Type of article: Original research

Title: Whole Exome Sequencing in dense families suggests genetic pleiotropy amongst Mendelian and complex neuropsychiatric syndromes

Running title: Pleiotropy in familial SMI

Authors:

Suhas Ganesh, MD, Schizophrenia Neuropharmacology Research Group, Yale University School of Medicine, New Haven, US

Alekhya Vemula, B.Tech, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bengaluru, India

Samsiddhi Bhattacharjee, PhD, National Institute of Biomedical Genomics, Kalyani, India

Kezia Mathew, MSc, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bengaluru, India

Dhruva Ithal, MD, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bengaluru, India

Karthick Navin, MD, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bengaluru, India

Ravi Kumar Nadella, MD, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bengaluru, India

Biju Viswanath, MD, PhD, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bengaluru, India

The ADBS consortium

Sanjeev Jain, MD, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bengaluru

Meera Purushottam*, PhD, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bengaluru

*corresponding author

The ADBS consortium

Naren P. Rao[1], Janardhanan C. Narayanaswamy[1], Palanimuthu T. Sivakumar[1], Arun Kandasamy[1], Muralidharan Kesavan[1], Urvakhsh Meherwan Mehta[1], Ganesan Venkatasubramanian[1], John P. John[1], Odity Mukherjee[2], Ramakrishnan Kannan[1], Bhupesh Mehta[1], Thennarasu Kandavel[1], B. Binukumar[1], Jitender Saini[1], Deepak Jayarajan[1], A. Shyamsundar[1], Sydney Moirangthem[1], K. G. Vijay Kumar[1], Bharath Holla[1], Jayant Mahadevan[1], Jagadisha Thirthalli[1], Prabha S. Chandra[1], Bangalore N. Gangadhar[1], Pratima Murthy[1], Mitradas M. Panicker[3], Upinder S. Bhalla[3], Sumantra Chattarji[3], Vivek Benegal[1], Mathew Varghese[1], Janardhan Y. C. Reddy[1], Padinjat Raghu[3] and Mahendra Rao[2], Biju Viswanath[1], Meera Purushottam[1], Sanjeev Jain[1]

1 National Institute of Mental Health and Neurosciences, Bangalore, India

2 Institute for Stem Cell Biology and Regenerative Medicine (InStem), Bangalore, India

3 National Center for Biological Sciences (NCBS), Bangalore, India

Word count:

Abstract – 254

Main text – 4610

Number of Tables: 4

Number of figures: 1

Corresponding Author:

Meera Purushottam, * PhD,

National Institute of Mental Health and Neurosciences

Bangalore, India

E-mail: meera.purushottam@gmail.com

Title: Whole Exome Sequencing in dense families suggests genetic pleiotropy amongst Mendelian and complex neuropsychiatric syndromes

Abstract

Introduction: Whole Exome Sequencing (WES) studies have provided important insights into the genetic architecture of neuropsychiatric syndromes identifying rare and novel variants in the protein-coding sequence of the genome that impact function. Variants and genes that are central to the shared biology of these clinical syndromes may be identified by WES in families with multiple affected individuals with serious mental illnesses (F-SMI).

Methods: We performed WES in 250 individuals (affected = 186, family-control = 64) from 100 families, each with ≥2 members with SMI, and 60 unrelated population-controls. Within pedigree prioritization employed criteria of 1. *rarity* (Minor Allele Frequency <0.1%, GnomAD South-Asian sample, 15308 exomes); 2. *functional consequence* ('Loss of Function' or 'missense deleterious in 4/5 in silico predictions). 3. *sharing* by ≥3 affected members within a family. Across the sample, gene-set-wide case-control association analysis was performed with Sequence Kernel Association Test, accounting for kinship.

Results: In 17 families with ≥3 exome samples, we identified 79 rare predicted deleterious variants in 79 unique genes shared by ≥3 affected members and absent in 60 unrelated controls. Twenty (25.32%) genes were implicated in monogenic neurodevelopmental syndromes in Online Mendelian Inheritance in Man representing a statistically significant overrepresentation (Fisher's Exact test OR = 2.47, p = 0.001). In gene-set wise SKAT, statistically significant association was noted for genes related to synaptic function (SKAT-p = 0.017).

Discussion: In F-SMI based WES study, we identify private, rare, protein altering variants in genes previously implicated in monogenic Mendelian neuropsychiatric syndromes; suggesting pleotropic influences in neurodevelopment between complex and monogenic syndromes.

Introduction

Neuropsychiatric syndromes, such as schizophrenia (SCZ), bipolar disorder (BD), obsessive compulsive disorder (OCD), substance use disorders (SUDs) and Alzheimer's dementia (AD) (referred hereafter as serious mental illness [SMI]), often cluster in families (Huang et al, 2020; Sreeraj et al, 2021a). Next generation sequencing (NGS) and analysis can now be used to explore the genetics of complex, common disorders, using both population and family-based designs (Glahn et al, 2019). These include an extension of population-based case-control genome wide association studies (GWAS) to include rare coding variants (Gulsuner et al, 2020; Singh et al, 2020). Such methods have provided useful details about the rare variant contribution to the genetic architecture of SMI supplementing the common variant contributions identified in GWAS.

Population based case-control studies typically exclude familial samples and related individuals to minimize sampling bias and potential false positive associations. However, the sharing, or increased occurrence of potentially disease relevant variants, in families, may offer clues towards genes and pathways involved in the neurobiology of SMI. Many studies employing whole exome sequencing (WES) in psychiatry have thus analyzed families with multiple affected members (reviewed in (Kato, 2015)). These studies have generally focused on pedigrees with a cluster of individuals affected with a specific SMI such as BD or SCZ.

Family studies: Using WES to analyze 65 families with SCZ, Li et al, identified several de novo mutations (DNMs) and risk alleles in genes previously implicated in SCZ as well as novel genes that contributed to risk (Li et al, 2020). Studies suggest that rare loss-of-function variants occur more often in those with SCZ compared to control samples (Legge et al, 2021). An analysis of parent-offspring trios with SCZ suggested that DNMs were detected much more often in genes expressed in the brain, implicated in neurodevelopmental disorders, or under evolutionary constraint (Howrigan et al, 2020). They also noted that many of these variants seem to be of

recent origin. A similar observation was also made with respect to inherited variants that contribute to risk of autism, perhaps attributable to variants that were likely to disrupt gene function, in certain pathways; and this suggests a multi-hit model (Wilfert et al, 2021). Overlaps between neurodevelopmental disorders, autism and risk of schizophrenia are thus observed quite often, and it is suggested that they may have an overlapping pathophysiology.

A WES analysis of 81 individuals from 27 families (with several members diagnosed as BD) again suggested that several rare variants that could contribute to risk of disease. While some of these could be construed as being highly penetrant (shared across families, present in most affected within a family), others too were thought to be plausible risk factors. In addition, overlaps with rare variants implicated in SZ and autism were also identified (Forstner et al, 2020). A family-based WES study in BD with 15 families noted that rare predicted pathogenic variants shared by 3 or more cases were overrepresented in genes coding for post-synaptic density proteins (Toma et al, 2018). Another family-based study noted small to moderate effect increase in the burden of rare deleterious variants in extended pedigrees with multiple affected members. This analysis was restricted to genes that were previously associated with BD. (Sul et al, 2020).

A trio-based WES study in OCD noted an increased burden of protein altering variations, which were statistically significant in the SLITRK5 gene CHD8 gene (Halvorsen et al, 2021). The SLITRK gene family in involved in synaptic regulation, while the CHD8 has been implicated as a risk gene in BD in a case control WES study (Palmer et al, 2021).

Interestingly, while most common and rare variant case-control studies in AD have implicated immune and amyloid processing genes (Jansen et al, 2019), a recent family based WGS implicated rare variants in neurodevelopmental and synaptic function genes (Prokopenko et al, 2021). In a family-based WES study with multiple affected members with SMI, we have

previously demonstrated segregation of rare deleterious variants in genes implicated in Mendelian neuropsychiatric syndromes, that are private to each pedigree (Ganesh et al, 2019).

Population based case-control studies: A WES study comparing around ~900 SCZ cases and matched controls South African Xhosa noted an increased burden of ultra-rare damaging variants in genes related to synaptic function, specifically those involved in glutamate, gamma aminobutyric acid, and dopamine signaling (Gulsuner et al, 2020). Another large metanalysis involving 24248 cases and 97322 controls identified increased burden of large effect ultra-rare variants in 10 exome wide significant genes again implicating molecular functions involving formation, structure and function of the synapse (Singh et al, 2020). A WES study in BD with ~14000 cases and controls also noted a significant burden of ultra-rare protein truncating variants along with significant enrichment of these signals with the gene-set implicated in SCZ WES (Palmer et al, 2021). However, another study that analyzed pathogenic or likely pathogenic variants across the exome in BD did not find a replicable signal for increased burden of such variants in cases compared to controls (Jia et al, 2021).

In summary, there seems to be a general propensity for a higher burden of rare, protein altering variations across cases with different SMI syndromes compared to controls (Ganna et al, 2018; Nishioka et al, 2021), barring some studies providing mixed evidence (Jia et al, 2021; Sul et al, 2020). Functionally, these variants tend to be overrepresented among genes integral to neurodevelopment and synaptic function, and these ontologies are often identified across SMI syndromes (Fromer et al, 2014; Genovese et al, 2016; Kenny et al, 2014). Often, some genes with a higher burden of rare variants, identified in family studies, have also been implicated in other severe neuropsychiatric syndromes, with features of anomalies in neurodevelopment and degeneration (Ganesh et al, 2019; Rees et al, 2021). There seems to be considerable overlap of rare variants that contribute to the risks of SMI, across syndromes. This convergence has also been observed for common variants (Brainstorm-Consortium et al, 2018), genes, gene

networks (Cristino et al, 2014; Grotzinger, 2021), molecular pathways (Gandal et al, 2018), and brain imaging endophenotypes (Radonjic et al, 2021). There is also frequent co-aggregation of SMI syndromes and overlaps in their symptom dimensions within a pedigree (Huang et al, 2020; Sreeraj et al, 2021a). Rare variants, identified in families with multiple ill members, may thus explain a proportion of the risk in the population. They are obviously of great heuristic value, to explore the pathobiology of the disease, and correlates of clinical features.

In this study, we examined the occurrence of rare, deleterious variants in individuals from families that had multiple affected members in families identified in Accelerator Program for Discovery in Brain Disorders using Stem Cells (ADBS) cohort (Viswanath et al, 2018). Within pedigree segregation as well as cross-sample case-control association tests were used to prioritize risk variants. We examined the functional and clinical significance of the prioritized genes and variants, including evolutionary conservation, mutation intolerance, brain expression, protein function and disease relevance. Specifically, we examined if the genes carrying the prioritized variants are overrepresented in Mendelian neuropsychiatric syndromes, and synaptic genes (as attempted for SCZ in the Xhosa population (Gulsuner et al, 2020)). In families segregating a variant in Mendelian disease genes, we scanned the clinical profile of affected individuals for symptoms and signs of the particular Mendelian syndrome.

Methods

Sample: The families with multiple members with SMI (F-SMI) were identified as part of ADBS. This program is aimed at prospective, trans-diagnostic characterization of clinical and neurobiological phenotype of neuropsychiatric syndromes and identification of genetic and molecular correlates of disease using cell models (Sreeraj et al, 2021a; Sreeraj et al, 2021b; Viswanath et al, 2018). The diagnosis of SMIs were established by independent clinical evaluation by two psychiatrists based on International Classification of Diseases – 10 criteria (WHO, 1992). Diagnosis and current and lifetime comorbidity were further evaluated and

confirmed with Mini International Neuropsychiatric Inventory 5.0.0 (Sheehan et al, 1998). The clinical status of the members belonging to the first two generations of all recruited participants was confirmed using Family Interview for Genetic Studies and pedigree charting (Maxwell, 1992).

Individuals with a diagnosis of SMI were defined as 'cases', and unaffected individuals from the same families without current or lifetime diagnosis of SMIs or related syndromes were defined as 'family-controls'. In addition, we identified unaffected unrelated individuals without family history of SMI, as 'population-controls'. Families represented by three or more 'cases' were designated as 'multi-sample F-SMI' and selected for analysis of within-family segregation of variants as described below.

Sequencing, alignment, and variant calling: The protocols for sequencing, alignment and variant calling including the quality controls (QC) at each step employed in the current study have been previously published (Ganesh et al, 2019). In brief, Illumina exome enrichment kits were used for library preparation and Illumina Hiseq NGS platform was used for paired-end sequencing. Raw-read QC was performed with FastQC.0.10.1 and low-quality reads (<Q20) were excluded. Alignment with reference human genome build hg19/GRCH37 was performed using Burrow's Wheeler Algorithm (BWA) 0.5.9 tool. Realignment was performed with 1000G phase1 INDELS (hg19) using GenomeAnalysisTK-3.6 software for removal of PCR duplicates and alignment artefacts. Single Nucleotide Polymorphisms (SNPs) and short insertion deletions (INDEL) variants were called with standard parameters (min coverage = 8, MAF ≥ 0.25 and P ≤ 0.001) using Varscan2 to generate a sample-wise Variant Called File (VCF).

Annotation: Variant Called Files (VCF) were annotated with ANNOVAR (Wang et al, 2010) tool for gene, region and filter based annotation options. For downstream analysis of rarity of identified variants in the population, minor allele frequencies in the Genome Aggregation Database - South Asian sample (gnomAD-SAS) were annotated. For *in silico* prediction of

deleteriousness of coding variants, five functional annotations (*SIFT* (Kumar et al, 2009), *LRT* (Chun & Fay, 2009), *MutationTaster* (Schwarz et al, 2014), *MutationAssessor* (Reva et al, 2011) and *MetaSVM* (Dong et al, 2015)) were used.

Variant prioritization: Variants were prioritized based on the criteria of 'rarity' and having a 'predicted-deleterious' functional consequence. Rare variants were defined as those with a minor allele frequency ≤ 0.001 in the gnomAD-SAS whole exome dataset (consisting of 15308 exomes of persons from South Asian ancestry) in the current study. Single nucleotide variants (SNVs) were defined as predicted-deleterious variants if these were: 1. Protein truncating variants (positional annotation of stop-gain, stop-loss or start loss consequence) 2. Missense variants predicted to have deleterious consequence *in silico* in ≥ 4 of the 5 functional prediction annotation tools. Protein truncating or out-of-frame small insertion-deletions (indels) were similarly prioritized. All subsequent analysis involved the prioritized rare predicted-deleterious (RPD) variants.

Analysis approach: To identify the RPDs that are putatively relevant to SMI syndromes we adopted two independent analytical approaches; a within-family prioritization of variants segregating with SMI and a case-control association analysis.

A. Within-family variant selection – In the subsample of 'multi-sample F-SMI', we examined RPD variants that were shared by three or more cases, that were absent in the in-house population-controls. The resulting lists of variants and the genes harboring these variants were queried with RPDs noted in the cases in previously excluded families (<3 samples from cases/family) to examine potential gene level overlaps in SMI associated genes.

B. Case control association analysis - We performed gene-set and gene-wise case-control association analysis between the entire sample of cases and population controls, after accounting for the kinship structure within the sample using Sequence Kernel Association Test

(SKAT) (Lee et al, 2012) as implemented in the R package SKAT (Lee, 2020). Initially, the SKAT_null_emmax() function was used to approximately adjust for the overall genomic correlation among the individuals by fitting a null model for the binary case/control status incorporating the kinship matrix. The residuals thus obtained were permuted to obtain resampled residuals using SKAT_Null_Model() function. Finally, association of RPD variants with case/control status was examined using SKAT() function. The analysis was run using up to 10e+6 permutations per variant to adjust the accuracy of the p values. Gene-set wide analysis p values were adjusted for Bonferroni correction for the two gene-sets tested. Gene-based p values were corrected for multiple testing using false a discovery cut off of 5% using the BH [Benjamini Hochberg, 1995] procedure.

Clinical and functional significance of the prioritized genes – The set of genes from within-family prioritization was examined for enrichment in monogenic neurodevelopmental syndromes in Online Mendelian Inheritance in Man (OMIM) database using two tailed Fisher's Exact test. The specificity of this enrichment to the CNS was verified by examining enrichment for 19 non-CNS gene lists among OMIM clinical synopsis categories (supplement). In families where RPD variants in OMIM genes for particular monogenic CNS syndromes were noted, we examined the medical records for any overlapping OMIM-like clinical features.

We examined the mutational constraint, tissue expression and evolutionary conservation using publicly available datasets. Mutational constraint was assessed by LOEUF ("loss-of-function observed/expected upper bound fraction") (Karczewski et al, 2020) and pLI ("probability of being loss-of-function intolerance") (Fuller et al, 2019) scores. Brain and cortical expression were examined using the Genotype tissue expression portal GTEx v8 with mean expression values per gene. Conservation scores derived from 240 species alignment were adopted from the Zoonomia consortium (Genereux et al, 2020). Comparisons between prioritized and the background list were performed using the Welch's t test.

In the across pedigree data, we performed a gene-set-wide association analysis between cases and population controls using SKAT modified for gene-set analysis. In this approach, we tested two *a priori* defined gene sets 1. OMIM-CNS genes – to test the findings of the within-pedigree analyses in the full sample 2. Synaptic genes from Syngo database (Koopmans et al, 2019) - to compare with the results of earlier WES studies that implicated synaptic genes in SMIs.

Results

Sample: We sequenced 310 (143 females) individuals, including 250 from 100 F-SMI, and 60 unrelated population-controls. Of the F-SMI samples, 186 were cases diagnosed with a SMI [SCZ (n = 63), BD (n = 80), AD (n = 25), OCD (n = 7), SUD (n = 7), complex SMI (n = 4)] and 64 family-controls were those without a current or lifetime diagnosis of mental illness. The demographic profile of the sample with age and gender distribution and illness profile of the affected sample is provided (Table 1). Fifty-four of the 100 families were constituted by cases with diagnosis across the 6 categories noted above (Figure 1a).

Within the F-SMI familial sample, the mean number of samples from cases per family were 2 (range:1 – 6) and family-controls was 1 (range:1 – 7). A subset of 17 multi-sample F-SMI (≥3 exome samples) were selected for analysis of within pedigree segregation of putatively deleterious variants.

Variant profile: Variant calling the exomes of 310 samples resulted in identification of 793818 unique variants. Among these, the mean (SD) number of synonymous and non-synonymous SNVs per sample were 8933 (2042) and 8646 (2173) respectively. The mean (SD) frequency of rare variants (MAF < 0.1%) per sample was 1009 (726). A breakdown of the variant profile in the sample is provided in figure 1b.

Within pedigree segregation of private variants: Among the 17 multi-sample high-density families, we identified 79 RPD variants in 79 different genes that were shared by three or more

affected members within a family and were absent in the population controls (supplementary table 1 (ST_01)). Of these 78 were private to 14 pedigrees. A variant at Chr5:154268943 in *GEMIN5* gene was shared by two different families among 3/6 and 3/3 (figure 1c) affected cases respectively. In the remaining set of 83 pedigrees with ≤2 samples, 15 of the 79 genes were noted to carry RPD variants that were absent in the in-house population controls (ST_02). Additionally, at the variant level, 5 of the same 79 RPD variants were noted in this set of families (ST_03).

Overrepresentation analysis: Twenty of these 79 genes (25.32%) have been implicated in severe forms of monogenic neurodevelopmental syndromes, with both dominant and recessive modes of inheritance in the Online Mendelian Inheritance in Man (OMIM) database. We performed an overrepresentation analysis of our gene-list on a list of 2450 genes having a central nervous system (CNS) phenotype annotation in OMIM clinical synopsis (OMIM-CNS) among 20203 protein coding genes in the human genome (supplement). The prioritized gene list from multi-sample F-SMI was significantly enriched for genes implicated in monogenic CNS syndromes (Fisher's Exact test OR = 2.47, 95%CI = 1.41 – 4.17, p = 0.001). These gene-phenotype relationships along with the genomic coordinates and pathogenicity prediction of the identified variants in this sample are described in table 2.

To check if the observed overrepresentation was specific to the CNS, we derived 19 additional gene-lists that encompassed genes implicated in OMIM syndromes affecting other organ systems (e.g. 'cs_head_and_neck_head', 'cs_cardiovascular' 'cs_respiratory' etc.). Among the 20 OMIM derived gene-lists (CNS + above 19), statistically significant overrepresentation at p threshold = .0025 (0.05/20) was noted for gene lists annotated with clinical synopsis terms 'central nervous system' (p_{cor} = 0.027), 'head and neck' (p_{cor} = 0.042), and a nominally significant overrepresentation for 'peripheral nervous system (0.057) (supplement and ST_04) suggesting relative specificity of prioritized genes to clinical conditions involving these systems.

Clinical significance of prioritized genes: Variants in 20 prioritized OMIM-CNS genes were noted in 14 of the 17 multi-sample F-SMI. Clinical features that overlapped with those described in the primary OMIM syndrome were noted in members of four of these families (table 3, Figure 1c, 1d).

Functional significance of prioritized genes: Seventy-nine prioritized genes in the within pedigree analysis were significantly more conserved, in comparison with the background list of 20164 remaining protein coding genes - mean (SD) conservation score 0.63 (0.13) vs 0.59 (0.18), p =0.014. Of these, the 20 OMIM-CNS genes were significantly more mutation intolerant as compared to the other 59 genes, as suggested by lower LOEUF score (LOEUF mean (SD) 0.6 (0.32) vs 0.99 (0.48), p = 0.001) and a higher pLI score (pLI scores mean (SD) 0.41 (0.49) vs 0.16 (0.36), p = 0.05). The mean expression of these 20 genes was also significantly higher compared to remaining 59 genes in the brain cortex (mean (SD) 15.9 (15.3) vs 5.2 (7.5), p = 0.006); and rest of the brain (mean (SD) 16.9 (15.2) vs 6.02 (8.4), p = 0.005).

Across pedigree association analysis using SKAT: We performed a gene-set-wise association analysis for higher burden of RPD variants in a synaptic gene-set consisting of 1233 genes derived from SynGO database (Koopmans et al, 2019) and the same OMIM-CNS gene-set (2450 genes). In the SKAT analysis accounting for kinship between cases and population controls (supplement), a statistically significant association was noted for synaptic gene-set (case burden = 172, population control burden = 41, SKAT-p = 0.017). The association for synaptic gene-set was significant in a sensitivity analysis employing 10*6 permutation tests on SKAT. The association for OMIM CNS gene-set was not statistically significant (SKAT-p = 0.19).

We performed a preliminary gene-wise association analysis to examine genes with higher burden of RPD variants compared to controls, within the limited power offered by the study sample. We noted a nominally significant association in gene-wise SKAT in *CTBP2*, *ZNF717*, *SLC9B1*, *POTEE*, *CISD2*, and *ABCD1* genes (table 4, supplement and ST_05).

Discussion

Employing WES in families with a high frequency of SMIs, we identify segregation of rare deleterious variants in several genes. A significant proportion of these genes are implicated in neuropsychiatric syndromes with Mendelian inheritance, such as *(GEMIN5, COASY, CACNA1B* etc, see table 2). Majority (78/79) of the prioritized variants were noted to be private to fourteen multi-sample pedigrees. In four families that harbored a variant in an OMIM-CNS gene, a review of health records revealed evidence of clinical features that overlapped with the primary OMIM syndrome. The 79 genes prioritized by pedigree analysis tended to be more conserved, and more intolerant of variation, when compared to the remaining set of 20124 genes. Lastly, in the across-pedigree association analysis, compared to unrelated controls, cases harbored an increased burden of rare deleterious variants in genes involved in the structural and functional integrity of synapses.

This overrepresentation of genes in within pedigree prioritization was largely specific to disorders of 'central nervous system' (CNS), as genes with a clinical synopsis annotation to this category showed the strongest overrepresentation. Among 19 additional clinical synopsis categories, significant overrepresentation was also noted for 'head and neck'. The overlap with 'head and neck' may be explained by tightly interlinked developmental and molecular processes that regulate cranio-facial and brain development (Naqvi et al, 2021). These results are broadly in keeping with our previous findings, in a smaller WES study (Ganesh et al, 2019).

Among the 20 variants in genes implicated in CNS syndromes, three variants had been previously reported in the ClinVar database in the context of the primary Mendelian syndrome

(Table 3). The variants in *ADAMTS2* and *JAG1* were noted to be of 'uncertain significance' and *AFG3L2* variant was recorded as 'likely benign' (Richards et al, 2015). The remaining 17 variants have not been previously reported in the ClinVar database, in the context of the primary Mendelian syndrome. Interestingly, a duplication at the same site where we find a G allele insertion (Chr21: 47860904) in the *PCNT* gene implicated in Microcephalic osteodysplastic primordial dwarfism type II, has been identified as a pathogenic variant in ClinVar, potentially suggesting the significance of this site in protein function.

Pleiotropic influence for *de novo* variants and genes between SCZ, BD, autism and neurodevelopmental syndromes has been reported in large trio-based studies (Howrigan et al, 2020; Nishioka et al, 2021; Rees et al, 2021). Our analysis provides further indication for genetic pleiotropy between Mendelian monogenic syndromes and complex SMI syndromes that aggregate in families. None of the ten families with variants in genes implicated in Mendelian neuropsychiatric syndrome had the 'classical features of the primary syndrome. However, in four of these families we noted partial overlaps in clinical features, in some of the individuals with SMI (table 3), suggesting incomplete penetrance and expression of the primary syndrome, and possible pleotropic expression of a psychiatric syndrome.

To examine if the observed pleiotropy between monogenic and complex SMI was specific to the sample structure, population, or variant/gene prioritization approach we curated a list of 142 genes identified in 10 recent NGS studies in SCZ, BD, OCD, AD in the literature (ST_06). With 37 of 142 (26.1%) genes implicated in a Mendelian syndrome, we noted a statistically significant overrepresentation (Fisher's Exact test OR = 2.29, $p < 0.0001$). Despite substantial heterogeneity in the methods across these 10 studies, consistent evidence of overlap of risk between Mendelian and complex disease, at the level of genes harboring rare variants, provides further reason to suspect pleotropic influence amongst these syndromes.

Thus, while certain mutations in Mendelian disease genes result in early onset severe neurodevelopmental phenotype, other variants in the same genes may result in more subtle anatomical and functional consequences, that perhaps predispose to late-onset neuropsychiatric disorders (Zhu et al, 2014). These phenotype level effects of a given mutation, in a given gene, may further be moderated by background genetic effects (Chen et al, 2016; Riordan & Nadeau, 2017) and the role of the gene in neurodevelopment over time (Ball et al, 2020). Lastly, behavioral symptoms are frequently encountered with many Mendelian neurodevelopmental syndromes, and familial SMI may represent a cumulative consequence of multiple rare variants in more than one gene (Wang et al, 2021).

We noted a higher burden of RPD variants in genes related to synaptic function, in the across-pedigree association analysis using SKAT. Both family-based and population-based genetic studies probing common and rare variants across a spectrum of SMI syndromes have implicated multiple synaptic genes with consequences on the structure and function of synapse (Forrest et al, 2018; Hall et al, 2015; Nishioka et al, 2021; Taoufik et al, 2018). Similar to the case-control WES study in South African Xhosa (Gulsuner et al, 2020), we observed rare predicted-deleterious variants in several (119) genes related to synaptic structure and function. Of these, variants in genes represented by the GO component 'integral component of presynaptic membrane' (18/119) and 'integral component of post synaptic membrane' (17/119), contributed the greatest proportion. Most (95%) of the rare-predicted deleterious variants in synaptic genes were private to a pedigree or an individual with only 6 genes having variants across two or more pedigrees (ST_07).

Nominally significant associations were noted in the gene-wise association test using SKAT in the across pedigree analysis for *CTBP2*, *ZNF717*, *SLC9B1*, *POTEE*, and *ABCD1* genes. *CTBP2* gene codes for C-terminal binding protein 2, an isoform of which is a major component of synaptic ribbons, and it also plays a critical role in regulating cell migration during neocortical

development (Wang et al, 2019). The *ZNF717* gene encodes a KRAB- zinc finger transcription factor that has been hypothesized to be critical for primate cortical evolution, specifically differentiating human cortical transcription factor expression from that of chimpanzees (Nowick et al, 2009). Post-zygotic mutations in *ZNF717* have also been proposed to explain the discordance for expression of schizophrenia among monozygotic twins (Castellani et al, 2017). *SLC9B1* gene was within the cis regulatory region of a genome wide significant association locus in a recently published GWAS of schizophrenia (Li et al, 2017). *ABCD1* gene on chromosome X encodes peroxisomal membrane protein called adrenoleukodystrophy protein involved in the transport of very long chain fatty acids. More than 650 mutations in this gene have been implicated in adrenoleukodystrophy syndrome, with highly variable clinical manifestations that often include cognitive and neuropsychiatric symptoms (Engelen et al, 2012). In summary, five of the six nominally significant genes with a higher variant burden among cases, identified in across-pedigree association analysis, had evidence in the literature with relevance to SMI and neurodevelopment.

Some limitations are to be considered while interpreting the results of our study. While each F-SMI pedigree had multiple members affected with SMI, we could sample 3 or more affected individuals in only 17 such pedigrees. Hence the within-pedigree segregation analysis could only be performed in this subsample, reducing the power of this analysis to detect additional disease relevant variants. Augmenting sampling in remaining pedigrees may yield larger number of SMI relevant signals. The alignment and variant calling were performed with hg19, an earlier version of reference sequence for the human genome. A recent study has demonstrated that 0.9% of exome targets and up to 206 genes may fall in regions that are susceptible to discrepancies in the reference assemblies between hg19 and hg38 (Li et al, 2021). We verified the prioritized genes in within-family segregation and across family association analysis and none of these overlapped with the discrepant genes or regions reported by Li et al. While our

sample was adequately powered to confirm across pedigree association in pre-identified gene sets, the novel gene-wise associations identified in this analysis are to be treated as tentative, and need confirmation in larger samples. However, five of the six genes where RPD were identified in cases, have a significant role in neurodevelopment and brain function.

Conclusion

In this WES study in F-SMI, we identify several private, rare, protein altering variants that segregate among the cases within a pedigree. These variants are overrepresented among genes implicated in monogenic forms of Mendelian neuropsychiatric syndromes suggesting pleotropic influences in neurodevelopment and functioning. We also note a greater frequency of variants in genes involved in structural and functional integrity of synapses in cases. The study demonstrates the usefulness of NGS approaches in F-SMI to identify disease relevant variants. Future WES studies in larger number of pedigrees as well as denser pedigrees will better elucidate the contribution of rare coding variant to F-SMI. Validation of the functional impact of identified variants in Mendelian genes in cell models, combined with the application of in silico approaches to model protein structure alterations may help elucidate convergent and divergent developmental mechanisms that underlie Mendelian genes to complex SMI.

References

Ball, G., Seidlitz, J., O'Muircheartaigh, J., Dimitrova, R., Fenchel, D., Makropoulos, A., Christiaens, D., Schuh, A., Passerat-Palmbach, J., Hutter, J., Cordero-Grande, L., Hughes, E., Price, A., Hajnal, J. V., Rueckert, D., Robinson, E. C. & Edwards, A. D. (2020) Cortical morphology at birth reflects spatiotemporal patterns of gene expression in the fetal human brain. *PLoS Biol*, 18(11), e3000976.

Brainstorm-Consortium, Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G. J., Gormley, P., Malik, R., Patsopoulos, N. A., Ripke, S., Wei, Z., Yu, D., Lee, P. H., Turley, P., Grenier-Boley, B., Chouraki, V., Kamatani, Y., Berr, C., Letenneur, L., Hannequin, D., Amouyel, P., Boland, A., Deleuze, J. F., Duron, E., Vardarajan, B. N., Reitz, C., Goate, A. M., Huentelman, M. J., Kamboh, M. I., Larson, E. B., Rogaeva, E., St George-Hyslop, P., Hakonarson, H., Kukull, W. A., Farrer, L. A., Barnes, L. L., Beach, T. G., Demirci, F. Y., Head, E., Hulette, C. M., Jicha, G. A., Kauwe, J. S. K., Kaye, J. A., Leverenz, J. B., Levey, A. I., Lieberman, A. P., Pankratz, V. S., Poon, W. W., Quinn, J. F., Saykin, A. J., Schneider, L. S., Smith, A. G., Sonnen, J. A., Stern, R. A., Van Deerlin, V. M., Van Eldik, L. J., Harold, D., Russo, G., Rubinsztein, D. C., Bayer, A., Tsolaki, M., Proitsi, P., Fox, N. C., Hampel, H., Owen, M. J., Mead, S., Passmore, P., Morgan, K., Nothen, M. M., Rossor, M., Lupton, M. K., Hoffmann, P., Kornhuber, J., Lawlor, B., McQuillin, A., Al-Chalabi, A., Bis, J. C., Ruiz, A., Boada, M., Seshadri, S., Beiser, A., Rice, K., van der Lee, S. J., De Jager, P. L., Geschwind, D. H., Riemenschneider, M., Riedel-Heller, S., Rotter, J. I., Ransmayr, G., Hyman, B. T., Cruchaga, C., Alegret, M., Winsvold, B., Palta, P., Farh, K. H., Cuenca-Leon, E., Furlotte, N., et al (2018) Analysis of shared heritability in common disorders of the brain. *Science*, 360(6395).

Castellani, C. A., Melka, M. G., Gui, J. L., Gallo, A. J., O'Reilly, R. L. & Singh, S. M. (2017) Post-zygotic genomic changes in glutamate and dopamine pathway genes may explain discordance of monozygotic twins for schizophrenia. *Clin Transl Med*, 6(1), 43.

Chen, R., Shi, L., Hakenberg, J., Naughton, B., Sklar, P., Zhang, J., Zhou, H., Tian, L., Prakash, O., Lemire, M., Sleiman, P., Cheng, W. Y., Chen, W., Shah, H., Shen, Y., Fromer, M., Omberg, L., Deardorff, M. A., Zackai, E., Bobe, J. R., Levin, E., Hudson, T. J., Groop, L., Wang, J., Hakonarson, H., Wojcicki, A., Diaz, G. A., Edelmann, L., Schadt, E. E. & Friend, S. H. (2016) Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat Biotechnol*, 34(5), 531-8.

Chun, S. & Fay, J. C. (2009) Identification of deleterious mutations within three human genomes. *Genome Research*, 19(9), 1553-1561.

Cristino, A. S., Williams, S. M., Hawi, Z., An, J. Y., Bellgrove, M. A., Schwartz, C. E., Costa Lda, F. & Claudianos, C. (2014) Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Mol Psychiatry*, 19(3), 294-301.

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. & Liu, X. (2015) Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*, 24(8), 2125-37.

Engelen, M., Kemp, S., de Visser, M., van Geel, B. M., Wanders, R. J., Aubourg, P. & Poll-The, B. T. (2012) X-linked adrenoleukodystrophy (X-ALD): clinical presentation and guidelines for diagnosis, follow-up and management. *Orphanet J Rare Dis*, 7, 51.

Forrest, M. P., Parnell, E. & Penzes, P. (2018) Dendritic structural plasticity and neuropsychiatric disease. *Nat Rev Neurosci*, 19(4), 215-234.

Forstner, A. J., Fischer, S. B., Schenk, L. M., Strohmaier, J., Maaser-Hecker, A., Reinbold, C. S., Sivalingam, S., Hecker, J., Streit, F., Degenhardt, F., Witt, S. H., Schumacher, J., Thiele, H., Nurnberg, P., Guzman-Parra, J., Orozco Diaz, G., Auburger, G., Albus, M., Borrmann-Hassenbach, M., Gonzalez, M. J., Gil Flores, S., Cabaleiro Fabeiro, F. J., Del Rio Noriega, F., Perez Perez, F., Haro Gonzalez, J., Rivas, F., Mayoral, F., Bauer, M., Pfennig, A., Reif, A., Herms, S., Hoffmann, P., Pirooznia, M., Goes, F. S., Rietschel, M., Nothen,

M. M. & Cichon, S. (2020) Whole-exome sequencing of 81 individuals from 27 multiply affected bipolar disorder families. *Transl Psychiatry*, 10(1), 57.
Fromer, M., Pocklington, A. J., Kavanagh, D. H., Williams, H. J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D. M., Carrera, N., Humphreys, I., Johnson, J. S., Roussos, P., Barker, D. D., Banks, E., Milanova, V., Grant, S. G., Hannon, E., Rose, S. A., Chambert, K., Mahajan, M., Scolnick, E. M., Moran, J. L., Kirov, G., Palotie, A., McCarroll, S. A., Holmans, P., Sklar, P., Owen, M. J., Purcell, S. M. & O'Donovan, M. C. (2014) De novo mutations in schizophrenia implicate synaptic networks. *Nature*, 506(7487), 179-84.
Fuller, Z. L., Berg, J. J., Mostafavi, H., Sella, G. & Przeworski, M. (2019) Measuring intolerance to mutation in human genetics. *Nat Genet*, 51(5), 772-776.
Gandal, M. J., Haney, J. R., Parikshak, N. N., Leppa, V., Ramaswami, G., Hartl, C., Schork, A. J., Appadurai, V., Buil, A., Werge, T. M., Liu, C., White, K. P., CommonMind, C., Psych, E. C., i, P.-B. W. G., Horvath, S. & Geschwind, D. H. (2018) Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*, 359(6376), 693-697.
Ganesh, S., Ahmed, P. H., Nadella, R. K., More, R. P., Seshadri, M., Viswanath, B., Rao, M., Jain, S., Consortium, A. & Mukherjee, O. (2019) Exome sequencing in families with severe mental illness identifies novel and rare variants in genes implicated in Mendelian neuropsychiatric syndromes. *Psychiatry Clin Neurosci*, 73(1), 11-19.
Ganna, A., Satterstrom, F. K., Zekavat, S. M., Das, I., Kurki, M. I., Churchhouse, C., Alfoldi, J., Martin, A. R., Havulinna, A. S., Byrnes, A., Thompson, W. K., Nielsen, P. R., Karczewski, K. J., Saarentaus, E., Rivas, M. A., Gupta, N., Pietilainen, O., Emdin, C. A., Lescai, F., Bybjerg-Grauholm, J., Flannick, J., Go, T. D. T. D. G. C., Mercader, J. M., Udler, M., Project, S. C. H. I. E. S., FinMetSeq, C., i, P.-B. C., Laakso, M., Salomaa, V., Hultman, C., Ripatti, S., Hamalainen, E., Moilanen, J. S., Korkko, J., Kuismin, O., Nordentoft, M., Hougaard, D. M., Mors, O., Werge, T., Mortensen, P. B., MacArthur, D., Daly, M. J., Sullivan, P. F., Locke, A. E., Palotie, A., Borglum, A. D., Kathiresan, S. & Neale, B. M. (2018) Quantifying the Impact of Rare and Ultra-rare Coding Variation across the Phenotypic Spectrum. *Am J Hum Genet*, 102(6), 1204-1211.
Genereux, D. P., Serres, A., Armstrong, J., Johnson, J., Marinescu, V. D., Murén, E., Juan, D., Bejerano, G., Casewell, N. R., Chemnick, L. G., Damas, J., Di Palma, F., Diekhans, M., Fiddes, I. T., Garber, M., Gladyshev, V. N., Goodman, L., Haerty, W., Houck, M. L., Hubley, R., Kivioja, T., Koepfli, K.-P., Kuderna, L. F. K., Lander, E. S., Meadows, J. R. S., Murphy, W. J., Nash, W., Noh, H. J., Nweeia, M., Pfenning, A. R., Pollard, K. S., Ray, D. A., Shapiro, B., Smit, A. F. A., Springer, M. S., Steiner, C. C., Swofford, R., Taipale, J., Teeling, E. C., Turner-Maier, J., Alfoldi, J., Birren, B., Ryder, O. A., Lewin, H. A., Paten, B., Marques-Bonet, T., Lindblad-Toh, K., Karlsson, E. K. & Zoonomia, C. (2020) A comparative genomics multitool for scientific discovery and conservation. *Nature*, 587(7833), 240-245.
Genovese, G., Fromer, M., Stahl, E. A., Ruderfer, D. M., Chambert, K., Landen, M., Moran, J. L., Purcell, S. M., Sklar, P., Sullivan, P. F., Hultman, C. M. & McCarroll, S. A. (2016) Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat Neurosci*, 19(11), 1433-1441.
Glahn, D. C., Nimgaonkar, V. L., Raventos, H., Contreras, J., McIntosh, A. M., Thomson, P. A., Jablensky, A., McCarthy, N. S., Charlesworth, J. C., Blackburn, N. B., Peralta, J. M., Knowles, E. E. M., Mathias, S. R., Ament, S. A., McMahon, F. J., Gur, R. C., Bucan, M., Curran, J. E., Almasy, L., Gur, R. E. & Blangero, J. (2019) Rediscovering the value of families for psychiatric genetics research. *Mol Psychiatry*, 24(4), 523-535.
Grotzinger, A. D. (2021) Shared genetic architecture across psychiatric disorders. *Psychol Med*, 1-7.
Gulsuner, S., Stein, D. J., Susser, E. S., Sibeko, G., Pretorius, A., Walsh, T., Majara, L., Mndini, M. M., Mqulwana, S. G., Ntola, O. A., Casadei, S., Ngqengelele, L. L., Korchina, V., van der Merwe, C., Malan, M., Fader, K. M., Feng, M., Willoughby, E., Muzny, D., Baldinger, A., Andrews, H. F., Gur, R. C., Gibbs, R. A., Zingela, Z., Nagdee, M., Ramesar, R. S., King, M. C. & McClellan, J. M. (2020) Genetics of schizophrenia in the South African Xhosa. *Science*, 367(6477), 569-573.

Hall, J., Trent, S., Thomas, K. L., O'Donovan, M. C. & Owen, M. J. (2015) Genetic risk for schizophrenia: convergence on synaptic pathways involved in plasticity. *Biol Psychiatry*, 77(1), 52-8.
Halvorsen, M., Samuels, J., Wang, Y., Greenberg, B. D., Fyer, A. J., McCracken, J. T., Geller, D. A., Knowles, J. A., Zoghbi, A. W., Pottinger, T. D., Grados, M. A., Riddle, M. A., Bienvenu, O. J., Nestadt, P. S., Krasnow, J., Goes, F. S., Maher, B., Nestadt, G. & Goldstein, D. B. (2021) Exome sequencing in obsessive-compulsive disorder reveals a burden of rare damaging coding variants. *Nat Neurosci*, 24(8), 1071-1076.
Howrigan, D. P., Rose, S. A., Samocha, K. E., Fromer, M., Cerrato, F., Chen, W. J., Churchhouse, C., Chambert, K., Chandler, S. D., Daly, M. J., Dumont, A., Genovese, G., Hwu, H. G., Laird, N., Kosmicki, J. A., Moran, J. L., Roe, C., Singh, T., Wang, S. H., Faraone, S. V., Glatt, S. J., McCarroll, S. A., Tsuang, M. & Neale, B. M. (2020) Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nat Neurosci*, 23(2), 185-193.
Huang, M.-H., Cheng, C.-M., Tsai, S.-J., Bai, Y.-M., Li, C.-T., Lin, W.-C., Su, T.-P., Chen, T.-J. & Chen, M.-H. (2020) Familial coaggregation of major psychiatric disorders among first-degree relatives of patients with obsessive-compulsive disorder: a nationwide study. *Psychological Medicine*, 51(4), 680-687.
Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., Sealock, J., Karlsson, I. K., Hagg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I. S., Andersen, F., Bergh, S., Bettella, F., Bjornsson, S., Braekhus, A., Brathen, G., de Leeuw, C., Desikan, R. S., Djurovic, S., Dumitrescu, L., Fladby, T., Hohman, T. J., Jonsson, P. V., Kiddle, S. J., Rongve, A., Saltvedt, I., Sando, S. B., Selbaek, G., Shoai, M., Skene, N. G., Snaedal, J., Stordal, E., Ulstein, I. D., Wang, Y., White, L. R., Hardy, J., Hjerling-Leffler, J., Sullivan, P. F., van der Flier, W. M., Dobson, R., Davis, L. K., Stefansson, H., Stefansson, K., Pedersen, N. L., Ripke, S., Andreassen, O. A. & Posthuma, D. (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*, 51(3), 404-413.
Jia, X., Goes, F. S., Locke, A. E., Palmer, D., Wang, W., Cohen-Woods, S., Genovese, G., Jackson, A. U., Jiang, C., Kvale, M., Mullins, N., Nguyen, H., Pirooznia, M., Rivera, M., Ruderfer, D. M., Shen, L., Thai, K., Zawistowski, M., Zhuang, Y., Abecasis, G., Akil, H., Bergen, S., Burmeister, M., Chapman, S., DelaBastide, M., Jureus, A., Kang, H. M., Kwok, P. Y., Li, J. Z., Levy, S. E., Monson, E. T., Moran, J., Sobell, J., Watson, S., Willour, V., Zollner, S., Adolfsson, R., Blackwood, D., Boehnke, M., Breen, G., Corvin, A., Craddock, N., DiFlorio, A., Hultman, C. M., Landen, M., Lewis, C., McCarroll, S. A., Richard McCombie, W., McGuffin, P., McIntosh, A., McQuillin, A., Morris, D., Myers, R. M., O'Donovan, M., Ophoff, R., Boks, M., Kahn, R., Ouwehand, W., Owen, M., Pato, C., Pato, M., Posthuma, D., Potash, J. B., Reif, A., Sklar, P., Smoller, J., Sullivan, P. F., Vincent, J., Walters, J., Neale, B., Purcell, S., Risch, N., Schaefer, C., Stahl, E. A., Zandi, P. P. & Scott, L. J. (2021) Investigating rare pathogenic/likely pathogenic exonic variation in bipolar disorder. *Mol Psychiatry*.
Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S., Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M. E., Genome Aggregation Database, C., Neale, B. M., Daly, M. J. & MacArthur, D. G. (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434-443.
Kato, T. (2015) Whole genome/exome sequencing in mood and psychotic disorders. *Psychiatry Clin Neurosci*, 69(2), 65-76.
Kenny, E. M., Cormican, P., Furlong, S., Heron, E., Kenny, G., Fahey, C., Kelleher, E., Ennis, S., Tropea, D., Anney, R., Corvin, A. P., Donohoe, G., Gallagher, L., Gill, M. & Morris, D. W. (2014) Excess of rare novel

loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Mol Psychiatry*, 19(8), 872-9.
Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M. P., Cornelisse, L. N., Farrell, R. J., Goldschmidt, H. L., Howrigan, D. P., Hussain, N. K., Imig, C., de Jong, A. P. H., Jung, H., Kohansalnodehi, M., Kramarz, B., Lipstein, N., Lovering, R. C., MacGillavry, H., Mariano, V., Mi, H., Ninov, M., Osumi-Sutherland, D., Pielot, R., Smalla, K. H., Tang, H., Tashman, K., Toonen, R. F. G., Verpelli, C., Reig-Viader, R., Watanabe, K., van Weering, J., Achsel, T., Ashrafi, G., Asi, N., Brown, T. C., De Camilli, P., Feuermann, M., Foulger, R. E., Gaudet, P., Joglekar, A., Kanellopoulos, A., Malenka, R., Nicoll, R. A., Pulido, C., de Juan-Sanz, J., Sheng, M., Sudhof, T. C., Tilgner, H. U., Bagni, C., Bayes, A., Biederer, T., Brose, N., Chua, J. J. E., Dieterich, D. C., Gundelfinger, E. D., Hoogenraad, C., Huganir, R. L., Jahn, R., Kaeser, P. S., Kim, E., Kreutz, M. R., McPherson, P. S., Neale, B. M., O'Connor, V., Posthuma, D., Ryan, T. A., Sala, C., Feng, G., Hyman, S. E., Thomas, P. D., Smit, A. B. & Verhage, M. (2019) SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron*, 103(2), 217-234 e4.
Kumar, P., Henikoff, S. & Ng, P. C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7), 1073-1081.
Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., Team, N. G. E. S. P.-E. L. P., Christiani, D. C., Wurfel, M. M. & Lin, X. (2012) Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet*, 91(2), 224-37.
Lee, S. S. a. Z. Z. (2020) *SNP-Set (Sequence) Kernel Association Test*.
Legge, S. E., Santoro, M. L., Periyasamy, S., Okewole, A., Arsalan, A. & Kowalec, K. (2021) Genetic architecture of schizophrenia: a review of major advancements. *Psychol Med*, 1-10.
Li, H., Dawood, M., Khayat, M. M., Farek, J. R., Jhangiani, S. N., Khan, Z. M., Mitani, T., Coban-Akdemir, Z., Lupski, J. R., Venner, E., Posey, J. E., Sabo, A. & Gibbs, R. A. (2021) Exome variant discrepancies due to reference-genome differences. *Am J Hum Genet*, 108(7), 1239-1250.
Li, M., Shen, L., Chen, L., Huai, C., Huang, H., Wu, X., Yang, C., Ma, J., Zhou, W., Du, H., Fan, L., He, L., Wan, C. & Qin, S. (2020) Novel genetic susceptibility loci identified by family based whole exome sequencing in Han Chinese schizophrenia patients. *Transl Psychiatry*, 10(1), 5.
Li, Z., Chen, J., Yu, H., He, L., Xu, Y., Zhang, D., Yi, Q., Li, C., Li, X., Shen, J., Song, Z., Ji, W., Wang, M., Zhou, J., Chen, B., Liu, Y., Wang, J., Wang, P., Yang, P., Wang, Q., Feng, G., Liu, B., Sun, W., Li, B., He, G., Li, W., Wan, C., Xu, Q., Li, W., Wen, Z., Liu, K., Huang, F., Ji, J., Ripke, S., Yue, W., Sullivan, P. F., O'Donovan, M. C. & Shi, Y. (2017) Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat Genet*, 49(11), 1576-1583.
Maxwell, M. E. (1992) *Family Interview for Genetic Studies (FIGS): a manual for FIGS* [Letter]. 1992.
Naqvi, S., Sleyp, Y., Hoskens, H., Indencleef, K., Spence, J. P., Bruffaerts, R., Radwan, A., Eller, R. J., Richmond, S., Shriver, M. D., Shaffer, J. R., Weinberg, S. M., Walsh, S., Thompson, J., Pritchard, J. K., Sunaert, S., Peeters, H., Wysocka, J. & Claes, P. (2021) Shared heritability of human face and brain shape. *Nat Genet*, 53(6), 830-839.
Nishioka, M., Kazuno, A. A., Nakamura, T., Sakai, N., Hayama, T., Fujii, K., Matsuo, K., Komori, A., Ishiwata, M., Watanabe, Y., Oka, T., Matoba, N., Kataoka, M., Alkanaq, A. N., Hamanaka, K., Tsuboi, T., Sengoku, T., Ogata, K., Iwata, N., Ikeda, M., Matsumoto, N., Kato, T. & Takata, A. (2021) Systematic analysis of exonic germline and postzygotic de novo mutations in bipolar disorder. *Nat Commun*, 12(1), 3750.
Nowick, K., Gernat, T., Almaas, E. & Stubbs, L. (2009) Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc Natl Acad Sci U S A*, 106(52), 22358-63.
Palmer, D. S., Howrigan, D. P., Chapman, S. B., Adolfsson, R., Bass, N., Blackwood, D., Boks, M. P. M., Chen, C.-Y., Churchhouse, C., Corvin, A. P., Craddock, N., Curtis, D., Di Florio, A., Dickerson, F., Goes, F. S.,

Jia, X., Jones, I., Jones, L., Jonsson, L., Kahn, R. S., Landén, M., Locke, A., McIntosh, A., McQuillin, A., Morris, D. W., O'Donovan, M. C., Ophoff, R. A., Owen, M. J., Pedersen, N., Posthuma, D., Reif, A., Risch, N., Schaefer, C., Scott, L., Singh, T., Smoller, J. W., Solomonson, M., St. Clair, D., Stahl, E. A., Vreeker, A., Walters, J., Wang, W., Watts, N. A., Yolken, R., Zandi, P. & Neale, B. M. (2021) Exome sequencing in bipolar disorder reveals shared risk gene AKAP11 with schizophrenia. *medRxiv*, 2021.03.09.21252930.
Prokopenko, D., Morgan, S. L., Mullin, K., Hofmann, O., Chapman, B., Kirchner, R., Alzheimer's Disease Neuroimaging, I., Amberkar, S., Wohlers, I., Lange, C., Hide, W., Bertram, L. & Tanzi, R. E. (2021) Whole-genome sequencing reveals new Alzheimer's disease-associated rare variants in loci related to synaptic function and neuronal development. *Alzheimers Dement*.
Radonjic, N. V., Hess, J. L., Rovira, P., Andreassen, O., Buitelaar, J. K., Ching, C. R. K., Franke, B., Hoogman, M., Jahanshad, N., McDonald, C., Schmaal, L., Sisodiya, S. M., Stein, D. J., van den Heuvel, O. A., van Erp, T. G. M., van Rooij, D., Veltman, D. J., Thompson, P. & Faraone, S. V. (2021) Structural brain imaging studies offer clues about the effects of the shared genetic etiology among neuropsychiatric disorders. *Mol Psychiatry*, 26(6), 2101-2110.
Rees, E., Creeth, H. D. J., Hwu, H. G., Chen, W. J., Tsuang, M., Glatt, S. J., Rey, R., Kirov, G., Walters, J. T. R., Holmans, P., Owen, M. J. & O'Donovan, M. C. (2021) Schizophrenia, autism spectrum disorders and developmental disorders share specific disruptive coding mutations. *Nat Commun*, 12(1), 5353.
Reva, B., Antipin, Y. & Sander, C. (2011) Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research*, 39(17), e118-e118.
Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L. & Committee, A. L. Q. A. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 17(5), 405-24.
Riordan, J. D. & Nadeau, J. H. (2017) From Peas to Disease: Modifier Genes, Network Resilience, and the Genetics of Health. *Am J Hum Genet*, 101(2), 177-191.
Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. (2014) MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods*, 11(4), 361-362.
Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R. & Dunbar, G. C. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*, 59 Suppl 20, 22-33;quiz 34-57.
Singh, T., Neale, B. M., Daly, M. J. & on behalf of the Schizophrenia Exome Meta-Analysis, C. (2020) Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv*, 2020.09.18.20192815.
Sreeraj, V. S., Holla, B., Ithal, D., Nadella, R. K., Mahadevan, J., Balachander, S., Ali, F., Sheth, S., Narayanaswamy, J. C., Venkatasubramanian, G., John, J. P., Varghese, M., Benegal, V., Jain, S., Reddy, Y. J., Consortium, A. & Viswanath, B. (2021a) Psychiatric symptoms and syndromes transcending diagnostic boundaries in Indian multiplex families: The cohort of ADBS study. *Psychiatry Res*, 296, 113647.
Sreeraj, V. S., Puzhakkal, J. C., Holla, B., Nadella, R. K., Sheth, S., Balachander, S., Ithal, D., Ali, F., Viswanath, B., Muralidharan, K., Venkatasubramanian, G., John, J. P., Benegal, V., Murthy, P., Varghese, M., Reddy, Y. J., Jain, S. & Accelerator Program for Discovery in Brain disorders using Stem cells, C. (2021b) Cross-diagnostic evaluation of minor physical anomalies in psychiatric disorders. *J Psychiatr Res*, 142, 54-62.
Sul, J. H., Service, S. K., Huang, A. Y., Ramensky, V., Hwang, S. G., Teshiba, T. M., Park, Y., Ori, A. P. S., Zhang, Z., Mullins, N., Olde Loohuis, L. M., Fears, S. C., Araya, C., Araya, X., Spesny, M., Bejarano, J., Ramirez, M., Castrillon, G., Gomez-Makhinson, J., Lopez, M. C., Montoya, G., Montoya, C. P., Aldana, I., Escobar, J. I., Ospina-Duque, J., Kremeyer, B., Bedoya, G., Ruiz-Linares, A., Cantor, R. M., Molina, J.,

Coppola, G., Ophoff, R. A., Macaya, G., Lopez-Jaramillo, C., Reus, V., Bearden, C. E., Sabatti, C. & Freimer, N. B. (2020) Contribution of common and rare variants to bipolar disorder susceptibility in extended pedigrees from population isolates. *Transl Psychiatry*, 10(1), 74.
Taoufik, E., Kouroupi, G., Zygogianni, O. & Matsas, R. (2018) Synaptic dysfunction in neurodegenerative and neurodevelopmental diseases: an overview of induced pluripotent stem-cell-based disease models. *Open Biol*, 8(9).
Toma, C., Shaw, A. D., Allcock, R. J. N., Heath, A., Pierce, K. D., Mitchell, P. B., Schofield, P. R. & Fullerton, J. M. (2018) An examination of multiple classes of rare variants in extended families with bipolar disorder. *Transl Psychiatry*, 8(1), 65.
Viswanath, B., Rao, N. P., Narayanaswamy, J. C., Sivakumar, P. T., Kandasamy, A., Kesavan, M., Mehta, U. M., Venkatasubramanian, G., John, J. P., Mukherjee, O., Purushottam, M., Kannan, R., Mehta, B., Kandavel, T., Binukumar, B., Saini, J., Jayarajan, D., Shyamsundar, A., Moirangthem, S., Vijay Kumar, K. G., Thirthalli, J., Chandra, P. S., Gangadhar, B. N., Murthy, P., Panicker, M. M., Bhalla, U. S., Chattarji, S., Benegal, V., Varghese, M., Reddy, J. Y. C., Raghu, P., Rao, M. & Jain, S. (2018) Discovery biology of neuropsychiatric syndromes (DBNS): a center for integrating clinical medicine and basic science. *BMC Psychiatry*, 18(1), 106.
Wang, H., Xiao, Z., Zheng, J., Wu, J., Hu, X. L., Yang, X. & Shen, Q. (2019) ZEB1 Represses Neural Differentiation and Cooperates with CTBP2 to Dynamically Regulate Cell Migration during Neocortex Development. *Cell Rep*, 27(8), 2335-2353 e6.
Wang, K., Li, M. & Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, 38(16), e164.
Wang, Y. C., Chen, C. H., Chen, S. J., Cheng, F. Y., Hsu, S. H., Tsai, H. Y. & Cheng, M. C. (2021) Identification of ultra-rare missense mutations associated with familial schizophrenia by whole-exome sequencing. *Schizophr Res*, 235, 60-62.
WHO, W. H. O. (1992) *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*World Health Organization.
Wilfert, A. B., Turner, T. N., Murali, S. C., Hsieh, P., Sulovari, A., Wang, T., Coe, B. P., Guo, H., Hoekzema, K., Bakken, T. E., Winterkorn, L. H., Evani, U. S., Byrska-Bishop, M., Earl, R. K., Bernier, R. A., Consortium, S., Zody, M. C. & Eichler, E. E. (2021) Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat Genet*, 53(8), 1125-1134.
Zhu, X., Need, A. C., Petrovski, S. & Goldstein, D. B. (2014) One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. *Nat Neurosci*, 17(6), 773-81.

Acknowledgements: The authors are grateful to all the patients, their family members and healthy volunteers who participated in the study. Financial support for the study was provided by Department of Biotechnology funded grants - BT/01/CEIB/11/VI/11/2012, entitled, “Targeted generation and interrogation of cellular models and networks in neuro-psychiatric disorders using candidate genes” and BT/PR17316/MED/31/326/2015 entitled, “Accelerator program for discovery in brain disorders using stem cells” (ADBS), Pratiksha Trust and The Institute of Stem Cells and Regenerative Medicine (InStem), Bengaluru, India.

The authors would like to thank the sequencing core facility at the Institute of Genomics and Integrative Biology (IGIB), Delhi (Dr. Faruq Mohammed) and the National Centre for Biological Sciences (NCBS), Bengaluru (Dr. Awadhesh Pandit) for sample processing and WES data generation.

The authors would like to thank Prof Patrick Sullivan, Psychiatric Genomics Consortium, for his inputs in data analysis and interpretation.

The authors would like to thank all investigators of ADBS consortia for providing valuable inputs to the manuscript and having final approval of the manuscript. Suhas Ganesh is currently affiliated with and supported by Schizophrenia Neuropharmacology Research Group at Yale university and is supported by a NARSAD young investigator grant from the Brain and Behavior Research Foundation.

Supplementary material

S1 Derivation of OMIM lists from clinical synopsis terms

S2 OMIM overrepresentation – specificity to CNS

S3 Variant quality assessment with IGV and read depth

Supplementary tables

ST_01: Within pedigree prioritization in multi-sample (≥3) families with SMI – list of 79 prioritized variants and genes

ST_02: Gene level overlaps for prioritized genes in less sample dense (≤2) pedigrees

ST_03: Variant level overlaps for prioritized genes in less sample dense (≤2) pedigrees

ST_04: Specificity of OMIM overrepresentation to Central Nervous System impacting genes

ST_05: Association results for gene-wise sequence kernel association test

ST_06: OMIM syndrome and complex SMI gene level overlap in 10 NGS studies in literature

ST_07: Gene ontology description of variant burden in synaptic genes

Table 1: Demographic and profile of the sample

1a. Demographic profile of cases, family controls and population controls					
Groups	**Cases**	**Family Controls**	**Population Controls**	**F/Chi**	**p**
Sample size	186	64	60		
Age mean (SD)	42.1(16)	44(15.4)	45.6(18.2)	1.17	0.31
Sex male/female	106/80	32/32	29/31	1.85	0.39
1b. Clinical profile by diagnosis					
Diagnosis	**Sample size**	**Age mean (SD)**	**Sex (male/female)**	**AAO mean (SD)**	**DOI mean (SD)**
Schizophrenia	63	37.1(12.9)	35/27	23.1(7.5)	12.4(9.8)
Bipolar disorder	80	39.9(13.5)	45/35	21.9(7.5)	19.9(11.3)
Dementia	24	65.1(14.2)	16/8	62(24)	5(1.4)
OCD	7	33.7(12.1)	2/6	24.4(7.7)	8.4(7.3)
Addiction	7	42.9(13.9)	7/0	33.8(14.4)	9(7.9)
Other/mixed SMI	5	39 (10.8)	1/4	41.6 (14.2)	8(8.2)

Table 2: Variants segregating within families with SMI in genes implicated in Mendelian Syndromes

chr:position	Ref > alt	Gene	Family Number	Burden - case	Gene description	OMIM_syndrome
chr18:12337480	G>C	***AFG3L2***	D002	3	AFG3 like matrix AAA peptidase subunit 2	Spinocerebellar Ataxia 28; SCA28
chr17:3384920	A>G	***ASPA***	D002	3	aspartoacylase	Canavan-Van Bogaert-Bertrand disease - spongy degeneration of central nervous system
chr9:141016353	C>T	***CACNA1B***	D002	4	calcium voltage-gated channel subunit alpha1 B	Neurodevelopmental disorder with seizures and nonepileptic hyperkinetic movements
chr11:64953762	G>A	***CAPN1***	D002	3	calpain 1	Spastic paraplegia 76, autosomal recessive
chr5:154268943	C>T	***GEMIN5***	D002	3	gem nuclear organelle associated protein 5	Neurodevelopmental disorder with cerebellar atrophy and motor dysfunction; nedcam
chr20:10625568	C>G	***JAG1***	D002	5	jagged canonical Notch ligand 1	Alagille syndrome 1
chr1:10035801	C>T	***NMNAT1***	D002	3	nicotinamide nucleotide adenylyltransferase 1	Spondyloepiphyseal dysplasia, sensorineural hearing loss, impaired intellectual development, and leber congenital amaurosis; shilca
chr21:47860904-47860904	- > G	***PCNT***	D003	6	pericentrin	Microcephalic osteodysplastic primordial dwarfism, type II
chr12:120241184	G>A	***CIT***	D004	3	citron rho-interacting serine/threonine kinase	microcephaly 17
chr10:70225532	G>A	***DNA2***	D006	4	DNA replication helicase/nuclease 2	seckel syndrome 8
chr2:166032822	G>A	***SCN3A***	D006	4	sodium voltage-gated channel alpha subunit 3	Epilepsy Familial Focal With Variable Foci 4, Developmental and Epipeltic Encephalopathy 62
chr12:23998998	C>T	***SOX5***	D006	4	SRY-box transcription factor 5	Lamb Shaffer Syndrome
chr2:69409769	G>T	***ANTXR1***	D007	3	ANTXR cell adhesion molecule 1	Growth retardation, alopecia, pseudoanodontia, and optic atrophy
chr17:40715328	C>T	***COASY***	D007	3	Coenzyme A synthase	Neurodegeneration with brain iron accumulation 6, pontocerebellar hypoplasia, type 12
chr3:143567048	C>T	***SLC9A9***	D007	3	solute carrier family	Autism Susceptibility AUTS 16

					9 member A9	
chr6:31827960	G>A	***NEU1***	D008	3	neuraminidase 1	Neuraminidase Deficiency
chr11:66029625	T>C	***KLC2***	D009	3	kinesin light chain 2	Spastic Paraplegia, Optic Atrophy, and Neuropathy; SPOAN
chr5:178581083	C>T	***ADAMTS2***	D010	3	ADAM metallopeptidase with thrombospondin type 1 motif 2	Ehlers-Danlos syndrome, dermatosparaxis type
chr5:154268943	C>G	***GEMIN5***	D012	3	gem nuclear organelle associated protein 5	Neurodevelopmental disorder with Cerebellar Atrophy and Motor dysfunction; NEDCAM
chr6:43018723	C>T	***CUL7***	D013	3	cullin 7	Three M syndrome 1
chr19:50826985	C>T	***KCNC3***	D013	3	potassium voltage-gated channel subfamily C member 3	Spinocerebellar ataxia 13; SCA13

Table 3: Partial expression of symptoms from OMIM syndrome within an SMI family

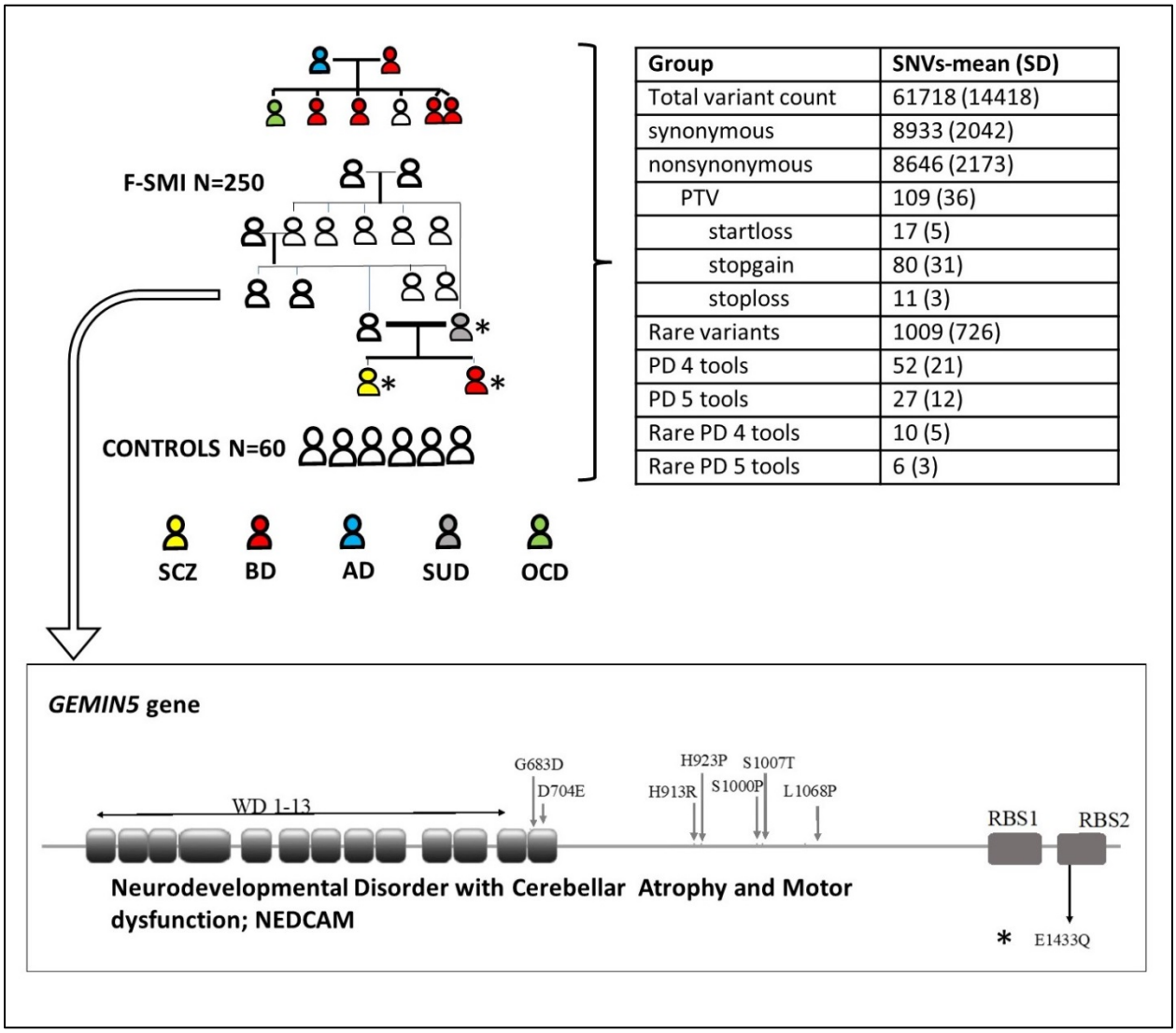
Family Number	**Gene**	**OMIM Clinical Syndrome**	**Partial overlaps in clinical features in WES sample**
D007	*COASY* (OMIM 615643, 618266)	Neurodegeneration with Brain Iron Accumulation 6 (Dystonia, Parkinsonism, Cognitive Decline)	Dystonia with low dose antipsychotic, Tremors
	ANTXR1 (OMIM 230740)	Growth Retardation, Alopecia, Pseudoanodontia and Optic Atrophy (IDD, Delayed motor development)	Intellectual disability
D012	*GEMIN5* (OMIM 619333)	Neurodevelopmental Disorder with Cerebellar Atrophy and Motor Dysfunction (Global developmental delay, IDD, speech Delay)	Learning disability Marfanoid features
D002	*JAG1* (OMIM 118450)	Alagille Syndrome 1 (Absent deep tendon reflexes)	Delayed deep tendon reflexes
D006	*SCN3A* (OMIM 617935, 617938)	Epilepsy Familial Focal with Variable Foci 4 (Learning disability, Speech delay)	Tardive dystonia Coarse tremors Academic difficulties
		Developmental Epileptic Encephalopathy 62 (GDD, hyperkinetic movements)	Tardive dystonia Coarse tremors
	SOX5 (OMIM 616803)	Lamb Shaffer Syndrome (IDD, Speech delay)	Academic difficulties
	DNA2 (OMIM 615807)	Seckel Syndrome (GDD, IDD)	Academic Difficulties

Table 4: Gene-set and gene-wide across pedigree association analysis

Gene-set wise burden					
Gene	**Variant count**	**case burden**	**FC burden**	**PC burden**	**SKAT p value**
Synaptic genes	211	172	63	41	0.014
OMIM CNS genes	636	173	64	59	0.19
Gene wise burden					
Gene	**Variant count**	**case burden**	**FC burden**	**PC burden**	**SKAT p value**
CTBP2	3	149	35	2	0.000001
ZNF717	3	93	44	15	0.000121
SLC9B1	2	100	59	10	0.000448
POTEE	3	69	47	10	0.000627
ABCD1	4	55	32	7	0.013

FC – Family control, PC – Population control, SKAT – sequence kernel association test

Figure 1: Conceptual overview



Group	SNVs-mean (SD)
Total variant count	61718 (14418)
synonymous	8933 (2042)
nonsynonymous	8646 (2173)
PTV	109 (36)
startloss	17 (5)
stopgain	80 (31)
stoploss	11 (3)
Rare variants	1009 (726)
PD 4 tools	52 (21)
PD 5 tools	27 (12)
Rare PD 4 tools	10 (5)
Rare PD 5 tools	6 (3)

Figure legend: Exome sequencing was done for individuals from families with multiple members with serious mental illness (SMI). The Table shows the meaningful (significant) variants detected. Details of one specimen gene (GEMIN5) are represented. The seven variants marked above have been implicated in Neurodevelopmental Disorder with Cerebellar Atrophy and Motor dysfunction: NEDCAM earlier. The variant marked below **rs544452250** (5-154268943-C-G; E144Q) was predicted to be deleterious, and was shared across three individuals with schizophrenia, BPAD and Substance use disorder, in one family.