

## ORIGINAL ARTICLE

# Association of *FAAH* p.Pro129Thr and *COMT* p.Ala72Ser with schizophrenia and comorbid substance use through next-generation sequencing: an exploratory analysis

José J. Martínez-Magaña,<sup>1,2</sup> Alma D. Genis-Mendoza,<sup>2,3</sup>  Vanessa González-Covarrubias,<sup>4</sup> Isela E. Juárez-Rojop,<sup>1</sup> Carlos A. Tovilla-Zárate,<sup>5</sup> Xavier Soberón,<sup>4</sup> Nuria Lanzagorta,<sup>6</sup> Humberto Nicolini<sup>1,6</sup> 

<sup>1</sup>División Académica de Ciencias de la Salud, Universidad Juárez Autónoma de Tabasco, Villahermosa, Mexico. <sup>2</sup>Laboratorio de Genómica de Enfermedades Psiquiátricas y Neurodegenerativas, Instituto Nacional de Medicina Genómica, Ciudad de México, Mexico. <sup>3</sup>Servicios de Atención Psiquiátrica, Hospital Psiquiátrico Infantil Juan N. Navarro, Ciudad de México, Mexico. <sup>4</sup>Laboratorio de Farmacogenómica, Instituto Nacional de Medicina Genómica, Ciudad de México, Mexico. <sup>5</sup>División Multidisciplinaria de Comalcalco, Universidad Juárez Autónoma de Tabasco, Comalcalco, Mexico. <sup>6</sup>Grupo de Estudios Médicos y Familiares Carracci, Ciudad de México, Mexico.

**Objective:** Individuals with schizophrenia and substance use disorders have a poor prognosis and increased psychiatric symptoms. The present study aimed to explore the association of 106 genes in individuals with schizophrenia and comorbid substance use through a next-generation sequencing (NGS) analysis and different *in silico* algorithms.

**Methods:** We included 105 individuals diagnosed with schizophrenia and a family history of schizophrenia, of whom 49 (46.67%) presented comorbid substance use. Using NGS, we sequenced 106 genes previously associated with schizophrenia. Logistic regression models were used to assess differences in allele frequencies, and a generalized gene-set analysis was performed at the gene level. Functional annotations were performed using different algorithms and databases.

**Results:** We identified a total of 3,109 variants, of which 25 were associated with schizophrenia and comorbid substance use and were located in regulatory and coding regions. We found low-frequency variants in *COMT* p.Ala72Ser, independently of p.Val158Met, that were associated with substance use. The endocannabinoid functional variant *FAAH* p.Pro129Thr was also associated with substance use.

**Conclusions:** Genetic variants of genes related to dopaminergic and cannabinoid neurotransmitter systems were associated with comorbid substance use in schizophrenia. Nevertheless, more studies with larger sample sizes are needed to confirm our findings.

**Keywords:** Dual diagnosis; next-generation sequencing; Mexican population; schizophrenia; p.Pro129Thr *FAAH*; p.Ala72Ser *COMT*

## Introduction

Epidemiological studies have shown a high comorbidity of substance use in schizophrenia.<sup>1-7</sup> Individuals affected by this comorbidity present increased psychotic symptoms, more severe psychiatric symptoms, a higher suicide risk, and sometimes decreased treatment response.<sup>8-11</sup> The etiology behind this comorbidity has not been widely studied, although some explanatory models have emerged.<sup>12,13</sup> One of these models is diathesis-stress, in which affected individuals could have a higher genetic risk that prompts this dual morbidity under stressful conditions.<sup>12</sup> Genomic studies aiming to determine the molecular basis of this comorbidity have focused on common genetic variations, considering

variants with a population minor allele frequency (MAF) > 1% and one phenotype in substance use disorder or schizophrenia. These studies have identified correlations between the genetic risk of schizophrenia and substance use and have selected genetically correlated *loci* as a proxy for the genetic risk of comorbidity.<sup>14-18</sup> Schizophrenia-associated *loci* are highly correlated with substance use, and individuals with a higher genetic risk also have a higher prevalence of schizophrenia according to genetic risk scores (i.e., the score is estimated by the schizophrenia-associated variants an individual carries, pondered by the estimated effect in genome-wide association studies).<sup>17,19</sup> Therefore, the variants used to define the genetic score for schizophrenia are also of great interest due to their association

Correspondence: Humberto Nicolini, Instituto Nacional de Medicina Genómica, Laboratorio de Genómica de las Enfermedades Psiquiátricas y Neurodegenerativas, Periférico Sur 4809, Arenal Tepepan, 14610, Tlalpan, Ciudad de México, México.  
E-mail: hnicolini@inmegen.gob.mx  
Submitted Oct 04 2020, accepted Mar 11 2021, Epub May 21 2021.

**How to cite this article:** Martínez-Magaña JJ, Genis-Mendoza AD, González-Covarrubias V, Juárez-Rojop IE, Tovilla-Zárate CA, Soberón X, et al. Association of *FAAH* p.Pro129Thr and *COMT* p.Ala72Ser with schizophrenia and comorbid substance use through next-generation sequencing: an exploratory analysis. Braz J Psychiatry. 2022;44:164-170. <http://dx.doi.org/10.1590/1516-4446-2020-1546>

with comorbid substance use. Additionally, higher genetic scores have been associated with the severity of both positive and negative schizophrenia symptoms. Thus, schizophrenia research that correlates substance use and schizophrenia symptoms could help us define the characteristics most closely associated with these variants.

The heritability of schizophrenia and comorbid substance use has been evaluated through genome-wide association studies that have analyzed common genetic variants.<sup>20</sup> However, since they characterize only a small proportion of heritability, it is a phenotype with large "missing heritability."<sup>21-23</sup> Furthermore, most genome-wide association studies cannot account for all worldwide variation. Rare genetic variants have also been proposed as part of this missing heritability,<sup>24</sup> and exploring allele differences among populations is more likely to detect population stratification and pinpoint relevant variants. Rare or population-specific variations can be detected by next-generation sequencing (NGS).<sup>25</sup> A previous limitation of rare variant detection by NGS was the lack of population-based references to compare allelic frequencies. This limitation has been overcome by international studies, including the Genome Aggregation Database (GnomAD).<sup>25</sup> The GnomAD can be used as reference to determine whether a genetic variant found in a particular phenotype by NGS is rare in the population in order to estimate its possible effect. The aims of the present study were: 1) to use NGS to explore common and rare genetic variations of 106 genes previously associated with schizophrenia; 2) to explore the association of these genetic variations with schizophrenia and comorbid substance use; and 3) to identify whether these genetic variations are potentially associated with clinical data such as positive or negative symptoms and age of onset in a sample of Mexican individuals diagnosed with schizophrenia.

## Methods

### Population

A total of 105 individuals were included in the present study. We only included individuals diagnosed with schizophrenia who also had a first-degree relative (parents or siblings) affected with schizophrenia. All participants had to be of Mexican descent for at least three generations. The participants were recruited at the Hospital Regional de Alta Especialidad Dr. Gustavo Rivirosa Perez and at the Grupo de Estudios Médicos y Familiares Carracci. The participants were evaluated using the Diagnostic Interview for Genetic Studies, and the diagnosis was determined by two psychiatrists using the mean of the best estimate. The diagnostic criteria were based on the DSM-IV-R. Comorbid substance use was defined as any of the following criteria: 1) illegal drug abuse/dependence; 2) alcohol abuse/dependence; and/or 3) tobacco abuse/dependence (Table 1). All 105 participants met the criteria for schizophrenia, and 49 also presented comorbid substance use.

### Next-generation sequencing

Genomic DNA was extracted from peripheral leukocytes using a Gentra Puregene kit (QIAGEN, Germantown, MD,

**Table 1** Clinical characteristics of the sample

	Sample (n = 105)
Sex	
Male	62 (59.05)
Female	43 (40.95)
Scale for the Assessment of Negative Symptoms	36.85 (34.38)
Scale for the Assessment of Positive Symptoms	16.05 (26.03)
Age (years), mean (SD)	33.28 (9.09)
Schizophrenia onset (years), mean (SD)	9.76 (13.75)
Substance comorbidity	49 (46.67)
Alcohol abuse/dependence	42 (40.00)
Tobacco abuse/dependence	33 (31.43)
Cannabis abuse/dependence	2 (1.90)
Cocaine abuse/dependence	5 (4.76)
Stimulant abuse/dependence	4 (3.81)

Data presented as n (%), unless otherwise specified.  
SD = standard deviation.

USA). Synthetic probes for NGS were designed to target genes associated with schizophrenia. This gene panel, which has been previously described by our group,<sup>26</sup> included genes that encode all the dopaminergic, serotonergic, and GABAergic receptors, as well as enzymes involved in the synthesis of these neurotransmitters. Gene capture was performed using a HaloPlex target enrichment system (Agilent Technologies, Wilmington, DE, USA). Sequencing libraries were generated according to manufacturer protocols. In short, all DNA samples were digested with eight-paired restriction enzymes and the fragmentation pattern was analyzed in a 2100 Bioanalyzer (Agilent Technologies). DNA fragments were hybridized with HaloPlex synthetic probes for library enrichment, and adapters were ligated by polymerase chain reaction. The sequencing was performed using a NextSeq500 system (Illumina, San Diego, CA, USA) aiming for a depth coverage of 200x in paired-end reads. For sequence quality control we used Trimmomatic to eliminate reads with poor quality scores (Phred-QS < 25.0). According to general practice, lengths shorter than 55 bp, indexes, adapters, and 5 bp at both read ends were trimmed. We then aligned the reads to the human genome (build: GRCh37/hg19) using BWA. InDel realignment, base recalibration, and variant calling were performed following best practices for GATK. Haplotype-Caller was used for single-nucleotide variant (SNV) detection.

### Statistical analysis

Allele frequency comparisons between individuals with comorbid substance use and those with schizophrenia alone were performed with a logistic regression model using PLINK (i.e., variant level analyses).<sup>27</sup> Single nucleotide polymorphism-wise generalized gene-set analyses were performed using MAGMA software.<sup>28</sup> For the latter, we assessed statistical power using PAGEANT<sup>29</sup> for an alpha of 0.05 and a power of 45.2%. We considered p-values < 0.05 for genetic variants statistically significant. Significant variants were functionally annotated using the ENCODE database<sup>30</sup> and GnomAD.<sup>25</sup> We used a variant effect predictor to define coding frames and allelic

frequencies.<sup>31</sup> We developed an in-house script to search for continuous associated variants (peak of associations) in each regulatory region. 3'-UTR variants were searched in the MirSNP database<sup>32</sup> for potential microRNAs binding sites, and pathway enrichment was conducted using miRPath.<sup>33</sup> Finally, we developed complementary linear regression models to identify potential associations between clinical data, including positive and negative symptoms (measured by the Scale for the Assessment of Positive Symptoms and the Scale for the Assessment of Negative Symptoms), as well as the age of disease onset. The associations were also evaluated in PLINK for variant level analysis and in MAGMA for gene-set analysis; genetic variants with  $p < 0.05$  were considered statistically significant.

### Ethics statement

Every participant provided written informed consent prior to inclusion. All protocols were performed according to Declaration of Helsinki guidelines. The protocols were revised and approved by the ethics and research committees of the Instituto Nacional de Medicina Genómica (no. IMG/DI/136/2014).

## Results

### Next-generation sequencing summary

We identified 3,109 genetic variants in 106 genes. Of these, 2,855 were SNVs and 254 were short insertion/deletion variants. Of the SNVs, 2,773 (89.19%) were intronic and 336 were exonic. Of the exonic variants, only 80 (2.57%) were missense variants and nine were predicted to be damaging; 308 (9.91%) were annotated as regulatory and 78 (2.51%) were novel variants (not previously reported). The genes with the highest number of novel variants were *CYP2D6* ( $n=4$ ) and *CYP2D7* ( $n=4$ ). We also observed that the extremely rare missense variant p.Ser1767Asn was overrepresented (sample MAF = 0.09) in the *TET2* gene (ENST00000540549.1, MAF = 6.4e-06).

### Gene-based associations with substance use

In the gene-based analysis, we identified seven genes associated with substance use in individuals diagnosed with schizophrenia: catechol-O-methyltransferase gene (*COMT*) ( $z = 1.8666$ ,  $p\text{-value} = 0.0309$ ), cannabinoid receptor 1 gene (*CNR1*) ( $z = 2.7928$ ,  $p\text{-value} = 0.0026$ ), dopamine transporter gene (*SLC6A3*) ( $z = 1.7823$ ,  $p\text{-value} = 0.0374$ ), flavin containing dimethylaniline monooxygenase 2 gene (*FMO2*) ( $z = 1.7443$ ,  $p\text{-value} = 0.0406$ ), potassium voltage-gated channel subfamily H member 2 gene (*KCNH2*) ( $z = 2.2480$ ,  $p\text{-value} = 0.0123$ ), monoamine oxidase B gene (*MAOB*) ( $z = 2.8369$ ,  $p\text{-value} = 0.00223$ ), and thiopurine S-methyltransferase gene (*TPMT*) ( $z = 1.7962$ ,  $p\text{-value} = 0.0362$ ). Pathway analysis showed that these genes were enriched for dopamine clearance from the synaptic cleft (hsa-379401, adjusted  $p\text{-value} = 0.0000$ ). The pathway involves four genes (*SLC6A3*, *MAOB*, *TPMT* and

*COMT*) and we found that two of them were associated with substance use.

### Variant level and functional analysis

The intron variant rs2119767 in the *GABRA2* gene (ENST00000510861.1) showed the strongest association with substance use. The MAF in individuals with schizophrenia and comorbid substance use was 0.5957, while in schizophrenia alone it was 0.3021 ( $p = 7.303e-05$ ); no regulatory element has been reported for this variant.

We also found three missense variants associated with substance use, p.Pro129Thr in the fatty acid amide hydrolase gene (*FAAH*), *FMO2* p.Glu314Gly, and *COMT* p.Ala72Ser (Table 2). The *FMO2* and *COMT* genes were also significantly associated in the gene-based analysis. *COMT* p.Ala72Ser was not in linkage disequilibrium compared with the most studied missense common genetic variant in *COMT*, p.Val158Met ( $r^2 = 0.0466$ ), and this variant was not associated with substance use in our study ( $p\text{-value} = 0.0785$ ).

We found 22 variants associated with substance use in annotated regulatory elements (Table 3): nine were annotated to CCCTC-binding factor (CTCF) or transcription factor binding sites, nine were annotated to microRNAs binding sites, one was annotated to open chromatin, and two were annotated to promoter-associated variants. Variants associated with CTCF binding-sites were found in *SLC6A3* and *KCNH2*; both genes were also identified in gene-based analyses. The associated promoter variants were found in the *NRG1* gene.

Variants significantly associated with substance use that were annotated to microRNAs were clustered in fatty acid biosynthesis (hsa00061,  $p\text{-value} < 1.8337e-14$ ) and Parkinson's disease (hsa05012,  $p\text{-value} < 2.1126e-05$ ) pathways. The microRNAs hsa-miR-182-5p and hsa-miR-1293 were associated with the fatty acid biosynthesis pathway. They were bound to the 3'-UTR regions in the *FMO2* gene and the solute carrier organic anion transporter family member 3A1 (*SLCO3A1*). hsa-miR-583 has been related to the Parkinson's disease pathway and seems to be enriched at the 3'-UTR of *MAOB*, a gene associated with substance use in this genetic analysis.

### Genetic association with schizophrenia and clinical data

To determine whether the genetic variants linked to substance use were also associated with schizophrenia symptoms, we developed regression models that accounted for clinical data, including positive and negative symptoms and age of onset. We found that positive symptoms were significantly associated with the variants rs61759708 of *ADRA1A* (Beta = 25.68,  $p\text{-value} = 1.69e-06$ ) and rs17333700 of *ADRA1A* gene (Beta = -6.593,  $p\text{-value} = 1.31e-05$ ). Negative symptoms were significantly associated with rs9147 and rs2088102 variants of the *AKR1A1* gene (Beta = 17.21,  $p\text{-value} = 4.79e-05$ , and Beta = 16.92,  $p\text{-value} = 1.68e-05$ ). On the other hand, variants rs12901528 and rs12901348 of the *GABRB3* gene were significantly associated with age of onset (Beta = 3.08,  $p\text{-value} = 2.75e-04$  and Beta = 3.00,  $p\text{-value} =$

**Table 2** Missense variants associated with comorbid substance use

Position (hg19)	Gene	Type	SNP	Ref/alt	Freq. com	Freq. no com	p-value	All freq	GnomAD	Effect	PolyPhen	SIFT
chr1:46870761	FAAH	SNV	rs324420	C/A	0.3936	0.2404	0.0222	0.3131	0.2405	p.Pro129Thr (ENST00000243167.8)	Probable Damaging	Tolerated
chr1:171174531	FMO2	SNV	rs2020863	A/G	0.0889	0.2444	0.0085	0.1667	0.1313	p.Glu314Gly (ENST00000441535.1)	Probable Damaging	Deleterious
chr22:19950263	COMT	SNV	rs6267	G/T	0.0204	0.09804	0.0335	0.0600	0.0136	p.Ala72Ser (ENST00000361682.6)	Possibly Damaging	Deleterious

All freq = frequency in the overall sample; Freq. com = frequency on individuals with substance use comorbidity; Freq. no com = frequency on individuals without substance use comorbidity; GnomAD = Genome Aggregation Database; PolyPhen = Polymorphism Phenotyping tool; Ref/alt = reference and alternative alleles; SIFT = Sorting Intolerant from Tolerant algorithm; SNP = single nucleotide polymorphisms; SNV = single-nucleotide variants.

**Table 3** Regulatory variants associated with comorbid substance use

Position (hg19)	Gene	Type	SNP	Ref/alt	Freq. com	Freq. no com	p-value	All freq	GnomAD	Transcript	Effect
chr1:171178257-171178258	FMO2	InDel	rs2234889	C/CT	0.0682	0.1979	0.0165	0.1359	0.0979	3'-UTR (ENST00000441535.1)	hsa-miR-182-5p/hsa-miR-3613-3p/hsa-miR-4715-5p
chr1:171178490	FMO2	SNV	rs28369914	C/T	0.0851	0.2400	0.0038	0.1649	0.0986	3'-UTR (ENST00000441535.1)	hsa-miR-3545-3p/hsa-miR-3688-3p
chr1:171179351	FMO2	SNV	rs16864194	A/T	0.0778	0.2447	0.0025	0.1630	0.0991	3'-UTR (ENST00000441535.1)	hsa-miR-3942-3p
chr5:1420306	SLC6A3	SNV	rs28382248	G/A	0.1809	0.3039	0.0481	0.2449	0.2112	Intron (ENST00000270349.9)	CTCF BS (ENSR00000402282)
chr5:1420346	SLC6A3	SNV	rs28382247	G/T	0.1809	0.3039	0.0481	0.2449	0.2152	Intron (ENST00000270349.9)	CTCF BS (ENSR00000402282)
chr5:1420476	SLC6A3	SNV	rs28382245	C/T	0.1915	0.0769	0.0206	0.1313	0.2732	Intron (ENST00000270349.9)	CTCF BS (ENSR00000402282)
chr6:18128659	TPMT	SNV	rs7886	T/A	0.2447	0.1154	0.0244	0.1768	0.2114	3'-UTR (ENST00000309983.4)	hsa-miR-4789-3p/hsa-miR-642a-3p/hsa-miR-642b-3p
chr6:18128999	TPMT	SNV	rs1142378	T/G	0.2553	0.1100	0.0093	0.1804	0.2091	3'-UTR (ENST00000309983.4)	
chr7:150645682	KCNH2	SNV	rs2968860	G/A	0.2889	0.4796	0.0107	0.3883	0.3765	Intron (ENST00000262186.5)	TF and CTCF BS (ENSR00000219754, ENSR00000414356)
chr7:150647569	KCNH2	SNV	rs4725983	T/C	0.3205	0.5333	0.0078	0.4345	0.3179	Intron (ENST00000262186.5)	Open Chromatin (ENSR00000219755)
chr7:150649603	KCNH2	SNV	rs740852	G/A	0.4333	0.2900	0.04891	0.3579	0.2988	Synonymous (ENST00000262186.5)	CTCF BS (ENSR00000414357)
chr7:150655624	KCNH2	SNV	rs758891	T/C	0.4318	0.2692	0.0223	0.3438	0.3899	Intron (ENST00000262186.5)	TF BS (ENSR00000219757)
chr7:150655643	KCNH2	SNV	rs758890	G/A	0.4872	0.3182	0.0385	0.3976	0.4172	Intron (ENST00000262186.5)	TF BS (ENSR00000219757)
chr7:150657095	KCNH2	SNV	rs56282717	G/A	0.1477	0.0556	0.0487	0.1011	0.1534	Intron (ENST00000262186.5)	CTCF BS (ENSR00000414361)
chr7:150657201	KCNH2	SNV	rs12668582	A/C	0.4886	0.2708	0.0036	0.3750	0.4222	Intron (ENST00000262186.5)	CTCF BS (ENSR00000414361)
chr8:32406148	NRG1	SNV	rs7834206	C/A	0.5319	0.3700	0.0302	0.4485	0.3948	5'-UTR (ENST00000356819.4)	Promoter (ENSR00000330589)
chr8:32406382	NRG1	SNV	rs73234136	T/C	0.5426	0.3600	0.0139	0.4485	0.4295	5'-UTR (ENST00000356819.4)	Promoter (ENSR00000330589)
chr15:92715167	SLC3A1	SNV	rs2270060	C/T	0.5532	0.3679	0.0105	0.4550	0.3928	3'-UTR (ENST00000424469.2)	hsa-miR-1321, hsa-miR-4270, hsa-miR-765
chr15:92715131	SLC3A1	SNV	rs2270061	A/T	0.5761	0.4000	0.0205	0.4844	0.4035	3'-UTR (ENST00000424469.2)	hsa-miR-1293, hsa-miR-1587, hsa-miR-4417, hsa-miR-4446-3p, hsa-miR-4463, hsa-miR-4721, hsa-miR-541-3p, hsa-miR-654-5p
chrX:43626085	MAOB		rs3027438	G/A	0.0106	0.1078	0.0054	0.0612	0.0207	3'-UTR (ENST00000378069.4)	hsa-miR-181c-3p
chrX:43626136	MAOB		rs3027439	G/A	0.0106	0.1346	0.0008	0.0758	0.0488	3'-UTR (ENST00000378069.4)	hsa-miR-3173-3p, hsa-miR-3689d, hsa-miR-4668-5p, hsa-miR-583
chrX:43626329	MAOB		rs2072745	T/A	0.0104	0.1300	0.0013	0.0714	0.0509	3'-UTR (ENST00000378069.4)	hsa-miR-4511

All freq = frequency in the overall sample; Freq. com = frequency in individuals with substance use comorbidity; Freq. no com = frequency in individuals without comorbid substance use; GnomAD = Genome Aggregation Database; Ref/alt = reference and alternative alleles; SNP = single nucleotide polymorphisms; SNV = single-nucleotide variants.

3.87e-04). There was a significant association between these six variants and schizophrenia symptoms or age of onset, but not substance use. In contrast, the variants rs73234136 and rs7834206 of *NRG1* were significantly associated with age of onset in regression models for schizophrenia and comorbid substance use.

## Discussion

The genomics of schizophrenia and substance use have mostly focused on one phenotype, and their co-occurrence has not been fully investigated. We investigated the potential association of schizophrenia and comorbid substance use in 106 schizophrenia-associated genes through targeted NGS. Our results revealed that this comorbidity has an association with genes related to the dopamine clearance pathway, including *SLC6A3*, *MAOB* and *COMT*. This is in agreement with a study that reported an acute increase of dopamine neurotransmission in the striatum in the presence of addictive drugs.<sup>34</sup> On the other hand, the main effects of addictive drugs show three major mechanisms: 1) neural firing; 2) reuptake mostly by the dopamine transporter; and 3) altered dopamine release. We inferred that the variants we identified in these three genes are associated with at least two of the above-mentioned mechanisms: dopamine transporter reuptake is associated with *SLC6A3* gene, while dopamine release is associated with *MAOB* and *COMT* genes. It is well known that the *SLC6A3* gene codes for the dopamine transporter; however, there is conflicting evidence regarding the transporter and its relation to substance dependence.<sup>35-39</sup> Some authors have proposed that these conflicting associations are due to the genetic position of the variants, since some are located in the first exons while others are in the 3'-UTR region, which affects either expression or proper transcription, hence the functional effect of the protein.<sup>35</sup> Our targeted sequencing approach was capable of detecting genetic variations across whole genes due to the fact that we included at least 50 bp of untranslated regions for sequencing. The association between substance use and intron 6 of *SLC6A3* has been reported to have a regulatory impact, promoted by the union of transcriptional repressor CTCF. CTCF is a zinc finger protein that acts as an insulator, mediating chromatin looping and attracting a diversity of tissue-specific transcription factors.<sup>40</sup> Other CTCF binding sites associated with substance use variants were found in introns 2 and 3 of *KCNH2*, a key gene associated with long QT syndrome. Although this gene has not been associated with addictive drug use, it has been reported to be involved in schizophrenia and cognition.<sup>41</sup>

Our regression analysis detected variants in the *COMT* and *MAOB* genes that are associated with schizophrenia and comorbid substance use. We hypothesized that those variants could affect dopamine release, because both *COMT* and *MAOB* genes code for the enzymes involved in dopamine degradation.<sup>42-45</sup> We observed a variant of low allele frequency in *COMT*, p.Ala72Ser (MAF: 0.0136), and it was not in linkage disequilibrium with *COMT* p.Val158Met, one of the most studied variants.<sup>46</sup>

It is possible that there are only a few reports about the variant p.Ala72Ser, since it is not always present in commercial microarrays. Thus, it has not been previously associated with psychiatric disorders, which highlights the relevance of NGS technology to better characterize diseases. Similarly, the *MAOB* gene, which also codes for an enzyme involved in dopamine catabolism, was associated with substance use and was present in the 3'-UTR region, showing a global MAF of 0.02 to 0.05. It is likely that this gene has also been underexplored. The mechanisms by which this variant may impact *MAOB* function and expression and ultimately alter dopamine catabolism in psychiatric disorders remain unknown and warrant further investigation.

The substance use analysis revealed that variants with microRNA binding-sites could be associated with microRNAs involved in the modulation of fatty acid biosynthesis. Fatty acid modulation has been related to addiction, especially cannabis dependence.<sup>47-49</sup> Previously reported associations have focused on how the *FAAH* gene affects the metabolism of cannabinoids. *FAAH* was first described as the enzyme responsible for metabolizing endogenous fatty acid amines, including anandamide.<sup>50</sup> Anandamide is an agonist of cannabinoid receptor 1 (*CNR1*), another gene found to be associated with substance use in this study. We found an association between the variant *FAAH* p.Pro129Thr, a functional missense variant, with substance use in individuals with schizophrenia. This variant, which has been previously associated with cannabis dependence,<sup>51</sup> promotes a reduction in *FAAH* activity, consequently enhancing the endocannabinoid system.<sup>51-53</sup> It has also been associated with alcohol and cocaine addiction,<sup>47</sup> which indicates that this gene has a pleiotropic effect in addictive behavior.

We are aware of the limitations of our study given its sample size. For example, the statistical power for genetic associations was estimated to be around 50%, which prompted us to classify our observations as preliminary. Thus, we are working on increasing the sample size and recruiting a replication cohort to confirm these associations. Nevertheless, we obtained NGS data with > 100x sequencing depth, which has contributed to the collection of genetic variants in admixed psychiatric individuals. Despite the small power of this initial study, using NGS to search for associations of variants present in schizophrenia-associated loci could help better characterize variants and replicate them in other samples as a fine-mapping approach. Additionally, we found genetic variants that have scarcely been studied. Moreover, the association of at least six variants in four genes could improve our understanding of schizophrenia with comorbid substance use by including variants functionally relevant to this clinical phenotype.<sup>54</sup> The lack of a non-psychotic group prevented us from detecting genetic variants associated exclusively with substance use; nevertheless, since we included variants previously associated with schizophrenia, we were able to determine whether these variants were associated with subtraits of individuals diagnosed with schizophrenia (substance use, positive and negative symptoms, etc.). To further explore this, we developed additional regression models to

identify potential associations with positive-negative symptoms, as well as with the age of schizophrenia onset, which may or may not be associated with substance use in individuals with schizophrenia. Interestingly, none of the variants significantly associated with substance use were associated with positive or negative symptoms of schizophrenia; although we identified variants of the *NRG1* gene relevant to age of schizophrenia onset that were present in the associations with substance use.

In conclusion, we found an association between variants in genes of the dopamine (*SLC6A3*, *MAOB*, and *COMT*) and cannabinoid pathways (*CNR1* and *FAAH*) in individuals with schizophrenia and comorbid substance use. We consider our observations preliminary and will seek to validate them through replication in a larger sample.

### Acknowledgements

We are deeply grateful to Alfredo Mendoza Vargas from the Sequencing Unit at the Instituto Nacional de Medicina Genómica for his technical assistance during the next-generation sequencing (NGS) analysis. This study received funding from the Instituto Nacional de Medicina Genómica (grant no. IMG/DI/136/2014).

### Disclosure

The authors report no conflicts of interest.

### References

- Jané-Llopis E, Matytsina I. Mental health and alcohol, drugs and tobacco: a review of the comorbidity between mental disorders and the use of alcohol, tobacco and illicit drugs. *Drug Alcohol Rev.* 2006;25:515-36.
- Munro I, Edward KL. Mental illness and substance use: an Australian perspective. *Int J Ment Health Nurs.* 2008;17:255-60.
- Vega WA, Sribney WM, Achara-Abrahams I. Co-occurring alcohol, drug, and other psychiatric disorders among Mexican-origin people in the United States. *Am J Public Health.* 2003;93:1057-64.
- Marín-Navarrete R, Medina-Mora ME, Horigian VE, Salloum IM, Vilalobos-Gallegos L, Fernández-Mondragón J. Co-occurring disorders: a challenge for Mexican community-based residential care facilities for substance use. *J Dual Diagn.* 2016;12:261-70.
- Kessler RC. The epidemiology of dual diagnosis. *Biol Psychiatry.* 2004;56:730-7.
- Medina-Mora ME, Borges G, Benjet C, Lara C, Berglund P. Psychiatric disorders in Mexico: lifetime prevalence in a nationally representative sample. *Br J Psychiatry.* 2007;190:521-8.
- Szerman N, Martínez-Raga J, Baler R, Roncero C, Vega P, Basurte I, et al. Joint statement on dual disorders: addiction and other mental disorders. *Salud Ment.* 2017;40:245-7.
- Farren CK, Mc Elroy S. Treatment response of bipolar and unipolar alcoholics to an inpatient dual diagnosis program. *J Affect Disord.* 2008;106:265-72.
- Gimelfarb Y, Natan Z. [Risk factors for suicide attempts in dual diagnosis patients]. *Harefuah.* 2009;148:355-8, 413.
- Oquendo MA, Currier D, Liu SM, Hasin DS, Grant BF, Blanco C. Increased risk for suicidal behavior in comorbid bipolar disorder and alcohol use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *J Clin Psychiatry.* 2010;71:902-9.
- Wusthoff LE, Waal H, Grawe RW. The effectiveness of integrated treatment in patients with substance use disorders co-occurring with anxiety and/or depression--a group randomized trial. *BMC Psychiatry.* 2014;14:67.
- Polimanti R, Agrawal A, Gelernter J. Schizophrenia and substance use comorbidity: a genome-wide perspective. *Genome Med.* 2017;9:25.
- Mueser KT, Drake RE, Wallach MA. Dual diagnosis: a review of etiological theories. *Addict Behav.* 1998;23:717-34.
- Pasman JA, Verweij KJ, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci.* 2018;21:1161-70.
- Otto JM, Gizer IR, Bizon C, Wilhelmsen KC, Ehlers CL. Polygenic risk scores for cigarettes smoked per day do not generalize to a Native American population. *Drug Alcohol Depend.* 2016;167:95-102.
- Carey CE, Agrawal A, Bucholz KK, Hartz SM, Lynskey MT, Nelson EC, et al. Associations between polygenic risk for psychiatric disorders and substance involvement. *Front Genet.* 2016;7:149.
- Hartz SM, Horton AC, Oehlert M, Carey CE, Agrawal A, Bogdan R, et al. Association between substance use disorder and polygenic liability to schizophrenia. *Biol Psychiatry.* 2017;82:709-15.
- Andersen AM, Pietrzak RH, Kranzler HR, Ma L, Zhou H, Liu X, et al. Polygenic scores for major depressive disorder and risk of alcohol dependence. *JAMA Psychiatry.* 2017;74:1153-60.
- Mallard TT, Harden KP, Fromme K. Genetic risk for schizophrenia is associated with substance use in emerging adulthood: an event-level polygenic prediction model. *Psychol Med.* 2019;49:2027-35.
- Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet.* 2017;100:635-49.
- Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet.* 2011;88:294-305.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461:747-53.
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010;11:446-50.
- 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015;526:68-74.
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581:434-43.
- Martínez-Magaña JJ, Genís-Mendoza AD, González-Covarrubias V, Jiménez-Guenchi J, Galindo-Chávez AG, Roche-Bergua A, et al. Exploratory analysis of rare and novel variants in Mexican patients diagnosed with schizophrenia and dementia. *Rev Invest Clin.* 2019;71:246-54.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
- de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* 2015;11:e1004219.
- Derkach A, Zhang H, Chatterjee N. Power Analysis for Genetic Association Test (PAGEANT) provides insights to challenges for rare variant association studies. *Bioinformatics.* 2018;34:1506-13.
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for the ENCODE Project. *Genome Res.* 2012;22:1760-74.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol.* 2016;17:122.
- Liu C, Zhang F, Li T, Lu M, Wang L, Yue W, et al. MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics.* 2012;13:661.
- Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, et al. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.* 2015;43:W460-6.
- Sulzer D. How addictive drugs disrupt presynaptic dopamine neurotransmission. *Neuron.* 2011;69:628-49.
- Huang CC, Kuo SC, Yeh YW, Chen CY, Yen CH, Liang CS, et al. The *SLC6A3* gene possibly affects susceptibility to late-onset alcohol dependence but not specific personality traits in a Han Chinese population. *PLoS One.* 2017;12:e0171170.

- 36 Hong CJ, Cheng CY, Shu LR, Yang CY, Tsai SJ. Association study of the dopamine and serotonin transporter genetic polymorphisms and methamphetamine abuse in Chinese males. *J Neural Transm (Vienna)*. 2003;110:345-51.
- 37 Lind PA, Eriksson CJ, Wilhelmsen KC. Association between harmful alcohol consumption behavior and dopamine transporter (DAT1) gene polymorphisms in a male Finnish population. *Psychiatr Genet*. 2009;19:117-25.
- 38 Xu M, Lin Z. Genetic influences of dopamine transport gene on alcohol dependence: a pooled analysis of 13 studies with 2483 cases and 1753 controls. *Prog Neuro-Psychopharmacology Biol Psychiatry*. 2011;35:1255-60.
- 39 Ma Y, Fan R, Li MD. Meta-analysis reveals significant association of the 3'-UTR VNTR in SLC6A3 with alcohol dependence. *Alcohol Clin Exp Res*. 2016;40:1443-53.
- 40 Holwerda SJ, de Laat W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci*. 2013;368:20120369.
- 41 Huffaker SJ, Chen J, Nicodemus KK, Sambataro F, Yang F, Mattay V, et al. A primate-specific, brain isoform of KCNH2 affects cortical physiology, cognition, neuronal repolarization and risk of schizophrenia. *Nat Med*. 2009;15:509-18.
- 42 Montag C, Jurkiewicz M, Reuter M. The role of the catechol-o-methyltransferase (COMT) gene in personality and related psychopathological disorders. *CNS Neurol Disord Drug Targets*. 2012; 11:236-50.
- 43 Rosengren E. On the role of monoamine oxidase for the inactivation of dopamine in brain. *Acta Physiol Scand*. 1960;49:370-5.
- 44 Tunbridge EM, Bannerman DM, Sharp T, Harrison PJ. Catechol-o-methyltransferase inhibition improves set-shifting performance and elevates stimulated dopamine release in the rat prefrontal cortex *J Neurosci*. 2004;24:5331-5.
- 45 Hong J, Shu-Leong H, Tao X, Lap-Ping Y. Distribution of catechol-o-methyltransferase expression in human central nervous system. *Neuroreport*. 1998;9:2861-4.
- 46 Tammimäki AE, Männistö PT. Are genetic variants of COMT associated with addiction? *Pharmacogenet Genomics*. 2010;20:717-41.
- 47 Hindocha C, Freeman TP, Schafer G, Gardner C, Bloomfield MA, Bramon E, et al. Acute effects of cannabinoids on addiction endophenotypes are moderated by genes encoding the CB1 receptor and FAAH enzyme. *Addict Biol*. 2020;25:e12762.
- 48 Palmer RH, McGeary JE, Knopik VS, Bidwell LC, Metrik JM. CNR1 and FAAH variation and affective states induced by marijuana smoking. *Am J Drug Alcohol Abuse*. 2019;45:514-26.
- 49 Haughey HM, Marshall E, Schacht JP, Louis A, Hutchison KE. Marijuana withdrawal and craving: influence of the cannabinoid receptor 1 (CNR1) and fatty acid amide hydrolase (FAAH) genes. *Addiction*. 2008;103:1678-86.
- 50 Cravatt BF, Giang DK, Mayfield SP, Boger DL, Lerner RA, Gilula NB. Molecular characterization of an enzyme that degrades neuromodulatory fatty-acid amides. *Nature*. 1996;384:83-7.
- 51 Monteleone P, Milano W, Petrella C, Canestrelli B, Maj M. Endo-cannabinoid pro129Thr FAAH functional polymorphism but not 1359G/A CNR1 polymorphism is associated with antipsychotic-induced weight gain. *J Clin Psychopharmacol*. 2010;30:441-5.
- 52 Sloan ME, Gowin JL, Yan J, Schwandt ML, Spagnolo PA, Sun H, et al. Severity of alcohol dependence is associated with the fatty acid amide hydrolase Pro129Thr missense variant. *Addict Biol*. 2018; 23:474-84.
- 53 Patel MM, Nielsen DA, Kosten TR, De La Garza R 2nd, Newton TF, Verrico CD. FAAH variant Pro129Thr modulates subjective effects produced by cocaine administration. *Am J Addict*. 2018;27:567-73.
- 54 Harrison RJ. Understanding genetic variation and function- the applications of next generation sequencing. *Semin Cell Dev Biol*. 2012;23:230-6.