Genomics 101 (2013) 273–281

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Review

The VNTR in complex disorders: The forgotten polymorphisms? A functional way forward?

K.J. Brookes

Faculty of Life Sciences, Michael Smith Building, University of Manchester, Manchester, UK

ARTICLE INFO

Article history:
Received 3 February 2013
Accepted 11 March 2013
Available online 19 March 2013

Keywords:
Variable Number Tandem Repeat
Functional
Polymorphism
Review
Gene expression

ABSTRACT

In the last few years, research has focused on single nucleotide polymorphisms (SNPs) in the search for underlying genetic aetiology of complex disorders. This has been afforded by the rapid technological advancement to enable the interrogation of hundreds of thousands of SNPs in one assay *via* microarrays. However SNPs are only one form of genetic variation and in the midst of the Genome-Wide Association Study (GWAS) explosion Variable Number Tandem Repeat (VNTR) polymorphism exploration has seemingly been left behind. This review will argue that VNTR investigations still hold substantial potential for a role in complex disorders *via* possible functional properties.

Contents

1. Introduction: what are complex disorders & VNTRs . . . 273
2. Current state of play . . . 274
3. The potential of VNTR polymorphisms . . . 274
4. Coding polymorphisms . . . 274
5. Promoter polymorphisms . . . 275
6. Other polymorphisms . . . 276
7. Points for consideration . . . 276
8. The future . . . 279
Acknowledgments . . . 279
References . . . 279

1. Introduction: what are complex disorders & VNTRs

Many heritable medical disorders, termed 'complex disorders' display no clear Mendelian pattern of inheritance, despite this, evidence from family studies, and twin concordance rates indicate strong heritability, and genetic components to their aetiology. Complex disorders are rarely the result of single gene mutations and have three main features. First, complex disorders are a product of many genes working with and against each other, for which none are essential or sufficient for disorder manifestation. In addition these genes will influence and be influenced by additional environmental components. Secondly, complex disorders are likely to have a high degree of both environmental and genetic heterogeneity within its aetiology. Finally, clinical heterogeneity such as symptom profile, severity, and age of onset creates many sub-types within one disorder [112].

Variable Number Tandem Repeat (VNTR) is an encompassing term for a DNA sequence motif that is repeated several times in the genome continuously and that are inherited in a Mendelian fashion. Despite these VNTRs often being categorised and ignored as 'junk' DNA; the number of times the sequence is repeated can differ within and between individuals therefore making these VNTRs a polymorphic entity. The term VNTR covers both micro-satellites that are classified as motifs of 1–6 bp long and mini-satellites that are made up of longer continuous motif sequence blocks which can span hundreds of bps. VNTRs are thought to have arisen due to slippages during DNA replication or as a consequence of unequal crossing-over, and due to the repetitive nature of the sequence have allowed further increases/decreases in

E-mail address: keeley.brookes@manchester.ac.uk.

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.03.003

the number of repeats resulting in the polymorphisms observed [106]. They are estimated to represent 3% of the genome and are frequently found in promoters and potential functional locations within genes, suggesting that they may have an impact on several phenotypes [35,68].

2. Current state of play

The decades of research to identify underlying gene variations responsible for complex disorders have; despite some moderate successes; largely failed in the task to reliably distinguish causal variants. Initial linkage studies of multi-affected families were generally underpowered to identify the genes, which underlie the aetiology of such disorders, suggesting that the original estimates of effect sizes for the genetic variations were inflated, and the number of genes involved was conservative [103]. A move towards the 'common disorder, common polymorphism' ideology followed, which believed that the underlying genetic aetiology would come from polymorphisms that were found frequently in the population, and not rare mutations that have often been identified in single-gene disorders [67].

With the advances made in genotyping technologies, laboratories moved to candidate gene association studies. Identifying genes with biological evidence to play a role in the aetiology and testing the polymorphisms that lay within for association with their disorder of choice. Over the years, projects grew from investigating single polymorphisms to several polymorphisms and genes at once. Due to the limited ability to only investigate single nucleotide polymorphisms that created/abolished restriction enzyme cut sites, VNTR polymorphisms became a popular choice for investigation. Sadly it became ever more obvious that the effect sizes of the genes involved were smaller than the field ever imagined, and to attain enough power to detect them bigger sample sizes were called for. Due to the cost of collecting larger samples, the cost in genotyping needed to decrease. SNP genotyping led the field with the advent of microarray technology allowing thousands of SNPs to be genotyped on a single array. As technology progressed, the genotyping of SNP markers became easier and cheaper, and soon hundreds of thousands of SNP markers could be genotyped at once, and the dawn of Genome-Wide Association Study (GWAS) began.

Despite the success in the development of high-throughput SNP genotyping, high-throughput VNTR genotyping has not been developed. The very nature of the VNTRs led to technical problems with genotyping on the array platforms that SNP genotyping utilises, complicating the scoring of the number of repeats present. Furthermore, VNTR polymorphisms were often named after their location in the gene and had no structured nomenclature or human specific databases such as SNPs with their rs# logged into national database such dbSNP for easy access to information [109]. Finally the complex mutation pattern observed in VNTR polymorphisms led to different alleles being associated with the same disorder between populations; for example the *DRD4* gene exon 3 VNTR association with ADHD is with different alleles for European (7-repeat) and Asian (2-repeat) populations due to when the alleles evolved and the selective pressures on them, therefore adding complexity in interpreting results of the causal risk variant [17,127].

3. The potential of VNTR polymorphisms

Although GWAS has fulfilled its task and identified a number of replicable novel SNP association findings [52,123]; it only explains a small amount of the heritability behind complex disorders [76–78]. In addition, the association findings tend to be isolated SNPs that are intergenic, and to assign functional content to them will not be easy. It is feasible that single base changes in the DNA sequence have functional consequences for example they may alter the binding affinity of proteins that either enhance or repress gene expression, however their effect is also likely to depend on the presence or absence of other factors. In addition SNPs identified in GWAS do not amount to the estimates calculated by quantitative genetic studies; and so we are left with "missing heritability", which has been hypothesised to lie with other polymorphism such as VNTRs [46], however, an in-depth discussion of this phenomenon is outside the scope of this review and is adequately discussed elsewhere in the literature [76–78,98].

For these reasons I would like to argue in favour of the investigation of VNTR polymorphisms and their potential functional consequences. If one takes a look into the literature, there are thousands of publications on VNTR polymorphisms; detailing their associations with disorders (or failures to find association), their interactions with the environment and other genes with multiple review articles for any given disorder, and therefore the association findings will not be discussed here. However there are also many publications about how VNTR polymorphisms affect the functionality of the gene.

Due to the nature of VNTR polymorphisms, they can be proposed to be fine tuners of gene expression. As common polymorphisms are likely to have small effect sizes edging the risk by a single 'notch' towards/away for a given phenotype; the multiple alleles of a given VNTR allow more scope for functional consequences than perhaps biallelic SNPs and therefore may be easier to observe. In addition the greater potential which VNTR polymorphisms have for mutation means that they may have a role to play in the evolution of specific higher organism traits, such as behaviour [35].

Using examples from the literature, I wish to highlight the functional consequences of the VNTR alleles and propose that VNTR polymorphisms should not be forgotten and hope that this review will re-ignite the study of VNTR polymorphisms within complex disorders.

4. Coding polymorphisms

The most likely candidates for functional VNTR polymorphisms are those that lie within the coding regions of the gene, therefore directly affecting the protein product. An example of this is the VNTR that lies within the third exon of the dopamine receptor D4 (*DRD4*) gene. Originally identified in the early 1990s [70], it has received ample attention due to its speculated involvement in Attention Deficient Hyperactivity Disorder [65], and its high affinity for clozapine, a treatment for schizophrenia [120].

This 48 bp repeat unit has been observed to extend between 2 and 11 imperfect tandem repeats, with the most common alleles being the 2-, 4- and 7-repeat, although there is population variation in frequency [29]. The tandem repeat affects the putative third cytoplasmic loop of the receptor protein, altering its length by 16 amino acids with each tandem repeat. When activated by an agonist, *via* G-protein coupling, such as dopamine, the receptor is thought to inhibit cAMP production by decreasing the activity of adenylyl cyclase [24,120] and therefore it might be feasible to believe that the VNTR polymorphism would affect the receptor function. Initial reports investigating the VNTR effect on pharmacological action in cell lines suggested that the binding affinity of clozapine, an antagonist of the receptor, was similar for the three common alleles of the VNTR (2-, 4- and 7-repeats). However, in the presence of sodium chloride, the binding affinity of clozapine increased for constructs containing the 2- and 4-repeat alleles, with the 7-repeat allele showing a similar binding affinity as in the absence of sodium chloride. This would suggest that the effect of the VNTR is only observed in certain conditions, and that the 7-repeat allele might have reduced sensitivity to the salt concentration in the cell. Despite the observation of these small pharmacological differences between alleles, further studies found little impact on pharmacological action in constructs with the smallest and longest tandem repeats [57], and those with the VNTR deleted [4].

A later study investigated the level of cAMP formation in order to observe if the alleles had functional relevance rather than of pharmacological action [3]. Observation of forskolin stimulated dopamine inhibition of cAMP levels in cell lines containing constructs of the three common alleles suggested that there was a blunted response in cells carrying the 7-repeat allele, with the 2- and 4-repeat alleles displaying

greater inhibition of adenylyl cyclase than the 7-repeat allele [3]. The 2-repeat allele has also been shown to exhibit a blunted response in comparison to the 4-repeat allele, but midway between that and the 7-repeat allele [29,127]. In addition, a study by Kazmi et al. [60] investigated the effects of the VNTR variants on the G-protein coupling properties of the DRD4 receptor. Using two different methods, the group confirmed that the third cytoplasm loop was required for G-protein coupling; however there was no quantitative difference in the coupling between constructs of the 2-, 4- and 7-repeat alleles.

Recently, it has been identified that dopamine itself acts as a chaperone protein to stimulate dopamine receptor D4 protein synthesis, in a dose-dependent manner [119]. The exon 3 polymorphism is thought to affect the extent to which dopamine can increase protein folding and therefore synthesis. Van Craenenbroeck et al. [119] investigated the three common alleles of the polymorphism and their chaperone induced up-regulation by dopamine. Dopamine was found to up-regulate the alleles to different extents, with the 2-repeat showing the least up-regulation from baseline (without dopamine), and the 7-repeat displaying the highest level of up-regulation, with the 4-repeat displaying an intermediate response. Interestingly the 7-repeat had the lowest levels of protein synthesis at baseline, although not significantly different from the other alleles.

In silico analysis of the VNTR sequence has identified the presence of potential Src Homology 3 (SH3) binding motifs in the repeated 16 amino acid sequence of the third cytoplasmic loop, which provides modular binding domains for protein–protein interactions. Synthesised labelled fragments, of the most common allele (4-repeat allele), were found to bind strongly to other proteins that had SH3 binding domains; NCK1 and Growth factor Receptor Bound protein 2 (GRB2) [90]. When full-length cDNA constructs of the three common alleles were compared, the 7-repeat allele was found to display the strongest interaction with NCK1 and GRB2 proteins *versus* the 2- and 4-repeat alleles [90]. Constructs with a deletion of the SH3 domains, failed to be able to inhibit cAMP formation *via* forskolin stimulated dopamine binding, although binding affinity did not differ, suggesting that the protein interactions facilitated through the SH3 domains, might regulate the action of the receptor [90].

Finally, other studies have investigated whether the VNTR affects *DRD4* gene expression, both *in vitro* [107] and *in vivo* [111]. Electrophoretic mobility shift assays of a 72 bp construct containing a single repeat unit and 12 bp flanking sequence either side, with nuclear extract from HeLa cells suggest that the VNTR repeat unit is capable of binding a protein. Increases in the concentration of the nuclear extract revealed several bands, suggesting that multiple different proteins can bind to the genomic sequence of the repeat unit [107]. Follow-up luciferase assays comparing constructs of the three most common alleles displayed little expression variation between them when placed within the promoter, but displayed a significant difference when residing in the 3′UTR. The observations suggested that there was decreased expression of the reporter when the construct contained the 7-repeat allele in comparison to the shorter alleles [107]. This observation has been supported by *in vivo* investigations of *DRD4* gene expression in human post-mortem pre-frontal cortex samples. Simpson et al. [111] compared expression of the *DRD4* gene between human samples with different exon 3 VNTR alleles. Despite no significant results, a weak trend was observed for lower gene expression in individuals with one or more 7-repeat alleles at the VNTR polymorphism.

5. Promoter polymorphisms

Gene promoters are prime targets in the search for regulatory polymorphisms due to their potential to affect the binding of the transcriptional machinery. Therefore it is unsurprising that many functional VNTR polymorphisms have been identified in this area. A prime example is that of tri-nucleotide expansion polymorphisms found in Fragile-X Syndrome and Huntington's disease [116].

Within the literature there are several examples of promoter VNTR polymorphisms associated with complex disorders with the potential to be functional [126,128,136]. In particular there are a number of studies that concentrate on the promoter VNTR polymorphisms within the Insulin (*INS*) gene, Serotonin Transporter (*5HTT*/*SLC6A4*) gene and the Monoamine Oxidase A (*MAOA*) gene. The functionality of the serotonin transporter gene polymorphism has recently been comprehensively covered elsewhere and therefore shall not be included here [44].

One of the most successful VNTR polymorphism associations has been between the VNTR within the promoter region of the insulin gene and Type 1 Diabetes (T1D). The insulin gene lies on the long arm of chromosome 11 with a VNTR polymorphism residing 365 bp upstream to the transcription start site, consisting of an imperfect 14–15 bp tandem repeat sequence [10,11]. The VNTR has been classed into three alleles dependant on its length: Class I consists of 28–44 repeat units; Class II 45–137 repeat units and Class III 138–159 repeat units. Class II alleles are very rare in populations other than in African samples, and in Caucasian samples the Class I allele has a frequency of approximately 70%. It is the Class I alleles that have been found to be associated with T1D, repeatedly across investigations and populations [6,10,41,101,125].

The same VNTR has also been associated with polycystic ovary syndrome [34], type 2 diabetes [80], and other related traits such as obesity, BMI [50,51] and weight at birth [91,93], however these associations have had less replicated success [15,47,74,133].

Despite some question over the association of the alleles with disorders and traits; functional analyses suggest potential for the VNTR to alter expression of the mRNA in a tissue specific fashion. *In vitro* reporter gene constructs of the VNTR alleles, in both fetal rat islets and hamster insulin producing β-cell lines, display similar evidence suggesting the Class III alleles result in higher expression of the insulin gene, in comparison to the shorter Class I alleles [61]. In addition, this group also found evidence suggesting that the Pur-1 transcription factor binds the repeat unit, and is the mechanism behind the transcription level. Compared to a reporter construct without the VNTR, transcription was increased 25% with the presence of Class I alleles, and 200% with Class III alleles when in the presence of Pur-1 [61]. This work was replicated in human fetal thymus cells, which also indicated that Class III alleles had higher expression than Class I [118]. However conflicting data from studies in both human adult and fetal pancreas cells suggest that Class III alleles have a marginally lower gene expression than Class I alleles [12,13,73,117], and may reflect the availability of Pur-1 in different cells.

In a more recent study, it has been proposed that the autoimmune regulator (*AIRE*) gene which underlies Autoimmune Polyendrocrinopathy Type 1 Syndrome (APS1) might regulate the insulin gene *via* the VNTR [19]. The insulin promoter Class I alleles have been found to be associated with APS1 in individuals with T1D [94]. Using hamster pancreatic β-cell lines, Class I allele constructs were found to have higher expression than Class III, although both classes were found to be similar when AIRE was overexpressed in these cells. In contrast, when using human thymic epithelial cell lines derived from six individual DNA samples, they observed that the Class III alleles displayed higher transcriptional activity than the Class I alleles in the presence of AIRE. In addition, DNA–protein binding experiments found that AIRE bound to both Class I and Class III constructs [19]. With this evidence and that from previous data, Cai and colleagues hypothesised that within the thymus, a critical region for identifying self proteins, low expression of the insulin gene might lead to T1D, which is due to an autoimmune destruction of the pancreatic β-cells.

A different promoter VNTR considered to be functional is that located approximately 1.2 kb upstream to the first exon of the Monoamine Oxidase A (*MAOA*) gene, positioned on the X-chromosome. The 30 bp repeat unit has been identified with tandem repeats in the order of 2, 3, 3.5, 4 and 5, with the 4-repeat allele being the common allele and the 2-repeat being very rare [104]. The VNTR has been associated with a varying number of behavioural disorders and traits,

including bi-polar disorder, impulsivity and antisocial behaviour [54,72,105] and in females the longer alleles with panic disorder [27].

Initial *in vitro* studies agree that in comparison to the 3.5 and 4-repeat alleles the 3-repeat allele displayed diminished transcriptional activity. Three groups comparing constructs of the alleles in neuroblastoma cell lines found that the 3-repeat allele was associated with decreased expression [27,104], although their findings disagreed with the extent of transcriptional ability of the 5-repeat, with Sabol et al. [104] reporting that the 5-repeat showed reduced expression similar to the 3-repeat, and [27], suggesting that the 5-repeat allele displayed expression activity similar to the 3.5 and 4-repeat alleles. The third conducted by Guo et al. [42], only compared the expression of 2-, 3- and 4-repeat allele constructs, finding that expression level correlated positively with the number of repeats [42].

A further study using protein extractions of fibroblast cells cultured from male human samples measured MAOA protein concentration and there too found that extracts from those with 3-repeat alleles were lower in activity than those with the 4-repeat allele [28]. In opposition, a comparison of *MAOA* gene expression within the pineal gland from post-mortem brains of Alzheimer's patients and controls suggested that although Alzheimer's patients with the 3.5- and 4-repeat alleles display higher levels of gene expression than the 3-repeat, this trend was not observed in the control population [130]. This highlights the fact that polymorphisms may act differently in diseased and healthy tissues.

Recently, new methods for determining transcriptional activity have been developed for identifying allele specific expression in *ex vivo* samples, such as post-mortem brain tissue. The utilisation of quantitative PCR methods detects an imbalance of expression of one allele over the other in mRNA, using genomic DNA as a reference. In the case of testing a promoter VNTR such as this, a coding marker SNP that is in high linkage disequilibrium with it, is required as the VNTR polymorphism is not transcribed. This method was employed in two studies [25,97], and both found that the *MAOA* transcript was subjected to allelic expression imbalance (AEI), proposing that one allele was transcribed in preference over the other, and this AEI was indicative of a continuous trait [25]. However both studies failed to be able to correlate the AEI with the number of promoter tandem repeats. As the *MAOA* gene is located on the X-chromosome X-inactivation may have played a role in these observations (despite the knowledge that the *MAOA* escapes X-inactivation), however, AEI was not found to coincide with which chromosome was inactivated. In addition, some methylation was observed at the promoter of the gene in female subjects but not in males, suggesting another form of gene compensatory mechanism.

Despite the failure of correlating AEI with genotype at the VNTR, *in silico* analyses have identified potential sites that the VNTR could control transcription of the *MAOA* gene. The repeat unit itself lies within a CpG island, and holds three speculative methylation sites, and therefore the degree for putative methylation of the promoter could be influenced by the number of repeats present [110]. Moreover, the VNTR has been found to contain Sp1 transcription factor binding sites, therefore suggesting further functional relevance [113]. However, in both cases these *in silico* investigations need to be experimentally verified.

6. Other polymorphisms

Although the functional relevance of VNTR polymorphisms located within introns and outside the gene coding sequence, such as the 3′ UTR is less certain, they still hold potential to play a role in alternative splicing, mRNA translocation, stability and translation efficiency [82]. An example of this is found in the dopamine transporter (*DAT1*/*SLC6A3*) gene, located on chromosome 5. Like many of the other VNTRs mentioned here, this one has also been associated with behavioural traits and namely a robust candidate for Attention Hyperactivity Deficit Disorder [26]. The 48 bp repeat unit is found in tandem repeats of 7 through to 11, with 9 and 10-repeat units being the most common. As with most VNTRs sequence variation has been observed amongst the repeats [84].

Unfortunately the functional studies of this polymorphism have shown less concord than the others discussed here. Over the years there have been numerous studies that have tested the functional effects of the VNTR polymorphism in the 3′UTR of the *DAT1* gene using reporter gene assays and transient transfection in mammalian cell cultures [39,40,81,84,85,121].

The VNTR was initially deemed a functional candidate when the insertion of the VNTR construct containing the 9-repeat allele upstream to a SV40 promoter with GFP reporter gene, transfected into midbrain derived cells from neonatal rats displayed enhanced expression of the gene in comparison to the construct without the VNTR [81].

Comparison of all observed alleles of the VNTR polymorphism suggested that the 10-repeat allele induced the highest expression of the reporter gene [39]. Contrastingly, Miller & Madras [85] found that the 9-repeat allele enhanced gene expression to a higher degree than the 10-repeat allele. More recently VanNess et al. [121], created four *DAT1* constructs; two constructs contained the 3′UTR VNTR one with the 9-repeat allele, one with the 10-repeat allele, the third with just the coding region and a fourth with a truncated 3′UTR prior to the VNTR location. DAT binding assays that were performed in human embryonic kidney HEK-293 cells indicated that the 10-repeat allele of the VNTR had 50% higher DAT binding than the 9-repeat allele. However constructs lacking the VNTR displayed the highest DAT binding of all, indicating that the VNTR may control the level of DAT protein produced. Yet other *in vitro* studies have revealed no significant differences in transcription between the 9 and 10-repeat alleles when cloned downstream of the luciferase gene and in the presence of either the SV40 or the *DAT1* homologous promoters in SK-N-SH, SN4741, HEK293 or SH-SY5Y cells [40,84], therefore indicating the VNTR may affect translation efficiency rather than transcription.

Yeast-one hybrid screens using the 10-repeat VNTR allele as bait identified a potential modulator of expression [38]. This led to a series of experiments that investigated the level of reporter gene expression in relation to the VNTR alleles and the presence of the transcription factor. Hairy/Enhancer of Split Related transcription factor 1 (*HESR1*/*HEY1*) was found to differentially negatively regulate the expression of a reporter gene construct dependant on the presence of different VNTR alleles. In comparison to the 10-repeat allele, the 7-repeat allele was found to result in significantly higher expression, whilst the 11-repeat allele was found to show significantly diminished expression. The 9-repeat allele displayed an intermediate level of expression between the 7- and 10-repeat alleles but did not reach significance [38]. Later research found that HESR2 and HESR3 also displayed a similar pattern of inhibition on gene expression *via* the 3′UTR polymorphism [58], potentially suggesting a negative relationship between VNTR length and gene expression.

Studies using post-mortem samples from different areas of the brain seem to converge suggesting that the 10-repeat allele conveys higher mRNA expression in comparison with individuals with the 9-repeat allele [16,83], however a recent study using a larger sample detected no association of VNTR genotype with mRNA expression levels [96]. Furthermore, allelic expression imbalance investigation of post-mortem brain tissue from Alzheimer's patients, displayed significant differences in expression between the coding marker SNP alleles, however these could not be associated with the VNTR polymorphism [96].

7. Points for consideration

As can be observed from the above examples, delineating the true functional effects of VNTR polymorphisms on gene function is by no means easy or complete. In general the picture painted by both

association and functional investigation for VNTRs is one of a high degree of complexity. Even if functional consequences can be established, and concur with directionality, the outcome of such functions might not correlate with a single disorder. Therefore many points must be considered when investigating the potential of VNTRs to play a role in the development of traits and disorders *via* functional means.

Firstly, the determination of functional properties for a VNTR will be contributed by the complement of factors, *e.g.* other proteins that act on the DNA sequence of the VNTR, present in the cell. Therefore the VNTR may seem to have functional effects in one cell type, none in another and opposite consequences in yet another. Likewise in one disorder compared to another. Therefore the observation of functionality within investigations can also depend on the methodology used, *in vivo versus in vitro* and which tissue or cell lines are utilised. Although *in silico* analyses and hypotheses are useful, they always need to be confirmed by experimental means, which as seen numerous times in the literature, generate conflicting results. These discrepancies may be due to the employment of different strategies, which offer different advantages and disadvantages.

Whilst *in vitro* investigations bring an element of control for experimental conditions, can be carried out without the need of human samples, and be easily replicated between studies; differences between cell lines, constructs and vectors between investigations can be a source for the conflicting data observed in the literature. The utilisation of different cell lines, location of the VNTR polymorphism construct within the reporter gene, and the type of reporter gene could all lead to differing effects. Additionally, different reporter constructs can exploit different promoters (viral *versus* endogenous promoters) and therefore could lead to variation between studies. Finally the different cell types may contain factors that interact with the VNTR polymorphisms that would not be found in the native *in vivo* cell, and *vice versa* [134].

For example in the exploration of the *DAT1* 3′UTR VNTR polymorphism the use of monkey epithelium cells in the Fuke et al. [39] study is a questionable substitute for representatives of the human brain cells. Whilst both the Michelhaugh et al. [40,81] studies utilised the substania nigra derived SN4741 cell lines, cells from the midbrain known to contain dopaminergic cells, the line was derived from mouse embryonic cells, which again may be questionable substitutes for human brain cells although it may be more representative of what is likely to be occurring in *in vivo* brain cells in comparison to other cells used in these studies. Furthermore, despite the utilisation of the same cell lines, the two studies observed conflicting results. Whilst Michelhaugh et al. [81] suggested that the presence of the VNTR enhanced transcription of the gene (at least with the 9-repeat allele); experiments by Greenwood and Kelsoe [40] failed to observe any increase in gene transcription. An important point to note here is that in the Michelhaugh et al. [81] study the VNTR construct was inserted upstream to the reporter gene and not in its native 3′ location. This may have altered its functional properties. The HEK-293 cells utilised in three studies [84,85,121] are derived from human embryonic kidney cells, although found to be naturally dopamine transporter producing cells, the function and usage of these transporters may not be exact to that of transporters found in the brain. In addition, the various conflicting observations between the studies could be due to the varying sequence lengths of the VNTR and surrounding DNA inserted into the reporter gene constructs.

Generally investigations using *in vivo* disease tissues would be expected to possess the same within-cell factors that impact on the VNTR DNA sequences however, investigations involving human samples, both those measuring mRNA expression and AEI although carried out in the subject and tissue of interest also have caveats that need to be addressed. Post-mortem sample sizes generally tend to be small in comparison to association studies, and could consist of varying unknown lifestyle factors, for example alcohol consumption and smoking are known to affect the level of *DAT1* gene expression [30,63,79]. Furthermore, pre-mortem (agonal) and post-mortem factors need to be considered carefully when collecting human autopsy samples. Firstly, agonal factors, which include specific conditions at the time of death and agonal duration, are known to decrease brain tissue pH and significantly affect RNA integrity and abundance and thus have a major impact on gene expression [8,99]. Secondly, post-mortem factors such as condition of the tissue after death, post-mortem interval (PMI) and duration of brain tissue storage in the freezer also affect RNA integrity [7,114]. These studies show that post-mortem intervals of 48 h do not affect expression levels. Furthermore post-mortem samples are not always from individuals that were affected with the disorder being investigated and therefore the mRNA levels observed may be due to the effects of the disorder rather than the polymorphism. However, controlling for these factors is not realistic. In addition different cell types exist in one tissue, and not all may express the gene of interest, and therefore the level of expression may depend on the number and type of cells present in the tissue. Furthermore in investigations of AEI it is questionable whether the bi-allelic SNPs used as markers truly can represent the multiple alleles of the VNTR polymorphisms.

Further complexity is added when considering that the DNA sequence of these tandem repeats is often imperfect copies of each other, which could potentially be different between constructs if not controlled for, adding yet more variation. Although this strengthens the hypotheses that VNTR polymorphisms may be functional as it gives a greater scope to create variable expression levels, it also makes it more difficult to determine the direction of the functionality and so great care must be taken when publishing to include such data as DNA sequence used in constructs, or detailed sequencing of human samples used.

For example in the case of the serotonin transport promoter VNTR where it has been reported that the short allele has lower transcriptional activity than the long allele [49,87]. However the long allele also contains an A/G SNP (rs25531) that is not present in the short allele; long alleles carrying the G-allele have been found to show activity similar to the short allele and therefore if not taken into consideration will impede investigations of functionality [53,88].

In summary the question becomes what is the best way to investigate functionality? Is it better to have every investigation using exactly the same method, or will this lead to bias in the results, due to experimental artefacts? Can we trust the observations made to be real or just due to experimental methods? So although differing procedures might not produce the same results all the time, if the effect of the VNTR polymorphism is robust enough to overcome experimental differences, so that all the data points in the same direction, it could then be construed as functional. Of course this should not mean that the experimental design can be anything, it should strive to try and mimic the endogenous *in vivo* conditions as much as possible, which means using endogenous promoters, inserting the VNTR to be explored in its natural context with as much of the native gene sequence as possible, as it is likely that the polymorphism in question works in tandem with other variants along the gene sequence. Trying to match cell lines to those where the VNTR polymorphism is speculated to have an effect is also important, so that endogenous transcription factors and other proteins that might contribute to the functionality are as similar to the *in vivo* cells as possible. It is anticipated that these types of methodologies will be used as technological advances are accomplished.

Once functionality of a VNTR polymorphism has been established to a level of certainty, the next step would be to try and determine the immediate outcome of this function. This may depend on whether the polymorphisms affect the levels of transcription, translation or some other measure of gene expression. For instance what does a decrease in the expression of the insulin gene in the thymus lead to, what are the knock-on effects, and the biochemical outcomes? Is less protein made? Is there a quantitative change in the number of T-cells produced? Likewise, how do the allelic variations in the pancreatic insulin gene expression change reactions to blood glucose levels? Can we correlate

different biological/biochemical outcomes? Although one might expect such changes to occur in logical directionality *e.g.* higher insulin production leads to a greater decrease in blood glucose levels, the question is to what extent does this occur in relation to the polymorphism under investigation, and what other factors contribute to this.

This nicely leads on to the question about how does the presence of other gene variants and environmental factors affect the consequences of the polymorphism under investigation. The original estimates of heritability of complex disorders involved only additive genetic and environmental effects, however it is unlikely to be this simple and interactions are speculated to play a role, therefore new strategies are being developed in order to address this [135]. For example, the effect of a drug used to treat a disorder may depend not only on the genotype at the point of action for the drug, but also on genes that may affect the metabolism of the drug (*e.g.* [18]). Theoretically polymorphisms within the same gene may also show both additive and interactive effects resulting in different outcomes. Hypothetically, if the presence of the 7-repeat allele at the *DRD4* gene leads to a blunted response, but another polymorphism in the promoter of the gene, such as the 120 bp duplication [108], leads to an increase in gene transcription, the elevated levels of receptor might overcome the effects of the blunted response and result in a normal downstream biochemical outcome.

With the accumulation of GWAS, there is ample data to analyse epistasis between genes and between polymorphisms, and in recent years many algorithms have been developed for this [20,132]. The results of this work may be the creation of SNP panels that are associated with disorders and traits in an empirical way, free of previous hypotheses [37,69].

The many thousands of samples present in GWAS suggest that they may be large enough to detect the effect size of the interactions [62] but even so may not appreciate the level of heterogeneity present in complex disorders. Therefore signals may be lost, in the background noise of so many samples with heterogeneous phenotypes; however, creating panels based on related functional polymorphisms such as VNTRs and investigating the biochemical/gene expression results, might overcome this, and produce clearer data.

As suggested by Caspi et al. [22,23] ignorance of environmental variables may cause some genetic influences of small effect to be overlooked in samples, therefore leading to the inconsistencies observed across samples for genetic association. For example, association studies for the *MAOA* promoter VNTR have used the classification of the low and high expressing alleles for *a priori* hypothesis for gene × environment investigations. There have been many studies investigating a modifying effect of the *MAOA* gene; with some finding that the low expressing genotype (3-repeat allele) leads to the presence of antisocial behaviour when childhood maltreatment has occurred [22,32,33,36]. However, even this has not been fully replicated with other studies failing to find the correlation [43,55]. Furthermore, the situation is complicated with investigations finding a modifying effect but with the opposing allele in female samples, hinting that gender may also play a role in this associations [5,100,124], adding yet another factor consideration.

The same can be said of when testing functionality of polymorphisms, the effect of a variant allele might only be observed when a certain environment/condition is present, for instance it could be possible that the extent of the VNTR effect on insulin gene expression might only be observed when there is a high amount of glucose in the blood.

Although complex, and difficult to interpret, investigation of interactions must continue in order to advance further in the field of complex disorders despite facing a number of hurdles. The first hurdle is that of methodology. Investigations into possible interactions have become the norm, with new ideas of how best to approach investigations forming constantly. The fact that even between a few factors, multiple interactions could occur, therefore, which ones are investigated, and if all are analysed, what of the multiple testing issues? The approach that has been suggested by Moffitt [86] might provide a good structure for such investigations. Moffitt [86] proposes that deliberate testing of gene–environmental interactions needs to take place with *a prior* hypotheses behind the interaction being tested, to limit multiple testing. They suggest that several steps be taken. The first must be consultation of quantitative genetic studies and the identification of a candidate environmental risk factor. Secondly optimisation of the environmental factor measurement must be addressed followed by identification of candidate genes relating with both the disorder and environmental factor, with plausible biological mechanisms. Testing for interaction can then commence with evaluation of whether the interaction can be extended beyond the gene–environment-disorder triad. Finally replication and meta-analyses are required to support the finding.

Another suggestion for testing such interactions and correlations [71] is to look for genetic contributions to specific environmental/biochemical triggered events, causing specific biological phenotypes rather than a disorder as a whole. The ascertainment of biochemical alterations in relation to genotype and environmental events would improve the biological plausibility of the finding, with measurement of a biochemical marker more reliable than diagnosis in some complex disorders.

In addition, and as a consequence of these interactions the epigenome may also be altered, and ultimately change the functional outcome of a polymorphic allele. Epigenetics is likely to be the mechanism by which the environment and other gene interactions exert their effects and therefore markers of epigenetic changes should also be taken into consideration when investigating functional effects of polymorphisms. Potentially, for example, methylation quantification in and around the polymorphism could be used as a proxy for any unknown variables affecting the functionality of the target polymorphism.

In the fall out of the GWAS era, it has become apparent that disorders share common underlying features, which are genetically controlled, leading to the same genes and polymorphisms being associated with different phenotypes [21,102]. In addition it is plausible that the effects of functional polymorphisms associated are both temporal and spatial, taking cues from other factors to determine one outcome over the other.

If we look closer at the different phenotypes the polymorphisms are associated with, it might convey information about the internal mechanisms they have in common and why in such cases as the insulin gene VNTR, one allele is associated with one disorder and the other with different features. T1D has been robustly associated with the presence of Class I (or absence of protective Class III) alleles at the promoter VNTR polymorphism of the insulin gene locus [6,10,41,101,125]. Although not affirmed it seems that tissue specificity may be important in the functional role of these alleles. In the thymus, which is part of the immune system, the Class I alleles seem to predispose for lower expression of the insulin gene. In the thymus, proteins expressed by the body are produced, and if the immune systems' T-cells respond to these proteins the T-cells are destroyed before circulating the body, to prevent autoimmunity. Hypothetically, if low levels of insulin are produced in the thymus, the T-cells might not detect the protein, and therefore will not mount a response, resulting in these T-cells progressing into the body, and leading to T1D. In reality this is likely to require further genetic components and environmental insults to either lower insulin production in the thymus to low enough levels to cause this kind of response.

In contrast, the Class III allele of the same polymorphism has been associated with disorders such as T2D, obesity and polycystic ovary syndrome, and with traits such as increased birth weight, head circumference and BMI [31,56,92,129]. These disorders and traits have been found to associate with each other [66,75,95]. T2D diabetes is a result of decreased insulin production from pancreatic β-cells accompanied with insulin resistance. Observations that the presence of Class III alleles at the insulin locus results in decreased expression of insulin in the pancreas [12,13,73,117], and added environmental

factors such as diet might explain the association. The observation that there is no agreement in the directionality of the associated alleles with connected biochemical events, such as response to glucose suggests that other factors, both genetic and/or environmental also play a role [2,9,15,51]. Although this is purely speculative and very much simplified it demonstrates how one polymorphism might have differing effects dependent on different factors and how if followed up one can move through biochemical changes and up to traits and disorders.

8. The future

VNTRs offer the potential to be functional units of variance, and therefore should not be forgotten when searching for genetic aetiology behind complex disorders. Although the identification and analyses of such polymorphisms present challenges, as technological methods improve like it has done for SNP detection, the exploration of their effects in the underlying biology of how systems function will become easier.

Currently the development of next-generation sequencing techniques has enabled the cheap sequencing of the entire genome. In order to be able to do this the sequence reads have become smaller covering only 50–150 bp, in comparison to the 800–900 bps produced by the older capillary based systems. Repeated regions in the DNA cause problems for the alignment of sequences, and genome assembly due to the assembly methods being unable to discern where the repeat lies resulting in difficulties for identification of tandem repeats *versus* interspersed repeats, their true copy numbers and whether the repeat is polymorphic. Next-generation sequencing with its small read sizes has made this more difficult, as ideally to be able to accurately align the sequence with the genomic reference the read has to be longer than the repeated region, incorporating unique sequences either side of the repeated unit [115].

The detection of copy number variants by sequencing methods gives hope that perhaps with the development of more sensitive algorithms (and along with increased read lengths), VNTRs will also be able to be aligned accurately [48,64,131]. Moreover, a recent study has shown that simple tandem repeats used for identification in forensic science have been genotyped with high accuracy on a short-read next generation sequencing platform [14]. It can then be hoped that the detection of *de novo* tandem repeats is not too far in the future.

Until a time where the high through-put of VNTR polymorphism genotyping is feasible it may be possible to infer VNTR alleles through LD with its surrounding SNP variants [45]. Given the number of SNP markers that can be genotyped for an individual it could be possible to use linkage disequilibrium and haplotype blocks to deduce the VNTR variant at any given locus. Recently this has been shown to be possible with the accurate calling of the serotonin promoter VNTR alleles by SNP haplotypes [122]. Single SNP markers may not be able to account for the multi-allelic variation of VNTR polymorphisms however multiple SNP haplotypes may be able to accurately call the likelihood of which allele is present. Vinkhuyzen et al. [122] were able to call the serotonin transporter promoter genotype using a two-SNP haplotype from commercially available platforms. The TA haplotype of SNP markers rs2129785 and rs11867581 was found to be in high LD ($r^2 > 0.75$) with the short allele of the VNTR, and therefore could be used as proxies to identify the presence of this variant. Using methodology such as this would allow the mass 'genotyping' of VNTRs for analysis. This could be aided by the genotyping of the samples within the International HapMap project for the most prevalent VNTRs within the literature.

Improved detection on a genome-wide scale is only half of the problem that needs to be addressed; better methods of determining functionality also need to be developed and current ones extended to different applications. For example, Chromatin Immuno-Precipitation (ChIP) assays could be a useful technique to test *in silico* predictions of transcription factor bindings that are altered by VNTR polymorphisms. Chromatin extracted from individuals with different genotypes may exhibit varying amounts of sequence bound to the transcription factor, and therefore combined with either sequencing or qPCR, could identify both SNP and VNTR polymorphisms that affect binding.

In addition, stem cell derivatives could offer an invaluable alternative to cell lines, and post-mortem *ex vivo* methodologies. Studies show that stem cells obtained from dental pulp have the potential to differentiate into several different cell types, including neural cells [1,59,89], and therefore could be used for more accurate *in vivo* conditions in reporter gene assays.

In summary, perhaps a novel framework of research into complex genetics will start with the identification of potential functional candidates, like VNTR polymorphisms. In alliance with GWAS and next-generation sequencing studies, providing gene networks to explore for functional VNTR polymorphisms, a bottom-up approach could be employed. With the utilisation of an array of techniques, functional properties, if present, with directionality could be accredited to the polymorphism being investigated. The knowledge would then be used to take educated guesses as to what biological systems will be affected and link observations of biochemical changes with other polymorphisms and environmental factors. The differences in underlying biological function, and/or structure, can then be related back to complex disorders, helping develop understanding, treatment and diagnosis along the way.

Acknowledgments

I would like to thank Professor Kevin Morgan with his help in the preparation of this manuscript.

References

[1] A. Achilleos, P.A. Trainor, Neural crest stem cells: discovery, properties and potential for therapy, Cell Res. 22 (2) (2012) 288–304.
[2] S. Ahmed, S.T. Bennett, et al., INS VNTR allelic variation and dynamic insulin secretion in healthy adult non-diabetic Caucasian subjects, Diabet. Med. 16 (11) (1999) 910–917.
[3] V. Asghari, S. Sanyal, et al., Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants, J. Neurochem. 65 (3) (1995) 1157–1165.
[4] V. Asghari, O. Schoots, et al., Dopamine D4 receptor repeat: analysis of different native and mutant forms of the human and rat genes, Mol. Pharmacol. 46 (2) (1994) 364–373.
[5] C. Aslund, N. Nordquist, et al., Maltreatment, MAOA, and delinquency: sex differences in gene–environment interaction in a large population-based cohort of adolescents, Behav. Genet. 41 (2) (2011) 262–272.
[6] T. Awata, S. Kurihara, et al., Evidence for association between the class I subset of the insulin gene minisatellite (IDDM2 locus) and IDDM in the Japanese population, Diabetes 46 (10) (1997) 1637–1642.
[7] S. Bahn, S.J. Augood, et al., Gene expression profiling in the post-mortem human brain—no cause for dismay, J. Chem. Neuroanat. 22 (1–2) (2001) 79–94.
[8] A.J. Barton, R.C. Pearson, et al., Pre- and postmortem influences on brain RNA, J. Neurochem. 61 (1) (1993) 1–11.
[9] R.A. Bazaes, C.J. Petry, et al., Insulin gene VNTR genotype is associated with insulin sensitivity and secretion in infancy, Clin. Endocrinol. (Oxf.) 59 (5) (2003) 599–603.
[10] G.I. Bell, S. Horita, et al., A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus, Diabetes 33 (2) (1984) 176–183.
[11] G.I. Bell, J.H. Karam, et al., Properties of a polymorphic DNA segment in the 5′ flanking region of the human insulin gene, Prog. Clin. Biol. Res. 103 (Pt A) (1982) 57–65.
[12] S.T. Bennett, A.M. Lucassen, et al., Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus, Nat. Genet. 9 (3) (1995) 284–292.
[13] S.T. Bennett, A.J. Wilson, et al., IDDM2-VNTR-encoded susceptibility to type 1 diabetes: dominant protection and parental transmission of alleles of the insulin gene-linked minisatellite locus, J. Autoimmun. 9 (3) (1996) 415–421.
[14] D.M. Bornman, M.E. Hester, et al., Short-read, high-throughput sequencing technology for STR genotyping, Biotechniques (2012) 1–6.
[15] N. Bouatia-Naji, F. De Graeve, et al., INS VNTR is not associated with childhood obesity in 1,023 families: a family-based study, Obesity (Silver Spring) 16 (6) (2008) 1471–1475.
[16] K.J. Brookes, B.M. Neale, et al., Relationship between VNTR polymorphisms of the human dopamine transporter gene and expression in post-mortem midbrain tissue, Am. J. Med. Genet. B Neuropsychiatr. Genet. 144B (8) (2007) 1070–1078.

[17] K.J. Brookes, X. Xu, et al., No evidence for the association of DRD4 with ADHD in a Taiwanese population within-family study, BMC Med. Genet. 6 (2005) 31.
[18] R. Cacabelos, R. Martinez, et al., Genomics of dementia: APOE- and CYP2D6-related pharmacogenetics, Int. J. Alzheimers Dis. 2012 (2012) 518901.
[19] C.Q. Cai, T. Zhang, et al., Both polymorphic variable number of tandem repeats and autoimmune regulator modulate differential expression of insulin in human thymic epithelial cells, Diabetes 60 (1) (2011) 336–344.
[20] R.M. Cantor, K. Lange, et al., Prioritizing GWAS results: a review of statistical methods and recommendations for their application, Am. J. Hum. Genet. 86 (1) (2010) 6–22.
[21] L.S. Carroll, M.J. Owen, Genetic overlap between autism, schizophrenia and bipolar disorder, Genome Med. 1 (10) (2009) 102.
[22] A. Caspi, J. McClay, et al., Role of genotype in the cycle of violence in maltreated children, Science 297 (5582) (2002) 851–854.
[23] A. Caspi, K. Sugden, et al., Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene, Science 301 (5631) (2003) 386–389.
[24] C.L. Chio, R.F. Drong, et al., D4 dopamine receptor-mediated signaling events determined in transfected Chinese hamster ovary cells, J. Biol. Chem. 269 (16) (1994) 11813–11819.
[25] E.T. Cirulli, D.B. Goldstein, In vitro assays fail to predict in vivo effects of regulatory polymorphisms, Hum. Mol. Genet. 16 (16) (2007) 1931–1939.
[26] E.H. Cook Jr., M.A. Stein, et al., Association of attention-deficit disorder and the dopamine transporter gene, Am. J. Hum. Genet. 56 (4) (1995) 993–998.
[27] J. Deckert, M. Catalano, et al., Excess of high activity monoamine oxidase A gene promoter alleles in female patients with panic disorder, Hum. Mol. Genet. 8 (4) (1999) 621–624.
[28] R.M. Denney, H. Koch, et al., Association between monoamine oxidase A activity in human male skin fibroblasts and genotype of the MAOA promoter-associated variable number tandem repeat, Hum. Genet. 105 (6) (1999) 542–551.
[29] Y.C. Ding, H.C. Chi, et al., Evidence of positive selection acting at the human dopamine receptor D4 gene locus, Proc. Natl. Acad. Sci. U. S. A. 99 (1) (2002) 309–314.
[30] A.E. Drew, A.E. Derbez, et al., Nicotinic receptor-mediated regulation of dopamine transporter activity in rat prefrontal cortex, Synapse 38 (1) (2000) 10–16.
[31] D.B. Dunger, K.K. Ong, et al., Association of the INS VNTR with size at birth. ALSPAC study team. Avon longitudinal study of pregnancy and childhood, Nat. Genet. 19 (1) (1998) 98–100.
[32] D.M. Fergusson, J.M. Boden, et al., Moderating role of the MAOA genotype in antisocial behaviour, Br. J. Psychiatry 200 (2) (2012) 116–123.
[33] D.M. Fergusson, J.M. Boden, et al., MAOA, abuse exposure and antisocial behaviour: 30-year longitudinal study, Br. J. Psychiatry 198 (6) (2011) 457–463.
[34] P. Ferk, M.P. Perme, et al., Insulin gene polymorphism in women with polycystic ovary syndrome, J. Int. Med. Res. 36 (6) (2008) 1180–1187.
[35] J.W. Fondon III, E.A. Hammock, et al., Simple sequence repeats: genetic modulators of brain function and behavior, Trends Neurosci. 31 (7) (2008) 328–334.
[36] G. Frazzetto, G. Di Lorenzo, et al., Early trauma and increased risk for physical aggression during adulthood: the moderating role of MAOA genotype, PLoS One 2 (5) (2007) e486.
[37] B.L. Fridley, J.M. Biernacka, Gene set analysis of SNP data: benefits, challenges, and future directions, Eur. J. Hum. Genet. 19 (8) (2011) 837–843.
[38] S. Fuke, N. Sasagawa, et al., Identification and characterization of the Hesr1/Hey1 as a candidate trans-acting factor on gene expression through the 3′ non-coding polymorphic region of the human dopamine transporter (DAT1) gene, J. Biochem. 137 (2) (2005) 205–216.
[39] S. Fuke, S. Suo, et al., The VNTR polymorphism of the human dopamine transporter (DAT1) gene affects gene expression, Pharmacogenomics J. 1 (2) (2001) 152–156.
[40] T.A. Greenwood, J.R. Kelsoe, Promoter and intronic variants affect the transcriptional regulation of the human dopamine transporter gene, Genomics 82 (5) (2003) 511–520.
[41] C. Guja, L. Guja, et al., Strong association of insulin gene INS-VNTR polymorphisms with type 1 diabetes in the Romanian population, Rom. J. Intern. Med. 42 (2) (2004) 313–323.
[42] G. Guo, X.M. Ou, et al., The VNTR 2 repeat in MAOA and delinquent behavior in adolescence and young adulthood: associations and MAOA promoter activity, Eur. J. Hum. Genet. 16 (5) (2008) 626–634.
[43] B.C. Haberstick, J.M. Lessem, et al., Monoamine oxidase A (MAOA) and antisocial behaviors in the presence of childhood and adolescent maltreatment, Am. J. Med. Genet. B Neuropsychiatr. Genet. 135B (1) (2005) 59–64.
[44] K. Haddley, V.J. Bubb, et al., Behavioural genetics of the serotonin transporter, Curr. Top. Behav. Neurosci. 12 (2012) 503–535.
[45] R.E. Handsaker, J.M. Korn, et al., Discovery and genotyping of genome structural polymorphism by sequencing on a population scale, Nat. Genet. 43 (3) (2011) 269–276.
[46] A.J. Hannan, Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability', Trends Genet. 26 (2) (2010) 59–65.
[47] S.K. Hansen, A.P. Gjesing, et al., Large-scale studies of the HphI insulin gene variable-number-of-tandem-repeats polymorphism in relation to type 2 diabetes mellitus and insulin release, Diabetologia 47 (6) (2004) 1079–1087.
[48] D. He, F. Hormozdiari, et al., Efficient algorithms for tandem copy number variation reconstruction in repeat-rich regions, Bioinformatics 27 (11) (2011) 1513–1520.
[49] A. Heils, A. Teufel, et al., Allelic variation of human serotonin transporter gene expression, J. Neurochem. 66 (6) (1996) 2621–2624.
[50] B. Heude, S. Dubois, et al., VNTR polymorphism of the insulin gene and childhood overweight in a general population, Obes. Res. 12 (3) (2004) 499–504.
[51] B. Heude, C.J. Petry, et al., The insulin gene variable number of tandem repeat: associations and interactions with childhood body fat mass and insulin secretion in normal children, J. Clin. Endocrinol. Metab. 91 (7) (2006) 2770–2775.
[52] L.A. Hindorff, P. Sethupathy, et al., Potential etiologic and functional implications of genome-wide association loci for human diseases and traits, Proc. Natl. Acad. Sci. U. S. A. 106 (23) (2009) 9362–9367.
[53] X.Z. Hu, R.H. Lipsky, et al., Serotonin transporter promoter gain-of-function genotypes are linked to obsessive–compulsive disorder, Am. J. Hum. Genet. 78 (5) (2006) 815–826.
[54] Y.Y. Huang, S.P. Cate, et al., An association between a functional polymorphism in the monoamine oxidase a gene promoter, impulsive traits and early abuse experiences, Neuropsychopharmacology 29 (8) (2004) 1498–1505.
[55] D. Huizinga, B.C. Haberstick, et al., Childhood maltreatment, subsequent antisocial behavior, and the role of monoamine oxidase A genotype, Biol. Psychiatry 60 (7) (2006) 677–683.
[56] S.J. Huxtable, P.J. Saker, et al., Analysis of parent–offspring trios provides evidence for linkage and association between the insulin gene and type 2 diabetes mediated exclusively through paternally transmitted class III variable number tandem repeat alleles, Diabetes 49 (1) (2000) 126–130.
[57] V. Jovanovic, H.C. Guan, et al., Comparative pharmacological and functional analysis of the human dopamine D4.2 and D4.10 receptor variants, Pharmacogenetics 9 (5) (1999) 561–568.
[58] K. Kanno, S. Ishiura, Differential effects of the HESR/HEY transcription factor family on dopamine transporter reporter gene expression via variable number of tandem repeats, J. Neurosci. Res. 89 (4) (2011) 562–575.
[59] E. Karaoz, P.C. Demircan, et al., Human dental pulp stem cells demonstrate better neural and epithelial stem cell properties than bone marrow-derived mesenchymal stem cells, Histochem. Cell Biol. 136 (4) (2011) 455–473.
[60] M.A. Kazmi, L.A. Snyder, et al., Selective reconstitution of human D4 dopamine receptor variants with Gi alpha subtypes, Biochemistry 39 (13) (2000) 3734–3744.
[61] G.C. Kennedy, M.S. German, et al., The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription, Nat. Genet. 9 (3) (1995) 293–298.
[62] P. Kraft, D. Hunter, Integrating epidemiology and genetic association: the challenge of gene–environment interaction, Philos. Trans. R. Soc. Lond. B Biol. Sci. 360 (1460) (2005) 1609–1616.
[63] K.H. Krause, S.H. Dresel, et al., Stimulant-like action of nicotine on striatal dopamine transporter in the brain of adults with attention deficit hyperactivity disorder, Int. J. Neuropsychopharmacol. 5 (2) (2002) 111–113.
[64] C.S. Ku, E.Y. Loy, et al., The discovery of human genetic variations and their use as disease markers: past, present and future, J. Hum. Genet. 55 (7) (2010) 403–415.
[65] G.J. LaHoste, J.M. Swanson, et al., Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder, Mol. Psychiatry 1 (2) (1996) 121–124.
[66] K. Lakhani, G.M. Prelevic, et al., Polycystic ovary syndrome, diabetes and cardiovascular disease: risks and risk factors, J. Obstet. Gynaecol. 24 (6) (2004) 613–621.
[67] E.S. Lander, The new genomics: global views of biology, Science 274 (5287) (1996) 536–539.
[68] E.S. Lander, L.M. Linton, et al., Initial sequencing and analysis of the human genome, Nature 409 (6822) (2001) 860–921.
[69] C. Li, J. Han, et al., Identifying disease related sub-pathways for analysis of genome-wide association studies, Gene 503 (1) (2012) 101–109.
[70] J.B. Lichter, C.L. Barr, et al., A hypervariable segment in the human dopamine receptor D4 (DRD4) gene, Hum. Mol. Genet. 2 (6) (1993) 767–773.
[71] J. Licinio, Gene–environment interactions in molecular psychiatry, Mol. Psychiatry 7 (2) (2002) 123–124.
[72] L.C. Lim, J. Powell, et al., Evidence for a genetic association between alleles of monoamine oxidase A gene and bipolar affective disorder, Am. J. Med. Genet. 60 (4) (1995) 325–331.
[73] A.M. Lucassen, G.R. Screaton, et al., Regulation of insulin gene expression by the IDDM associated, insulin locus haplotype, Hum. Mol. Genet. 4 (4) (1995) 501–506.
[74] J.A. Maas, D.O. Mook-Kanamori, et al., Insulin VNTR and IGF-1 promoter region polymorphisms are not associated with body composition in early childhood: the generation R study, Horm. Res. Paediatr. 73 (2) (2010) 120–127.
[75] M. Magnotti, W. Futterweit, Obesity and the polycystic ovary syndrome, Med. Clin. North Am. 91 (6) (2007) 1151–1168, (ix–x).
[76] B. Maher, Personal genomes: the case of the missing heritability, Nature 456 (7218) (2008) 18–21.
[77] T.A. Manolio, F.S. Collins, et al., Finding the missing heritability of complex diseases, Nature 461 (7265) (2009) 747–753.
[78] A.J. Marian, Elements of 'missing heritability', Curr. Opin. Cardiol. 27 (3) (2012) 197–201.
[79] R.D. Mayfield, R. Maiya, et al., Ethanol potentiates the function of the human dopamine transporter expressed in *Xenopus* oocytes, J. Neurochem. 79 (5) (2001) 1070–1079.
[80] J.B. Meigs, J. Dupuis, et al., The insulin gene variable number tandem repeat and risk of type 2 diabetes in a population-based sample of families and unrelated men and women, J. Clin. Endocrinol. Metab. 90 (2) (2005) 1137–1143.
[81] S.K. Michelhaugh, C. Fiskerstrand, et al., The dopamine transporter gene (SLC6A3) variable number of tandem repeats domain enhances transcription in dopamine neurons, J. Neurochem. 79 (5) (2001) 1033–1038.
[82] F. Mignone, C. Gissi, et al., Untranslated regions of mRNAs, Genome Biol. 3 (3) (2002), (REVIEWS0004).

[83] J. Mill, P. Asherson, et al., Expression of the dopamine transporter gene is regulated by the 3′ UTR VNTR: evidence from brain and lymphocytes using quantitative RT-PCR, Am. J. Med. Genet. 114 (8) (2002) 975–979.
[84] J. Mill, P. Asherson, et al., Transient expression analysis of allelic variants of a VNTR in the dopamine transporter gene (DAT1), BMC Genet. 6 (2005) 3.
[85] G.M. Miller, B.K. Madras, Polymorphisms in the 3′-untranslated region of human and monkey dopamine transporter genes affect reporter gene expression, Mol. Psychiatry 7 (1) (2002) 44–55.
[86] T.E. Moffitt, The new look of behavioral genetics in developmental psychopathology: gene–environment interplay in antisocial behaviors, Psychol. Bull. 131 (4) (2005) 533–554.
[87] O.V. Mortensen, M. Thomassen, et al., Functional analysis of a novel human serotonin transporter gene promoter in immortalized raphe cells, Brain Res. Mol. Brain Res. 68 (1–2) (1999) 141–148.
[88] M. Nakamura, S. Ueno, et al., The human serotonin transporter gene linked polymorphism (5-HTTLPR) shows ten novel allelic variants, Mol. Psychiatry 5 (1) (2000) 32–38.
[89] N. Nourbakhsh, M. Soleimani, et al., Induced in vitro differentiation of neural-like cells from human exfoliated deciduous teeth-derived stem cells, Int. J. Dev. Biol. 55 (2) (2011) 189–195.
[90] J. Oldenhof, R. Vickery, et al., SH3 binding domains in the dopamine D4 receptor, Biochemistry 37 (45) (1998) 15726–15736.
[91] K.K. Ong, D.B. Dunger, Birth weight, infant growth and insulin resistance, Eur. J. Endocrinol. 151 (Suppl. 3) (2004) U131–U139.
[92] K.K. Ong, D.I. Phillips, et al., The insulin gene VNTR, type 2 diabetes and birth weight, Nat. Genet. 21 (3) (1999) 262–263.
[93] H. Osada, K. Seki, et al., Genetic variations within the insulin gene region are associated with accelerated fetal growth, Tohoku J. Exp. Med. 212 (1) (2007) 27–34.
[94] J. Paquette, D.S. Varin, et al., Risk of autoimmune diabetes in APECED: association with short alleles of the 5′ insulin VNTR, Genes Immun. 11 (7) (2010) 590–597.
[95] R. Pasquali, A. Gambineri, Polycystic ovary syndrome: a multifaceted disease from adolescence to adult age, Ann. N. Y. Acad. Sci. 1092 (2006) 158–174.
[96] J.K. Pinsonneault, D.D. Han, et al., Dopamine transporter gene variant affecting expression in human brain is associated with bipolar disorder, Neuropsychopharmacology 36 (8) (2011) 1644–1655.
[97] J.K. Pinsonneault, A.C. Papp, et al., Allelic mRNA expression of X-linked monoamine oxidase a (MAOA) in human brain: dissection of epigenetic and genetic factors, Hum. Mol. Genet. 15 (17) (2006) 2636–2649.
[98] R. Plomin, Child development and molecular genetics: 14 years later, Child Dev. 84 (1) (2013) 104–120.
[99] P. Preece, N.J. Cairns, Quantifying mRNA in postmortem human brain: influence of gender, age at death, postmortem interval, brain pH, agonal state and inter-lobe mRNA variance, Brain Res. Mol. Brain Res. 118 (1–2) (2003) 60–71.
[100] E.C. Prom-Wormley, L.J. Eaves, et al., Monoamine oxidase A and childhood adversity as risk factors for conduct disorder in females, Psychol. Med. 39 (4) (2009) 579–590.
[101] R. Rani, A. Sood, et al., Molecular basis of predisposition to develop type 1 diabetes mellitus in North Indians, Tissue Antigens 64 (2) (2004) 145–155.
[102] C. Richard-Miceli, L.A. Criswell, Emerging patterns of genetic overlap across autoimmune disorders, Genome Med. 4 (1) (2012) 6.
[103] N. Risch, K. Merikangas, The future of genetic studies of complex human diseases, Science 273 (5281) (1996) 1516–1517.
[104] S.Z. Sabol, S. Hu, et al., A functional polymorphism in the monoamine oxidase A gene promoter, Hum. Genet. 103 (3) (1998) 273–279.
[105] J. Samochowiec, K.P. Lesch, et al., Association of a regulatory polymorphism in the promoter region of the monoamine oxidase A gene with antisocial alcoholism, Psychiatry Res. 86 (1) (1999) 67–72.
[106] C. Schlotterer, The evolution of molecular markers—just a matter of fashion? Nat. Rev. Genet. 5 (1) (2004) 63–69.
[107] O. Schoots, H.H. Van Tol, The human dopamine D4 receptor repeat sequences modulate expression, Pharmacogenomics J. 3 (6) (2003) 343–348.
[108] M.I. Seaman, J.B. Fisher, et al., Tandem duplication polymorphism upstream of the dopamine D4 receptor gene (DRD4), Am. J. Med. Genet. 88 (6) (1999) 705–709.
[109] S.T. Sherry, M.H. Ward, et al., dbSNP: the NCBI database of genetic variation, Nucleic Acids Res. 29 (1) (2001) 308–311.
[110] E. Shumay, J.S. Fowler, Identification and characterization of putative methylation targets in the MAOA locus using bioinformatic approaches, Epigenetics 5 (4) (2010) 325–342.
[111] J. Simpson, G. Vetuz, et al., The DRD4 receptor Exon 3 VNTR and 5′ SNP variants and mRNA expression in human post-mortem brain tissue, Am. J. Med. Genet. B Neuropsychiatr. Genet. 153B (6) (2010) 1228–1233.
[112] H.K. Tabor, N.J. Risch, et al., Candidate-gene approaches for studying complex genetic traits: practical considerations, Nat. Rev. Genet. 3 (5) (2002) 391–397.
[113] I. Thanseem, A. Anitha, et al., Elevated transcription factor specificity protein 1 in autistic brains alters the expression of autism candidate genes, Biol. Psychiatry 71 (5) (2012) 410–418.
[114] H. Tomita, M.P. Vawter, et al., Effect of agonal and postmortem factors on gene expression profile: quality control in microarray analyses of postmortem human brain, Biol. Psychiatry 55 (4) (2004) 346–352.
[115] T.J. Treangen, S.L. Salzberg, Repetitive DNA and next-generation sequencing: computational challenges and solutions, Nat. Rev. Genet. 13 (1) (2012) 36–46.
[116] K. Usdin, The biological effects of simple tandem repeats: lessons from the repeat expansion diseases, Genome Res. 18 (7) (2008) 1011–1019.
[117] P. Vafiadis, S.T. Bennett, et al., Imprinted and genotype-specific expression of genes at the IDDM2 locus in pancreas and leucocytes, J. Autoimmun. 9 (3) (1996) 397–403.
[118] P. Vafiadis, S.T. Bennett, et al., Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus, Nat. Genet. 15 (3) (1997) 289–292.
[119] K. Van Craenenbroeck, S.D. Clark, et al., Folding efficiency is rate-limiting in dopamine D4 receptor biogenesis, J. Biol. Chem. 280 (19) (2005) 19350–19357.
[120] H.H. Van Tol, J.R. Bunzow, et al., Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine, Nature 350 (6319) (1991) 610–614.
[121] S.H. VanNess, M.J. Owens, et al., The variable number of tandem repeats element in DAT1 regulates in vitro dopamine transporter density, BMC Genet. 6 (2005) 55.
[122] A.A. Vinkhuyzen, T. Dumenil, et al., Identification of tag haplotypes for 5HTTLPR for different genome-wide SNP platforms, Mol. Psychiatry 16 (11) (2011) 1073–1075.
[123] P.M. Visscher, M.A. Brown, et al., Five years of GWAS discovery, Am. J. Hum. Genet. 90 (1) (2012) 7–24.
[124] L.S. Wakschlag, E.O. Kistner, et al., Interaction of prenatal exposure to cigarettes and MAOA genotype in pathways to youth antisocial behavior, Mol. Psychiatry 15 (9) (2010) 928–937.
[125] M. Walter, E. Albert, et al., IDDM2/insulin VNTR modifies risk conferred by IDDM1/HLA for development of type 1 diabetes and associated autoimmunity, Diabetologia 46 (5) (2003) 712–720.
[126] D. Wang, X. Sun, et al., CYP2C9 promoter variable number tandem repeat polymorphism regulates mRNA expression in human livers, Drug Metab. Dispos. 40 (5) (2012) 884–891.
[127] E. Wang, Y.C. Ding, et al., The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus, Am. J. Hum. Genet. 74 (5) (2004) 931–944.
[128] S. Wang, M. Wang, et al., A novel variable number of tandem repeats (VNTR) polymorphism containing Sp1 binding elements in the promoter of XRCC5 is a risk factor for human bladder cancer, Mutat. Res. 638 (1–2) (2008) 26–36.
[129] D.M. Waterworth, S.T. Bennett, et al., Linkage and association of insulin gene VNTR regulatory polymorphism with polycystic ovary syndrome, Lancet 349 (9057) (1997) 986–990.
[130] Y.H. Wu, D.F. Fischer, et al., A promoter polymorphism in the monoamine oxidase A gene is associated with the pineal MAOA activity in Alzheimer's disease patients, Brain Res. 1167 (2007) 13–19.
[131] C. Xie, M.T. Tammi, CNV-seq, a new method to detect copy number variation using high-throughput sequencing, BMC Bioinforma. 10 (2009) 80.
[132] B.L. Yaspan, O.J. Veatch, Strategies for pathway analysis from GWAS data, Curr. Protoc. Hum. Genet. (2011), (Chapter 1: Unit1 20).
[133] J.H. Yun, B.H. Gu, et al., Association between INS-VNTR polymorphism and polycystic ovary syndrome in a Korean population, Gynecol. Endocrinol. 28 (7) (2012) 525–528.
[134] A.A. Zarrin, L. Malkin, et al., Comparison of CMV, RSV, SV40 viral and Vlambda1 cellular promoters in B and T lymphoid and non-lymphoid cell lines, Biochim. Biophys. Acta 1446 (1–2) (1999) 135–139.
[135] O. Zuk, E. Hechter, et al., The mystery of missing heritability: genetic interactions create phantom heritability, Proc. Natl. Acad. Sci. U. S. A. 109 (4) (2012) 1193–1198.
[136] B. Zukic, M. Radmilovic, et al., Functional analysis of the role of the TPMT gene promoter VNTR polymorphism in TPMT gene transcription, Pharmacogenomics 11 (4) (2010) 547–557.