Schizophrenia Research xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

Whole genome sequencing of 91 multiplex schizophrenia families reveals increased burden of rare, exonic copy number variation in schizophrenia probands and genetic heterogeneity

Fayeza F. Khan [a,*], Phillip E. Melton [a], Nina S. McCarthy [a,b,c], Bharti Morar [b,c,d], John Blangero [e], Eric K. Moses [a], Assen Jablensky [b,c]

[a] *Centre for Genetic Origins of Health and Disease, School of Biomedical Sciences, The University of Western Australia and School of Biomedical Sciences, Curtin University, MRF Building, 50 Murray St., Perth, WA 6000, Australia*
[b] *Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, University of Western Australia, MRF Building, 50 Murray Street, Perth 6000, Australia*
[c] *Cooperative Research Centre for Mental Health, Carlton, South Victoria, Australia*
[d] *Harry Perkins Institute of Medical Research, Centre for Medical Research, The University of Western Australia, 6 Verdun Street, Nedlands, WA 6009, Australia*
[e] *South Texas Diabetes and Obesity Institute, The University of Texas Rio Grande Valley, One West University Blvd, Brownsville, TX 78520, USA*

ARTICLE INFO

Article history:
Received 9 October 2017
Received in revised form 19 December 2017
Accepted 15 February 2018
Available online xxxx

Keywords:
Schizophrenia
Family study
Whole-genome sequence
Copy number variation

ABSTRACT

The importance of genomic copy number variants (CNVs) has long been recognized in the etiology of neurodevelopmental diseases. We report here the results from the CNV analysis of whole-genome sequences from 91 multiplex schizophrenia families. Employing four algorithms (CNVnator, Cn.mops, DELLY and LUMPY) to identify CNVs, we find 1231 rare deletions and 287 rare duplications in 300 individuals (77 with schizophrenia (SZ), 32 with schizoaffective disorder (SAD), 82 with another neuropsychiatric diagnosis and 109 unaffected). The size of the CNVs ranges from a few hundred base-pairs to about 1.3 Mb. The total burden of CNVs does not differ significantly between affected (SZ and SAD) and unaffected individuals. Parent-to-child transmission rate for rare CNVs affecting exonic regions is significantly higher for affected (SZ and SAD) probands as compared to their siblings, but rates for all CNVs is not. We observe heterogeneity between families in terms of genes involved in CNVs, and find several CNVs involving genes previously implicated in either schizophrenia or other neuropsychiatric disorders.

1. Introduction

Schizophrenia (SZ) is a complex neuropsychiatric disorder with about 1% life-time risk and high heritability estimates of up to 80% (Sullivan et al., 2003). Despite the high heritability of SZ and years of genetic studies, only a small proportion of the genetic contribution to its causation has to date been accounted for.

Over the last decade, studies of genomic copy number variants (CNVs) have indicated that CNVs play an important role in the etiology of SZ (Stone et al., 2008) and generally have much higher penetrance than Single Nucleotide Variants (SNVs) (Kirov et al., 2014; Schizophrenia Working Group of the Psychiatric Genomics, 2014). Several SZ-associated CNVs have been found to also increase susceptibility to other neuropsychiatric disorders, such as autism and developmental delay. Interestingly, even when these CNVs do not result in any disorder, they still contribute to cognitive deficiencies in unaffected carriers (Stefansson et al., 2014).

To date CNV studies in SZ have relied on microarray technology for detection. This has meant that the sizes of CNVs detected and investigated has been larger than 10 kb (Malhotra et al., 2011). These studies have shown that both the global burden of rare CNVs and de novo CNVs are increased in SZ and related disorders (International Schizophrenia, 2008; Malhotra et al., 2011). Moreover, due to reduced fecundity and therefore negative selective pressure on causative variants, the highly penetrant CNVs discovered so far have been rare and recurrent. So far, most studies of SZ have been underpowered for individual rare variant detection, however 11 rare CNVs have been shown to increase risk for SZ in microarray studies (Kirov, 2015). All of these SZ-associated CNVs are large (100 s of kb to several Mb in size) and very rare, such that their observed cumulative frequency is <3% in SZ cohorts and even less in matched controls (Kirov, 2015; Rees et al., 2014).

---

* Corresponding author at: 50 Murray Street, MRF Building, Level 5, Perth, 6000, Western Australia, Australia.
E-mail addresses: fayeza.khan@uwa.edu.au (F.F. Khan), phillip.melton@uwa.edu.au (P.E. Melton), nina.mccarthy@uwa.edu.au (N.S. McCarthy), bharti.morar@perkins.uwa.edu.au (B. Morar), john.blangero@utrgv.edu (J. Blangero), eric.moses@uwa.edu.au (E.K. Moses), assen.jablensky@uwa.edu.au (A. Jablensky).

https://doi.org/10.1016/j.schres.2018.02.034

Please cite this article as: Khan, F.F., et al., Whole genome sequencing of 91 multiplex schizophrenia families reveals increased burden of rare, exonic copy number variation in s..., Schizophr. Res. (2018), https://doi.org/10.1016/j.schres.2018.02.034

More recently, the CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium (Cnv et al., 2017) called CNVs from the GWAS data of 41,321 subjects, and genome-wide significance was reached for eight CNV loci and suggestive support found for another eight loci. However, the top eight loci had been implicated previously (Rees et al., 2014). Unlike microarray and SNP datasets, whole genome sequence (WGS) data offers the potential to assay the full spectrum of CNVs across the genome, including smaller CNVs not robustly picked up by microarrays. There are no published studies to date using WGS data to call CNVs in SZ cohorts.

In the current study, we used WGS to detect CNVs (deletions and duplications) in multiplex families with SZ. This dataset provides a comprehensive picture of rare CNVs of all sizes, and the family structure allows the assessment of inherited and de novo CNVs. Having family data also improves detection of rare variants, as multiple family members may carry the same rare variant, increasing our confidence in the results. From our curated list of rare CNVs we test for statistically significant differences in the burden of CNVs between SZ probands and their family members. We also assess de novo CNVs in a smaller subset of families that have quality data from both parents. We compare all SZ probands to their siblings to compute statistical differences in the transmission rate of CNVs from their parents. Finally, we ascertain CNVs that involve previously identified candidate genes for SZ.

2. Methods

2.1. Samples

The Western Australia Family Study of Schizophrenia (WAFSS) has been described in detail elsewhere (Hallmayer et al., 2005). It was initiated in 1996 with the aim of comprehensively assessing families with ≥1 member affected with a disorder within the ICD-10 and DSM-IV schizophrenia spectrum. The majority of probands were recruited from consecutive admissions to a psychiatric hospital.

The present study used a subset of the WAFSS cohort. All families are of European ancestry. Affectedness status was divided into four categories: unaffected, schizophrenia diagnosis (SZ), schizoaffective disorder diagnosis (SAD), and other neuropsychiatric diagnosis (other). The "other" category includes diagnoses of depressive episodes of any severity (21 individuals), adjustment disorder (11 individuals), recurrent depressive disorder (9 individuals), dysthymia (6 individuals), anxiety disorder (6 individuals), two cases each of agoraphobia, avoidant personality, mild cognitive disorder, mixed anxiety and depressive disorder, panic disorder, and bipolar disorder, and single cases of social phobia, Tourette's syndrome, and intermittent explosive disorder.

The average number of members from each family included in the study is 3.3 and the average number of SZ or SAD individuals is 1.2 per family. The breakdown of individuals in this study is as follows: 87 SZ/SAD probands, 93 siblings (13 SZ/SAD, 37 other and 43 unaffected), 117 parents (8 SZ/SAD, 45 other and 64 unaffected), and 3 other relatives (1 SAD and 2 unaffected). There were 13 families with data from one parent and proband, 37 families with data from one parent, proband and at least one sibling, 11 families with trio data, and 19 families with data from trios and at least one sibling. No parent data were available for 7 families and no proband data for four families (filtered in quality control). Complete family data is provided in Supplementary Table 1.

2.2. Whole genome sequencing

Genomic DNA extracted from blood was used to perform WGS in 317 WAFSS participants through the commercial provider Macrogen (South Korea) using Illumina HiSeqX technology. Sequencing was to an average of 16× sequence depth (Supplementary Table 2). Resulting sequencing files were aligned to hg19 using the Isaac aligner (Raczy et al., 2013).

2.3. CNV calling

Four algorithms were chosen to analyse the sorted and indexed BAM files, with duplicate reads marked and unmapped reads removed. Cn.mops (Klambauer et al., 2012) and CNVnator (Abyzov et al., 2011) use read-depth to infer copy number states, while DELLY (Rausch et al., 2012) and LUMPY (Layer et al., 2014) use discordant paired-end and split reads to identify structural variants.

CNV calls from all tools for each sample were merged such that lists of CNVs called by any two tools were obtained for each sample. Further details of the CNV calling and quality control are provided in Supplementary Methods.

2.4. Selecting rare or novel variants

Common variants were removed following comparison with the Database of Genomic Variants (DGV) Gold Standard Variants and 1000 Genomes CNV calls, and using a criterion of 50% or more reciprocal overlap with population CNVs with 1% or higher frequency. BEDTools (Quinlan and Hall, 2010) was used to identify called CNVs that overlapped with variants in databases.

2.5. Reducing false calls

To minimize false calls, rare CNV calls (consensus of at least two tools) from each individual were first used to query single tool calls in family members and non-members for the same or similar breakpoints. If the CNVs were found to be called by only a single tool in members of the same family, they were added to the final list. If the same CNV was found to be called by any number of tools in members of more than two families and was not a known population CNV, it was excluded from further consideration.

2.6. Annotation of CNV calls

ANNOVAR (Wang et al., 2010) was used for the gene-based annotation of all CNV calls using the hg19 refGene database, and transcription factor binding site scoring using the hg19 tfbsConsSites database. Fig. 1 presents the workflow from sequence data to the final annotated list of CNVs.

2.7. DeNovo CNV calling

In families with sequence data available from both parents, CNV calls in progeny were compared against those in the parents. Calls present in parents were ignored. Calls only made in progeny (putative de novo CNVs) were checked for having been called by a single tool in any other family member (parents or siblings) and removed from the de novo list if found to be shared.

2.8. PCR-based validation of CNVs

To verify deletion calls, PCR primers in the flanking sequence and deleted sequence were designed. Amplified products were resolved by agarose gel electrophoresis. Real-time quantitative PCRs with commercially designed primers were used for duplication calls (Supplementary Methods).

2.9. Statistical analysis

Comparisons were made between SZ/SAD cases and unaffected/other family members in terms of burden of total CNVs, exonic CNVs, CNVs larger than 50 kb, CNVs containing a transcription factor binding site (TFBS) and parent-to-child transmission rates of all and just exonic CNVs. Students *t*-tests were performed to compare the two groups. Family structure was accounted for using the R package "kinship2"

Please cite this article as: Khan, F.F., et al., Whole genome sequencing of 91 multiplex schizophrenia families reveals increased burden of rare, exonic copy number variation in s..., Schizophr. Res. (2018), https://doi.org/10.1016/j.schres.2018.02.034