Identification of putative second genetic hits in schizophrenia carriers of high‑risk copy number variants and resequencing in additional samples

Julio Rodriguez‑Lopez[1] · Beatriz Sobrino[1,2,3] · Jorge Amigo[1,2,3] · Noa Carrera[1,3,5] · Julio Brenlla[1,4] · Santiago Agra[1,4] · Eduardo Paz[1,4] · Ángel Carracedo[1,2,3] · Mario Páramo[1,4] · Manuel Arrojo[1,4] · Javier Costas[1]

Abstract
Copy number variants (CNVs) conferring risk of schizophrenia present incomplete penetrance, suggesting the existence of second genetic hits. Identification of second hits may help to find genes with rare variants of susceptibility to schizophrenia. The aim of this work was to search for second hits of moderate/high risk in schizophrenia carriers of risk CNVs and resequencing of the relevant genes in additional samples. To this end, ten patients with risk CNVs at cytobands 15q11.2, 15q11.2-13.1, 16p11.2, or 16p13.11, were subjected to whole-exome sequencing. Rare single nucleotide variants, defined as those absent from main public databases, were classified according to bioinformatic prediction of pathogenicity by CADDscores. The average number of rare predicted pathogenicvariants per sample was 13.6 (SD 2.01). Two genes, *BFAR* and *SYNJ1*, presented rare predicted pathogenic variants in more than one sample. Follow-up resequencing of these genes in 432 additional cases and 432 controls identified a significant excess of rare predicted pathogenic variants in case samples at *SYNJ1*. Taking into account its function in clathrin-mediated synaptic vesicle endocytosis at presynaptic terminals, our results suggest an impairment of this process in schizophrenia.

Introduction

Recently, several rare recurrent copy number variants(CNVs) involved in the risk of having neurodevelopmental disorders have been identified. They show incomplete penetrance,i.e., there are asymptomatic carriers, and variable clinical expressivity, i.e., they confer risk of developing different disorders, such as intellectual disability, autism spectrum disorders, schizophrenia, or epilepsy [1]. These CNVs appear recurrently due to high mutation rates, but are eliminated in a few generations due to the action of purifying selection. Thus, their frequencies in general populations are very low, showing an inverse correlation with their risk effects. For instance, the 16p11.2. duplication presents a frequency of 0.03% in control populations and an odds ratio (OR) for schizophrenia of 11.5. By contrast, the 15q11.2 deletion that presents a considerably lower effect (OR = 2.15) reaches a higher frequency (0.28%) in control populations [2].

The incomplete penetrance of risk CNVs suggests the existence of second genetic hits [3–7]. Empirical data support this suggestion. For instance, 10 of 42 (24%) patients with the 16p12.1 microdeletion (involving ~520 kb) diagnosed with intellectual disability/developmental delay, presented an additional CNV larger than 500 kb. By comparison, only 4.4% of 471 controls having at least one CNV larger than 500 kb presented a second CNV larger than 500 kb ($P = 5.7 \times 10^{-5}$) [5].

The second hit hypothesis has been tested in a large sample of subjects with developmental delay, congenital malformations, and/or autism spectrum disorders; confirming that approximately 10% of carriers of risk CNVs present an additional rare CNV larger than 500 kb or an additional pathogenic CNV [8]. The “second hit” hypothesis explains as many cases of schizophrenia as cases of developmental delay, congenital malformations, and/or autism spectrum disorders [9].

In addition to CNVs, other types of genetic variants, such as rare single nucleotide variants (SNVs), may act as second hits. Currently, it is possible to study SNVs across the genome making use of the next-generation sequencing technologies. Following the pioneering work of Need et al. [10], Purcell et al. [11] was the first study to publish whole exome sequencing studies involving thousands of schizophrenia patients and controls. Although they did not find any gene with an experiment wise significant excess of rare variants in schizophrenia at the level of candidate gene sets. Recently, Genovese et al. [12] found an increase burden of ultra-rare disruptive and damaging SNVs in schizophrenia cases versus controls, mainly in genes expressed in the brain. Studies of de novo mutations in trios also detected a role for rare variants in schizophrenia [13, 14]. In fact, Fromer et al. [15] confirmed the importance of two of the gene-sets identified by Purcell et al. [11], comprising the actin-mediated cytoskeleton protein (ARC) complex and the *n*-methyl d-aspartate receptor (NMDAR) complex. More specific applications of whole-exome sequencing in schizophrenia has been tested,

such as co-segregation studies in pedigrees [16, 17]; comparison of patients with clozapine- induced agranulocitosis versus other schizophrenia patients under clozapine treatment [18]; and comparison of 22q.11.2 deletion carriers differing in the presence of psychosis [19].

The main difficulty of whole-exome sequencing studies is the differentiation of pathogenic and benign SNVs among the tens of thousands of SNVs present in a single exome. Recently, the combined annotation-dependent depletion (CADD) method has been proposed to estimate the relative pathogenicity of SNVs, making use of manydifferent sources of information. This estimate, called the C score, may be used to prioritize SNVs from whole-exome/genome sequencing data according to their relative deleteriousness[20].

Here, we present a strategy for the identification of new schizophrenia loci consisting of three steps. First, we performed whole-exome sequencing in ten schizophrenic carriers of risk CNVs, under the hypothesis that risk CNVs with incomplete penetrance are silenced in general genetic backgrounds but are fully expressed in those subjects that present high impact SNVs acting as second genetic hits. Then, those genes with more than one sample presenting putative second genetic hits were genotyped in our collection of cases and controls to confirm the presence of the SNV, its low frequency, and to identify additional carriers. Finally, a random subset of the whole collection was sequenced in a case–control design to search for an excess of rare putative functional variants in schizophrenia patients. A schematic diagram of the overall experimental approach is shown in Supplementary Figure 1.

Methods

Samples

The study is based on our collection of 558 schizophrenia patients from the Santiago de Compostela healthcare area (Galicia, NW Spain), meeting the DSM-IV criteria for schizophrenia, described in Carrera et al. [21]. A total of ten schizophrenia samples detected in our previous work as carriers of schizophrenia-associated CNVs were examined in the present study, including two carriers of the 15q11.2 deletion, one of the 15q11.2-13.1 duplication, four of the 16p11.2 duplication, one of the 16p11.2 distal deletion, and two of the 16p13.11 duplication [22]. Nine samples were identified by quantitative interspecies competitive (qic) PCR and the remaining one by MLPA during the confirmation step of the qicPCR screening with the MLPA design SALSA MLPA KIT P343-C1 AUTISM-1 from MRC-Holland (Amsterdam, The Netherlands). All sample donors gave their written informed consent for this study. The study was performed in accordance with the latest version of the Declaration of Helsinki and was approved by the Galician Ethical Committee for Clinical Research. Main clinical and genetic characteristics of the ten schizophrenia patients carrying risk CNVs are shown in Supplementary Table 1. In addition, 583 samples from the same geographical region, collected at the Galician Transfusion Center at Santiago de Compostela, were used as controls. Previous analysis using multidimensional scaling showed that population stratification is not a problem in this case–control sample [21].

Whole‑exome sequencing

Whole-exome capture of the ten carriers of schizophrenia- associated CNVs was performed using Agilent Sure-SelectXT Human All Exon v4 or v5 (Agilent Technologies, Inc, Santa Clara, USA). Captured fragments were sequenced on a SOLiD 5500XL platform (Applied BiosystemsTM) following manufacturer instructions. Exome sequencing raw data were mapped relative to the human reference genome GRCh37 following the Genomic Analysis Software pipeline (Applied Biosystems, Waltham, Massachusetts, USA) with default parameters. Variant calling was also performed by Life- Scope. The mapping generated by LifeScope was used as input for the Genome Analysis Toolkit (GATK) software [23]. SNV detection was carried out following the GATK Best Practices recommendations, including detection of duplicated reads by Picard Tools v 1.106 (http://picard. sourceforge.net), local realignment around indels, base quality score recalibration (BQSR), and variant quality score recalibration (VQSR), to obtain the final sequence alignment data [24, 25]. Only those SNVs detected by both LifeScope and GATK were considered for further analyses. Systematic sequencing or mapping errors were discarded by comparison with an in-house database of wholeexome sequences using the same technology from the Galician Public Foundation of Genomic Medicine (sample size = 117) or by visual inspection of interesting SNVs with the Integrative Genomics Viewer (IGV) software [26].

SNV annotation

Relevant annotations of SNVs were obtained from the CADD score web server (http://cadd.gs.washington.edu/). Annotations included normalized C score, frequency in the NHLBI GO Exome Sequencing Project (data release ESP6500SI-V2) [27] and The 1000 Genomes Project Phase I v3 (T1000G) [28], and consequence to transcript based on the Ensembl Variant Effect Predictor.

CNV detection by exome depth coverage

ExomeDepth [29] was used to detect CNVs based on a comparison of the depth of coverage using the ten exome sequences from our patients as well as a reference set of other in-house sequences. Default values of transition probability and the over-dispersion parameter of the binomial model were used in the analysis. To avoid false positives, only those CNVs affecting more than two exons were considered. The criteria described by Kearney et al. [30] were followed to identify pathogenic CNVs or uncertain clinical significance-likely pathogenic CNVs.

Identification of putative second hit SNVs

Putative relevant SNVs were considered based on three filters: (1) location at coding regions or at the intronic canonical splice sites; (2) absence from ESP6500 and T1000G projects; and (3) normalized C score ≥23. This is the mean score for pathogenic variants compiled by the ClinVar database [20]. The SNVs fulfilling these three criteria are referred to herein as "putative second hit SNVs".

For comparison purposes, we also identified coding SNVs or SNVs in intronic canonical splice sites with normalized C score ≥23 present just once (singletons) in the 379 European samples from T1000G and absent from other T1000G populations and from the ESP6500 database.

SNV Genotyping

Relevant putative second hit SNVs were input into the Sequenom MassArray Assay Design software (Sequenom, San Diego, CA, USA) for genotyping confirmation and their detection in additional samples. All samples were genotyped by mass spectrometry, using the Sequenom MassArray technology according to manufacturer's instructions. Sequenom MassArray technology was also used to confirm relevant variants identified after targeted exon resequencing.

Targeted exon resequencing of candidate genes

Those genes with putative second hit SNVs at more than one sample were subjected to target exon resequencing. Ion AmpliSeq Designer (Thermo Fisher Scientific) was used to create 83 PCR assays for amplicons ranging in size from 124 to 274 bp, combined in two multiplex panels. A total of 432 schizophrenia patients and 432 controls randomly chosen from the whole sample were selected for resequencing. Prior to PCR, samples were combined in 144 pools, each pool containing equimolar quantities of DNA from six individuals. Estimation of DNA concentration was made by three independent measurements using the Qubit ® dsDNA BR assay (Invitrogen, Carlsbad, CA, USA). Sequencing libraries were prepared according to manufacturer's instructions. After determination of molarity, the 144 libraries were split into three subsets of 48 barcoded libraries, with equal number of patients and controls per subset. For each subset, all 48 final libraries were pooled together and standardized to 50 pM in low TE buffer (Life Technologies, Carlsbad, California, USA) for sequencing on the same Ion PI Chip using Ion Proton™ next-generation sequencing (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Data from the Ion Proton runs were processed by the Torrent Suite Software v3.0 and variant calling made with the Torrent Variant Caller Plugin applying low stringency default values for somatic variant detection. These parameters were adapted to minimize false-negative calls and to fit our DNA pooling approach.

ANNOVAR [31] was used to annotate the detected SNVs. Missense SNVs predicted to be pathogenic by at least one of five prediction algorithms (Polyphen Hum- Div, Polyphen HunVar, SIFT, MutationTaster, and LRT) were considered pathogenic. In addition, loss-of-function variants, i.e., nonsense variants, frameshift variants, or variants at canonical splicing sites, were also considered pathogenic. To increase sensitivity and specificity in SNV calling, all putative relevant SNVs were visually inspected using the Integrative Genomics Viewer (IGV) software [26] and those SNVs with reference and alternative alleles appearing at least once on each strand were selected for validation by genotyping.

Statistical analysis

Comparison of the burden of putative pathogenic SNVs at frequencies <0.1% in T1000G and ESP6500 databases in cases versus controls was analyzed using the one-tailed Fisher's exact test. Comparison of the distribution of rare variants at *SYNJ1* between cases and controls used the Cochran–Armitage test for trend under the alternative hypothesis of a greater number of rare SNV in cases using the DescTools package in R.

Table 1 Putative second hit SNVs at genes with more than one hit discovered by wholeexome sequencing of the ten carriers of pathogenic CNVs

Position (hg19)	Alleles[a]	Gene	Transcript	Consequence	Normalized C score
Chr11:66322054	G/A	*ACTN3*	ENST00000513398	p.Ala173Thr	29.3
Chr11:66323648	G/C	*ACTN3*	ENST00000513398	p.Ala251Pro	27.2
Chr16:14761565	C/A	*BFAR*	ENST00000261658	p.Pro412Thr	26.6
Chr16:14761643	G/A	*BFAR*	ENST00000261658	p.Val438Met	27.4
Chr21:34038789	C/T	*SYNJ1*	ENST00000322229	p.Val636Ile	24
Chr21:34072188	G/A	*SYNJ1*	ENST00000322229	p.Arg147Cys	29.4
Chr22:20101006	C/A	*TRMT2A*	ENST00000403707	p.Glu461Ter	39
Chr22:20102907	C/T	*TRMT2A*	ENST00000403707	p.Val258Met	27.9

[a] Shown as reference/alternative

Results

Exome sequencing of the ten carriers of risk CNVs

The ten carriers of schizophrenia-associated CNVs were subjected to exome sequencing to search for additional variants that could represent a second hit. In any particular patient, 79.10% of captured bases were covered by at least 10× sequence reads (average 86%) and 59.43% by at least 30× sequence reads (average 69%). All CNVs were confirmed based on depth of coverage, as previously described [22].

Searching for second hit variants

No samples presented additional CNVs that fulfilled the criteria for pathogenic or likely pathogenic CNVs. There were 136 SNVs absent from databases and with normalized C scores ≥23 (Supplementary Table 2). The distribution of SNVs by sample was very similar, ranging from 11 to 16 (mean 13.6, SD 2.01) (Supplementary Table 1). Among the genes with putative second hit SNVs, there were several interesting genes, such as *CACNA2D2*, *ANKS1B,* and *CAMK2B,* belonging to the list of 31 genes more likely to harbor large effect alleles according to a large exome sequencing study [11]. Four genes presented more than one putative second hit SNV (Table 1). For comparison, we performed 100,000 random permutations of 136 SNVs taken from the 3930 putative damaging SNVs at T1000G European samples after applying similar filters. Only 4.3% of permutations presented at least four genes with more than two SNVs, suggesting that the presence of four genes twice in our samples is unlikely by chance. Among the four genes, *ACTN3* is a sarcomeric protein expressed exclusively in skeletal muscle, and a common nonsense polymorphism is present at this gene [32]. The two SNPs of *TRMT2A* were carried by the same patient. In addition, there is a loss-of-function variant at a frequency higher than 1% in the 6500ESP data set at *TRMT2A*, suggesting low functional relevance. Therefore, both genes were excluded from further analysis. The two remaining genes, *BFAR* and *SYNJ1*, were selected for resequencing. There were no singleton SNVs at the two genes in T1000G European samples with normalized C scores ≥23. Prior to sequencing, the presence of the putative second hit SNVs was confirmed by genotyping the whole sample. None of the controls presented the rare alleles at any of the SNVs. While no additional cases presented the SNVs at *SYNJ1*, three and one additional cases carried the SNVs Pro412Thr and Val432Met at *BFAR* respectively.

Resequencing of *BFAR* and *SYNJ1*

A total of 432 schizophrenia patients and 432 controls randomly chosen from the whole sample were subjected to targeted resequencing. More than 90% of pools presented more than 90% of positions at a depth of coverage >240×, both in cases and controls. The depth of coverage was similar between cases and controls (*t* test $P = 0.21$) (Fig. 1). Table 2 shows the putative pathogenic SNVs at frequencies lower than 0.1% at these two genes, confirmed by genotyping. There was an excess of carriers of putative pathogenic SNVs at frequencies lower than 0.1% at these genes in cases ($P = 0.037$). At *BFAR*, three cases and one control presented putative pathogenic SNVs at frequencies lower than 0.1% ($P = 0.31$). At *SYNJ1*, there were nine cases and three controls presenting putative pathogenic SNVs at frequencies lower than 0.1%, including one nonsense variant in a schizophrenia patient. One of the cases presented two mutations, 61 bp apart (Table 2). Visual inspection of sequencing reads revealed that the sample is a compound heterozygote, leading to a significant association in the comparison of the number of alleles carried by cases versus controls (trend test, $P = 0.034$). The excess was more significant when the analysis was focused on the proline-rich domain ($P = 0.017$) (Fig. 2).

Discussion

In this study, we described a strategy for the identification of new schizophrenia susceptibility genes harboring rare risk variants. The strategy is based on exome sequencing of schizophrenia patients with risk CNVs to detect putative second genetic hits, following by resequencing of the involved genes in additional samples. This approach led to the identification of an excess of rare putative functional variants at *SYNJ1* in schizophrenia. One patient was a compound heterozygote.

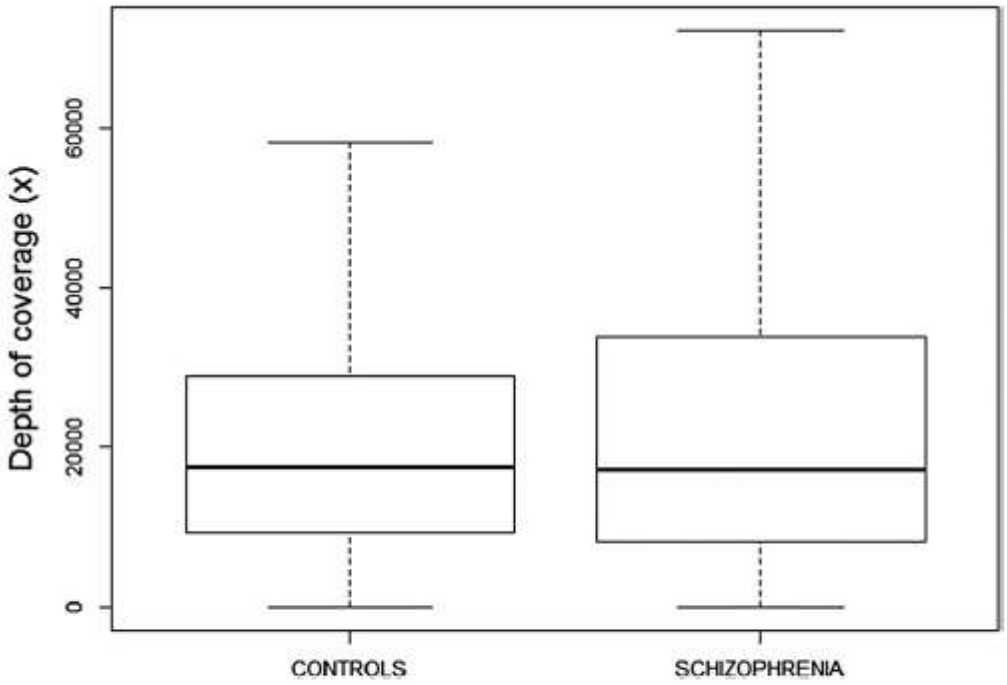

Fig. 1 *Boxplot* representing depth of coverage per base in case and control samples

Resequencing of *BFAR* and *SYNJ1*

A total of 432 schizophrenia patients and 432 controls randomly chosen from the whole sample were subjected to targeted resequencing. More than 90% of pools presented more than 90% of positions at a depth of coverage >240×, both in cases and controls (*t* test $P = 0.21$) (Fig. 1). Table 2 shows the putative pathogenic SNVs at frequencies lower than 0.1% at these two genes, confirmed by genotyping. There was an excess of carriers of putative pathogenic SNVs at frequencies lower than 0.1% at these genes in cases ($P = 0.037$). At *BFAR*, three cases and one control presented putative pathogenic SNVs at frequencies lower than 0.1% ($P = 0.31$). At *SYNJ1*, there were nine cases and three controls presenting putative pathogenic SNVs at frequencies lower than 0.1%, including one nonsense variant in a schizophrenia patient. One of the cases presented two mutations, 61 bp apart (Table 2). Visual inspection of sequencing reads revealed that the sample is a compound heterozygote, leading to a significant association in the comparison of the number of alleles carried by cases versus controls (trend test, $P = 0.034$). The excess was more significant when the analysis was focused on the proline-rich domain ($P = 0.017$) (Fig. 2).

Discussion

In this study, we described a strategy for the identification of new schizophrenia susceptibility genes harboring rare risk variants. The strategy is based on exome sequencing of schizophrenia patients with risk CNVs to detect putative second genetic hits, following by resequencing of the involved genes in additional samples. This approach led to the identification of an excess of rare putative functional variants at *SYNJ1* in schizophrenia. One patient was a compound heterozygote.

Interestingly, Purcell et al. [11] found eight nonsynonymous SNVs at *SYNJ1* with bioinformatic evidence of functionality at a frequency lower than 0.1% in 2536 schizophrenia patients versus just one in 2543 controls

(P = 0.0283, as shown in the genebook web server, http:// research.mssm.edu/statgen/sweden/), using strict criteria for prediction of damaging mutation. SYNJ1 is a phosphoinositide phosphatase involved in clathrin-mediated synaptic vesicle endocytosis at presynaptic terminals [33, 34]. This mechanism has been recently proposed as a core pathophysiological process in schizophrenia [35]. In addition, SYNJ1 is also located postsynaptically, regulating endocytosis of AMPA receptors in dendrites [36]. SYNJ1 presents three domains: a Sac1 phosphatase domain involved in the actin cytoskeleton dynamic, an inositol-5-phosphatase domain, and a proline-rich domain. The association between *SYNJ1* and schizophrenia seems to involve mainly the proline-rich domain both in our data as well as in the Purcell et al.'s data [11], which showed five SNVs in cases and none in controls at this domain. However, the two SNVs detected by exome sequencing of CNV carriers in the first step of the study lay within the other domains (Fig. 2).

The proline-rich domain is a protein binding domain implicated in interactions with many partners, such as Grb2, amphiphysin, syndapin/pacsin, intersectin, and endophilin [37]. In general, this binding controls subcellular location of SYNJ1 and enhances its phosphatase activity. The enzymatic activity of SYNJ1 is mainly regulated through phosphorylation/dephosphorylation at specific sites within the proline-rich motif by Cdk5, EphB2, or calcineurin, and the protein is activated at synapses by calcium influx [37]. Interestingly, a recessive mutation at the Sac1 phosphatase domain has been implicated in early-onset Parkinson disease with atypical features in three different families, including occasional seizures and delayed child developmental milestones [38–40].

Table 2 Putative pathogenic SNVs at frequency lower than 0.1% in T1000G and ESP6500 databases, discovered by resequencing

Gene	Position (hg19)	Alleles[a]	Consequence[b]	Carriers (432 SCZ, 432 CO)[c]	Normalized C score
BFAR	Chr16:14738456	A/G	p.Ile85Val	1 CO	20.2
BFAR	Chr16:14749014	G/C	p.Glu244Gln	1 SCZ	15.48
BFAR	Chr16:14761565	C/A	p.Pro412Thr	1 SCZ	26.6
BFAR	Chr16:14761643	G/A	p.Val438Met	1 SCZ	27.4
SYNJ1	Chr21:34003924	C/T	p.Ser1368Asn	1 SCZ	12.23
SYNJ1	Chr21:34004005	G/A	p.Ser1341Phe	2 SCZ	17.7
SYNJ1	Chr21:34007188	G/A	p.Gln1292Ter	1 SCZ	42
SYNJ1	Chr21:34011369	G/A	p.Pro1255Leu	1SCZ/1CO	7.68
SYNJ1	Chr21:34017997[d]	G/A	p.Pro1117Leu	1 SCZ	14.5
SYNJ1	Chr21:34018060[d]	T/G	p.Gln1096Pro	1 SCZ	8.58
SYNJ1	Chr21:34038353	C/T	p.Ser682Asn	1 CO	36
SYNJ1	Chr21:34045731	G/A	p.Arg549Cys	1 CO	32
SYNJ1	Chr21:34050957	C/T	p.Arg503His	1 SCZ	11.26
SYNJ1	Chr21:34059345	T/C	p.Ile297Met	1 SCZ	14.14
SYNJ1	Chr21:34099169	A/G	p.Leu13Ser	1 SCZ	16.44

[a] Shown as reference/alternative

[b] Referred to the same transcript as in Table 1

[c] CO, control; SCZ, schizophrenia. All the SNVs were confirmed by genotyping

[d] These two SNVs are carried by the same patient

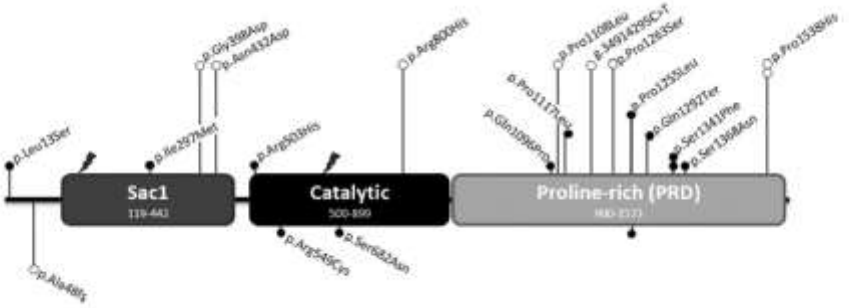

Fig. 2 Schematic representation of the domain structure of SYNJ1 protein showing the position of rare putative pathogenic SNVs. Filled circles represent SNVs detected in our samples. Open circles represent the variants reported by Purcell et al. [11] for comparison. SNVs detected in schizophrenia cases are shown above the protein and SNVs detected in healthy controls are shown below the protein. All SNVs are represented on the canonical isoform O43426-1 (1573aa), except the nonsense SNV p.Gln1292Ter, which was no present in this canonical isoform and was placed on this schematic representation in its corresponding position on the isoform O43426-4 (1295aa). SNV g.34914295C>T is an intronic variant located at a splicing donor site. *Lightning arrows* represent the two second hit SNVs found in risk CNV carriers

The other gene with putative second hit SNVs in more than one CNV carrier, *BFAR,* is mainly expressed in neurons, protecting them from different cell death stimuli [41, 42]. Four additional cases and no controls were carriers of one of the two putative second hit SNVs at *BFAR*. However, our resequencing approach did not find any significant excess of rare putative damaging nonsynonymous SNVs in schizophrenia. Similarly, Purcell et al. did not find any significant result at this gene [11].

Some limitations of our study should be considered, such as the small sample size, the inexact classification of pathogenicity inherent in current bioinformatics approaches, or the lack of clinical, and, mainly, genetic data from parents to determine if the second hits were inherited or de novo mutations. Among the strengths of the study, the use of Iberian samples provided a higher proportion of rare SNVs in comparison with other European samples from T1000G [28], which gave an increase in power at the same sample size.

Conclusion

To conclude, our approach to find additional risk variants in schizophrenia carriers of risk CNVs, consisting on whole-exome sequencing and follow-up analyses of genes with putative second hit SNVs in a resequencing step, led to the identification of *SYNJ1* as a possible new schizophrenia risk gene. This result points to impairment in clathrin-mediated synaptic vesicle endocytosis in schizophrenia.

Acknowledgements This work was supported by Grant CP11/00163 from Instituto de Salud Carlos III, cofounded by FEDER; to JC, by agreement between SERGAS and Fundación Pública Galega de Medicina Xenómica, and by the Innopharma project (USC). The founders had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. The genotyping service was carried out at CEGEN-PRB2-ISCIII; it is supported by Grant PT13/0001, ISCIII-SGEFI/FEDER. The authors would like to thank Centro de Supercomputación de Galicia (CESGA) for the use of their computing facilities and the NHLBI GO Exome Sequencing Project and its ongoing studies which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL 102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926), and the Heart GO Sequencing Project (HL-103010).

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Malhotra D, Sebat J (2012) CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 148:1223–1241. doi:10.1016/j.cell.2012.02.039
2. Rees E, Walters JT, Georgieva L, Isles AR, Chambert KD, Richards AL, Mahoney-Davies G, Legge SE, Moran JL, McCarroll SA, O'Donovan MC, Owen MJ, Kirov G (2014) Analysis of copy number variations at 15 schizophrenia-associated loci. Br J Psychiatry 204:108–114. doi:10.1192/bjp. bp.113.131052
3. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM et al (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320:539–543. doi:10.1126/science.1155174
4. Girirajan S, Eichler EE (2010) Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet 19:R176– R187. doi:10.1093/hmg/ddq366
5. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A et al (2010) A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet 42:203–209. doi:10.1038/ng.534
6. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S et al (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43:585– 589. doi:10.1038/ng.835
7. Mitchell KJ, Porteous DJ (2011) Rethinking the genetic architecture of schizophrenia. Psychol Med 41:19–32. doi:10.1017/ S003329171000070X
8. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A et al (2012) Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med 367:1321– 1331. doi:10.1056/NEJMoa1200395
9. Kirov G, Rees E, Walters JT, Escott-Price V, Georgieva L, Richards AL et al (2013) The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry 75:378–385. doi:10.1016/j.biopsych.2013.07.022
10. Need AC, McEvoy JP, Gennarelli M, Heinzen EL, Ge D, Maia JM et al (2012) Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. Am J Hum Genet 91:303–312. doi:10.1016/j.ajhg.2012.06.018

11. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P et al (2014) A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506:185–190. doi:10.1038/ nature12975
12. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M et al (2016) Increased burden of ultra-rare proteinaltering variants among 4,877 individuals with schizophrenia. Nat Neurosci 19:1433–1441. doi:10.1038/nn.4402
13. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S et al (2011) Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nat Genet 43:864–868. doi:10.1038/ ng.902
14. McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y et al (2014) De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Mol Psychiatry 19:652–658. doi:10.1038/ mp.2014.29
15. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P et al (2014) De novo mutations in schizophrenia implicate synaptic networks. Nature 506:179–184. doi:10.1038/ nature12929
16. Egawa J, Hoya S, Watanabe Y, Nunokawa A, Shibuya M, Ikeda M et al (2016) Rare UNC13B variations and risk of schizophrenia: whole-exome sequencing in a multiplex family and followup resequencing and a case-control study. Am J Med Genet B Neuropsychiatr Genet 171:797–805. doi:10.1002/ajmg.b.32444
17. Timms AE, Dorschner MO, Wechsler J, Choi KY, Kirkwood R, Girirajan S et al (2013) Support for the *N* methyl-d-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families. JAMA Psychiatry 70:582–590. doi:10.1001/jamapsychiatry.2013.1195
18. Tiwari AK, Need AC, Lohoff FW, Zai CC, Chowdhury NI, Müller DJ et al (2014) Exome sequence analysis of Finnish patients with clozapine-induced agranulocytosis. Mol Psychiatry 19:40- 405.doi:10.1038/mp.2013.74
19. Balan S, Iwayama Y, Toyota T, Toyoshima M, Maekawa M, Yoshikawa T (2014) 22q11.2 deletion carriers and schizophrenia- associated novel variants. Br J Psychiatry 204:398–399. doi:10.1192/bjp.bp.113.138420
20. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46:310–315.
21. Carrera N, Arrojo M, Sanjuán J, Ramos-Ríos R, Paz E et al (2012) Association study of nonsynonymous single nucleotide polymorphisms in schizophrenia. Biol Psychiatry 71:169–177. doi:10.1016/j.biopsych.2011.09.032
22. Rodriguez-Lopez J, Carrera N, Arrojo M, Amigo J, Sobrino B, Páramo M et al (2015) An efficient screening method for simultaneous detection of recurrent copy number variants associated with psychiatric disorders. Clin Chim Acta 445:34–40. doi:10.1016/j.cca.2015.03.013
23. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297 1303. doi:10.1101/gr.107524.110
24. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498. doi:10.1038/ng.806
25. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A et al (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11.10.1-33. doi:10.1002/0471250953.bi1110s43
26. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. doi:10.1093/bib/bbs017
27. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM et al (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 493:216–220. doi:10.1038/nature11690
28. The Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073. doi:10.1038/nature09534
29. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S et al (2012) A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics 28:2747–2754. doi:10.1093/ bioinformatics/bts526
30. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST, Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee (2011) American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med 13:680–685. doi:10.1097/ GIM.0b013e3182217a3a
31. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164. doi:10.1093/nar/gkq603
32. MacArthur DG, North KN (2004) A gene for speed? The evolution and function of alpha-actinin-3. BioEssays 26:786–795
33. Cremona O, Di Paolo G, Wenk MR, Lüthi A, Kim WT, Takei K et al (1999) Essential role of phosphoinositide metabolism in synaptic vesicle recycling. Cell 99:179–188
34. Kononenko NL, Haucke V (2015) Molecular mechanisms of presynaptic membrane retrieval and synaptic vesicle reformation. Neuron 85:484–496. doi:10.1016/j.neuron.2014.12.016

35. Schubert KO, Föcking M, Prehn JH, Cotter DR (2012) Hypothesis review: are clathrin-mediated endocytosis and clathrindependent membrane and protein trafficking core pathophysiological processes in schizophrenia and bipolar disorder? Mol Psychiatr 17:669–681. doi:10.1038/mp.2011.123
36. Gong LW, De Camilli P (2008) Regulation of postsynaptic AMPA responses by synaptojanin 1. Proc Natl Acad Sci USA 105:17561–17566. doi:10.1073/pnas.0809221105
37. Krauss M, Haucke V (2007) Phosphoinositide-metabolizing enzymes at the interface between membrane traffic and cell signalling. EMBO Rep 8:241–246. doi:10.1038/sj.embor.7400919
38. Quadri M, Fang M, Picillo M, Olgiati S, Breedveld GJ, Graafland J et al (2013) Mutation in the SYNJ1 gene associated with autosomal recessive, early-onset Parkinsonism. Hum Mutat 34:1208–1215. doi:10.1002/humu.22373
39. Krebs CE, Karkheiran S, Powell JC, Cao M, Makarov V, Darvish H et al (2013) The Sac1 domain of SYNJ1 identified mutated in a family with early-onset progressive Parkinsonism with generalized seizures. Hum Mutat 34:1200–1207. doi:10.1002/ humu.22372
40. Olgiati S, De Rosa A, Quadri M, Criscuolo C, Breedveld GJ, Picillo M et al (2014) PARK20 caused by SYNJ1 homozygous Arg258Gln mutation in a new Italian family. Neurogenetics 15:183–188. doi:10.1007/s10048-014-0406-0
41. Roth W, Kermer P, Krajewska M, Welsh K, Davis S, Krajewski S et al (2003) Bifunctional apoptosis inhibitor (BAR) protects neurons from diverse cell death pathways. Cell Death Differ 10:1178–1187. doi:10.1038/sj.cdd.4401287
42. Zhang H, Xu Q, Krajewski S, Krajewska M, Xie Z, Fuess S et al (2000) BAR: an apoptosis regulator at the intersection of caspases and Bcl-2 family proteins. Proc Natl Acad Sci USA 97:2597–2602