Rare variant analysis in multiply affected families, association studies and functional analysis suggest a role for the *ITGB4* gene in schizophrenia and bipolar disorder

N.L. O'Brien [a,1], A. Fiorentino [a,1,2], D. Curtis [b,c], C. Rayner [b], C. Petrosellini [a], M. Al Eissa [a], N.J. Bass [a], A. McQuillin [a,*], S.I. Sharp [a]

[a] UCL Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London, UK
[b] UCL Genetics Institute, University College London, London, UK
[c] Centre for Psychiatry, Barts and the London School of Medicine and Dentistry, London, UK

ARTICLE INFO

ABSTRACT

Recent results imply that rare variants contribute to the risk of schizophrenia. Exome sequence data from the UK10K project was used to identify three rare, amino acid changing variants in the *ITGB4* gene which segregated with schizophrenia in two families: rs750367954, rs147480547 and rs145976111. Association analysis was carried out in the exome-sequenced Swedish schizophrenia study and in UCL schizophrenia and bipolar cases and controls genotyped for these variants. A gene-wise weighted burden test was performed on a trio sample of schizophrenia cases and their parents. rs750367954 was seen in two Swedish cases and in no controls. The other two variants were commoner in cases than controls in both Swedish and UCL cohort samples and an overall burden test was significant at p = 0.0000031. The variants were not observed in the trio sample but *ITGB4* was most highly ranked out of 14,960 autosomal genes in a gene-wise weighted burden test. The effect of rs147480547 and rs145976111 was studied in human neuroblastoma SH-SY5Y cells. Cells transfected with both variants had increased proliferation at both 24 and 48 h (p = 0.013 and p = 0.05 respectively) compared to those with wild-type *ITGB4*. Taken together, these results suggest that rare variants in *ITGB4* which affect function may contribute to the aetiology of schizophrenia and bipolar disorder.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The results of exome sequence studies imply a role for rare variants in the aetiology of schizophrenia (SCZ). Most studies to date have implicated pathways and sets of genes rather than individual genes (Curtis and UK10K Consortium, 2016; Genovese et al., 2016; Purcell et al., 2014; Singh et al., 2017), however, a combined analysis of case control and trio sequence data has implicated rare loss of function variants in the *SETD1A* gene (p = 3.3×10^{-9}) in the aetiology of SCZ (Singh et al., 2016).

Analysis of sequencing data from multiply affected extended pedigrees can be valuable for identifying extremely rare variants that are transmitted to affected individuals (Curtis, 2011) and this approach has been successful in identifying variants in the *PLD3* (phospholipase

D3) gene in late-onset Alzheimer's disease (Cruchaga et al., 2014). Recently a whole genome-sequencing study identified two different rare, protein-truncating variants in the *RBM12* gene cosegregating with psychosis in two different pedigrees (Steinberg et al., 2017). Here we report the discovery and follow-up of rare, protein-changing variants which were observed to cosegregate with SCZ in subjects from two multiply-affected families which were sequenced as part of the UK10K project (Muddyman et al., 2013; UK10K Consortium et al., 2015).

2. Material and methods

2.1. Subjects

Four independent datasets were used. These comprised:

2.1.1. UK10K cohort

We used British subjects from the UK10K dataset comprising 1392 with SCZ and 982 with severe childhood obesity (SCOOP) not known to have mental illness which have been described elsewhere (Curtis and UK10K Consortium, 2016). UK10K subjects were exome sequenced with coverage of 72× as described in full elsewhere (UK10K Consortium

* Corresponding author at: University College London, Molecular Psychiatry Laboratory, Division of Psychiatry, Gower Street, London WC1E 6BT, UK.
E-mail address: a.mcquillin@ucl.ac.uk (A. McQuillin).
[1] Drs O'Brien and Fiorentino contributed equally to the work presented.
[2] Current address: UCL Institute of Ophthalmology, University College London, London, UK.

et al., 2015). Among the UK10K SCZ subjects were 48 subjects from 16 published and unpublished families multiply affected with SCZ that had been collected at UCL (Kalsi et al., 1994; Sherrington et al., 1988). Exome data was available for between 2 and 5 affected members per pedigree.

2.1.2. Swedish SCZ study

The Swedish SCZ study consisted of the 2545 controls and 2545 cases with SCZ from Sweden for whom whole exome sequence data is available *via* dbGaP (Purcell et al., 2014). Detailed sample descriptions have been provided previously (Purcell et al., 2014).

2.1.3. Bulgarian trio sample

The Bulgarian trio sample consisted of probands with SCZ and their parents (Fromer et al., 2014; Rees et al., 2015). The sample comprised whole exome sequence data from 591 trios, consisting of probands with SCZ and their parents, five of whom were also affected (Fromer et al., 2014; Rees et al., 2015). The short read files were downloaded from dbGaP along with family structure and phenotype information.

2.1.4. UCL case-control samples

The UCL case-control samples sample and the recruitment methods have been previously described (Al Eissa et al., 2017; Fiorentino et al., 2014). Briefly, the sample comprised 1917 bipolar disorder (BP) participants, 1304 SCZ participants and 1348 control participants recruited from the UK.

2.2. Variant selection and bioinformatic prediction of variant impact

Custom-written software was used to annotate and identify rare, possibly functional variants shared between affected pedigrees members in the UCL samples included in UK10K. The impact of these variants was predicted using the PolyPhen-2 (Adzhubei et al., 2013) and Sort Intolerant from Tolerant (SIFT)(Kumar et al., 2009) bioinformatics tools.

2.3. Sanger sequencing in families F047 and F158

Exome sequence data was available for three members of family F158 (subjects 3, 5 and 7) and for two members of family F047 (subjects 3 and 5; Fig. 1) from the UK10K sample. Transmission of rs750367954 (allele T) in family F158 and of rs147480547 (allele A) and rs145976111 (allele T) in family F047 was sought in all additional family members with available DNA by PCR and Sanger sequencing (Fig. 1). There was a limited amount of DNA available for individual 4 from family F047 and it was not possible to verify the genotype status of rs147480547 in this individual (Fig. 1).
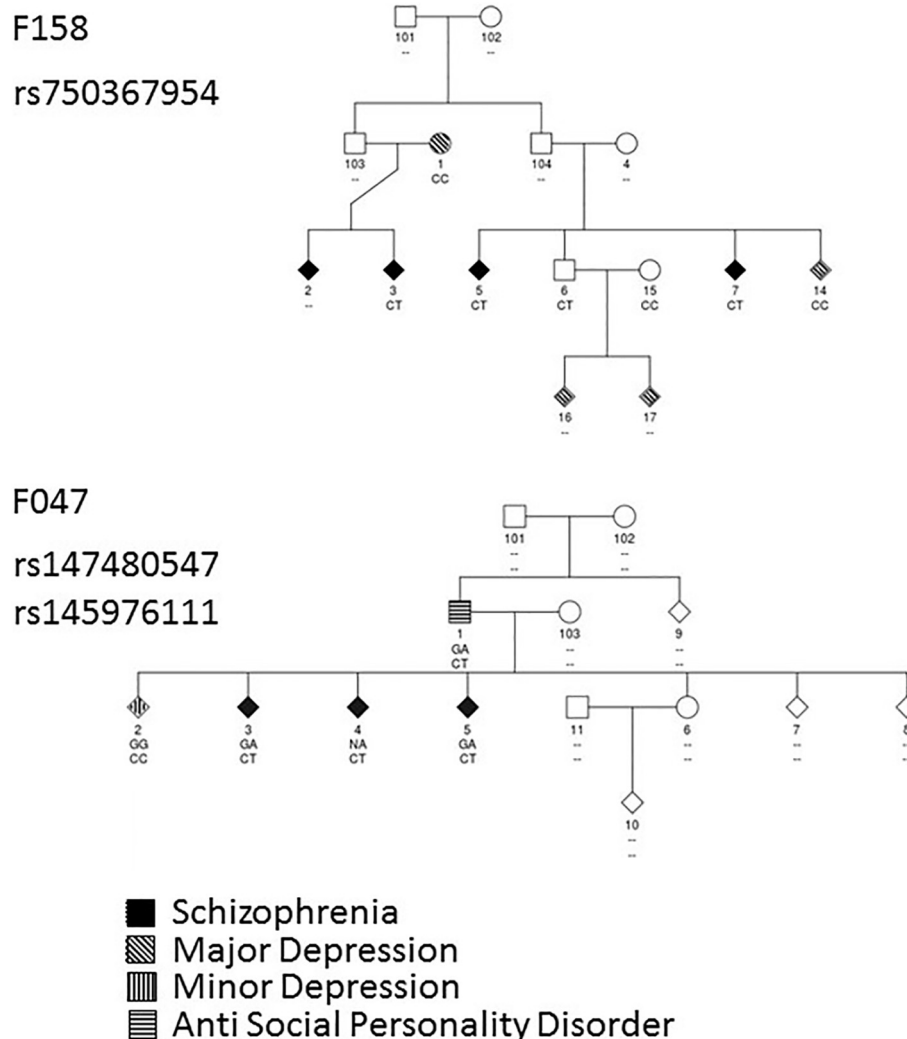

Fig. 1. Non-synonymous *ITGB4* variants in two families multiply affected by schizophrenia and other psychopathology. unk Genotype not determined; filled shapes represent individuals suffering from schizophrenia; shaded shapes represent other psychiatric diagnoses as indicated.

2.4. Variant selection and genotyping

Non-synonymous variants segregating with SCZ in families F047 and F158 were selected for genotyping in the SCZ, BP and control samples. Genotyping was performed in-house with allele-specific PCR using KASPar reagents (LGC Genomics, Hoddesdon, UK) on a LightCycler 480 (Roche, Burgess Hill, UK) real-time PCR machine. Allele-specific primers were designed for each of the SNPs using Primer Picker (LGC Genomics) as shown in Supplementary Table 1. Genotyping of each heterozygote sample was repeated at least twice and known heterozygote individuals were included on each genotyping plate to ensure reliable calling of heterozygote subjects.

2.5. Variant calling in trios

The short read files from the Bulgarian trio sample were converted to fastq files using the fastq-dump utility of the dbGaP SRA toolkit. Reads were then aligned to the hg19 human reference sequence (build GRCh37) using Novoalign V3.02.08 (NovoCraft Technologies), duplicate reads were marked using SAMBLASTER (Faust and Hall, 2014) and the BAM files were sorted using Novosort V1.03.09 (NovoCraft Technologies). Genotypes were called according to GATK best practices (Broad Institute). The HaplotypeCaller module of GATK V3.6 was used to produce gVCF files and these were then combined using the CombineGVCFs module. Initial calls were made using the GenotypeGVCFs module and then SNPs were filtered based on accuracy estimates produced by VariantRecalibrator and indels were filtered using the VariantFiltration module and the filtering expression "QD < 2.0 || FS > 50.0 || ReadPosRankSum < −20.0".

2.6. Statistical analysis

2.6.1. Case-control analysis

Allelic association analyses for the genotyped SNPs were performed using Fisher's exact test in the Swedish and UCL samples separately and in combination with each other. A burden test was performed by combining information from all genotyped SNPs to compare the total variant allele counts between cases and controls. Statistical analyses were performed using R (R Core Team, 2017).

2.6.2. Weighted burden analysis of whole exome sequence data from schizophrenia trios

The exome-sequenced trios were used to perform a gene-wise weighted burden analysis across all autosomal genes. Genotypes of autosomal variants in each gene were exported to SCOREASSOC (Curtis, 2012; Curtis and UK10K Consortium, 2016), which was used to construct a sample of pseudo-controls with genotypes consisting of the untransmitted alleles from the parents of each affected proband. A number of QC measures were applied. Variants were excluded if they did not have a PASS in the information field and individual genotype calls were excluded if they had a genotype quality score < 60. Variants were also excluded if >10% of cases or of pseudo-controls failed this quality threshold or if the heterozygote count was smaller than both homozygote counts in both cohorts. Indel calls seemed to produce artefactual results so were excluded. SCOREASSOC was used to carry out gene-wise weighted burden tests as previously described (Curtis, 2012; Curtis and UK10K Consortium, 2016) on the 591 pairs of cases and pseudo-controls to test whether there was an excess of rare, functional variants among cases. Only variants with a frequency of <0.01 in cases or pseudo-controls were considered. Functional weights were assigned so that variant types deemed more likely to have a functional effect were assigned higher weights in an approach similar to that recently suggested for dealing with *de novo* mutations (Jiang et al., 2015). Stop gain mutations were allocated a weight of 20. Non-synonymous variants were assigned a weight of 10. Stop loss and splice site variants are assigned a weight of 5. Other variants were not considered. To carry out a weighted burden analysis a score was calculated for each subject consisting of the sum of the weights of the variant alleles carried by that subject. A gene-wise test for association was then performed using a *t*-test to see if the average scores for the cases were higher than for the pseudo-controls, indicating that the cases tend to have an excess of rare, functional variants. The result of this weighted burden test was summarised as a signed log p (SLP), which is the logarithm base 10 of the two-tailed p value, given a positive sign if the mean score is higher for cases than pseudo-controls.

2.7. Cell lines

Undifferentiated SH-SY5Y human neuroblastoma cell lines were obtained from Sigma Aldrich (Gillingham, UK). The cells were cultured as mono layers in DMEM (Dulbecco's Modified Eagle Media), 10% FBS (Fetal Bovine Serum), 1% Penicillin (50 U/ml)/Streptomycin (50 μg/ml) maintained at 37 °C and 5% CO_2. Cells were not passaged beyond passage 25.

2.8. ITGB4 plasmid constructs

A TrueORF gold *ITGB4* plasmid with a c-terminal Myc-DDK tag was obtained from Origene (RC220541, Origene Technologies, Rockville, MD, USA). The *ITGB4* genetic variants rs147480547 allele A and rs145976111 allele T were introduced into the plasmid using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) and following the manufacturer's protocol. A plasmid containing both of the genetic variants was also constructed using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs, UK).

2.9. Cell transfections

SH-SY5Y cells were transfected using lipofectamine 2000 (Life technologies, UK). Growth media was removed and cells were washed with PBS. Growth media was then replaced with Opti-MEM® reduced media serum (Life technologies, UK). Cell samples were transfected with four different DNA plasmid constructs, wildtype *ITGB4*, *ITGB4* rs147480547 allele A, *ITGB4* rs145976111 allele T and a plasmid construct containing both variants in the *ITGB4* cDNA. Successful transfection was determined using quantitative RT-PCR (not shown).

2.10. Cell proliferation assay

Cells were seeded at a density of $1 \times 10^5 cells/cm^2$ and incubated for 24 and 48 h after transfection before cell proliferation was assayed. MTT (3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide) was used to measure cellular proliferation (Sigma Aldrich, Gillingham, UK)(Mosmann, 1983). MTT was dissolved in PBS at 5 mg/ml and filtered-sterilised and then added to cells at 24 h and 48 h. Cells were incubated at 37 °C for 3 h. After 3 h, the media was removed and 0.04 N HCL-isopropanol was added to solubilise the converted formazan. Absorbance of formazan was measured at a wavelength of 570 nm with a background absorbance of 630-690 nm. *t*-Tests were performed on the MTT data using R.

3. Results

3.1. Variant identification

Among the sixteen UCL pedigrees exome sequenced as part of the UK10K study, the only gene in which rare (observed at a frequency of <1% in the UK10K dataset), protein-changing variants co-segregated with SCZ in more than one pedigree was *ITGB4*. One variant, rs750367954 (C > T), had not been previously reported at the time of analysis and was found in three affected members, two siblings and one first cousin, from family F158 (Fig. 1). The variant was also found

in another sibling from family F158 who was not known to have a mental disorder but no follow up of mental health status was possible in this individual. This variant was not found in any other UK10K subjects. Two previously reported rare variants were identified in family F047: rs147480547 (G > A) which was shared by three affected siblings and rs145976111 (C > T) which was shared by two of them but the third sibling had an unknown genotype (Fig. 1). The minor allele (A) of rs147480547 was present in two additional UK10K SCZ subjects (minor allele frequency (MAF) = 0.0007) while minor allele (T) of rs145976111 was present in 7 additional UK10K SCZ subjects (MAF = 0.0025) and 3 SCOOP subjects (MAF = 0.0015).

rs750367954 C > T leads to a non-conservative amino acid change from alanine to valine at position 1689 (A1689V) of the main isoform of ITGB4 (NM_000213; Supplementary Fig. 1). The A1689V amino acid substitution was predicted to be "Benign" with a score between 0 and 0.011 depending on the isoform by PolyPhen-2 (Supplementary Table 2) and "Deleterious" with a score of 0.01 by SIFT (Table 1). A1689V is present in an experimentally demonstrated protein interaction domain for ERBIN (Favre et al., 2001) and in Fibronectin type III and Immunoglobulin-like fold domains as predicted by INTERPRO (https://www.ebi.ac.uk/interpro/protein/P16144; Supplementary Fig. 1). rs147480547 G > A causes a non-conservative amino acid change from alanine to threonine at position 808 (A808T) of NM_000213. The A808T amino acid substitution was predicted to be "Benign" with a score between 0.006 and 0.335 depending on the isoform by PolyPhen-2 and "Deleterious" with a score of 0.01 by SIFT. rs145976111 C > T causes a non-conservative amino acid change from arginine to cysteine at position 977 (R977C) of NM_000213 (Supplementary Fig. 1). The R977C amino acid substitution was predicted to be "Probably Damaging" with a score between 0.95 and 0.999 depending on the isoform by PolyPhen-2 (Supplementary Table 2) and "Deleterious" with a score of 0 by SIFT.

3.2. Association analysis

Genotype counts for the variants in the Swedish SCZ cases and controls and the UCL samples are shown in Table 2. Results for the UCL SCZ and BP cases are shown separately and combined. BP subjects were included in the genotyping panel to allow assessment of the frequency of variants across the psychosis spectrum. rs750367954 C > T was detected in two Swedish cases but no other subjects and the other variants were commoner in cases than controls in both the Swedish and UCL cohorts. The data shows evidence for association with schizophrenia and bipolar disorder separately (p = 0.0022 and p = 0.0011 respectively for rs145976111) and for both disorders when the data was combined (p = 0.0004 for rs145976111). Table 2 also shows the results obtained from comparing combined genotype counts for the Swedish and UCL cases with other subjects, both for individual variants and for a burden test combining counts across all variants. The burden test was significant at p = 0.0000031.

3.3. Weighted burden analysis of trios

In the Bulgarian trio dataset no variants were observed at rs147480547 G > A, rs145976111 C > T or rs750367954 C > T. However, a gene-wise weighted burden test was carried out and provided informative results for 14,960 autosomal genes (Supplementary Table 3). A QQ plot showed that the SLPs obtained closely conformed to the null hypothesis distribution and no gene was exome-wide significant (Supplementary Fig. 2). However, out of all these genes ITGB4 was most highly ranked with an SLP of 4.3, corresponding to a p value of 0.00005. Table 3 shows the genotype counts for ITGB4 in cases and pseudo-controls. It can be seen that the result is driven by an excess of variants among cases at a variety of different positions (Supplementary Fig. 1).

3.4. The impact of ITGB4 coding variants rs147480547 and rs145976111 on SH-SY5Y cell proliferation

Quantitative PCR analysis confirmed overexpression of ITGB4 in SH-SY5Y cells transfected with the ITGB4 plasmid constructs compared to SH-SY5Y cells transfected with an empty vector.

No gross morphological changes were observed in the cells transfected with the wildtype or variant ITGB4 plasmid constructs. Next we tested whether the variants would have an impact on SH-SY5Y cell proliferation on the basis that Integrin β4 is thought to play a role in neuronal survival and apoptosis signal transduction pathways (Lv et al., 2008; Lv et al., 2006; Su et al., 2007). Cells transfected with ITGB4 constructs containing variant alleles for both rs147480547 and rs145976111 had increased MTT-formazan formation at both 24 and 48 h (p = 0.013 and p = 0.05 respectively) (Fig. 2). Transfection with ITGB4 constructs expressing these variants separately produced intermediate values.

4. Discussion

One approach to detect rare variants is to utilise segregation information from multiple affected individuals within a family (Cruchaga et al., 2014; Curtis, 2011; Timms et al., 2013). Here we describe how exome sequencing data from related affected subjects was used to initially identify candidate variants in ITGB4. Follow up studies in other samples provide additional support for the involvement of this gene in the aetiology of SCZ and bipolar disorder. Further support comes from data from 872 subjects from the Western Australian Family Study of Schizophrenia and 36,355 controls who reported two SCZ case only ITGB4 variants (McCarthy et al., 2017). However, it should be noted that for rs147480547 and rs145976111 the MAF (minor allele frequency) observed in both sets of controls is lower than in the Non-Finnish European samples in the ExAC database of aggregated exome sequencing data from large-scale sequencing projects (Lek et al., 2016). Indeed, the ExAC allele frequencies are similar to what we observe in the cases even in the subset of samples that do not include individuals with known psychiatric diagnoses. Thus, the findings for these variants appear to be driven by the fact that they are unexpectedly uncommon in the control samples. By contrast, rs750367954 is very rare and occurs at a frequency of 0.00012 in the non-Finnish European samples in the current version (0.3.1) of ExAC and at a frequency of 0.000025 in the non-psychiatric subjects.

ITGB4 encodes the integrin beta 4 subunit. Integrins are a large family of heterodimeric receptors for laminin, an extracellular matrix protein which plays a fundamental role in cellular interactions, motility and signalling during development (Tarone et al., 2000). Integrins

Table 1
Prediction of variant impact on protein structure from Poly-Phen2 and SIFT for isoform 1 of ITGB4 (NCBI reference sequence, NM_000213).

Variant ID [nucleotide change]	Amino acid change	Poly-Phen2	SIFT	Likely function & domain
rs147480547 [G/A]	Ala/Thr (A808T)	Benign (0.006–0.335)	Deleterious (0.01)	Unknown/Undefined
rs145976111 [C/T]	Arg/Cys (R977C)	Probably damaging (0.95–0.99)	Deleterious (0)	CALX-BETA Domain
rs750367954 [C/T]	Ala/Val (A1689V)	Benign (0–0.011)	Deleterious (0.014)	Protein:protein interaction with ERBIN; fibronectin type-III & immunoglobulin-like fold domains

Table 2

Genotypes of *ITGB4* variants in the cases and controls from the Swedish schizophrenia cohort (SE); the UCL cohorts (UK); the two cohorts combined; and for a burden analysis using genotypes combined across all three variants. Significance values were obtained using Fisher's exact test. Variant positions are shown according to hg19 (GRCh37).

Variant ID (Position on Chr17) [Base pair change]	Sample	Genotype counts AA/AB/BB	MAF	P	Odds ratio (95% CI)
rs750367954	SE SCZ	2534/2/0	0.00039	0.25	Inf (0.19-Inf)
(73753036)	SE control	2540/0/0	0		
[C/T]	UK SCZ	1275/0/0	0	NA	NA
	UK BP	1893/0/0	0	NA	NA
	UK SCZ&BP	3168/0/0	0	NA	NA
	UK control	1324/0/0	0		0
	Combined SCZ	3802/2/0	0.0002	0.25	Inf (0.19-Inf)
	Combined BP	1892/0/0	0	NA	NA
	Combined BP&SCZ	5694/2/0	0.00018	0.52	Inf (0.13-Inf)
	Combined control	3844/0/0	0		
rs147480547	SE SCZ	2532/4/0	0.00079	0.062	Inf (0.66-Inf)
(73736128)	SE control	2540/0/0	0		
[G/A]	UK SCZ	1257/4/0	0.0016	0.21	4.16 (0.41–204.92)
	UK BP	1885/6/0	0.0016	0.25	4.16 (0.50–191.48)
	UK BP&SCZ	3142/10/0	0.0016	0.19	4.16 (0.59–180.66)
	UK control	1310/1/0	0.0004		
	Combined SCZ	3789/8/0	0.0011	0.021	8.12 (1.09–359.83)
	Combined BP	1885/6/0	0.0016	0.0064	12.24 (1.48–561.34)
	Combined BP&SCZ	5674/14/0	0.0012	0.0069	9.49 (1.44–400.60)
	Combined control	3850/1/0	0.00013		
rs145976111	SE SCZ	2522/12/0	0.0024	0.021	4.02 (1.08–22.19)
(73738809)	SE control	2537/3/0	0.00059		
[C/T]	UK SCZ	1258/10/0	0.0039	0.054	3.45 (0.89–19.53)
	UK BP	1876/14/0	0.0037	0.081	3.24 (0.90–17.60)
	UK SCZ&BP	3134/24/0	0.0038	0.035	3.32 (1.01–17.26)
	UK control	1306/3/0	0.0011		
	Combined SCZ	3780/22/0	0.0029	0.0022	3.72 (1.46–11.22)
	Combined BP	1876/14/0	0.0037	0.0011	4.76 (1.72–15.14)
	Combined BP&SCZ	5656/36/0	0.0032	0.0004	4.07 (1.70–11.81)
	Combined control	3843/6/0	0.00078		
Burden analysis	SCZ	11,371/32/0	0.0014	0.000038	
	BP	5653/20/0	0.0018	0.000015	
	BP&SCZ	17,024/52/0	0.0015	0.0000031	
	Control	11,537/7/0	0.00078		

consist of α and β subunits. The main isoform of the β4 subunit consists of 1882 amino acids. The first 27 amino acids form a signal peptide, the next 683 amino acids form an extracellular domain that is followed by a short transmembrane domain. The C-terminus of the protein is an exceptionally long cytoplasmic domain of 1089 residues. The length of the cytoplasmic domain suggests the importance of the interaction of the β subunit with cytoplasmic proteins (Hogervorst et al., 1990;

Suzuki and Naitoh, 1990; Tamura et al., 1990). There is experimental evidence for interaction between the cytoplasmic domain of *ITGB4* with *COL17A1*, *ERBIN*, *BP230*, *BP180*, *DST* and *RAC1* (Favre et al., 2001; Hamill et al., 2009; Hopkinson and Jones, 2000; Koster et al., 2003). Integrin β4 is found primarily in epithelial cells and binds with integrin alpha (α) 6 (*ITGA6*). It is of interest that the *ITGB4*/*ITGA6* complex binds to the EGF-like domain in Neuregulin-1 (*NRG1*) (Ieguchi et al., 2010)

Table 3

Genotype counts for rare variants in *ITGB4* in trio cases and pseudo-controls consisting of non-transmitted parental alleles. Using the weighted burden test implemented in SCOREASSOC, the excess in cases is significant at p = 0.00005. Amino acid changes are shown for NM_000213. NS Nonsynonymous.

| Position on chr17 (hg19)	Variant identifier, base pair and amino acid change	Type	Pseudo-controls			Schizophrenia cases		
			AA	AB	BB	AA	AB	BB
73724461	rs775438066 G/A; R158Q	NS	589	0	0	586	3	0
73724543	17:73724543 G/A; M185I	NS	576	0	0	575	1	0
73726959	17:73726959 C/T; L336F	NS	585	0	0	584	1	0
73726989	rs778405226 G/C; G346R	NS	583	0	0	581	2	0
73727379	rs762334700 G/A; R382Q	NS	534	0	0	533	1	0
73728226	rs373717293 T/C;	Splice site	587	0	0	586	1	0
73728300	rs8079267 G/T; Q478H	NS	586	2	0	581	7	0
73729749	rs772615108 C/T; R545C	NS	546	1	0	547	0	0
73736505	rs760130077 G/A; R838Q	NS	542	0	0	541	1	0
73738731	rs764035400 C/T; R951W	NS	533	0	0	531	2	0
73746859	17:73746859 C/G; D1191E	NS	576	1	0	577	0	0
73746884	rs75129664 G/A; G1200R	NS	567	4	0	560	11	0
73750828	17:73750828 G/A; R1497H	NS	583	0	0	582	1	0
73752635	rs538467153 C/T; R1612C	NS	576	0	0	575	1	0
73753143	rs762581242 G/A; G1725R	NS	535	0	0	534	1	0
73753152	rs759302516 C/T; R1728C	NS	543	0	0	542	1	0
73753278	rs200010813 C/T;	Splice site	532	3	0	531	4	0
73753310	rs151327791 G/A; G1750R	NS	540	0	0	539	1	0
73753326	rs140577812 C/T; P1755L	NS	543	1	0	543	1	0
73753574	rs773923807 C/T; R1803W	NS	557	0	0	556	1	0
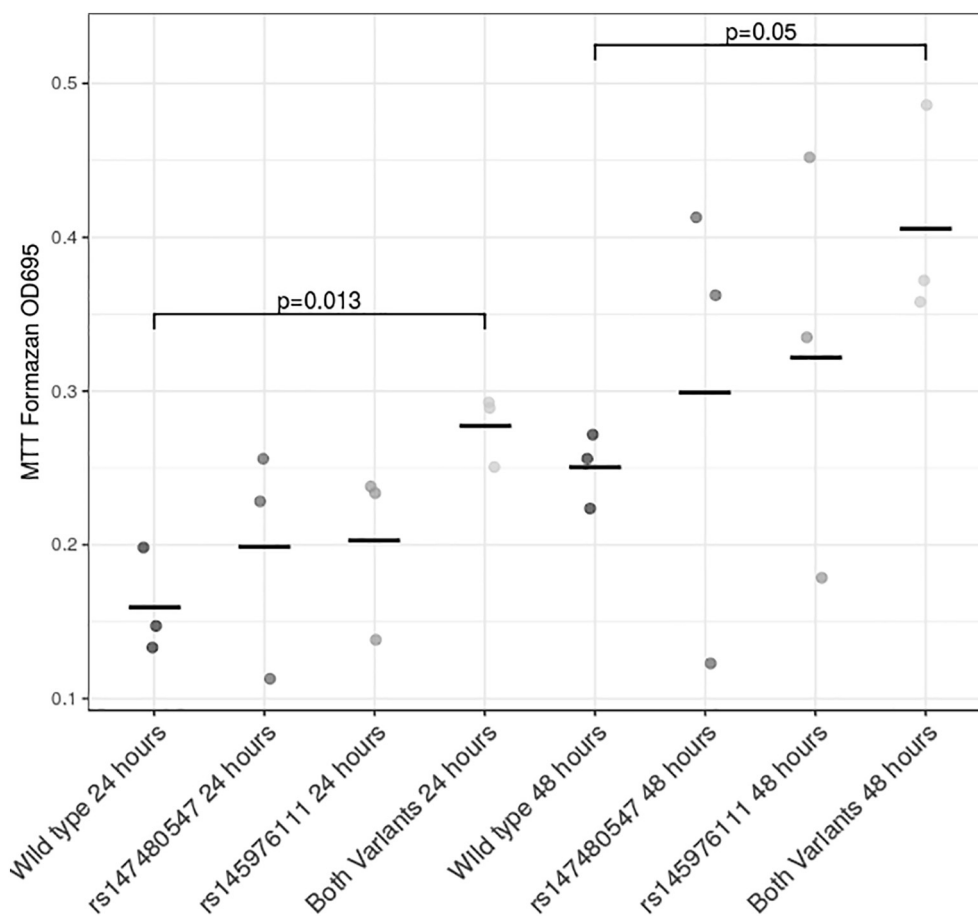
Fig. 2. MTT-formazan formation from SH-SY5Y cells overexpressing the wildtype, rs147480547, rs145976111 and both variants of the *ITGB4* gene. The data was obtained from three independent experiments.

and variants in the *NRG1* gene have been previously associated with SCZ (Craddock et al., 2005; Mei and Xiong, 2008). All three *ITGB4* variants identified here are located in the long cytoplasmic domain, making them likely to have a functional impact on the interaction of integrin β4 with other cytoplasmic proteins.

Integrin β4 is involved in cell-cell and cell-matrix adhesion. In addition to epithelial cells it is expressed in astrocytes, Schwann cells and neurons (Jaakkola et al., 1993; Milner and Campbell, 2006; Su et al., 2007) and it is abundantly expressed in the developing cerebral cortex (Lv et al., 2008). Integrin β4 is important for the interaction of Schwann cells with the basal lamina and axons, and it could influence the structure of myelinating Schwann cells (Su et al., 2008). Integrin β4 is also believed to be a key factor in neuronal survival and apoptosis signal transduction pathways (Lv et al., 2008; Lv et al., 2006; Su et al., 2007).

We also present *in vitro* evidence from our functional studies to suggest that haplotypes comprising the minor alleles of rs147480547 and rs145976111 lead to an increase in cell proliferation with increased cellular metabolic activity in human neuroblastoma cells transfected with both variant alleles together. Integrin β4 is expressed in human neuronal stem/precursor cells (Flanagan et al., 2006) and downregulation inhibits mouse neural stem cell differentiation (Su et al., 2008). It is well established that the integrin β4 and α6 complex induces c-Src signalling and mTOR activation which is important for stimulating transcription and translation of cancer related genes (Muranen et al., 2017; Soung et al., 2013). Interestingly, the antipsychotic penfluridol has been shown to significantly reduce tumor growth and induce apoptosis in cancer cell lines through inhibition of integrin signalling. This effect was mediated by a decrease in the expression of *ITGB4* (Ranjan et al., 2016). Our functional studies demonstrate that *ITGB4* SNP variants rs147480547 and rs145976111 further increase cytoplasmic activity

and cell growth in comparison with wild type *ITGB4*. It is thus likely that expression of these variants in *ITGB4* increase cell proliferation by facilitating cell-cell interactions through upregulation of cell signalling. An important limitation of this study is that our functional assays relied on overexpression of ITGB4 in a proliferative cell line. The use of induced pluripotent stem cells derived from patients carrying these variants would likely provide a clearer picture of the functional effects of these variants.

Here we report co-segregation of three rare, protein-changing variants in the *ITGB4* gene with SCZ in two families, with one family presenting a haplotype consisting of two of these variants. We provide further evidence from case control cohorts of subjects that had been exome sequenced or directly genotyped that variants in these families may contribute to susceptibility to SCZ and/or BP in the wider population. We also find support for a role of the *ITGB4* gene in susceptibility to SCZ through a weighted burden test in exome data from a SCZ trio sample and from the literature. Finally, we demonstrate that when these variants are present on the same haplotype there is an increase in cell proliferation *in vitro*. Whilst none of the results obtained provides unequivocal evidence for the role of *ITGB4* in susceptibility to psychotic illness, considered together they are of interest and justify further investigation this gene and its variants in schizophrenia.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.schres.2018.03.001.

References

Adzhubei, I., Jordan, D.M., Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. 76 (7.20), 7.20.1–7.20.41.

Al Eissa, M.M., Fiorentino, A., Sharp, S.I., O'Brien, N.L., Wolfe, K., Giaroli, G., Curtis, D., Bass, N.J., McQuillin, A., 2017. Exome sequence analysis and follow up genotyping implicates rare ULK1 variants to be involved in susceptibility to schizophrenia. Ann. Hum. Genet. 82 (2), 88–92.

Craddock, N., O'Donovan, M.C., Owen, M.J., 2005. The genetics of schizophrenia and bipolar disorder: dissecting psychosis. J. Med. Genet. 42 (3), 193–204.

Cruchaga, C., Karch, C.M., Jin, S.C., Benitez, B.A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., Jeng, A.T., Cooper, B., Skorupa, T., Carrell, D., Levitch, D., Hsu, S., Choi, J., Ryten, M., Sassi, C., Bras, J., Gibbs, R.J., Hernandez, D.G., Lupton, M.K., Powell, J., Forabosco, P., Ridge, P.G., Corcoran, C.D., Tschanz, J.T., Norton, M.C., Munger, R.G., Schmutz, C., Leary, M., Demirci, F.Y., Bamne, M.N., Wang, X., Lopez, O.L., Ganguli, M., Medway, C., Turton, J., Lord, J., Braae, A., Barber, I., Brown, K., Alzheimer's Research, U.K.C., Pastor, P., Lorenzo-Betancor, O., Brkanac, Z., Scott, E., Topol, E., Morgan, K., Rogaeva, E., Singleton, A., Hardy, J., Kamboh, M.I., George-Hyslop, P.S., Cairns, N., Morris, J.C., Kauwe, J.S.K., Goate, A.M., 2014. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 505 (7484), 550–554.

Curtis, D., 2011. Assessing the contribution family data can make to case-control studies of rare variants. Ann. Hum. Genet. 75 (5), 630–638.

Curtis, D., 2012. A rapid method for combined analysis of common and rare variants at the level of a region, gene, or pathway. Adv. Appl. Bioinforma. Chem. 5, 1–9.

Curtis, D., UK10K Consortium, 2016. Practical experience of the application of a weighted burden test to whole exome sequence data for obesity and schizophrenia. Ann. Hum. Genet. 80 (1), 38–49.

Faust, G.G., Hall, I.M., 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics 30 (17), 2503–2505.

Favre, B., Fontao, L., Koster, J., Shafaatian, R., Jaunin, F., Saurat, J.H., Sonnenberg, A., Borradori, L., 2001. The hemidesmosomal protein bullous pemphigoid antigen 1 and the integrin beta 4 subunit bind to ERBIN. Molecular cloning of multiple alternative splice variants of ERBIN and analysis of their tissue expression. J. Biol. Chem. 276 (35), 32427–32436.

Fiorentino, A., O'Brien, N.L., Locke, D.P., McQuillin, A., Jarram, A., Anjorin, A., Kandaswamy, R., Curtis, D., Blizard, R.A., Gurling, H.M., 2014. Analysis of ANK3 and CACNA1C variants identified in bipolar disorder whole genome sequence data. Bipolar Disord. 16 (6), 583–591.

Flanagan, L.A., Rebaza, L.M., Derzic, S., Schwartz, P.H., Monuki, E.S., 2006. Regulation of human neural precursor cells by laminin and integrins. J. Neurosci. Res. 83 (5), 845–856.

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), 179–184.

Genovese, G., Fromer, M., Stahl, E.A., Ruderfer, D.M., Chambert, K., Landén, 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.

Hamill, K.J., Hopkinson, S.B., DeBiase, P., Jones, J.C., 2009. BPAG1e maintains keratinocyte polarity through beta4 integrin-mediated modulation of Rac1 and cofilin activities. Mol. Biol. Cell 20 (12), 2954–2962.

Hogervorst, F., Kuikman, I., von dem Borne, A.E., Sonnenberg, A., 1990. Cloning and sequence analysis of beta-4 cDNA: an integrin subunit that contains a unique 118 kd cytoplasmic domain. EMBO J. 9 (3), 765–770.

Hopkinson, S.B., Jones, J.C., 2000. The N terminus of the transmembrane protein BP180 interacts with the N-terminal domain of BP230, thereby mediating keratin cytoskeleton anchorage to the cell surface at the site of the hemidesmosome. Mol. Biol. Cell 11 (1), 277–286.

Ieguchi, K., Fujita, M., Ma, Z., Davari, P., Taniguchi, Y., Sekiguchi, K., Wang, B., Takada, Y.K., Takada, Y., 2010. Direct binding of the EGF-like domain of neuregulin-1 to integrins ({alpha}v{beta}3 and {alpha}6{beta}4) is involved in neuregulin-1/ErbB signaling. J. Biol. Chem. 285 (41), 31388–31398.

Jaakkola, S., Savunen, O., Halme, T., Uitto, J., Peltonen, J., 1993. Basement membranes during development of human nerve: Schwann cells and perineurial cells display marked changes in their expression profiles for laminin subunits and beta 1 and beta 4 integrins. J. Neurocytol. 22 (3), 215–230.

Jiang, Y., Han, Y., Petrovski, S., Owzar, K., Goldstein, D.B., Allen, A.S., 2015. Incorporating functional information in tests of excess de novo mutational load. Am. J. Hum. Genet. 97 (2), 272–283.

Kalsi, G., Mankoo, B.S., Brynjolfsson, J., Curtis, D., Read, T., Murphy, P., Sharma, T., Petursson, H., Gurling, H.M., 1994. The Marfan syndrome gene locus as a favoured locus for susceptibility to schizophrenia. Psychiatr. Genet. 4 (4), 219–227.

Koster, J., Geerts, D., Favre, B., Borradori, L., Sonnenberg, A., 2003. Analysis of the interactions between BP180, BP230, plectin and the integrin alpha6beta4 important for hemidesmosome assembly. J. Cell Sci. 116 (Pt 2), 387–399.

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4 (7), 1073–1081.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., Exome Aggregation, C., 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536 (7616), 285–291.

Lv, X., Wang, N., Su, L., Zhang, S., Miao, J., 2006. Inhibition of PC-PLC blocked the survival of mouse neural cells by up-regulating the expression of integrin beta4 and Rb. Dev. Neurosci. 28 (6), 499–504.

Lv, X., Su, L., Yin, D., Sun, C., Zhao, J., Zhang, S., Miao, J., 2008. Knockdown of integrin beta4 in primary cultured mouse neurons blocks survival and induces apoptosis by elevating NADPH oxidase activity and reactive oxygen species level. Int. J. Biochem. Cell Biol. 40 (4), 689–699.

McCarthy, N.S., Melton, P.E., Ward, S.V., Allan, S.M., Dragovic, M., Clark, M.L., Morar, B., Rubio, J.P., Blangero, J., Badcock, J.C., Morgan, V.A., Moses, E.K., Jablensky, A., 2017. Exome array analysis suggests an increased variant burden in families with schizophrenia. Schizophr. Res. 185, 9–16.

Mei, L., Xiong, W.C., 2008. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. Nat. Rev. Neurosci. 9 (6), 437–452.

Milner, R., Campbell, I.L., 2006. Increased expression of the beta4 and alpha5 integrin subunits in cerebral blood vessels of transgenic mice chronically producing the pro-inflammatory cytokines IL-6 or IFN-alpha in the central nervous system. Mol. Cell. Neurosci. 33 (4), 429–440.

Mosmann, T., 1983. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. J. Immunol. Methods 65 (1–2), 55–63.

Muddyman, D., Smee, C., Griffin, H., Kaye, J., 2013. Implementing a successful data-management framework: the UK10K managed access model. Genome Med. 5 (11), 100.

Muranen, T., Iwanicki, M.P., Curry, N.L., Hwang, J., DuBois, C.D., Coloff, J.L., Hitchcock, D.S., Clish, C.B., Brugge, J.S., Kalaany, N.Y., 2017. Starved epithelial cells uptake extracellular matrix for survival. Nat. Commun. 8, 13989.

Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kahler, A., Duncan, L., Stahl, E., Genovese, G., Fernandez, E., Collins, M.O., Komiyama, N.H., Choudhary, J.S., Magnusson, P.K., Banks, E., Shakir, K., Garimella, K., Fennell, T., DePristo, M., Grant, S.G., Haggarty, S.J., Gabriel, S., Scolnick, E.M., Lander, E.S., Hultman, C.M., Sullivan, P.F., McCarroll, S.A., Sklar, P., 2014. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506 (7487), 185–190.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ranjan, A., Gupta, P., Srivastava, S.K., 2016. Penfluridol: an antipsychotic agent suppresses metastatic tumor growth in triple-negative breast cancer by inhibiting integrin signaling axis. Cancer Res. 76 (4), 877–890.

Rees, E., Kirov, G., Walters, J.T., Richards, A.L., Howrigan, D., Kavanagh, D.H., Pocklington, A.J., Fromer, M., Ruderfer, D.M., Georgieva, L., Carrera, N., Gormley, P., Palta, P., Williams, H., Dwyer, S., Johnson, J.S., Roussos, P., Barker, D.D., Banks, E., Milanova, V., Rose, S.A., Chambert, K., Mahajan, M., Scolnick, E.M., Moran, J.L., Tsuang, M.T., Glatt, S.J., Chen, W.J., Hwu, H.G., Taiwanese Trios Exome Sequencing, C., Neale, B.M., Palotie, A., Sklar, P., Purcell, S.M., McCarroll, S.A., Holmans, P., Owen, M.J., O'Donovan, M.C., 2015. Analysis of exome sequence in 604 trios for recessive genotypes in schizophrenia. Transl. Psychiatry 5, e607.

Sherrington, R., Brynjolfsson, J., Petursson, H., Potter, M., Dudleston, K., Barraclough, B., Wasmuth, J., Dobbs, M., Gurling, H., 1988. Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature 336 (6195), 164–167.

Singh, T., Kurki, M.I., Curtis, D., Purcell, S.M., Crooks, L., McRae, J., Suvisaari, J., Chheda, H., Blackwood, D., Breen, G., Pietilainen, O., Gerety, S.S., Ayub, M., Blyth, M., Cole, T., Collier, D., Coomber, E.L., Craddock, N., Daly, M.J., Danesh, J., DiForti, M., Foster, A., Freimer, N.B., Geschwind, D., Johnstone, M., Joss, S., Kirov, G., Korkko, J., Kuismin, O., Holmans, P., Hultman, C.M., Iyegbe, C., Lonnqvist, J., Mannikko, M., McCarroll, S.A., McGuffin, P., McIntosh, A.M., McQuillin, A., Moilanen, J.S., Moore, C., Murray, R.M., Newbury-Ecob, R., Ouwehand, W., Paunio, T., Prigmore, E., Rees, E., Roberts, D., Sambrook, J., Sklar, P., Clair, D.S., Veijola, J., Walters, J.T., Williams, H., Swedish Schizophrenia, S., Study, I., Study, D.D.D., Consortium, U.K., Sullivan, P.F., Hurles, M.E., O'Donovan, M.C., Palotie, A., Owen, M.J., Barrett, J.C., 2016. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat. Neurosci. 19 (4), 571–577.

Singh, T., Walters, J.T.R., Johnstone, M., Curtis, D., Suvisaari, J., Torniainen, M., Rees, E., Iyegbe, C., Blackwood, D., McIntosh, A.M., Kirov, G., Geschwind, D., Murray, R.M., Di Forti, M., Bramon, E., Gandal, M., Hultman, C.M., Sklar, P., Study, I., Consortium, U.K., Palotie, A., Sullivan, P.F., O'Donovan, M.C., Owen, M.J., Barrett, J.C., 2017. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat. Genet. 49 (8), 1167–1173.

Soung, Y.H., Korneeva, N., Kim, T.H., Chung, J., 2013. The role of c-Src in integrin (alpha6beta4) dependent translational control. BMC Cell Biol. 14, 49.

Steinberg, S., Gudmundsdottir, S., Sveinbjornsson, G., Suvisaari, J., Paunio, T., Torniainen-Holm, M., Frigge, M.L., Jonsdottir, G.A., Huttenlocher, J., Arnarsdottir, S., Ingimarsson, O., Haraldsson, M., Tyrfingsson, T., Thorgeirsson, T.E., Kong, A., Norddahl, G.L., Gudbjartsson, D.F., Sigurdsson, E., Stefansson, H., Stefansson, K., 2017. Truncating mutations in RBM12 are associated with psychosis. Nat. Genet. 49 (8), 1251–1254.

Su, L., Zhao, B., Lv, X., Wang, N., Zhao, J., Zhang, S., Miao, J., 2007. Safrole oxide induces neuronal apoptosis through inhibition of integrin beta4/SOD activity and elevation of ROS/NADPH oxidase activity. Life Sci. 80 (11), 999–1006.

Su, L., Lv, X., Miao, J., 2008. Integrin beta 4 in neural cells. NeuroMolecular Med. 10 (4), 316–321.

Suzuki, S., Naitoh, Y., 1990. Amino acid sequence of a novel integrin beta 4 subunit and primary expression of the mRNA in epithelial cells. EMBO J. 9 (3), 757–763.

Tamura, R.N., Rozzo, C., Starr, L., Chambers, J., Reichardt, L.F., Cooper, H.M., Quaranta, V., 1990. Epithelial integrin alpha 6 beta 4: complete primary structure of alpha 6 and variant forms of beta 4. J. Cell Biol. 111 (4), 1593–1604.

Tarone, G., Hirsch, E., Brancaccio, M., De Acetis, M., Barberis, L., Balzac, F., Retta, S.F., Botta, C., Altruda, F., Silengo, L., 2000. Integrin function and regulation in development. Int. J. Dev. Biol. 44 (6), 725–731.

Timms, A.E., Dorschner, M.O., Wechsler, J., Choi, K.Y., Kirkwood, R., Girirajan, S., Baker, C., Eichler, E.E., Korvatska, O., Roche, K.W., Horwitz, M.S., Tsuang, D.W., 2013. Support for the N-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families. JAMA Psychiat. 70 (6), 582–590.

UK10K Consortium, Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R., Xu, C., Futema, M., Lawson, D., Iotchkova, V., Schiffels, S., Hendricks, A.E., Danecek, P., Li, R., Floyd, J., Wain, L.V., Barroso, I., Humphries, S.E., Hurles, M.E., Zeggini, E., Barrett, J.C., Plagnol, V., Richards, J.B., Greenwood, C.M., Timpson, N.J., Durbin, R., Soranzo, N., 2015. The UK10K project identifies rare variants in health and disease. Nature 526 (7571), 82–90.