Microduplications of 16p11.2 are Associated with Schizophrenia

Shane McCarthy, Vladimir Makarov, George Kirov, Anjene Addington, Jon McClellan, Seungtai Yoon, Dianna Perkins, Diane E. Dickel, Mary Kusenda, Olga Krastoshevsky, Verena Krause, Ravinesh A. Kumar, Detelina Grozeva, Dheeraj Malhotra, Tom Walsh, Elaine H. Zackai, Paige Kaplan, Jaya Ganesh, Ian D. Krantz, Nancy B. Spinner, Patricia Roccanova, Abhishek Bhandari, Kevin Pavon, B. Lakshmi, Anthony Leotta, Jude Kendall, Yoon-ha Lee, Vladimir Vacic, Sydney Gary, Lilia Iakoucheva, Timothy J. Crow, Susan L. Christian, Jeffrey Lieberman, Scott Stroup, Terho Lehtimäki, Kaija Puura, Chad Haldeman-Englert, Justin Pearl, Meredith Goodell, Virginia L. Willour, Pamela DeRosse, Jo Steele, Layla Kassem, Jessica Wolff, Nisha Chitkara, Francis McMahon, Anil K. Malhotra, James B. Potash, Thomas G. Schulze, Markus M. Nöthen, Sven Cichon, Marcella Rietschel, Ellen Leibenluft, Vlad Kustanovich, Clara M. Lajonchere, James S. Sutcliffe, David Skuse, Michael Gill, Louise Gallagher, Nancy R. Mendell, Wellcome Trust Case Control Consortium, Nick Craddock, Michael J. Owen, Michael C. O'Donovan, Tamim H. Shaikh, Ezra Susser, Lynn E. DeLisi, Patrick F. Sullivan, Curtis K. Deutsch, Judith Rapoport, Deborah L. Levy, Mary-Claire King, Jonathan Sebat.

Supplementary Table 1. Sample-by-Array Distribution. The institutions contributing case-control samples or data to the association analysis of 16p11.2 in schizophrenia, autism and bipolar disorder are provided with a breakdown of how many samples from each institution were analyzed by one of the five microarray platforms (ROMA 85K, NimbleGen HD2, Affymetrix 500K, Affymetrix 5.0 and Affymetrix 6.0). In addition we summarize the patient and control ascertainment schemes used by studies that contributed to the primary and replication samples. Cases that were drawn from larger pedigrees are referred to as “Family-based”. “Early onset Schizophrenia” was defined as having an age at onset younger than 13. References provided for each set sample set describe the ascertainment of these samples in greater detail.

Supplementary Table 1.

Sample Source	Total	ROMA 85K	NimbleGenHD2	Affymetrix 500K	Affymetrix 5.0	Affymetrix 6.0	Ascertainment (Reference)
Primary Cases							
Cardiff University	471	0	0	471	0	0	(3)
CATIE	738	0	0	738	0	0	(2, 14)
Columbia University	19	19	0	0	0	0	Family based (8)
McLean Hospital	161	9	152	0	0	0	Family based (7)
NIMH	83	0	0	83	0	0	Early Onset (1)
New York University	259	22	237	0	0	0	Family based (9)
University of Washington	175	51	124	0	0	0	In Patient or Early onset (1)
Total	**1906**	**101**	**513**	**1292**	**0**	**0**	
Primary Controls							
CATIE	289	0	0	289	0	0	Negative History (2, 14)
CSHL	29	29	0	0	0	0	Not Screened (11)
HapMap	117	0	117	0	0	0	Not Screened (4,5)
NINDS	262	262	0	0	0	0	Negative History (6)
NIMH	96	0	96	0	0	0	Negative History (11)
NYCP	386	20	366	0	0	0	Not Screened (10)
WTCCC	2792	0	0	2792	0	0	Not Screened (12)
Total	**3971**	**311**	**579**	**3081**	**0**	**0**	
Replication Cases							
GAIN (phs000021.v2.p1)	**2645**	**0**	**0**	**0**	**0**	**2645**	Multisite Recruitment (15)
Replication Controls							
GAIN (phs000021.v2.p1)	**2420**	**0**	**0**	**0**	**0**	**2420**	Absence of History (15)
Bipolar Disorder Cases							
GEM	161	0	161	0	0	0	Multisite Recruitment (see section 1)
WTCCC	1697	0	0	1697	0	0	Multisite Recruitment (12)
GAIN (phs000017.v2.p1)	1457	0	0	0	0	1457	Multisite Recruitment (19,20)
Total	**3315**	**0**	**161**	**1697**	**0**	**1457**	
Autism Cases							
AGRE	692	9	0	0	683	0	(16,17)
Columbia University	32	32	0	0	0	0	(17)
Inst.Child Health	28	28	0	0	0	0	(17)
NIMHAU	49	10	39	0	0	0	(17)
N. Shore. Uni. Hospital	10	9	1	0	0	0	(17)
Trinity College Dublin	60	60	0	0	0	0	(18)
Vanderbilt University	63	63	0	0	0	0	(17)
Total	**934**	**211**	**40**	**0**	**683**	**0**	

Supplementary Table 2. HMM identified 16p11.2 Microduplications and Microdeletions. The 16p11.2 microduplications and microdeletions detected by HMM segmentation are summarized in detail by sample, dataset and microarray platform (Supplementary Note). The genome coordinates of each event are given for the discovery platform and coordinates are also given for the results of the fine mapping using the Nimblegen HD2 platform. Coordinates are based on the UCSC human genome version 18 (hg18). We also report the percentage of the 16p11.2 region (chr16:29557498-30107355) that was overlapped by each segment. The number of MeZOD genotyping probes and MeZOD Z-Score for each segment is also given. (Supplementary Note)

[a] Data from Cardiff University, NIMH and WTCCC were not analyzed by MeZOD. This table lists only CNVs detected in our analysis of data on the ROMA, NimbleGen HD2, Affymetrix 500K, 5.0 and 6.0 platforms. Not included here are CNVs reported in published studies [16,22,23] where raw data were not publicly available.

[b] Events detected in WTCCC bipolar cases and controls were detected independently on the NSP and STY arrays; thus CNVs were detected and validated on independent microarrays. Fine mapping of these samples using the Nimblegen HD2 platform was not possible due to insufficient DNA.

[c] Insufficient DNA was available to validate these 16p11.2 rearrangements in autism (one duplication and two deletions) detected by ROMA 85K arrays.

Supplementary Table 2

Sample	Data Set	Platform	Gender	Ethnicity	16p11.2 Genotype	Median Probe Ratio	hg18 Start	hg18 End	CNV-Target Overlap	Number of Genotyping Probes	MeZOD Z-Score	Valdiation Platform	hg18 Start	hg18 End
03C18520	Primary Case	Affymetrix 500K	Male	EA	Del	0.69	29559989	29926001	67%	19	-3.03	NimbleGen HD2	29564890	30132589
03C15581	Primary Case	Affymetrix 500K	Male	EA	Dup	1.24	29559989	30085308	96%	20	1.70	NimbleGen HD2	29534144	30098184
03C15536	Primary Case	Affymetrix 500K	Female	AA	Dup	1.25	29559989	30085308	96%	20	1.25	NimbleGen HD2	29280728	30287854
03C15896	Primary Case	Affymetrix 500K	Male	EA	Dup	1.17	29580704	30085308	92%	20	1.18	NimbleGen HD2	29295104	30287854
AV-27-05	Primary Case	ROMA	Female	EA	Dup	1.31	29686691	30037311	64%	6	2.37	NimbleGen HD2	29345816	30252311
OX-100-01	Primary Case	NimbleGen HD2	Male	EA	Dup	1.33	29281696	30207681	100%	177	1.80	Agilent 244K		
MC235	Primary Case	NimbleGen HD2	Male	EA	Dup	1.33	29481419	30252311	100%	177	1.24	Agilent 244K		
NWP-110-02	Primary Case	NimbleGen HD2	Male	EA	Dup	1.34	29553777	30105580	100%	177	2.00	Agilent 244K		
676	Primary Case	Affymetrix 500K[a]	Female	EA	Dup	1.24	29652656	30085308	79%	17	NA	NimbleGen HD2	29557094	30107502
2011	Primary Case	Affymetrix 500K[a]	Female	EA	Dup	1.20	29657405	30235818	82%	16	NA	NimbleGen HD2	29534144	30098184
19326A3	Primary Case	Affymetrix 500K[a]	Male	EA	Dup	1.26	29559989	30227808	100%	20	NA	NimbleGen HD2	29325222	30133525
19328A3	Primary Case	Affymetrix 500K[a]	Female	EA	Dup	1.26	29559989	30085308	96%	20	NA	NimbleGen HD2	28548146	30309077
19328G5	Primary Case	Affymetrix 500K[a]	Male	EA	Dup	1.19	29559989	29926001	67%	19	NA	NimbleGen HD2	29512728	30105580
11158E8	Primary Control	Affymetrix 500K[a]	Female	EA	Dup	1.27	29559989	29926001	67%	19	NA	Affymetrix 500K[b]		
129998A	Primary Control	Affymetrix 500K[a]	Male	EA	Del	0.74	29559989	30227808	100%	20	NA	Affymetrix 500K[b]		
15051D6	Primary Control	Affymetrix 500K[a]	Male	EA	Del	0.70	29559989	30227808	100%	20	NA	Affymetrix 500K[b]		
15387D8	Primary Control	Affymetrix 500K[a]	Male	EA	Del	0.68	29559989	30227808	100%	20	NA	Affymetrix 500K[b]		
37540	Replication Case	Affymetrix 6.0	Male	EA	Dup	1.23	29158416	30038055	87%	137	1.68	NimbleGen HD2	29156718	30217554
40245	Replication Case	Affymetrix 6.0	Female	AA	Dup	1.23	29425200	30099396	99%	163	1.58	NimbleGen HD2	29470406	30105580
37612	Replication Case	Affymetrix 6.0	Female	EA	Dup	1.24	29474798	30214457	100%	163	1.52	NimbleGen HD2	29351484	30252311
851	Replication Case	Affymetrix 6.0	Female	EA	Dup	1.26	29487523	30099396	99%	163	1.46	NimbleGen HD2	29528158	30111196
40832	Replication Case	Affymetrix 6.0	Female	AA	Dup	1.23	29487523	30084010	96%	159	1.59	NimbleGen HD2	29553777	30105580
1669	Replication Case	Affymetrix 6.0	Male	EA	Dup	1.25	29488112	30099396	99%	163	1.35	NimbleGen HD2	29351484	30256447
40350	Replication Case	Affymetrix 6.0	Male	AA	Dup	1.26	29488603	30058600	91%	148	1.42	NimbleGen HD2	29204284	30275324
38499	Replication Case	Affymetrix 6.0	Female	EA	Dup	1.25	29498829	30099396	99%	163	1.71	NimbleGen HD2	29459174	30119610
38492	Replication Case	Affymetrix 6.0	Male	EA	Dup	1.26	29498891	30097531	98%	162	1.37	NimbleGen HD2	29521802	30170962
36604	Replication Control	Affymetrix6.0	Female	EA	Del	0.72	29309802	30085308	96%	161	-2.78	NimbleGen HD2	29293200	30133525
39636	Replication Control	Affymetrix6.0	Male	AA	Dup	1.26	29425200	30099396	99%	163	1.59	NimbleGen HD2	29537669	30107502
2201-0	Bipolar Disorder	NimbleGen HD2	Male	EA	Dup	1.34	29453358	30252311	100%	177	1.68	Agilent 244K		
F440_4	Bipolar Disorder	Affymetrix 500K[a]	Male	EA	Dup	1.17	29559989	29926001	67%	19	NA	Affymetrix 500K[b]		
8026_1	Bipolar Disorder	Affymetrix 500K[a]	Female	EA	Dup	1.26	29559989	30085308	96%	20	NA	Affymetrix 500K[b]		
6023_9	Bipolar Disorder	Affymetrix 500K[a]	Female	EA	Dup	1.23	29559989	30085308	96%	20	NA	Affymetrix 500K[b]		
JS-6059-900	Autism	ROMA	Female		Del	0.71	27000632	45495496	100%	6	-1.87	NA[C]		
JS-6056-900	Autism	ROMA	Male		Del	0.74	27051731	45671669	100%	6	-2.29	NA[C]		
CG20261	Autism	ROMA	Female	EA	Del	0.67	29560786	30037311	87%	6	-2.98	NimbleGen HD2	29564890	30102245
AU0938301	Autism	Affymetrix 5.0	Male		Del	0.72	29530260	30011863	83%	51	-2.27	NimbleGen HD2	29367869	30226378
AU041905	Autism	Affymetrix 5.0	Male		Del	0.72	29530260	30214457	100%	63	-2.33	NimbleGen HD2	29511515	30119610
AU0154302	Autism	Affymetrix 5.0	Male		Del	0.73	29530260	30038195	87%	56	-2.55	NimbleGen HD2	29561239	30124017
AU029803	Autism	Affymetrix 5.0	Male	EA	Del	0.72	29534002	30035231	87%	55	-2.64	NimbleGen HD2	29367869	30132589
JS-2100-3	Autism	ROMA	Male		Dup	1.34	29560786	30037311	87%	6	2.65	NA[C]		
AU011004	Autism	Affymetrix 5.0	Male	EA	Dup	1.22	29530260	30097531	98%	62	1.28	NimbleGen HD2	29459174	30107502
AU002903	Autism	Affymetrix 5.0	Male	EA	Dup	1.21	29559011	30055840	90%	57	1.36	NimbleGen HD2	29497757	30105580
AU032704	Autism	Affymetrix 5.0	Male	EA	Dup	1.16	29606126	30007823	73%	45	1.23	NimbleGen HD2	29537669	30105580

Supplementary Table 3. Studies Reporting 16p11.2 Variants used in the Meta-analysis of 16p11.2 Variants and Psychiatric Disease. Listed for each source (i.e. study) are the number of individuals of each disease type and the number of controls that were included in the meta-analysis. Efforts were made to eliminate sample overlap between studies. For the published studies by the ISC [22] and Marshall et al. [23], no cases or controls overlapped with the samples in our primary or replication samples, and thus the numbers of each corresponded exactly to what is described in the methods of the respective study. From the study by Weiss et al, we included only data from Children's Hospital Boston, deCode and STEP-BD. We did not include any data from AGRE or NIMH controls that overlapped with sources used in our study. Last, because of the removal of the corresponding NIMH controls, the STEP-BD patients were combined with the bipolar sample from deCode for the purposes of the meta-analysis.

Study	Disease	Deletion	Duplication	Normal	Total
Current Study	Schizophrenia	1	12	1893	1906
	Bipolar Disorder	0	4	1854	1858
	Global Deveopmental Delay or Autism	7	4	923	934
	Controls	3	1	3967	3971
GAIN	Schizophrenia	0	9	2636	2645
	Bipolar Disorder	0	0	1457	1457
	Controls	1	1	2418	2420
Weiss et al	Schizophrenia (deCode)	1	0	647	648
	Global Deveopmental Delay or Autism (deCode)	3	0	296	299
	Bipolar Disorder (STEP-BD and deCode)	4	2	1501	1507
	Controls (deCode)	2	5	18827	18834
Weiss et al (CHB)	Global Deveopmental Delay or Autism (CHB)	5	4	503	512
	Controls (CHB)	0	0	434	434
Marshall et al	Global Deveopmental Delay or Autism	2	2	423	427
	Controls	0	0	1652	1652
ISC	Schizophrenia	1	5	3385	3391
	Controls	3	1	3177	3181

Supplementary Table 4. Quantitative Clinical Data for 16p11.2 Microduplication and Microdeletion Carriers. Presented here are the available data on height, weight, occipital-frontal-circumference (OFC) or non-verbal IQ on patients with microduplications or microdeletions of 16p11.2. Data were obtained from patients identified in this study and in 3 previous publications [16,27,28]. Also included are data on three microdeletion carriers referred for psychiatric evaluation at the Children's Hospital in Philadelphia (CHOP). Height and OFC data were collected in centimeters (cm) and weight data werecollected in kilograms (kg). Standardized Z-scores were calculated as described in the Supplementary Note. Height and head circumference measurements were taken at the same age for each subject with one exception CG20261, whose height and OFC were collected at ages 25 and 5 years respectively. The reported diagnoses of individual patients are listed, which included schizophrenia (SCZ), developmental delay (DD), mental retardation (MR), autism spectrum disorder (ASD) and attention-deficit disorder (ADD), as well as atypical autism and dysmorphism. Diagnoses for all cases from Children's Hospital Boston (See reference 16) were classified as "developmental delay, mental retardation or autism" (which we refer to as "DD/ASD").

[a] CHOP: Children's Hospital of Philadelphia; CATIE: Clinical Antipsychotic Trials of Intervention Effectiveness; AGRE: Autism Genetics Resource Exchange; UW: University of Washington; NIMH: National Institute of Mental Health; MCHG: Marshfield Clinical Human Genetics, WI; CHB : Children's Hospital Boston

[b] European Ancestry (EA); Ancestry Unknown (-)

c Autism Spectrum Disorder (ASD): Global Developmental Delay (DD); Schizophrenia (SCZ); Mental Retardation (MR); No diagnosis (-).

Study	Patient	Source[a]	16p11.2. Genotype	Gender	Ethnicity[b]	Primary Diagnosis[c]	Age at Examination	WISC Performance IQ	Ravens IQ	Estimated IQ	Weight (kg)	Height (cm)	Height Z-score (CDC)	Height Z-score (Farkas)	OFC (cm)	OFC Z-score (CDC)	OFC Z-Score (Farkas)
Referal	CHOP 1	CHOP	Deletion	Female	EA	DD	3.25				12.40	86.30	-2.33	-2.59	50	0.74	-0.17
Referal	CHOP 3	CHOP	Deletion	Male	EA	DD	19								58.5	2.40	0.42
Referal	CHOP 4	CHOP	Deletion	Male	EA	DD	14.75				74.40	178.70	1.23	2.17	58	2.05	1.99
Current	03C15581	CATIE	Duplication	Male	EA	SCZ	62				99.79	185.42	1.04	1.09			
Current	03C15896	CATIE	Duplication	Male	EA	SCZ	44				58.06	182.88	0.71	0.78			
Current	03C18520	CATIE	Deletion	Male	EA	SCZ	23				96.16	175.26	-0.25	-0.17			
Current	AU002903	AGRE	Duplication	Male	EA	ASD	15.46		90		69.40	179.07	0.92	0.86	54.8	-0.13	-0.99
Current	AU002904	AGRE	Duplication	Female	EA	-	13.36		94		49.44	156.21	-0.47	-0.13	53.3	-0.33	-0.35
Current	AU002905	AGRE	Duplication	Female	EA	ASD	11.94				39.46	153.67	0.08	0.97	53.4	0.31	0.26
Current	AU011004	AGRE	Duplication	Male	EA	ASD	11.04		110						53.5	0.20	-0.14
Current	AU032704	AGRE	Duplication	Male	EA	ASD			75								
Current	AU032705	AGRE	Duplication	Male	EA	ASD			50								
Current	AU032706	AGRE	Duplication	Male	EA	ASD			107								
Current	AU032707	AGRE	Duplication	Male	EA	-			107								
Current	AU041905	AGRE	Deletion	Male	EA	ASD	7.96		108		24.49	124.46	-0.61	-0.02	54	1.28	1.32
Current	AU0938301	AGRE	Deletion	Male	EA	ASD	14								56.5	1.23	1.06
Current	CG20261	Columbia	Deletion	Female	EA	ASD	5					157.48	-0.95	-0.76	51.5	0.84	-0.18
Current	MC235	UW	Duplication	Male	EA	SCZ				80-90							
Current	Rap-2011	NIMH	Duplication	Female	EA	SCZ	15	81			48.10	152.40	-1.44	-1.44	54	-0.18	-0.13
Current	Rap-676	NIMH	Duplication	Female	EA	SCZ	13	72			34.50	157.50	0.00	0.02	52.5	-0.88	-0.94
Ghebranious et al	Twin1	MCHG	Deletion	Male	-	DD	28								59.5	3.10	1.11
Ghebranious et al	Twin2	MCHG	Deletion	Male	-	DD	28								57.3	1.55	-0.42
Weiss et al	CHBDel1	CHB	Deletion	Male	-	DD/ASD	6y6m				19.00	108.10	-1.88	-1.65	51	-0.55	-0.60
Weiss et al	CHBDel2	CHB	Deletion	Male	-	DD/ASD	2y9m								52	1.64	1.32
Weiss et al	CHBDel3	CHB	Deletion	Male	-	DD/ASD	17m				9.00	74.20	-2.33	-1.42	48.5	0.67	-0.53
Weiss et al	CHBDel4	CHB	Deletion	Male	-	DD/ASD	9y2m				65.40	143.30	0.67	0.89	54.5	1.41	1.17
Weiss et al	CHBDel5	CHB	Deletion	Male	-	DD/ASD	9y2m				71.90	150.00	2.33	1.73	56	2.50	2.25
Weiss et al	CHBDup1	CHB	Duplication	Male	-	DD/ASD	14m				11.00					0.84	0.84
Weiss et al	CHBDup2	CHB	Duplication	Female	-	DD/ASD	3y3m				14.70	97.20	0.25	-0.45	46.5	-1.48	-2.91
Weiss et al	CHBDup4	CHB	Duplication	Female	-	DD/ASD	9y9m				33.50				51	-0.88	-0.90
Weiss et al	ICEDel1	Iceland	Deletion	Female	EA	ASD/MR	5y2m				18.00	106.00	-0.67	-1.87	52.5	1.48	1.71
Weiss et al	ICEDel2	Iceland	Deletion	Male	EA	Atypical Autism	10y6m				69.00	152.00	1.55	1.76	54	0.71	0.38
Bijlsma	Case 1	-	Deletion	Male	-	DD	44				84.9	172	-0.67	-0.57			
Bijlsma	Case 2	-	Deletion	Male	-	DD	17yr 2m				130.5	180	0.67	0.52	60	3.4	3.81
Bijlsma	Case 3	-	Deletion	Female	-	DD	8yr 2m				45	130	0.18	0.17	54.5	2.05	1.90
Bijlsma	Case 4	-	Deletion	Male	-	ADD	13yr							2.00			
Bijlsma	Case 5	-	Deletion	Male	-	ASD	3yr							0.00		-0.80	-0.80
Bijlsma	Case 6	-	Deletion	Female	-	DD	7yr				25	122	0.05	-0.28	52.5	0.88	0.67
Bijlsma	Case 7	-	Deletion	Male	-	DD	18m				8.2	72	-2.33	-1.85	47.5	-0.20	-1.43
Bijlsma	Case 8	-	Deletion	Female	-	DD	11yr				42.8	146	0.13	-0.04			
Bijlsma	Case 10	-	Deletion	Female	-	DD	8yr				24	128	0.05	-0.14	53	1.04	0.85
Bijlsma	Case 11	-	Deletion	Male	-	DD	4yr				15	101.5	-0.39	-1.67	51.5	0.74	-0.23
Bijlsma	Case 12	-	Deletion	Female	-	Dysmorphism	11m				8.4	77.5	1.75	1.79	45.5	0.52	0.21
Bijlsma	Case 13	-	Deletion	Male	-	DD	4yr 6m				21	112	1.41	0.47			

Supplementary Table 5. Rearrangements of 16p11.2 are Significantly Associated with Head Circumference. Measurements of orbital frontal circumference (OFC) were available for 23 patients with microdeletions and 9 patients with microduplications. OFC was adjusted for age and gender. Listed here are the mean OFC and standard error for the microdeletion and microduplication groups respectively and the increase in mean OFC of the deletion group compared to that of the duplication group (ΔOFC). Significance was tested using the Wilcoxon two-sample Rank Sum Test. Analyses were performed on all patients, on patients with developmental delay or autism (DD/ASD), and separately in the subgroups of patients with primary diagnoses of autism (ASD) and developmental delay (DD). All cases from one center (Children's Hospital Boston) were listed under the broad category, "developmental delay, mental retardation or autism", and these were assigned to the DD subgroup.

		Mean OFC			
Case group	**Del:Dup**	**Deletion**	**Duplication**	**ΔOFC**	**P-value**
All subjects	23:9	1.25±0.22	-0.28±0.24	1.53	0.0007
DD/ASD	22:6	1.25±0.23	-0.19±0.35	1.44	0.0056
DD	16:3	1.46±0.28	-0.50±0.69	1.96	0.038
ASD	6:3	0.78±0.34	0.13±0.13	0.65	0.167

Supplementary Table 6. Description and Function of Genes within the 16p11.2 Critical Region. This table describes the function for each gene within the 16p11.2 critical region (Supplementary Figure 1) as recorded in the UCSC genome browser (hg18). The expression of the gene in either the murine or human brain is also given based on the GeneBrainAtlas or publication else where. Diseases associated with each gene are given as well as any known mouse models associated diseases with the gene, and references to the literature are provided.[29-39]

Supplementary Table 6.

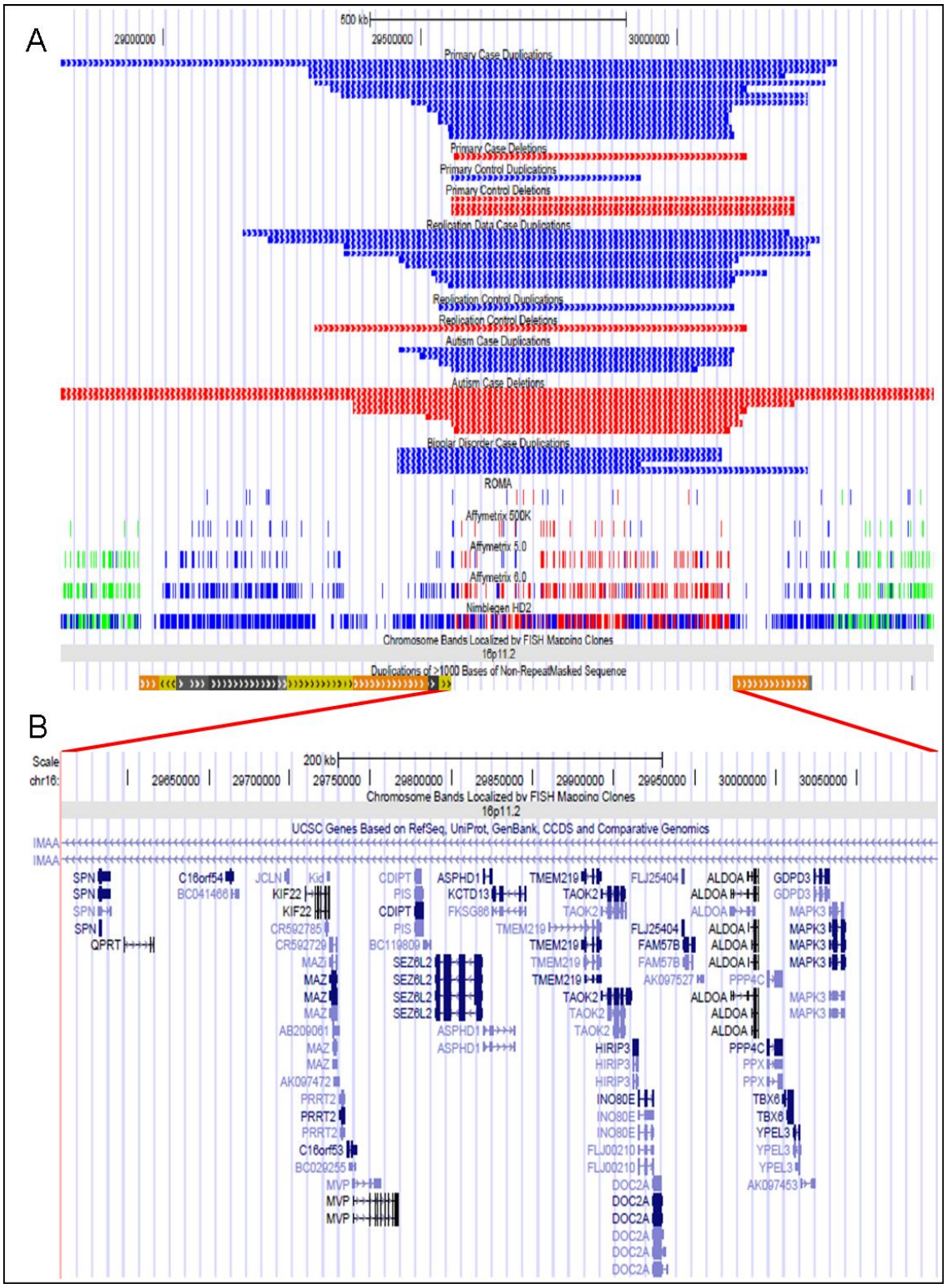
GeneName	Description	Known/Predicted Function	Expressed in Human/Mouse Brain	Associated Disease (OMIM)	Mouse Model	Model Phenotype
ALDOA	fructose-bisphosphate aldolase A	catalyzes the reversible conversion of fructose-1,6-bisphosphate to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate	Y	aldoA deficiency or red cell aldolase deficiency		
ASPHD1	aspartate beta-hydroxylase domain containing 1	catalyses oxidative reactions in a range of metabolic processes	Y			
C16orf53	PAXIP1-associated protein 1	component of a Set1-like multiprotein histone methyltransferase complex	Y			
C16orf54	Transmembrane protein C16orf54.	hypothetical protein LOC283897	Y			
CDIPT	CDP-diacylglycerol--inositol	in the biosynthesis of phosphatidylinosito	Y (29)			
CORO1A						
DOC2A	double C2-like domains, alpha	May be involved in calcium dependent neurotransmitter release through the interaction with UNC13A. May be involved in dynein-dependent intracellular vesicle transport. In vitro, binds calcium and phospholipids	Y		doc2a-/-	abnormal Passive Avoidance behavior/reduced long term potentiation (30)
FAM57B	hypothetical protein LOC83	intergral membrane protein	Y			
FLJ25404	hypothetical protein LOC146378 isoform 2	Uncharacterized protein FLJ25404.				
GDPD3	glycerophosphodiester phosphodiesterase domain	glycerol metabolic process	N			
HIRIP3	HIRA interacting protein 3	May play a role in chromatin function and histone metabolism via its interaction with HIRA and histones	Y			
IN080E	INO80 complex subunit E		Y			
KCTD13	potassium channel tetramerisation domain	voltage-gated potassium channel activity	Y	Upregulated in 15q11-q13 duplication autism cell lines (31)		
KIF22	kinesin family member 22	spindle formation and the movements of chromosomes during mitosis and meiosis			kif22-/-	abnormal development/premature death
MAPK3	mitogen-activated protein kinase 3 isoform 1					Hyperactive/abnormal active and passive avoidance behavior/reducted long term potentiation (32)
MAZ	MYC-associated zinc finger protein isoform 2	transcription factor, may be important for neuronal differentiation	Y	Expression Altered in lymphocytes in Schizophrenia (33)		
MVP	major vault protein	Required for normal vault structures may act as scaffolds for proteins involved in signal transduction and nucleo-cytoplasmic transport	Y (pyramidal neurons, especially axons and principle dendrites) (34)	upregulated in refractory epilepsy (35)	mvp-/-	increased mortality after P. aeruginosa infection (36)
PPP4C	Serine/threonine-protein phosphatase 4 catalytic subunit	Protein phosphatase that regulates many processes such as microtubule organization or maturation of spliceosomal snRNPs			ppp4c	leathal
PRRT2	proline-rich transmembrane protein 2	Multi-pass membrane protein	Y			
QPRT	quinolinate phosphoribosyltransferase	key enzyme in catabolism of quinolinate,	Y	Elevation of quinolinate levels in the brain has been linked to the pathogenesis of neurodegenerative disorders such as epilepsy, Alzheimer's disease, and Huntington's disease		
SEZ6L2	seizure related 6 homolog (mouse)-like 2 isoform		Y	Seizures (37)	sez6l2-/-	motor discoordination/Purkin+G19je cell dendritic innervation by climbing fibers is abnormal (38)
SPN	sialophorin	One of the major glycoproteins of thymocytes and T lymphocytes		Wiskott-Aldrich syndrome (301000)	spn-/-	Immune and hematopoietic System, decrease tumor incidence
TAOK2	TAO kinase 2 isoform 1	activates the JNK MAP kinase pathway	Y			
TBX6	T-box 6	involved in paraxial mesoderm formation and somitogenesis	Y		tbx6-/-	kinked neuronal tube (39)
TMEM219	transmembrane protein 219	Multi-pass membrane protein				
YPEL3	yippee-like 3		Y		Y	?

Supplementary Table 8. Summary of Psychiatric Symptoms in 16p11.2 Microduplication Carriers. The psychiatric symptoms for each 16p11.2 microduplication carrier were ascertained from medical records and supplementary clinical data as described in Supplementary Note.

ID	Approx. Age of Onset or 1st Admisson	Gender	Diagnosis	Delusions	Thought Disorder	Hallucinations	Disorganized or Catatonic Behaviour	Grandiose or Manic symptoms	Negative Symptoms	Family History
OX-100-1	**21**	M	SZ	+						
NWP-110-02	**32**	M	SZ	+	+	+	+		+	+
AV-27-05	**28**	F	SA							+
19326A3	**25**	M	SZ				+	+		+
19328A3	**22**	F	SZ	+	+		+			
19328G5	**19**	M	SZ	+					+	-
03C15581		M	SZ	+	+	+		+	+	
03C15536	**27**	F	SZ	+	+	+		+	+	
03C15896	**29**	M	SZ	+	+	+		+	+	+
MC235	**21**	M	SZ	+	+	+	+	+	+	+
676	**10**	F	SZ	+	+	+	+	+	+	+
2011	**8**	F	SZ	+	+	+	+	-	-	+
38499	**15**	F	SA	+	+	+			+	
37540	**16**	M	SA	+	+	+				
37612	**21**	F	SZ	+	+	+	+		+	
851	**13**	F	SA	+	+	+	+		+	+
38492	**40**	M	SZ	+	+	+			+	
1669	**16**	M	SZ	+		+	+		+	
40832	**17**	F	SZ	+	+	+	+		+	+
40245	**19**	F	SZ	+	+	+	+		+	
40350	**21**	M	SZ	+	+	+	+		+	
2201	**21**	M	BD	+	+	-		+		
6023_9	17	F	BD	-	-	-	-	+	-	+
F400_4	16	M	BD	-	-	-	-	+	-	-
8026_1	27	F	BD	+	+	+	+	+	-	-

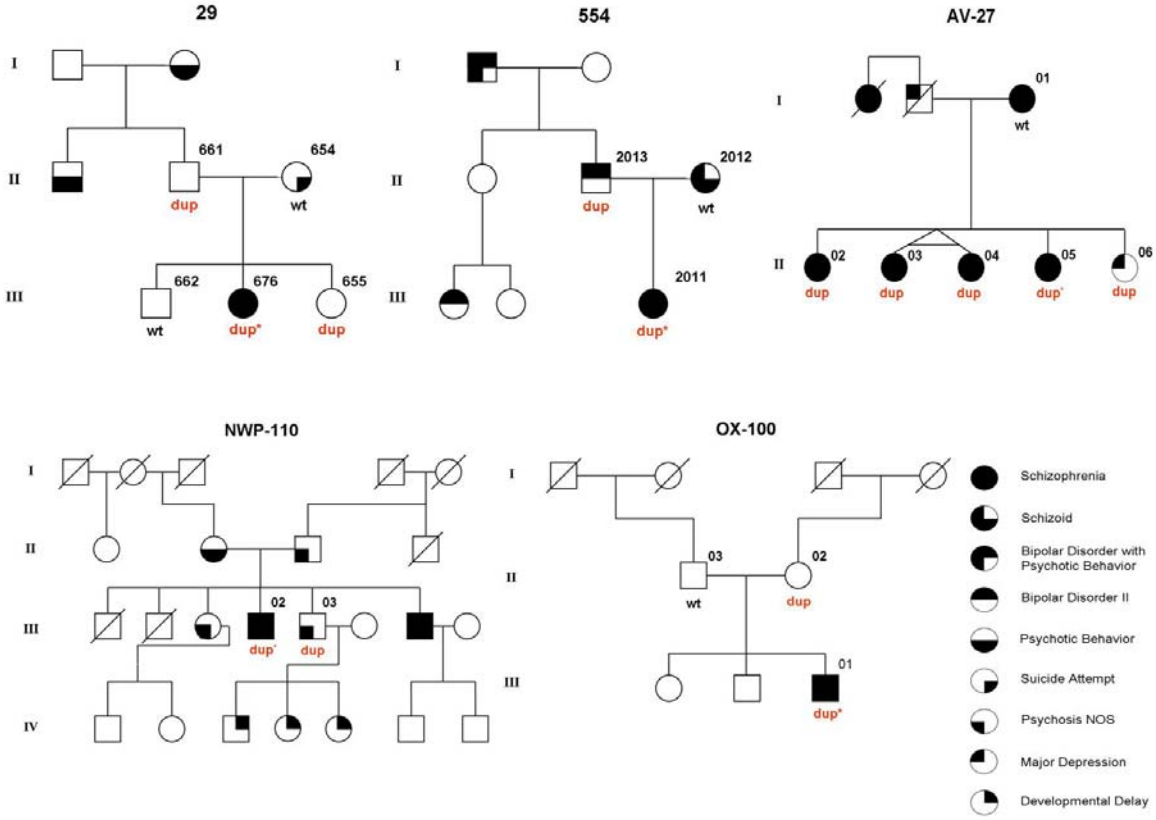
Supplementary Figure 1: HMM detection of 16p11.2 CNVs in Autism Spectrum Disorder, Bipolar Disorder and Controls. Rearrangements of 16p11.2 were detected by HMM segmentation in 4551 SCZ, 934 ASD and 3315 bipolar disorder cases, and in 6391 controls (Panel A) Not listed here are CNVs reported in published studies [16,22,23] where raw data were not publicly available. Probe coverage of the target region on each platform is displayed as 5 additional tracks. Probes highlighted in red were used for genotyping the target region. Probes highlighted in green were used to determine copy number of the flanking invariant regions. The accompanying UCSC bed file contains all the probes used for genotyping on all microarray platforms. Panel B displays the genes within the 16p11.2 critical region flanked by segmental duplications.

Supplementary Figure 1: HMM detection of 16p11.2 CNVs in Autism Spectrum Disorder, Bipolar Disorder and Controls.

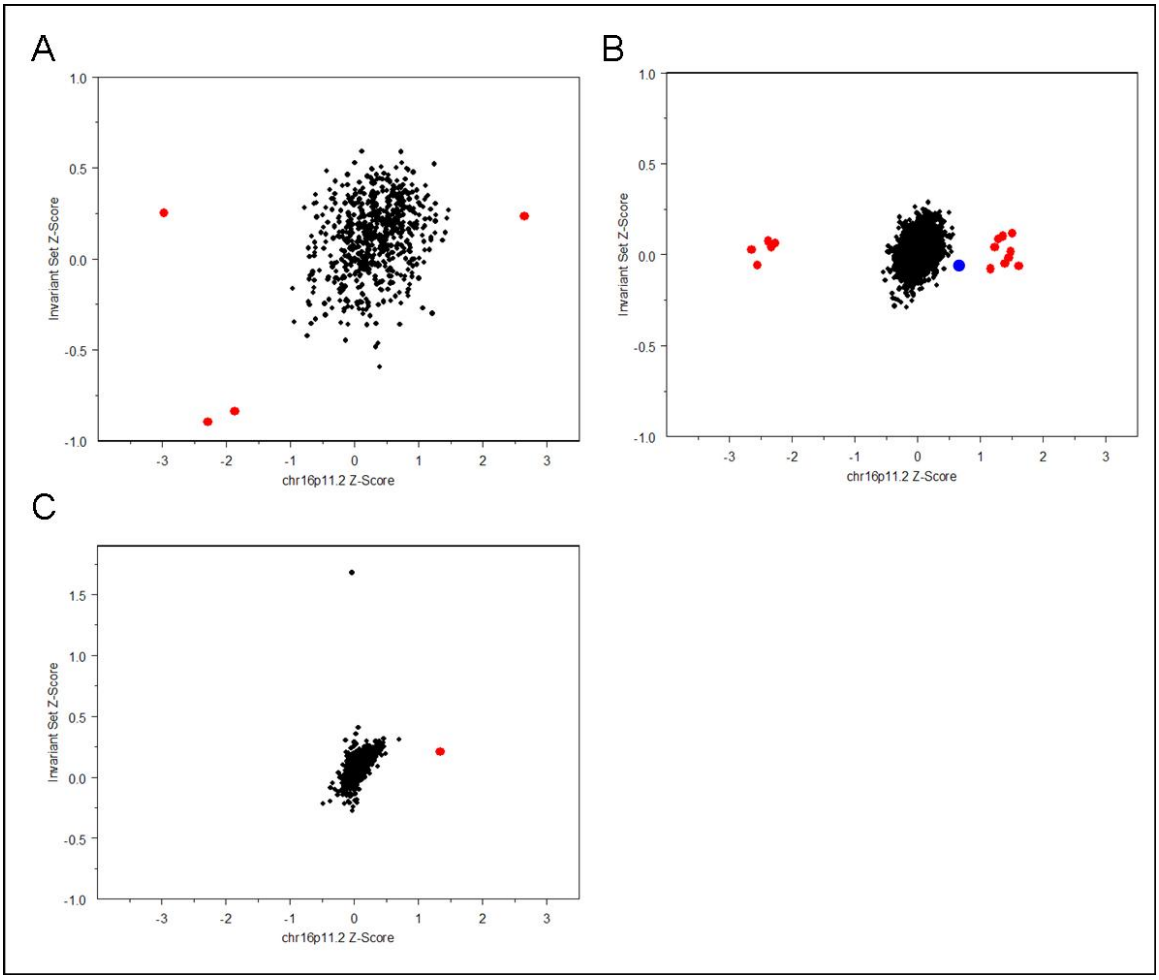


Supplementary Figure 2: Variability in Psychiatric Phenotype among Carriers of the 16p11.2 Microduplication in Families. Of the 12 microduplication carriers who were identified in our primary sample, five were derived from families where genomic DNAs were available on additional relatives. Mutation status was determined for all relatives by microarray CGH (Nimblegen HD2 platform). Affected status is shown in the key. Mutation status is indicated for subjects where genomic DNA was available. In three families, the microduplications were inherited from a non-schizophrenic parent. In the remaining 2 families, the microduplication was detected in one or more siblings of the proband, suggesting that that these microduplications were also in the parental germline. In addition to the 5 probands, 9 additional microduplication carriers were identified. Of these 9 additional carriers, the diagnoses were as follows: schizophrenia (n=3), bipolar disorder II (n=1), psychosis not otherwise specified (NOS) (n=1), major depressive disorder (N=1), and no mental illness (n=3).

Supplementary Figure 2: Variability in Psychiatric Phenotype among Carriers of the 16p11.2 Microduplication in Families.



Supplementary Figure 3: MeZOD detection of 16p11.2 CNVs in Autism Spectrum Disorder, Bipolar Disorder and Controls Results of MeZOD genotyping of the ROMA, Affymetrix 5.0 and NimbleGen HD2 data are shown in panels A, B, and C respectively. The X-axis represents the median Z-score of the target region and the Y-axis represents the median Z-score of invariant probes outside of the target region and segmental duplications. CNVs detected by HMM and MeZOD are shown in red (●). CNVs detected only by HMM are shown in blue (●).



Supplementary Note

Data Collection

a. Primary Data Set

The Primary Data Set for this study was composed of 5877 independent case and control samples analyzed on one of three microarray platforms (ROMA, NimbleGen HD2, Affymetrix500K). The primary sample consisted of 1906 patients (1329 male, 577 female) and 3971 controls (1966 male, 2005 female) collected at twelve sites. Ascertainment of these samples is described in previous publications [1-14]

Many of the same NIMH controls were analyzed using multiple platforms, and in each case we eliminated the overlapping samples from the dataset by including only the experiment with the greatest probe density. A Sample-by-Array breakdown of the final data set is provided in Supplementary Table 1. Approximately 74% of the Primary sample set was analyzed on the Affymetrix 500K microarray platform accounting for 67% and 78% of cases and controls respectively. NimbleGen HD2 data constituted 26.9% of cases and 14.6% of controls while 5.3% of cases were assayed by ROMA versus 7.8% of controls.

Information documenting the microarray hybridization methods for the ROMA [12] and Affymetrix 500K [1-3,13] platforms has been described previously. Microarray Hybridizations using the NimbleGen HD2 platform were performed at the service laboratory of Roche NimbleGen according to the manufacturer's specifications.

b. Replication Data Set

We used the Genetic Association Information Network (GAIN) supported case-control "Genome-Wide Association Study of Schizophrenia" (phs000021.v2.p1) as the replication data set in our analysis (Supplementary Table 1). This data set consisted of 2645 cases (1749 male, 896 female) and 2420 controls (1038 male, 1382 female) including individuals with European American (1404 cases, 1442 controls) and African American ancestry (1241 cases, 978 controls) genotyped on Affymetrix 6.0. Genotyping of these 5065 samples was performed at the Broad Institute Center for Genotyping and Analysis (see URLs). Further details concerning the collection sites, principal investigators and case-control inclusion criteria are available through dbGap (see URLs): and in a previous publication[15].

c. Autism Spectrum Disorder (ASD) Data Set

The ASD dataset was composed of 934 independent cases ascertained by public and private sample collections for analysis on one of three microarray platforms (ROMA 85K, Affymetrix 5.0, NimbleGen HD2) (Supplementary Table 1). Six hundred and ninety three cases were unrelated subjects from the Autism Genetic Resource Exchange (AGRE), 683 of which were analyzed on Affymetrix 5.0 and ten of which were analyzed on NimbleGen HD2 and ROMA arrays. Ascertainment of AGRE Affymetrix 5.0 data has been described previously [16] while the NimbleGen HD2 and ROMA experiments were performed at CSHL. None of the 49 NIMH Autism Genetics Initiative (NIMHAU) samples analyzed on ROMA and NimbleGen HD2 overlapped with the AGRE data set. Documentation of the NIMH autism sample collection and contributors is described elsewhere

(see URLs). Ascertainment of the remaining ASD samples has been published previously [17]. Samples from Trinity College Dublin have been described in previous publications [18].

In total Affymetrix 5.0 data composed 73% of the ASD data collection while 23% and 4% of the remaining samples were represented by ROMA and NimbleGen HD2 respectively.

d. Bipolar Disorder Data Set

The Bipolar Disorder (BD) data set was composed of 3315 unrelated cases, including 161 that were collected as part of the Genetics of Early Onset Mania (GEM) Study of Bipolar Disorder and analyzed using the NimbleGen HD2 platform, 1697 cases analyzed by Cardiff University as part of the WTCCC study of bipolar disorder on the Affymetrix 500K platform and 1457 cases were obtained from the GAIN supported study: "Whole Genome Association of Bipolar Disorder" (phs000017.v2.p1), analyzed on the Affymetrix 6.0 platform. GAIN bipolar cases that overlapped with the GEM cases were removed from the analysis (N = 2).

The GEM Study of Bipolar Disorder is a collaborative project between CSHL, Johns Hopkins School of Medicine, NIMH and the Central Institute of Mental Health, in Mannheim Germany (CIMH). All procedures were approved by the Internal Review Boards at the respective institutions. In total the GEM sample consisted of 161 patients, with 50% of cases having an age-at-onset <18 (mean and median age at onset were 19 and 18, respectively). Of these 161 cases, 88 were collected as part of a family-based study by TGS and MR. Parent-offspring trios were collected where the offspring had a clinical diagnosis of bipolar 1

disorder and there was no bipolar I or schizoaffective-bipolar illness in first-degree relatives. Families were recruited from consecutive admissions to the inpatient units of several German University Departments of Psychiatry within a common research project on the genetics of bipolar disorders. DSM-IV diagnosis was made by psychiatrists/psychologists applying a consensus best-estimate procedure based on all available information, including a structured interview (SCID-I), medical records, and the family history method. Thirty-five of these 88 patients had an early age at onset of mania defined as onset below 21 years. All patients are of Caucasian descent, based on patient- and parent-based self report. For all patients and parents, written informed consent was obtained prior to study participation. Patients and controls also consented to transfer of DNA and phenotypic data in a pseudonymised way for collaboration with other researchers from foreign countries. Protocols and procedures were approved by the local ethics committees.

Collection and diagnosis of the 1697 WTCCC Bipolar Disorder cases have been previously described [13]. All cases were >16 years of age and ascertained at 5 locations across the United Kingdom. Written consent was obtained from each subject. Diagnosis was assigned based on the Research Diagnostic Criteria and best estimate ratings for key phenotypic measurements on the basis of the OPCRIT checklist [13]

Ascertainment of the 1457 GAIN Bipolar Disorder cases (1057 of European Ancestry and 400 of African Ancestry), has been described previously [19,20]. Data collection was also performed at the Broad Institute Center for Genotyping and Analysis (see URLs). Further details concerning the collection

sites, principal investigators and case-control inclusion criteria are available through dbGap (see URLs)

Association of 16p11.2 Microduplication with Schizophrenia

Our primary sample consisted of data from multiple microarray platforms. These platforms vary in probe density. The ROMA, Affymetrix 500K and Nimblegen HD2 platform consist of 85,000, 500,000 and 2.1 million probes respectively. Although all of these platforms have good sensitivity to detect CNVs that are 600 Kb in size, even subtle differences in sensitivity could influence the overall frequency of the 16p11.2 microduplication in patients and controls when all platforms are combined into a single dataset. Therefore, we controlled for array-type when testing for association. We applied the Cochran-Mantel-Haenszel (CMH) exact test using array type as a stratifying variable, and logistic regression was used to estimate the combined odds ratio (OR) and 95% confidence intervals (95%CI). Based on the logistic regression, the effect of array type was not significant (Deviance P-value = 0.76).

In our replication dataset, which consisted of data from a single study using a single array type, we tested the association of the 16p11.2 microduplication using a Fisher's exact test. The Breslow-Day-Tarone test was used to assess the homogeneity of the odds ratios between the primary and replication data sets. Odds ratios of the primary and replication datasets were not significantly different (P = 0.45).

In both the primary and replication data the ratio of males to females in cases is approximately 2, while the male-to-female ratio in control samples is approximately 1. We examined whether sex had an effect on the association. We

tested the association in the combined primary and replication sample using gender as a covariate. Gender did not have an effect (P=0.6) and the association was still significant (P = 3.75 X 10-7, OR = 15.7[3.7,67.3]).

Meta-Analysis and Strength of 16p11.2 Associations in Multiple Psychiatric Disorders

To refine the spectrum of psychiatric illnesses associated with 16p11.2 rearrangements, we examined the association of microduplications and microdeletions independently with schizophrenia, bipolar disorder and autism. By combining data from this study and three independent published studies [16,22,23], we obtained combined samples of 8590 patients with schizophrenia 4822 with bipolar disorder and 2172 with autism or developmental delay, and a combined sample of 30,492 controls. We tested the association of the microduplication and microdeletion with each disorder with a Cochran Mantel Haenszel exact test, and we controlled for variation by study as described below. Because we were using "study" as a stratifying variable, the control samples used for a particular disorder were derived only from those studies contributing cases of the same disorder. Thus the control samples for schizophrenia, bipolar disorder and autism/developmental delay consisted of 28,406, 25,225 and 24,891 individuals, respectively.

Supplementary Table 3 describes the patients and controls from each source that were used in the meta-analysis. In addition, we list the number of individuals with microdeletions, microduplications and normal copy number for each group.

We examined the association of 16p11.2 microduplications and microdeletions with schizophrenia, bipolar disorder and autism or developmental delay by meta-analysis of data from this study and additional studies of schizophrenia, bipolar disorder and autism spectrum disorders. We examined the association of the microduplication and microdeletion independently in each disorder with the Cochran-Mantel-Haenszel exact test using source as a stratifying variable. Logistic regression was used to determine the combined odds ratio from multiple studies for the meta-analysis of each disorder.

The Breslow-Day-Tarone test was used to assess the homogeneity of the odds ratios between the studies used in the meta-analysis. A common *P*-value is reported from the CMH exact test and a common odds-ratio is reported from the logistic regression only if there is homogeneity in the odds ratios between the studies in the meta-analysis (Table 2). With two exceptions, there was no significant heterogeneity in odds ratios for the microdeletion or microduplication in schizophrenia, bipolar disorder and autism/developmental delay. Because significant differences between studies were observed for microdeletions in schizophrenia (Breslow-Day Tarone $P = 0.003$) and bipolar disorder (Breslow-Day Tarone $P = 0.0004$), the odds ratios were reported independently for each source, and a common odds ratio was not reported (Table 2). The partial odds ratios [95%CI] for the deletion in schizophrenia were 0.69 [0.1, 4.9], 0.3 [0.05, 2.2], 14.6 [1.9, 111.2], 0.3 [0.03, 3.7] in this study, ISC, Weiss et al and GAIN, samples, respectively. Partial odds ratios [95% CI] for the microdeletion in BD were 0.3[0.03,3.3], 0.55[0.05,6.7], 25[5.4,117] for our study, the GAIN study and the Weiss et al. studies, respectively. Given the very small number of deletion observations in the GAIN schizophrenia study and in each of the BD studies,

approximate odds ratios were calculated by replacing the number of deletions *n* with {*n* + 0.5}. In light of contrasting results between studies, our meta-analysis does not show an unequivocal association of the microdeletion with schizophrenia or bipolar disorder.

Analysis of Quantitative Clinical Features with 16p11.2 Rearrangements

To perform a more systematic analysis of clinical features potentially associated with 16p11.2 rearrangements, we collected quantitative clinical data on height, weight, and head circumference from records on 16p11.2 carriers in this study, 16p11.2 carriers in previously published studies (Weiss et al. and Ghebranious et al. [16, 27]) and from unpublished 16p11.2 carriers ascertained by referral for global developmental delay (Tamim Shaikh, personal communication, Supplementary Table 4). We excluded from our analysis subjects with known Hispanic, Polynesian and African American ethnicity. In addition, we excluded subjects with documented cytogenetic abnormalities.

Data on height (cm) were available for 23 deletion carriers (5 with a primary diagnosis of autism, 15 with developmental delay, 1 with ADD, 1 with dysmorphism and 1 with schizophrenia) and nine microduplication carriers (2 with a primary diagnosis of ASD, 1 with developmental delay, 5 schizophrenia and 1 unaffected).

Head Circumference (HC) is typically measured just above the glabella to the area near the top of the occipital bone (opisthocranion) and hence is often called Occipital-Frontal Circumference (OFC). HC/OFC data were recorded for nine microduplication (2 with childhood onset schizophrenia, 3 with ASD, 3 with developmental delay and 1 with no diagnosis) and 23 microdeletion (6 with a

primary diagnosis of autism, 16 with developmental delay and 1 with dysmorphism) carriers. See supplementary table 4.

Occipital-Frontal Circumference and body height measurements were converted to percentile rankings, conditioned on age and gender using clinical growth charts from the Center for Disease Control's (CDC) National Center for Health Statistics (see URLs). OFC percentile rankings were further verified using the online tool (see URLs), which is based on the same reference database. Height and OFC percentiles were converted to Z-scores using online resources (see URLs). Continuously-distributed dependent measures were contrasted among 16p11.2 microdeletions versus microduplications carriers, using the Wilcoxon two-sample Rank Sum test. Last, we repeated the above analysis using, instead of the CDC growth carts, the craniofacial normative database of Farkas et al.[24,25] which is based on a population with European ancestry.

Mean head circumference was 1.25 [1.05] and -0.28 [0.71] for patients with microdeletions and microduplications, respectively. Difference in OFC was statistically significant (two-tailed Wilcoxon rank sum test $P = 0.0007$). In order to rule out the possibility that differences in head circumference were related to ethnicity, we examined OFC for a subset of patients with unambiguous European-American (EA) ancestry (8 with microdeletions and 6 with microduplications), and similar results were obtained. In this group, the mean [SD] OFC values of patients with microdeletions and microduplications were 1.34 [0.62] and -0.17 [0.42] based on CDC converted Z-scores and 0.82 [0.83] and -0.38 [0.49] based on Standard Normal Database converted Z-scores (CDC, two-tailed Wilcoxon rank sum test $P = 0.0007$).

We repeated our analysis using the Farkas normative database instead of the CDC growth charts. Mean OFC was 0.69[1.2] and -0.58[1.06] for patients with microdeletions and microduplications respectively, which was statistically significant (P = 0.013). The effect size (d) for this contrast was 1.12, which is very large according to Cohen's criteria [26]. In the subset of patients with unambiguous European ancestry, the difference was also significant (P = 0.02, d = 2.0).

Group contrast for height was not significantly different. Mean height was -0.02 [1.37] and 0.13 [0.81] for patients with the microdeletion and microduplication, respectively (P = 0.68). Due to limited availability of data on IQ, we did not examine intellectual disability of microdeletion and microduplication cases. Because of the known influences of antipsychotic medication on body weight, differences in weight between microdeletion and duplication carriers were not examined.

The above analysis included patients with diagnoses of schizophrenia (SCZ), developmental delay (DD) or autism spectrum disorders (ASD). We further examined whether the observed effect was present within each individual group (Supplementary Table 9). In the broad group of developmental disorders (DD/ASD) mean OFCs of patients with deletions and duplications were significantly different (ΔOFC = 1.44, P = 0.0056). In the subgroup of patients with developmental delay, the effect was also significant (ΔOFC = 1.96, P = 0.038), and in the smaller subgroup of patients with a primary diagnosis of ASD, the effect was present but not significant (ΔOFC = 0.65, P = 0.167). OFC measurements were available for only 2 schizophrenic patients with duplications, which were -0.88 and -0.18 respectively. Thus, the association of 16p11.2 reciprocal rearrangements with head circumference was observed in multiple

diagnostic categories, and the effect was not limited the subgroup of patients with autism.

Intensity Data Processing

a) Array-CGH Intensity Data

1. NimbleGen HD2

NimbleGen HD2 dual color intensity data were normalized in a two step process: first, a "spatial" normalization of probes was performed to adjust for regional differences in intensities across the surface of the array, and second, the Cy5 and Cy3 intensities were adjusted to a fitting curve by invariant set normalization.

Briefly, spatial normalization was performed using an R-module provided by Kyle Munn (Roche-NimbleGen Inc). Invariant set normalization of intensity data involves the selection of a probe set with minimal variability between the ranked test and reference autosomal probe intensities as described in Li et al [21]. The test intensities of this invariant autosomal probe set are then adjusted to the reference distribution. Based on these adjustments, a fitting curve is established to which all other intensities are shifted, preserving the variability in the data. The intensities of XY-chromosomes are then extrapolated to the fitting curve. The process is repeated exchanging the test and reference to simulate a dye swap experiment. The $\log_2$ ratio for probe *i* is then estimated using the geometric mean of normalized and raw intensity data of test (*Tst*) and reference (*Rfn*) as follows:

$$\log_2\left(Ratio_i\right) = \log_2\left(\sqrt{\frac{NormalizedTst_i}{RawRfn_i} \times \frac{RawTst_i}{NormalizedRfn_i}}\right)$$

b) SNP Genotyping Data

1. Affymetrix 500K, Affymetrix 5.0 & Affymetrix 6.0

We developed a two-step normalization method to process Affymetrix SNP Array single color intensity data (Affymetrix 500K, Affymetrix 5.0 & Affymetrix 6.0). First, all arrays are normalized by invariant set normalization to a single reference array. The common reference array was selected by comparing all arrays pair-wise based on the Pearson's correlation coefficient of a defined probe set representing 5% of all probes on the array. This defined probe set was composed of every 20th autosomal SNP probe and the adjacent copy number (CN) probe. Second, for each experiment we calculate the ratio of intensities in comparison to a "virtual reference genome" (VRG). The optimal VRG for each "test" consisted of the median of the 10 most highly-correlated experiments in the dataset based on the Pearson correlation of 5% of probes on the array. Separate virtual reference genomes were computed for males and females in order to have a sex-matched VRG.

$$\log_2(Ratio) = \log_2\left(\frac{NormalizedTst}{VRG}\right)$$

Further Assessment of HMM Detection and MeZOD Genotyping

To further assess the reliability of HMM detection and MeZOD genotyping, we compared HMM segmentation and MeZOD genotype calls made in a subset of the AGRE Affymetrix 5.0 data set originally used to identify the association 16p11.2 CNVs in autism [16]. The authors of this study reported fifteen 16p11.2

rearrangements (ten microduplications and five microdeletions) in seven AGRE families. All fifteen expected HMM rearrangements were detected by the HMM segmentor and fourteen were detected by MeZOD genotyping (Supplementary Figure 3). One microduplication was not detected by MeZOD, although it was detected by the HMM in our study and it was detected and validated in Weiss et al [16]. Examination of the probe ratio data in this subject revealed high variance in probe ratios genome-wide, which contributed to the negative result by MeZOD.

Description of Psychiatric Symptoms in 16p11.2 Microduplication Carriers

Data on the psychiatric symptoms of each microduplication carrier were acquired from medical case records or supplementary clinical data (Supplementary Table 8). With the exception of samples from the CATIE and GAIN studies all psychiatric symptom data were obtained from individual medical records. Clinical data for CATIE microduplication carriers were ascertained from baseline measures recorded in the PANSS (Positive and Negative Syndrome Scale) and SCID (Structured Clinical Interview for DSM Disorders) Public Access Files. Psychiatric symptoms for GAIN microduplication carriers were based on consensus ratings for DSM-IV criteria provided in the supplementary clinical data for project version *phs000021.v2.p1*. With the exception of early onset in some cases, no particular positive or negative symptom distinguished microduplication carriers from typical cases of schizophrenia.

URLS

The Broad Center for Genotyping and Analysis, www.broad.mit.node/306

dbGap (Schizophrenia), http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000021.v2.p1

NIMH Center for Genetics Studies, http://www.nimhgenetics.org/

Rutgers University Cell and DNA Repository (Autism), http://www.rucdr.org/ASHG/Autism.htm

dbGap (Bipolar), http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000017.v2.p1

CDC Clinical Growth Charts, http://www.cdc.gov/nchs/about/major/nhanes/growthcharts/clinical_charts.htm)

Simconsult, http://www.simulconsult.com/resources/ftemp20.html

HyperStat Online, http://davidmlane.com/hyperstat/z_table.html

Members of the Wellcome Trust Case Control Consortium

Management Committee: Paul R Burton[1], David G Clayton[2], Lon R Cardon[3], Nick Craddock[4],Panos Deloukas[5], Audrey Duncanson[6], Dominic P Kwiatkowski[3,5], Mark I McCarthy[3,7], Willem H Ouwehand[8,9], Nilesh J Samani[10], John A Todd[2], Peter Donnelly (Chair)[11]

Data and Analysis Committee: Jeffrey C Barrett[3], Paul R Burton[1], Dan Davison[11], Peter Donnelly[11], Doug Easton[12], David M. Evans[3], Hin-Tak Leung[2],

Jonathan L Marchini[11], Andrew P Morris[3], Chris CA Spencer[11], Martin D Tobin[1], Lon R Cardon (Co-chair)[3], David G Clayton (Co-chair)[2]

UK Blood Services & University of Cambridge Controls: Antony P Attwood[5,8], James P Boorman[8,9], Barbara Cant[8], Ursula Everson[13], Judith M Hussey[14], Jennifer D Jolley[8], Alexandra S Knight[8], Kerstin Koch[8], Elizabeth Meech[15], Sarah Nutland[2], Christopher V Prowse[16], Helen E Stevens[2], Niall C Taylor[8], Graham R Walters[17], Neil M Walker[2], Nicholas A Watkins[8,9], Thilo Winzer[8], John A Todd[2], Willem H Ouwehand[8,9]

1958 Birth Cohort Controls: Richard W Jones[18], Wendy L McArdle[18], Susan M Ring[18], David P Strachan[19], Marcus Pembrey[18,20]

Bipolar Disorder (Aberdeen): Gerome Breen[21], David St Clair[21]; (**Birmingham):** Sian Caesar[22], Katherine Gordon-Smith[22,23], Lisa Jones[22]; **(Cardiff):** Christine Fraser[23], Elaine K Green[23], Detelina Grozeva[23], Marian L Hamshere[23], Peter A Holmans[23], Ian R Jones[23], George Kirov[23], Valentina Moskvina[23], Ivan Nikolov[23], Michael C O'Donovan[23], Michael J Owen[23], Nick Craddock[23]; **(London):** David A Collier[24], Amanda Elkin[24], Anne Farmer[24], Richard Williamson[24], Peter McGuffin[24]; **(Newcastle):** Allan H Young[25], I Nicol Ferrier[25]

Coronary Artery Disease (Leeds): Stephen G Ball[26], Anthony J Balmforth[26], Jennifer H Barrett[26], D Timothy Bishop[26], Mark M Iles[26], Azhar Maqbool[26], Nadira Yuldasheva[26], Alistair S Hall[26]; **(Leicester):** Peter S Braund[10], Paul R Burton[1], Richard J Dixon[10], Massimo Mangino[10], Suzanne Stevens[10], Martin D Tobin[1], John R Thompson[1], Nilesh J Samani[10]

Crohn's Disease (Cambridge): Francesca Bredin[27], Mark Tremelling[27], Miles Parkes[27]; **(Edinburgh):** Hazel Drummond[28], Charles W Lees[28], Elaine R Nimmo[28], Jack Satsangi[28]; **(London):** Sheila A Fisher[29], Alastair Forbes[30], Cathryn M Lewis[29], Clive M Onnie[29], Natalie J Prescott[29], Jeremy Sanderson[31], Christopher G Mathew[29]; **(Newcastle):** Jamie Barbour[32], M Khalid Mohiuddin[32], Catherine E Todhunter[32], John C Mansfield[32]; **(Oxford):** Tariq Ahmad[33], Fraser R Cummings[33], Derek P Jewell[33]

Hypertension (Aberdeen): John Webster[34]; **(Cambridge):** Morris J Brown[35], David G Clayton[2]; **(Evry, France): G** Mark Lathrop[36]; **(Glasgow):** John Connell[37], Anna Dominiczak[37]; **(Leicester):** Nilesh J Samani[10]; **(London):** Carolina A Braga Marcano[38], Beverley Burke[38], Richard Dobson[38], Johannie Gungadoo[38], Kate L Lee[38], Patricia B Munroe[38], Stephen J Newhouse[38], Abiodun Onipinla[38], Chris Wallace[38], Mingzhan Xue[38], Mark Caulfield[38]; **(Oxford):** Martin Farrall[39]

Rheumatoid Arthritis: Anne Barton[40], Ian N Bruce[40], Hannah Donovan[40], Steve Eyre[40], Paul D Gilbert[40], Samantha L Hider[40], Anne M Hinks[40], Sally L John[40], Catherine Potter[40], Alan J Silman[40], Deborah PM Symmons[40], Wendy Thomson[40], Jane Worthington[40]

Type 1 Diabetes: David G Clayton[2], David B Dunger[2,41], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], Barry Widmer[2,41], John A Todd[2]

Type 2 Diabetes (Exeter): Timothy M Frayling[42,43], Rachel M Freathy[42,43], Hana Lango[42,43], John R B Perry[42,43], Beverley M Shields[43], Michael N Weedon[42,43], Andrew

T Hattersley[42,43]; **(London):** Graham A Hitman[44]; **(Newcastle):** Mark Walker[45]; **(Oxford):** Kate S Elliott[3,7], Christopher J Groves[7], Cecilia M Lindgren[3,7], Nigel W Rayner[3,7], Nicholas J Timpson[3,46], Eleftheria Zeggini[3,7], Mark I McCarthy[3,7]

Tuberculosis (Gambia): Melanie Newport[47], Giorgio Sirugo[47]; **(Oxford):** Emily Lyons[3], Fredrik Vannberg[3], Adrian VS Hill[3]

Ankylosing Spondylitis: Linda A Bradbury[48], Claire Farrar[49], Jennifer J Pointon[48], Paul Wordsworth[49], Matthew A Brown[48,49]

AutoImmune Thyroid Disease: Jayne A Franklyn[50], Joanne M Heward[50], Matthew J Simmonds[50], Stephen CL Gough[50]

Breast Cancer: Sheila Seal[51], Michael R Stratton[51,52], Nazneen Rahman[51]

Multiple Sclerosis: Maria Ban[53], An Goris[53], Stephen J Sawcer[53], Alastair Compston[53]

Gambian Controls (Gambia): David Conway[47], Muminatou Jallow[47], Melanie Newport[47],Giorgio Sirugo[47]; **(Oxford):** Kirk A Rockett[3], Dominic P Kwiatkowski[3,5]

DNA, Genotyping, Data QC and Informatics (Wellcome Trust Sanger Institute, Hinxton): Suzannah J Bumpstead[5], Amy Chaney[5], Kate Downes[2,5], Mohammed JR Ghori[5], Rhian Gwilliam[5], Sarah E Hunt[5], Michael Inouye[5], Andrew Keniry[5], Emma King[5], Ralph McGinnis[5], Simon Potter[5], Rathi Ravindrarajah[5], Pamela Whittaker[5], Claire Widden[5], David Withers[5], Panos Deloukas[5];

(Cambridge): Hin-Tak Leung[2], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], John A Todd[2]

Statistics (Cambridge): Doug Easton[12], David G Clayton[2]; **(Leicester):** Paul R Burton[1], Martin D Tobin[1]; **(Oxford):** Jeffrey C Barrett[3], David M Evans[3], Andrew P Morris[3], Lon R Cardon[3]; **(Oxford):** Niall J Cardin[11], Dan Davison[11], Teresa Ferreira[11], Joanne Pereira-Gale[11], Ingeleif B Hallgrimsdóttir[11], Bryan N Howie[11], Jonathan L Marchini[11], Chris CA Spencer[11], Zhan Su[11], Yik Ying Teo[3,11], Damjan Vukcevic[11], Peter Donnelly[11]

PIs: David Bentley[5,54], Matthew A Brown[48,49], Lon R Cardon[3], Mark Caulfield[38], David G Clayton[2], Alistair Compston[53], Nick Craddock[23], Panos Deloukas[5], Peter Donnelly[11], Martin Farrall[39], Stephen CL Gough[50], Alistair S Hall[26], Andrew T Hattersley[42,43], Adrian VS Hill[3],Dominic P Kwiatkowski[3,5], Christopher G Mathew[29], Mark I McCarthy[3,7], Willem H Ouwehand[8,9], Miles Parkes[27], Marcus Pembrey[18,20], Nazneen Rahman[51], Nilesh J Samani[10], Michael R Stratton[51,52], John A Todd[2], Jane Worthington[40]

[1] Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester, LE1 7RH, UK;

[2] Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Cambridge, CB2 0XY, UK;

[3] Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK;

[4] Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK;

[5] The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK;

[6] The Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK;

[7] Oxford Centre for Diabetes, Endocrinology and Medicine, University of Oxford, Churchill Hospital, Oxford, OX3 7LJ, UK;

[8] Department of Haematology, University of Cambridge, Long Road, Cambridge, CB2 2PT, UK;

[9] National Health Service Blood and Transplant, Cambridge Centre, Long Road, Cambridge, CB2 2PT, UK;

[10] Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Groby Road, Leicester, LE3 9QP, UK;

[11] Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK;

[12] Cancer Research UK Genetic Epidemiology Unit, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK;

[13] National Health Service Blood and Transplant, Sheffield Centre, Longley Lane, Sheffield S5 7JN, UK;

[14] National Health Service Blood and Transplant, Brentwood Centre, Crescent Drive, Brentwood, CM15 8DP, UK;

[15] The Welsh Blood Service, Ely Valley Road, Talbot Green, Pontyclun, CF72 9WB, UK;

[16] The Scottish National Blood Transfusion Service, Ellen's Glen Road, Edinburgh, EH17 7QT, UK;

[17] National Health Service Blood and Transplant, Southampton Centre, Coxford Road, Southampton, SO16 5AF, UK;

[18] Avon Longitudinal Study of Parents and Children, University of Bristol, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK;

[19] Division of Community Health Services, St George's University of London, Cranmer Terrace, London SW17 0RE, UK;

[20] Institute of Child Health, University College London, 30 Guilford St, London WC1N 1EH, UK;

[21] University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen, AB25 2ZD, UK;

[22] Department of Psychiatry, Division of Neuroscience, Birmingham University, Birmingham, B15 2QZ, UK;

[23] Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK;

[24] SGDP, The Institute of Psychiatry, King's College London, De Crespigny Park Denmark Hill London SE5 8AF, UK;

[25] School of Neurology, Neurobiology and Psychiatry, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne, NE1 4LP, UK;

[26] LIGHT and LIMM Research Institutes, Faculty of Medicine and Health, University of Leeds, Leeds, LS1 3EX, UK;

[27] IBD Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, CB2 2QQ, UK;

[28] Gastrointestinal Unit, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU UK;

[29] Department of Medical & Molecular Genetics, King's College London School of Medicine, 8th Floor Guy's Tower, Guy's Hospital, London, SE1 9RT, UK;

[30] Institute for Digestive Diseases, University College London Hospitals Trust, London, NW1 2BU, UK;

[31] Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London, SE1 7EH, UK;

[32] Department of Gastroenterology & Hepatology, University of Newcastle upon Tyne, Royal Victoria Infirmary, Newcastle upon Tyne, NE1 4LP, UK;

[33] Gastroenterology Unit, Radcliffe Infirmary, University of Oxford, Oxford, OX2 6HE, UK;

[34] Medicine and Therapeutics, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, Grampian AB9 2ZB, UK;

[35] Clinical Pharmacology Unit and the Diabetes and Inflammation Laboratory, University of Cambridge, Addenbrookes Hospital, Hills Road, Cambridge CB2 2QQ, UK;

[36] Centre National de Genotypage, 2, Rue Gaston Cremieux, Evry, Paris 91057.;

[37] BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow, G12 8TA, UK;

[38] Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London, Queen Mary's School of Medicine, Charterhouse Square, London EC1M 6BQ, UK;

[39] Cardiovascular Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK;

[40] Epidemiology Research Unit, University of Manchester, Stopford Building, Oxford Rd, Manchester, M13 9PT, UK;

[41] Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 2QQ, UK;

[42] Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, Magdalen Road, Exeter EX1 2LU UK;

[43] Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula Medical School, Barrack Road, Exeter EX2 5DU UK;

[44] Centre for Diabetes and Metabolic Medicine, Barts and The London, Royal London Hospital, Whitechapel, London, E1 1BB UK;

[45] Diabetes Research Group, School of Clinical Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK;

[46] The MRC Centre for Causal Analyses in Translational Epidemiology, Bristol University, Canynge Hall, Whiteladies Rd, Bristol BS2 8PR, UK;

[47] MRC Laboratories, Fajara, The Gambia;

[48] Diamantina Institute for Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Woolloongabba, Qld 4102, Australia;

[49] Botnar Research Centre, University of Oxford, Headington, Oxford OX3 7BN, UK;

[50] Department of Medicine, Division of Medical Sciences, Institute of Biomedical Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK;

[51] Section of Cancer Genetics, Institute of Cancer Research, 15 Cotswold Road, Sutton, SM2 5NG, UK;

[52] Cancer Genome Project, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK;

[53] Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK;

[54] PRESENT ADDRESS: Illumina Cambridge, Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex, CB10 1XL, UK.

References

1. Walsh, T. et al. Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia. *Science* **320**, 539-543 (2008).
2. Sullivan, P.F. et al. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* **13**, 570-84 (2008).
3. Kirov, G. et al. Support for the involvement of large cnvs in the pathogenesis of schizophrenia. *Hum Mol Genet* **14**, 796-803 (2009).
4. Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E. & Pritchard, J.K. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* **38**, 75-81 (2006).
5. McCarroll, S.A. et al. Common deletion polymorphisms in the human genome. *Nat Genet* **38**, 86-92 (2006).
6. Fung, H.C. et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol* **5**, 911-6 (2006).
7. Brownstein, J. et al. Antisaccade performance is abnormal in schizophrenia patients but not in their biological relatives. *Schizophr Res* **63**, 13-25 (2003).
8. Susser, E., Hoek, H.W. & Brown, A. Neurodevelopmental disorders after prenatal famine: The story of the Dutch Famine Study. *Am J Epidemiol* **147**, 213-6 (1998).
9. DeLisi, L.E. et al. A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. *Am J Psychiatry* **159**, 803-12 (2002).
10. Mitchell, M.K., Gregersen, P.K., Johnson, S., Parsons, R. & Vlahov, D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J Urban Health* **81**, 301-10 (2004).
11. Talati, A., Fyer, A.J. & Weissman, M.M. A comparison between screened NIMH and clinically interviewed control samples on neuroticism and extraversion. *Mol Psychiatry* **13**, 122-30 (2008).
12. Sebat, J. et al. Large-scale copy number polymorphism in the human genome. *Science* **305**, 525-8 (2004).
13. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
14. Stroup, T.S. et al. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. *Schizophr Bull* **29**, 15-31 (2003).
15. Shi, J. et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753-7 (2009).
16. Weiss, L.A. et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* **358**, 667-75 (2008).
17. Sebat, J. et al. Strong association of de novo copy number mutations with autism. *Science* **316**, 445-9 (2007).
18. Conroy, J. et al. Fine mapping and association studies in a candidate region for autism on chromosome 2q31-q32. *Am J Med Genet B Neuropsychiatr Genet* **150B**, 535-44 (2009).

19. Smith, E.N. et al. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry* **14**, 755-63 (2009).
20. Zhang, D. et al. Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol.Psychiatry* **14**, 376-380 (2009).
21. Li, C. & Hung Wong, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* **2**, RESEARCH0032 (2001).
22. Stone, J.L. et al. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237-241 (2008).
23. Marshall, C.R. et al. Structural Variation of Chromosomes in Autism Spectrum Disorder. *The American Journal of Human Genetics* **82**, 477-488 (2008).
24. Deutsch, C.K. Head Circumference in autism. in *Autism and PDD Dictionary* (eds. Neisworth, J. & Wolfe, P.) (Brookes, Baltimore, MD, 2004).
25. Deutsch, C.K. & Farkas, L.G. Quantitative methods of dysmorphology diagnosis. in *Anthropometry of the head and face* (ed. Farkas, L.G.) 151-158 (Raven, New York, 1994).
26. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, (Lawrence Erlbaum, Hillside, NJ, 1988).
27. Ghebranious, N., Giampietro, P.F., Wesbrook, F.P. & Rezkalla, S.H. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *Am J Med Genet A* **143**, 1462-71 (2007).
28. Bijlsma, E.K. et al. Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *Eur J Med Genet* (2009).
29. Saito, S., Sakagami, H., Tonosaki, A. & Kondo, H. Localization of mRNAs for CDP-diacylglycerol synthase and phosphatidylinositol synthase in the brain and retina of developing and adult rats. *Brain Res Dev Brain Res* **110**, 21-30 (1998).
30. Sakaguchi, G. et al. Doc2alpha is an activity-dependent modulator of excitatory synaptic transmission. *Eur J Neurosci* **11**, 4262-8 (1999).
31. Baron, C.A. et al. Genomic and functional profiling of duplicated chromosome 15 cell lines reveal regulatory alterations in UBE3A-associated ubiquitin-proteasome pathway processes. *Hum Mol Genet* **15**, 853-69 (2006).
32. Mazzucchelli, C. et al. Knockout of ERK1 MAP kinase enhances synaptic plasticity in the striatum and facilitates striatal-mediated learning and memory. *Neuron* **34**, 807-20 (2002).
33. Bowden, N.A. et al. Preliminary investigation of gene expression profiles in peripheral blood lymphocytes in schizophrenia. *Schizophr Res* **82**, 175-83 (2006).
34. Paspalas, C.D. et al. Major vault protein is expressed along the nucleus-neurite axis and associates with mRNAs in cortical neurons. *Cereb Cortex* **19**, 1666-77 (2009).
35. Sisodiya, S.M. et al. Major vault protein, a marker of drug resistance, is upregulated in refractory epilepsy. *Epilepsia* **44**, 1388-96 (2003).

36. Kowalski, M.P. et al. Host resistance to lung infection mediated by major vault protein in epithelial cells. *Science* **317**, 130-2 (2007).
37. Yu, Z.L. et al. Febrile seizures are associated with mutation of seizure-related (SEZ) 6, a brain-specific gene. *J Neurosci Res* **85**, 166-72 (2007).
38. Miyazaki, T. et al. Disturbance of cerebellar synaptic maturation in mutant mice lacking BSRPs, a novel brain-specific receptor-like protein family. *FEBS Lett* **580**, 4057-64 (2006).
39. Chapman, D.L. & Papaioannou, V.E. Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* **391**, 695-7 (1998).