible cell types (e.g. white blood cells). However, advances in the capacity to sequence genomic DNA rapidly from small tissue biopsies should make the exploration of mitotic and postmitotic repeat instability feasible in the near future.

Can TRPs help us find 'missing heritability' not accounted for by genome-wide SNPs?

Many common human disorders are polygenic in origin, and their genetic complexity is reflected in the difficulties encountered when attempting to identify the disease genes and variants involved. One likely explanation for the complexity of the genetics underlying such disorders is that they consist, at least partly, of polymorphisms at quantitative trait loci (QTLs) which, in particular genomic combinations and associated with specific environmental factors, lead to disease states. In such cases, genetic susceptibility loci might be expected to greatly overlap with QTLs for biological traits of relevance to each disorder.

Recent genome-wide association (GWA) studies have revealed candidates for many complex polygenic disorders. For example, a recent study has collated a database over 110 GWA studies that includes >56,000 genotype–phenotype associations, including disorders ranging from depression and schizophrenia to cardiovascular disease and cancer [66]. GWA studies continue unabated and their reported genetic associations are accumulating in various databases (e.g. http://www.ncbi.nlm.nih.gov/gap). However, the reported SNPs do not fully account for the major genetic contributions to such disorders, leaving the conundrum of 'missing heritability' [24]. This might be because GWA studies have relied on analysis of SNPs, comparing numerous SNPs between case and control groups, but missing much of the genetic variability that constitutes non-SNP polymorphisms, including TRPs. Similarly, the vast majority of candidate-gene studies of complex diseases in recent decades have focused almost entirely on SNPs [67]. Therefore, in order to ensure that future 'genome wide' association studies capture the full spectrum of genomic variance, they will need to investigate SNPs and other pivotal polymorphisms, particularly TRPs and other structural variants (Box 2).

Concluding remarks

TRPs can modulate a variety of biological processes, including gene transcription, protein function, morphological development, behaviour and physiology, and they are responsible for many unstable repeat expansion disorders in humans. An initial focus on SNPs in GWA studies have led to TRPs being largely overlooked as a major source of genetic variability with functional consequences. Recent discoveries of TRPs associated with common polygenic disorders have provided tantalising initial evidence supporting a proposed role in disease susceptibility, however genome-wide approaches in large cohorts are required to fully test this idea.

Specific tandem repeat lengths at different loci have been associated with various brain disorders, as well as cancers and metabolic and cardiovascular disorders. Tandem repeats are distributed across a variety of genes, including promoters, exons and introns, emphasising the need to consider all genomic possibilities in the search for genetic associations with complex diseases. Of course, for a particular complex polygenic disorder, a wide array of SNPs, TRPs and other polymorphisms will be expected to contribute to quantitative traits associated with increased genetic susceptibility. Therefore, more sophisticated approaches

Box 2. Key questions for future research

1. **What is the extent of polymorphism of tandem repeats across the whole genome?** It would be anticipated that current international approaches, including the 1000 Genomes Project (http://www.1000genomes.org/), will allow the full range of repeat lengths at all tandem repeat loci to be characterised. A recent study has compared trinucleotide repeats across three human genomes, the 'reference genome' and those of James Watson and J. Craig Ventor [69], providing a framework for more extensive investigations of this kind, involving much larger numbers of genomes.
2. **Are specific repeat lengths at TRP loci associated with increased disease susceptibility?** This question was raised earlier [16,75] but it has remained largely unexplored because most studies have continued to focus on SNPs [76]. The relationship between TRPs and polygenic diseases is probably complex, as it is for SNPs. For example, repeat lengths affecting neuromodulatory systems that mediate elevated basal anxiety states can increase susceptibility to depressive disorders via complex effects at the molecular, cellular and systems levels. Many years of sophisticated subsequent functional genomics will be required to show how particular combinations of TRPs, SNPs and other polymorphisms are necessary or sufficient in conferring genetic susceptibility to particular diseases, allowing for the complexity of potential non-linear gene–gene and gene–environment interactions.
3. **Are the current genome-wide approaches appropriate?** There are several challenges that will be associated with the analysis of TRPs at the whole genome level. If genome-wide TRPs are to be examined systematically for potential associations, then statistical approaches will need to factor in this added complexity, while fully utilizing tandem repeat databases [68,69] as they evolve. For example, for some diseases a specific repeat length at a given allele can contribute differentially to susceptibility (and thus be functionally multiallelic), whereas others will show a threshold effect (and thus be functionally biallelic). Thus, the current statistical approaches used to analyse SNPs in GWA studies might not be adequate to assess genome-wide associations of TRPs with specific diseases.
4. **What is the relationship between TRPs and other forms of structural polymorphism associated with human diseases?** TRPs in the human genome could be structurally and functionally related to copy number variants (CNVs). As the role of CNVs in the regulation and dysregulation of gene expression begins to be explored in more detail, their functional parallels with TRPs can be fully investigated. For example, one key question involves the effects of cell type and age on the stability of TRPs and CNVs in meiotic, mitotic and postmitotic cells (see Box 1 for a discussion of these forms of instability). Thus, more detailed investigation is required regarding the roles of variation in repetitive DNA sequences, both as TRPs and CNVs, in biological function and disease susceptibility.
5. **How have TRPs evolved?** Studies on the evolution of TRPs and other structural polymorphisms across species will be highly informative with respect to their function. For example, comparison of human TRP loci with their homologues in related species such as chimpanzees [69], with different cognitive, behavioural, physiological and morphological endophenotypes, will be highly informative in identifying those tandem repeats that are functionally conserved in evolution, with respect to species exhibiting both similar and different phenotypic traits.

are required, which analyse and integrate these various types of polymorphisms, as well as modelling gene–gene interactions and gene–environment interactions associated with each disorder.

Whole genome sequencing across large numbers of individuals will be needed to establish the extent of polymorphism at each tandem repeat locus. Existing evidence suggests that the majority of such tandem repeats will be polymorphic [68], and thus able to differentially influence molecular processes such as gene expression and protein function. Therefore, just as many recent GWA studies have compared SNPs across the genome of large numbers of subjects suffering from a given disorder with control data sets, genome-wide association studies of TRPs are not only becoming feasible and could be considered essential. Next-generation DNA sequencing approaches will facilitate exploration of structural variation across the human genome and the functional associations of TRPs with polygenic disorders, and thus future development of novel diagnostic, prophylactic and therapeutic approaches.

Acknowledgements

I am greatly indebted to the anonymous referees for their highly constructive feedback and excellent suggestions. Thanks to S.M. Miller for his insightful comments on an earlier draft of the manuscript. I appreciate the support of the NHMRC and a Pfizer Australia Research Fellowship. I apologise to those colleagues whose work has not been cited owing to space constraints.

References

1 Fondon, J.W., III *et al.* (2008) Simple sequence repeats: genetic modulators of brain function and behavior. *Trends Neurosci.* 31, 328–334
2 Usdin, K. (2008) The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome Res.* 18, 1011–1019
3 Kashi, Y. and King, D.G. (2006) Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 22, 253–259
4 Sutherland, G.R. *et al.* (1998) Fragile sites still breaking. *Trends Genet.* 14, 501–516
5 Nithianantharajah, J. and Hannan, A.J. (2007) Dynamic mutations as digital genetic modulators of brain development, function and dysfunction. *BioEssays* 29, 525–535
6 Coffey, S.M. *et al.* (2008) Expanded clinical phenotype of women with the FMR1 premutation. *Am. J. Med. Genet. A* 146A, 1009–1016
7 Bourgeois, J.A. *et al.* (2009) A review of fragile X premutation disorders: expanding the psychiatric perspective. *J. Clin. Psychiatry* 70, 852–862
8 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
9 Kashi, Y. *et al.* (1997) Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* 13, 74–78
10 Vinces, M.D. *et al.* (2009) Unstable tandem repeats in promoters confer transcriptional evolvability. *Science* 324, 1213–1216
11 Fondon, J.W., III and Garner, H.R. (2004) Molecular origins of rapid and continuous morphological evolution. *Proc. Natl Acad. Sci. U. S. A.* 101, 18058–18063
12 Hammock, E.A.D. and Young, L.J. (2005) Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* 308, 1630–1634
13 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
14 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
15 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
16 Comings, D.E. *et al.* (1998) Polygenic inheritance and micro/minisatellites. *Mol. Psychiatry* 3, 21–31
17 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. Psychiatry* 10, 686–698
18 de Quervain, DJ-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

19 Prichard, Z.M. *et al.* (2007) Association analysis of 15 polymorphisms within 10 candidate genes for antisocial behavioural traits. *Psychiatr. Genet.* 17, 299–303
20 Prichard, Z.M. *et al.* (2007) AVPR1A and OXTR polymorphisms are associated with sexual and reproductive behavioral phenotypes in humans. Mutation in brief no. 981. Online. *Hum. Mutat.* 28, 1150
21 Riley, D.E. and Krieger, J.N. (2009) Embryonic nervous system genes predominate in searches for dinucleotide simple sequence repeats flanked by conserved sequences. *Gene* 429, 74–79
22 Munafò, M.R. *et al.* (2008) Association of the dopamine D4 receptor (DRD4) gene and approach-related personality traits: meta-analysis and new data. *Biol. Psychiatry* 63, 197–206
23 Munafò, M.R. *et al.* (2009) 5-HTTLPR genotype and anxiety-related personality traits: a meta-analysis and new data. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 150B, 271–281
24 Manolio, T.A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature* 461, 747–753
25 Madsen, B.E. *et al.* (2008) Short tandem repeats in human exons: a target for disease mutations. *BMC Genomics* 9, 410
26 Bacolla, A. *et al.* (2008) Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. *Genome Res.* 18, 1545–1553
27 Ogilvie, A.D. *et al.* (1996) Polymorphism in serotonin transporter gene associated with susceptibility to major depression. *Lancet* 347, 731–733
28 Schulze, T.G. *et al.* (2000) Association between a functional polymorphism in the monoamine oxidase A gene promoter and major depressive disorder. *Am. J. Med. Genet.* 96, 801–803
29 Kishida, I. *et al.* (2007) Monoamine metabolites level in CSF is related to the 5-HTT gene polymorphism in treatment-resistant depression. *Neuropsychopharmacology* 32, 2143–2151
30 Wray, N.R. *et al.* (2009) Accurate, large-scale genotyping of 5HTTLPR and flanking single nucleotide polymorphisms in an association study of depression, anxiety, and personality measures. *Biol. Psychiatry* 66, 468–476
31 Rees, M. *et al.* (1997) Association studies of bipolar disorder at the human serotonin transporter gene (hSERT; 5HTT). *Mol. Psychiatry* 2, 398–402
32 Neves-Pereira, M. *et al.* (2002) The brain-derived neurotrophic factor gene confers susceptibility to bipolar disorder: evidence from a family-based association study. *Am. J. Hum. Genet.* 71, 651–655
33 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 Neuropsychiatr. Genet.* 133B, 110–115
34 Benedetti, F. *et al.* (2008) A length polymorphism in the circadian clock gene Per3 influences age at onset of bipolar disorder. *Neurosci. Lett.* 445, 184–187
35 Krebs, M.O. *et al.* (2000) Brain derived neurotrophic factor (BDNF) gene variants association with age at onset and therapeutic response in schizophrenia. *Mol. Psychiatry* 5, 558–562
36 Reif, A. *et al.* (2006) A neuronal nitric oxide synthase (NOS-I) haplotype associated with schizophrenia modifies prefrontal cortex function. *Mol. Psychiatry* 11, 286–300
37 Passos Gregorio, S. *et al.* (2006) Analysis of coding-polymorphisms in NOTCH-related genes reveals NUMBL poly-glutamine repeat to be associated with schizophrenia in Brazilian and Danish subjects. *Schizophr. Res.* 88, 275–282
38 Pedrosa, E. *et al.* (2007) Positive association of schizophrenia to JARID2 gene. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 144B, 45–51
39 Gysin, R. *et al.* (2007) Impaired glutathione synthesis in schizophrenia: convergent genetic and functional evidence. *Proc. Natl Acad. Sci. U. S. A.* 104, 16621–16626
40 Ohi, K. *et al.* (2009) TATA box-binding protein gene is associated with risk for schizophrenia, age at onset and prefrontal function. *Genes Brain Behav.* 8, 473–480
41 Franke, B. *et al.* (2008) Association of the dopamine transporter (SLC6A3/DAT1) gene 9-6 haplotype with adult ADHD. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 147B, 1576–1579
42 Johnson, K.A. *et al.* (2008) Absence of the 7-repeat variant of the DRD4 VNTR is associated with drifting sustained attention in children with ADHD but not in controls. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 147B, 927–937
43 Lehmann, D.J. *et al.* (2003) Association of the androgen receptor CAG repeat polymorphism with Alzheimer's disease in men. *Neurosci. Lett.* 340, 87–90
44 Hill, D. *et al.* (2007) Differential enrichment of simple sequence repeats in selected Alzheimer-associated genes. *Cell Mol. Biol. (Noisy-le-grand)* 53, 23–31
45 Pritchard, A.L. *et al.* (2007) Role of serotonin transporter polymorphisms in the behavioural and psychological symptoms in probable Alzheimer disease patients. *Dement. Geriatr. Cogn. Disord.* 24, 201–206
46 Ueki, A. *et al.* (2007) Serotonin transporter gene polymorphism and BPSD in mild Alzheimer's disease. *J. Alzheimers Dis.* 12, 245–253
47 Galimberti, D. *et al.* (2008) Association of a NOS1 promoter repeat with Alzheimer's disease. *Neurobiol. Aging* 29, 1359–1365
48 Lieberman, A.P. and Robins, D.M. (2008) The androgen receptor's CAG/glutamine tract in mouse models of neurological disease and cancer. *J. Alzheimers Dis.* 14, 247–255
49 Cervera, A. *et al.* (2007) The BC genotype of the VNTR polymorphism of platelet glycoprotein Ibalpha is overrepresented in patients with recurrent stroke regardless of aspirin therapy. *Cerebrovasc. Dis.* 24, 242–246
50 Worrall, B.B. *et al.* SWISS, ISGS, and MSGD Investigators (2007) IL1RN VNTR polymorphism in ischemic stroke: analysis in 3 populations. *Stroke* 38, 1189–1196
51 Worrall, B.B. *et al.* (2003) Interleukin-1 receptor antagonist gene polymorphisms in carotid atherosclerosis. *Stroke* 34, 790–793
52 Kamstrup, P.R. *et al.* (2008) Pentanucleotide repeat polymorphism, lipoprotein(a) levels, and risk of ischemic heart disease. *J. Clin. Endocrinol. Metab.* 93, 3769–3776
53 Hiromine, Y. *et al.* (2007) Trinucleotide repeats of programmed cell death-1 gene are associated with susceptibility to type 1 diabetes mellitus. *Metabolism* 56, 905–909
54 Galanakis, E. *et al.* (2008) Intron 4 a/b polymorphism of the endothelial nitric oxide synthase gene is associated with both type 1 and type 2 diabetes in a genetically homogeneous population. *Hum. Immunol.* 69, 279–283
55 Laken, S.J. *et al.* (1997) Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat. Genet.* 17, 79–83
56 McIntyre, M.H. *et al.* (2007) Prostate cancer risk and ESR1 TA, ESR2 CA repeat polymorphisms. *Cancer Epidemiol. Biomarkers Prev.* 16, 2233–2236
57 Schildkraut, J.M. *et al.* (2007) Trinucleotide repeat polymorphisms in the androgen receptor gene and risk of ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* 16, 473–480
58 Haberman, Y. *et al.* (2008) Trinucleotide repeats are prevalent among cancer-related genes. *Trends Genet.* 24, 14–18
59 Verstrepen, K.J. *et al.* (2005) Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37, 986–990
60 Faux, N.G. *et al.* (2005) Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res.* 15, 537–551
61 van Den Hurk, W.H. *et al.* (2001) Novel frameshift mutations near short simple repeats. *J Biol. Chem.* 276, 11496–11498
62 Krol, J. *et al.* (2007) Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Mol. Cell* 25, 575–586
63 Mattick, J.S. *et al.* (2009) RNA regulation of epigenetic processes. *BioEssays* 31, 51–59
64 Dion, V. and Wilson, J.H. (2009) Instability and chromatin structure of expanded trinucleotide repeats. *Trends Genet.* 25, 288–297
65 Gonitel, R. *et al.* (2008) DNA instability in postmitotic neurons. *Proc. Natl Acad. Sci. U. S. A.* 105, 3467
66 Johnson, A.D. and O'Donnell, C.J. (2009) An open access database of genome-wide association results. *BMC Med. Genet.* 10, 6
67 Tabor, H.K. *et al.* (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet.* 3, 391–397
68 O'Dushlaine, C.T. and Shields, D.C. (2008) Marked variation in predicted and observed variability of tandem repeat loci across the human genome. *BMC Genomics* 9, 175
69 Molla, M. *et al.* (2009) Triplet repeat length bias and variation in the human transcriptome. *Proc. Natl Acad. Sci. U. S. A.* 106, 17095–17100

70 Subramanian, S. *et al.* (2003) Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol.* 4, R13
71 Salichs, E. *et al.* (2008) Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PloS Genet.* 5, e1000397. doi:10.1371/journal.pgen.1000397 (www.plosgenetics.org)
72 Conrad, D.F. *et al.* (2009) Origins and functional impact of copy number variation in the human genome. *Nature* doi:10.1038/nature08516 (www.nature.com), Epub ahead of print, doi:10.1038/nature08516z
73 Zhang, F. *et al.* (2009) Complex human chromosomal and genomic rearrangements. *Trends Genet.* 25, 298–307
74 Bruce, H.A. *et al.* (2009) Long tandem repeats as a form of genomic copy number variation: structure and length polymorphism of a chromosome 5p repeat in control and schizophrenia populations. *Psychiatr. Genet.* 19, 64–71
75 Armour, J.A. (2006) Tandemly repeated DNA: why should anyone care? *Mutat. Res.* 598, 6–14
76 Bodmer, W. and Bonilla, C. (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* 40, 695–701