ARTICLE

The influence of X chromosome variants on trait neuroticism

Michelle Luciano [1] · Gail Davies[1] · Kim M. Summers[2] · W. David Hill[1] · Caroline Hayward [3] · David C. Liewald [1] ·
David J. Porteous[4] · Catharine R. Gale[1,5] · Andrew M. McIntosh [1,6] · Ian J. Deary[1]

Abstract

Autosomal variants have successfully been associated with trait neuroticism in genome-wide analysis of adequately powered samples. But such studies have so far excluded the X chromosome from analysis. Here, we report genetic association analyses of X chromosome and XY pseudoautosomal single nucleotide polymorphisms (SNPs) and trait neuroticism using UK Biobank samples ($N = 405,274$). Significant association was found with neuroticism on the X chromosome for 204 markers found within three independent loci (a further 783 were suggestive). Most of the lead neuroticism-related X chromosome variants were located in intergenic regions ($n = 397$). Involvement of *HS6ST2*, which has been previously associated with sociability behaviour in the dog, was supported by single SNP and gene-based tests. We found that the amino acid and nucleotide sequences are highly conserved between dogs and humans. From the suggestive X chromosome variants, there were 19 nearby genes which could be linked to gene ontology information. Molecular function was primarily related to binding and catalytic activity; notable biological processes were cellular and metabolic, and nucleic acid binding and transcription factor protein classes were most commonly involved. X-variant heritability of neuroticism was estimated at 0.22% (SE = 0.05) from a full dosage compensation model. A polygenic X-variant score created in an independent sample (maximum $N \approx 7,300$) did not predict significant variance in neuroticism, psychological distress, or depressive disorder. We conclude that the X chromosome harbours significant variants influencing neuroticism, and might prove important for other quantitative traits and complex disorders.

✉ Michelle Luciano
   michelle.luciano@ed.ac.uk

1   Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Edinburgh, UK

2   Mater Research Institute-University of Queensland, Translational Research Institute, Brisbane, QLD, Australia

3   MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

4   Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

5   MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

6   Division of Psychiatry, University of Edinburgh, Edinburgh, UK

Introduction

Neuroticism is a dispositional personality trait tapping negative emotion, like experience of anxiety, mood swings, and negative affect. It is a predictor of a range of psychiatric traits and disorders [1, 2]. Our understanding of the genetic influences on people's differences in neuroticism has increased substantially with the availability of large biological resources like UK Biobank. In a genome-wide autosomal analysis of neuroticism in UK Biobank, Luciano and colleagues [3] found 116 independent single nucleotide polymorphisms (SNPs) associated with neuroticism and replicated 15 of them in a large independent sample, comprising the Genetics of Personality Consortium and 23andMe. These variants and the biological pathways that were identified provided some understanding about the biological processes and molecular mechanisms of importance, for example, neural genesis and differentiation pathways. There was genetic overlap between neuroticism and depressive disorder and anti-depressant drug targets. Large-scale association analysis of neuroticism and the genetic variants on the sex chromosomes has not yet been

performed; therefore, our current understanding of the genetic factors that contribute to neuroticism is incomplete. The analysis of X chromosome variation (including pseudoautosomal variation) and neuroticism, in UK Biobank, is the focus of the current study.

In humans, phenotypic sex is determined by the X and Y sex chromosomes, with females having two X chromosomes and males having a single X chromosome and a single Y chromosome. To prevent overexpression of genes on the X chromosome in females, one X is randomly inactivated during development so that females have only one active X in each cell. This has implications for genetic mutations in genes located on the X chromosome and can result in phenotypes having a different prevalence between the sexes [4]. Sexual dimorphism is present in disease, including psychiatric disorder [5], and the X chromosome has a role in the development of the human brain (see [6]), suggesting that it may harbour variants associated with behaviour. X chromosome association analysis of behavioural traits is limited, but in a well-powered genome-wide association study (GWAS) of educational attainment, 10 independent X chromosome single-nucleotide polymorphism (SNP) associations were significant, and X chromosome SNP heritability was estimated at 0.3% in both sexes [7]. For 20 mostly physical/physiological traits in UK Biobank the mean X chromosome SNP heritability was 0.62% in males and 0.30% in females [8]. Given that men and women differ in the mean level of various personality traits [9] and that X chromosome variants have been associated with complex traits, we expect that the X chromosome will be important in understanding the genetic architecture of human personality.

Neuroticism scores are generally higher in women than men [9, 10], and neuroticism-related mood disorders such as anxiety and depression show greater prevalence in women than in men [11]. Findings from twin studies suggested that the genetic influences on neuroticism might be larger for women than men [12, 13]. In 6,828 twin brothers and 8,104 twin sisters from Finland aged 24–53 years, heritability of neuroticism (9-item Eysenck scale) differed significantly between men (.31) and women (.42) [14]. Opposite-sex twins were not sampled, precluding analysis of qualitative sex limitation, that is, where trait variation is influenced by different factors between the sexes. An integrated analysis of this Finnish sample with twins and extended family members from Australia ($n = 7,532$) and the US ($n = 20,554$) showed that the same genes influenced neuroticism in both sexes, but that non-additive genetic effects were significantly larger in men than women (21.9% versus 13.1%; additive genetic variance of 15.7% versus 38.2%) [15]. In 48,850 Australian and US individuals from extended twin families, broad heritability of neuroticism was slightly higher in women (41%) than men (35%) [16].

An alternative method of estimating trait heritability in unrelated individuals using observed genotype data found broadly consistent results to the pedigree-based analyses. In a population sample (5,016 male, 6,945 female), the heritability of neuroticism, estimated by common autosomal genetic variants, did not differ between sexes [17]. However, point estimates of heritability were higher for men than women (.16, SE .07 versus .06, SE .05; SNP × sex interaction effect P-value = .08). In this same study, analysis of the X chromosome was not-significant, that is, X variants did not contribute to the heritability of neuroticism. Here, we reconsider the contribution of X chromosome variance in neuroticism based on a much larger sample of 405,274 participants.

With regard to the identification of X chromosome loci influencing neuroticism, the largest genetic linkage study of neuroticism, in around 5,000 sibling pairs, did not show evidence of linked QTLs on the X chromosome [18]. However, linkage analysis is underpowered to detect the small QTL effects expected for complex traits. Similarly, genetic association studies including the X chromosome have so far lacked statistical power. Several small genome-wide association studies (GWAS) of neuroticism (e.g., N ranging ≈2,200 to ≈4,000) have analysed the X chromosome with null results [19, 20]. The present X chromosome association analysis (XWAS) comprises over 400,000 UK Biobank participants and tests whether X chromosome variants are associated with neuroticism scores. Further, we investigate the XY pseudoautosomal region (the ends of each arm of the chromosome that are common to both sex chromosomes, escape X-inactivation and participate in recombination) [21]. Results will be of special value to research on mood disorders, which are genetically highly correlated with neuroticism [22].

Materials and methods

Neuroticism was measured in UK Biobank, an open resource including more than 500,000 adults aged 40–69 years residing in the UK (http://www.ukbiobank.ac.uk), with data collected from 2006 through to 2010 [23]. We have previously published genome-wide autosomal results on neuroticism in unrelated individuals from this cohort ($N = 329,824$) [3]; here we include related individuals and focus on the X chromosome ($N = 405,274$) [24]. X chromosome and XY pseudoautosomal genetic variants were analysed for their association with neuroticism, measured as the summed score of 12 items from the Eysenck Personality Questionnaire-Revised Short Form (EPQ-R-SF) [25]. Information on the imputation of missing data and its distribution can be found in our previous report [3]. Neuroticism scores were residualised for age, assessment centre,

genotype batch, array, and 40 genetic principal components. 408,091 participants (45.94% male) of British descent had high-quality genotyping imputed to the UK10K haplotype, 1000 Genomes Phase 3, and Haplotype Reference Consortium panels. Quality control of this dataset was performed as described by Bycroft and colleagues [24], and individuals were removed sequentially based on non-British ancestry, high missingness, high relatedness (samples which have more than 10 putative third-degree relatives), and gender mismatch. All bi-allelic variants with a minor allele frequency ≥ 0.000009 and an information/imputation quality score of ≥0.1 were analysed.

Regression models of neuroticism on genetic variants (accounting for genotype uncertainty) were run separately for men and women, specifying male genotypes as 0/2 and female genotypes as 0/1/2 which assumes random X inactivation. Association analyses were performed using BOLT-LMM software, using a linear mixed model [26]. Meta-analysis of the association results of men ($n = 186,015$) and women ($n = 219,259$) allowed increased statistical power and by incorporating heterogeneity tests we could investigate differences in the magnitude of variant effects between the sexes [27]. On the X chromosome, 1,686,693 SNPs were meta-analysed. Pseudoautosomal regions were treated as autosomes and analysed under an additive model ($n = 40,764$ SNPs).

Downstream mapping of X chromosome association signals

FUnctional Mapping and Annotation of genetic associations (FUMA) was used to interrogate the association results, including identification of independent association signals, gene-based analysis, functional annotation, and gene expression [28]. Independent significant SNPs were those with a $P \leq 5 \times 10^{-8}$ and not in linkage disequilibrium (LD; $r^2 < .6$) with other significant SNPs. Tagging SNPs were those with MAF ≥ .0005 and in LD ($r^2 < .6$) with one or more of the independent significant SNPs; they were from the 1000G reference panel and not necessarily included in the present association analysis. Any associated loci within 250 kb of each other were considered a single locus. Lead SNPs were those independent significant SNPs that were even less closely related (LD $r^2 < .1$). Gene-based analysis is performed in FUMA using MAGMA [29]. Using the default settings, SNPs were allocated to genes according to the NCBI 37.3 build positions without extending the boundary beyond the gene. LD was taken into account with reference to the European panel of the 1000G data (phase 1, release 3). Gene-based statistics were available for 762 genes, but given that autosomal genes have been previously tested in this sample [3], a genome-wide threshold for significance was adopted α < 2.65×10^{-6} (.05/18,842). Expression quantitative trait loci

(eQTL) and functional annotation in FUMA relies on the publicly available Genotype-Tissue Expression Portal (GTEx) and Regulome DB databases.

Estimation of X chromosome SNP-based heritability

A genetic relationship matrix (GRM) was created using the autosomal SNPs; a relatedness cut-off of 0.05 was applied to this GRM. The resulting set of unrelated individuals were used in the following GCTA-GREML analyses [30]. We created an X chromosome GRM using the previously defined set of unrelated individuals ($N = 313,467$). This GRM was then parametrised under three assumptions of dosage compensation [31]: (1) equal X-linked variance for males and females; (2) no dosage compensation, meaning that each allele has a similar effect in both males and females, but, as both X chromosomes are active for females, and only one in males, there is twice the amount of genetic variance linked to the X chromosome in females; (3) full dosage compensation, i.e., that one of the X chromosomes is completely inactive for females, meaning that in females each allele has only half the effect of an allele in a male. Under full dosage compensation for females, genetic variance on the X chromosome will be around half of what will be observed in males. To estimate the proportion of variance explained by X chromosome common SNPs (h_X^2), we fitted the parameterised GRMs for the X chromosome in a mixed linear model whilst simultaneously estimating the proportion of variance explained by all common autosomal SNPs (h^2). These analyses were performed on a random sub-sample of ≈150,000 individuals due to computing resource capacity.

Polygenic prediction into generation Scotland

Neuroticism polygenic X chromosome prediction of neuroticism, psychological distress, and depressive disorder was carried out in Generation Scotland (GS) [32]. Polygenic scores were calculated in PLINK [33] based on the female and male SNP association results for neuroticism in UK Biobank. Prediction was carried out in unrelated individuals (n ranged 3,908–4,189 for women; 3,004–3,179 for men). Only SNPs in linkage equilibrium ($r^2 < .10$ within a 250-kb window) were included in the polygenic score. Individuals from GS who participated in UK Biobank were excluded ($n = 290$ of those who were unrelated). P-value thresholds of .01, .05, .1, .5 and 1 for the X chromosome association results were used to create 5 polygenic scores. Polygenic scores for each threshold were based on 163 (female)/166 (male) SNPs, 527/514 SNPs, 893/879 SNPs, 2,788/2,786 SNPs, and 4,122/4,137 SNPs, respectively. The associations between the polygenic score and neuroticism/psychological distress in GS were tested using linear regression, controlling

for age and the first 10 genetic principal components. Depression status was analysed by logistic regression with the same covariate controls. To correct for multiple testing across threshold and trait, a false discovery rate (FDR) method was used [34].

Phylogenetic analysis

A phylogenetic tree for the *HS6ST2* gene was retrieved using the Ensembl database *Gene tree* function (Ensembl release 93, accessed July 2018; http://www.ensembl.org). All nodes except those containing the human and dog genes were collapsed for ease of visualisation. Predicted protein sequences for HS6ST2 from a range of species were downloaded from Ensembl. The longest available predicted protein sequence was captured for each species. The sequences were aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) with the default parameters.

Results

X chromosome and XY pseudoautosomal meta-analysis association results showed evidence of genomic inflation (quantile-quantile plots in Supplementary Figures 1 and 2), with respective lambdas of 1.07 and 1.06. There were 204 SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$) for the X chromosome, but none for the XY pseudoautosomal region. At a suggestive level ($P < 1 \times 10^{-5}$), a further 783 X chromosome variants and 3 XY pseudoautosomal variants were identified (Supplementary Tables 1 and 2). X and XY pseudoautosomal association P-values are depicted in Fig. 1; Supplementary Table 3 shows annotation of the significant 1000G SNPs. Sex-stratified association results can be found in Supplementary Figures 3 and 4; none were significant. Among the significant X chromosome SNPs from the meta-analysis, there were three distinct loci:

Xp21.3 (183.8 kb, containing 65 SNPs), Xq25 (295.2 kb, 309 SNPs), and Xq26.2 (122.4 kb, 41 SNPs). Regional association plots for these loci are shown in Supplementary Figures 5–7. The respective lead SNPs (significant and with $r^2 < .1$) within these regions were rs6630665 ($P = 1.023 \times 10^{-8}$), rs177010 ($P = 2.594 \times 10^{-11}$) and rs5977754 ($P = 2.521 \times 10^{-8}$). Of the lead SNPs, 397 SNPs were intergenic and 48 intronic. One protein-coding gene was mapped to the associated part of Xq26.2 (131817931–131940379bp): *HS6ST2* (Heparan Sulphate 6-*O*-Sulfotransferase 2); there were no protein coding genes in the loci identified within Xp21.3 and Xq25.

None of the four independent SNPs (significant and with $r^2 < .6$ rather than $r^2 < .1$ for lead SNPs; 2 in Xq25) were expression quantitative trait loci (eQTLs; Supplementary Table 3). Regulome DB scores (http://www.regulomedb.org/) were available for two of the independent SNPs (rs6630665, rs5977754), which showed minimal DNA binding evidence, and thus, unlikely to have a regulatory role (Supplementary Table 3). The Protein Analysis Through Evolutionary Relationships (PANTHER) classification system [35] was used to describe the genes located near our suggestive X chromosome variants. Of the 30 genes identified, 19 could be mapped to different biological (23 hits), cellular (11), molecular (13), and protein (17) classes (Supplementary Figures 8–11). Consistent with classification of the prominent genes from autosomal analysis [3], metabolic and cellular processes were the primary biological classes, nucleic acid binding and transcription factor were among the most important protein classes, and binding and catalytic activity were the principal molecular functions.

Gene-based tests for the X chromosome showed that seven genes were significantly associated with neuroticism (Supplementary Table 4 shows all gene results). They included: *HS6ST2* (containing 1190 SNPs), *IL1RAPL1* (6670), *HPRT1* (226), *PHF6* (187), *DMRTC1* (50), *CTPS2*

Fig. 1 X Chromosome association results for neuroticism in UK Biobank ($N = 405{,}274$). XY pseudoautosomal SNPs are in darker blue
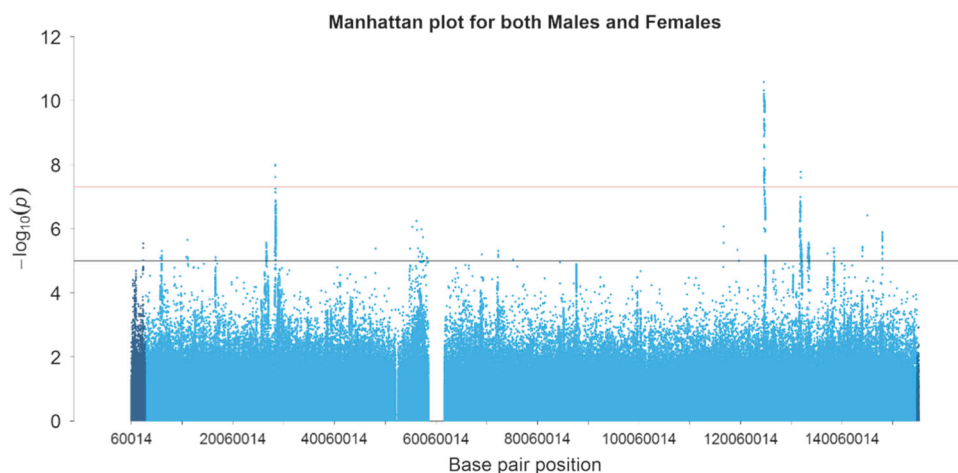
(528) and *NLGN4X* (2139). Only one—*HS6ST2*—showed single SNPs attaining significance. For the XY pseudoautosomal region, the *XG* gene (including 45 SNPs) was mapped and was not significant ($P = .5$).

Genome Complex Trait Analysis (GCTA) [31] SNP heritability estimates did not differ significantly between models with differing dosage compensation specification, as judged by their log likelihoods ($P > .05$). In the model with equal X-linked variance for males and females, the estimate for the X chromosome variance was 0.34% (SE = 0.07); the variance from the autosomes was 14.37% (SE = 0.42). The model with no dosage compensation (greater female X-linked variance) produced estimates of 0.43% (SE = 0.09) for X-linked variance, and 14.36% (SE = 0.42) for autosomal variance. The model with full dosage compensation (greater male X-linked variance) gave variance estimates of 0.22% (SE = 0.05) for X chromosome, and 14.39% (SE = 0.43) for the autosomes. X-linked variance was significant for all models ($P < 6.43 \times 10^{-8}$).

X chromosome polygenic scores were calculated in unrelated individuals from Generation Scotland [32] separately for men and women based on the sex-stratified association analyses. These were used to predict Eysenck Personality Questionnaire neuroticism scores, General Health Questionnaire 28-item [36] total scores (i.e., general psychological distress), and Structured Clinical Interview for DSM-IV (SCID)-diagnosed depressive disorder [37]. Following FDR correction, none of the results were significant ($P > .05$; Supplementary Table 5).

Because *HS6ST2* was the only gene within which single SNPs attained significance, was mapped within the key region at Xp26.2, and had been implicated in a relevant behavioural trait in the dog [38], it was investigated further. A phylogenetic tree was constructed from all available orthologous sequences (Supplementary Figure 12) and showed that the dog sequence was in a group with the cat and leopard, and close to the horse, another companion animal, and to the alpaca, a domesticated species. Somewhat more distant were other domesticated species—the goat, pig and sheep—but these were noted to have fragmented and incomplete sequences which could affect the analysis. In all mammalian species where a chromosome had been assigned, the orthologous gene was on the X chromosome; where there was no chromosome assigned, the gene was in a region syntenic with the X chromosome region of human and was flanked by the same set of genes (*MBNL3*, *RAP2C* and *FRMD7* on one side and *USP26*, *GPC4* and *GPC3* on the other). Alignment of the *HS6ST2* sequence from a selection of mammals, birds and fish revealed that there was strong conservation of the amino acid sequence across the protein, particularly within the putative transmembrane domain, the P-loop containing NTP hydrolase domain, and the region omitted by alternative splicing (highlighted in yellow, green and pink respectively in Supplementary Figure 13). The dog and human predicted protein and cDNA sequences for HS6ST2 showed 93% identity (Fig. 2), whereas the similarity between dog, cat and horse was 96–98% (Supplementary Figure 13). In contrast, for another X chromosome gene, *HPRT1*, which maps within 2 mbp of *HS6ST2* in most mammals, dog and human showed 100% identity whereas cat showed 93% identity with human/dog and 89% with horse. The dN/dS ratio between dog and human was 0.17, indicating that there has been strong selection pressure to conserve the amino acid sequence [39, 40]; this is consistent with the alignment of all species (Supplementary Figure 13).

Discussion

We have identified three independent loci on the X chromosome associated with scores on the personality trait of neuroticism. In gene-based tests, 7 genes were associated with neuroticism. We could not discriminate between different gene dosage compensation models, although X chromosome variance was significant in each one. Note that a recent study in UK Biobank estimated an X chromosome dosage compensation ratio for neuroticism that was consistent with a full dosage compensation model [8]. Under this model, we found that 0.22% of variance in neuroticism was explained by the X chromosome, which represents 1.5% of the total variance due to common genetic variants. A composite score of these variants did not predict variance in neuroticism and related traits in an independent sample. The XY pseudoautosomal region did not uncover any significant variants.

Significant SNPs were mapped to a protein coding gene, *HS6ST2*; this gene was also the most significant one arising from the gene-based tests. This gene encodes a member of the heparan sulphate sulfotransferase gene family and has a role in cranial nerve development [41]. Two isoforms identified by cDNA cloning in humans showed the long form to be expressed in brain-related tissue [42]. Sociability traits (attitude to human strangers) have been previously mapped to *HS6ST2* in dogs [38], with the neurobiological pathways involved in these emotions highly conserved in vertebrates [43, 44]. We found that the amino acid and nucleotide sequences are highly conserved between dogs and humans, particularly in the transmembrane and NTP hydrolase domains. The canine sequence was even more similar to other companion animals, the cat and horse. Since the similarity for another X chromosome gene, *HPRT1*, was lower, it is possible that this reflects selection for trait(s) associated with this gene. Sociability would have been strongly selected for in the evolution of companion animals, which may explain the strong conservation of this sequence.

```
human_HS6ST2-201    MALPACAVREFEPPRQPERGAPVRTTCPRRHSRVEAELAASRPGSVAASVRAGPPRGVSH    60
dog_HS6ST2-201      MALPACAAQALGPPLQPERGAPARTTCPRRHSRVEAELAGSRPGSVAASVRAGPPRGVSR    60
                    *******.: : ** *******.***************.********************:

human_HS6ST2-201    GFHTRPLLDKPRKASSSLAGAACAPLFALLSRGRRRRMHVLRRRWDLGSLCRALLTRGLA    120
dog_HS6ST2-201      GFNSQPLLDEPLKASSSPAGAARTPLFALLPRGRRRRMHDLRRRWDLGSLCRALLTRGLA    120
                    **:;:****:* ***** **** :****** ******* ********************

human_HS6ST2-201    ALGHSLKHVLGAIFSKIFGPMASVGNMDEKSNKLLLALVMLFLFAVIVLQYVCPGTECQL    180
dog_HS6ST2-201      ALGHSLKHVLGAIFSKIFGPLASVGNMDEKSNKLLLALVMLFLFAVIVLQYVCPGTECQL    180
                    ********************:***************************************

human_HS6ST2-201    LRLQAFSSPVPDPYRSEDESSARFVPRYNFTRGDLLRKVDFDIKGDDLIVFLHIQKTGGT    240
dog_HS6ST2-201      LRLQAFSSPMPDPYRSEDESSSRFVPRYNFSRGDLLRKVDFDIKGDDLIVFLHIQKTGGT    240
                    *********:***********:*******:******************************

human_HS6ST2-201    TFGRHLVRNIQLEQPCECRVGQKKCTCHRPGKRETWLFSRFSTGWSCGLHADWTELTSCV    300
dog_HS6ST2-201      TFGRHLVRNIQLEQPCECRVGQKKCTCHRPGKRETWLFSRFSTGWSCGLHADWTELTSCV    300
                    ************************************************************

human_HS6ST2-201    PSVVDGKRDARLRPSRWRIFQILDAASKDKRGSPNTNAGANSPSSTKTRNTSKSGKNFHY    360
dog_HS6ST2-201      PAVVDGKRDARLRPSRWRIFQILDAASKDRRGSPNTNPGANSPSSTKARNTSKSGKNFHY    360
                    *:*************************** *******.*********:*:**********

human_HS6ST2-201    ITILRDPVSRYLSEWRHVQRGATWKASLHVCDGRPPTSEELPSCYTGDDWSGCPLKEFMD    420
dog_HS6ST2-201      ITILRDPVSRYLSEWRHVQRGATWKASLHVCDGRPPTSEELPSCYTGDDWSGCPLKEFMD    420
                    ************************************************************

human_HS6ST2-201    CPYNLANNRQVRMLSDLTLVGCYNLSVMPEKQRNKVLLESAKSNLKHMAFFGLTEFQRKT    480
dog_HS6ST2-201      CPYNLANNRQVRMLSDLTLVGCYNLSVMPEKQRNKVLLESAKSNLKHMAFFGLTEFQRKT    480
                    ************************************************************

human_HS6ST2-201    QYLFEKTFNMNFISPFTQYNTTRASSVEINEEIQKRIEGLNFLDMELYSYAKDLFLQRYQ    540
dog_HS6ST2-201      QYLFEKTFNMNFISPFTQYNTTRASSVEINEEIQKRIEGLNFLDMELYSYAKDLFLQRYQ    540
                    ************************************************************

human_HS6ST2-201    FMRQKEHQEARRKRQEQRKFLKGRLLQTHFQSQGQGQSQNPNQNQSQNPNPNANQNLTQN    600
dog_HS6ST2-201      FMRQKEHQEARQKRQEQRKFLKGRFLQTHFQSQGQGQSQNPSQNQSQNPNLNVNQNVTQN    600
                    ***********:************:************** :********. ******* *.***:***

human_HS6ST2-201    LMQNLTQSLSQKENRESPKQNSGKEQND-NTSNGTNDYIGSVEKWR 645
dog_HS6ST2-201      LIQNLTQNLSHKENRESQKQNPGQEQSDGNTSNGTNDYIGSVEKWR 646
```

Fig. 2 Alignment of human and canine *HS6ST2* genes. The alignment was created using Clustal Omega. The human sequence accession 10 is ENSP00000429473/1·645 and the canine sequence accession 10 is ENSCAFP00000038323/1/646. The putative transmembrane domain is highlighted in yellow; the two P-loop containing NTP hydrolase domains are highlighted in green and the region removed by alternate splicing that produces the short form is highlighted in pink (information derived from Habuchi et al. [42] and Ensembl). Asterisks indicate positions with a conserved amino acid; colons indicate conservation of amino acids with highly similar properties; periods indicate conservation of amino acids with weakly similar properties

The high degree of conservation across vertebrates shown by the phylogenetic tree and the dN/dS ratios indicate that this gene encodes a protein which is under strong evolutionary constraint, particularly in the catalytic domain.

Given that X-linked heritabilities for dog agitation and noise fear are significant [45], there may be other X chromosome variants that will be useful to follow up with phylogenetic analysis once power has increased to detect individual gene associations. Indeed, phylogenetic analysis of autosomes will help to hone in on evolutionary conserved regions influencing neuroticism. A previous GWAS of neuroticism found that SNPs situated in conserved regions were those most significantly enriched in a test of 22 unique genomic groupings [46]. Phylogenetic analysis can provide insight into the biology of human neuroticism. For example, with regard to the function of *HS6ST2*, the study of a knock out mouse line did not find any gross phenotype other than obesity in elderly males [47], but future studies could involve behavioural tests that measure anxiety (a correlate of neuroticism) or breeding experiments could be performed to test whether a behavioural phenotype tracks with the null allele. CRISPR gene editing could be

used to create deletions or single base changes in the mouse gene in the conserved regions based on the phylogenetic analysis and subsequent behaviour and impact on neurons (differentiated in vitro from embryonic stem cells or induced pluripotent stem cells) interrogated.

Our gene-based tests of the XWAS results identified an additional six genes significantly associated with neuroticism. *IL1RAPL1* (Interleukin 1 Receptor Accessory Protein Like 1) is involved in neurite outgrowth, synapse formation and stabilisation. Deletions and mutations in this gene have been associated with intellectual disability and autism spectrum disorder [48, 49]. Further, there has been consistent support for a link between methylation of *IL1RAPL1* and major depressive disorder (MDD), with identified probes showing higher methylation in MDD cases than controls [48, 50]. *PHF6* (Plant Homeodomain (PHD)-like Finger protein 6) mutations are associated with Börjeson–Forssman–Lehmann syndrome with some documented cases characterised by behavioural disturbance related to anxiety, compulsive behaviour, and emotional attachment [51]. A variant close to *NLGN4X* (neuroligin 4 X-linked) has been associated (at a nominal level) with a quantitative index of suicidality in 2,023 depressive cases [52]. Mutations in this gene have been associated with autism spectrum disorder [53] and a role of neuroligins in cognitive function has been theorised to occur through cell adhesion and synapse formation pathways [54].

Our analysis has identified a number of X chromosome genes that are promising to follow up in future analysis. The large increase in power from a previous study of neuroticism that did not find significant X chromosome SNP heritability [17] has enabled us to detect this here. 1.5% of the total common SNP heritability was explained by the X chromosome, which is lower than might be expected given that the X chromosome is moderately gene dense and contains up to 4% of protein coding genes in the human genome [55]. However, the X chromosome heritability estimate is similar to the 2% of total SNP heritability reported for educational attainment [7]. This estimate contrasts noticeably with male pattern baldness, a physical trait where almost 9% of common SNP heritability can be explained by the X chromosome [56], likely due to the importance of sex hormones influencing this trait. Although for physical traits—like height and BMI—observed in both sexes, the X chromosome explains close to 1% of variance, representing around 1.3% and 4.8% of total common SNP heritability, respectively [31]. Given the low contribution of variance in neuroticism from the X chromosome, it is unsurprising that polygenic scores did not predict neuroticism in an independent sample. Polygenic scores created from autosomes explain up to 3% of variance in neuroticism [3], but the heritability due to common autosomal SNPs is

14% compared to about 0.3% for X chromosome variants. Further, polygenic scores were created from the sex-stratified association analyses which had less power than the meta-analysis; with larger sample sizes, X chromosome polygenic prediction may improve.

In UK Biobank, we have found that common SNPs on the X chromosome contribute to the heritability of neuroticism and we identified SNPs and genes that were associated with this important personality trait. In particular, *HS6ST2*—a gene whose protein is highly conserved across vertebrates—stands out as a promising candidate given its association with personality in dog (replication in other breeds might identify coding variant differences in dogs with different genotypes and behaviours). Further research might focus on replication of *HS6ST2* variants in non-human primates whose personality structure is more similar to humans than is dog. Given the strong autosomal genetic correlations between neuroticism and psychiatric traits [3], our study suggests that X chromosome analysis for related traits would successfully identify further genes/biological pathways involved in their aetiology.

URLs

UK Biobank Resource: http://www.ukbiobank.ac.uk
Gene Ontology: http://geneontology.org
METAL: http://csg.sph.umich.edu/abecasis/metal/index.html
PLINK V2: https://www.cog-genomics.org/plink2
Ensembl: http://www.ensembl.org
Clustal Omega: https://www.ebi.ac.uk/Tools/msa/clustalo/

Data availability

The X chromosome association results generated by this analysis can be downloaded from the CCACE data sharing resource; http://www.ccace.ed.ac.uk/node/335.

Compliance with ethical standards

References

1. Hettema JM, Neale MC, Myers JM, Prescott CA, Kendler KS. A population-based twin study of the relationship between neuroticism and internalizing disorders. Am J Psychiatry. 2006; 163:857–64.
2. Lahey BB. Public health significance of neuroticism. Am Psychol. 2009;64:241–56.
3. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat Genet. 2018;50:6–11.
4. Arnold Arthur P. A general theory of sexual differentiation. J Neurosci Res. 2016;95:291–300.
5. Chang D, Gao F, Slavney A, Ma L, Waldman YY, Sams AJ, et al. Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. PLoS ONE. 2014;9:e113684.
6. Zhao C, Gong G. Mapping the effect of the X chromosome on the human brain: neuroimaging evidence from Turner syndrome. Neurosci Biobehav Rev. 2017;80:263–75.
7. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50:1112–21.
8. Sidorenko J, Kassam I, Kemper K, Zeng J, Lloyd-Jones L, Montgomery GW, et al. The effect of X-linked dosage compensation on complex trait variation. *bioRxiv* 2018: 433870.
9. Weisberg YJ, DeYoung CG, Hirsh JB. Gender differences in personality across the ten aspects of the Big Five. Front Psychol. 2011;2:178.
10. Schmitt DP, Realo A, Voracek M, Allik J. Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. J Pers Soc Psychol. 2008;94:168–82.
11. Weissman MM, Bland RC, Canino GJ, et al. Cross-national epidemiology of major depression and bipolar disorder. JAMA. 1996;276:293–9.
12. Eaves LJ, Eysenck HJ, Martin NG. Genes, culture and personality: an empirical approach. San Diego: Academic Press; 1989. 465 pp.
13. Martin NG, Jardine R. Eysenck's contribution to behavior genetics. In: Modgil S, Modgil C, editors. Hans Eysenck: consensus and controversy. Lewes, England: Falmer Press; 1986. p. 13–47.
14. Viken RJ, Rose RJ, Kaprio J, Koskenvuo M. A developmental genetic analysis of adult personality: extraversion and neuroticism from 18 to 59 years of age. J Pers Soc Psychol. 1994;66:722–30.
15. Eaves LJ, Heath AC, Neale MC, Hewitt JK, Martin NG. Sex differences and non-additivity in the effects of genes on personality. Twin Res. 2012;1:131–7.
16. Lake RIE, Eaves LJ, Maes HHM, Heath AC, Martin NG. Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45,850 twins and relatives on two continents. Behav Genet. 2000;30:223–33.
17. Vinkhuyzen AAE, Pedersen NL, Yang J, Lee SH, Magnusson PKE, Iacono WG, et al. Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. Transl Psychiatry. 2012;2:e102.
18. Wray NR, Middeldorp CM, Birley AJ, Gordon SD, Sullivan PF, Visscher PM, et al. Genome wide linkage analysis of multiple measures of neuroticism of two large cohorts from Australia and the Netherlands. Arch Gen Psychiatry. 2008;65:649–58.
19. Calboli FC, Tozzi F, Galwey NW, Antoniades A, Mooser V, Preisig M, et al. A genome-wide association study of neuroticism in a population-based sample. PLoS ONE. 2010;5:e11504.
20. Terracciano A, Sanna S, Uda M, Deiana B, Usala G, Busonero F, et al. Genome-wide association scan for five major dimensions of personality. Mol Psychiatry. 2010;15:647–56.
21. Helena Mangs A, Morris BJ. The Human Pseudoautosomal Region (PAR): origin, function and future. Curr Genom. 2007;8:129–36.
22. Gale CR, Hagenaars SP, Davies G, Hill WD, Liewald DCM, Cullen B et al. Pleiotropy between neuroticism and physical and mental health: findings from 108038 men and women in UK Biobank. Transl Psychiatry 2016;6:e791.
23. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779.
24. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203–9.
25. Eysenck SB, Eysenck HJ, Barrett P. A revised version of the psychoticism scale. Pers Individ Diff. 1985;6:21–9.
26. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nat Genet. 2018;50:906–8.
27. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.
28. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.
29. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comp Biol. 2015;11:e1004219.
30. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–9.
31. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome-partitioning of genetic variation for complex traits using common SNPs. Nat Genet. 2011;43:519–25.
32. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013;42:689–700.
33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R S Soc Ser B. 1995;57:289–300.

35. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 2013;41 (D1):D377–86.

36. Goldberg D. General health questionnaire. Windsor: NFER Publishing Company; 1978.

37. Luciano M, Pujals AM, Marioni RE, Campbell A, Hayward C, MacIntyre DJ, et al. Current versus lifetime depression, APOE variation, and their interaction on cognitive performance in younger and older adults. Psychosom Med. 2015;77:480–92.

38. Zapata I, Serpell JA, Alvarez CE. Genetic mapping of canine fear and aggression. BMC Genom. 2016;17:572.

39. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. PLoS Genet. 2008;4:e1000304.

40. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends Ecol & Evol. 2000;15:496–503.

41. Tillo M, Charoy C, Schwarz Q, Maden CH, Davidson K, Fantin A, et al. 2- and 6-O-sulfated proteoglycans have distinct and complementary roles in cranial axon guidance and motor neuron migration. Development. 2016;143:1907.

42. Habuchi H, Miyake G, Nogami K, Kuroiwa A, Matsuda Y, Kusche-Gullberg M, et al. Biosynthesis of heparan sulphate with diverse structures and functions: two alternatively spliced forms of human heparan sulphate 6-O-sulphotransferase-2 having different expression patterns and properties. Biochem J. 2003;371 (Pt 1):131–42.

43. Anderson DJ, Adolphs R. A framework for studying emotions across phylogeny. Cell. 2014;157:187–200.

44. LeDoux JE. Evolution of human emotion: a view through fear. Prog Brain Res. 2012;195:431–42.

45. Ilska J, Haskell MJ, Blott SC, Sánchez-Molano E, Polgar Z, Lofgren SE, et al. Genetic characterization of dog personality traits. Genetics. 2017;206:1101.

46. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat Genet. 2018;50:920–7.

47. Nagai N, Habuchi H, Sugaya N, Nakamura M, Imamura T, Watanabe H, et al. Involvement of heparan sulfate 6-O-sulfation in the regulation of energy metabolism and the alteration of thyroid hormone levels in male mice. Glycobiology. 2013;23:980–92.

48. Jin H, Gardner RJ, Viswesvaraiah R, Muntoni F, Roberts RG. Two novel members of the interleukin-1 receptor gene family, one deleted in Xp22.1–Xp21.3 mental retardation. Eur J Hum Genet. 2000;8:87.

49. Piton A, Gauthier J, Hamdan FF, Lafrenière RG, Yang Y, Henrion E, et al. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. Mol Psychiatry. 2011;16:867–80.

50. Walker Rosie M, Sussmann Jessika E, Whalley Heather C, Ryan Niamh M, Porteous David J, McIntosh Andrew M, et al. Preliminary assessment of pre-morbid DNA methylation in individuals at high genetic risk of mood disorders. Bipolar Disord. 2016;18:410–22.

51. Jahani-Asl A, Cheng C, Zhang C, Bonni A. Pathogenesis of Börjeson-Forssman-Lehmann Syndrome: insights from PHF6 function. Neurobiol Dis. 2016;96:227–35.

52. Schosser A, Butler AW, Ising M, Perroud N, Uher R, Ng MY, et al. Genomewide association scan of suicidal thoughts and behaviour in major depression. PLoS ONE. 2011;6:e20690.

53. Jamain S, Quach H, Betancur C, Råstam M, Colineaux C, Gillberg IC, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet. 2003;34:27–9.

54. Dean C, Dresbach T. Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function. Trends Neurosci. 2006;29:21–9.

55. Naidoo N, Pawitan Y, Soong R, Cooper DN, Ku C-S. Human genetics and genomics a decade after the release of the draft sequence of the human genome. Hum Genom. 2011;5:577.

56. Hagenaars SP, Hill WD, Harris SE, Ritchie SJ, Davies G, Liewald DC, et al. Genetic prediction of male pattern baldness. PLoS Genet. 2017;13:e1006594.