

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

Clemens Falker-Gieske^{1*}, Jörn Bennewitz², and Jens Tetens^{1,3}

¹Department of Animal Sciences, Georg-August-University, Burckhardtweg 2, 37077 Göttingen, Germany

²Institute of Animal Science, University of Hohenheim, Garbenstr. 17, 70599 Stuttgart, Germany

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

*Correspondence

Clemens Falker-Gieske

Georg-August-Universität Göttingen

Department of Animal Sciences

Division of Functional Breeding

Burckhardtweg 2

37077 Göttingen

Germany

Tel.: +49 551 39 23087

clemens.falker-gieske@uni-goettingen.de

ORCID: Clemens Falker-Gieske 0000-0001-9160-1909; Jörn Bennewitz 0000-0001-6450-1160; Jens Tetens 0000-0001-5352-464X

Statements and Declarations

Availability of data and materials

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

Competing interests

The authors declare to have no competing interests of any kind.

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

Funding

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. Publication fee was covered by the Open Access Publication Funds of the Göttingen University.

Contributions

Conceptualization: Clemens Falker-Gieske, Jens Tetens; Methodology: Clemens Falker-Gieske, Jens Tetens; Formal analysis and investigation: Clemens Falker-Gieske; Writing - original draft preparation: Clemens Falker-Gieske; Writing - review and editing: Jens Tetens, Jörn Bennewitz; Funding acquisition: Jens Tetens, Jörn Bennewitz; Resources: Jens Tetens, Jörn Bennewitz; Supervision: Jens Tetens.

Acknowledgements

We acknowledge support by the Open Access Publication Funds of the Göttingen University.

Abstract

Feather pecking (FP) is a damaging non-aggressive 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 re-analyzed whole genome sequencing data from two experimental populations, which have been divergently selected for FP behavior for over more than 15 generations, and 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 micro RNAs targeting several GABA receptor genes, and dystrophin (*DMD*), a direct regulator of GABA receptor clustering. Furthermore, we found the transcription factor *ETV1* to be associated with differential expression of 23 genes, which points towards a role of *ETV1*, 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

Feather pecking (FP) in chickens is a behavioral disorder, which 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 have 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_2 cross and 12 half-sib families (in the preceding text referred to as F_2 and HS). A detailed description 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 γ -aminobutyric acid (GABA) 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 respond 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 down- 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 it 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 on synapse development and lead to behavioral abnormalities [16]. The evidence suggests that this mechanism is

responsible for disturbed embryonic brain development in HFP chickens, which contributes to feather pecking 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. Structural genomic variants (SVs) play a notable role in human psychiatric disorders [17], which is also the case for tandem repeats (TRs) [18]. Commonly investigated classes of SVs include insertions, deletions, duplications, inversions, and translocations, which arise from various combinations of DNA gains, losses, or rearrangements [19]. 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

All experimental procedures [1], rearing and husbandry conditions [20] as well as phenotyping [2] were described in previous studies. Briefly, the phenotypic data were generated by direct observations made by seven independent investigators. The phenotypes are expressed as the number of FP bouts actively delivered during a standardized time span. As these count data 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 (< 50bp) insertions and deletions (InDels) in our previous study [2]. SVs were called as described by Blaj *et al.* [21] with slightly different settings: A high confidence SV call set was produced from the output of three variant callers: smove v0.2.6 (Brent, P. (2018). Smoove. <https://brentp.github.io/post/smove/>), DELLY v0.7.7 [22], and manta v1.6.0 [23]. SURVIROR v1.0.7 [24] 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. Tandem repeats were called with GangSTR [25]. 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 of 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 5.2 [26] and the setting ne=1000. SVs and TRs were merged with SNPs and InDels from our previous study, which were acquired

with the genome analysis toolkit (GATK) [27]. This merged call set was phased with Beagle 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. To remove SNPs and InDels from the imputed SVs/TRs GATK v4.1.8.1 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) [28]. All GWAS were conducted with gcta v. 1.92.3 beta3 [29] applying a mixed linear model association analysis with a leaving-one-chromosome-out approach. 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]. Information on the hatch was used as a covariate in all GWAS and eGWAS. Genomic relationship matrices were created from the target 60k SNP Chip genotypes. Meta-analyses of GWAS results were performed with METAL [30] as previously described [2]. The proportion of variance in phenotype explained by a given SNP (PVE) was calculated according to Shim *et al.* [31].

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 [32] followed by the detection of significant gene-gene interactions with their PCIT algorithm [33]. 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 least ten of the eGWAS. That way, variants affecting the main phenotype and gene expression were, were 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 1×10^{-4} . The gene-gene interaction map was constructed with Cytoscape and gene class information was acquired with PANTHER [34] and UniProt [35].

Transcription factor enrichment

To detect significant binding site enrichment for the transcription factor *ETV1* the CiiDER software (build May 15th, 2020) [36] was employed with frequency matrix MA0761.2 (<https://jaspar.genereg.net/matrix/MA0761.2/>). 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 *ETV1* variant (P value $> 1 \times 10^{-4}$; Supplementary Information S1) were used as input and DEGs with a Log₂ fold change < 0.2 were used as background genes. The analysis was performed as in our previous study [8].

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 F₀ chickens. The SVs contained 6,014 deletions, 2,741 inversions, 1,334 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) [26] and set the effective population size to 1000, which is known to improve the imputation accuracy in small populations [37]. GWAS results for the trait FPD_BC were analyzed for both experimental designs, the F₂ 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 Figure 1a. Common peaks for both variants classes on GGA1 and GGA2 were observable in the F₂ 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 ($-\log_{10}P > 1.41 \times 10^{-6}$) was a TR in the HS population 126,821 bp downstream of *GABRB3*. The proportion of variance in the phenotype FPD_BC, which is explained by this variant is 0.047. Given heritability estimates of around 0.15 [38–40] 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 (Figure 1b). Numerous genes showed high association signals in both variant classes in the same experimental population. For instance, *ENSGALG00000047084*, *GABRB3*, *INPP4A*, *ATP10A*, *TSGA10*, *LIPT1*, *COA5*, *UNC50*, *SHOX*, *CRLS2*, and *AFF3* showed high association in the HS population for both variant classes. The same holds true for the F₂ cross (upper block of 20 genes in Figure 1b), although associations were not as strong as in the HS population. Of considerable interest are those genes, which show association in all GWAS: *ENSGALG00000048825*, *ENSGALG00000051557*, *ENSGALG00000051256*, *ENSGALG00000016495*, *FHOD3*, *gga-mir-187*, *GALNT1*, and *PIEZO2*. Predicted targets of the micro RNA (miRNA) *gga-mir-187* are *GABRA1*, *GABRB2*, *GABRB3*, *GABRG1*, and *GABRG2*. *GABRB2* is also a predicted target of *gga-mir-3524b* (www.targetscan.org accessed January 27, 2022).

Expression quantitative trait loci

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 S2). 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 [32] followed by the detection of significant correlations with the PCIT algorithm [33]. Central to the gene-gene interaction map is the transcription factor *ETV1*. We selected DEGs with a P value $< 1 \times 10^{-4}$ for association with the variant, a tandem repeat with the sequence CCCGGCCCG 70 bp upstream of *ETV1* (GGA2: 27337541). 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 *ETV1* binding sites (Figure 3a) in proximity to those genes (Figure 3b, Supplementary Information S3). 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 S4.

Discussion

Numerous studies have been conducted to unravel the genetics behind FP behavior in chickens, as reviewed in [12, 41], but none of these focused on the analysis of structural genetic variation. We combined data of two well-described experimental crosses, an F_2 design and a HS population, and performed multiple GWAS and an eQTL analysis on different classes of genetic variants. 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. Furthermore, miRNA *gga-mir-187* is strongly associated in all conducted GWAS (Figure 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 feather pecking 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 into 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 an 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 [42], 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 (Figure 2) have been connected to the GABAergic system. These include *COQ4* [43], *ETV1* [44], *NEURL1* [45], *SLC25A26*, and *SLC27A4* [46]. In this regard, *ETV1* is of considerable interest since it is the only transcription factor that we identified in our eQTL analysis with SVs and TRs. According to the Human Protein Atlas *ETV1* expression is specific to salivary glands and the brain, specifically the cerebellum and thalamus (<https://www.proteinatlas.org/ENSG00000006468-ETV1>; accessed August 2022) *ETV1* is associated with

the differential expression of 23 genes between HFP and LFP and *ETV1* binding sites are enriched in proximity to these genes (Figure 3). Among these putative *ETV1* targets are *CSF2RB*, which has been associated with major depression and schizophrenia [47], and *MIS18BP1*, a gene implicated in autism spectrum [48]. Furthermore, *ETV1* antibody coimmunoprecipitated the GABAA receptor $\alpha 6$ (GABAAR $\alpha 6$) promoter region in mice [44] and is a regulator of gene expression in CD4 and CD8 T cells [49]. Hence, *ETV1* 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 enrichment of binding sites for that transcription factor in proximity to its associated DEGs. However, in the light of the fact that *ETV1* forms functional complexes with Smad4 [50] 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 towards a central role of Krüppel-Like factors in neurodevelopment and behavior [51–55]. 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 [56].

Apart from *DMD* we also discovered the genes *PPP1R9A*, *INPP4A*, and *COA5* in both eQTL analyses. One of which was performed with SVs and TRs (Figure 2) and one with SNPs and InDels [8]. In schizophrenia and bipolar disorder, the prefrontal cortical expression of *PPP1R9A* was altered [57] and its gene product Neurabin regulates anxiety-like behavior in adult mice [58]. *INPP4A* has been implicated in multiple neurological conditions, namely schizophrenia, autism, epilepsy, and intellectual disability [59–61]. 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 [62]. We discovered multiple links of FP to human psychiatric disorders in our previous studies [1, 2, 14]. The fact that methylation of *KLF14* correlates with psychosis severity in schizophrenia patients [63] 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, which common mechanisms these conditions share, which is a current matter of debate [64]. Blokhin *et al.* proposed that future treatment of human psychiatric disorders should be tailored under the use of a genome-wide “targetome”. 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.

Tables

Table1: Positional ranges in base pairs of QTL that were discovered in both experimental populations (F₂ cross and half-sib (HS) families) for both groups of variant classes: Single nucleotide polymorphisms (SNPs) and short insertions/deletions (InDels) as well as structural variants (SVs) and tandem repeats (TRs).

F ₂ SNPs and short InDels	F ₂ 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

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 F₂ 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 ₂	INV	1,403	GGA1:59,216,858-59,218,260	1.14 x 10 ⁻⁵	intron_variant	<i>BICD1</i>
F ₂	TR		GGA1:59,138,261	2.29 x 10 ⁻⁵	intron_variant	<i>BICD1</i>
F ₂	TR		GGA2:83,217,700	2.11 x 10 ⁻⁵	intron_variant	<i>FHOD3</i>
F ₂	DEL	509	GGA2:83,493,152-83,493,661	2.11 x 10 ⁻⁵	intergenic_region	<i>ENSGALG00000051475-RPRD1A</i>
F ₂	TR		GGA2:83,497,153	2.30 x 10 ⁻⁵	intergenic_region	<i>ENSGALG00000051475-RPRD1A</i>
F ₂	DEL	209	GGA2:83,497,939-83,498,148	2.30 x 10 ⁻⁵	intergenic_region	<i>ENSGALG00000051475-RPRD1A</i>
F ₂	TR		GGA2:89,486,485	6.90 x 10 ⁻⁵	intergenic_region	<i>ENSGALG00000053898-ENSGALG0000005246</i>
F ₂	TR		GGA2:94,489,210	7.28 x 10 ⁻⁵	intergenic_region	<i>ENSGALG00000051191-ENSGALG0000003284</i>
HS	TR		GGA1:132,889,302	7.79 x 10 ⁻⁷	intergenic_region	<i>GABRB3-ENSGALG00000047084</i>
HS	TR		GGA1:130,621,356	2.64 x 10 ⁻⁶	intergenic_region	<i>CRLF2-SHOX</i>
HS	TR		GGA1:133,533,130	4.00 x 10 ⁻⁶	intron_variant	<i>INPP4A</i>
HS	TR		GGA1:133,533,741	4.09 x 10 ⁻⁶	intron_variant	<i>INPP4A</i>
HS	DEL	368	GGA1:134,077,752-134,077,908	1.37 x 10 ⁻⁵	intron_variant	<i>AFF3</i>
HS	TR		GGA1:133,606,912	2.30 x 10 ⁻⁵	downstream_gene_variant	<i>COA5</i>
HS	TR		GGA1:132,889,302	6.42 x 10 ⁻⁵	intergenic_region	<i>GABRB3-ENSGALG00000047084</i>
HS	TR		GGA1:134,020,467	9.02 x 10 ⁻⁵	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>

Table 3: Large deletions (DEL) and tandem repeats (TR) that showed genome-wide significant association 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

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 [65, 66]. Age-dependent increase of anxiety in knockout mice [67] and association with major depressive disorder [68].
<i>AGPAT4</i>	1-acyl-sn-glycerol-3-phosphate acyltransferase delta	Autism susceptibility locus [69].
<i>ATP10A</i>	Phospholipid-transporting ATPase VA	Regulator of compulsive behavior [70].
<i>CBLN2</i>	Cerebellin-2	Associated with major depressive disorder [71].
<i>CDH7</i>	Cadherin-7	Altered expression in patients with epilepsy and regulation of seizures in mice [72].
<i>DENND5B</i>	DENN domain-containing protein 5B	Involved in clustering and trafficking of GABA receptors [73].
<i>FGD4</i>	FYVE, RhoGEF and PH domain-containing protein 4	Controls the Dendritic Spine Morphology [74].
<i>FHOD3</i>	FH1/FH2 domain-containing protein 3	Associated with FP in chickens [2, 14] and schizophrenia[75].
<i>GABRB3</i>	Gamma-aminobutyric acid receptor subunit beta-3	Mediator of cocaine addiction [76].
<i>INO80C</i>	INO80 complex subunit C	Decreased expression in epilepsy [60] and nonsense mutation in intellectual disability [59]. Associated with depression, bipolar disorder and schizophrenia [77] and opioid use [78]
<i>INPP4A</i>	Inositol polyphosphate-4-phosphatase type I A	
<i>KIAA1211L</i> (<i>ENSGALG00000016755</i>)	CRACD-like protein	
<i>LIPT1</i>	Lipoyltransferase 1	Involved in the development of Leigh syndrome [79].
<i>MC2R</i>	Adrenocorticotrophic hormone receptor	Risk factor for schizophrenia [80].
<i>MGAT4A</i>	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	Abnormal expression in prefrontal cortex in schizophrenia [81].
<i>MRPL30</i>	39S ribosomal protein L30	Associated with major depressive disorder [68].
<i>NECTIN4</i>	Nectin-4	Associated with opioid abuse [82].
<i>NETO1</i>	Neuropilin and tolloid-like protein 1	NMDA receptor-interacting protein required for synaptic plasticity and learning [83].
<i>SHOX</i>	Short stature homeobox protein	Autism locus [84].
<i>TGFB2</i>	Transforming growth factor beta-2 proprotein	Involved in signaling towards age of onset and cognitive functioning in schizophrenia [85].
<i>TSGA10</i>	Testis-specific gene 10 protein	miRNA target in the development of schizophrenia [86].
<i>UBE3A</i>	Ubiquitin-protein ligase E3A	Autism susceptibility locus [69] [87]. Associated with bipolar disorder [88] and involvement in the cell-surface expression of neuronal nicotinic receptors [89].
<i>UNC50</i>	Protein unc-50 homolog	

Figure Legends

Figure 1: (a) Genome-wide association studies (GWAS) performed with single nucleotide polymorphisms (SNPs) in combination with short (< 50bp) insertions and deletions (InDels) and structural variants (SVs) in combination with tandem repeats (TRs). All variant classes were investigated in the F₂ cross and half-sib (HS) families of chickens divergently selected for feather pecking and subsequently combined in a meta-analysis. (b) Heatmap of the 20 highest associated genes from each GWAS. The $-\text{Log}_{10}P$ values of the highest associated variants for each gene were used.

Figure 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.

Figure 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 $> 1 \times 10^{-4}$.

Supplementary Information

Supplementary Information S1: Differentially expressed genes between high and low feather pecking chickens, which were associated with a tandem repeat 70 bp downstream of the *ETV1* transcription factor.

Supplementary Information S2: Manhattan plots of expression genome wide associations studies conducted on genes, which were differentially expressed between high and low feather pecking chickens.

Supplementary Information S3: Graphical representation of *ETV1* binding sites in proximity to associated differentially expressed genes.

Supplementary Information S4: Results of transcription factor binding site enrichment analysis with Ciider for transcription factors of the ETS (E twenty-six) family to differentially expressed genes that were associated with a tandem repeat downstream of *ETV1*.

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-analyses 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 Poultry Science Journal* 69:775–788. <https://doi.org/10.1017/S0043933913000809>
5. Rodenburg TB, van Krimpen MM, Jong IC de et al. (2013) The prevention and control of feather pecking in laying hens: identifying the underlying principles. *World's Poultry Science Journal* 69:361–374. <https://doi.org/10.1017/S0043933913000354>
6. Parmentier HK, Rodenburg TB, Vries Reilingh G de 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, Vries Reilingh G de 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, Vries H de, 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 Poultry Science Journal* 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. Rucker JJH, McGuffin P (2012) Genomic structural variation in psychiatric disorders. *Dev Psychopathol* 24:1335–1344. <https://doi.org/10.1017/S0954579412000740>
18. 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>
19. 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>
20. 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>
21. 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>
22. 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>
23. 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>
24. 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>
25. 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>
26. 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>
27. 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>
28. Danecek P, Bonfield JK, Liddle J et al. (2021) Twelve years of SAMtools and BCFtools. *Gigascience* 10. <https://doi.org/10.1093/gigascience/giab008>
29. 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>
30. 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>
31. 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>
32. 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

33. 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>
34. 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>
35. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
36. 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>
37. 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>
38. 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>
39. 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>
40. 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>
41. 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>
42. 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>
43. Earls LR, Hacker ML, Watson JD et al. (2010) Coenzyme Q protects *Caenorhabditis elegans* GABA neurons from calcium-dependent degeneration. *Proceedings of the National Academy of Sciences* 107:14460–14465. <https://doi.org/10.1073/pnas.0910630107>
44. 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. *Proceedings of the National Academy of Sciences* 108:12497–12502. <https://doi.org/10.1073/pnas.1109940108>
45. 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>
46. 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>
47. 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>
48. Gerges P, Bitar T, Laumonnier 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>

49. 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>
50. Oh S, Shin S, Song H et al. (2019) Relationship between ETS Transcription Factor ETV1 and TGF- β -regulated SMAD Proteins in Prostate Cancer. *Sci Rep* 9:8186. <https://doi.org/10.1038/s41598-019-44685-3>
51. 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>
52. Qin S, Liu M, Niu W et al. (2011) Dysregulation of Kruppel-like factor 4 during brain development leads to hydrocephalus in mice. *Proceedings of the National Academy of Sciences* 108:21117–21121. <https://doi.org/10.1073/pnas.1112351109>
53. 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>
54. 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>
55. 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>
56. Antin PB, Pier M, Seseasara 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>
57. 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>
58. 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>
59. 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>
60. 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>
61. 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, vol 1
62. 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>
63. 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>
64. 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>

65. 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>
66. Voisin N, Schnur RE, Douzgou S et al. (2021) Variants in the degron 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>
67. Chan JZ, Fernandes MF, Hashemi A et al. (2021) Age-associated increase in anxiety-like behavior in Lpaatδ/Agpat4 knockout mice. *Current Research in Behavioral Sciences* 2:100042. <https://doi.org/10.1016/j.crbeha.2021.100042>
68. 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>
69. 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>
70. 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>
71. 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>
72. 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.eplepsyres.2021.106817>
73. 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>
74. 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>
75. 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>
76. 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>
77. 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>
78. 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>
79. 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>

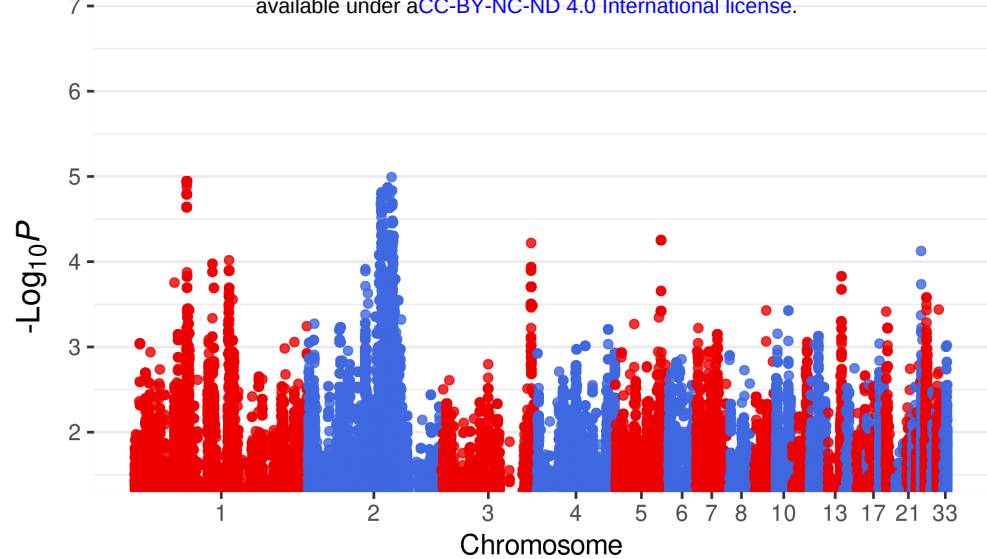
80. Tang L, Xiang Q, Xiang J et al. (2021) A variant in the 3'-untranslated region of the MC2R gene decreases the risk of schizophrenia in a female Han Chinese population. *J Int Med Res* 49:3000605211029504. <https://doi.org/10.1177/03000605211029504>
81. Kippe JM, Mueller TM, Haroutunian V et al. (2015) Abnormal N-acetylglucosaminyltransferase expression in prefrontal cortex in schizophrenia. *Schizophr Res* 166:219–224. <https://doi.org/10.1016/j.schres.2015.06.002>
82. 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>
83. 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>
84. 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>
85. Frydecka D, Misiak B, Pawlak-Adamska E et al. (2015) Sex differences in TGFB- β 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>
86. 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>
87. Vatsa N, Jana NR (2018) UBE3A and Its Link With Autism. *Front Mol Neurosci* 11:448. <https://doi.org/10.3389/fnmol.2018.00448>
88. 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>
89. Fitzgerald J, Kennedy D, Viseshakul N et al. (2000) UNCL, the mammalian homologue of UNC-50, is an inner nuclear membrane RNA-binding protein11Published on the World Wide Web on 10 August 2000. *Brain Research* 877:110–123. [https://doi.org/10.1016/S0006-8993\(00\)02692-5](https://doi.org/10.1016/S0006-8993(00)02692-5)

a

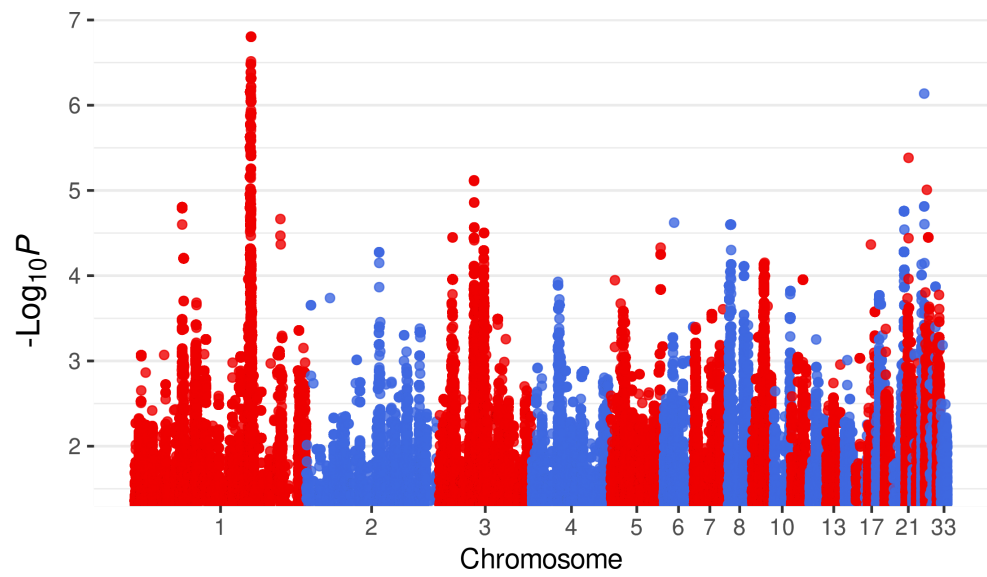
SNPs / short InDels

bioRxiv preprint doi: <https://doi.org/10.1101/2022.09.16.508210>; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

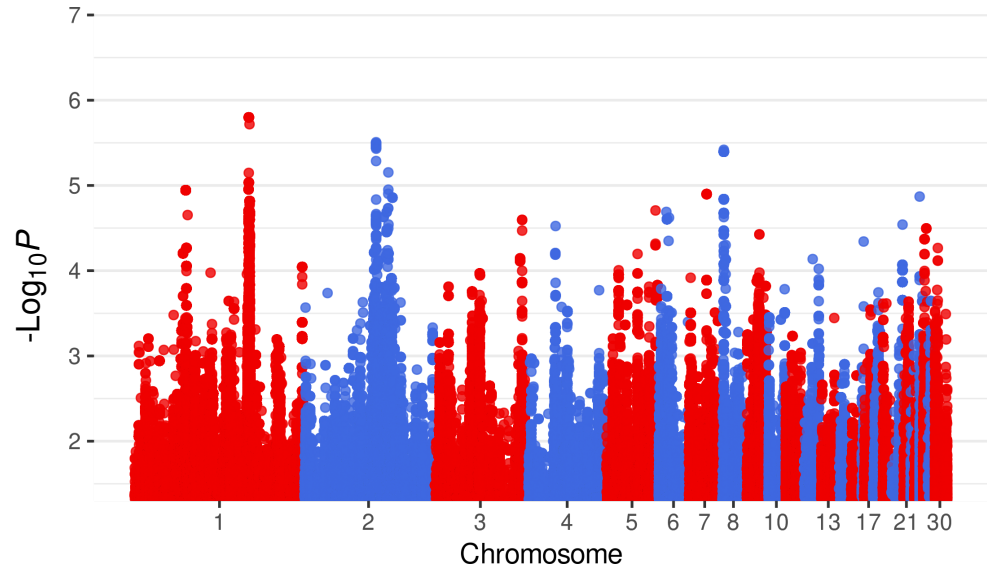
F₂



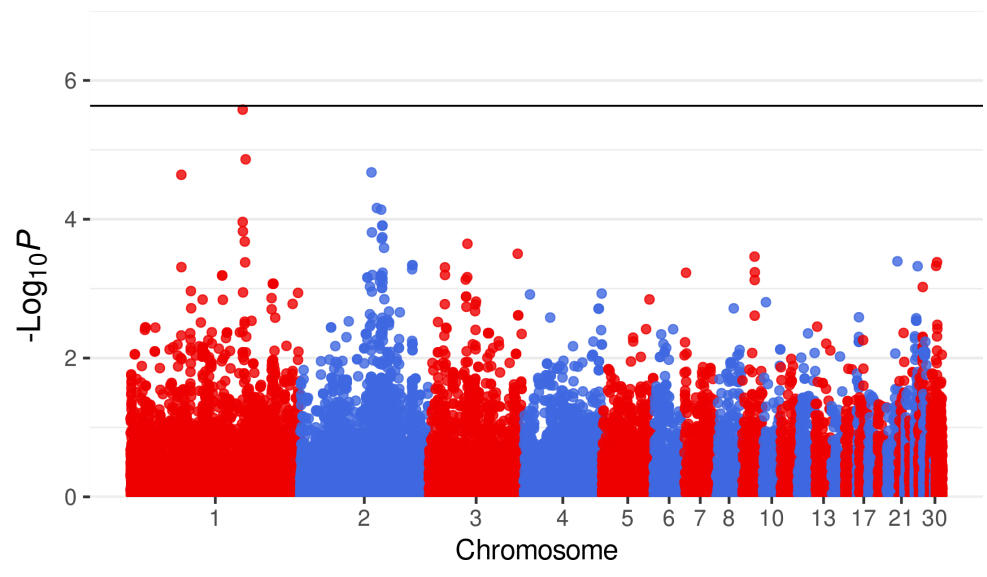
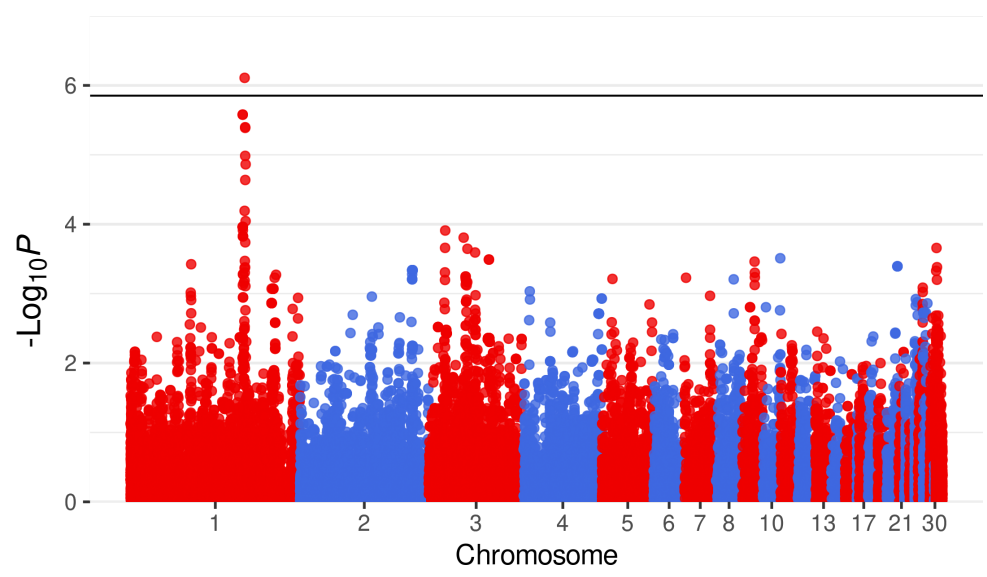
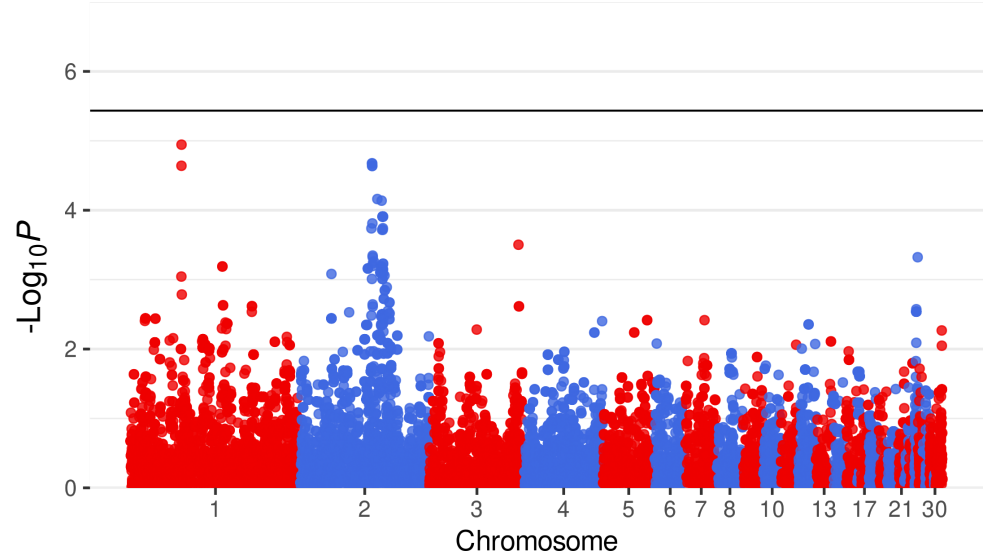
HS



meta



SVs / TRs



b

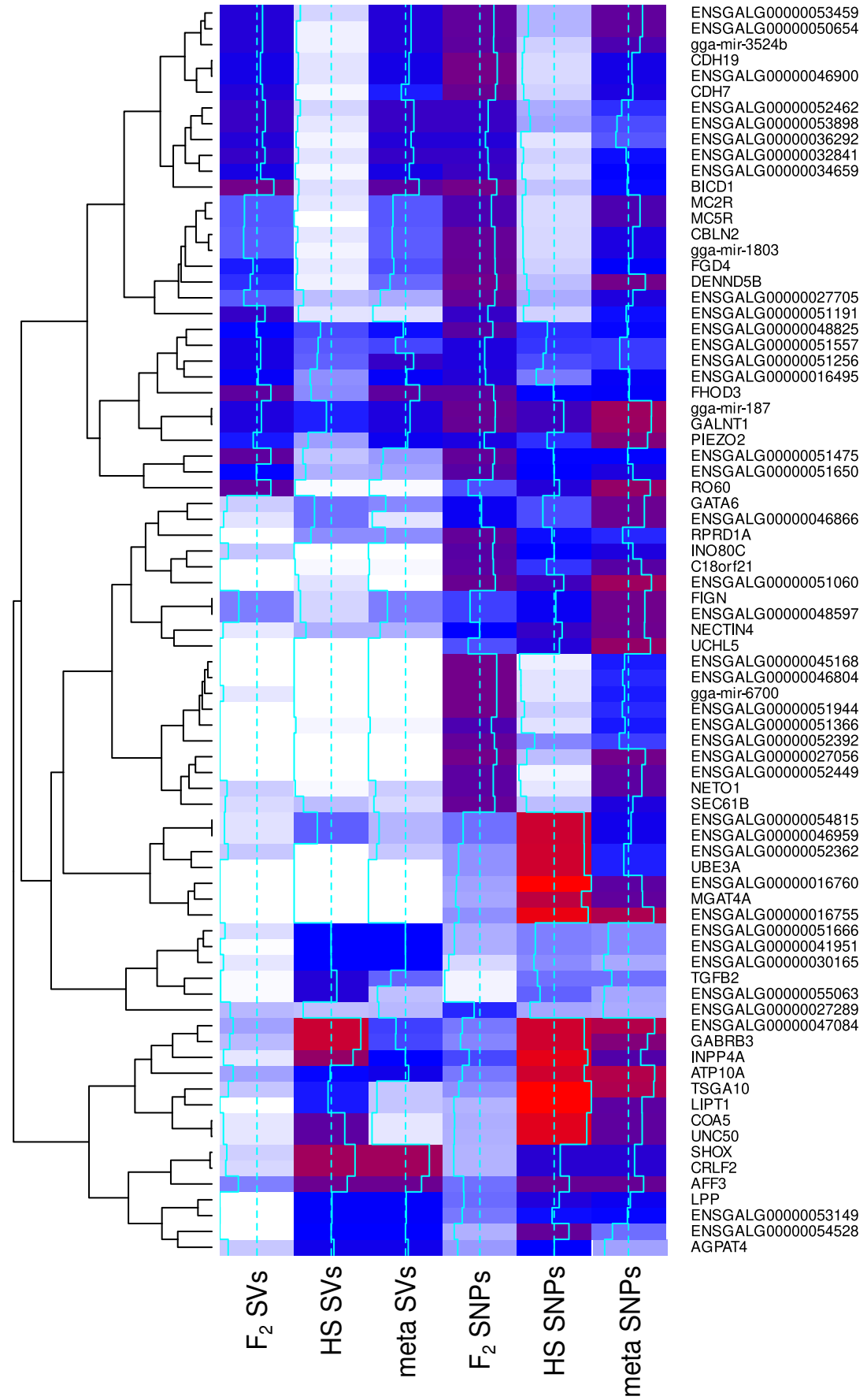
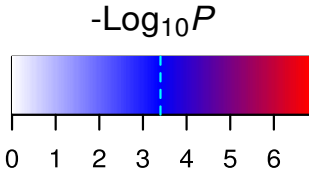
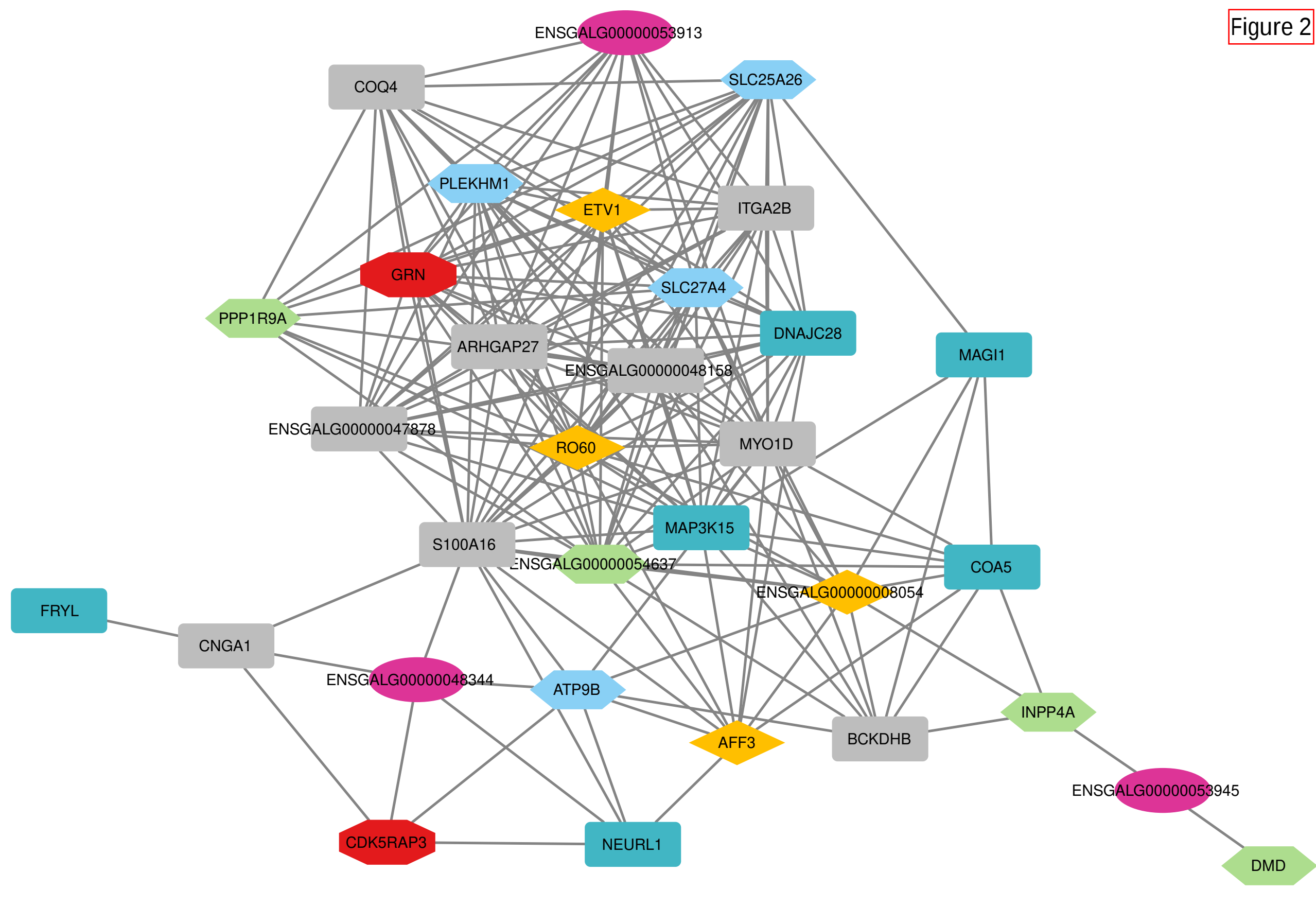
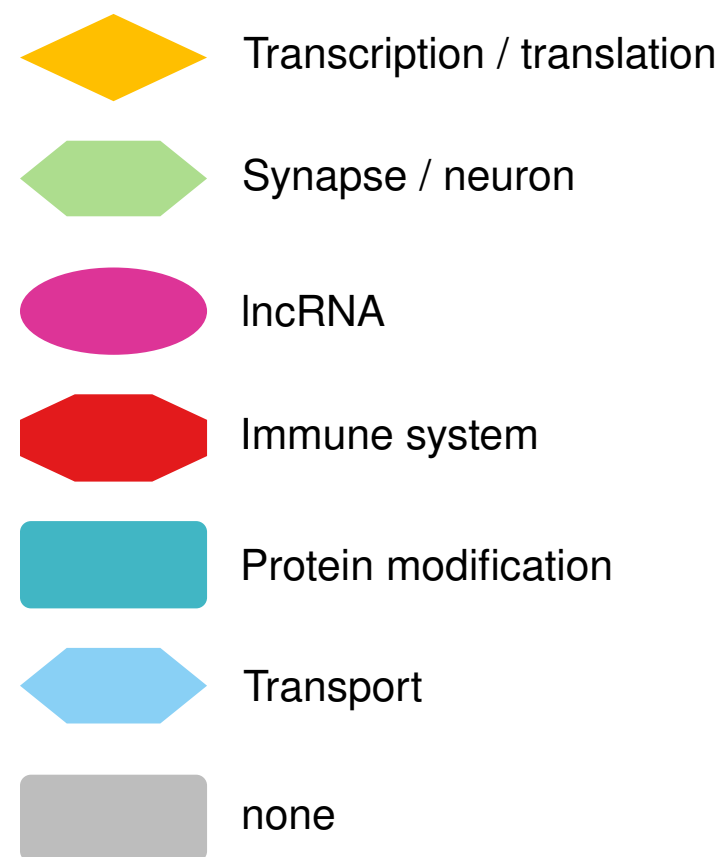


Figure 2



bioRxiv preprint doi: <https://doi.org/10.1101/2022.09.16.508210>; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

