An inherited duplication at the gene p21 Protein-Activated Kinase 7 (PAK7) is a risk factor for psychosis

Title	An inherited duplication at the gene p21 Protein-Activated Kinase 7 (PAK7) is a risk factor for psychosis
Author(s)	Morris, Derek W.
Publication Date	2014-01-28
Publisher	Oxford University Press (OUP)

An inherited duplication at the gene *p21 Protein-Activated Kinase 7 (PAK7)* is a risk factor for psychosis

Derek W Morris PhD[1*], Richard D Pearson PhD[2*], Paul Cormican PhD[1*], Elaine M Kenny PhD[1], Colm T O'Dushlaine PhD[3], Louis-Philippe Lemieux Perreault PhD[2,4], Eleni Giannoulatou PhD[2], Daniela Tropea PhD[1], Brion S Maher PhD[5], Brandon Wormley PhD[5], Eric Kelleher MD[1], Ciara Fahey BSc[1], Ines Molinos PhD[1], Stefania Bellini BSc[1], Matti Pirinen PhD[2], Amy Strange PhD[2], Colin Freeman PhD[2], Dawn L Thiselton PhD[5], Rachel L Elves PhD[5], Regina Regan PhD[6], Sean Ennis PhD[6], Timothy G Dinan MD PhD[7], Colm McDonald MD PhD[8], Kieran C Murphy MD PhD[9], Eadbhard O'Callaghan MD PhD[10]†, John L Waddington PhD[11], Dermot Walsh MD[12], Michael O'Donovan MD PhD[13], DetelinaGrozeva MD[13], Nick Craddock MD PhD[13], Jennifer Stone PhD[3], Ed Scolnick PhD[3], Shaun Purcell PhD[3, 14], Pamela Sklar MD PhD[3, 14], Bradley Coe PhD[15], Evan E Eichler PhD[15], RoelOphoff MD PhD[16], JacobineBuizer PhD[16], Jin Szatkiewicz PhD[17], Christina Hultman PhD[17], Patrick Sullivan MD FRANZCP[17], Hugh Gurling MD PhD[18], Andrew McQuillin MD PhD[18], David St Clair MD PhD[19], Elliott Rees[13], George Kirov MD PhD[13], James Walters MD PhD[13], Douglas Blackwood MD PhD[20], Mandy Johnstone MD PhD[20], Gary Donohoe PhD[1], International Schizophrenia Consortium, SGENE+ Consortium, Francis A O'Neill MD PhD[21], Wellcome Trust Case Control Consortium 2[23],Kenneth S Kendler MD PhD[5], Michael Gill MD[1], Brien P Riley PhD[5], Chris C A Spencer PhD[2‡], Aiden Corvin MD PhD[1‡].

1. Dept of Psychiatry & Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Ireland.
2. Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.
3. Broad Institute and Center for Human Genetics Research of Massachusetts General Hospital, Boston, Massachussets, US.
4. Montreal Heart Institute, Université de Montréal, Montréal, Québec, Canada
5. Depts of Psychiatry and Human Genetics, Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond VA, US
6. School of Medicine and Medical Science, University College Dublin, Ireland
7. Dept of Psychiatry, University College Cork, Cork, Ireland
8. Dept of Psychiatry, National University of Ireland, Galway, University Road, Galway, Ireland
9. Dept of Psychiatry, RCSI Education and Research Centre, Beaumont Hospital, Dublin 9, Ireland
10. DETECT Early Intervention in Psychosis Services, Dun Laoghaire, Co. Dublin, Ireland
11. Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland

12. Health Research Board, 73 Lower Baggot St, Dublin 2, Ireland.
13. MRC Centre for Neuropsychiatric Genetics and Genomics, and Neuroscience and Mental Health Research Institute, Cardiff University, Heath Park, Cardiff CF4 4XN, UK
14. The Mount Sinai Hospital, New York, NY 10029, US
15. University of Washington School of Medicine, Howard Hughes Medical Institute, Seattle, US.
16. UCLA School of Medicine, Los Angeles, CA 900966, US.
17. University of North Carolina, Chapel Hill, NC 27599-7264, US.
18. Rudolf Magnus Institute, University of Utrecht, 3584 CG Utrecht, Netherlands.
19. Molecular Psychiatry Laboratory, Mental Health Sciences Unit, University College London, London, UK
20. Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen, UK
21. Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, UK
22. Dept of Psychiatry, Queen's University, Belfast, BT7 1NN, Northern Ireland
23. Membership is listed in Supplementary Material

* ‘The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors’.

† Deceased.

‡Corresponding authors:

Aiden Corvin, Dept of Psychiatry & Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Ireland.

Email acorvin@tcd.ie

Chris Spencer, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN.

Emailspencer@well.ox.ac.uk

Abstract

Identifying rare, highlypenetrant risk mutations may be an important step in dissecting the molecular etiology of schizophrenia. We conducted a gene-based analysis of large (>100kb), rare copy number variants (CNVs) in the Wellcome Trust Case Control Consortium 2 (WTCCC2) schizophrenia sample of 1,564 cases and 1,748 controls all from Ireland, and furtherextended the analysis to include an additional 5,196 UK controls. We foundassociation with duplications at chr20p12.2 (P = 0.007) and evidence of replication in large independent European schizophrenia(P=0.052) and UK bipolar disorder case-control cohorts (P = 0.047). A combined analysis of Irish/UK subjects including additional psychosis cases (schizophrenia and bipolar disorder) identified 22 carriers in 11,707 cases and 10 carriers in 21,204 controls (meta-analysis CMH Pvalue=$2x10^{-4}$ (odds ratio (OR)=11.3, 95% CI=*3.7, ∞*)). Nineteenof the 22 cases and 8 of the 10 controlscarried duplications starting at 9.68Mb with similar breakpoints across samples. By haplotype analysis and sequencing we identified a tandem ~149kb duplication overlapping the gene p21 *Protein-Activated Kinase 7* (*PAK7*, also called*PAK5*)which was in linkage disequilibrium with local haplotypes (P=$2.5x10^{-21}$),indicative of a single ancestral duplication event. We confirmed the breakpoints in 8/8 carriers tested and found co-segregation of the duplication with illness in two additional family members of one of the affected probands.We demonstrate that *PAK7* is developmentally co-expressed withanother known psychosisrisk gene (*DISC1*) suggesting a potential molecular mechanism involving aberrant synapse development and plasticity.

Introduction

Schizophrenia [MIM 181500] is a poorly understood, but severe, heritable mental disorder with a lifetime risk of ~1%. The emerging genetic architecture includes a spectrumof risk variation from rare mutations of large effect, to common risk variants of small effect (odds ratio (OR) <1.15)whichcollectively account for at least 25% of susceptibility.[1-3]Through genome-wide association study (GWAS)and subsequent meta-analysis, more than 20 independent common loci have been confirmed, but there are likely to be many more.[4-6] A much smaller number of rare mutations of moderate or large effect have been identified, but these will be particularly important in facilitating dissection of the risk phenotypein model systems.[7]

The list of rare, highly penetrant schizophrenia mutations (OR= 2-30) includes recurrent *de novo*copy number variants (CNVs) involving deletions or duplications of large, genomic regions (>100kb),[8,9] but also the accumulation of differentCNV events at specific loci implicating single genes(*NRXN1* [MIM 600565][10-12] and *VIPR2* [MIM 601970])[13,14]. Almost all of the confirmed CNVs are also risk factors for other psychiatric or developmental phenotypes (e.g. intellectual disability, attention deficit hyperactivity disorder (ADHD), autism).[15,16] This is in keeping with epidemiological and GWAS data supporting shared genetic liability between schizophrenia and other psychiatric disorders, in particular bipolar disorder.[17-19]As current findings explain a modest

proportion of total schizophrenia susceptibility, expanding the number of risk mutations will be important in understanding molecular etiology but also the relationships between these clinical disorders.[20]

We report a gene-based analysis of large (>100kb), rare (<1% minor allele frequency (MAF)) CNVs in the Wellcome Trust Case Control Consortium 2 (WTCCC2) schizophrenia sample of 1,564 cases and 1,748 controls, all from Ireland. We find evidence of association with duplications at 20p12.2 and further supportin a large independent European sample. All of the carriers were from the British Isles and an extended analysis including 32,911Irish/UK subjects provided further association support. By haplotype analysis and sequencing we show the CNV is not the result of a repeated *de novo*eventsand is an inherited risk mutation potentially inherited from a single European ancestor. The mutation involves a tandem duplication of 148,951bp at chr20:9,684,767-9,833,717(hg18) overlapping the gene p21 *Protein-Activated Kinase 7* (*PAK7*, also known as *PAK5*). Other PAK family members modulate synaptic network development through a signaling pathway regulated by the schizophrenia risk gene *DISC1*.[21]We demonstrate that *PAK7* is co-expressed with *DISC1*in developing brain and further investigation of the role of the mutation in synaptic mechanisms salient to schizophrenia is warranted.

Results

Discovery phase evidence of association at four loci

We observed evidence of association (*P* < 0.05) at fourloci in the Irish discovery dataset of 1,564 cases and 1,748 controls; 20p12.2(*PAK7* [MIM 608038]); 2cen-q13 (*ANKRD36B, COX5B* [MIM 123866]; *ACTR1B* [MIM 605144]); 3p25.1 (*MRPS25* [MIM 611987], *ZFYVE20* [MIM 609511]) and a previously confirmedschizophrenia risk locus (chr1q.21 (*CHD1L* [MIM 613039]) locus.[8,9] Association results for the three novel loci are presented in Table 1.To improve our estimate of the population frequency ofwhat are rare events, we examined the three novel loci in an extended control sample (n=6,944) including UK controls from WTCCC2 (2,533 controls from the UK National Blood Service (NBS) and 2,663 from the UK 1958 Birth Cohort (58C)). Details for previously reported schizophrenia CNVs and all CNVs identified in our dataset are included with locus co-ordinates in Supplementary Table S3.

Replication phase association evidence at chr20p12.2

Three loci were carried forward for replication (Table 1) in theindependent International Schizophrenia Consortium (ISC) dataset of 3,111 cases and 2,267 controls from the UK, Portugal, Sweden and Bulgaria.[8]Only the chr20p12.2 locus containing p21 *Protein-Activated Kinase 7* (*PAK7*, also known as *PAK5*) showed evidence for association in the replication sample, with 6 duplication events,all in UK cases, and none in controls (*P* = 0.052). Because this is a rare event, and absent in the control population, the reported association result was the most significant *P*-value obtainable without a larger sample size.

As all of the carriers were identified in samples from the British Isles,we investigated thelocus for additional evidence of association in an independent sample ofbipolar disorder in the WTCCC1 (n=1,697)[22] and University College London (UCL) (n=546) samples.[23] The controls included independent UK controls from the UCL study (n=510) and 10,259 cases ascertained for non-psychiatric disorders in WTCCC1. We identified4duplication carriers in cases and 5carriers in the control sample, provided further nominal evidence of association (P = 0.047).

We performed a combined analysis of all Irish/UK subjects including additional psychosis cases from the CLOZUK study[24] (n=6,223) and independent controls from the WTCCC2 UK control sample(n=7,703).[25,26]Combining the evidence across studies, under the assumption of the same effect on risk in each strata, gave a meta-analysis CMH Pvalue=2×10^{-4} (odds ratio (OR)=11.3, 95% CI=*3.7, ∞)*with22 carriers in 11,707 psychosis cases and 10 carriers in 21,204 controls.

Evidence for a single founder risk mutation at PAK7

Based on the array data, we examined the inferred breakpoints for the chr20p12.2 duplications, and the estimated start and stop coordinates for the 32duplication carriers (Figure 1, Supplementary Table S5).Nineteen of the 22 duplication carriers in the cases and 8 of 10 carriers from the controls exhibited duplications of 132-146.5kb with very similar start positions (~9.68Mb), apparently similar breakpoints across samples and no evidence of flanked

segmental duplications. The evidence for association at the locus in the Irish/UK sample appeared to be driven by theseduplication events.

To explore whether this could represent a single ancestral mutation event, we examined haplotype sharing in the 5 Irish case samples in linkage disequilibrium (LD) blocks (termed Hap Block 1 and Hap Block 2) immediately 5' and 3' of the inferred duplication.Haplotypes were phasedin PLINKwith a probability >0.98. All 5 Irish cases carried a copy of the same haplotype in both Hap Block 1(CCTT, f = 0.188) and Hap Block 2 (TCAA, f = 0.208) (Supplementary Tables S6, S7). We extended the analysis to include 18 of the 22 UK carriers (as these had been genotyped on the same Affymetrix platform) and found that all carried copies of the CCTT and TCAA haplotypes flanking the duplication (Supplementary Table S8). Given the low level ofLDbetween these blocks (D'=0.22, r^2=0.02) the probability of making these observations if the duplication was the result of repeated *de novo* mutations is extremely low(P=2.5x10^{-21}),providing support for a single ancestral duplication event.We did not find an increased level of relatedness among the Irish *PAK7* duplication carriers compared to the level of relatedness of random Irish individuals, arguing against a very recent event (see Supplementary Figure 1).

Next, we confirmed the duplication event in one individual (IRL_101) using an Agilent custom designed comparative genomic hybridization (CGH) microarray with high probe density in the *PAK7* region. The breakpoints identified by CGH

(chr20:9,684,902-9,833,151) map closely to those predicted by the SNP array analyses (chr20:9,685,413-9,831,947; Supplementary Figure S2).The event overlaps exon 1 and enhancer and promoter regions adjacent to exon 1 in all four known *PAK7* transcripts, and exon 2 in two alternative transcripts and is absent from the Database of Genomic Variants (http://projects.tcag.ca/variation/). By capillary sequencingthis individualwe identified a tandem duplication of 148,951bp at chr20:9,684,767-9,833,717(hg18) with a 15bp sequence inserted between the first and second copies of the repeated sequence (Figure 2). By sequencing the five remaining Irish carriers and twoof the UK carriers (SCOT_101, SCOT_201) weconfirmedthat all shared exactly the same breakpoints with the same 15bp sequence inserted between the two copies of the repeat sequence, suggesting that this represents the same duplication event.

Assessment of additional family members of carriers

We were able to access additional family members of two cases to test forevidence of co-segregation of the duplication with illness(see Figure 3). In the family of IRL-101 there were three other surviving siblings, none of whom had a diagnosis of schizophrenia or a related major mental disorder. One sister (IRL_103) agreed to be tested. She had experienced a depressive episode after bereavement and did not carry the duplication. By going back to the original GWAS data we identified another Irish case (IRL_601) that had been excluded from the primary analysisbecause they are related to IRL_101 based on a proportion of identity-by descent (IBD) estimated at 0.033. IRL_601is a female

patient with a history of Schizoaffective Disorder. She is the youngest in a sibship of five children of which four were affected with a major psychotic disorder. We confirmed that two affected brothers (IRL_603 (Bipolar I Disorder), IRL_604 (Psychotic Disorder, Not Otherwise Specified)) also carry the *PAK7* duplication. The remaining affected brother (IRL_602 (Bipolar I Disorder)) died by suicide. We tested one of the eight offspring of IRL_601 and this unaffected individual did not carry the duplication.In the SCOT_1 family we confirmed that the duplication was inherited from the mother (SCOT_103). Neither of the parents wasaffected clinically, but there was a family history of depression in both families, and a brother with a history of psychosis did not carry the duplication (SCOT_104).

Is the mutation present in other European ancestry populations?

To assess whether the mutation was exclusive to the British Isles or present elsewhere in Europe we examined additional European populations. Details of the cohorts analyzed are included in Supplementary Table S1. The estimated control frequency for the mutation in the UK/Irish population is low (f= 0.037%) and most of these cohorts were underpowered to provide an accurate frequency estimate forwhat is a rare event. We failed to identify any carriers in relatively large Icelandic andFinnish samples, but did find carriers in populations from the Eastern US, the Netherlands, Sweden and Denmark. We were able to unambiguously phase haplotypes in 5 US control carriers with available Affymetrix data, confirming that all of these individuals shared the same haplotype background as the Irish/UK cases, suggesting that they carry the same

mutational event although the samples were not available for confirmatory sequencing. These data suggest that the mutation is present in other European populations. The limited number of cases available precluded formal association analyses, but we identified 4 additional cases with the duplication in Dutch, Danish and Swedish cohorts. Information on the clinical characteristics, illness course, co-morbidity and family history for all mutation case carriers is provided in Supplementary Table S9.

Functional investigation of PAK7

The p21-activated kinases (PAKs) are a family of serine/threonine protein kinases, which are regulated by the Rho family of small G proteins and are involved in multiple intracellular signaling pathways. Six PAK genes are expressed in human and based on their regulatory functions are classified into Group I (*PAK 1-3*) and Group II (*PAK4-6*) members.[27]Group I PAKs are activated by RAC-PAK signaling to promote axon connectivity, and synapse formation, in the developing brain in a pathway regulated by another schizophrenia risk gene *DISC1* [MIM:60521].[28] Aberrant synaptic network development represents a plausible molecular mechanism for psychosis susceptibility and mutations involving other Group I PAKs (*PAK2* (the 3q29 microdeletion syndrome locus [MIM 609425]) and *PAK3* [MIM:300558]) are known to be associated with neurodevelopmental syndromes characterized by psychosis.[28] *PAK7*is a brain-specific isoform within the cell it is localized to filipodia, where it has been shown to promote the induction of neurite outgrowth, filopodium formation and synaptic vesicle trafficking.[29]The Group II PAKs, including *PAK7*, differ

structurally from Group I members and their mechanism of activation requires clarification. To test whether *PAK7*may be under similar regulatory control to Group I PAKs during early brain development, we investigated the relationship between *PAK7* and *DISC1*. Ina co-immunoprecipitation experiment we confirmed interaction between *PAK7* and *DISC1* in synaptoneurosomal preparations from full mouse brain at post-natal day 8-10 (Figure 4). This suggests that *PAK7*is developmentally co-expressed with *DISC1* and may have a specific functional role at the time of synapse formation.

Discussion

We report evidence that rare, chr20p12.2 duplications increase risk for the major psychotic disorders, schizophrenia and bipolar disorder. Previous schizophrenia studies have identified at least eight rare copy number variants of strong effect (genotypic relative risks of 4-20).[5] Most of these CNVs have been identified because they recur in the population, either as relatively large events implicating many genes and arising from non-allelic homologous recombination (NAHR), or through the accumulation of different, generally *de novo* mutations at a single locus (e.g. *NRXN1*, *VIPR2*). We identified 2,653 genes impacted by a CNV (>100kb but <10Mb) in at least one individual in our study. However, only 382 genes had the4 or more overlapping CNVs required to achieve $P<0.05$ in our discovery sample. Based on our data, the overall contribution of CNVs to psychosis susceptibility risk is difficult to estimate as we could not exclude more rare, or smaller events. Ongoing CNV analysis of the Psychiatric Genomics

Consortium dataset will be important in more comprehensively investigating rare, large CNV events.[17]

By haplotype analysis and sequencing we established that the association is driven by a ~149kb duplication at the gene *PAK7* with a likely common European ancestor. Inherited mutations have been described contributing to other neurological disorders, but their relative contribution to schizophrenia susceptibility is unknown.[31,32]Such events are likely to be recent in origin and differentiated between populations making them difficult to replicate across diverse collections usingstandard gene association methods in large multi-cohort CNV studies.[33-35]By examining cases excluded, through relatedness QC, from our original Irish GWAS dataset we identified an additional duplication carrier related at a distance of five meiosis to one of the index cases. Within the family of this additional carrier we found evidence of co-segregation of the duplication with psychosis in two affected siblings. Identifying cryptic relationships between individuals in available GWAS datasets may be helpful in identifying extended pedigrees for analysis as a complementary approach to existing association methods.

The risk duplication identified in this study overlaps the first two exons and enhancer and promoters regions in the known transcripts of *PAK7*. Because *PAK7* is exclusively brain expressed the impact of the duplication on gene expression and the genetic mechanism involved are yet to be determined. This is

important as we identified other larger duplication events at the locus in two control samples, but do not know if these would have the same or different effects on gene function. From existing animal models of *PAK7* we do know that knockout mice are viable with no obvious developmental abnormalities but *PAK6/PAK7* double knockout mice show behavioral and learning deficits suggesting functional redundancy between these isoforms.[36] Genomic studies in carriers of the founder mutation will be important as additional risk variants or mutations (e.g. in *DISC1* or *PAK6*) may be contributing to risk in this extended pedigree.[37]

The *PAK7* risk mutation is rare, but based on data from the British Isles we estimate the exposed attributable risk for carriers is likely to be substantial(~68%).[38]Almost all of the risk CNVs identified to date in the schizophrenia literature also increase susceptibility to other developmental phenotypes including intellectual disability, autism andseizure disorder. Only four of our cases received a full cognitive assessment (including IQ measurement) and all were assessed as being within the normal range of cognitive function. One individual had a history of co-morbid language delay and another had a history of seizure disorder. Although most of the cases had schizophrenia, seven had bipolar disorder or schizoaffective disorder. This is in keeping with evidence of shared genetic etiology between these disorders, although the evidence for CNV involvement in bipolar disorder has until now been more equivocal.[39,40]No information on cognitive function, psychiatric morbidity or developmental history was available for any of the controls

identified. Further clinical and genetic assessment of patient carriers, their families and carriers identified in other disease or normal populationswill be important in quantifyingthe penetrance and range of phenotypic expression of the *PAK7*duplication.

Schizophrenia is a neurodevelopmental disorder and a disease model involving aberrant synaptic network development has been proposed.[41]Group I PAKs play a role in the development of these networks through a signaling pathway regulated by an established susceptibility gene (DISC1).[42]*PAK7* is a brain-specific isoform, within the cell it is localized to filipodia, where it has been shown to promote the induction of neurite outgrowth, filopodium formation and synaptic vesicle trafficking.[29]To clarify its mechanism of activation we performed functional studies that confirmed interaction between *PAK7* and *DISC1*, suggesting that this gene may be under similar regulatory control to other PAK family members.

In conclusion, we have identified a psychosis risk duplication at the gene *PAK7*, with evidence that it is inherited from a common ancestor. The gene functions in the development of synaptic networks and may be under the regulatory control of *DISC1*. Many of the cases segregating the DISC1 translocation in the original Scottish schizophrenia pedigree were affected with schizoaffective disorder or mood disorders as was the case in the largest*PAK7* pedigree we investigated and we also found evidence for association with bipolar disorder.[43]This suggests a

broader molecular risk mechanism for psychosis. A critical next step will be to understand how the duplication impacts on gene expression and function so that it can be investigated in model systems. As *PAK7* is exclusively brain-expressed,this will require further experimental work.

Materials & Methods

Study subjects

The discovery sample included 1,564 cases and 1,748 controls from the Irish Schizophrenia Genomics Consortium/WTCCC2 GWAS study which has previously been described.[44]Participants, from the Republic of Ireland or Northern Ireland, were interviewed using a structured clinical interview and diagnosis of schizophrenia (n=1,418) or a related disorder (schizoaffective disorder (n=182); schizophreniform disorder (n=6)) was made by the consensus lifetime best estimate method using DSM-IV criteria. Control subjects were ascertained with written informed consent from the Irish GeneBank and represented blood donors from the Irish Blood Transfusion Service. Cases and controls met the same ethnicity criteria (Irish origin with all 4 grandparents born in Ireland or the UK). To improve our estimate of the population frequency ofwhat are rare events, we examined nominally significant loci in an extended control sample (n=6,944) including UK controls from WTCCC2 (2,533 controls from the UK National Blood Service (NBS) and 2,663 from the UK 1958 Birth Cohort (58C)).[25]

The first phase of replication examined non-overlapping subjects from the International Schizophrenia Consortium dataset.[8] To follow up evidence of association in the British Isles we performed further association analysis in bipolar disorder cases and controls from the UK/Ireland. The WTCCC1 bipolar disorder sample included 1,697 cases recruited throughout the UK. Most of the cases had a diagnosis of bipolar I disorder (71%) or schizoaffective disorder (15%). non-psychiatric disease case participants in the WTCCC1 study (n=10,769).[22]The UCL bipolar disorder sample included 546 individuals (97% bipolar I disorder) and comparison subjects (n=510) with no personal or first-degree history of any mental disorder and from a similar UK or Irish ancestry, based on the origin of all four grandparents.[23]

Finally in an extended analysis we included additional psychosis cases and controls from the UK. The additional psychosis cases were of Caucasian origin and came from the CLOZUK sample (n=6,223).[24]The CLOZUK sample consists of patients taking the antipsychotic medication Clozapine, a drug reserved for the treatment-resistant psychosis patients in the UK. The remaining control subjects were ascertained as described in previous WTCCC2 studies from the UK National Blood Service (NBS) (n=2,553); the UK 1958 Birth Cohort (n=2,633); the People of the British Isles (POBI) study) and from.[25,26] We also investigated additional European ancestry cohorts for evidence of the PAK7 duplication and details on the samples tested, case and/or control numbers and genotyping platform are provided in Supplementary Table S1.[5,9,45,46]

CNV calling and validation

We report a schizophrenia study investigating large (>100kb), rare (<1% minor allele frequency (MAF) in all samples) CNVs using data from the Wellcome Trust Case Control Consortium 2 (WTCCC2). All discovery samples were genotyped using the Affymetrix 6.0 platform either at the Affymetrix (Santa Clara, California, USA) or Broad Institute (Cambridge, Massachusetts, USA) laboratories. Samples were processed using the WTCCC2 pipeline and quality control details have been reported previously.[24, 42]

CNV calls were created using Birdseye from Birdsuite (version 1.5.5)[47]for autosomes only and we excluded calls where lengths were <100kb or >10Mb, or LOD score <10. We excluded CNVs with at least 50% overlap with a region copy number variable in at least 1% of samples; individuals with >30 CNV calls or a total event length >10Mbp and calls for samples from plates containing fewer than 40 samples (Supplementary Table S2). Calls with a copy number of 0 or 1 were considered to be deletions and calls with a copy number of 3 or 4 were considered duplications.

For the replication analysis, the International Schizophrenia Consortium data was called using the same calling protocol as in the discovery analysis. Details on genotyping platforms used for the other samples are provided in Supplementary Table S1. For the identified *PAK7* risk duplication we performed standard qPCR validation of the CNV calls in the discovery sample (Supplementary Table S4). Verifying the presence of CNVs by this method addresses sensitivity, but not the specificity of CNV calling. To test the specificity and sensitivity of calling across genotyping platforms we examined probe intensity data for 6,542 control

subjects who had been genotyped on Affymetrix 6.0 and Illumina 1.2M-Duo arrays as part of the wider WTCCC2 study. We called CNVs on the Illumina platform using QuantiSNP[48]and excluded those with log Bayes factor < 10, length <100 kb or > 10Mb. We excluded samples with more than 10 CNVs in total, with more than 10 MB of total CNV length, or failed SNP quality control. This identified the same four individuals as the only PAK7 duplication carriers across different arrays and calling algorithms. For the additional European ancestry cohorts CNV calls were made as described in the primary publications (see Supplementary Table S1 and references [5,9,45,46]). The CLOZUK calling method is described in Guha et al (2013).[49]

CNV analysis

We conducted a gene-based analysis using gene boundaries from UCSC Genome Browser refGene (hg18) and identified 2,653 genes impacted (from the transcription start to end point) by a CNV (>100kb but <10Mbp) in at least one individual. 382 genes had at least 4 overlapping CNVs (required to achieve $P<0.05$) in the discovery data, representing a smaller number of loci, as some CNVs overlap multiple genes. Fisher's exact tests were used to calculate *P* values in the discovery analysis and a fixed-effects Cochran-Mantel-Haenszel (CMH) test to assess the evidence across replication cohorts.

Haplotype analysis

We analyzed carriers of the 'common' *PAK7* duplication at chr20:9685413-9831947, which is carried by 27 of 32 individuals with duplications overlapping

the *PAK7* locus in the British Isles samples. Haplotype analysis used a core set of Affymetrix SNPs that had been genotyped in 23 of the samples and all genotype data was converted to the forward strand. The 4 samples omitted were from the CLOZUK sample genotyped on Illumina where no suitable proxy SNPs were available for analysis.FromHapMap data, the start of the duplication sits in a haplotype block that extends from rs2423462 (9682770) to rs2423467 (9689876). The part of this block that is outside the region, termed Hap Block 1 is from rs2423462 (9682770) to rs742450 (9684493). The duplication ends in a region between haplotype blocks. Immediately 3' of the duplication is a haplotype block that extends from rs6057009 (9840327) to rs6118819 (9867572), (Hap Block 2). We have focused on haplotypes outside the duplication region, to analyze diploid genotypes, and limited the analysis to regions of high LD where haplotypes could be phased. In the absence of family data, extended haplotypes outside these local blocks could not be accurately phased. Haplotype frequencies and phased haplotypes, with a probability >0.98, were estimated using the --hap and --hap-phase functions in Plink (see Supplementary Tables S6, S7). For either Hap Block, the chances that all 23 samples would carry at least one copy of a specific haplotype, i.e. be either heterozygous or homozygous for this haplotype q = (q^2*(probability of homozygous q)* + 2*q*(1-q) *(probability of homozygous q)*)^{23} = $(2q - q^2)^{23}$ where q is the haplotype frequency. This probability was calculated for both Hap Blocks and in the absence of LD between the blocks, these numbers were multiplied together to determine the combined probability of all 23 carriers carry copies of the same haplotypes either side of the duplication.

CNV breakpoint sequencing

Using an Agilent custom designed comparative genomic hybridization (CGH) microarray with high probe density in the *PAK7* region, we confirmed the duplication event in one individual and the breakpoints identified by CGH (chr20:9,684,902-9,833,151) map closely to those predicted by the SNP array analyses (chr20:9,685,413-9,831,947; Supplementary Figure S4). Based on these estimated breakpoints and assuming that this is a tandem duplication that is not inverted, we attempted to sequence from one copy of the duplication into the second copy in order to identify the precise breakpoints. Using a forward primer positioned at the end of the duplication (TCTCTGTTGGATGGAGCTTCT) and a reverse primer positioned at the start of the duplication (CGATGTAAAAAGACACAAGAGAAA), we successfully PCR-amplified this unique region in a carrier sample. The PCR did not amplify in non-carrier samples. Using capillary sequencing, we identified the event as being a tandem duplication of 148,951bp at chr20:9,684,767-9,833,717(hg18) with a 15bp sequence inserted between the first and second copies of the repeated sequence (Figure 2).

Functional analyses

Protein extraction

Synaptoneurosome preparation was performed from full mouse brain at post-natal day 8-10 (P8-10) as described elsewhere with modifications.[50]Briefly, brains were glass/glass homogenized in ice; the lysis buffer was daily fresh-

made in ultrapure water containing 10mM HEPES, 2mM EDTA, 2mM EGTA, 150mM sodium chloride and added with Protease Inhibitor Cocktail (Roche Diagnostic) and Phosphatase Inhibitor Cocktail 3 (Sigma). The homogenate was centrifuged at 1000g X 10 minutes in order to separate heavy cellular membranes. The resulting supernatant was filtered through two 105 μm polypropylene mesh (Amazon US) and a 5μm nitrocellulose filter (Millipore) and finally centrifuged at 1000g X 15 minutes. Synaptoneurosome pellet was resuspended in lysis buffer and stored at -80 degrees.

Co-Immunoprecipitation

For the immunodetection of PAK7 in synaptoneurosome samples, Western Blotting was carried out by standard procedures with the following specifications: samples were fractioned through 10% acrylamide gel and blotted onto PVDF membranes. The membranes were probed with rabbit polyclonal IgG anti-PAK7 (Sigma, 3500335, 1:1000) and rabbit polyclonal anti-βactin (Cell Signaling Technology, 4967, 1:5000), followed by biotinylated anti-rabbit IgG (VECTOR, BA-1000, 1:2000) and anti-rabbit IgG HRP-linked (Cell Signaling Technology, 7074, diluted 1:10000) for PAK5 and β-actin respectively. The signal for PAK7 was further amplified with Vectastain ABC kit (Vector, PK-6100). The amount of PAK7 in each sample was measured with ImageJ and normalized to the endogenous level of β-actin. The co-immunoprecipitation procedure was adapted from a previously published method.[51]A full mouse brain was glass/glass homogenized in non-denaturing lysis buffer containing 10mM Tris-HCl (pH adjusted at 8), 0.1M sodium chloride, 4mM EDTA, 0.1% NP40 with

Protease Inhibitor Cocktail (Roche). After preliminary experiments to confirm the ability of the antibody (goat polyclonal IgG anti-PAK5, sc-22155, Santa Cruz Biotechnology Inc.) to immunoprecipitate the PAK7 protein, we performed a co-immunoprecipitation experiment in order to assess the interaction between PAK7 and DISC1. In brief, 1000μg of total protein extract was immunoprecipitated with goat anti-PAK7 and incubated with protein A/G PLUS-agarose (sc-2003, Santa Cruz Biotechnology, Inc.). Normal goat IgG (Sigma, I5256) was used as a negative control. Samples were then resolved by Western Blotting as described above and probed with rabbit polyclonal IgG anti-DISC1 (Invitrogen, 40-6900, 1:1200), followed by incubation with biotinylated anti-rabbit IgG diluted 1:3000 and amplification with Vectastain ABC kit (see Figure 4).

Acknowledgements

The authors sincerely thank all patients who contributed to this study and all staff who facilitated their involvement. Funding for this study was provided by the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z), the Wellcome Trust (072894/Z/03/Z, 090532/Z/09/Z and 075491/Z/04/B), NIMH grants (MH 41953 and MH083094) and Science Foundation Ireland (08/IN.1/B1916). We acknowledge use of the Trinity Biobank sample from the Irish Blood Transfusion Service, the Trinity Centre for High Performance Computing, British 1958 Birth Cohort DNA collection funded by the Medical Research Council (grant G0000934) and the Wellcome Trust (grant 068545/Z/02) and of the UK National Blood Service controls funded by

the Wellcome Trust. We thank W. Bodmer and B. Winney for use of the People of the British Isles DNA collection, which was funded by the Wellcome Trust. Chris Spencer is supported by a Wellcome Trust Career Development Fellowship [097364/Z/11/Z].

Conflict of interest

The authors report no conflicts of interest

References

1. Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J.A., Karayiorgou, M. (2012) De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.*, **44**:1365-1369.
2. Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., Thibodeau, P., et al. (2011) Increased exonic do novo mutation rate in individuals with schizophrenia. Nat. Genet., **43**: 860-863.
3. Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Schizophenia Psychiatric Genome-wide Association Study Consortium (PGC-SCZ), International Schizophrenia Consortium (ISC), Molecular Genetics of Schizophrenia Consortium (MGS), Sullivan, P.F., Goddard, M.E., Keller, M.C., et al. (2012)

Estimated proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet*, **44**: 247-250.

4. Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.Y., Duan, J., Ophoff, R.A., Andreassen, O.A., et al. (2011) Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.*, **43**: 969-976.

5. Ripke, S., O'Dushaline, C., Chambert, K., Moran, J.L., Kahler, A.K., Aakterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al. (2012). Genome-wide association analysis identified 13 new risk loci for schizophrenia. *Nat. Genet.*, in press.

6. Shi, Y., Li, Z., Xu, Q., Wang, T., Li, T., Shen, J., Zhang, F., Chen, J., Zhou, G., Ji, W., et al. (2011) Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat. Genet.,* **43**: 1224-1227.

7. Karyiorgou, M., Flint, J., Gogos, J.A., Malenka, R.C., Genetic and Neural Complexity in Psychiatry 2011 Working Group (2012) The best of times, the worst of times for psychiatric disease. *Nat. Neurosci*. **15**: 811-812.

8. International Schizophrenia Consortium (2008). Rare chromsomal deletions and duplications increase risk of schizophrenia. *Nature,* **455**: 237-241.

9. Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O.P., Ingason, A., Steinberg, S., Fossdal, R. Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., et al.(2008) Large recurrent microdeletions associated with schizophrenia. *Nature,* **455**: 232-236.

10. Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R., Toulopoulo, T., Picchioni, M., Vassos, E., Ettinger, U., Braemon, E., et al. (2008) Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol. Genet.,* **18**: 988-996.

11. Kirov., G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., O'Donovan, M.C., Erdogan, F., Owen, M.J., Ropers, H.H., et al. (2008) Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum. Mol. Genet.,***17**: 458-465.

12. Kirov, G., Rujescu, D., Ingason, A., Collier, D.A., O'Donovan, M.C., Owen, M.J. (2009) Neurexin 1 (NRXN1) Deletions in Schizophrenia. *Schizophr. Bull.,* **35**: 851–854.

13. Vacic,V., McCarthy, S., Malhotra, D., Murray, F., Chou, H.H., Peoples, A., Makarov, V., Yoon, S., Bhandari, A., Corominas, R., et al. (2011) Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature,* **471**: 499-503.

14. Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., Zhang, N., Mowry, B.J., Olincy, A., Amin, F., et al.(2011) Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am. J. Psychiatry,* **16**: 302-316.

15. Mefford, H.C., Sharp, A.J., Baker, C., Itsara, A., Jiang, Z., Buysse, K., Huang, S., Maloney, V.K., Crolla, J.A., Baralle. D., Collins, A., et al. (2008) Recurrent

rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.,* **359**: 1685-1699.

16. Sullivan, P.F., Daly, M.J., O'Donovan, M. (2012) Genetic architecture of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.*, **13**: 537-551.
17. Cross Disorder Group of the Psychiatric Genomics Consortium, Smoller, J.W., Craddock, N., Kendler, K.S., Lee, P.H., Neale, B.M., Nurnberger, J.I., Ripke, S., Santangelo, S., Sullivan, P.F. (2013) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet***381**: 1379-1379.
18. Morgan. V.A., Croft, M.L., Valuri, G.M., Zubrick, S.R., Bower, C., MCNeil T.F., Jablensky, A.V. (2012) Intellectual disability and other neuropsychiatric outcomes in high-risk children of mothers with schizophrenia, bipolar disorder and unipolar major depression. *Br. J. Psychiatry,***200**: 282-289.
19. Lichtenstein, P., Yip, B.H., Bjork, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F, Hultman, C.M. (2009) Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*, **373**: 234-9.
20. Corvin, A. (2013) Schizophrenia at a genetics crossroads: where to now? *Schizopr. Bull.***39**: 490-495.
21. Porteous, D.J., Millar, J.K., Brandon, N.J., Sawa, A. (2011) DISC1 at 10: connecting psychiatric genetics and neuroscience *Trends. Mol. Med.,* **17**: 699-706.

22. Wellcome Trust Case Control Consortium (2007). Genome-wide study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* **447**: 661-678.

23. McQuillin, A., Bass, N., Anjorin, A., Lawrence, J., Kandaswamy,R., Lawrence, J., Curtis, D., Sklar, P., Purcell, S.M., Gurling, H.M. (2011) Analysis of genetic deletions and duplication in the University College London bipolar disorder case control sample. *Eur. J. Hum. Genet.,* **19**: 588-592.

24. Hamshere, M.L., Walters, J.T., Smith, R., Richards, A.L., Green, E., Grozeva, D., Jones, I., Forty, L., Gordon-Smith, K., Riley, B., et al (2013). Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Mol. Psychiatry***18**: 708-712.

25. UK IBD Genetics Consortium, Barrett, J.C., Lee, J.C., Lees, C.W., Prescott, N.J., Anderson, C.A., Phillips, A., Wesley, E., Parnell, K., Zhang, H., et al.(2009) Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat. Genet.,***41**: 1330-1334.

26. Winney, B., Boumertit, A., Day, T., Davison, D., Echeta, C., Evseeva, I., Hutnik, K., Leslie, S., Nicodemus, K., Royrvik, E.C., et al. (2012) People of the British Isles: preliminary analysis of genotypes and surnames in a UK-control population. *Eur. J. Hum. Genet.,***20**: 203-210.

27. Arias-Romero, L.E., Chernoff, J. (2008) A tale of two Paks. *Biol. Cell.*, **100**: 97-108.

28. Chen, S.Y., Huang, P.H., Cheng, H.J. (2011) Disrupted-in-Schizophrenia 1-mediated axon guidance involves TRIO-RAC-PAK small GTPase pathway signaling. *Proc. Natl. Acad. U.S.A*., **108**: 5861-5866.

29. Strochlic, T.I., Concilio, S., Viaud, J., Eberwine, R.A., Wong, L.E., Minden, A., *et al*. Identification of neuronal substrates implicates Pak5 in synaptic vesicle trafficking. *Proc Natl Sci USA 2012*; **109**: 4116-4121.

30. Peippo, M., Koivistro, A.M., Sarkamo, T., Sipponen, M., von Koskull, H., Ylisaaukko-oja, T.,Rehnstrom, K., Froyen, G., Ignatuis, J., Jarvela, I. (2007) PAK3 related mental disability: further characterization of the phenotype. *Am. J. Med. Genet.,***143A**: 2406-2416.

31. Smith, B.N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., Miller, J., Lee, Y., Troakes, C., Scott, K.M., et al. (2013) The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. *Eur. J. Hum. Genet.,* **21**: 102-108.

32. Magana, J.J., Gomez, R., Maldonado-Rodriguez, M., Velasquez-Perez, L.,Tapia-Guerrero, Y.S., Cortes, H., Levya-Garcia, N., Hernandez-Hernandez, O., Cisneros, B. (2013) Origin of the Spinocerebellar ataxia type 7 mutation in Mexican population. *Cerebellum* [Epub ahead of print].

33. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., MCVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**: 56-65.

34. Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., 1000 Genomes Project, Bustamante, C.D. (2011) Demographic history and rare allele sharing among human populations. *Proc. Natl. Acad. U.S.A.*, **108**:11983-11988.

35. Mathieson, I. & McVean, G. (2012) Differential confounding of rare and common variants in spatially structured populations. *Nat. Genet.*, **44**: 243-246.

36. Nekrasova, T., Jobes, M.L., Ting, J.H., Wagner, G.C., Minden, A. (2008) Targeted disruption of the Pak5 and Pak6 genes in mice leads to deficits in learning and locomotion. *Dev. Biol.,* **322**: 95-108.

37. Lupski, J.R., Belmont, J.W., Boerwinkle, E., Gibbs, R.A., (2011) Clan genomics and the complex architecture or human disease. *Cell***147**:32-43.

38. Benichou, J. (2007) Biostatistics and epidemiology: measuring the risk attributable to an environmental or genetic factor. *C.R. Biol.*, **330**: 281-298.

39. Grozeva, D., Kirov, G., Ivanov, D., Jones, I.R., Jones, L., Green, E.K., St Clair, D.M., Young, A.H., Ferrier, N., Farmer, A.E. et al. (2010) Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch. Gen. Psychiatry,* **67**: 318-372.

40. Malhotra, D., McCarthy, S., Michaelson, J.J., Vacic, V., Burdick, K.E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E.S., et al.(2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron,* **72**: 951-963.

41. Meyer-Lindenberg A. From maps to mechanisms through neuroimaging of schizophrenia (2010). *Nature,* **468**: 194-202.

42. Hayashi-Takagi, A., Takaki, M., Graziance, N., Seshadri, S., Murdoch, H., Dunlop, A.J., Makino, Y., Seshadri, A.J., Ishizuka, K., Srivastava, D.P., et al.(2010) Disrupted-in –Schizophrenia-1 (DISC1) regulates spines of the glutamate synapse via Rac1. *Nat. Neurosci.,* **13**: 327-332.

43. St Clair, D., Blackwood, D., Muir, W., Carothers, A., Walker, M., Spowart, G., Godson, C., Evans, H.J. (1990) Association within a family of a balanced translocation with major mental illness. *Lancet*, **336**: 13-6.

44. Irish Schizophrenia Genomics Consortium and the Wellcome Trust Case Control Consortium 2 (2012) Genome-wide association study implicates HLA-C*01:02 as a risk factor at the MHC locus in schizophrenia. *Biol Psychiatry,* **72**: 620-628.

45. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R.,Hannig, V., et al (2011) A copy number variation morbidity map of developmental delay. *Nat. Genet.,* **43**: 838-46.

46. The ARIC Investigators (1989) The atherosclerosis risk in communities (ARIC) study: design and objectives. *Am J Epidemiol*., **129**: 687-702.

47. Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., et al. (2008) Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet,* **40**: 1253-1260.

48. Collela, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller., A., Holmes, C.C., Ragoussis, J. (2007) QuantiSNP: an objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res,* **35**: 2013-25.

49. Guha, S., Rees, E., Darvasi, A., Ivanov, D., Ikeda, M., Bergen, S.E., Magnusson, P.K., Cormican, P., Morris, D., Gill, M., et al. (2013) Implications of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA Psychiatry*, **70**:252-60.

50. Hollingsworth, E.B., McNeal, E.T., Burton, J.L., Williams, R.J., Daly, J.W., Creveling, C.R. (1985) Biochemical characterization of a filtered synaptoneurosome preparation from guinea pig cerebral cortex: cyclic adenosine 3':5'-monophosphate-generating systems, receptors, and enzymes. *J Neurosci*, **5**:2240-2253.

51. Felli, M.P., Vacca, A., Calce, A., Bellavia, D., Campese, A.F., Grillo, R., Giovine, M., Checquolo, S., Talora, C., Palermo, R. et al. (2005) PKθ mediates pre-TCR signaling and contributes to Notch3-induced T-cell leukemia.*Oncogene*, **24**: 992–1000.

Figures

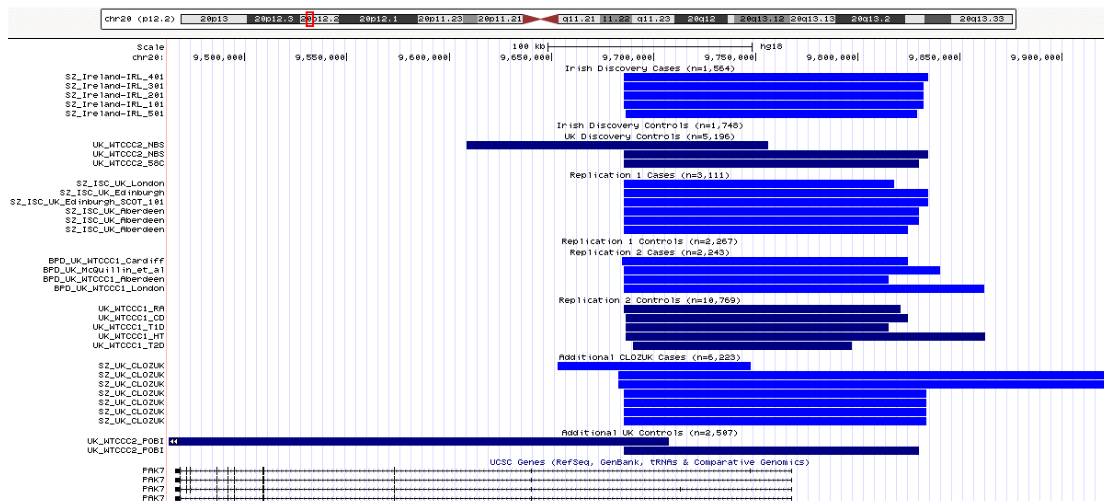

Figure 1: Inferred position of duplication events at the chr20p12.2 locus in build hg18. SZ=schizophrenia. BD=bipolar disorder. See Supplementary table 5 for exact start/stop coordinates. Duplications in controls are colored light blue and SZ or BP cases are colored dark blue.

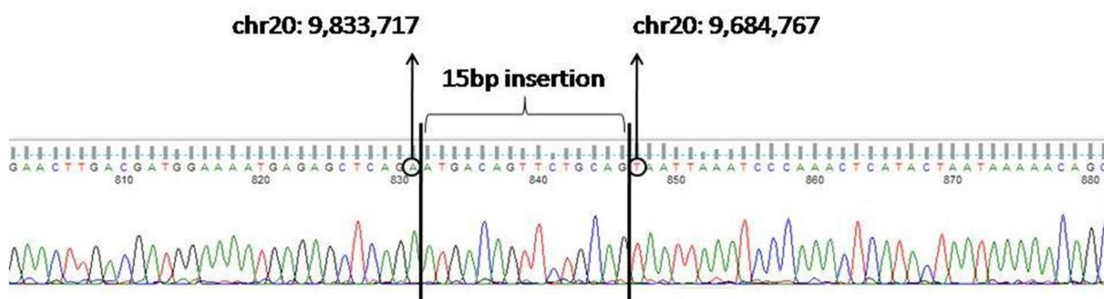

Figure 2: Capillary sequence of the unique region in the middle of the two copies of the tandem repeat sequence at the *PAK7* duplication. The first copy of the sequence ends at chr20:9,833,717(hg18) and the second copy starts with sequence from chr20:9,684,767(hg18), thereby defining the breakpoints. A 15bp sequence is present between the two copies.

Figure 3: Familial pedigrees for patients from 3 families (IRL_6, IRL_1, SCOT_1) who were found to be positive for *PAK7* duplications. The index individual in each family screened is indicated with an arrow. Individuals positive (PAK7+VE) or negative (PAK7-VE) for PAK7 are indicated. Filled symbols indicate individuals with Psychotic disorders (SCZ= Schizophrenia, SCA= Schizoaffective disorder, BPA=Bipolar affective disorder, PDNOS=Psychotic Disorder Not Otherwise Specified). Partial vertically shaded symbol indicates Depressive disorder (DEP= Depression). Partial horizontal shaded symbol indicates Learning disability (LD). Cause of death where known: SUIC=Suicide, BA=Birth Asphyxia, CT=Cerebral tumour.

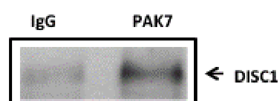

Fig. XXX: PAK7 co-localize with DISC1 in mouse brain extracts.
The anti-PAK (right lane) was used to immuno-precipitate the protein from brain extracts and the membrane was then blotted for anti-DISC1. In the control lane (left), the IgG were used as controls and the anti-DISC1 did not detect any significant staining.

Figure 4: Increase in neuronal activity induces expression of PAK7. Immunostaining in Primary cortical neurons. A: representative images of primary cortical neurons in control conditions (left) or after stimulation with K=100mM for 20 minutes. B: plot of cytoplasmic PAK7 immunostaining in control conditions and after K+ stimulation (p=0.035).

Supplementary material

An inherited duplication at the gene *p21 Protein-Activated Kinase 7 (PAK7)* is a risk factor for psychosis.

SUPPLEMENTARY INFORMATION

Table S1: Details of all samples in the study

Table S2: Sample exclusion details for Discovery sample

Table S3: Previously reported associated CNVs and all CNVs in discovery

Table S4: Results of qPCR validation

Table S5: Start and stop coordinates for *PAK7* duplication in carriers

Table S6: SNP haplotypes at *PAK7* locus

Table S7: Haplotype frequencies and phased haplotypes estimation

Table S8: Phased haplotypes for each block at *PAK7* locus

Table S9: Clinical details and family history in *PAK7* carriers

Supplementary Figures S1-3: Legends

Acknowledgments Wellcome Trust Case Control Consortium 2

TABLE S1: Details of all samples used in the study				
Study	**Cohort (Case Phenotype)**	**Array**	**# Cases**	**# Controls**
Discovery				
WTCCC2	Ireland (SZ)	Affymetrix 6.0	1,564	1,748
WTCCC2	UK - 58BC	Affymetrix 6.0		2,663
WTCCC2	UK - NBS	Affymetrix 6.0		2,533
Replication 1				
ISC	UK - Aberdeen (SZ)	Affymetrix 5.0	727	694
ISC	UK - London (SZ)	Affymetrix 5.0	547	0
ISC	UK - Edinburgh (SZ)	Affymetrix 6.0	403	290
ISC	Portugal (SZ)	Affymetrix 5.0	333	200
ISC	Sweden (SZ)	Affymetrix 5.0/6.0	622	437
ISC	Bulgaria (SZ)	Affymetrix 6.0	479	646
Replication 2				
WTCCC1	UK (BPD)	Affymetrix 500K	1,697	10,259[a]
McQuillin et al.	UK (BPD)	Affymetrix 500K	546	510
Additional UK Samples				
CLOZUK	UK (SZ)	Illumina OmniExpress/Combo	6,223	
WTCCC2	UK - POBI	Affymetrix 6.0		2,507
Other European and European Ancestry Samples				
SGENE	Iceland (SZ)	Illumina HumanHap317/370	627	33,729
SGENE	Germany (SZ)	Illumina HumanHap317/370	1,600	1,600
SGENE	Holland (SZ)	Illumina HumanHap550	614	3,687
SGENE	Finland (SZ)	Illumina HumanHap550	580	3,273
SGENE	Denmark (SZ)	Illumina HumanHap610	765	493
SGENE	Norway (SZ)	Illumina HumanHap610	363	367
SGENE	UK (SZ)	Illumina HumanHap610	92	83
SGENE	Italy (SZ)	Illumina HumanHap610	84	86
Ripke et al.	Sweden (SZ)	Affy6.0/Illumina OmniExpress	4,719	5,918
Cooper et al.	US - NINDS (Coriell 500K)	HumanHap550v3_A		441
Cooper et al.	US - NINDS (317K +240K)	Illumina 317K+240K		227
Cooper et al.	US - PARC (CAP and PRINCE)	Illumina 550K		936
Cooper et al.	US - PARC2 (CAP2)	Illumina 550K		232
Cooper et al.	US - PARC2(PRINCE2)	Illumina610K Quad		534
Cooper et al.	US - FHCRC	Human 610v1_B		1,334
Cooper et al.	Italy - InChianti	Illumina HumanHap550v3_a		695
ARIC[b]	US	Affymetrix 6.0		11,305
Total			22,585	77,168

[a] Non-psychiatric cases from WTCCC1

[b] For the ARIC controls the arrays were obtained from dbGAP (phs000090.v1.p1). CNV calls were performed using the Affymetrix Genotyping Console 4.1, based on a minimum of 20 probes and a minimum genomic size of 10kbp. Additionally samples with an excess number of CNV calls (> 72) were filtered as potentially unreliable samples based on an outlier detection approach for skewed data (http://onlinelibrary.wiley.com/doi/10.1002/cem.1123/pdf).

TABLE S2: Details on WTCCC2 sample exclusions.

	Irish SZ Cases	Irish Controls	UK Controls (58BC)	UK Controls (NBS)
SNP call rate <0.95	6	0	4	3
<40 sample on plate[a]	67	29	47	147
>30 large CNVs[b]	13	34	1	14
Large rare events >10Mbp[c]	7	2	6	5
SNP calling[d]	302	74	276	285
Total excluded	395	139	334	454
Final included	1564	1748	2663	2533

[a] excluded from analysis as they were on a hybridization containing less than 40 samples.

[b] samples contained more than 30 large (>100kb) rare (<1% population frequency) CNVs.

[c] excluded because the total event length of large (>100kb), rare (<1% population frequency) CNVs was >10Mbp.

[d] samples excluded by previously described WTCCC2 SNP QC procedures.

TABLE S3:A:Identified carriers of known schizophrenia risk CNVs in our discovery set.B:All large (>100kb) and rare (MAF<0.01) CNVs identified in the discovery dataset.

See accompanying Excel file

Table S4: Results of qPCR validation

See accompanying Excel file with details of qPCR results at PAK7.

TABLE S5: All carriers of duplications at *PAK7* on chromosome 20 (hg18)				
Sample	**Origin**	**Local ID**	**Start**	**End**
WTCCC2 SZ cases	Ireland	IRL_101	9685413	9831947
WTCCC2 SZ cases	Ireland	IRL_201	9685413	9831947
WTCCC2 SZ cases	Ireland	IRL_301	9685413	9831947
WTCCC2 SZ cases	Ireland	IRL_401	9685413	9834500
WTCCC2 SZ cases	Ireland	IRL_501	9686531	9829109
WTCCC2 SZ cases	Ireland	IRL_601[a]	9685413	9831947
ISC SZ cases	UK_Aberdeen		9685413	9824688
ISC SZ cases	UK_Aberdeen		9685413	9829748
ISC SZ cases	UK_Aberdeen		9685413	9829748
ISC SZ cases	UK_Edinburgh	SCOT_101	9685413	9834500
ISC SZ cases	UK_Edinburgh		9685413	9834500
ISC SZ cases	UK_London		9685413	9817570
CLOZUK SZ cases	UK		9652943	9747765
CLOZUK SZ cases	UK		9682770	9924721
CLOZUK SZ cases	UK		9682770	9924721
CLOZUK SZ cases	UK		9685413	9833653
CLOZUK SZ cases	UK		9685413	9833653
CLOZUK SZ cases	UK		9685413	9833653
CLOZUK SZ cases	UK		9685413	9833653
WTCCC1BPD cases	UK_Cardiff		9684363	9824688
WTCCC1BPD cases	UK_London		9685413	9861678
WTCCC1BPD cases	UK_Aberdeen		9685413	9815136
McQuillin_et_al BPD cases	UK		9685413	9840328
WTCCC2 controls	UK_58BC		9685413	9829748
WTCCC2 controls	UK_NBS		9608359	9756235
WTCCC2 controls	UK_NBS		9685413	9834500
WTCCC2 controls	UK_POBI		9685413	9829748
WTCCC2 controls	UK_POBI		9082549	9707414
WTCCC1 non-psychiatric cases	UK_WTCCC1_RA		9685413	9820828
WTCCC1 non-psychiatric cases	UK_WTCCC1_HT		9686531	9862400
WTCCC1 non-psychiatric cases	UK_WTCCC1_T1D		9686531	9815136
WTCCC1 non-psychiatric cases	UK_WTCCC1_CD		9686531	9824688
WTCCC1 non-psychiatric cases	UK_WTCCC1_T2D		9689876	9796854
SGENE SZ cases	Denmark		9686531	9833653
SGENE SZ cases	Holland		9688441	9833653
SGENE controls	Denmark		9682770	9923954
SGENE controls	Holland		9688441	9833653
SGENE controls	Holland		9712263	9833653
Ripke et al SZ cases	Sweden		9679193	9895068
Ripke et al SZ cases	Sweden		9682770	9842874
Ripke et al controls	Sweden		9682770	9919125
Ripke et al controls	Sweden		9689876	9915394
Ripke et al controls	Sweden		9686531	9831948
Ripke et al controls	Sweden		9682770	9924722
Ripke et al controls	Sweden		9685413	9829110
Ripke et al controls	Sweden		9686531	9824689
Ripke et al controls	Sweden		9682770	9917372
European ancestry controls	US_Cooper_et_al		9545000	9930000
European ancestry controls	US_ARIC		9684363	9829109
European ancestry controls	US_ARIC		9684363	9831947
European ancestry controls	US_ARIC		9684363	9831947
European ancestry controls	US_ARIC		9685413	9831947

European ancestry controls	US_ARIC		9685413	9831947
European ancestry controls	US_Cooper_et_al		9686531	9834500
[a] IRL_601 is related to IRL_101 and thus not included in association analysis				

TABLE S6: Hap Block 1 and 2 core SNPs				
Hap Block	**SNP**	**Position**	**A1**	**A2**
1	rs2423464	9683106	G	C
1	rs742452	9684363	C	G
1	rs742451	9684473	T	C
1	rs742450	9684493	T	C
2	rs6057009	9840327	C	T
2	rs6516523	9860923	T	C
2	rs6039636	9861678	G	A
2	rs6118819	9867572	G	A

TABLE S7: Haplotype frequencies		
Hap Block	**Hap**	**Freq**
1	CCTT	0.188
1	CCCC	0.022
1	GGCC	0.301
1	CGCC	0.487
2	TCAG	0.397
2	CTGA	0.178
2	CCGA	0.089
2	CCAA	0.119
2	TCAA	0.208

TABLE S8: Phasing of haplotypes in Hap Blocks 1 and 2				
Sample	**Hap Block 1**		**Hap Block 2**	
	Hap 1	**Hap 2**	**Hap 1**	**Hap 2**
WTCCC2 SZ cases Ireland	CCTT	CCCC	TCAA	CCAA
WTCCC2 SZ cases Ireland	CCTT	CGCC	TCAA	TCAA
WTCCC2 SZ cases Ireland	CCTT	CGCC	TCAA	TCAG
WTCCC2 SZ cases Ireland	CCTT	CGCC	TCAA	TCAG
WTCCC2 SZ cases Ireland	CCTT	CGCC	TCAA	CCGA
ISC SZ cases UK Aberdeen	CCTT	CGCC	TCAA	TCAA
ISC SZ cases UK Aberdeen	CCTT	CGCC	TCAA	CCAA
ISC SZ cases UK Aberdeen	CCTT	CGCC	TCAA	CTGA
ISC SZ cases UK London	CCTT	GGCC	TCAA	TCAA
ISC SZ cases UK Edinburgh	CCTT	CGCC	TCAA	CCAA
ISC SZ cases UK Edinburgh	CCTT	GGCC	TCAA	CTGA
WTCCC1 BPD cases UK Cardiff	CCTT	CGCC	TCAA	TCAG
WTCCC1 BPD cases UK London	CCTT	CCTT	TCAA	CTGA
WTCCC1 BPD cases UK Aberdeen	CCTT	CGCC	TCAA	TCAA
McQuillin et al BPD cases UK	CCTT	CGCC	TCAA	TCAG
WTCCC2 controls NBS UK	CCTT	CCTT	TCAA	TCAA
WTCCC2 controls 58BC UK	CCTT	CGCC	TCAA	CCGA
WTCCC2 controls POBI UK	CCTT	CGCC	TCAA	CTGA
WTCCC1 RA case UK	CCTT	GGCC	TCAA	TCAA
WTCCC1 CD case UK	CCTT	GGCC	TCAA	TCAG
WTCCC1 T2D case UK	CCTT	CCTT	TCAA	CCAA
WTCCC1 HT case UK	CCTT	CCTT	TCAA	CCGA
WTCCC1 T1D case UK	CCTT	GGCC	TCAA	CTGA
EA controls ARIC US	CCTT	CGCC	TCAA	TCAG
EA controls ARIC US	CCTT	CGCC	TCAA	TCAA
EA controls ARIC US	CCTT	CCTT	TCAA	CTGA
EA controls ARIC US	CCTT	CGCC	TCAA	CCAA
EA controls ARIC US	CCTT	CGCC	TCAA	TCAG

TABLE S9 (details on clinical features of identified PAK7 duplication carriers)

See accompanying Excel file