



# Structural variation and eQTL analysis in two experimental populations of chickens divergently selected for feather-pecking behavior

Clemens Falker-Gieske<sup>1</sup> · Jörn Bennewitz<sup>2</sup> · Jens Tetens<sup>1,3</sup>

Received: 5 September 2022 / Accepted: 19 November 2022 / Published online: 30 November 2022  
© The Author(s) 2022

## Abstract

Feather pecking (FP) is a damaging nonaggressive behavior in laying hens with a heritable component. Its occurrence has been linked to the immune system, the circadian clock, and foraging behavior. Furthermore, dysregulation of miRNA biogenesis, disturbance of the gamma-aminobutyric acid (GABAergic) system, as well as neurodevelopmental deficiencies are currently under debate as factors influencing the propensity for FP behavior. Past studies, which focused on the dissection of the genetic factors involved in FP, relied on single nucleotide polymorphisms (SNPs) and short insertions and deletions < 50 bp (InDels). These variant classes only represent a certain fraction of the genetic variation of an organism. Hence, we reanalyzed whole-genome sequencing data from two experimental populations, which have been divergently selected for FP behavior for over more than 15 generations, performed variant calling for structural variants (SVs) as well as tandem repeats (TRs), and jointly analyzed the data with SNPs and InDels. Genotype imputation and subsequent genome-wide association studies, in combination with expression quantitative trait loci analysis, led to the discovery of multiple variants influencing the GABAergic system. These include a significantly associated TR downstream of the GABA receptor subunit beta-3 (*GABRB3*) gene, two microRNAs targeting several GABA receptor genes, and dystrophin (*DMD*), a direct regulator of GABA receptor clustering. Furthermore, we found the transcription factor *ETVI* to be associated with the differential expression of 23 genes, which points toward a role of *ETVI*, together with *SMAD4* and *KLF14*, in the disturbed neurodevelopment of high-feather pecking chickens.

**Keywords** Feather pecking · GABA · Structural variation · Schizophrenia · Genome-wide association study · eQTL

## Introduction

FP in chickens is a behavioral disorder that severely impacts animal welfare and causes significant economic losses. It has been proposed that FP is obsessive–compulsive-like behavior [3]. In the past, the damage has been controlled by beak trimming, which has now been prohibited in many

European countries. Numerous studies found an involvement of environmental factors such as light intensity, nutrition, stocking density, and lack of foraging (reviewed by [4] and [5]). Furthermore, evidence has been accumulating that the immune system plays a major role in the development of FP behavior [1, 6–8]. The gut microbiota has also been proposed to be involved in FP behavior, but this has been disproven in several studies [9, 10]. Since feather pecking is a complex heritable trait (reviewed by [11]), the dissection of the genetic causes is essential for the development of effective breeding strategies to eradicate the causative alleles. To achieve that goal, chickens were divergently selected for FP behavior over more than 15 generations based on their estimated breeding values for the behavior. Breeding of these lines was initiated in Denmark and continued in Hohenheim, Germany. In Hohenheim, two populations were established—an F<sub>2</sub> cross and 12 half-sib families (in the preceding text referred to as F<sub>2</sub> and HS). A detailed description

✉ Clemens Falker-Gieske  
clemens.falker-gieske@uni-goettingen.de

<sup>1</sup> Department of Animal Sciences, Georg-August-University, Burckhardtweg 2, 37077 Göttingen, Germany

<sup>2</sup> Institute of Animal Science, University of Hohenheim, Garbenstr. 17, 70599 Stuttgart, Germany

<sup>3</sup> Center for Integrated Breeding Research, Georg-August-University, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany

of the experimental populations and the research conducted with them was reviewed by Bennewitz and Tetens [12]. Based on the results that were acquired by whole-genome and transcriptome sequencing with the Hohenheim selection lines, FP appears to be a disorder of the  $\gamma$ -aminobutyric acid (GABAergic) system in conjunction with a disturbance of embryonic neurodevelopment by a lack of leukocytes in the developing brain. Several variants in or in close proximity to GABA receptor genes were identified in genome-wide association studies (GWAS) conducted on medium-density single nucleotide polymorphism (SNP) array and imputed-sequence level genotypes [2, 13, 14]. Furthermore, brain transcriptome analysis of high and low feather peckers (HFP and LFP) before and after light stimulation revealed that HFP responds with very few changes in gene expression in comparison to LFP, with numerous GABA receptor genes upregulated in LFP. Only *GABRB2* (gamma-aminobutyric acid type A receptor subunit beta2) was among differentially expressed genes (DEGs) in HFP, but it was downregulated, instead of upregulated, a pattern that we observed for most DEGs in HFP brains [14]. We attribute this low level of gene expression changes in response to light to a high level of excitation in HFP brains due to the lack of multiple GABA receptors. Since GABA is the major inhibitory neurotransmitter, a lack of its receptors in the brain would lead to a high neuronal excitatory state, which explains the observed hyperactivity and the obsessive–compulsive-like behavior observed in HFP. Since *Dicer1* was among the downregulated DEGs, we assume that miRNA processing is disturbed, as is also the case in schizophrenia patients [15], which in turn leads to low GABA receptor expression levels. In the general comparison of brain transcriptomes between HFP and LFP hens, we observed an enrichment of immune system-related DEGs [1], which we could further pinpoint in an expression quantitative trait loci (eQTL) analysis to a small deletion 652 bp downstream of the *KLF14* gene. In total, the differential expression of 40 genes between HFP and LFP chickens was significantly associated with this *KLF14* variant, a majority of which are involved in leukocyte biology [8]. It has been shown in mice that CD4 T cells are essential for healthy development from the fetal to the adult brain. A defect in CD4 T cell maturation affected synapse development and led to behavioral abnormalities [16]. The evidence suggests that this mechanism is responsible for disturbing embryonic brain development in HFP chickens, which contributes to FP behavior.

One commonality of all the studies that we conducted on the genetics involved in FP is the overlap in associated genes with human psychiatric disorders, most prevalent schizophrenia. SVs play a notable role in human psychiatric disorders [17–19], which is also the case for tandem repeats (TRs) [20]. Commonly investigated classes of SVs include insertions, deletions, duplications, inversions, and

translocations, which arise from various combinations of DNA gains, losses, or rearrangements [21]. Here, we present the first in-depth study on the potential role of SVs and TRs in FP, which led to a deeper understanding of the mechanisms responsible for this behavioral disorder.

## Material and methods

### Animals and husbandry

The F<sub>2</sub> and HS lines of White Leghorn chickens were divergently selected for feather pecking behavior for over more than 15 generations at Hohenheim University [22, 23]. Animals, which were used for genotyping in the study presented here were as follows: from the F<sub>2</sub> design 25 founder (whole-genome sequenced), 89 F<sub>1</sub> (SNP chip genotyped), and 817 F<sub>2</sub> animals (SNP chip genotyped) and from the HS population, 24 animals were whole-genome sequenced and 494 animals were SNP chip genotyped. RNA from the whole brains of 167 HS chickens was used for Fluidigm gene expression analysis for the eGWAS approach [8]. All experimental procedures [1], rearing and husbandry conditions [24], as well as phenotyping [2], were described in previous studies. Briefly, the phenotypic data were generated by direct observations made by seven independent investigators at approximately 32 weeks of age. Observations were recorded in 20 min sessions in average group sizes of about 42 animals. The phenotypes are expressed as the number of FP bouts actively delivered during a standardized time span. As these count data are heavily distributed from normality, Box–Cox-transformation was applied [2, 13].

### Structural variants and short tandem repeats discovery

Illumina whole genome sequencing data were mapped to chicken genome version GRCg6a (GCF\_000002315.5 RefSeq assembly) and used to call SNPs and short (< 50 bp) insertions and deletions (InDels) in our previous study [2]. This was achieved with the genome analysis toolkit (GATK) v. 4.0, according to the best practice guideline of the broad institute [25]. SVs were called as described by Blaj et al. [26] with slightly different settings: A high-confidence SV call set was produced from the output of three variant callers: smove v. 0.2.6 (Brent, P. (2018) Smoove. <https://brentp.github.io/post/smove/>), DELLY v. 0.7.7 [27], and manta v. 1.6.0 [28]. SURVIROR v. 1.0.7 [29] was used to combine the output of the three variant callers with the following settings: maximum distance between breakpoints of 1000 bp, minimum number of supporting callers 2, SV type and strands were taken into account, and the minimum SV size was set to 30 bp. Variants with a call rate < 0.8 and variants

with  $QUAL < 1000$  were removed. TRs were called with GangSTR v. 2.5 [30]. A library of known TRs as input for GangSTR was acquired from the UCSC data repository (<https://hgdownload.soe.ucsc.edu/goldenPath/galGal6/bigZips/>). TRs were filtered to retain genotypes with a minimum sequence depth (DP) of 10, a quality score (Q) higher than 0.8, and a call rate  $< 0.8$ .

### Haplotype construction and imputation

To impute medium-density chip genotypes to whole genome-level SNPs and InDels, we employed the same strategy as we previously described [2], with the deviation that all imputation steps were performed with Beagle v. 5.2 [31] and the setting  $ne = 1000$ . SVs and TRs were merged with SNPs and InDels from our previous study, which were acquired with the GATK [25]. This merged call set was phased with Beagle v. 5.2 for the estimation of haplotypes and used as a reference panel to impute medium-density chip genotypes to SVs and TRs with the same strategy as for SNPs and InDels. Chip genotypes from HS animals were directly imputed using the WGS reference panel. For the  $F_2$  design, we first imputed chip-genotyped  $F_1$  animals to the WGS level, merged the output with the initial reference panel, phased the merged dataset, and used it as a reference panel for the imputation of  $F_2$  SNP chip-genotyped animals. To remove SNPs and InDels from the imputed SVs/TRs, GATK SelectVariants was utilized.

### Detection of quantitative trait loci

Prior to GWAS, multiallelic variants from the SVs/TR dataset were converted to biallelic variants with the norm function from bcftools v. 1.14 [32]. All GWAS were conducted with gcta v. 1.92.3 beta3 [33], applying a mixed linear model association analysis with a leaving-one-chromosome-out (LOCO) approach and a minor allele frequency (MAF) threshold of 0.01. In brief, relatedness between animals and stratification [6] was corrected by including a random genetic term based on a genomic relationship matrix calculated only from SNP-chip data and following a LOCO approach [34]. Briefly, a model of the form  $y = W\alpha + X\beta + u + \epsilon$  is fitted ( $y = n \times 1$  vector of phenotyped ( $n$ ) hens;  $W = n \times c$  incidence matrix of fixed effects with  $c$  being the number of effects;  $\alpha =$  vector of corresponding coefficients including the mean;  $X = n \times 1$  vector of marker genotypes at the locus tested;  $\beta =$  corresponding effect size;  $u =$  vector of random genetic effects, with  $u \sim N(0, A^- \sigma_g^2)$ , where  $\sigma_g^2$  represents genetic variance and  $A^-$  is the genomic relationship matrix based on all SNP-chip markers except those on the chromosome currently analyzed;  $\epsilon =$  random residual term, with  $\epsilon \sim N(0, I \sigma_e^2)$ , where  $\sigma_e^2$  represents the residual variance and  $I$  represents an identity matrix). Line effects (HFP and LFP) and hatch were included

in the analysis of the HS design; for the  $F_2$  design, only the hatch was included as a fixed effect. The phenotype used for GWAS, “feather pecks delivered Box–Cox transformed” (FPD\_BC), has been described by Iffland et al. [13]. Phenotypes for expression GWAS (eGWAS) were normalized gene expression data from our previous study [8]. There we analyzed the expression of 86 genes in 167 HS chickens, which we discovered in a transcriptome analysis performed on 48 of the HS birds [1]. Information on the hatch was used as a covariate in all GWAS and eGWAS. Genomic relationship matrices were created from the target 60 k SNP Chip genotypes. Meta-analyses of GWAS results were performed with METAL v. 1.1 [35] using the sample size-based approach with default settings. The proportion of variance in phenotype explained by a given SNP (PVE) was calculated according to the formula  $\text{Var}(Y) = \beta^2 \text{Var}(X) + \sigma^2$  by Shim et al. [36] ( $\text{Var}(Y) =$  variance in phenotype;  $\beta =$  effect size of genetic variant  $X$ ;  $\sigma^2 =$  remaining variance). To correct for multiple testing, the threshold for genome-wide significance of variants was calculated by Bonferroni correction ( $\frac{\text{number of variants}}{0.05}$ ).

### Association weight matrix construction

The AWM was created as described in our previous study [8] by deploying the strategy for AWM construction by Reverter and Fortes [37], followed by the detection of significant gene–gene interactions with their PCIT algorithm [38]. Input variants were chosen as follows:  $p$ -value  $< 1 \times 10^{-4}$  for the main phenotype (FPD\_BC) or  $p$ -value  $< 1 \times 10^{-4}$  in at least ten of the eGWAS. That way, variants affecting the main phenotype and gene expression were both considered in the analysis. This led to the selection of 57 input variants, 0.16% of all detected SVs and TRs, for the HS population. The R script by Reverter and Fortes was modified by setting the  $p$ -value threshold for primary and secondary SNP selection to  $1 \times 10^{-4}$ . The gene–gene interaction map was constructed with Cytoscape [39], and gene class information was acquired with PANTHER [40] and UniProt [41].

### Transcription factor enrichment

To detect significant binding site enrichment for the transcription factor *ETVI*, the CiiiDER software (build May 15th, 2020) [42] was employed with frequency matrix MA0761.2 (<https://jaspar.genereg.net/matrix/MA0761.2/>) and the following settings:  $p$ -value threshold for gene coverage enrichment = 0.05; base position upstream scan limit = 1500 bp; base position downstream scan limit = 500 bp. Additional members of the ETS (E twenty-six) family of transcription factors used in the analysis were *ETV2* (MA0762.1), *ETV3* (MA0763.1), *ETV4* (MA0764.2), *ETV5* (MA0765.1), *ETV6* (MA0645.1), and *ETV7* (MA1708.1). Genes that were associated with the *ETVI*

variant ( $p$ -value  $< 1 \times 10^{-4}$ ; Supplementary Information S1) were used as input, and DEGs with a  $\log_2$  fold change  $< 0.2$  were used as background genes (14,514 genes).

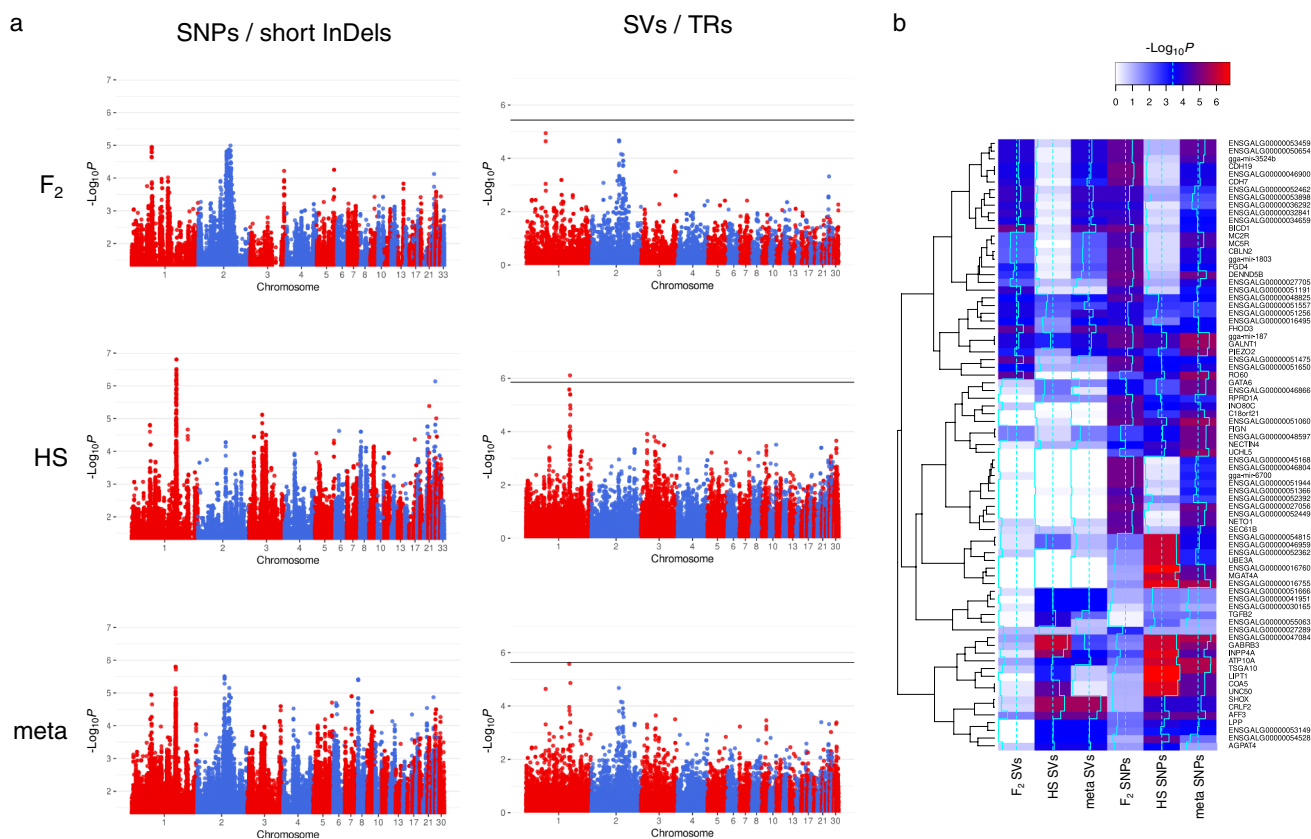
## Results

### Genome-wide association studies with different classes of genetic variants

In total, 63,824 TRs and 11,098 SVs were discovered in the joint variant calling of whole genome sequenced HS and  $F_2$  chickens. The SVs contained 6014 deletions, 2741 inversions, 1334 duplications, and 995 translocations. The variant calling of SNPs and InDels was reported in our previous study and yielded 12,864,421 SNPs and 2,142,539 InDels [2]. To grasp the whole depth of the datasets at hand, we repeated the imputation of SNPs and InDels with the most recent Beagle version (v. 5.2) [31] and set the effective population size to 1000, which is known to improve the imputation accuracy in small populations [43]. GWAS results

for the trait FPD\_BC were analyzed for both experimental designs, the  $F_2$  cross, and the HS population, separately and after combining the results in a meta-analysis. Manhattan plots from GWAS with imputed SNPs/InDels and imputed SVs/TRs are shown in Fig. 1a. Common peaks for both variant classes on GGA1 and GGA2 were observable in the  $F_2$  cross, as well as on GGA1 and GGA3 of the HS population. QTL are summarized in Table 1, and lead SVs and TRs with their predicted effects are listed in Table 2.

The only variant that reached genome-wide significance after Bonferroni correction ( $-p < 1.41 \times 10^{-6}$ ) was a TR in the HS population, 126,821 bp downstream of *GABRB3*. The proportion of variance in the phenotype FPD\_BC that is explained by this variant is 0.047. Given heritability estimates of around 0.15 [22, 44, 45], the amount of phenotypic variance explained by the lead variant is considerable. To visualize overlaps between the six different sets of results, we created a heat map, which shows the  $-\log_{10}p$ -values of the top 20 associated genes from the different GWAS (Fig. 1b) and summarized all  $p$ -values in Supplementary Information S2. Numerous



**Fig. 1 a** Genome-wide association studies (GWAS) performed with single nucleotide polymorphisms (SNPs) in combination with short ( $< 50$  bp) insertions and deletions (InDels) and structural variants (SVs) in combination with tandem repeats (TRs). All variant classes were investigated in the  $F_2$  cross and half-sib (HS) families of chick-

ens divergently selected for feather pecking and subsequently combined in a meta-analysis. **b** Heatmap of the 20 highest associated genes from each GWAS. The  $-\log_{10} p$ -values of the highest associated variants for each gene were used

**Table 1** Positional ranges in base pairs of QTL that were discovered in both experimental populations (F<sub>2</sub> cross and half-sib (HS) families) for both groups of variant classes: single nucleotide polymorphisms

F <sub>2</sub> SNPs and short InDels	F <sub>2</sub> SVs and TRs	HS SNPs and short InDels	HS SVs and TRs
GGA1: 59,032,772–59,548,358	GGA1: 59,138,261–59,216,858	GGA1: 132,578,705–133,803,185	GGA1: 130,621,356–134,077,752
GGA2: 81,060,432–97,072,903	GGA2: 83,493,152–94,489,210	GGA3: 18,434,015–54,601,143	GGA3: 19,120,538–70,385,553

(SNPs) and short insertions/deletions (InDels) as well as structural variants (SVs) and tandem repeats (TRs)

**Table 2** Structural variants (inversions (INV) and deletions (DEL)) and tandem repeats (TRs) that showed the highest association in genome wide association studies for feather pecking behavior in an

F<sub>2</sub> cross and half-sib (HS) families. Variant effects and closest genes were predicted with SnpEff

GWAS	Type	Size (bp)	Position	p-value	Effect	Gene(s)
F <sub>2</sub>	INV	1403	GGA1: 59,216,858–59,218,260	1.14 × 10 <sup>-5</sup>	intron_variant	<i>BICD1</i>
F <sub>2</sub>	TR		GGA1: 59,138,261	2.29 × 10 <sup>-5</sup>	intron_variant	<i>BICD1</i>
F <sub>2</sub>	TR		GGA2: 83,217,700	2.11 × 10 <sup>-5</sup>	intron_variant	<i>FHOD3</i>
F <sub>2</sub>	DEL	509	GGA2: 83,493,152–83,493,661	2.11 × 10 <sup>-5</sup>	intergenic_region	<i>ENSGALG00000051475-RPRD1A</i>
F <sub>2</sub>	TR		GGA2: 83,497,153	2.30 × 10 <sup>-5</sup>	intergenic_region	<i>ENSGALG00000051475-RPRD1A</i>
F <sub>2</sub>	DEL	209	GGA2: 83,497,939–83,498,148	2.30 × 10 <sup>-5</sup>	intergenic_region	<i>ENSGALG00000051475-RPRD1A</i>
F <sub>2</sub>	TR		GGA2: 89,486,485	6.90 × 10 <sup>-5</sup>	intergenic_region	<i>ENSGALG00000053898-ENSGALG00000052462</i>
F <sub>2</sub>	TR		GGA2: 94,489,210	7.28 × 10 <sup>-5</sup>	intergenic_region	<i>ENSGALG00000051191-ENSGALG00000032841</i>
HS	TR		GGA1: 132,889,302	7.79 × 10 <sup>-7</sup>	intergenic_region	<i>GABRB3-ENSGALG00000047084</i>
HS	TR		GGA1: 130,621,356	2.64 × 10 <sup>-6</sup>	intergenic_region	<i>CRLF2-SHOX</i>
HS	TR		GGA1: 133,533,130	4.00 × 10 <sup>-6</sup>	intron_variant	<i>INPP4A</i>
HS	TR		GGA1: 133,533,741	4.09 × 10 <sup>-6</sup>	intron_variant	<i>INPP4A</i>
HS	DEL	368	GGA1: 134,077,752–134,077,908	1.37 × 10 <sup>-5</sup>	intron_variant	<i>AFF3</i>
HS	TR		GGA1: 133,606,912	2.30 × 10 <sup>-5</sup>	downstream_gene_variant	<i>COA5</i>
HS	TR		GGA1: 132,889,302	6.42 × 10 <sup>-5</sup>	intergenic_region	<i>GABRB3-ENSGALG00000047084</i>
HS	TR		GGA1: 134,020,467	9.02 × 10 <sup>-5</sup>	intron_variant	<i>AFF3</i>
HS	TR		GGA3: 19,445,704	0.00012	downstream_gene_variant	<i>TGFB2</i>
HS	TR		GGA3: 40,921,527	0.00016	intron_variant	<i>DLL1</i>
HS	TR		GGA3: 19,264,158	0.00022	intergenic_region	<i>LYPLAL1-ENSGALG00000051126</i>
HS	TR		GGA3: 45,338,399	0.00023	intron_variant	<i>AGPAT4</i>
HS	DEL	531	GGA3: 54,243,436–54,244,326	0.00026	intron_variant	<i>REPS1</i>
HS	TR		GGA3: 70,359,487	0.00032	intron_variant	<i>GRIK2</i>
HS	TR		GGA3: 70,385,553	0.00032	intron_variant	<i>GRIK2</i>
HS	TR		GGA3: 19,120,538	0.00049	intergenic_region	<i>ENSGALG00000047985-LYPLAL1</i>

genes showed high association signals ( $p < 0.01$ ) in the meta-analyses for both variant classes: *AFF3*, *ATP10A*, *BICD1*, *CDH19*, *CDH7*, *CRLF2*, *ENSGALG00000016495* (*KLHL29*), *ENSGALG00000032841* (lncRNA), *ENSGALG00000034659* (miRNA), *ENSGALG00000046900* (lncRNA), *ENSGALG00000048825* (lncRNA), *ENSGALG00000050654* (lncRNA), *ENSGALG00000053149* (miRNA), *ENSGALG00000053459* (lncRNA), *FHOD3*, *GALNT1*, *gga-mir-187*, *gga-mir-3524b*, *INPP4A*, *LPP*, *PIEZO2*, and *SHOX*. We assume that these genes are burdened with multiple mutations that contribute to the phenotype, which might be a result of the divergent

selection for feather-pecking behavior over multiple generations. Predicted targets of the microRNA (miRNA) *gga-mir-187* are *GABRA1*, *GABRB2*, *GABRB3*, *GABRG1*, and *GABRG2*. *GABRB2* is also a predicted target of *gga-mir-3524b* ([www.targetscan.org](http://www.targetscan.org), accessed January 27, 2022).

### Expression quantitative trait loci (eQTL) analyses

To clarify whether the SVs and TRs that we detected influence the expression of transcripts that we identified in a previous study to be differentially expressed between LFP and HFP [1], we performed an eQTL analysis. We employed the

same strategy as in our past study [8] and performed eGWAS for 86 genes from 167 chickens (84 HFP and 83 LFP) from the HS population (Manhattan plots are summarized in Supplementary Information S3). A total of 35,571 SVs and TRs were screened for association with gene expression, and we detected 909 genome-wide significant associated signals. SVs and TRs for which we detected genome-wide association with at least 10 DEGs are shown in Table 3. To identify significant gene–gene interactions from those 86 eGWAS, we constructed an association weight matrix [37], followed by the detection of significant correlations with the PCIT algorithm [38]. Central to the gene–gene interaction map is the transcription factor *ETVI*. We selected DEGs with  $p$ -values  $< 1 \times 10^{-4}$  for association with the variant, a tandem repeat with the sequence CCCGGCCCG 70 bp upstream of *ETVI* (GGA2: 27,337,541). This led to the selection of 23 genes that were associated with this variant, with  $p$ -values ranging from  $1.41 \times 10^{-6}$  to  $9.89 \times 10^{-5}$  with the top associated DEG *CERS4L* almost reaching genome-wide significance ( $p$ -value =  $1.41 \times 10^{-6}$ ). With the 23 selected genes (Supplementary Information S1), we performed a transcription factor binding site enrichment analysis with Ciiider and found a significant enrichment ( $p$ -value = 0.013,  $\log_2$ -enrichment = 0.494) of *ETVI* binding sites (Fig. 3a) in proximity to those genes (Fig. 3b, Supplementary Information S4). To demonstrate specificity, we included all available PWMs for members of the ETS family of transcription factors: *ETV2–ETV6*. The only other member of the TF family for which transcription factor binding site enrichment was detected was *ETV4*, but the  $p$ -value was not significant. The complete results of the analysis are summarized in Supplementary Information S5. Another noteworthy gene, which

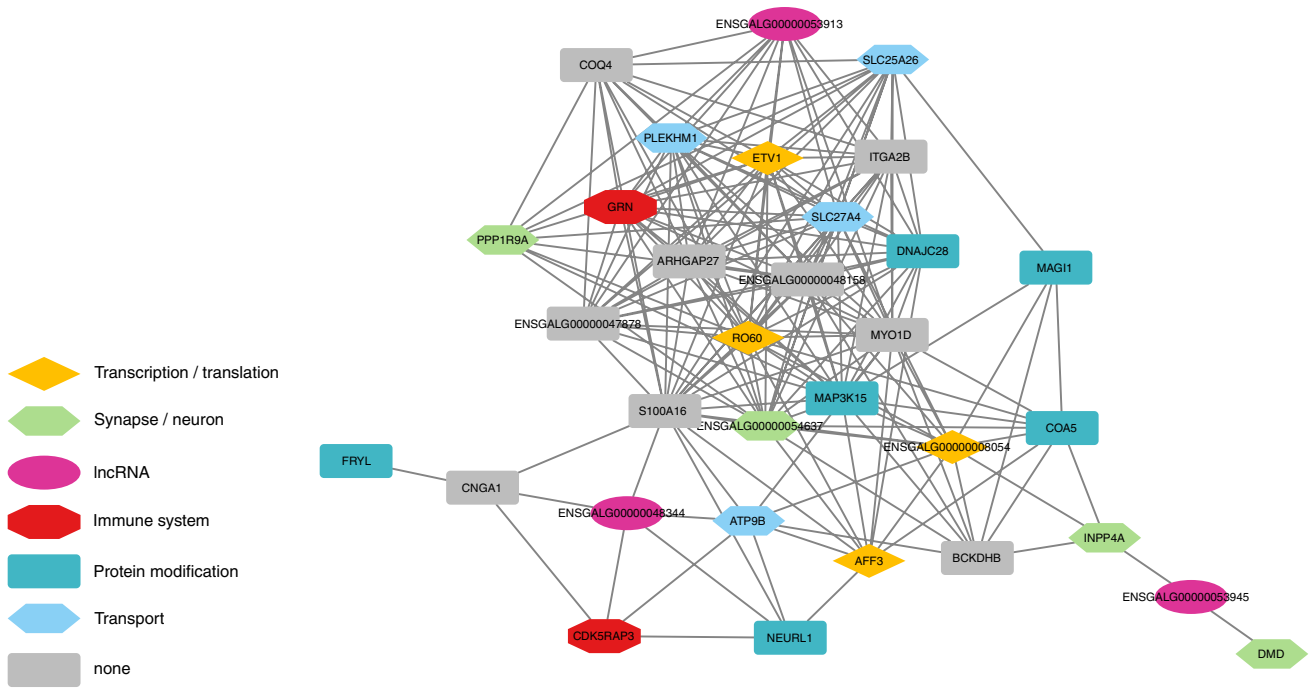
we discovered in an eQTL study with SNPs and InDels [8], is *dystrophin (DMD)*. The variant at position GGA1: 116,965,973, which is a biallelic intronic TR between exons 7 and 8, was associated with genome-wide significance to the DEGs *LOC112531493* ( $p$ -value =  $1.36 \times 10^{-8}$ ), *LOC422393* ( $p$ -value =  $5.57 \times 10^{-6}$ ), *RASSF8* ( $p$ -value =  $5.76 \times 10^{-6}$ ), and *LOC112533169* ( $p$ -value =  $8.48 \times 10^{-6}$ ). The 2 bp *DMD* intron deletion between exons 1 and 2 we discovered in our previous study (rs735635304, GGA1: 117,179,438) was associated with genome-wide significance to the DEGs *LOC112532977* ( $p$ -value =  $7.34 \times 10^{-10}$ ) and *LOC107049114* ( $p$ -value =  $1.60 \times 10^{-9}$ ). The two *DMD* variants are 213 kb apart and not in linkage disequilibrium ( $R^2 = 0.022868$ ).

## Discussion

Numerous studies have been conducted to unravel the genetics behind FP behavior in chickens, as reviewed in [12, 46], but none of these focused on the analysis of structural genetic variation. SVs contribute to complex traits to a higher degree than SNPs [21] and have been the focus of studies on neuropsychiatric disorders in recent years [17–19]. Here, we combined data from two well-described experimental crosses, an  $F_2$  design and a HS population, and performed multiple GWAS and eQTL analyses on different classes of genetic variants. By applying conventional GWAS approaches on the two experimental populations with two sets of variant classes (SNPs and InDels as well as SVs and TRs) followed by meta-analysis, we identified strong associations with numerous putative candidate genes

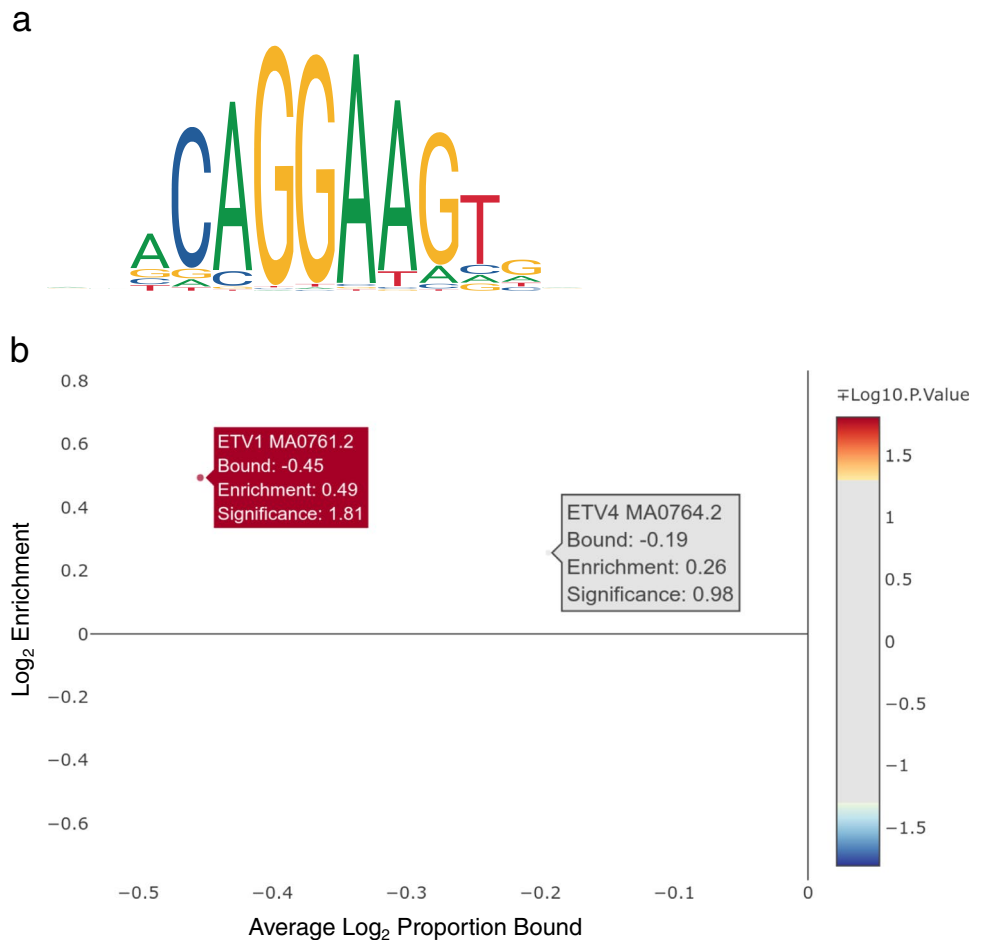
**Table 3** Large deletions (DEL) and tandem repeats (TR) that showed genome-wide significant association ( $p$ -values  $< 1.41 \times 10^{-6}$ ) in at least 10 expression genome-wide association studies for genes that are differentially expressed in brains of high and low feather pecking chickens

Type	Size (bp)	Position	Effect	Gene(s)	No. of associations
TR		GGA25: 1,465,942	intergenic_region	<i>ENSGALG00000055092–IQGAP3</i>	53
TR		GGA25: 3,317,155	intron_variant	<i>S100A16</i>	47
TR		GGA17: 6,359,797	intergenic_region	<i>ENSGALG00000035908–ENSGALG00000047338</i>	39
DEL	294	GGA8: 3,889,333–3,889,627	intron_variant	<i>RO60</i>	30
TR		GGA2: 56,803,419	intron_variant	<i>ATP9B</i>	22
TR		GGA2: 56,944,739	intergenic_region	<i>SALL3–ENSGALG00000054175</i>	22
TR		GGA17: 5,355,006	upstream_gene_variant	<i>SLC27A4</i>	15
TR		GGA27: 3,942,783	intron_variant	<i>GPATCH8</i>	11
TR		GGA27: 3,976,916	frameshift_variant	<i>GPATCH8</i>	11
TR		GGA27: 3,977,142	intron_variant	<i>GRN</i>	11
TR		GGA27: 3,988,533	intergenic_region	<i>TMEM98–ARHGAP27</i>	11
TR		GGA27: 4,009,079	intron_variant	<i>PLEKHM1</i>	11
TR		GGA27: 4,022,832	intron_variant	<i>ENSGALG00000047878</i>	11



**Fig. 2** Gene–gene interaction map of significant correlations that were identified from an association weight matrix on expression genome-wide association studies with the PCIT algorithm

**Fig. 3 a** Graphical representation of the binding motif frequency matrix of the *ETV1* transcription factor. **b** Result of transcription factor binding site enrichment analysis for *ETV1* with genes from expression genome-wide associations studies with  $p$ -values  $< 1 \times 10^{-4}$



for both variant classes. GALNT1 catalyzes glycosylation of target proteins, and aberrant glycosylation of i.e. the GABA<sub>A</sub> receptor is a pathological hallmark in postmortem brains of patients suffering from schizophrenia [47]. FHOD3 controls dendritic spine morphology [48] and might contribute to FP behavior by causing aberrant neuronal development in the cerebral cortex. Similarly, *KLHL29* has already been discussed to be involved in a neurodevelopmental disorder [49]. *AFF3* is also a new promising candidate gene that was previously associated with intellectual disability and cellular migration in the cerebral cortex [50, 51]. Regarding its cellular function, *PIEZO2* is of considerable interest. It is a mechanically activated cation channel, required for light/touch sensation and proprioception, and is abundantly expressed in dorsal root ganglion and sensory endings of proprioceptors in mice [52]. Since one of the major triggers of FP behavior is light stimulation [14, 53], genetic variation in *PIEZO2* might lead to disturbed light perception in HFP.

However, the lead variant from these analyses is a TR 126 kb downstream of the *GABRB3* gene, which explains a considerable amount of the observed phenotypic variance. But, since this variant was detected with a sample size of about 500 animals, its contribution to the phenotype should be interpreted with caution. According to the Beavis effect, the results of QTL studies with sample sizes of around 500 individuals are slightly overestimated [54]. Furthermore, miRNA *gga-mir-187* is strongly associated with FP behavior in all conducted GWAS (Fig. 1b), and among predicted targets of this regulatory RNA are several GABA receptor genes. Based on whole-brain transcriptome analyses of the light response of HFP, we previously postulated that downregulation or missing upregulation of GABA receptor expression is caused by miRNA dysregulation due to low levels of *Dicer1* expression in HFP brains [14]. The findings presented here provide further evidence for a disturbance in miRNA regulation and involvement of the GABAergic system in FP behavior. Since these chickens have been selected for FP behavior based on their estimated breeding values for multiple generations, it is possible that mutations leading to low expression of several GABA receptors have been accumulating. Previous studies with the same experimental population already pointed in that direction [2, 13]. But these are not only limited to GABA receptor genes, miRNAs, or genes encoding miRNA processing proteins. By conducting an eQTL analysis with SVs and TRs, we were able to confirm the involvement of *DMD* (dystrophin) in the regulation of genes that are differentially expressed between HFP and LFP. We already discovered *DMD* in an eQTL analysis with SNPs and InDels [8]. Since dystrophin is a direct regulator of GABA receptor clustering [55], this adds another layer to the GABA receptor disturbance in HFP brains. Several other genes that appear in the gene–gene interaction map based on significant AWM correlations (Fig. 2) have been connected to the GABAergic

system. These include *COQ4* [56], *ETVI* [57], *NEURL1* [58], *SLC25A26*, and *SLC27A4* [59]. In this regard, *ETVI* is of considerable interest since it is the only transcription factor that we identified in the AWM analysis with SVs and TRs. According to the Human Protein Atlas, *ETVI* expression is specific to salivary glands and the brain, specifically the cerebellum and thalamus (<https://www.proteinatlas.org/ENSG0000006468-ETV1>; accessed August 2022). *ETVI* is associated with the differential expression of 23 genes between HFP and LFP, and *ETVI* binding sites are enriched in proximity to these genes (Fig. 3). Although genome-wide significance was closely missed for these associations, we strongly believe that the AWM approach by Reverter and Fortes [37] led to the discovery of a high-confidence set of variants for this complex trait. Among these putative *ETV1* targets are *CSF2RB*, which has been associated with major depression and schizophrenia [60], and *MIS18BP1*, a gene implicated in the autism spectrum [61]. Furthermore, an *ETVI* antibody coimmunoprecipitated the GABA<sub>A</sub> receptor  $\alpha 6$  (GABAAR $\alpha 6$ ) promoter region in mice [57], and *ETVI* is a regulator of gene expression in CD4 and CD8 T cells [62]. Hence, *ETVI* may not only impact GABA receptor expression but might also participate in neurodevelopment. As demonstrated by Pasciuto *et al.*, a lack of CD4 T cells in the brain of mice leads to excess immature neuronal synapses and behavioral abnormalities. Among the top DEGs identified by single-cell RNAseq were the transcription factors *Klf2* and *Klf4* [16]. We previously postulated that FP behavior is the result of disturbances in embryonic neurodevelopment and identified *KLF14* as a major regulator in this regard [8]. In that study, we also identified a Smad4 domain-containing transcription factor, namely *ENSGALG00000042129* (*CHTOP*), but we were not able to detect the enrichment of binding sites for that transcription factor in proximity to its associated DEGs. However, in light of the fact that *ETV1* forms functional complexes with Smad4 [63], we propose a model in which the transcription factors *ETV1*, *CHTOP*, and *KLF14* have an additive effect on the density of T cells in the developing brain of HFP. This might lead to structural abnormalities in the brains of HFP and may explain, at least in part, their abnormal behavior. Multiple studies point toward a central role of Krüppel-Like factors in neurodevelopment and behavior [64–68]. However, conclusive results on the function of Krüppel-like factors in the neurodevelopment in chickens have not been reported yet, and research is hampered by erroneous annotation of KLF orthologues [69].

Apart from *DMD*, we also discovered the genes *PPP1R9A*, *INPP4A*, and *COA5* in both eQTL-AWM analyses, one of which was performed with SVs and TRs (Fig. 2) and one with SNPs and InDels [8]. In schizophrenia and bipolar disorder, the prefrontal cortical expression of *PPP1R9A* was altered [70] and its gene product Neurabin regulates anxiety-like behavior in adult mice [71]. *INPP4A* has been implicated in multiple



neurological conditions, namely schizophrenia, autism, epilepsy, and intellectual disability [72–74]. An additional link to schizophrenia is a frameshift variant in the *GPATCH8* gene in exon 9 of transcript GPATCH8-201 (Table 3), an ortholog of *ZNF804A*, which has been shown to impact various mental illnesses via pre-mRNA processing [75]. Multiple TRs and one large deletion in *RO60* showed genome-wide association with at least 30 DEGs between HFP and LFP (Table 3). With 53 genome-wide associations, an intergenic TR seemed to have an impact on more than half of the top DEGs between HFP and LFP whole brains. The variant is located 2.7 kb downstream of the E3 ubiquitin ligase *ENSGALG00000055092 (DCAF11)* and 7.5 kb upstream of *IQGAP3*, which is required for proper cell cycle progression [76] and which influences immune cell infiltration and immune modulators [77]. A 294-bp intron

deletion in *RO60*, a gene involved in the regulation of inflammatory gene expression [78], was associated with 30 DEGs, which might be responsible for the observation that 48.1% of DEGs between HFP and LFP belong to the PANTHER protein class defense/immunity (PC00090) [1]. A TR within an *S100A16* intron yielded 47 significantly associated DEGs, which is also a gene related to the immune system [79]. An intergenic TR with 39 genome-wide associated DEGs was located between the two lncRNAs *ENSGALG00000035908* and *ENSGALG00000047338*. They probably belong to the class of psychiatric ncRNAs, a term scored by Gandal *et al.* [80].

We discovered multiple links between FP and human psychiatric disorders in our previous studies [1, 2, 14]. The fact that methylation of *KLF14* correlates with psychosis severity

**Table 4** Summary of published data related to neurological disorders on the 20 highest associated genes from each genome wide association study on feather pecking behavior in laying hens

Gene	Gene product	Published data related to neurological disorders
<i>AFF3</i>	AF4/FMR2 family member 3	Association with intellectual disability [50, 51]
<i>AGPAT4</i>	1-acyl-sn-glycerol-3-phosphate acyltransferase delta	Age-dependent increase of anxiety in knockout mice [84] and association with the major depressive disorder [85]
<i>ATP10A</i>	Phospholipid-transporting ATPase VA	Autism susceptibility locus [86]
<i>CBLN2</i>	Cerebellin-2	Regulator of compulsive behavior [87]
<i>CDH7</i>	Cadherin-7	Associated with the major depressive disorder [88]
<i>DENND5B</i>	DENN domain-containing protein 5B	Altered expression in patients with epilepsy and regulation of seizures in mice [89]
<i>FGD4</i>	FYVE, RhoGEF and PH domain-containing protein 4	Involved in clustering and trafficking of GABA receptors [90]
<i>FHOD3</i>	FH1/FH2 domain-containing protein 3	Controls the dendritic spine morphology [48]
<i>GABRB3</i>	Gamma-aminobutyric acid receptor subunit beta-3	Associated with FP in chickens [2, 14] and schizophrenia [91]
<i>INO80C</i>	INO80 complex subunit C	Mediator of cocaine addiction [92]
<i>INPP4A</i>	Inositol polyphosphate-4-phosphatase type I A	Decreased expression in epilepsy [73] and a nonsense mutation in intellectual disability [72]
<i>KIAA1211L</i> ( <i>ENSGALG00000016755</i> )	CRACD-like protein	Associated with depression, bipolar disorder, schizophrenia [93], and opioid use [94]
<i>LIPT1</i>	Lipoyltransferase 1	Involved in the development of Leigh syndrome [95]
<i>MC2R</i>	Adrenocorticotrophic hormone receptor	Risk factor for schizophrenia [96]
<i>MGAT4A</i>	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	Abnormal expression in the prefrontal cortex in schizophrenia [97]
<i>MRPL30</i>	39S ribosomal protein L30	Associated with the major depressive disorder [85]
<i>NECTIN4</i>	Nectin-4	Associated with opioid abuse [98]
<i>NETO1</i>	Neuropilin and tolloid-like protein 1	NMDA receptor-interacting protein is required for synaptic plasticity and learning [99]
<i>SHOX</i>	Short stature homeobox protein	Autism locus [100]
<i>TGFB2</i>	Transforming growth factor beta-2 proprotein	Involved in signaling toward the age of onset and cognitive functioning in schizophrenia [101]
<i>TSGA10</i>	Testis-specific gene 10 protein	miRNA target in the development of schizophrenia [102]
<i>UBE3A</i>	Ubiquitin-protein ligase E3A	Autism susceptibility locus [86] [103]
<i>UNC50</i>	Protein unc-50 homolog	Associated with bipolar disorder [104] and involvement in the cell-surface expression of neuronal nicotinic receptors [105]

in schizophrenia patients [81] strengthens our argument for the applicability of HFP chickens in the research of the basic mechanisms involved in human psychiatric disorders. Furthermore, by conducting multiple GWAS on two experimental populations of FP chickens with different variant classes with whole-genome marker density, we identified numerous genes that have previously been linked to psychiatric disorders or other neurological conditions or phenotypes (Table 4). These genes coherently fit the mechanistic hypotheses raised so far and, at the same time, underpin the complex nature of this trait. Among these are anxiety, depression, intellectual disability, autism, compulsive behavior, drug addiction, bipolar disorder, and schizophrenia. This raises the question of what common mechanisms these conditions share, which is a current matter of debate. Blokhin *et al.* proposed that future treatment of human psychiatric disorders should be tailored under the use of a genome-wide “targetome” [82], and Johnsson *et al.* discovered multiple genetic overlaps between anxiety behavior in chickens and numerous human psychiatric disorders [83]. The strong genetic burden of HFP chickens with mutations affecting neuropsychiatric genes makes this line of chickens a valid model system to study the effects of tailored drug treatment strategies.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10048-022-00705-5>.

**Acknowledgements** We acknowledge support from the Open Access Publication Funds of Göttingen University. We further thank the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) for the access to their high-performance computing cluster.

**Author contribution** Conceptualization: Clemens Falker-Gieske and Jens Tetens; methodology: Clemens Falker-Gieske and Jens Tetens; formal analysis and investigation: Clemens Falker-Gieske; writing—original draft preparation: Clemens Falker-Gieske; writing—review and editing: Jens Tetens and Jörn Bennewitz; funding acquisition: Jens Tetens and Jörn Bennewitz; resources: Jens Tetens and Jörn Bennewitz; supervision: Jens Tetens.

**Funding** Open Access funding enabled and organized by Projekt DEAL. The study was funded by the German Research Foundation (DFG) under file numbers TE622/4–2 and BE3703/8–2. The funders had no role in study design, data collection and analysis, and interpretation of data and in writing the manuscript. The publication fee was covered by the Open Access Publication Funds of Göttingen University.

**Data availability** All methods applied here have been outlined in previous studies [1, 2, 8]. The raw RNA sequencing data have been deposited at the NCBI Sequence Read Archive (BioProject ID PRJNA656654) and the raw whole genome sequencing data as well (BioProject ID PRJNA664592).

## Declarations

**Ethics approval** The research protocol was approved by the German Ethical Commission of Animal Welfare of the Provincial Government of Baden-Wuerttemberg, Germany (code: HOH 35/15 PG, date of approval: April 25, 2017).

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Falker-Gieske C, Mott A, Preuß S et al (2020) Analysis of the brain transcriptome in lines of laying hens divergently selected for feather pecking. *BMC Genomics* 21. <https://doi.org/10.1186/s12864-020-07002-1>
2. Falker-Gieske C, Iffland H, Preuß S et al (2020) Meta-analysis of genome wide association studies in lines of laying hens divergently selected for feather pecking using imputed sequence level genotypes. *BMC Genet* 21. <https://doi.org/10.1186/s12863-020-00920-9>
3. Brunberg E, Jensen P, Isaksson A et al (2011) Feather pecking behavior in laying hens: hypothalamic gene expression in birds performing and receiving pecks. *Poult Sci* 90:1145–1152. <https://doi.org/10.3382/ps.2010-00961>
4. Nicol CJ, Bestman M, Gilani A-M et al (2013) The prevention and control of feather pecking: application to commercial systems. *World's Poult Sci J* 69:775–788. <https://doi.org/10.1017/S0043933913000809>
5. Rodenburg TB, van Krimpen MM, de Jong IC et al (2013) The prevention and control of feather pecking in laying hens: identifying the underlying principles. *World's Poult Sci J* 69:361–374. <https://doi.org/10.1017/S0043933913000354>
6. Parmentier HK, Rodenburg TB, de Vries Reilingh G et al (2009) Does enhancement of specific immune responses predispose laying hens for feather pecking? *Poult Sci* 88:536–542. <https://doi.org/10.3382/ps.2008-00424>
7. van der Eijk JAJ, Verwoolde MB, de Vries Reilingh G et al (2019) Chicken lines divergently selected on feather pecking differ in immune characteristics. *Physiol Behav* 212:112680. <https://doi.org/10.1016/j.physbeh.2019.112680>
8. Mott AC, Mott A, Preuß S et al (2022) eQTL analysis of laying hens divergently selected for feather pecking identifies KLF14 as a potential key regulator for this behavioral disorder. *Front Genet* 13. <https://doi.org/10.3389/fgene.2022.969752>
9. van der Eijk JAJ, de Vries H, Kjaer JB et al (2019) Differences in gut microbiota composition of laying hen lines divergently selected on feather pecking. *Poult Sci* 98:7009–7021. <https://doi.org/10.3382/ps/pez336>
10. Borda-Molina D, Iffland H, Schmid M et al (2021) Gut microbial composition and predicted functions are not associated with feather pecking and antagonistic behavior in laying hens. *Life (Basel)* 11. <https://doi.org/10.3390/life11030235>

11. Wysocki M, Bessei W, Kjaer JB et al (2010) Genetic and physiological factors influencing feather pecking in chickens. *World's Poult Sci J* 66:659–672. <https://doi.org/10.1017/S0043933910000644>
12. Bennewitz J, Tetens J (2021) Feather pecking behavior in the Hohenheim selection lines of laying hens – a selective review of genomic, transcriptomic and gut microbiota analyses. *Züchtungskunde* 93:229–243
13. Iffland H, Wellmann R, Preuß S et al (2020) A novel model to explain extreme feather pecking behavior in laying hens. *Behav Genet* 50:41–50. <https://doi.org/10.1007/s10519-019-09971-w>
14. Falker-Gieske C, Bennewitz J, Tetens J (2021) The light response in chickens divergently selected for feather pecking behavior reveals mechanistic insights towards psychiatric disorders. *Mol Biol Rep.* <https://doi.org/10.1007/s11033-021-07111-4>
15. Rey R, Suaud-Chagny M-F, Dorey J-M et al (2020) Widespread transcriptional disruption of the microRNA biogenesis machinery in brain and peripheral tissues of individuals with schizophrenia. *Transl Psychiatry* 10. <https://doi.org/10.1038/s41398-020-01052-5>
16. Pasciuto E, Burton OT, Roca CP et al (2020) Microglia require CD4 T cells to complete the fetal-to-adult transition. *Cell* 182:625–640.e24. <https://doi.org/10.1016/j.cell.2020.06.026>
17. Mulle JG (2008) Genomic structural variation and schizophrenia. *Curr Psychiatry Rep* 10:171–177. <https://doi.org/10.1007/s11920-008-0029-y>
18. Rucker JJH, McGuffin P (2012) Genomic structural variation in psychiatric disorders. *Dev Psychopathol* 24:1335–1344. <https://doi.org/10.1017/S0954579412000740>
19. Sullivan PF, Geschwind DH (2019) Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell* 177:162–183. <https://doi.org/10.1016/j.cell.2019.01.015>
20. Xiao X, Zhang C-Y, Zhang Z et al (2021) Revisiting tandem repeats in psychiatric disorders from perspectives of genetics, physiology, and brain evolution. *Mol Psychiatry.* <https://doi.org/10.1038/s41380-021-01329-1>
21. Ho SS, Urban AE, Mills RE (2020) Structural variation in the sequencing era. *Nat Rev Genet* 21:171–189. <https://doi.org/10.1038/s41576-019-0180-9>
22. Grams V, Wellmann R, Preuß S et al (2015) Genetic parameters and signatures of selection in two divergent laying hen lines selected for feather pecking behaviour. *Genet Sel Evol* 47:77. <https://doi.org/10.1186/s12711-015-0154-0>
23. Kjaer JB, Sørensen P, Su G (2001) Divergent selection on feather pecking behaviour in laying hens (*Gallus gallus domesticus*). *Appl Anim Behav Sci* 71:229–239. [https://doi.org/10.1016/s0168-1591\(00\)00184-2](https://doi.org/10.1016/s0168-1591(00)00184-2)
24. Bennewitz J, Bögelein S, Stratz P et al (2014) Genetic parameters for feather pecking and aggressive behavior in a large F2-cross of laying hens using generalized linear mixed models. *Poult Sci* 93:810–817. <https://doi.org/10.3382/ps.2013-03638>
25. McKenna A, Hanna M, Banks E et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>
26. Blaj I, Tetens J, Bennewitz J et al (2022) Structural variants and tandem repeats in the founder individuals of four F2 pig crosses and implications to F2 GWAS results. *BMC Genomics* 23. <https://doi.org/10.1186/s12864-022-08716-0>
27. Rausch T, Zichner T, Schlattl A et al (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28:i333–i339. <https://doi.org/10.1093/bioinformatics/bts378>
28. Chen X, Schulz-Trieglaff O, Shaw R et al (2016) Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32:1220–1222. <https://doi.org/10.1093/bioinformatics/btv710>
29. Jeffares DC, Jolly C, Hoti M et al (2017) Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* 8:14061. <https://doi.org/10.1038/ncomms14061>
30. Mousavi N, Shleizer-Burko S, Yanicky R et al (2019) Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res* 47:e90. <https://doi.org/10.1093/nar/gkz501>
31. Browning BL, Zhou Y, Browning SR (2018) A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet* 103:338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
32. Danecek P, Bonfield JK, Liddle J et al (2021) Twelve years of SAMtools and BCFtools. *Gigascience* 10. <https://doi.org/10.1093/gigascience/giab008>
33. Yang J, Lee SH, Goddard ME et al (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88:76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
34. Falker-Gieske C, Blaj I, Preuß S et al (2019) GWAS for meat and carcass traits using imputed sequence level genotypes in pooled F2-designs in pigs. G3 (Bethesda) 9:2823–2834. <https://doi.org/10.1534/g3.119.400452>
35. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190–2191. <https://doi.org/10.1093/bioinformatics/btq340>
36. Shim H, Chasman DI, Smith JD et al (2015) A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS One* 10:e0120758. <https://doi.org/10.1371/journal.pone.0120758>
37. Reverter A, Fortes MRS (2013) Association weight matrix: a network-based approach towards functional genome-wide association studies. *Methods Mol Biol* 1019:437–447. [https://doi.org/10.1007/978-1-62703-447-0\\_20](https://doi.org/10.1007/978-1-62703-447-0_20)
38. Reverter A, Chan EKF (2008) Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24:2491–2497. <https://doi.org/10.1093/bioinformatics/btn482>
39. Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
40. Mi H, Huang X, Muruganujan A et al (2017) PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 45:D183–D189. <https://doi.org/10.1093/nar/gkw1138>
41. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
42. Gearing LJ, Cumming HE, Chapman R et al (2019) CiiiDER: a tool for predicting and analysing transcription factor binding sites. *PLoS One* 14:e0215495. <https://doi.org/10.1371/journal.pone.0215495>
43. Pook T, Mayer M, Geibel J et al (2020) Improving imputation quality in BEAGLE for Crop and Livestock Data. G3 (Bethesda) 10:177–188. <https://doi.org/10.1534/g3.119.400798>
44. Rodenburg TB, Buitenhuis AJ, Ask B et al (2003) Heritability of feather pecking and open-field response of laying hens at two different ages. *Poult Sci* 82:861–867. <https://doi.org/10.1093/ps/82.6.861>
45. Kjaer JB, Sørensen P (1997) Feather pecking behaviour in White Leghorns, a genetic study. *Br Poult Sci* 38:333–341. <https://doi.org/10.1080/00071669708417999>
46. Wysocki M, Preuss S, Stratz P et al (2013) Investigating gene expression differences in two chicken groups with variable propensity to feather pecking. *Anim Genet* 44:773–777. <https://doi.org/10.1111/age.12050>
47. Williams SE, Mealer RG, Scolnick EM et al (2020) Aberrant glycosylation in schizophrenia: a review of 25 years of post-mortem brain studies. *Mol Psychiatry* 25:3198–3207. <https://doi.org/10.1038/s41380-020-0761-1>

48. Sulistomo HW, Nemoto T, Kage Y et al (2021) Fhod3 controls the dendritic spine morphology of specific subpopulations of pyramidal neurons in the mouse cerebral cortex. *Cereb Cortex* 31:2205–2219. <https://doi.org/10.1093/cercor/bhaa355>
49. Lu G, Ma L, Xu P et al (2022) A de novo ZMIZ1 pathogenic variant for neurodevelopmental disorder with dysmorphic facies and distal skeletal anomalies. *Front Genet* 13:840577. <https://doi.org/10.3389/fgene.2022.840577>
50. Moore JM, Oliver PL, Finelli MJ et al (2014) *Laf4/Aff3*, a gene involved in intellectual disability, is required for cellular migration in the mouse cerebral cortex. *PLoS One* 9:e105933. <https://doi.org/10.1371/journal.pone.0105933>
51. Voisin N, Schnur RE, Douzou S et al (2021) Variants in the degen of *AFF3* are associated with intellectual disability, mesomelic dysplasia, horseshoe kidney, and epileptic encephalopathy. *Am J Hum Genet* 108:857–873. <https://doi.org/10.1016/j.ajhg.2021.04.001>
52. Ranade SS, Woo S-H, Dubin AE et al (2014) *Piezo2* is the major transducer of mechanical forces for touch sensation in mice. *Nature* 516:121–125. <https://doi.org/10.1038/nature13980>
53. Shi H, Li B, Tong Q et al (2019) Effects of LED light color and intensity on feather pecking and fear responses of layer breeders in natural mating colony cages. *Animals (Basel)* 9. <https://doi.org/10.3390/ani9100814>
54. Xu S (2003) Theoretical basis of the Beavis effect. *Genetics* 165:2259–2268. <https://doi.org/10.1093/genetics/165.4.2259>
55. Knuesel I, Mastrocola M, Zuellig RA et al (1999) Short communication: altered synaptic clustering of GABAA receptors in mice lacking dystrophin (*mdx* mice). *Eur J Neurosci* 11:4457–4462. <https://doi.org/10.1046/j.1460-9568.1999.00887.x>
56. Earls LR, Hacker ML, Watson JD et al (2010) Coenzyme Q protects *Caenorhabditis elegans* GABA neurons from calcium-dependent degeneration. *Proc Natl Acad Sci* 107:14460–14465. <https://doi.org/10.1073/pnas.0910630107>
57. Abe H, Okazawa M, Nakanishi S (2011) The *Etv1/Er81* transcription factor orchestrates activity-dependent gene regulation in the terminal maturation program of cerebellar granule cells. *Proc Natl Acad Sci* 108:12497–12502. <https://doi.org/10.1073/pnas.1109940108>
58. Woo R-S, Li X-M, Tao Y et al (2007) Neuregulin-1 enhances depolarization-induced GABA release. *Neuron* 54:599–610. <https://doi.org/10.1016/j.neuron.2007.04.009>
59. Ayka A, Şehirli AÖ (2020) The role of the SLC transporters protein in the neurodegenerative disorders. *Clin Psychopharmacol Neurosci* 18:174–187. <https://doi.org/10.9758/cpn.2020.18.2.174>
60. Chen P, Huang K, Zhou G et al (2011) Common SNPs in *CSF2RB* are associated with major depression and schizophrenia in the Chinese Han population. *World J Biol Psychiatry* 12:233–238. <https://doi.org/10.3109/15622975.2010.544328>
61. Gerges P, Bitar T, Laumonier F et al (2022) Identification of novel gene variants for autism spectrum disorders in the Lebanese population using whole-exome sequencing. *Genes* 13:186. <https://doi.org/10.3390/genes13020186>
62. Kiaii S, Clear AJ, Ramsay AG et al (2013) Follicular lymphoma cells induce changes in T-cell gene expression and function: potential impact on survival and risk of transformation. *J Clin Oncol* 31:2654–2661. <https://doi.org/10.1200/JCO.2012.44.2137>
63. Oh S, Shin S, Song H et al (2019) Relationship between ETS transcription factor *ETV1* and TGF- $\beta$ -regulated SMAD proteins in prostate cancer. *Sci Rep* 9:8186. <https://doi.org/10.1038/s41598-019-44685-3>
64. Bonett RM, Hu F, Bagamasbad P et al (2009) Stressor and glucocorticoid-dependent induction of the immediate early gene *kruppel-like factor 9*: implications for neural development and plasticity. *Endocrinology* 150:1757–1765. <https://doi.org/10.1210/en.2008-1441>
65. Qin S, Liu M, Niu W et al (2011) Dysregulation of *Kruppel-like factor 4* during brain development leads to hydrocephalus in mice. *Proc Natl Acad Sci* 108:21117–21121. <https://doi.org/10.1073/pnas.1112351109>
66. Qin S, Zhang C-L (2012) Role of *Kruppel-like factor 4* in neurogenesis and radial neuronal migration in the developing cerebral cortex. *Mol Cell Biol* 32:4297–4305. <https://doi.org/10.1128/MCB.00838-12>
67. Besnard A, Langberg T, Levinson S et al (2018) Targeting *Kruppel-like factor 9* in excitatory neurons protects against chronic stress-induced impairments in dendritic spines and fear responses. *Cell Rep* 23:3183–3196. <https://doi.org/10.1016/j.celrep.2018.05.040>
68. Ávila-Mendoza J, Subramani A, Denver RJ (2020) *Krüppel-like factors 9 and 13* block axon growth by transcriptional repression of key components of the cAMP signaling pathway. *Front Mol Neurosci* 13:602638. <https://doi.org/10.3389/fnmol.2020.602638>
69. Antin PB, Pier M, Sesepasara T et al (2010) Embryonic expression of the chicken *Krüppel-like (KLF)* transcription factor gene family. *Dev Dyn* 239:1879–1887. <https://doi.org/10.1002/dvdy.22318>
70. Konopaske GT, Subburaju S, Coyle JT et al (2015) Altered prefrontal cortical *MARCKS* and *PPP1R9A* mRNA expression in schizophrenia and bipolar disorder. *Schizophr Res* 164:100–108. <https://doi.org/10.1016/j.schres.2015.02.005>
71. Kim SS, Wang H, Li X-Y et al (2011) Neurabin in the anterior cingulate cortex regulates anxiety-like behavior in adult mice. *Mol Brain* 4:6. <https://doi.org/10.1186/1756-6606-4-6>
72. Banihashemi S, Tahmasebi-Birgani M, Mohammadiasl J et al (2020) Whole exome sequencing identified a novel nonsense *INPP4A* mutation in a family with intellectual disability. *Eur J Med Genet* 63:103846. <https://doi.org/10.1016/j.ejmg.2020.103846>
73. Wang L, Luo J, Fang M et al (2012) A new trick of *INPP4A*: decreased expression of *INPP4A* in patients with temporal lobe epilepsy and pilocarpine-induced rat model. *Synapse* 66:533–541. <https://doi.org/10.1002/syn.21540>
74. John J, Kukshal P, Bhatia T et al (2019) Oligogenic rare variant contributions in schizophrenia and their convergence with genes harbouring de novo mutations in schizophrenia, autism and intellectual disability: evidence from multiplex families 1
75. Chapman RM, Tinsley CL, Hill MJ et al (2019) Convergent evidence that *ZNF804A* is a regulator of pre-messenger RNA processing and gene expression. *Schizophr Bull* 45:1267–1278. <https://doi.org/10.1093/schbul/sby183>
76. Leone M, Cazorla-Vázquez S, Ferrazzi F et al (2021) *IQGAP3*, a YAP target, is required for proper cell-cycle progression and genome stability. *Mol Cancer Res* 19:1712–1726. <https://doi.org/10.1158/1541-7786.MCR-20-0639>
77. Yuan Y, Jiang X, Tang L et al (2022) Comprehensive analyses of the immunological and prognostic roles of an *IQGAP3AR/let-7c-5p/IQGAP3* axis in different types of human cancer. *Front Mol Biosci* 9:763248. <https://doi.org/10.3389/fmolb.2022.763248>
78. Hung T, Pratt GA, Sundararaman B et al (2015) The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science* 350:455–459. <https://doi.org/10.1126/science.aac7442>
79. Zhang Y, Yang X, Zhu X-L et al (2021) *S100A* gene family: immune-related prognostic biomarkers and therapeutic targets for low-grade glioma. *Aging (Albany NY)* 13:15459–15478. <https://doi.org/10.18632/aging.203103>
80. Gandal MJ, Zhang P, Hadjimichael E et al (2018) Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and

- bipolar disorder. *Science* 362. <https://doi.org/10.1126/science.aat8127>
81. Walton E, Cecil CAM, Suderman M et al (2017) Longitudinal epigenetic predictors of amygdala:hippocampus volume ratio. *J Child Psychol Psychiatry* 58:1341–1350. <https://doi.org/10.1111/jcpp.12740>
  82. Blokhin IO, Khorkova O, Saveanu RV et al (2020) Molecular mechanisms of psychiatric diseases. *Neurobiol Dis* 146:105136. <https://doi.org/10.1016/j.nbd.2020.105136>
  83. Johnsson M, Williams MJ, Jensen P et al (2016) Genetical genomics of behavior: a novel chicken genomic model for anxiety behavior. *Genetics* 202:327–340. <https://doi.org/10.1534/genetics.115.179010>
  84. Chan JZ, Fernandes MF, Hashemi A et al (2021) Age-associated increase in anxiety-like behavior in Lpaat $\delta$ /Agpat4 knockout mice. *Curr Res Behav Sci* 2:100042. <https://doi.org/10.1016/j.crbeha.2021.100042>
  85. Hori H, Sasayama D, Teraishi T et al (2016) Blood-based gene expression signatures of medication-free outpatients with major depressive disorder: integrative genome-wide and candidate gene analyses. *Sci Rep* 6:18776. <https://doi.org/10.1038/srep18776>
  86. Guffanti G, Strik Lievers L, Bonati MT et al (2011) Role of UBE3A and ATP10A genes in autism susceptibility region 15q11-q13 in an Italian population: a positive replication for UBE3A. *Psychiatry Res* 185:33–38. <https://doi.org/10.1016/j.psychres.2010.04.057>
  87. Seigneur E, Wang J, Dai J et al (2021) Cerebellin-2 regulates a serotonergic dorsal raphe circuit that controls compulsive behaviors. *Mol Psychiatry*. <https://doi.org/10.1038/s41380-021-01187-x>
  88. Li X, Wang Q, He K et al (2014) Common variants in the CDH7 gene are associated with major depressive disorder in the Han Chinese population. *Behav Genet* 44:97–101. <https://doi.org/10.1007/s10519-014-9645-y>
  89. Huang Y, Zhu Y, He W et al (2021) Altered expression of DENND5B in patients with epilepsy and its regulation of seizures in mice. *Epilepsy Res* 178:106817. <https://doi.org/10.1016/j.epilepsyres.2021.106817>
  90. Pei Y-F, Zhang L, Yang T-L et al (2012) Genome-wide association study of copy number variants suggests LTBP1 and FGD4 are important for alcohol drinking. *PLoS One* 7:e30860. <https://doi.org/10.1371/journal.pone.0030860>
  91. Chiapponi C, Piras F, Piras F et al (2016) GABA system in schizophrenia and mood disorders: a mini review on third-generation imaging studies. *Front Psychiatry* 7:61. <https://doi.org/10.3389/fpsy.2016.00061>
  92. Werner CT, Mitra S, Martin JA et al (2019) Ubiquitin-proteasomal regulation of chromatin remodeler INO80 in the nucleus accumbens mediates persistent cocaine craving. *Sci Adv* 5:eaay0351. <https://doi.org/10.1126/sciadv.aay0351>
  93. Iwamoto K, Kakiuchi C, Bundo M et al (2004) Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. *Mol Psychiatry* 9:406–416. <https://doi.org/10.1038/sj.mp.4001437>
  94. Alblooshi H, Al Safar H, El Kashef A et al (2020) Stratified analyses of genome wide association study data reveal haplotypes for a candidate gene on chromosome 2 (KIAA1211L) is associated with opioid use in patients of Arabian descent. *BMC Psychiatry* 20:41. <https://doi.org/10.1186/s12888-019-2425-8>
  95. Schubert Baldo M, Vilarinho L (2020) Molecular basis of Leigh syndrome: a current look. *Orphanet J Rare Dis* 15:31. <https://doi.org/10.1186/s13023-020-1297-9>
  96. Tang L, Xiang Q, Xiang J et al (2021) A variant in the 3'-untranslated region of the MC2R gene decreases the risk of schizophrenia in a female Han Chinese population. *J Int Med Res* 49:3000605211029504. <https://doi.org/10.1177/03000605211029504>
  97. Kippe JM, Mueller TM, Haroutunian V et al (2015) Abnormal N-acetylglucosaminyltransferase expression in prefrontal cortex in schizophrenia. *Schizophr Res* 166:219–224. <https://doi.org/10.1016/j.schres.2015.06.002>
  98. Fang C-P, Liu T-H, Chung R-H et al (2020) Genetic variants in NECTIN4 encoding an adhesion molecule are associated with continued opioid use. *PLoS One* 15:e0234549. <https://doi.org/10.1371/journal.pone.0234549>
  99. Ng D, Pitcher GM, Szilard RK et al (2009) Neto1 is a novel CUB-domain NMDA receptor-interacting protein required for synaptic plasticity and learning. *PLoS Biol* 7:e41. <https://doi.org/10.1371/journal.pbio.1000041>
  100. Tropeano M, Howley D, Gazzellone MJ et al (2016) Microduplications at the pseudoautosomal SHOX locus in autism spectrum disorders and related neurodevelopmental conditions. *J Med Genet* 53:536–547. <https://doi.org/10.1136/jmedgenet-2015-103621>
  101. Frydecka D, Misiak B, Pawlak-Adamska E et al (2015) Sex differences in TGFB- $\beta$  signaling with respect to age of onset and cognitive functioning in schizophrenia. *Neuropsychiatr Dis Treat* 11:575–584. <https://doi.org/10.2147/NDT.S74672>
  102. Warnica W, Merico D, Costain G et al (2015) Copy number variable microRNAs in schizophrenia and their neurodevelopmental gene targets. *Biol Psychiatry* 77:158–166. <https://doi.org/10.1016/j.biopsych.2014.05.011>
  103. Vatsa N, Jana NR (2018) UBE3A and its link with autism. *Front Mol Neurosci* 11:448. <https://doi.org/10.3389/fnmol.2018.00448>
  104. Gamazon ER, Badner JA, Cheng L et al (2013) Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol Psychiatry* 18:340–346. <https://doi.org/10.1038/mp.2011.174>
  105. Fitzgerald J, Kennedy D, Viseshakul N et al (2000) UNCL, the mammalian homologue of UNC-50, is an inner nuclear membrane RNA-binding protein1 Published on the world wide web on 10 August 2000. *Brain Res* 877:110–123. [https://doi.org/10.1016/S0006-8993\(00\)02692-5](https://doi.org/10.1016/S0006-8993(00)02692-5)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.