DNA sequence level analyses reveal potential phenotypic modifiers in a large family with psychiatric disorders.

Citation for published version:
Ryan, N, Lihm, J, Kramer, M, McCarthy, S, Morris, S, Arnau soler, A, Davies, G, Duff, B, Ghiban, E, Hayward, C, Deary, I, Blackwood, D, Lawrie, S, McIntosh, A, Evans, K, Porteous, D, McCombie, WR & Thomson, P 2018, 'DNA sequence level analyses reveal potential phenotypic modifiers in a large family with psychiatric disorders.' Molecular Psychiatry. DOI: 10.1038/s41380-018-0087-4

Digital Object Identifier (DOI):
10.1038/s41380-018-0087-4

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published in:
Molecular Psychiatry

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
DNA sequence-level analyses reveal potential phenotypic modifiers in a large family with psychiatric disorders

Niamh M. Ryan1 · Jayon Lihm2 · Melissa Kramer2 · Shane McCarthy2 · Stewart W. Morris1 · Aleix Arnau-Soler1 · Gail Davies3 · Barbara Duff4 · Elena Ghiban2 · Caroline Hayward2 · Ian J. Deary3 · Douglas H. R. Blackwood4 · Stephen M. Lawrie4 · Andrew M. McIntosh3,4 · Kathryn L. Evans1,3 · David J. Porteous4,1,3,6 · W. Richard McCombie2 · Pippa A. Thomson2,1,3

Received: 2 November 2017 / Revised: 30 March 2018 / Accepted: 9 April 2018
© The Author(s) 2018. This article is published with open access

Abstract
Psychiatric disorders are a group of genetically related diseases with highly polygenic architectures. Genome-wide association analyses have made substantial progress towards understanding the genetic architecture of these disorders. More recently, exome- and whole-genome sequencing of cases and families have identified rare, high penetrant variants that provide direct functional insight. There remains, however, a gap in the heritability explained by these complementary approaches. To understand how multiple genetic variants combine to modify both severity and penetrance of a highly penetrant variant, we sequenced 48 whole genomes from a family with a high loading of psychiatric disorder linked to a balanced chromosomal translocation. The (1;11)(q42;q14.3) translocation directly disrupts three genes: DISC1, DISC2, DISC1FP and has been linked to multiple brain imaging and neurocognitive outcomes in the family. Using DNA sequence-level linkage analysis, functional annotation and population-based association, we identified common and rare variants in GRM5 (minor allele frequency (MAF) > 0.05), PDE4D (MAF > 0.2) and CNTN5 (MAF < 0.01) that may help explain the individual differences in phenotypic expression in the family. We suggest that whole-genome sequencing in large families will improve the understanding of the combined effects of the rare and common sequence variation underlying psychiatric phenotypes.

Introduction
Schizophrenia (SCZ), bipolar disorder (BD) and major depressive disorder (MDD) are debilitating psychiatric disorders, with MDD ranked by the World Health Organisation as the single largest contributor to global disability [1]. Family, twin and adoption studies have shown a strong heritable component to these disorders [2, 3], with heritability estimates of 0.37 for MDD, 0.75 for BD and 0.81 for...
SCZ [4]. Genome-wide association studies (GWAS) have identified over a hundred common single nucleotide polymorphisms (SNPs) robustly associated with these disorders, most for SCZ, but all with small effect sizes [5–7]. Individually rare copy number variants from case–control studies and exome sequencing-derived missense variants with large effects have also been reported [8–12]. However, even within families, individuals with these variants show variability in clinical phenotype [13–16]. Thus, the evidence to date suggests that psychiatric disorders have a complex and heterogeneous genetic architecture, with many potential genetic routes leading to the same clinical outcome, some variants associated with a broad range of conditions, and with both common and rare variants playing a role [4, 17].

Family-based studies can be used to study the effects of common and rare variants in the context of a less heterogeneous genetic background, and could allow us to better understand the intra-family variation and reduced penetrance commonly seen in complex traits. We have analysed the whole-genome sequences of 48 out of 107 members of a large Scottish family with a high loading of broad spectrum major mental illness. Members of the family carry a balanced translocation between chromosomes 1 and 11 (1q42; 11q14.3, hereafter t(1;11)) that is significantly linked to major mental illness, with a maximum logarithm of the odds (LOD) score of 7.9 for a broad model including individuals with SCZ, BD, schizoaffective disorder (SCZAFF), recurrent MDD (rMDD), cyclothymia, single episode MDD (sMDD) plus minor diagnoses (including: alcoholism, adolescent conduct disorder and anxiety) [18]. However, nine family members without the translocation also have psychiatric diagnoses, whereas two carriers show no evidence of psychiatric disorder, suggesting that additional factors may be involved in the clinical phenotype. We hypothesised that the variation in phenotypic expression in the family may reflect the inheritance of the translocation and a variable subset of predisposing/modifier variants. We combined genome-wide variance component multipoint linkage, regional two-point linkage and haplotype analyses on the full spectrum of variants within the t(1;11) family to identify additional potential risk/modifier loci and tested the association of these loci in two case–control samples drawn from the UK population.

**Materials and methods**

The materials and methods are described in full in the Supplementary Information.

Details of the translocation family are given in Thomson et al. [18] (see also: Supplementary methods: Diagnoses and phenotypic models in the family). This study was approved by the Multi-Centre Research Ethics Committee in Scotland (09/MRE00/81). All study participants gave their written, informed consent.

DNA samples were sequenced to a median coverage of > 30 over three lanes of a HiSeq2000 using 101 bp paired-end reads. After local realignment around indels and base quality score recalibration of each library, single-nucleotide variants (SNVs) were called with GATK v2.4.9 [19-21], using the multi-sample joint calling mode to achieve consistent calling across samples (Supplementary table 1: GATK summary table; Supplementary figure 1a: Individual sample coverage; Supplementary methods: WGS sequencing and variant calling). Variant quality score recalibration (VQSR) parameters were applied as recommended in the GATK best practices documentation for GATK v2.4.9. The “truth sensitivity filter level” was set at 99.0. Deletions ≥ 1 kb that map to a single genomic location were detected by event-wise testing based on read depth [22] (Supplementary methods: Copy Number Variation (CNV) calling).

Validation of GATK VQSR modelling showed that 94% pass variants and 52% of failed variants were validated by custom designed Taqman assays (Supplementary methods: Variant validation-VQSR filter). All variants were therefore retained initially and an additional filter for Mendelian segregation in the family applied.

Variance component linkage analyses were performed using the SOLAR software package [23] (Supplementary methods: Linkage analysis). LOD scores were adjusted for deviation of the phenotype distribution from normal but, due to the nested nature of the phenotypes, were not adjusted for multiple testing. SNVs within the genomic regions under the linkage peaks were phased using the software SHAPEIT (v2.r83725) and the 1000 Genomes Phase 1 integrated reference panel26, incorporating the pedigree information to increase accuracy (Supplementary methods: Haplotype phasing). Minimal haplotype regions were defined by recombination breakpoints in the affected individuals in the family. Mixed logistic or linear regression models, fitting the inverse relationship matrix as a random effect to control for familial structure, were used to test phenotype associations of haplotypes using ASREML-R (www.vsni.co.uk/software/asreml). The significance of fixed effects within the model was assessed using a conditional Wald F-test. A model p-value of < 0.05 was considered significant. In addition to family structure, age and sex were fitted in analyses of global assessment of functioning (GAF), current IQ and attention/processing speed.

Association studies of affective disorder and related traits were performed in two population-based cohorts: Generation Scotland: Scottish Family Health Study (GS:SFHS) [24–26] and UK Biobank (UKB) [27], fitting principal components and cohort/phenotypes appropriate covariates using subsets of unrelated individuals (see Supplementary methods: UK Population-based cohorts: GS:SFHS and UKB and region-wide association analyses).
Whole-genome sequencing in the translocation family

Whole-genome sequencing was performed on comprising all 48 individuals from the t(1;11) family for whom DNA was available. The final dataset included 9.76 million SNVs present in at least one individual of which 16.46 and 5.80% were not found in the 1000 Genomes phase 3 European [28] or gnomAD non-Finnish European subsets, respectively, and 5.41% were found in neither repository (July 2017; Supplementary figure 1b: Minor allele frequencies (MAF) of SNVs in 1000 Genomes phase 3 European sample (EUR) and GnomAD non-Finnish European (NFE)). In addition, using a read depth approach, 27 unique deletions were identified with sizes ranging from 1.2 to 145 kb.

Evidence for the effects of additional regions on risk of psychiatric disorder in the family

To study the genetic contribution to the range of diagnoses present in this family, nine phenotypic models were evaluated using genome-wide linkage analyses of the 48 t(1;11) family members; of whom 19 are translocation carriers (18 affected; 1 unaffected) and 29 non-carriers (6 affected; 22 unaffected; 1 unknown). Table 1 shows the numbers of cases and controls included in each model (Supplementary Table 2: Phenotype models split by translocation status). The maximum theoretical LOD (mtLOD) and the observed two-point LOD for the translocation was generated for each model using only the set of family members sequenced in this study. The translocation explained the majority of the mtLOD, particularly for Model A: SCZ, BD, SCZAFF (number affected = 6, 6 translocation carriers, 0 non-carriers) and Model H: SCZ and SCZAFF (number affected = 4, 4 translocation carriers, 0 non-carriers). However, the translocation LOD explained less than half of the mtLOD for affective disorders (Model F; number affected = 14, 9 translocation carriers, 5 non-carriers), Psychosis (number affected = 8, 8 translocation carriers, 0 non-carriers), and the models containing minor diagnoses (Model C and Model D; Table 1). This implies additional segregating genetic factors, over and above the translocation, which impact on the clinical presentation for these models.

Multipoint linkage analyses identified four genome-wide significant (LOD ≥ 3.3) peaks: three on the translocation chromosomes: one located on chr1q and two peaks on chr11q (chr11q1 and chr11q2); and a fourth peak on chr5q (Supplementary methods; Fig. 1, Table 2). The chr1q and the chr11q1 linkage peaks show evidence for linkage across all models, with maximum LODs for the broad phenotypic models (Model C: any psychiatric diagnosis, and Model B: SCZ, BD, SCZAFF, rMDD and cyclothymia). In contrast, the chr11q2 and chr5q peaks were only significant for the affective disorder phenotype (Model F: BD, rMDD and MDD). Five further regions with LOD ≥ 2 were also identified, each of which was specific to a phenotype model (Fig. 1, Table 2; plots for all chromosomes are given in Supplementary Figure 2).

Minimum haplotype regions do not span the translocation breakpoints but contribute to prediction of psychiatric disorder in the family

The minimum haplotype regions were defined for each genome-wide significant peak and the diagnoses of the carriers identified (Supplementary Figure 3). Table 2 shows the boundaries of these linkage regions. Information on all the genes in the regions is given in Supplementary Table 3.
The minimum haplotypes under the linkage peaks on chromosomes 1 and 11 (chr1q, chr11q1 and chr11q2), although adjacent to the translocation, do not include the breakpoints (hg19: chr1: 231,950,368 and chr11: 90,361,108; Supplementary Figure 3a and 3b). The maximum multipoint LOD scores in the region of the translocation are driven both by individuals with the translocation, who share a wide flanking region, and by individuals with psychiatric diagnoses who carry discrete regions shared with the translocation carriers (Supplementary Information: Haplotype phasing). The combination of haplotypes carried by each family member and their diagnosis are shown in Supplementary Table 4 (see also Fig. 2). The chr1q haplotype and the chr11q1 haplotypes are defined by all 19 of the translocation carriers and additional affected individuals who do not carry the translocation (Supplementary Figures 3a and 3b). The chr11q2 haplotype is not present in three of the translocation carriers; this region is defined by recombination events in two translocation carriers (ID 18 and 19), beyond the recombination event inherited from the married-in parent, and an individual with rMDD who does not carry the translocation. The chr5q haplotype is shared by 10 translocation carriers and 7 non-carriers (Supplementary Figure 3c).

An indirect effect of the translocation in the family is to “lock together” the translocation, chr1q, chr11q1 and chr11q2 regions in translocation carriers. In translocation carriers, the minimum shared haplotypes spanning the

**Fig. 1** Multipoint linkage analysis of the t(1;11) family. Plots of the chromosomes with multipoint linkage peaks with LOD ≥ 2 and a summary table of the phenotype models (the colour coding reflecting the colours of the models in the multipoint plots). The plots represent multipoint LOD scores vs. chromosome position in Mb. Multipoint LODs = 1, 2 and 3 annotated with grey, blue and red horizontal dotted lines, respectively. The translocation breakpoints, T, are marked by a red vertical line. For plots of all chromosomes, see Supplementary Figure 2.
breakpoints were defined (Supplementary Figures 2d and 2e). The haplotypes shared by the translocation carriers spanned 17.5 Mb on the derived chr1 and 22.2 Mb on the derived chr11. The size of these shared regions is consistent with the lack of evidence for reduced recombination around the translocation reported by He et al. [29].

Sequence-level analyses of the regions on chr1q and chr11q1, shared by the translocated chromosomes that have undergone recombination (IDs 18 and 19; Supplementary Figure 3b) and three non-translocation carriers (siblings with IDs 44 and 47, and the child of 44, ID 87; Fig. 2; Supplementary Table 4: Summary of haplotype segregation in the t(1:11) family), show that these haplotypes are near identical to those on the derived chromosomes. Although it is not possible to definitely show that these haplotypes are inherited from the familial parent, these regions span multiple haplotype blocks in the Scottish population and the combinations found on the derived chromosomes occur at <1% frequency, suggesting that these haplotype arose once in the family. One individual in the family appears to carry two copies that are 97.3% identical across the chr11q1 region suggesting that regions homologous to the chromosomes flanking the translocation are present within the Scottish population.

We next used mixed regression models to investigate the contribution of each haplotype to phenotype prediction. Given the relationships between the family members, and the total number of individuals, the correlation structure between the haplotypes is complex (Supplementary information: Haplotype correlations). This must be considered when interpreting the results of phenotypic prediction using backward selection, as the translocation event, chr1q and chr11q1 regions are likely to explain the same variance in phenotype and be retained randomly in the models (Supplementary Information: Regional correlations; Supplementary Figure 4: Intra-family regional correlations).

Analyses of Model F (affective disorder) and Psychosis suggest that additional haplotypes (chr2p, chr3q and chr5q), in combination with the haplotypes on the derived chromosomes, may contribute to the phenotypic heterogeneity seen in the translocation family (Supplementary Information: Phenotype prediction; Supplementary Table 5a and 5c). Age of onset was available for 23 of 24 of the individuals with a diagnosis of affective disorder (Model F). The minimum models for this trait retained both chr11q2 and chr16p (Supplementary Table 5c). In analyses of translocation carriers only, no individual haplotype predicted phenotype; however, combined analyses of all haplotypes not on the translocated arms (chr5q–chr16p) did predict Model F and Psychosis. One of the haplotypes that contribute to Model F has entered the family through an unaffected individual who married into a sub-branch of the family (chr3q, sub-branch 2b, Fig. 2). In contrast, chr16p

| Region          | Start          | End            | Size       | Max 2pt LOD | Model | Closset gene distance | Top 2 pt LOD | Model | N genes | MAX mLOD | Start and position given for ID19: MAX mLOD, maximum multipoint LOD score, Top 2 pt LOD, two-point LOD score; and the closest gene to the highest 2 pt LOD with the brackets denoting the relative position. Model definitions are shown in Table 1 |
|-----------------|----------------|----------------|------------|-------------|-------|-----------------------|--------------|-------|---------|----------|----------------------------------|
| chr1q           | 223,324,661    | 224,922,062    | 1,597,401  | 5.6         | B     | C/C                   |              |       | 14      | 9.3      | chr1: 223,804,982 [CAPN8: intronic] |
| chr5q           | 58,712,609     | 58,721,609     | 91,000     | 3.5         | F     | F/C                   |              |       | 3      | 8.3      | chr5: 58,378,127 [C5orf64: intronic] |
| chr11q1         | 87,350,867     | 87,357,790     | 71,000     | 5.9         | C     | C/F                   |              |       | 5       | 10.8     | chr11: 88,219,121 [LRRC5: intronic] |
| chr11q2         | 90,730,867     | 90,749,921     | 2,000,000  | 3.7         | F     | F/C                   |              |       | 1       | 8.5      | chr11: 90,927,990 [LRRC7C: intronic] |
| chr1p           | 69,000,000     | 69,000,000     | 100,000    | 2.2         | F     | F/C                   |              |       | 7       | 3.7      | chr1: 69,000,000 [CAPN8: intronic] |
| chr10p          | 162,000,000    | 162,000,000    | 100,000    | 2.3         | C     | C/F                   |              |       | 5       | 2.3      | chr10: 162,000,000 [CAPN8: intronic] |
| chr16p          | 15,800,000     | 15,800,000     | 100,000    | 2.2         | C     | C/F                   |              |       | 6       | 2.2      | chr16: 15,800,000 [CAPN8: intronic] |
| chr16q           | 16,000,000     | 16,000,000     | 100,000    | 2.6         | G     | G/C                   |              |       | 15      | 2.6      | chr16: 16,000,000 [CAPN8: intronic] |

DNA sequence-level analyses reveal potential phenotypic modifiers in a large family with psychiatric...
for Model F and chr2p for psychosis were present in the founders of the family, whereas it is unknown whether chr5q was inherited from the founders or through the non-familial parent founder of branch 2 (Fig. 2: Segregation of linked haplotypes within the pedigree).

Fine-mapping linked regions using two-point linkage analyses

Detailed two-point linkage analysis was performed across the haplotype-associated regions under the four genome-wide significant multipoint linkage peaks and the five regions with multipoint LODs ≥ 2. Two-point linkage was performed on those variants that passed QC filters: Hardy–Weinberg equilibrium (HWE) \( p > 0.001 \) and present in at least three individuals with an additional raw case–control odds ratio filter for the LOD ≥ 2 regions (Supplementary Methods: Linkage analysis; Supplementary Figures 5 and 6: Regional two-point summaries). The top 100 LOD scores for each of the haplotype regions, and for all variants analysed in genome order, are listed in the Supplementary Tables 6a-e.

With the caveat that two-point linkage can be unstable if the population allele frequencies are mis-specified, the top two-point LOD scores and the respective model for each linkage region are shown in Table 2. No two-point LOD score (mLODs) achieved the maximum theoretical LOD scores (mtLODs, Table 1); however, the highest two-point LODs for the chr11q1 (mLOD = 10.8; mtLOD = 12, Model C: any diagnosis), chr11q2 (mLOD = 8.3; mtLOD = 8.9, Model F: affective disorder) and chr4q (mLOD = 4.7; mtLOD = 5.7, Model G: BD and rMDD) regions explain the majority of the maximum theoretical linkage signal. The four genome-wide significant linkage regions and chr16p achieved higher maximum two-point LODs than those of the translocation for the same models (Table 1). Examination of the two-point linkage results across the regions revealed multiple blocks of linkage in each region with evidence that different subregions show strongest linkage under different models (Supplementary Figures 5 and 6; Supplementary Information: Two-point linkage analyses). This may reflect the presence of multiple risk loci within each region.

Many GWAS significant SNPs have been identified in non-coding regions and are predicted to tag regulatory or missense variants [30–33]. Functional annotation of variants under the genome-wide significant linkage peaks detected evidence of missense variants and brain expression quantitative trait loci (eQTL) (Supplementary Information: Functional annotation of variants under the genome-wide...
DNA sequence-level analyses reveal potential phenotypic modifiers in a large family with psychiatric disorders. 

Sequence-level analyses of the linked haplotypes indicated that these haplotypes may be present in the Scottish population, and may be independently associated with risk. Evidence for association between SNVs in the linked regions and affective disorder outside of the family (Model F, MDD and BD combined) was investigated using unrelated individuals from two population-based UK cohorts: GS:SFHS (2060 cases and 4495 controls) and UKB (8294 cases and 15,872 controls) (Supplementary Figures 5 and 6, Supplementary Tables 9a). No SNV reached genome-wide significance, but nominal evidence ($p<5 \times 10^{-4}$) was identified for association of $GRM5$ (chr11q1), $CNTN5$ (chr11q2), $PDE4D$ (chr5q), $LRRC7$ (chr1p) and $VSNL1$ (chr2p) (Table 3).

To investigate the potential pleiotropy of the associated haplotypes, cognitive, personality and mental health-related traits associated with psychiatric disorders were tested for association in GS:SFHS (Supplementary methods: Region-wide association analyses; Table 3; Supplementary Tables 9b; Supplementary Figures 7 and 8). Nominal association ($p<5 \times 10^{-4}$) was identified for multiple traits and genes. Of these, $CNTN5$ in chr11q2 showed evidence for nominal association with number of episodes of depression (SCID), psychological distress (GHQ-28B) and the cognitive variables: digit symbol coding and Mill Hill vocabulary. Similarly, $PDE4D$ in chr5q was nominally associated with MDD, and Mill Hill vocabulary. Multiple cognitive traits were also nominally associated with SNPs in $RAPGEF2$ in chr4q (digit symbol coding, Mill Hill vocabulary and the general cognitive factor $g$).

Although these cross-trait associations are in some cases within the same subregion, as defined by the recombination rates across the region in the 1000 Genome UK population,
for example, for affective disorder, cognitive variable and number of episodes in CNTN5 (Supplementary Figure 7c), or affective disorder, MDD and Mill Hill vocabulary in PDE4D (Supplementary Figure 7d), this is not true for affective disorder and psychological distress in CNTN5 (Supplementary Figure 7c), or affective disorder and MDD in chr4q (Supplementary Figure 8d). These results suggest that variants in the genes within the family-defined haplotypes are associated, at low penetrance, in the UK population and may modulate aspects of phenotype such as number of episodes and cognitive ability.

Evidence of nominal association with cognitive and mental health-related traits were therefore examined in the t (1;11) family. Haplotype carrier status was used to predict global function (GAF) [34], current IQ (Wechsler Abbreviated Scale of Intelligence) [35] and attention/processing speed (Cambridge Neuropsychological Test Automated Battery, CANTAB) [36]. No association was seen with any haplotype and either current IQ or attention/processing speed (p > 0.05). However, chr1q, chr11q1 and chr5q, but not the translocation, predicted GAF (p < 0.05 adjusting for age and sex (Supplementary Table 10: Association of threshold status and all haplotypes retained haplotypes on the derived selection of models containing all three PRS, translocation chromosomes under all four phenotypic models and were the only variables retained in Model B, Model C and Model F (Supplementary Table 5c: Phenotype prediction). PRS were not significantly associated with age of onset in the family, suggesting that there was no evidence of an increased load of common variants modifying the onset of the phenotypes. To test whether there was an accumulation of common risk variants with subsequent generations, ‘generation’ was used to predict PGC PRS. There was no significant association between PGC PRS and generations in the family (PRS-BD p = 0.063, PRS-MDD p = 0.148, PRS-SCZ p = 0.691). These results suggest that common risk variants, particularly those associated with BD and SCZ, may contribute to psychosis within the family, but there is no evidence of accumulation or loss of polygenic risk over generations.

**Discussion**

The evidence for high heritability and familial clustering in psychiatric disorders is counter balanced by the limitations of current diagnostic criteria, over-lapping symptomatology and absence of definitive biomarkers, physiology or pathology. Using whole-genome sequencing, we investigated the existence of disease-modifying loci in a large Scottish pedigree in which a balanced t(1;11) translocation predisposes to major psychiatric disorders. The Scottish t (1;11) family is exceptional because of its size, longitudinal clinical follow-up and detailed molecular genetic study. The foundational finding is of a t(1;11) translocation that disrupts three genes: DISC1, DISC2 and DISC1FP, alters DISC1 expression and results in production of abnormal fusion transcripts [39, 40]. Direct disruption of DISC1 impacts on neurodevelopment, glutamate-signalling, cognitive ability and liability to psychiatric disorder [41–50]. Studies on individuals from the t(1;11) family have identified abnormalities in brain structure particularly white matter integrity [51] and cortical thickness [18, 52], and in brain activation identified through P300 amplitude and latency [53], as well as activation during working memory tasks, and altered glutamate signalling [18]. These functions are congruent with prevailing hypotheses of neurodevelopmental and synaptogenic origins of SCZ and related disorders. However, the penetrance of the t(1;11) is incomplete and the variability in both age of onset and presentation of symptoms remained unexplained.

Comparison of the two-point LOD scores for the translocation and the theoretical maximum LODs across multiple phenotypes indicated that, while the translocation event and its disruption of the breakpoint genes explains the majority of the linkage to SCZ, SCAFF and BD in the family, additional variants predisposing to affective disorder and
DNA sequence-level analyses reveal potential phenotypic modifiers in a large family with psychiatric…

psychosis may be identified through linkage analyses. Genome-wide multipoint linkage analyses identified four genome-wide significant linkage peaks (LOD > 3.3), spanning 11 Mb and 51 genes, and five peaks with LOD scores >2, spanning 35.5 Mb and 285 genes. To prioritise variants, we combined information across linkage analyses in the t(1;11) family, brain cis-eQTLs, and association in two UK population-based cohorts and identified consistent, although nominal, evidence of association between variants in GRM5, PDE4D and CNTN5. All three genes have previously been associated with psychiatric or neurodevelopmental disorders (Supplementary Information: GRM5, PDE4D and CNTN5). PDE4D is a direct protein–protein interactor of DISC1 [54, 55], GRM5 modulates glutamatergic signalling, and CNTN5 is a neurodevelopmental gene implicated in the specification of dendritic arbors [56, 57]. In the family, it is not possible to separate the direct effects of the translocation and the linked loci on the derived chromosomes. However, the identification of these variants suggests that the family may be segregating multiple genetic hits on causal pathways for psychiatric disorders, consistent with the polygenic nature of these complex traits. Studies of the effects of disrupting DISC1 in cell and animal models show a clear effect of the translocation on risk of psychiatric illness through this direct effect [49]. The genetic effects of variants in the linkage regions identified in this study will require similar validation.

Although we have highlighted these three genes, multiple subregions within the linked haplotypes may contain independent variants that, in sum, are responsible for different aspects of clinical presentation in the family (e.g. MAST4 in chr5q and RAPGEF2 in chr4q; Supplementary Figures 7 and 8). Further, we have shown that PRS derived from PGC summary data for BD and SCZ can predict liability to psychosis in the family, suggesting a contribution from additional common variants. Although, it is important to note that all cases of SZ, BD and psychosis carry the mutation caused by the translocation.

This study is limited by the number of family members available for study, restricting the power to detect all variants contributing to the phenotype. However, this is counter balanced by the advantage of reduced genetic heterogeneity and the ability to examine multiple copies of ultra-rare variants in a single pedigree. The contribution of a broad spectrum of allele frequencies to psychiatric disorders, as in other complex traits, argues against the present practise of analysing either very rare or common variants in isolation. It suggests that emerging methods that combine both, such as those searching for compound heterozygosity [58], may be beneficial in predicting complex phenotypes and will allow us to test the prediction that a combination of rare disruptive and common polymorphisms explains a measurable fraction of the individual differences in phenotype. Further, that the separate and joint effects of potentially modulating loci and associated genetic variants need to be assessed for biological consequence and phenotypic impact [59, 60]. We do not rule out the influence of differential environmental exposures on psychiatric outcome, but our study clearly indicates that the high density of psychiatric disorder in the t(1;11) family involves a combination of: direct effects of the translocation on the genes located at the breakpoint (DISC1, DISC2, DISC1FP), the “locking together” of familial risk factors, additional unlinked loci and common risk variants.

In conclusion, although GWAS has identified well over 100 statistically robust findings at the population level for psychiatric disorder, a substantial proportion of variance remains unexplained and this is more so at the individual and family level. Family studies have identified rare, high penetrant variants (akin to the t(1;11), but typically large copy number variants), which have provided valuable pointers to biological targets. Between these two genetic mainstays, classical linkage analysis has had less impact, reflecting the failure to identify consistent linkage at the population level. However, as for the complex neurological trait of Hirschsprung’s Disease [61], our study suggests that an oligogenic model best describes the spectrum of psychiatric presentations in the t(1;11) family. We suggest the application of a combined strategy justifies the linkage approach, not in a sib-pair study design, but an extended within-family design [62]. This strategy has the added benefit of identifying, as we have done here with DISC1, CNTN5, PDE4D and GRM5, a subset of variants with both statistical support and biological plausibility. Our study suggests that genetic modifiers may play a significant role in psychiatric disorders and that this is now tractable by whole-genome sequencing in families.

Acknowledgements We are indebted to members of the family, for their continuing support and participation in this long-term study. The work was supported by NIH grant MH102068. The whole-genome sequencing was performed at the CSHL Cancer Center Next-Generation Sequencing Shared Resource, supported by Cancer Center (grant number S3CA045508). Generation Scotland (GS:SFHS) received core support.
from the Chief Scientist Office (CSO) of the Scottish Government Health Directorates (grant number CZD/116/6) and the Scottish Funding Council (HR03006). Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council of UK and the Wellcome Trust (Wellcome Trust Strategic Award “STRatifying Resilience and Depression Longitudinally” (STRADL) Reference 104036/2/14Z). Ethics approval for the study was given by the NHS Tayside committee on research ethics (reference 05/S1401/89). We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. This research has also been conducted using the UKB Resource (Application 48/44). Part of the work was undertaken in The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE), part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1); funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. Age UK (The Disconnected Mind project) also provided support for the work undertaken at CCACE. AMM acknowledges support from MRC grant reference MR/J004367/1. This work was funded in part by the Eva Lester bequest to the University of Edinburgh.

Author contributions PAT, WRM and DJP conceived the work, oversaw the analyses, and wrote the manuscript. NMR, JL and MK performed analyses and wrote the manuscript. SM, SWM, AA-S, GD, EG and CH performed analyses. JD, DHRB, SML, AMM and KLE contributed scientific insight and edited the manuscript. B.D. provided data and contributed scientific insight. All authors commented on drafts of the paper.

Compliance with ethical standards

Conflict of interest WRM has participated in Illumina sponsored meetings over the past 4 years, and received travel reimbursement and an honorarium for presenting at these events. Illumina had no role in decisions relating to the study/work to be published, data collection and analysis of data and the decision to publish. WRM has participated in Pacific Biosciences sponsored meetings over the past 3 years and received travel reimbursement for presenting at these events. WRM is a founder and shareholder of Orion Genomics. WRM is a member of the scientific Advisory Board of RainDance, Inc.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

References

1. World Health Organization. Depression and other common mental disorders: global health estimates. Geneva 2017. Report no. WHO reference number: WHO/MSD/MER/2017.2.
2. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet. 2009;373:234–9.
3. Uher R. Gene-environment interactions in severe mental illness. Front Psychiatry. 2014;5:48.
4. Sullivan PF, Daly MJ, O’Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat Rev Genet. 2012;13:537–51.
5. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. 2016;48:1031–6.
6. Mulhieisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F, Treutlein J, et al. Genome-wide association study reveals two new risk loci for bipolar disorder. Nat Commun. 2014;5:3339.
7. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–7.
8. Green EK, Rees E, Walters JT, Smith KG, Forty L, Grozeva D, et al. Copy number variation in bipolar disorder. Mol Psychiatry. 2016;21:89–93.
9. Kos MZ, Carless MA, Peralta J, Blackburn A, Atmeida M, Roalf D, et al. Exome sequence data from multigenerational families implicate AMPA receptor trafficking in neurocognitive impairment and schizophrenia risk. Schizophr Bull. 2016;42:288–300.
10. Pardinas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. Nat Genet. 2018;50(3):381–9. https://doi.org/10.1038/ s41588-018-0059-2.
11. Rees E, Kendall K, Pardinas AF, Legge SE, Pocklington A, Escott-Price V, et al. Analysis of intellectual disability copy number variants for association with schizophrenia. JAMA Psychiatry. 2016;73:963–9.
12. Singh T, Walters JTR, Johnstone M, Curtis D, Suvisaari J, Tor-niainen, M et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat Genet. 2017;49(8):1167–73.
13. Costain G, Lionel AC, Fu F, Stavropoulos DJ, Gazzellone MJ, Marshall CR, et al. Adult neuropsychiatric expression and familial segregation of 2q13 duplications. Am J Med Genet B Neuropsychiatr Genet. 2014;165B:337–44.
14. Kember RL, Georgi B, Bailey-Wilson JE, Stambolian D, Paul SM, Bucan M. Copy number variants encompassing Mendelian disease genes in a large multigenerational family segregating bipolar disorder. BMC Genet. 2015;16:27.
15. Todarello G, Feng N, Kolachana BS, Li C, Vakkalanka R, Bertolino A, et al. Incomplete penetrance of NRXN1 deletions in families with schizophrenia. Schizophr Res. 2014;155B:1–7.
16. Woodbury-Smith M, Nicolson R, Zarrei M, Yuen RKC, Walker S, Howe J, et al. Variable phenotype expression in a family segregating microdeletions of the NRXN1 and MBD5 autism spectrum disorder susceptibility genes. NPJ Genom Med. 2017;2 pii:17.
17. Tansey KE, Rees E, Linden DE, Ripke S, Chambert KD, Morian JL, et al. Common allelic contribute to schizophrenia in CNV carriers. Mol Psychiatry. 2016;21:1085–9.
18. Thomson PA, Duff B, Blackwood DH, Romanuk L, Watson A, Whalley HC, et al. Balanced translocation linked to psychiatric disorder, glutamate, and cortical structure/function. NPJ Schizophren. 2016;2:16024.
19. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
20. Van der Auweera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;43:11.10.1–33.

21. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.

22. Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. Sensitive and accurate detection of copy number variants using read depth of coverage. Genome Res. 2009;19:1586–92.

23. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998;62:1198–211.

24. Nagy R, Boutin TS, Marten J, Huffman JE, Kerr SM, Campbell A, et al. Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. Genome Med. 2017;9:23.

25. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013;42:689–700.

26. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. BMC Med Genet. 2006;7:74.

27. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. PLoS ONE. 2013;8:e75362.

28. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. Global reference for human genetic variation. Nature. 2015;526:68–74.

29. He L, Carothers A, Blackwood DH, Teague P, Maclean AW, Brown J, et al. Recombination patterns around the breakpoints of a balanced 1:1 autosomal translocation associated with major mental illness. Psychiatr Genet. 1996;6:201–8.

30. Croteau-Chonka DC, Rogers AJ, Raj T, McGeachie MJ, Qiu W, Zinti JP, et al. Expression quantitative trait loci information improves predictive modeling of disease relevance of non-coding genetic variation. PLoS ONE. 2015;10:e0140758.

31. Cusanovich DA, Caliskan M, Billstrand C, Michelini K, Charvaria C, De Leon S, et al. Integrated analyses of gene expression and genetic association studies in a founder population. Hum Mol Genet. 2016;25:2104–12.

32. Fath KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518:337–43.

33. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2015;47:226–33.

34. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2015;47:226–33.

35. DNA sequence-level analyses reveal potential phenotypic modifiers in a large family with psychiatric...
regulation of the cAMP-PDE4 cascade by DISC1. Mol Psychiatry. 2013;18:898–908.

56. Oguro-Ando A, Zuko A, Kleijer KTE, Burbach JPH. A current view on contactin-4, -5, and -6: implications in neurodevelopmental disorders. Mol Cell Neurosci. 2017;81:72–83.

57. Peng YR, Tran NM, Krishnaswamy A, Kostadinov D, Martersteck EM, Sanes JR. Satb1 regulates contactin 5 to pattern dendrites of a mammalian retinal ganglion cell. Neuron. 2017;95:869–83.e6.

58. Zhong K, Karssen LC, Kayser M, Liu F. CollapsABEL: an R library for detecting compound heterozygote alleles in genome-wide association studies. BMC Bioinform. 2016;17:156.

59. Bossi S, Musante I, Bonfiglio T, Bonifacino T, Emionite L, Cerminara M, et al. Genetic inactivation of mGlu5 receptor improves motor coordination in the Grm1(crv4) mouse model of SCAR13 ataxia. Neurobiol Dis. 2018;109(Pt A):44–53.

60. Heffer A, Marquart GD, Aquilina-Beck A, Saleem N, Burgess HA, Dawid IB. Generation and characterization of Kctd15 mutations in zebrafish. PLoS ONE. 2017;12:e0189162.

61. Gabriel SB, Salomon R, Pelet A, Angrist M, Amiel J, Fornage M, et al. Segregation at three loci explains familial and population risk in Hirschsprung disease. Nat Genet. 2002;31:89–93.

62. Glahn DC, Nimgaonkar VL, Raventós H, Contreras J, McIntosh AM, Thomson PA, et al. Rediscovering the value of families for psychiatric genetics research. Mol Psychiatry (in press).