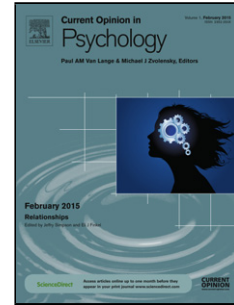# Accepted Manuscript

Title: Non-coding genetic variation shaping mental health

Authors: John P Quinn, Abigail Savage, Vivien J Bubb

To appear in:

Non-coding genetic variation shaping mental health

John P Quinn*, Abigail Savage, Vivien J Bubb,

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

John Quinn: Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, The University of Liverpool, Liverpool L69 3BX, UK, jquinn@liverpool.ac.uk; Tel: (+44) 151 794 5498.

Highlights (maximum 85 characters, including spaces, per bullet point)

- Gene expression determined by the genome mediating a response to cell environment
- Genetic variation results in distinct individual response in gene expression
- Non-coding DNA is an important site for such functional genetic variation
- Gene expression is a major modulator of brain chemistry and thus behaviour

Abstract 173 words

Main Text 2288 words

References 65 with 9 highlighted

Also includes

1. **4 highlights**2 Boxes
    2. 2 Figures

Abstract

Over 98% of our genome is non-coding and is now recognised to have a major role in orchestrating the tissue specific and stimulus inducible gene expression pattern which underpins our wellbeing and mental health. The non-coding genome responds functionally to our environment at all levels, encompassing the span from psychological to physiological challenge. The gene expression pattern, termed the transcriptome, ultimately gives us our neurochemistry. Therefore a major modulator of mental wellbeing is how our genes are regulated in response to life experiences. Superimposed on the aforementioned non-coding

DNA framework is a vast body of genetic variation in the elements that control response to challenges. These differences, termed polymorphisms, allow for a differential response from a specific DNA element to the same challenge thus potentially allowing 'individuality' in the modulation of our transcriptome. This review will focus on a fundamental mechanism defining our psychological and psychiatric wellbeing, namely how genetic variation can be correlated with differential gene expression in response to specific challenges, thus resulting in altered neurochemistry which consequently may shape behaviour.

Introduction

The human genome has evolved to include a combination of both highly conserved regions of regulatory non-coding DNA (ncDNA) found across many species and human-specific regulatory DNA elements which together act to regulate expression of mRNA. This combination of DNA elements allows determination of where, when, how much and for how long, genes are expressed in the human brain in response to normal developmental, psychological and physiological cues, Figure 1. Many of these elements exhibit genetic variation which is not only associated with risk for a specific condition, but has also been demonstrated to alter the regulatory properties of the gene. The functional interpretation and analysis of ncDNA variation can be initially addressed *in silico* by overlaying its position on databases containing characterised and predicted functional elements within the genome, Box 1. The most easily accessible free database is the Encyclopaedia of DNA Elements (ENCODE; https://www.encodeproject.org/) which is a collaboration of research groups funded by the National Human Genome Research Institute [1,2], this can be used in combination with a plethora of other database browsers [3] such as the University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/)[4]. This review will begin with an introduction to the most conserved regulatory regions in the genome and how these may be functionally modified by the simplest and most extensively studied class of genetic variation, single nucleotide polymorphisms (SNPs). The review will then focus on human regulatory elements that are associated with neuropsychiatric conditions which are larger blocks of DNA variation such as variable number tandem repeats (VNTRs) and non-long terminal repeat (non-LTR) retrotransposons, Box 2.

Evolutionary Conserved Regions.

Evolutionary conserved regions (ECRs) in the genome can be easily found using the ECR browser (https://ecrbrowser.dcode.org/) [5]. ECRs in this browser are typically defined as regions of sequence within the human genome that retain 70% or more sequence identity over a window of 100 bases when compared to the corresponding region of sequence in other species, this will frequently include exons in coding DNA. However, Pennacchio et al., were amongst the first to demonstrate that ECRs in the non-coding DNA (ncECRs) could be important, particularly in directing gene expression in the CNS. They determined by use of a transgenic mouse model that whilst ncECRs could direct expression in a broad range of

anatomical structures in the embryo, the majority of the ncECRs tested directed expression to various regions of the developing nervous system [6]. Subsequently, consistent with this, a third of paralogous ncECRs examined were predicted to have regulatory activity in the brain [7], for example, deletion of ncECRs in the neuronal transcription factor *Arx* resulted in substantial alterations of neuron populations and structural brain defects in a trangenic model [8]. Furthermore, the combinatorial complexity of gene expression was exquistely demonstrated in a transgenic model of craniofacial morphology in which the action of multiple ncECRs driving expression of many genes resulted in a vast array of facial differences [9]**. These studies demonstrated that ncECRs can have important transcripitonal regulatory properties, therefore the expectation is that polymorphism in such domains has the potential to modify interactions with transcription factors and thus affect regulatory function.

Single nucleotide polymorphisms

Early studies of genetic variation correlated with mental health focused on DNA variation in exons encoding proteins. Most of these studies addressed SNPs; thus a SNP that changed an amino acid (non-synonymous change) or resulted in a truncation of the protein could be mechanistically relevant as it could alter protein function. However, with technological advances the ability to address SNPs in genome wide association studies (GWAS) rather than solely exons, demonstrated that the vast bulk of SNP variation associated with behavioural and psychiatric conditions was in ncDNA [10,11]. GWAS has led to significant discoveries in defining some of the genes involved in neuropsychiatric disorders and demonstrated there is genetic overlap between many of the major psychiatric disorders [12,13]. In several examples the proteins identified can work together to alter a key pathway underpinning wellbeing and mental health, such as those modifying calcium signalling [14].

Understanding the mechanistic significance of SNPs in ncDNA for a specific condition has been a much more difficult task than for SNPs found in exons. A SNP in ncDNA could be tagging a regulatory domain 10K+ bases from itself (a tagging SNP is representative of a large section of DNA that is inherited as one, thus the SNP is not the causative agent but rather highlights a region of DNA). Analysis of SNP variation within ENCODE and associated data sets can determine if it is present in a genomic region defined as a regulatory domain.

In this scenario, the SNP could affect the efficiency or specificity with which a transcription factor, proteins which modulate the process of transcription, would bind to this regulatory DNA sequence, Figure 2.

We and others demonstrated that SNP polymorphisms in ncECRs which correlated with known behavioural problems could modify the regulatory properties of the ECR including those associated with depression located in BDNF, BICC1 and galanin genes [15-19]. Furthermore multiple ncECRs may be required for appropriate gene expression, for example eight conserved ncECRs were identified at the schizophrenia-associated MIR137/DPYD locus of these, six were shown to be positive transcriptional regulators, and two negative transcriptional regulators in a human cell line model [18,19]. Bioinformatic analysis of this locus using the Psychiatric Genomics Consortium GWAS dataset for schizophrenia highlighted five of the ncECRs had genome-wide significant SNPs in, or adjacent to their sequence [11].

Epigenetic marks which are indicative of active or inactive chromatin, are often found at regulatory DNA. Genetic variation such as GWAS risk SNPs, can effect such epigenetic parameters impacting on long term regulatory changes in response to challenge [20]. Both local (gene specific) and global (multigene) epigenetic changes have been implicated in neuropsychiatric disorders [21,22] and the NIH Roadmap Epigenomics Consortium (http://www.roadmapepigenomics.org/) data can be utilised to analyse such data. For example, local methylation variation at the glucocorticoid receptor gene has been associated with prenatal and postnatal depression [23]* and global differences in methylation in astrocytes have been associated with depression [24]. Simplistically, methylation of regulatory regions is considered a repressor of transcription as it interferes with transcription factor binding by limiting the accessibility of specific DNA recognition sequences. The ability to rapidly address genetic variation on a 'road map' of regulatory domains has allowed the development of a significantly better understanding of how the ncDNA GWAS SNPs can be mechanistically involved in mental health issues [25]. This can be further updated within the UCSC browser which permits new, novel data to be overlaid on the existing data from ENCODE.

Variable Number Tandem Repeats

SNP variation is not the only example of ncDNA variation that can affect the regulation of gene expression. Many of the best characterised genetic polymorphisms correlating with mental health issues are found in repetitive DNA, Figure 2. These include the VNTRs [26], examples of which have been identified in key behavioural and mental health-related genes. VNTRs have been demonstrated to be both biomarkers and transcriptional regulators in genes such as the serotonin transporter, the dopamine transporter and monoamine oxidase A [21,22,27-32]. In these three examples, the primary DNA sequences of the VNTRs are rapidly evolving such that humans have their own specific VNTR sequences. All three of these monoaminergic genes contain a minimum of two VNTRs that have been demonstrated to act both independently and synergistically as transcriptional regulators whose function is further modulated by the repeat copy number within the VNTR [21,29,31]. The copy number of the repeat itself is also a biomarker for good mental health and wellbeing thus correlating function with phenotype [27,28,31,33,34]. Perhaps not unexpectedly VNTRs and GWAS SNPs in the same promoter may act additively or synergistically to regulate gene expression. This is exemplified by one of the promoters of the schizophrenia candidate risk gene, MIR137 [35,36]*, where experimentally *in vitro,* the VNTR in the promoter can support differential reporter gene expression based on the copy number of the repeat within the VNTR, and inclusion of the promoter region encompassing the GWAS SNP can further modulate expression depending on the allele of the SNP present. This illustrates a route to identifying the potential functional significance of non-coding variants in transcriptional or post transcriptional regulatory mechanisms in areas distinct from the region of the DNA in which the GWAS SNP is found.

The rapid evolution of VNTRs has been noted more globally for contributing to primate evolution; analysis in humans and non-human great apes identified that genes with VNTRs have higher expression divergence than those without [37]. The association of VNTRs with gene expression is reflected in the finding that VNTRs are enriched in promoter regions and locations close to transcriptional start sites for mRNA expression [26,38,39]. Generally, VNTRs have not been analysed as extensively as SNPs which may be attributed to the requirement to perform PCR to genotype each VNTR target and the inability to accurately identify such regions in the initial short read whole genome sequencing protocols. Improved depth and coverage in whole genome sequence combined with the

development of bioinformatic programmes such as ExpansionHunter may improve the association of VNTRs and other repeat variants with neuropsychiatric conditions [40].

Non-LTR Retrotransposons

Non-long terminal repeat (non-LTR) retrotransposons are mobile DNA elements that can copy and paste themselves into new genomic loci and are therefore polymorphic for their presence or absence at specific loci in the genome. These retrotransposable elements (RTEs), also known as 'jumping genes', can range in size from a few hundred to 6000 base pairs and have been shown to be major modulators of gene expression at several levels. Non-LTR retrotransposons comprise three classes, long interspersed nuclear element 1 (LINE-1), *Alu* and 'SINE-VNTR-Alu' (SVA), Figure 2. LINE-1 expression has been implicated in many neuropsychiatric conditions such as depression [41], addiction [42], schizophrenia [43-45] and autism [46]. The other two classes have also been implicated in CNS function, for example variation in the *Alu* sequence within an intron of the TOMM40 gene is associated with non-pathogenic cognitive decline [47] and X-Linked Dystonia-Parkinsonism is associated with the presence or absence of a SVA in the TAF1 gene [48]**. SVAs contain several distinct VNTR elements and have properties consistent with transcriptional regulation [49-51]*.

There has been tremendous interest in RTEs, due to their ability to make a copy of themselves which then can reinsert at a different locus in the genome of that cell. Depending on when this occurs, it results in either novel heritable germline variation or somatic mutation that can alter cellular function in only the individual affected. The former generates a large reservoir of *de novo* genomic variation in the population which, to date, is poorly characterised as it is very seldom annotated properly in the DNA sequence databases. The mobilisation or 'jumping' of RTEs is proposed to increase both with age within the adult CNS and to comprise one of the key mechanisms underpinning age-related CNS problems [52-54]**; several instances where this variation has been addressed, have associated it with disease [48,55]*. More recently bioinformatic analysis has improved to allow robust identification of RTEs using programmes such as TEBreak (https://github.com/adamewing/tebreak) and MELT [56].

Development

The basic transcriptional mechanisms outlined in this review will also operate during development. Conditions such as schizophrenia and autism are often referred to as having a neurodevelopmental origin [57,58]. Modulation of the transcriptome during development could have a significant effect on the wiring of the brain and therefore how information is processed in the future. Early *in vivo* work using a mouse transgenic model indicated how a human serotonin transporter VNTR could differentially affect gene expression in key areas of serotonergic lineage based on the copy number of the repeat unit [32,59]. The regulation of regulatory domains during development will be determined by the co-expression of transcription factors as exemplified by the schizophrenia associated gene CACNA1C whose expression in development mirrors that of the transcription factor EZH2 an important regulator of the CACNA1C gene promoter [14]. An argument can be made that alterations in the transcriptome at specific times in foetal development could result in a physical change in neuronal connections that would be more difficult to correct than transcriptome changes in the adult [60].

Summary

The identification of variation in the non-coding part of the genome which affects the regulation of gene expression in part explains the often episodic nature of mental health conditions. In addition, it offers the potential for resolution of these conditions by a variety of interventions ranging from pharmaceutical to cognitive behavioural therapy which modify the signalling pathways targeting specific gene regulatory domains, modulation of the 'stress' driving such pathways would alter the transcriptome and hence brain chemistry, Figure 1. A prior exposure to trauma or stress could leave a molecular scar of that event, represented by an epigenetic change which alters parameters of transcriptional or post transcriptional regulation in the medium to long term [61]. It is often considered that the environmental challenge needed to affect mental health should be severe, which is not necessarily correct. For example 'normal' child development could also have an effect on mental health and wellbeing [23,27,33,62]. Similarly a more general approach to maintaining good mental health via diet and exercise could play a role as they could affect the cellular signalling pathways that affect mental health [63,64]. However these issues only illustrate the complexity of defining 'life style/environment' and its effect on our wellbeing given the complex nature of life-long experiences in defining our transcriptome, which in

turn affects the neurochemistry that ultimately shapes CNS function. It's often said that the genome is the roadmap through which 'life style' shapes the individual, however one can argue that the roadmap is unique for each one of us and we all have our own route to travel [65].

Conflict of interest

None of the authors have interests to declare, thus we are stating officially 'Declarations of interest: none'.

References

1. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, et al.: **ENCODE data in the UCSC Genome Browser: year 5 update**. *Nucleic Acids Res* 2013, **41**:D56-63.
2. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, et al.: **The Encyclopedia of DNA elements (ENCODE): data portal update**. *Nucleic Acids Res* 2018, **46**:D794-D801.**
The ENCODE consortia have given us a tool box for interpreting the potential function of DNA element
3. Rigden DJ, Fernandez XM: **The 2018 Nucleic Acids Research database issue and the online molecular biology database collection**. *Nucleic Acids Res* 2018, **46**:D1-D7.
4. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC**. *Genome Res* 2002, **12**:996-1006.
5. Ovcharenko I, Nobrega MA, Loots GG, Stubbs L: **ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes**. *Nucleic Acids Res* 2004, **32**:W280-286.
6. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al.: **In vivo enhancer analysis of human conserved non-coding sequences**. *Nature* 2006, **444**:499-502.
7. Matsunami M, Saitou N: **Vertebrate paralogous conserved noncoding sequences may be related to gene expressions in brain**. *Genome Biol Evol* 2013, **5**:140-150.
8. Dickel DE, Ypsilanti AR, Pla R, Zhu Y, Barozzi I, Mannion BJ, Khin YS, Fukuda-Yuzawa Y, Plajzer-Frick I, Pickle CS, et al.: **Ultraconserved Enhancers Are Required for Normal Development**. *Cell* 2018, **172**:491-499 e415.
9. Attanasio C, Nord AS, Zhu Y, Blow MJ, Li Z, Liberton DK, Morrison H, Plajzer-Frick I, Holt A, Hosseini R, et al.: **Fine tuning of craniofacial morphology by distant-acting enhancers**. *Science* 2013, **342**:1241006.**
A tour de force of how regulatary DNA can result in no two faces being the same

10. Van der Auwera S, Peyrot WJ, Milaneschi Y, Hertel J, Baune B, Breen G, Byrne E, Dunn EC, Fisher H, Homuth G, et al.: **Genome-wide gene-environment interaction in depression: A systematic evaluation of candidate genes: The childhood trauma working-group of PGC-MDD**. *Am J Med Genet B Neuropsychiatr Genet* 2018, **177**:40-49.

11. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Borglum AD, Breen G, Cichon S, Edenberg HJ, Faraone SV, Gelernter J, et al.: **Psychiatric Genomics: An Update and an Agenda**. *Am J Psychiatry* 2018, **175**:15-27.

12. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS, et al.: **Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs**. *Nat Genet* 2013, **45**:984-994.

13. Power RA, Tansey KE, Buttenschon HN, Cohen-Woods S, Bigdeli T, Hall LS, Kutalik Z, Lee SH, Ripke S, Steinberg S, et al.: **Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium**. *Biol Psychiatry* 2017, **81**:325-335.

14. Billingsley KJ, Manca M, Gianfrancesco O, Collier DA, Sharp H, Bubb VJ, Quinn JP: **Regulatory characterisation of the schizophrenia-associated CACNA1C proximal promoter and the potential role for the transcription factor EZH2 in schizophrenia aetiology**. *Schizophr Res* 2018.

15. Davidson S, Shanley L, Cowie P, Lear M, McGuffin P, Quinn JP, Barrett P, MacKenzie A: **Analysis of the effects of depression associated polymorphisms on the activity of the BICC1 promoter in amygdala neurones**. *Pharmacogenomics J* 2016, **16**:366-374.

16. Hing B, Davidson S, Lear M, Breen G, Quinn J, McGuffin P, Mackenzie A: **A Polymorphism Associated with Depressive Disorders Differentially Regulates Brain Derived Neurotrophic Factor Promoter IV Activity**. *Biol Psychiatry* 2012, **71**:618-626.

17. Davidson S, Lear M, Shanley L, Hing B, Baizan-Edge A, Herwig A, Quinn JP, Breen G, McGuffin P, Starkey A, et al.: **Differential activity by polymorphic variants of a remote enhancer that supports galanin expression in the hypothalamus and amygdala: implications for obesity, depression and alcoholism**. *Neuropsychopharmacology* 2011, **36**:2211-2221.

18. Gianfrancesco O, Warburton A, Collier DA, Bubb VJ, Quinn JP: **Novel brain expressed RNA identified at the MIR137 schizophrenia-associated locus**. *Schizophr Res* 2017, **184**:109-115.

19. Gianfrancesco O, Griffiths D, Myers P, Collier DA, Bubb VJ, Quinn JP: **Identification and Potential Regulatory Properties of Evolutionary Conserved Regions (ECRs) at the Schizophrenia-Associated MIR137 Locus**. *J Mol Neurosci* 2016, **60**:239-247.

20. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, et al.: **Integrative analysis of 111 reference human epigenomes**. *Nature* 2015, **518**:317-330.

21. Manca M, Pessoa V, Lopez AI, Harrison PT, Miyajima F, Sharp H, Pickles A, Hill J, Murgatroyd C, Bubb VJ, et al.: **The Regulation of Monoamine Oxidase A Gene Expression by Distinct Variable Number Tandem Repeats**. *J Mol Neurosci* 2018.

22. 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* 2012, **17**:156-170.

23. Murgatroyd C, Quinn JP, Sharp HM, Pickles A, Hill J: **Effects of prenatal and postnatal depression, and maternal stroking, at the glucocorticoid receptor gene**. *Transl Psychiatry* 2015, **5**:e560.*

Interaction of mother and child affecting epigenetic variation in the child

24. Nagy C, Suderman M, Yang J, Szyf M, Mechawar N, Ernst C, Turecki G: **Astrocytic abnormalities and global DNA methylation patterns in depression and suicide**. *Mol Psychiatry* 2015, **20**:320-328.

25. Zhao J, Cheng F, Jia P, Cox N, Denny JC, Zhao Z: **An integrative functional genomics framework for effective identification of novel regulatory variants in genome-phenome studies**. *Genome Med* 2018, **10**:7.

26. 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* 2008, **27**:103-104, 108.

27. 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* 2013, **54**:1308-1317.

28. Haddley K, Bubb VJ, Breen G, Parades-Esquivel UM, Quinn JP: **Behavioural genetics of the serotonin transporter**. *Curr Top Behav Neurosci* 2012, **12**:503-535.

29. Ali FR, Vasiliou SA, Haddley K, Paredes UM, Roberts JC, Miyajima F, Klenova E, Bubb VJ, Quinn JP: **Combinatorial interaction between two human serotonin transporter gene variable number tandem repeats and their regulation by CTCF**. *J Neurochem* 2010, **112**:296-306.

30. 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* 2007, **27**:2793-2801.

31. Guindalini C, Howard M, Haddley K, Laranjeira R, Collier D, Ammar N, Craig I, O'Gara C, Bubb VJ, Greenwood T, et al.: **A dopamine transporter gene functional variant associated with cocaine abuse in a Brazilian sample**. *Proc Natl Acad Sci U S A* 2006, **103**:4552-4557.

32. 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**. *Proc Natl Acad Sci U S A* 1999, **96**:15251-15255.

33. 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 Behav* 2013, **12**:388-396.

34. Brotons O, O'Daly OG, Guindalini C, Howard M, Bubb J, Barker G, Dalton J, Quinn J, Murray RM, Breen G, et al.: **Modulation of orbitofrontal response to amphetamine by a functional variant of DAT1 and in vitro confirmation**. *Mol Psychiatry* 2010.

35. Warburton A, Breen G, Bubb VJ, Quinn JP: **A GWAS SNP for Schizophrenia Is Linked to the Internal MIR137 Promoter and Supports Differential Allele-Specific Expression**. *Schizophr Bull* 2016, **42**:1003-1008.

36. Warburton A, Breen G, Rujescu D, Bubb VJ, Quinn JP: **Characterization of a REST-Regulated Internal Promoter in the Schizophrenia Genome-Wide Associated Gene MIR137**. *Schizophrenia Bulletin* 2015, **41**:698-707.*
Demonstration of how a polymorphic regulatory variant could drive differential expression of the microRNA MIR137 by the action of a transcription previously predicted to be a major modulator of chromatin structure in schizophrenia

37. Bilgin Sonay T, Carvalho T, Robinson MD, Greminger MP, Krutzen M, Comas D, Highnam G, Mittelman D, Sharp A, Marques-Bonet T, et al.: **Tandem repeat variation in human and great ape populations and its impact on gene expression divergence**. *Genome Res* 2015, **25**:1591-1599.

38. Quilez J, Guilmatre A, Garg P, Highnam G, Gymrek M, Erlich Y, Joshi RS, Mittelman D, Sharp AJ: **Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans**. *Nucleic Acids Res* 2016, **44**:3750-3762.

39. Sawaya S, Bagshaw A, Buschiazzo E, Kumar P, Chowdhury S, Black MA, Gemmell N: **Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements**. *PLoS ONE* 2013, **8**:e54710.

40. Dolzhenko E, van Vugt J, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G, Ajay SS, Rajan V, Lajoie BR, Johnson NH, et al.: **Detection of long repeat expansions from PCR-free whole-genome sequence data**. *Genome Res* 2017, **27**:1895-1903.

41. Liu S, Du T, Liu Z, Shen Y, Xiu J, Xu Q: **Inverse changes in L1 retrotransposons between blood and brain in major depressive disorder**. *Sci Rep* 2016, **6**:37530.

42. Doyle GA, Doucet-O'Hare TT, Hammond MJ, Crist RC, Ewing AD, Ferraro TN, Mash DC, Kazazian HH, Jr., Berrettini WH: **Reading LINEs within the cocaine addicted brain**. *Brain Behav* 2017, **7**:e00678.

43. Doyle GA, Crist RC, Karatas ET, Hammond MJ, Ewing AD, Ferraro TN, Hahn CG, Berrettini WH: **Analysis of LINE-1 Elements in DNA from Postmortem Brains of Individuals with Schizophrenia**. *Neuropsychopharmacology* 2017, **42**:2602-2611.

44. Guffanti G, Gaudi S, Klengel T, Fallon JH, Mangalam H, Madduri R, Rodriguez A, DeCrescenzo P, Glovienka E, Sobell J, et al.: **LINE1 insertions as a genomic risk factor for schizophrenia: Preliminary evidence from an affected family**. *Am J Med Genet B Neuropsychiatr Genet* 2016, **171**:534-545.

45. Bundo M, Toyoshima M, Okada Y, Akamatsu W, Ueda J, Nemoto-Miyauchi T, Sunaga F, Toritsuka M, Ikawa D, Kakita A, et al.: **Increased l1 retrotransposition in the neuronal genome in schizophrenia**. *Neuron* 2014, **81**:306-313.

46. Shpyleva S, Melnyk S, Pavliv O, Pogribny I, Jill James S: **Overexpression of LINE-1 Retrotransposons in Autism Brain**. *Mol Neurobiol* 2018, **55**:1740-1749.

47. Payton A, Sindrewicz P, Pessoa V, Platt H, Horan M, Ollier W, Bubb VJ, Pendleton N, Quinn JP: **A TOMM40 poly-T variant modulates gene expression and is associated with vocabulary ability and decline in nonpathologic aging**. *Neurobiol Aging* 2016, **39**:217 e211-217.

48. Aneichyk T, Hendriks WT, Yadav R, Shin D, Gao D, Vaine CA, Collins RL, Domingo A, Currall B, Stortchevoi A, et al.: **Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly**. *Cell* 2018, **172**:897-909 e821.**
Intergration of molecular, genetic and clinical information to support a retrotransposon insertion is underpinning X-Linked Dystonia-Parkinsonism

49. Savage AL, Wilm TP, Khursheed K, Shatunov A, Morrison KE, Shaw PJ, Shaw CE, Smith B, Breen G, Al-Chalabi A, 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**:e90833.

50. Quinn JP, Bubb VJ: **SVA retrotransposons as modulators of gene expression**. *Mob Genet Elements* 2014, **4**:e32102.*
The evolution of the regulatory components of the primate specific retrotransposon termed an SVA building on the function of VNTR domains

51. 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**:101.

52. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan P, Rizzu P, Smith S, Fell M, et al.: **Somatic retrotransposition alters the genetic landscape of the human brain**. *Nature* 2011.**
Evidence for somatic retrotransposition mutation in the adult CNS

53. Cardelli M: **The epigenetic alterations of endogenous retroelements in aging**. *Mech Ageing Dev* 2018.

54. Van Meter M, Kashyap M, Rezazadeh S, Geneva AJ, Morello TD, Seluanov A, Gorbunova V: **SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age**. *Nat Commun* 2014, **5**:5011.*
Mechanisms which can alter retrotransposon activity in the brain directed by age and possibly nutrition.

55. Hancks DC, Kazazian HH, Jr.: **Roles for retrotransposon insertions in human disease**. *Mob DNA* 2016, **7**:9.

56. Ewing AD: **Transposable element detection from whole genome sequence data**. *Mob DNA* 2015, **6**:24.*
Great review of the different types of bioinformatic programmes for addressing genomic variation

57. Ziats MN, Grosvenor LP, Rennert OM: **Functional genomics of human brain development and implications for autism spectrum disorders**. *Transl Psychiatry* 2015, **5**:e665.

58. Owen MJ, O'Donovan MC: **Schizophrenia and the neurodevelopmental continuum:evidence from genomics**. *World Psychiatry* 2017, **16**:227-235.

59. MacKenzie A, Quinn JP: **Post-genomic approaches to exploring neuropeptide gene mis-expression in disease**. *Neuropeptides* 2004, **38**:1-15.

60. Gaspar P, Cases O, Maroteaux L: **The developmental role of serotonin: news from mouse molecular genetics**. *Nat Rev Neurosci* 2003, **4**:1002-1012.

61. Bryant RA, Edwards B, Creamer M, O'Donnell M, Forbes D, Felmingham KL, Silove D, Steel Z, Nickerson A, McFarlane AC, et al.: **The effect of post-traumatic stress disorder on refugees' parenting and their children's mental health: a cohort study**. *Lancet Public Health* 2018, **3**:e249-e258.

62. Eun JD, Paksarian D, He JP, Merikangas KR: **Parenting style and mental disorders in a nationally representative sample of US adolescents**. *Soc Psychiatry Psychiatr Epidemiol* 2018, **53**:11-20.

63. Warburton A, Vasieva O, Quinn P, Stewart JP, Quinn JP: **Statistical analysis of human microarray data shows that dietary intervention with n-3 fatty acids, flavonoids and resveratrol enriches for immune response and disease pathways**. *Br J Nutr* 2018, **119**:239-249.

64. Marx W, Moseley G, Berk M, Jacka F: **Nutritional psychiatry: the present state of the evidence**. *Proc Nutr Soc* 2017, **76**:427-436.

65. Mac Giollabhui N, Hamilton JL, Nielsen J, Connolly SL, Stange JP, Varga S, Burdette E, Olino TM, Abramson LY, Alloy LB: **Negative cognitive style interacts with negative life events to predict first onset of a major depressive episode in adolescence via hopelessness**. *J Abnorm Psychol* 2018, **127**:1-11.

Figure and Table Caption

Figures and Boxes for Current Opinions in Psychology

Box1: Genome Browsers to interrogate DNA function

Encyclopaedia of DNA Elements (https://www.encodeproject.org/)

Evolutionary conserved regions browser (https://ecrbrowser.dcode.org/

NIH Roadmap Epigenomics Consortium (http://www.roadmapepigenomics.org/)

University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/)

Box 2: Genetic variation commonly found in the genome

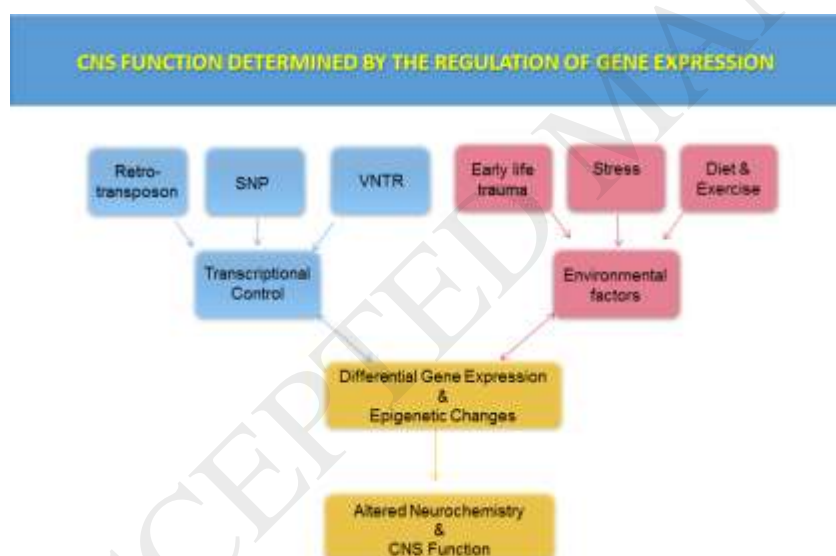**SNP: single nucleotide polymorphism**

The most common form of genetic variation in the genome. When located in a coding exon it may change the amino acid encoded thus altering the protein. In non-coding DNA it may have a variety of functions, however since transcription factors are sequence specific DNA binding proteins it could change the affinity or specificity of that interaction thus altering the function of a specific DNA element.
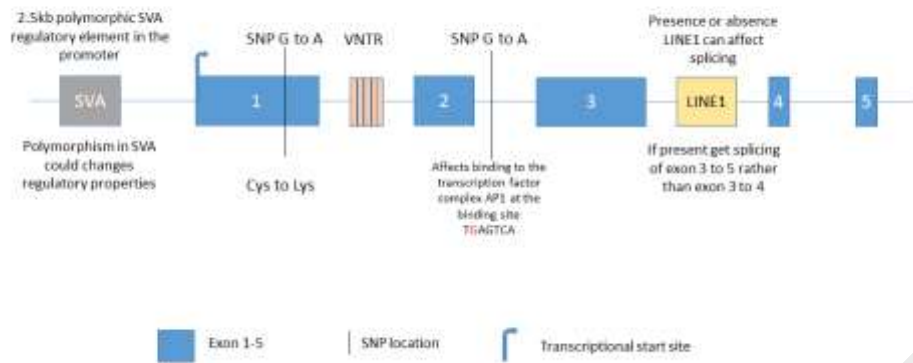
VNTR: variable number tandem repeats

These are adjacent repeats of single or more nucleotides and are a large component of the non-coding DNA. Often used in forensic DNA analysis as a genetic fingerprint given their variation in the genome. They are routinely separated into microsatellites (ranging in repeat length from 1–9 base pairs) and minisatellites (ranging in length from 10–60 base pairs), both classes are generally repeated 3-50 times. They are also found in exons such as in the Huntington gene.

RTE: retrotransposon

RTEs constitute approximately 42% of the human genome. Retrotransposons can be subdivided in long terminal repeat (LTR) retrotransposons (also termed 'endogenous retroviruses') and non-LTR retrotransposons lacking LTRs. Non-LTR retrotransposons propagate via a copy-and-paste mechanism, meaning that retrotransposon transcripts are reverse transcribed into a cDNA intermediate which is integrated into a new site of the host genome. Non-LTR retrotransposons can be sub-divided into LINE-1, *Alu* and SINE-VNTR-*Alu* (SVA) families which have the capacity to modify gene expression at transcriptional and post transcriptional levels. They constitute 21%, 13% and 0.2% respectively of the human genome.

Different types of polymorphism over a gene

2.5kb polymorphic SVA regulatory element in the promoter

SVA

Polymorphism in SVA could changes regulatory properties

SNP G to A

1

Cys to Lys

VNTR

SNP G to A

2

Affects binding to the transcription factor complex AP1 at the binding site TGAGTCA

3

Presence or absence LINE1 can affect splicing

LINE1

If present get splicing of exon 3 to 5 rather than exon 3 to 4

4

5

Exon 1-5   SNP location   Transcriptional start site