Current Trends in Neurology
Vol. 7, 2013

Review

What role do human specific retrotransposons play in mental health and behaviour?

Abigail L. Savage, Vivien J. Bubb and John P. Quinn*

Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Liverpool, L69 3BX, UK

ABSTRACT

The transposition of mobile DNA elements has contributed to nearly half of the human genome and influenced human evolution. Members of the retrotransposable elements are still altering the human genome through their active retrotransposition and a number of these insertions have led directly to disease. They can also impact on genomic function by introducing regulatory domains potentially altering the epigenetic landscape and transcriptome. SINE-VNTR-Alus (SVAs) are the youngest of the retrotransposons and are unique to the hominids with many of the SVAs human specific. We might envisage that such human specific domains could be involved in altering expression patterns that would underpin human traits associated with, for example, higher order cognitive function. SVAs are a composite element consisting of a hexamer CCCTCT repeat, Alu-like sequence, a guanine-cytosine (GC) rich variable number tandem repeat (VNTR), a short interspersed element and poly A-tail. The structure and sequence of an SVA indicate its potential regulatory properties which include splice sites, multiple cytosine-guanine dinucleotides for methylation and runs of guanines with potential for G-quadruplex DNA formation. The differential regulation of gene expression, the response of an individual to his/her environment and predisposition to disease can all be affected by the genotype of an individual at a specific locus. SVAs have generated genetic differences between individuals whether through their presence or absence or the difference in the repeat copy number of their VNTRs, which could be a source of genetic variation that is important in modulating human behaviour and mental health. It is the ability of SVA insertions, both in the germline and in somatic cells, to affect gene expression and their contribution to genetic variation of an individual that is the focus of this review. Amongst the genes highlighted in this review are those involved in Parkinson's disease containing SVAs within their genomic locus.

KEYWORDS: retrotransposons, SVAs, genetic variation, gene expression

ABBREVIATIONS

SVA, SINE-VNTR-Alu; VNTR, variable number tandem repeat; TE, transposable element; LINE, long interspersed element; SINE, short interspersed element

INTRODUCTION

Nearly half of the human genome consists of mobile DNA termed transposable elements (TEs). TEs, despite having long been thought of as 'junk' DNA, have influenced the human genome during its evolution through mechanisms such as insertional mutagenesis, recombination events, exonisation and modulation of gene expression, contributing to the potential adaptability of the species and may be instrumental in some of the

*Corresponding author: jquinn@liv.ac.uk

mechanisms that make us human [1-5]. However these domains have not been analysed to the same extent as proximal promoters and exons for their functional significance. Whereas exons and proximal promoters represent approximately 2% of the genome, the large number of transposable elements has represented a practical barrier to their analysis. However with a better appreciation of structural and epigenetic parameters in genome regulation we are beginning to appreciate the vast potential for these domains as key to genomic regulation. To allow for a more practical and functional analysis of the role of retrotransposons we will focus on a primate specific subset termed SINE-VNTR-Alu (SVA), many of which are human specific, and constitute only 0.13% of the human genome.

There are two classes of TEs: class I or retrotransposable elements that move within the genome through a 'copy and paste' mechanism and class II elements or DNA transposons that move through a 'cut and paste' mechanism. DNA transposons encode a transposase that removes or 'cuts' the transposon from its locus in the host genome and inserts it at a different site and are thought to be no longer active within the human genome [3]. Retrotransposable elements are mobilised through a RNA intermediate that is reverse transcribed and it is this cDNA 'copy' that is inserted back into the host genome at a different locus than the source element. The retrotransposable elements can be further subdivided into two main groups: long terminal repeats (LTR) retrotransposons and non-LTR retrotransposons.

The non-LTR retrotransposons contain the only known currently active TEs in the human genome and include long interspersed elements (LINEs), short interspersed elements (SINEs) and SINE-VNTR-Alus (SVAs). The active retrotransposition of these elements have the potential to create human specific traits within our genome and even between human individuals, whether harmful or beneficial, these differences can impact on our phenotype. The cell type in which new insertions of retrotransposons occur will determine if they are passed onto the next generation [1]. New insertions into primordial germ cells or very early in development in germ cell progenitors will be passed onto following generations. Insertions into other cell types during development will not be heritable but will contribute to somatic mosaicism of the individual. This process of somatic mosaicism through retrotransposition could introduce genetic variability between individual cells [4]. Depending on the site of insertion the effect of this process could be 1) neutral, 2) positive, as may provide novel and distinct variation in regulatory control for the cell, or 3) detrimental through mutation and leading to diseases such as cancer. Somatic insertions may be affecting genomic function of individuals and even contributing to disease without being readily detectable through genetic analysis of the individuals' genomic DNA from a single source.

LINEs are the only autonomous non-LTR retrotransposons in the human genome. The majority of these are LINE-1 (L1) elements with approximately 500,000 copies constituting 18% of the human genome [2]. A full length L1 element is 6 kb in size with two open reading frames (ORF) [6] and both ORF encoded proteins are required for retrotransposition [7]. ORF1 encodes for a 40 kDa protein (ORF1p) that binds to single stranded RNA [8] and ORF2 encodes for a 150 kDa protein (ORF2p) with reverse transcriptase and endonuclease functions [9, 10]. The L1 encoded proteins demonstrate a *cis* preference for their encoding RNA to ensure functioning L1 RNA is more likely to be inserted into the host genome [11]. Retrotransposition of the L1 elements occurs through a process called target primed reverse transcription (TPRT) [12]. More than 99.9% of the L1s in the human genome are no longer active due to mutations in their ORFs or rearrangements in their structure such as inversions and truncations [3]. There are a predicted 80-100 L1 elements that are retrotransposition competent in a given human genome with a smaller number of highly active elements that are responsible for the majority of retrotransposition in the human population [13]. L1 elements have not only expanded the human genome through their own proliferation but also mobilised non-autonomous retrotransposons including SINEs and SVAs. The most successful SINE to populate the human genome is the primate specific Alu element with more than 1 million copies [3] and many different subfamilies that have been actively expanding our genome for the past 65 million years [14]. Alus are

300 bp long and their sequence originates from a processed 7SL RNA gene and contain an internal RNA polymerase III promoter to regulate their transcription [14, 15].

SVAs are the youngest of the retrotransposable elements in the human genome and are hominid specific with 2676 SVAs identified in the Hg19 using the UCSC genome browser. SVAs are a composite element with a canonical SVA consisting of a hexamer repeat (CCCTCT), an Alu-like sequence, a GC-rich variable number tandem repeat (VNTR), a SINE derived from part of the *env* gene and a 3'LTR from the HERV-K10 endogenous retrovirus and a poly A-tail [16-18] (Figure 1). SVAs vary in length from 700-4000 bp with 63% of SVA insertions in the human genome full length, containing all five domains within the canonical element [19]. A precursor of the VNTR domain found within the SVAs is present within the rhesus macaque genome. Many of these precursor elements are also present in the human genome suggesting they were retrotransposing prior to the divergence of the old world monkeys and the hominoids [20]. The precursor sequence was termed SVA2 and contains a GC-rich VNTR, a unique 3' sequence and a poly A tail with 40 copies identified in the rhesus macaque genome [21].

The SINE region of the SVAs has been used to divide them into subtypes (A-F) with the estimated age of the subtypes ranging from the oldest at 13.56 million years (A) to 3.18 million years for the youngest subtype (F) [19]. In addition to the six subtypes defined by Wang *et al.* a seventh has been identified that contains a 5' transduction of the sequence from the first exon of the MAST2 gene and the associated CpG island, and has been referred to as either CpG-SVA, MAST2 SVA or SVA F1 [22-24] contributing to the success of the subtype in its retrotransposition [25]. Over 40% of SVAs belong to the subtype D with the fewest belonging to subtypes E and F1. Subtypes E, F and F1 are human specific as are some members of SVA subtype D with a total of 864 SVA insertions within the human genome since the human-chimpanzee divergence ~6 million years ago [26]. An analysis of the human and chimpanzee genomes revealed that 46537 bp had been deleted from the human genome through the processes of SVA insertion mediated deletions and SVA recombination associated deletions [27]. The retrotransposition of SVA elements has altered our genome, creating regions of DNA unique to humans.

To date eight diseases in humans have been caused by the insertion of an SVA element: Fukuyama-type congenital muscular dystrophy, X-linked dystonia-parkinsonism, autosomal recessive hyper-cholesterolemia, X-linked agammaglobulinemia, hereditary eliptocytosis and pyropoikiolcytosis, neutral lipid storage disease with subclinical myopathy, Lynch Syndrome and leukaemia (Table 1) [28-41]. These are human specific insertions with seven located within an exon or intron highlighting some of the mechanisms, such as alternative splicing and a reduction of mRNA expression, through which SVAs may affect the human genome.

SVA retrotransposition

SVAs are mobilised by the L1 encoded protein machinery demonstrated by two separate studies [42, 43] and their retrotransposition rate is estimated at 1 in every 916 births [44]. SVA insertions show the hallmarks of LINE-1 mediated retrotransposition such as insertion at a consensus L1 endonuclease recognition motif (5'TTTTAA 3'), poly A-tails, inversions and rearrangements, target site duplications, truncations and transductions [21]. In the studies by Hancks *et al.* and Raiz *et al.*, SVAs from subtypes D, E and F1 were shown to be retrotransposition competent

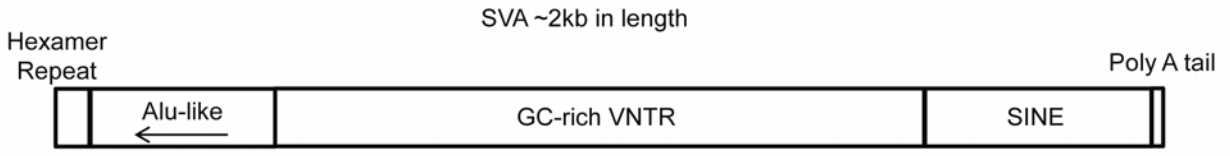

Figure 1. A schematic illustrating the structure of a canonical SVA. The diagram shows the components of an SVA which include a CCCTCT hexamer repeat, Alu-like sequence, GC-rich VNTR, SINE and a poly A-tail.

Table 1. This table summarises the SVA insertions that have been linked to disease. The information shown in this table summarises data regarding SVA insertions and their link to disease from the publications listed and the following reviews [36, 41]. Diseases: FCMD - Fukuyama-type congenital muscular dystrophy, XDP - X-linked dystonia-parkinsonism, ARH - Autosomal Recessive Hypercholesterolemia, XLA – X-linked agammaglobulinemia, HE – hereditary eliptocytosis, HPP – hereditary pyropoikiolcytosis, NLSDM – neutral lipid storage disease with subclinical myopathy.

Disease	Gene	SVA subtype	Size (kb)	Loci of insertion	Effect of insertion	Reference
FCMD	FKTN	E	3	3'UTR	Alternative splicing with exonisation of the SVA	Kobayashi *et al.* 1998 [28] Watanabe *et al.* 2005 [37] Taniguchi-Ikeda *et al.* 2011 [39]
Leukaemia	HLA-A	F1	2	-	14 kb deletion	Takasu *et al.* 2007 [35]
XDP	TAF-1	F	2.6	Intron	Tissue specific mRNA reduction	Makino *et al.* 2007 [29]
ARH	LDRAP1	F	2.6	Intron	Alternative splicing with exonisation of the SVA	Wilund *et al.* 2002 [30] Taniguchi-Ikeda *et al.* 2011[39]
XLA	BTK	-	0.25	Exon	Exon skipping	Rohrer *et al.* 1999 [31] Conley *et al.* 2005 [40]
HE and HPP	SPTA1	E	0.63	Exon	Exon skipping	Hassoun *et al.* 1994 [32] Ostertag *et al.* 2003 [38]
NLSDM	PNPLA2	F	1.8	Exon	Alternative splicing with exonisation of the SVA	Akman *et al.* 2010 [33] Taniguchi-Ikeda *et al.* 2011 [39]
Lynch Syndrome	PMS2	F	2.2	Intron	Alternative splicing with exonisation of the SVA	van der Klift *et al.* 2012 [34]

in multiple cell lines but at differing frequencies with the ORF2p essential for retrotransposition. L1 retrotransposition requires both of its ORF encoded proteins but Alus, like SVAs mobilised in *trans* by the L1 machinery, do not require the ORFp1 [45]. The L1 encoded proteins show a *cis* preference for their encoded RNA; therefore some of the non-autonomous elements may have evolved to require only ORFp2 in an attempt to increase their success. The regulation of transcription of the SVA mRNA is yet to be fully defined unlike the regulation of L1 and Alu elements. A recent study to determine the nature of SVA retrotransposition revealed that no individual domain of an SVA is fundamental for this to occur, but each domain differentially affected the rate at which retrotransposition can take place in the human osteosarcoma cell line (U2OS) [46].

Gene duplication is an important mechanism in the evolution of a species and the generation of new genes, allowing the genome to evolve with less risk as the function of genes already present can be maintained. Transduction events (both 3' and 5') during retrotransposition can result in flanking sequence of an SVA being transcribed and retrotransposed along with the SVA duplicating sections of the genome and integration at a different locus; 10% of SVA insertions have transduced sequence at their 3' end [19]. These processes provide mechanisms for creation of new exons or even duplication of genes. Approximately 53 kb of genomic sequence has been duplicated by 143 different SVA mediated 3' transduction events including the duplication of the entire acyl-malonyl condensing enzyme (*AMAC*) gene three times with at least two of the SVA transduced

genes expressed in humans and all four have retained their open reading frames [47].

Analysis of the genomic distribution of SVAs has shown that SVA density is positively correlated with gene density and GC content [19] and that they are found more frequently in genic regions as opposed to gene deserts [48]. This distribution of SVA insertions has placed them in the regions of the genome where they have the potential to influence the regulation of gene expression. The complex process of controlling gene expression is important for maintaining normal physiological processes and genetic variation in the regulatory domains play a role in phenotypic differences not only between species but also among the individuals of the same species.

Contribution of SVAs to genetic variation

Actively retrotransposing elements cause inter-individual variation among humans with elements being polymorphic for their absence or presence, SVAs included. This has been analysed for a group of human specific SVAs which estimated that 37.5% of SVA Es and 27.6% of SVA Fs were polymorphic for their presence in the genome [19] and the average human is estimated to have 56 SVA absence/presence polymorphisms [49]. The frequency of SVA retrotransposition is estimated at 1 in 916 births [44] which would result in approximately 7×10^6 private SVA insertions worldwide [4]. The exact locus or impact of these SVA insertions may only come to light if they result in disease; however these unique insertions, depending on their site of insertion, may be playing a more subtle role modulating the levels of gene expression and responding to environmental cues.

SVAs not only provide genetic variation through their presence or absence but also by the repetitive nature of their sequence within the VNTR domain of their structure. The central GC-rich VNTR was known to be polymorphic in terms of the copy number of the repeats and it has also more recently been identified that the hexamer repeat located at the 5' end of an SVA can be classed as a VNTR [17, 48]. Analysis of the structure and sequence of an SVA located upstream of the Parkinson's disease associated gene, *PARK7*, revealed four alleles in a CEU (Utah Residents with Northern and Western European Ancestry from the CEPH collection) HapMap cohort with variation in copy number of repeats located in the central GC-rich VNTR and in the 5' hexamer CCCTCT repeat [48]. This analysis also demonstrated that the central VNTR consisted of a tandem repeat (TR) and a VNTR with similar yet distinct sequences. The polymorphic nature of SVAs was assessed further in an SVA upstream of the fused in sarcoma (*FUS*) gene. This SVA, like the *PARK7* element, was found to contain a central TR and VNTR with two alleles identified; however this particular SVA is lacking a hexamer CCCTCT repeat at its 5' end (personal observations). Genetic variation is an important factor in modulating the response of an individual to his or her environment and the genetic variation located within VNTRs has been shown to regulate gene expression in a stimulus inducible and tissue specific manner [50-52]. Several VNTRs have been linked to a genetic predisposition to a specific disease, including those of the serotonin and dopamine transporter genes that have demonstrated 'risk' alleles for a variety of disorders including depression, addiction and Parkinson's disease [reviewed in 53, 54]. The VNTRs of the SVA could show similar functional properties regulating gene expression in an allele dependent manner and in response to the environment contributing to a genetic predisposition of disease or impacting on human behaviour.

We speculated that due to the primary sequence homology within distinct classes of SVAs that these elements might be able to respond to similar cellular signalling pathways. We therefore extrapolated our analysis of the *PARK7* gene to other genes associated with Parkinson's disease using a recent review by Corti *et al.* 2011 [55] as a basis; of the 13 genes in that review demonstrating a genetic association with Parkinson's disease, we found that 5 genes contained an SVA at their genomic locus (Table 2). This over representation of genes in this disease gene pathway might suggest that in part the cellular dysregulation could be directed by a concerted change in the genomic locus of genes involved in this pathway. A better understanding of genomic regulation in the future may allow us to mechanistically understand the combinatorial effect of changes by SVAs (and indeed other retrotransposons) in specific pathways rather than solely in a specific gene.

Table 2. Five of these Parkinson's disease associated genes contain an SVA. This table lists the genes that are known to be associated with Parkinson's Dissease taken from table 1 of Corti *et al.* 2011 [55]. The UCSC genome browser (Hg19) was used to identify the presence of an SVA within the gene or within 10kb. N= no SVA present (+) – on sense strand, (-) – on antisense strand. SNCA – alpha synuclien, UCHL1 – ubiquitin COOH-terminal hydrolase 1, PINK 1 – PTEN-induced kinase 1, LRRK2 – leucine rich repeat kinase 2, ATP13A2 – ATPase type 13A2, GIGYF2 – GRB10-interacting GYF protein 2, HTRA2 - HtrA serine peptidase 2, PLA2GB – group VI phopholipase A2, FBXO7 – F-box protein 7, ATXN2/SCA2 – Ataxin 2, GBA – beta glucocerebrosidase.

PARK Loci	Gene	Chr position	Involvement in PD	SVA	Chr Loci of SVA	SVA locus to gene
PARK1/4	SNCA	4q21	Early onset, dominant	N	-	-
PARK2	Parkin (-)	6q25-q27	Juvenile and early onset, recessive and sporadic	SVA F (+)	Chr6:162759277-162761189	Intron
PARK5	UCHL 1	4p14	Late onset, dominant	N	-	-
PARK6	PINK 1	1p35-p36	Early onset, recessive	N	-	-
PARK7	DJ-1 (+)	1p36	Early onset, recessive	SVA D (+)	Chr1:8012111-8013640	8kb upstream
PARK8	LRRK2 (+)	12q12	Late onset, dominant and sporadic	SVA C (-)	Chr12:40746271-40747834	Intron
PARK9	ATP13A2	1p36	Early onset recessive	N	-	-
PARK11	GIGYF2	2q36-q37	Late onset, dominant	N	-	-
PARK13	Omi/HTRA2	2p13	Unclear	N	-	-
PARK14	PLA2G6	22q12-q13	Atypical PD, recessive	Fragment of SVA F	Chr22:38549308-38549389	Intron
PARK15	FBXO7	22q12-q13	Atypical PD, recessive	N	-	-
-	ATXN2/ SCA2 (-)	12q24.1	Unclear	SVA D (+)	Chr12:111944423-111945974	Intron
-	GBA	1q21	Unclear	N	-	-

Impact of SVAs on transcriptional regulation

Retrotransposons are a source of regulatory elements providing promoters (sense and antisense), binding sites for transcription factors, donor and acceptor splice sites and polyadenylation signals that could affect gene expression [56]. Retrotransposons play an important role in the transcriptome of mammalian cells. Retrotransposons located 5' of protein coding regions can function as alternative promoters and retrotransposon derived transcriptional start sites are generally tissue specific and associate with gene dense regions [57]. 433 SVAs were identified within 10 kb upstream of the transcriptional start site of a gene within the UCSC genome browser (personal observations).

SVAs provide mechanisms to regulate transcription such as the introduction of splices sites, the formation of secondary DNA structures and multiple sites for methylation and transcription factor binding. SVAs can cause alternative splicing and exon skipping resulting in the production of differential transcripts of a gene, as illustrated by disease causing SVA insertions (see Table 1). Six out of the seven SVA disease causing insertions that are located within exons or introns have inserted into the same strand as the gene introducing additional splice sites [36, 41]. 47% of SVAs have inserted in the same orientation to the gene when located within the gene's 10 kb flank whereas only 26% of SVAs insert on the same strand as the gene when inserted within an intron or exon

(personal observations). This suggests that the insertion of an SVA in the same orientation of a gene into one of its exon or introns is more detrimental than if it had inserted into the opposite strand, perhaps due to the splice splices located in the sense strand of the SVA's sequence. The polyadenylation signal present at the 3' end of a canonical SVA insertion, if located on the same strand as a gene could also affect the transcriptional machinery causing pausing or termination of transcription. The SINE region of the SVA contains LTR sequences from the HERV-K10 which are known to contain regulatory domains and have been hypothesised to be involved in the expression of the human specific processed pseudogene *NANOGP8* and the duplicated *AMAC* genes [47, 58].

SVAs contain large domains of repetitive DNA (VNTRs) similar in copy number and size of individual repeats, that have been found to direct differential tissue specific and stimulus inducible gene expression in many genes and the copy number of those repeats have been correlated to disease predisposition [50, 51, 59-63]. Due to the young age of the SVAs they still share many similarities even across subtypes; therefore they could respond to similar stimuli throughout the genome to give a concerted response to the environment. Two SVAs have demonstrated the ability to modulate gene expression in a reporter gene model, the *PARK7* SVA *in vitro* and the *FUS* SVA both *in vitro* and *in vivo* containing multiple regulatory domains within their structure [48] (Savage *et al.* in press).

SVAs are highly GC rich, approximately 60%, with the central VNTR having a GC content of above 70% and are hypothesised to act as mobile CpG islands [19]. CpG islands are located generally at the 5' and 3' ends of genes and are associated with promoters, in particular with those of genes that are widely expressed [64, 65]. CpG islands are involved in gene regulation, genomic imprinting and X-chromosome inactivation, with hypermethylation of CpG islands associated with stable repression of transcription [66, 67]. Retrotransposons, including SVAs, are targeted for methylation to prevent their retrotransposition and potential detrimental effects associated with their insertions. SVAs could therefore potentially act as CpG islands at the site of their insertion, influencing the neighbouring genomic locus repressing the expression of the nearby genes. This is exemplified by the disease causing SVA insertion into the *TAF1* gene. The insertion of an SVA into intron 32 of the *TAF1* gene has been associated with the disease X-linked dystonia-parkinsonism (XDP) in males of a Philippine population [29]. XDP was associated with a neuron-specific reduction of mRNA of the *TAF1* gene which may interfere with the transcription of neuronal genes such as the dopamine receptor D2. The authors suggested this tissue specific reduction in mRNA of the *TAF1* gene was linked to the hypermethylation of the SVA insertion. This demonstrates the potential of SVAs to influence gene expression through their methylation state within the brain and in a tissue specific manner.

The nature of the sequence contained within SVAs also shows the potential for formation of secondary structures such as cruciforms and G-quadruplexes [21]. Cruciform formation requires perfect or imperfect inverted repeats of 6 or more bases, like those seen in the central VNTR of the SVAs, and are involved in processes such as DNA replication and gene regulation [68]. G4 DNA is a secondary structure formed in guanine-rich sequences and is abundant in promoter regions [69, 70]. G4 structures are hypothesised to interfere with replication of DNA and a host of regulatory functions including gene expression, genome stability and telomerase activity [71-74]. Sequences with potential to form G4 are located in the promoters of several genes such as the *KRAS*, *HRAS* and *c-MYC* genes, and their ability to decrease transcription has been demonstrated [75-77]. The nature of the sequence of SVAs provides the potential for the formation of G4 DNA which could be involved in the regulation of nearby genes in a similar process. The potential of SVAs to form G4 DNA is located within the 5' hexamer repeat and the central VNTR. The amount of G4 potential within each SVA subtype increased as the age of the subtype decreased [48]. The human specific SVAs show the greatest potential for G4 formation and therefore the regulatory properties of this type of secondary structure could play a more predominant role in these elements over the subtypes that are found within other primates.

The modulation of retrotransposons by the environment

The histone marks across retrotransposons and their rate of retrotransposition has been shown to be modulated by environmental factors such as exercise, stress and cocaine in animal models [78-80]. In a mouse model, a labelled L1 reporter element showed a higher number of insertion events in the brain of animals that were able to undertake voluntary exercise over the sedentary animals indicating L1 retrotransposition was affected by exercise [80]. The repeated exposure of mice to cocaine resulted in a decrease in the heterochromatic histone H3K9me3 mark across L1 elements in the nucleus accumbens, an important area of the brain for reward, and an increase in the expression of L1 elements in the same region which may contribute to genomic instability [79]. Another study demonstrated that in response to acute stress in the rat there was an enrichment of the repressive histone modification, H3K9me3, over transposable elements [78]. Although the studies discussed were carried out in animal models, they illustrate how retrotransposons in the mammalian brain can be modulated by the environment, suggesting mechanisms by which the activity of retrotransposons of the human genome may also be affected. There is evidence for global changes across retrotransposons in humans, for example a loss of epigenetic silencing of these elements in the tumour and a change to a relatively more open chromatin state in replicatively senescent cells *in vitro*, which were more prominently associated with the evolutionary young elements [81, 82]. The SVAs may be subject to modulation by their environment which may result in an increase in their rate of retrotransposition or activating regulatory domains within their structure that were previously silenced, impacting on genomic function.

Somatic retrotransposition modifying the genetics of the human brain

Retrotransposons have demonstrated the ability to generate somatic mosaicism, resulting in differences at the genetic level between tissues or cells of the same individual. L1 retrotransposition has been shown to be possible in non-dividing human somatic cells, neural progenitor cells and neuronal cells *in vitro* and *in vivo* [83-85]. It was first demonstrated that an engineered human L1 element within an indicator cassette retrotransposed within adult rat neuronal progenitor cells (NPC) *in vitro* altering the expression of neuronal genes and within the brains of transgenic mice *in vivo* [83]. Further studies showed that L1 retrotransposition can occur within human NPCs and in the adult human brain [85, 86]. A high-throughput analysis of somatic retrotransposition identified 7743 L1, 13692 Alu and 1350 SVA putative somatic insertions across the brains of three individuals, with these insertions occurring at a higher frequency in protein coding genes expressed in the brain, providing evidence for somatic retrotransposition of all active non-LTR elements in the human brain [87]. It would be likely that the SVAs, like the L1 elements discussed above could be regulated by the environment and in part modify genomic function. These new retrotransposon insertions then have the ability to impact on gene regulation within that particular cell, potentially affecting its levels of expression or mRNA splicing. The generation of genetic diversity and the potential addition of new regulatory domains across the cells of an individual's brain could alter the transcriptome of those cells and ultimately impact on the phenotype produced.

CONCLUSIONS

The insertion of SVAs, the youngest of the retrotransposons, into genes or regulatory domains of the human genome could impact on gene regulation and expression through alternative splicing, methylation or the binding of transcription factors. SVAs have contributed to the genetic differences of humans from other primates and among human individuals due to their polymorphic nature in terms of their absence or presence, and the differences in the copy number of their VNTR domains. These genetic differences could contribute to phenotypic and behavioural differences within humans due to differential gene regulation and response to the environment. Greater focus will be required to expand on the functional consequences of the insertion of the SVA elements, both in the germ line and in somatic cells, to enable the extent of their impact to be understood. However the relatively small number of human specific SVAs in the genome

allows us to more practicably determine how these domains not only alter our evolution but also if they in part generate human traits.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

REFERENCES

1. Muotri, A. R., Marchetto, M. C., Coufal, N. G. and Gage, F. H. 2007, Hum. Mol. Genet., 16 Spec. No. 2, R159-67.
2. Goodier, J. L. and Kazazian, H. H. Jr. 2008, Cell, 135, 23-35.
3. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi S. and Chen, Y. J. 2001, Nature, 409, 860-921.
4. Faulkner, G. J. 2011, FEBS Lett., 585, 1589-94.
5. Cordaux, R. and Batzer, M. A. 2009, Nat. Rev. Genet., 10, 691-703.

6. Scott, A. F., Schmeckpeper, B. J., Abdelrazik, M., Comey, C. T., O'Hara, B., Rossiter, J. P., Cooley, T., Heath, P., Smith, K. D. and Margolet, L. 1987, Genomics, 1, 113-25.
7. Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D. and Kazazian, H. H. Jr. 1996, Cell, 87, 917-27.
8. Hohjoh, H. and Singer, M. F. 1997, EMBO J., 16, 6034-43.
9. Mathias, S. L., Scott, A. F., Kazazian, H. H. Jr., Boeke, J. D. and Gabriel, A. 1991, Science, 254, 1808-10.
10. Feng, Q., Moran, J. V., Kazazian, H. H. Jr. and Boeke, J. D. 1996, Cell, 87, 905-16.
11. Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., Boeke, J. D. and Moran, J. V. 2001, Mol. Cell. Biol., 21, 1429-39.
12. Babushok, D. V. and Kazazian, H. H. Jr. 2007, Hum. Mutat., 28, 527-39.
13. Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V. and Kazazian, H. H. Jr. 2003, Proc. Natl. Acad. Sci. USA, 100, 5280-5.
14. Batzer, M. A. and Deininger, P. L. 2002, Nat. Rev. Genet., 3, 370-9.
15. Ullu, E. and Tschudi, C. 1984, Nature, 312, 171-2.
16. Ono, M., Kawakami, M. and Takezawa, T. 1987, Nucleic Acids Res., 15, 8725-37.
17. Zhu, Z. B., Hsieh, S. L., Bentley, D. R., Campbell, R. D. and Volanakis, J. E. 1992, J. Exp. Med., 175, 1783-7.
18. Shen, L., Wu, L. C., Sanlioglu, S., Chen, R., Mendoza, A. R., Dangel, A. W., Carroll, M. C., Zipf, W. B. and Yu, C. Y. 1994, J. Biol. Chem., 269, 8466-76.
19. Wang, H., Xing, J., Grover, D., Hedges, D. J., Han, K., Walker, J. A., and Batzer, M. A. 2005, J. Mol. Biol., 354, 994-1007.
20. Han, K., Konkel, M. K., Xing, J., Wang, H., Lee, J., Meyer, T. J., Huang, C. T., Sandifer, E., Hebert, K., Barnes, E. W., Hubley, R., Miller, W., Smit, A. F., Ullmer, B. and Batzer, M. A. 2007, Science, 316, 238-40.
21. Hancks, D. C. and Kazazian, H. H. Jr. 2010, Semin. Cancer Biol., 20, 234-45.
22. Bantysh, O. B. and Buzdin, A. A. 2009, Biochemistry (Mosc.), 74, 1393-9.
23. Hancks, D. C., Ewing, A. D., Chen, J. E., Tokunaga, K. and Kazazian, H. H. Jr. 2009, Genome Res., 19, 1983-91.
24. Damert, A., Raiz, J., Horn, A. V., Lower, J., Wang, H., Xing, J., Batzer, M. A., Lower, R. and Schumann, G. G. 2009, Genome Res., 19, 1992-2008.
25. Zabolotneva, A. A., Bantysh, O., Suntsova, M. V., Efimova, N., Malakhova, G. V., Schumann, G. G., Gayfullin, N. M. and Buzdin, A. A. 2012, Gene, 505, 128-36.
26. Mills, R. E., Bennett, E. A., Iskow, R. C., Luttig, C. T., Tsui, C., Pittard, W. S. and Devine, S. E. 2006, Am. J. Hum. Genet., 78, 671-9.
27. Lee, J., Ha, J., Son, S. Y. and Han, K. 2012, Comp. Funct. Genomics, 2012, 807270.
28. Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Iida, E., Nomura, Y., Segawa, M., Yoshioka, M., Saito, K., Osawa, M., Hamano, K., Sakakihara, Y., Nonaka, I., Nakagome, Y., Kanazawa, I., Nakamura, Y., Tokunaga, K. and Toda, T. 1998, Nature, 394, 388-92.
29. Makino, S., Kaji, R., Ando, S., Tomizawa, M., Yasuno, K., Goto, S., Matsumoto, S., Tabuena, M. D., Maranon, E., Dantes, M., Lee, L. V., Ogasawara, K., Tooyama, I., Akatsu, H., Nishimura, M. and Tamiya, G. 2007, Am. J. Hum. Genet., 80, 393-406.
30. Wilund, K. R., Yi, M., Campagna, F., Arca, M., Zuliani, G., Fellin, R., Ho, Y. K., Garcia, J. V., Hobbs, H. H. and Cohen, J. C. 2002, Hum. Mol. Genet., 11, 3019-30.
31. Rohrer, J., Minegishi, Y., Richter, D., Eguiguren, J. and Conley, M. E. 1999, Clin. Immunol., 90, 28-37.
32. Hassoun, H., Coetzer, T. L., Vassiliadis, J. N., Sahr, K. E., Maalouf, G. J., Saad, S. T., Catanzariti, L. and Palek, J. 1994, J. Clin. Invest., 94, 643-8.
33. Akman, H. O., Davidzon, G., Tanji, K., Macdermott, E. J., Larsen, L., Davidson, M. M., Haller, R. G., Szczepaniak, L. S., Lehman, T. J., Hirano, M. and DiMauro, S. 2010, Neuromuscul. Disord., 20, 397-402.
34. van der Klift, H. M., Tops, C. M., Hes, F. J., Devilee, P. and Wijnen, J. T. 2012, Hum. Mutat., 33, 1051-5.

35. Takasu, M., Hayashi, R., Maruya, E., Ota, M., Imura, K., Kougo, K., Kobayashi, C., Saji, H., Ishikawa, Y., Asai, T. and Tokunaga, K. 2007, Tissue Antigens, 70, 144-50.
36. Hancks, D. C. and Kazazian, H. H. Jr. 2012, Curr. Opin. Genet. Dev., 22, 191-203.
37. Watanabe, M., Kobayashi, K., Jin, F., Park, K. S., Yamada, T., Tokunaga, K. and Toda, T. 2005, Am. J. Med. Genet. A, 138, 344-8.
38. Ostertag, E. M., Goodier, J. L., Zhang, Y. and Kazazian, H. H. Jr. 2003, Am. J. Hum. Genet., 73, 1444-51.
39. Taniguchi-Ikeda, M., Kobayashi, K., Kanagawa, M., Yu, C. C., Mori, K., Oda, T., Kuga, A., Kurahashi, H., Akman, H. O., DiMauro, S., Kaji, R., Yokota, T., Takeda, S. and Toda, T. 2011, Nature, 478, 127-31.
40. Conley, M. E., Partain, J. D., Norland, S. M., Shurtleff, S. A. and Kazazian, H. H. Jr. 2005, Hum. Mutat., 25, 324-5.
41. Kaer, K. and Speek, M. 2013, Gene, 518, 231-41.
42. Hancks, D. C., Goodier, J. L., Mandal, P. K., Cheung, L. E. and Kazazian, H. H. Jr. 2011, Hum. Mol. Genet., 20, 3386-400.
43. Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Lower, J., Stratling, W. H., Lower, R. and Schumann, G. G. 2012, Nucleic Acids Res., 40, 1666-83.
44. Xing, J., Zhang, Y., Han, K., Salem, A. H., Sen, S. K., Huff, C. D., Zhou, Q., Kirkness, E. F., Levy, S., Batzer, M. A. and Jorde, L. B. 2009, Genome Res., 19, 1516-26.
45. Dewannieux, M., Esnault, C. and Heidmann, T. 2003, Nat. Genet., 35, 41-8.
46. Hancks, D. C., Mandal, P. K., Cheung, L. E. and Kazazian, H. H. Jr. 2012, Mol. Cell Biol., 32, 4718-26.
47. Xing, J., Wang, H., Belancio, V. P., Cordaux, R., Deininger, P. L. and Batzer, M. A. 2006, Proc. Natl. Acad. Sci. USA, 103, 17608-13.
48. Savage, A. L., Bubb, V. J., Breen, G. and Quinn, J. P. 2013, BMC Evol. Biol., 13, 101.
49. Bennett, E. A., Coleman, L. E., Tsui, C., Pittard, W. S. and Devine, S. E. 2004, Genetics, 168, 933-51.
50. MacKenzie, A. and Quinn, J. 1999, Proc. Natl. Acad. Sci. USA, 96, 15251-5.
51. Vasiliou, S. A., Ali, F. R., Haddley, K., Cardoso, M. C., Bubb, V. J. and Quinn, J. P. 2012, Addict. Biol., 17, 156-70.
52. Brotons, O., O'Daly, O. G., Guindalini, C., Howard, M., Bubb, J., Barker, G., Dalton, J., Quinn, J., Murray, R. M., Breen, G. and Shergill, S. S. 2011, Mol. Psychiatry, 16, 124-6.
53. Haddley, K., Vasiliou, A. S., Ali, F. R., Paredes, U. M., Bubb, V. J. and Quinn, J. P. 2008, Neurochem. Res., 33, 652-67.
54. Haddley, K., Bubb, V. J., Breen, G., Parades-Esquivel, U. M. and Quinn, J. P. 2012, Curr. Top. Behav. Neurosci., 12, 503-535.
55. Corti, O., Lesage, S. and Brice, A. 2011, Physiol. Rev., 91, 1161-218.
56. Rebollo, R., Romanish, M. T. and Mager, D. L. 2012, Annu. Rev. Genet., 46, 21-42.
57. Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., Schroder, K., Cloonan, N., Steptoe, A. L., Lassmann, T., Waki, K., Hornig, N., Arakawa, T., Takahashi, H., Kawai, J., Forrest, A. R., Suzuki, H., Hayashizaki, Y., Hume, D. A., Orlando, V., Grimmond, S. M. and Carninci, P. 2009, Nat. Genet., 41, 563-71.
58. Fairbanks, D. J., Fairbanks, A. D., Ogden, T. H., Parker, G. J. and Maughan, P. J. 2012, G3 (Bethesda), 2, 1447-57.
59. Roberts, J., Scott, A. C., Howard, M. R., Breen, G., Bubb, V. J., Klenova, E. and Quinn, J. P. 2007, J. Neurosci., 27, 2793-801.
60. Anguelova, M., Benkelfat, C. and Turecki, G. 2003, Mol. Psychiatry, 8, 646-53.
61. Herman, A. I., Kaiss, K. M., Ma, R., Philbeck, J. W., Hasan, A., Dasti, H. and DePetrillo, P. B. 2005, Am. J. Med. Genet. B Neuropsychiatr. Genet., 133B, 74-8.
62. Guindalini, C., Howard, M., Haddley, K., Laranjeira, R., Collier, D., Ammar, N., Craig, I., O'Gara, C., Bubb, V. J., Greenwood, T., Kelsoe, J., Asherson, P., Murray, R. M., Castelo, A., Quinn, J. P., Vallada, H. and Breen, G. 2006, Proc. Natl. Acad. Sci. USA, 103, 4552-7.
63. Munafo, M. R. and Johnstone, E. C. 2008, Addict. Biol., 13, 435-9.
64. Larsen, F., Gundersen, G., Lopez, R. and Prydz, H. 1992, Genomics, 13, 1095-107.

65. Gardiner-Garden, M. and Frommer, M. 1987, J. Mol. Biol., 196, 261-82.
66. Reik, W. 2007, Nature, 447, 425-32.
67. Bird, A., 2002, Genes Dev., 16, 6-21.
68. Brazda, V., Laister, R. C., Jagelska, E. B. and Arrowsmith, C. 2011, BMC Mol. Biol., 12, 33.
69. Huppert, J. L. and Balasubramanian, S. 2007, Nucleic Acids Res., 35, 406-13.
70. Zhao, Y., Du, Z. and Li, N. 2007, FEBS Lett., 581, 1951-6.
71. Clark, D. W., Phang, T., Edwards, M. G., Geraci, M. W. and Gillespie, M. N. 2012, Free Radic. Biol. Med., 53, 51-9.
72. De, S. and Michor, F. 2011, Nat. Struct. Mol. Biol., 18, 950-5.
73. Huppert, J. L. and Balasubramanian, S. 2005, Nucleic Acids Res., 33, 2908-16.
74. Fletcher, T. M., Sun, D., Salazar, M. and Hurley, L. H. 1998, Biochemistry, 37, 5536-41.
75. Cogoi, S. and Xodo, L. E. 2006, Nucleic Acids Res., 34, 2536-49.
76. Membrino, A., Cogoi, S., Pedersen, E. B. and Xodo, L. E. 2011, PLoS One, 6, e24421.
77. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. and Hurley, L. H. 2002, Proc. Natl. Acad. Sci. USA, 99, 11593-8.
78. Hunter, R. G., Murakami, G., Dewell, S., Seligsohn, M., Baker, M. E., Datson, N. A., McEwen, B. S. and Pfaff, D. W. 2012, Proc. Natl. Acad. Sci. USA, 109, 17657-62.
79. Maze, I., Feng, J., Wilkinson, M. B., Sun, H., Shen, L. and Nestler, E. J. 2011, Proc. Natl. Acad. Sci. USA, 108, 3035-40.
80. Muotri, A. R., Zhao, C., Marchetto, M. C. and Gage, F. H. 2009, Hippocampus, 19, 1002-7.
81. Szpakowski, S., Sun, X., Lage, J. M., Dyer, A., Rubinstein, J., Kowalski, D., Sasaki, C., Costa, J. and Lizardi, P. M. 2009, Gene, 448, 151-67.
82. De Cecco, M., Criscione, S. W., Peckham, E. J., Hillenmeyer, S., Hamm, E. A., Manivannan, J., Peterson, A. L., Kreiling, J. A., Neretti, N. and Sedivy, J. M. 2013, Aging Cell, 12, 247-56.
83. Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V. and Gage, F. H. 2005, Nature, 435, 903-10.
84. Kubo, S., Seleme, M. C., Soifer, H. S., Perez, J. L., Moran, J. V., Kazazian, H. H. Jr. and Kasahara, N. 2006, Proc. Natl. Acad. Sci. USA, 103, 8036-41.
85. Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., Morell, M., O'Shea, K. S., Moran, J. V. and Gage, F. H. 2009, Nature, 460, 1127-31.
86. Evrony, G. D., Cai, X., Lee, E., Hills, L. B., Elhosary, P. C., Lehmann, H. S., Parker, J. J., Atabay, K. D., Gilmore, E. C., Poduri, A., Park, P. J. and Walsh, C. A. 2012, Cell, 151, 483-96.
87. Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., Brennan, P. M., Rizzu, P., Smith, S., Fell, M., Talbot, R. T., Gustincich, S., Freeman, T. C., Mattick, J. S., Hume, D. A., Heutink, P., Carninci, P., Jeddeloh, J. A. and Faulkner, G. J. 2011, Nature, 479, 534-7.