Title page

SVA retrotransposons as modulators of gene expression

John P. Quinn* and Vivien J. Bubb

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

*To whom correspondence should be addressed: Department of Molecular and Clinical Pharmacology, The University of Liverpool, Crown Street, Liverpool, UK, L69 3BX. Tel: +44 151 794 5498. Fax: +44 151 794 5517. E-mail: jquinn@liverpool.ac.uk.

Abstract: 216 words

Text body: 2006 without references

Commentary to

Savage AL, Wilm TP, Khursheed K, Shatunov, A, Morrison, K. E, Shaw, P. J, Shaw, C. E, Smith, B, Breen, G, Al-Chalabi, A, Moss, D, Bubb, V. J, Quinn, J. P. An evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS. *PLoS ONE* 2014;9(3):e9083

Abstract

Endogenous mobile genetic elements can give rise to *de novo* germline or somatic mutations that can have dramatic consequences for genome regulation both local and possibly more globally based on the site of integration. However if we consider them as 'normal genetic' components of the reference genome then they are likely to modify local chromatin structure which would have an effect on gene regulation irrelevant of their ability to further transpose. As such they can be treated as any other domain involved in a gene x environment interaction. Similarly their evolutionary appearance in the reference genome would supply a driver for species specific responses/traits. Our recent data would suggest the hominid specific subset of retrotransposons, SINE-VNTR-Alu (SVA), can function as transcriptional regulatory domains both *in vivo* and *in vitro* when analysed in reporter gene constructs. Of particular interest in the SVA element, were the variable number tandem repeat (VNTR) domains which as their name suggests can be polymorphic. We and others have previously shown that VNTRs can be both differential regulators and biomarkers of disease based on the genotype of the repeat. Here, we provide an overview of why polymorphism in the SVA elements, in particular the VNTRs, could alter gene expression patterns that could be mechanistically associated with different traits in evolution or disease progression in humans.

Repetitive DNA in the Genome

In this article we will use the term variable number tandem repeat (VNTR) to encompass both microsatellites and minisatellites however not all tandem repeat domains are polymorphic in repeat copy number. Nevertheless there are estimated to be approximately 1 million VNTRs in the human genome [1]. Their function has been demonstrated to include modulation of alternative splicing, thus altering protein isoform levels and effects in gene regulation. It's the latter of these functions that our group has the most experience analysing. We have demonstrated in several genes involved in monoaminergic transmission in the nervous system that, based on copy number, VNTRs can be both biomarkers of predisposition to neuropsychiatric disorders direct differential tissue specific and stimulus inducible expression both *in vitro* and *in vivo* [2-12]. We began to address potential VNTRs in the limited number of candidate genes associated with Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Lobar Degeneration (FTLD). We identified a large tandem repeat domain approximately 10kb 5' of the major transcriptional start site of the FUS (Fused in sarcoma) gene using the UCSC genome bioinformatic site (http://genome.ucsc.edu). Indeed, closer inspection of the primary sequence showed the potential for two adjacent VNTRs which varied between 37-50bp in the size of their repeat sequence. When BLAT was performed using this repeat sequence we identified a host of related sequences all sharing significant homology. This tandem repeat represented the central domain of the hominid retrotransposon family termed SINE-VNTR-Alu (SVA).

SVA retrotransposons

SVA elements only represent 0.13% of the genome, representing ~2700 elements, constituting the youngest of the retrotransposable elements in the human genome and are hominid specific. They consist of a hexamer repeat (CCCTCT), an Alu-like sequence, a GC-rich VNTR, a SINE and a poly A-tail [13]. Many SVA elements contain two central GC-rich VNTRs and it should be noted that the flanking hexamer is also a VNTR[14]. Generally SVA elements vary in length from approximately 1000-4000bp with 63% of SVA element insertions in the human genome being full length, containing all five domains[13, 15]. SVA elements are divided into subtypes (A-F) by the SINE region and more recently a seventh subtype was identified containing a 5' transduction of the sequence from MAST2 gene referred to as either CpG-SVA, MAST2 SVA or SVA F1 element [16-18]. Similarly to other actively retrotransposing elements SVA elements show inter-individual variation in man and can be polymorphic for their absence or presence in the genome. It has been estimated that 37.5 % of SVA E elements and 27.6% of SVA F elements are polymorphic for their presence in the genome [13] and the average human is estimated to have 56 SVA absence/presence polymorphisms [19].

As with other retrotransposons much effort has been expended on the retrotransposition event itself giving rise to disease. This has resulted in the identification of at least 8 well characterised diseases in

which the retrotransposition event has caused the disease, in the majority of cases by affecting alternative slicing[15 20]. However our hypothesis is that SVA elements would have an effect on epigenetic and transcriptional parameters at the locus in which they are found in the human reference genome without the need for retrotransposition. These parameters are embedded within the primary sequence of the SVA element and in particular the tandem repeat domains.

SVA elements as a model for Gene x Environment (GxE) interaction in disease progression and evolution

Our hypothesis is that the SVA element could impart a hominid specific regulatory twist on regulation of the gene in which they have inserted (assuming that insertion is not in an exon or destroys intron/exon boundaries etc). In the promoter they would act as classical mediators of gene expression. Analysis of Human Genome release 19 (Hg19) indicates that of the 2676 SVA elements in the human genome, 433 are present within 10kb 5' of the main transcriptional start site [15]. We have published on two such promoter domains approximately 8 and 10kb respectively from the major transcriptional start site of the nearest gene, namely the PARK 7 and FUS genes. The following points all support a regulatory role for SVA elements in the genome:

1. SVA elements are functional in conventional reporter gene assay both *in vitro* (cell line) and *in vivo* (chick embryo)[14, 21 22].
2. SVA elements are polymorphic, most clearly via VNTR domains (flanking and internal)[14, 15, 21]. Although SNPs could also exist.
    a. We have previously shown in other genes (as have many others) that VNTRs are both polymorphic biomarkers associated with disorders and transcriptional regulatory domains *in vivo* and *in vitro*[2-12].
    b. VNTRs are common regulatory domains in viruses;
        i. They in part control HSV-1 virus latency in dorsal root ganglia (via a similar CCCCTC repeat to that found in the flanking sequence of the SVA element)[23-25].
        ii. VNTRs are present in the enhancers of many viruses including retroviruses [26, 27].
3. SVA elements can be associated with active chromatin adjacent to:
    a. Positive affymetrix probe arrays in the human CNS
    b. ENCODE active histone marks
4. SVA elements have the potential to form structures which may be strong modulators of genome structure/function:
    a. Potential G4 quadruplex[14]
    b. Strong CpG component approaching the classical requirements to be defined as CpG islands, which is consistent with their differential methylation in tumours [28, 29]

How do hominid specific SVA regulatory domains utilise the existing transcription machinery in the cell

In our analysis of the FUS SVA element we generated reporter gene constructs that were analysed in both the neuroblastoma cell line, SK-N-AS and a chicken embryo model[21]. In the former, the data was consistent with our previous data on the PARK7 SVA element in which we were able to demonstrate that the SVA element contained multiple regulatory domains[14]. This would also be consistent with previous data demonstrating promoter function with the SVA F1 element [22]. This regulatory analysis was extended with the FUS SVA element to demonstrate that this domain supported expression *in vivo* using the chicken embryo model. This is also an example of hominid specific regulatory domains functioning in other species. Previously, we have demonstrated the human serotonin transporter VNTR, which is not present in rodents, nonetheless supported both tissue specific and differential reporter gene expression in mouse embryos which was directed by genotype of the VNTR [12]. In this model the expression in the embryo was in the cells which at that point in development in the rodent first begin to develop a serotoninergic lineage with serotonin transporter (SLC6A4) expression. We hypothesised that the observed reporter expression would result from the novel regulatory VNTR which had evolved taking advantage of the existing transcriptional machinery in particular subsets of cells that were important for serotonin transporter expression and that the VNTR had been maintained in the genome as it contributed a favourable selection in evolution. We propose that a similar mechanism would operate at the VNTRs in the SVA element and in part this would lead to novel regulatory responses via the SVA element at the adjacent endogenous gene. It should also be noted that many VNTRs such as the SLC6A4 VNTR in the mouse transgenic model above are intronic and therefore the location of the SVA element to impact such regulatory properties can be quite variable at the gene loci.

The distance of the SVA element from its site of action on a specific promoter or transcriptional start site can be quite large. It has been demonstrated by us and others when attempting to reproduce endogenous gene expression in transgenic models, that regulatory domains utilised by a locus can be 100kb+ away from the gene itself. Our own studies on the endogenous expression of the neuropeptide, substance P encoded by the TAC1 gene demonstrated that although the gene itself from 5' to 3' UTR was 8kb, we required a DNA fragment of 350kb to reproduce the appropriate expression patterns[30, 31]. Subsequently our collaborators demonstrated key regulatory domains were as far as 250kb and 100kb 5'of the TAC1 gene[32]. The 350kb TAC1 fragment as a transgenic insertion demonstrated not only the well characterised rodent expression of TAC1, but as was the case with the SLC6A4 VNTR, it also demonstrated human specific TAC1 expression patterns and we again argued that the DNA was taking advantage of the complement of transcription factor that was present in the cells when the DNA in the regulatory regions evolved to support distinct transcription profiles for this gene. By the same mechanism we would argue

that SVA elements even 100kb away from the transcription start site could modulate transcriptional regulation distinct from an allele without a SVA element.

The recent evolution of SVA elements has imposed restricted sequence variation in each of the SVA subclasses; this may allow a concerted response to challenge by groups of SVA elements with similar primary sequence based on sequence specific DNA binding proteins. In this manner it might be possible for SVA elements containing similar primary sequence such as is present in the hexamer repeat or central VNTRs to bind similar transcription factors based on shared consensus DNA binding domains. Thus alterations in active transcription factor complement in the cell might modulate several SVA elements by the same signal transduction pathways.

Can SVA elements be responsible for emerging diseases?

The frequency of SVA element retrotransposition is estimated to be 1:916 births which would correspond to $7x10^6$ distinct insertions worldwide in the population [33]. These could result in modulation of transcriptional properties without modifying protein structure or splicing parameters. Could these constitute a proportion of the rare event genetic components that underlie disease processes? We can argue that SVA elements should be assigned a greater role in disease without the retrotransposition event and that due in part to the nature of the VNTR elements; they could be biomarkers of predisposition to a specific disorder. In our analysis of SVA elements located close to FUS and PARK7 genes, we have addressed the polymorphic variation in the VNTRs to establish the range of VNTR repeat number in these domains. We found in both of these genes that only one of the central tandem repeat was polymorphic whilst the other was a fixed repeat length. We also confirmed in PARK7 that the hexamer was a repeat and 3 length variants were observed in the HapMap CEU cohort. For the FUS analysis we performed polymorphic analysis in sporadic ALS versus controls (241 and 228 respectively). The FUS SVA element does not contain a flanking hexamer so we analysed only the central repeat domain and found 2 variants. The data implied a difference in the homozygous short vs heterozygous long/short SVA element allele carriers in the two cohorts for association with sporadic ALS. This might suggest that in a larger cohort we might reach significance for these genotypes. A technical problem of such analysis is that PCR over large repeat units is not simplistic and each PCR requires significant optimisation, therefore the requirement to do this PCR for large cohorts might prove a hindrance to such studies. However, it should be possible to find tagging SNPs for SVA elements which demonstrate limited polymorphism as in the FUS SVA element. We have done this by converting the long and short SVA element genotype of the individuals from the CEU cohort to 'SNPs' so they could be uploaded into the Haploview software to allow for identification of tagging SNPs that can be used to address correlation with disease in larger cohorts in the future. This approach will allow us in the future to generate tagging SNPs for SVA elements for addressing association to disease. This is of course dependent on a restricted number of SVA 'VNTR' alleles in a specifically targeted SVA element. In the

limited number of SVA elements we have addressed, including PARK7 and FUS, we have evidence for two major VNTR alleles in the majority with none exhibiting greater variation than PARK7; however the analysis is far from complete. Nevertheless such a strategy may allow for integration of SVA element genotype as a correlate of human disease more rapidly than PCR analysis of 1000s of individual SVA elements.

Concluding Remarks

Our recent work suggests that SVA elements can in part modulate genome function via their action as a transcriptional or epigenetic modulator. Polymorphism in the VNTRs within these retrotransposons demonstrates a mechanism not only for differential regulatory function associated with genotype but a way to rapidly integrate polymorphism in the SVA element as a potential biomarker for disease association without the requirement for retrotransposition.

References

1. Breen G, Collier D, Craig I, Quinn J. Variable number tandem repeats as agents of functional regulation in the genome. *IEEE Eng Med Biol Mag* Mar-Apr 2008;27(2):103-104, 108.

2. Quinn JP, Warburton A, Myers P, Savage AL, Bubb VJ. Polymorphic variation as a driver of differential neuropeptide gene expression. *Neuropeptides* Dec 2013;47(6):395-400.

3. Pickles A, Hill J, Breen G, Quinn J, Abbott K, Jones H, Sharp H. Evidence for interplay between genes and parenting on infant temperament in the first year of life: monoamine oxidase A polymorphism moderates effects of maternal sensitivity on infant anger proneness. *J Child Psychol Psychiatry* Dec 2013;54(12):1308-1317.

4. Hill J, Breen G, Quinn J, Tibu F, Sharp H, Pickles A. Evidence for interplay between genes and maternal stress in utero: monoamine oxidase A polymorphism moderates effects of life events during pregnancy on infant negative emotionality at 5 weeks. *Genes, brain, and behavior* Jun 2013;12(4):388-396.

5. Vasiliou SA, Ali FR, Haddley K, Cardoso MC, Bubb VJ, Quinn JP. The SLC6A4 VNTR genotype determines transcription factor binding and epigenetic variation of this gene in response to cocaine in vitro. *Addict Biol* Jan 2012;17(1):156-170.

6. Brotons O, O'Daly OG, Guindalini C, et al. Modulation of orbitofrontal response to amphetamine by a functional variant of DAT1 and in vitro confirmation. *Molecular psychiatry* Feb 2011;16(2):124-126.

7. Talkowski ME, McCann KL, Chen M, et al. Fine-mapping reveals novel alternative splicing of the dopamine transporter. *Am J Med Genet B Neuropsychiatr Genet* Dec 5 2010;153B(8):1434-1447.

8. Ali FR, Vasiliou SA, Haddley K, et al. Combinatorial interaction between two human serotonin transporter gene variable number tandem repeats and their regulation by CTCF. *Journal of neurochemistry* Jan 2010;112(1):296-306.

9. Haddley K, Vasiliou AS, Ali FR, Paredes UM, Bubb VJ, Quinn JP. Molecular genetics of monoamine transporters: relevance to brain disorders. *Neurochem Res* Apr 2008;33(4):652-667.

10. Roberts J, Scott AC, Howard MR, Breen G, Bubb VJ, Klenova E, Quinn JP. Differential regulation of the serotonin transporter gene by lithium is mediated by transcription factors, CCCTC binding protein and Y-box binding protein 1, through the polymorphic intron 2 variable number tandem repeat. *J Neurosci* Mar 14 2007;27(11):2793-2801.

11. Klenova E, Scott AC, Roberts J, et al. YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. *J Neurosci* Jun 30 2004;24(26):5966-5973.

12. MacKenzie A, Quinn J. A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer- like properties in the mouse embryo. *Proceedings of the National Academy of Sciences of the United States of America* 1999;96(26):15251-15255.

13. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA elements: a hominid-specific retroposon family. *J Mol Biol* Dec 9 2005;354(4):994-1007.

14. Savage AL, Bubb VJ, Breen G, Quinn JP. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol Biol* 2013;13(1):101.

15. Savage AL, Bubb VJ, Quinn JP. What role do human specific retrotransposons play in mental health and behaviour? *Current Trends in Neurology* 2013;7:57-68.

16. Bantysh OB, Buzdin AA. Novel family of human transposable elements formed due to fusion of the first exon of gene MAST2 with retrotransposon SVA. *Biochemistry (Mosc)* Dec 2009;74(12):1393-1399.

17. Damert A, Raiz J, Horn AV, et al. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome research* Nov 2009;19(11):1992-2008.

18. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH, Jr. Exon-trapping mediated by the human retrotransposon SVA. *Genome Res* Nov 2009;19(11):1983-1991.

19. Bennett EA, Coleman LE, Tsui C, Pittard WS, Devine SE. Natural genetic variation caused by transposable elements in humans. *Genetics* Oct 2004;168(2):933-951.

20. Hancks DC, Kazazian HH, Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* Jun 2012;22(3):191-203.

21. Savage AL, Wilm TP, Khursheed K, et al. An evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS. *PLoS ONE* 2014;9(3):e90833.

22. Zabolotneva AA, Bantysh O, Suntsova MV, Efimova N, Malakhova GV, Schumann GG, Gayfullin NM, Buzdin AA. Transcriptional regulation of human-specific SVAF(1) retrotransposons by cis-regulatory MAST2 sequences. *Gene* Aug 15 2012;505(1):128-136.

23. Stevens HC, Fiskerstrand C, Bubb VJ, Dalziel R, Quinn JP. A regulatory domain spanning the repeat sequence RE1 from herpes simplex virus type 1 has cell specific differential functions in trigeminal neurons and fibroblasts. *FEBS letters* Oct 20 2009;583(20):3335-3338.

24. Quinn JP, Dalziel RG, Nash AA. Herpes virus latency in sensory ganglia--a comparison with endogenous neuronal gene expression. *Prog Neurobiol* 2000;60(2):167-179.

25. Quinn JP, McGregor RA, Fiskerstrand CE, Davey C, Allan J, Dalziel RG. Identification of a novel multifunctional structural domain in the herpes simplex virus type 1 genome: implications for virus latency. *J Gen Virol* 1998;79(Pt 10):2529-2532.

26. Quinn JP, Farina AR, Gardiner K, Krutzsch H, Levens D. Multiple components are required for sequence recognition of the AP1 site in the Gibbon Ape leukemia virus enhancer. *Molecular and cellular biology* 1989;9:4713-4721.

27. Quinn JP, Holbrook N, Levens D. Binding of a cellular protein to the Gibbon ape leukemia virus enhancer. *Molecular and cellular biology* 1987;7:2735-2744.

28. De Cecco M, Criscione SW, Peckham EJ, et al. Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* Apr 2013;12(2):247-256.

29. Konkel MK, Batzer MA. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Seminars in cancer biology* Aug 2010;20(4):211-221.

30. MacKenzie A, Payne C, Boyle S, Clarke AR, Quinn JP. The human preprotachykinin-A gene promoter has been highly conserved and can drive human-like marker gene expression in the adult mouse CNS. *Mol Cell Neurosci* Nov 2000;16(5):620-630.

31. MacKenzie A, Quinn J. A yeast artificial chromosome containing the human preprotachykinin-A gene expresses substance P in mice and drives appropriate marker-gene expression during early brain embryogenesis. *Mol Cell Neurosci* Jan 2002;19(1):72-87.

32. Shanley L, Davidson S, Lear M, Thotakura AK, McEwan IJ, Ross RA, MacKenzie A. Long-range regulatory synergy is required to allow control of the TAC1 locus by MEK/ERK signalling in sensory neurones. *Neurosignals* 2010;18(3):173-185.

33. Xing J, Zhang Y, Han K, et al. Mobile elements create structural variation: analysis of a complete human genome. *Genome research* Sep 2009;19(9):1516-1526.