Copy number variants in schizophrenia: confirmation of five previous findings and possible association of 3q29 microdeletions and VIPR2 duplications

Supplementary Information

Table S1: Published studies of CNVs in Schizophrenia

First author, Year	Sample	Platforms	Analysis package	Genome-wide vs. focused	Major findings	Case:control ratio for reported CNVs in each region: 1q21.1	2p16.3 (*NRXN1*)	15q11.2	15q13.2	16p11.2	22q11.21
Bassett, 2008	42 cases/53 controls	Affy 250K	dChip, CNAG, GEMCA	Genome-wide	No rare CNV other than 22q11.2 deletion confer susceptibility to SZ	NA	NA	NA	NA	NA	NA
Bruce, 2009	34 probands	Array-CGH	NA	Genome-wide	Detected a deletion on 5p15.1 in two probands; but no association	NA	NA	NA	NA	NA	NA
Brunet, 2008	190 cases	MLPA	NA	Focused	Foud two 22q11.2 deletions but no duplication	NA	NA	NA	NA	NA	2:0 del
Friedman, 2008	335 cases/512 controls	Array-CGH	NA	Genome-wide	Two SZ cases had deletions in CNTNAP2 gene	NA	NA	NA	NA	NA	NA
Ikeda, 2009	575 cases/ 564 controls	Diverse	Birdsuite	Genome-wide	Nonsignificant excess of rare CNVs in SZ (p = .087)	1:0 del	1:0 del	NA	NA	NA	NA
Ingason, 2009	4,345 cases/ 35,079 controls	Diverse	Dosage Miner, Quanti SNP	Focused	Three-fold excess of duplications and deletions of 16p13.1 in SZ cases	NA	NA	NA	NA	NA	NA
Kirov, 2008	93 trios	Array CGH	CGHPRO	Genome-wide	Two CNVs likely to be pathogenic	NA	1 del	NA	1 dup	NA	NA
Kirov, 2009	471 cases / 2,792 controls	Affy 500K	Genotyping Console v2.1	Genome-wide	Large CNVs (>1Mb) were 2.26 times over-represented in cases	0:2 del	1:3 del	4:14 del	0:0 del	NA	2:0 del
Lee, 2010	20 cases	Array-based CGH	NA	Genome-wide	Cases with negative symptomes have more genic CNVs (13 vs. 6)	NA	NA	NA	NA	NA	NA
McCarthy, 2009	1,906 cases/3,971 controls (discovery); 2,645 cases/2,420 controls (replication)	Diverse	a modified HMM	Focused	16p11.2 micorduplication is strongly associated with SZ	NA	NA	NA	NA	21: 2 dup	NA
Moon, 2006	30 cases/ 20 controls	Array-CGH	NA	Genome-wide	No specific CNV was associated	NA	NA	NA	NA	NA	
Mulle, 2010	245 cases, 490 controls (discovery); available published data	Affy 6.0	GLAD, GADA, BEAST	Genome-wide	Focused on 3q29 deletions (among deletions >500Mb, not in DGV, found only in cases	NA	NA	NA	NA	NA	2:0
Need, 2009	1,013 cases / 1,084 controls	HumanHap300 ,550,or 610 chips	PennCNV	Genome-wide	Large CNVs (>2Mb) are enriched in cases	1:0 del	3:1 del	NA	NA	NA	4:0 del
Rodriguez-Santiago, 2009	654 cases/ 604 controls	Diverse	PennCNV	Focused	Common CNVs at two glutathione S-transferase (GST) genes asscoiated with SZ	NA	NA	NA	NA	NA	NA

Rujescu, 2008	2,977 cases/ 33,746 controls	Diverse	Dosage Miner, Quanti SNP	Focused	NRXN1 deletions affecting exons confer risk of SZ	NA	12:49 del;2:3 dup	NA	NA	NA	NA
Stefansson, 2008	1,433 cases / 33,250 controls; 3 CNVs (1q21.1, 15q11.2 and 15q13.3) were followed up in 3,285 cases / 7,951 controls	Diverse	Dosage Miner	Genome-wide	Three rare CNVs (1q21.1, 15q11.2, and 15q13.3) showed nominal association	11:8 del	0:2 del	26:79 del	7:8 del	2:11 del	8:0 del
Steinbert, 2010	4,235 cases (psychosis) / 3,9481 controls	Diverse	PennCNV	Focused	Two CNVs in ZNF804A in psychosis patients and none in controls	NA	NA	NA	NA	NA	NA
Stone, 2008	3,391 cases / 3,181 controls	Affy 5.0/6.0	Birdsuite	Genome-wide	Rare (<1%) and large CNVs (>100kb) are enriched in cases (1.15-fold); 3 regions (1q21.1, 15q13.2, and 22q11.21) showed significant association	10:1 del	5:6 del	26:11 del	9:0 del	5:1 dup	13:0 del
Walsh, 2008	150 cases / 268 controls; 92 childhood onset cases	Array CGH	ROMA	Genome-wide	Rare CNVs in 15% cases vs. 5% controls	1:0 del	1:0 del	NA	NA	2:0 dup	NA
Wilson, 2006	35 cases/ 35 controls	Array-CGH	NA	Genome-wide	4 loci with CNV were only in cases	NA	NA	NA	NA	NA	NA
Xu, 2008	359 trios (screening); 152 cases / 159 controls	Affy 5.0	dCHIP	Genome-wide	In sporadic cases, frequency of rare de novo CNVs was 10% vs. 1.3% in controls	1:0 del	NA	NA	NA	NA	3:0 del
Xu, 2009	48 familial, 152 sporadic cases/159 controls	Affy 5.0	Birdsuite	Genome-wide	Rare genic CNVs are enriched in familial cases vs. controls.	NA	NA	NA	NA	NA	NA

See references below. The table provides summaries of findings in published papers on CNVs in schizophrenia. Note that in the main text, we only use data from large studies or meta-analyses which are technically comparable to the MGS analysis (Affymetrix or Illumina GWAS chips; data reported genome-wide or for CNVs of interest in large samples). We have not attempted to combine all data from the studies described in the table, many of which are focused studies of specific regions, or involved small samples, or overlap with samples reported in the large studies cited in the main text and tables.

References for Table S1

1. Bassett AS, Marshall CR, Lionel AC, Chow EW, Scherer SW. Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. Hum Mol Genet. 2008 Dec 15;17(24):4045-53. Epub 2008 Sep 20. Erratum in: Hum Mol Genet. 2009 May 1;18(9):1717. PubMed PMID: 18806272; PubMed Central PMCID: PMC2638574.
2. Bruce HA, Sachs N, Rudnicki DD, Lin SG, Willour VL, Cowell JK, Conroy J, McQuaid DE, Rossi M, Gaile DP, Nowak NJ, Holmes SE, Sklar P, Ross CA, Delisi LE, Margolis RL. Long tandem repeats as a form of genomic copy number variation: structure and length polymorphism of a chromosome 5p repeat in control and schizophrenia populations. Psychiatr Genet. 2009 Apr;19(2):64-71. PubMed PMID: 19672138.
3. Brunet A, Armengol L, Pelaez T, Guillamat R, Vallès V, Gabau E, Estivill X, Guitart M. Failure to detect the 22q11.2 duplication syndrome rearrangement among patients with schizophrenia. Behav Brain Funct. 2008 Feb 19;4:10. PubMed PMID: 18284679; PubMed Central PMCID: PMC2278148.
4. Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH, Knoers NV, Cahn W, Kahn RS, Edelmann L, Davis KL, Silverman JM, Brunner HG, van Kessel AG, Wijmenga C, Ophoff RA, Veltman JA. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Mol Psychiatry. 2008 Mar;13(3):261-6. Epub 2007 Jul 24. PubMed PMID: 17646849.
5. Ikeda M, Aleksic B, Kirov G, Kinoshita Y, Yamanouchi Y, Kitajima T, Kawashima K, Okochi T, Kishi T, Zaharieva I, Owen MJ, O'Donovan MC, Ozaki N, Iwata N. Copy number variation in schizophrenia in the Japanese population. Biol Psychiatry. 2010 Feb 1;67(3):283-6. Epub 2009 Oct 31. PubMed PMID: 19880096.
6. Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietiläinen OP, Buizer-Voskamp JE, Strengman E, Francks C, Muglia P, Gylfason A, Gustafsson O, Olason PI, Steinberg S, Hansen T, Jakobsen KD, Rasmussen HB, Giegling I, Möller HJ, Hartmann A, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Bramon E, Kiemeney LA, Franke B, Murray R, Vassos E, Toulopoulou T, Mühleisen TW, Tosato S, Ruggeri M, Djurovic S, Andreassen OA, Zhang Z, Werge T, Ophoff RA; GROUP Investigators, Rietschel M, Nöthen MM, Petursson H, Stefansson H, Peltonen L, Collier D, Stefansson K, Clair DM. Copy number variations of chromosome 16p13.1 region associated with schizophrenia. Mol Psychiatry. 2009 Sep 29. [Epub ahead of print] PubMed PMID: 19786961.
7. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008 Sep 11;455(7210):237-41. Epub 2008 Jul 30. PubMed PMID: 18668038.
8. Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P; International Schizophrenia Consortium; Wellcome Trust Case Control Consortium, Craddock N, Owen MJ, O'Donovan MC. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Hum Mol Genet. 2009 Apr 15;18(8):1497-503. Epub 2009 Jan 29. PubMed PMID: 19181681; PubMed Central PMCID: PMC2664144.
9. Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, O'Donovan MC, Erdogan F, Owen MJ, Ropers HH, Ullmann R. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. Hum Mol Genet. 2008 Feb 1;17(3):458-65. Epub 2007 Nov 6. PubMed PMID: 17989066.
10. Lee CH, Liu CM, Wen CC, Chang SM, Hwu HG. Genetic copy number variants in sib pairs both affected with schizophrenia. J Biomed Sci. 2010 Jan 11;17:2. PubMed PMID: 20064257; PubMed Central PMCID: PMC2843606.
11. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, Perkins DO, Dickel DE, Kusenda M, Krastoshevsky O, Krause V, Kumar RA, Grozeva D, Malhotra D, Walsh T, Zackai EH, Kaplan P, Ganesh J, Krantz ID, Spinner NB, Roccanova P, Bhandari A, Pavon K, Lakshmi B, Leotta A, Kendall J, Lee YH, Vacic V, Gary S, Iakoucheva LM, Crow TJ, Christian SL, Lieberman JA, Stroup TS, Lehtimäki T, Puura K, Haldeman-Englert C, Pearl J, Goodell M, Willour VL, Derosse P, Steele J, Kassem L, Wolff J, Chitkara N, McMahon FJ, Malhotra AK, Potash JB, Schulze TG, Nöthen MM, Cichon S, Rietschel M, Leibenluft E, Kustanovich V, Lajonchere CM, Sutcliffe JS, Skuse D, Gill M, Gallagher L, Mendell NR; Wellcome Trust Case Control Consortium, Craddock N, Owen MJ, O'Donovan MC, Shaikh TH, Susser E, Delisi LE, Sullivan PF, Deutsch CK, Rapoport J, Levy DL, King MC, Sebat J. Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet. 2009 Nov;41(11):1223-7. Epub 2009 Oct 25. PubMed PMID: 19855392.

12. Moon HJ, Yim SV, Lee WK, Jeon YW, Kim YH, Ko YJ, Lee KS, Lee KH, Han SI, Rha HK. Identification of DNA copy-number aberrations by array-comparative genomic hybridization in patients with schizophrenia. Biochem Biophys Res Commun. 2006 Jun 2;344(2):531-9. Epub 2006 Apr 3. PubMed PMID: 16630559.
13. Mulle JG, Dodd AF, McGrath JA, Wolyniec PS, Mitchell AA, Shetty AC, Sobreira NL, Valle D, Rudd MK, Satten G, Cutler DJ, Pulver AE, Warren ST: Microdeletions of 3q29 confer high risk for schizophrenia. Am J Hum Genet 2010; 87(2):229-36. Pubmed PMID: 20691406.
14. Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen EL, Shianna KV, Yoon W, Kasperaviciūte D, Gennarelli M, Strittmatter WJ, Bonvicini C, Rossi G, Jayathilake K, Cola PA, McEvoy JP, Keefe RS, Fisher EM, St Jean PL, Giegling I, Hartmann AM, Möller HJ, Ruppert A, Fraser G, Crombie C, Middleton LT, St Clair D, Roses AD, Muglia P, Francks C, Rujescu D, Meltzer HY, Goldstein DB. A genome-wide investigation of SNPs and CNVs in schizophrenia. PLoS Genet. 2009 Feb;5(2):e1000373. Epub 2009 Feb 6. Erratum in: PLoS Genet. 2009 Mar;5(3). doi: 10.1371/annotation/e0196ebb-de40-453f-8f8c-791b126618da. PubMed PMID: 19197363; PubMed Central PMCID: PMC2631150.
15. Rodríguez-Santiago B, Brunet A, Sobrino B, Serra-Juhé C, Flores R, Armengol L, Vilella E, Gabau E, Guitart M, Guillamat R, Martorell L, Valero J, Gutiérrez-Zotes A, Labad A, Carracedo A, Estivill X, Pérez-Jurado LA. Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia. Mol Psychiatry. 2009 Jun 16. [Epub ahead of print] PubMed PMID: 19528963.
16. Rujescu D, Ingason A, Cichon S, Pietiläinen OP, Barnes MR, Toulopoulou T, Picchioni M, Vassos E, Ettinger U, Bramon E, Murray R, Ruggeri M, Tosato S, Bonetto C, Steinberg S, Sigurdsson E, Sigmundsson T, Petursson H, Gylfason A, Olason PI, Hardarsson G, Jonsdottir GA, Gustafsson O, Fossdal R, Giegling I, Möller HJ, Hartmann AM, Hoffmann P, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Djurovic S, Melle I, Andreassen OA, Hansen T, Werge T, Kiemeney LA, Franke B, Veltman J, Buizer-Voskamp JE; GROUP Investigators, Sabatti C, Ophoff RA, Rietschel M, Nöthen MM, Stefansson K, Peltonen L, St Clair D, Stefansson H, Collier DA. Disruption of the neurexin 1 gene is associated with schizophrenia. Hum Mol Genet. 2009 Mar 1;18(5):988-96. Epub 2008 Oct 22. PubMed PMID: 18945720; PubMed Central PMCID: PMC2695245.
17. Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Möller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Toulopoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Mühleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeney LA, Franke B; GROUP, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nöthen MM, Peltonen L, Collier DA, St Clair D, Stefansson K. Large recurrent microdeletions associated with schizophrenia. Nature. 2008 Sep 11;455(7210):232-6. PubMed PMID: 18668039; PubMed Central PMCID: PMC2687075.
18. Steinberg S, Mors O, Børglum AD, Gustafsson O, Werge T, Mortensen PB, Andreassen OA, Sigurdsson E, Thorgeirsson TE, Böttcher Y, Olason P, Ophoff RA, Cichon S, Gudjonsdottir IH, Pietiläinen OP, Nyegaard M, Tuulio-Henriksson A, Ingason A, Hansen T, Athanasiu L, Suvisaari J, Lonnqvist J, Paunio T, Hartmann A, Jürgens G, Nordentoft M, Hougaard D, Norgaard-Pedersen B, Breuer R, Möller HJ, Giegling I, Glenthøj B, Rasmussen HB, Mattheisen M, Bitter I, Réthelyi JM, Sigmundsson T, Fossdal R, Thorsteinsdottir U, Ruggeri M, Tosato S, Strengman E; GROUP, Kiemeney LA, Melle I, Djurovic S, Abramova L, Kaleda V, Walshe M, Bramon E, Vassos E, Li T, Fraser G, Walker N, Toulopoulou T, Yoon J, Freimer NB, Cantor RM, Murray R, Kong A, Golimbet V, Jönsson EG, Terenius L, Agartz I, Petursson H, Nöthen MM, Rietschel M, Peltonen L, Rujescu D, Collier DA, Stefansson H, St Clair D, Stefansson K. Expanding the range of ZNF804A variants conferring risk of psychosis. Mol Psychiatry. 2010 Jan 5. [Epub ahead of print] PubMed PMID: 20048749.
19. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Roccanova P, Makarov V, Lakshmi B, Findling RL,

Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gochman P, Long R, Chen Z, Davis S, Baker C, Eichler EE, Meltzer PS, Nelson SF, Singleton AB, Lee MK, Rapoport JL, King MC, Sebat J. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008 Apr 25;320(5875):539-43. Epub 2008 Mar 27. PubMed PMID: 18369103.

20. Wilson GM, Flibotte S, Chopra V, Melnyk BL, Honer WG, Holt RA. DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling. Hum Mol Genet. 2006 Mar 1;15(5):743-9. Epub 2006 Jan 24. PubMed PMID: 16434481.
21. Xu B, Woodroffe A, Rodriguez-Murillo L, Roos JL, van Rensburg EJ, Abecasis GR, Gogos JA, Karayiorgou M. Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans. Proc Natl Acad Sci U S A. 2009 Sep 29;106(39):16746-51. Epub 2009 Sep 11. PubMed PMID: 19805367; PubMed Central PMCID: PMC2757863.
22. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet. 2008 Jul;40(7):880-5. Epub 2008 May 30. PubMed PMID: 18511947.

Table S2: CNV calling criteria

	----------Broad criteria----------				----------Narrow criteria----------			
CN	**Probes**	**LOD**	**N**	**Concordance**	**Probes**	**LOD**	**N**	**Concordance**
0	3	2	431	0.83	5	6	152	0.93
1	5	2	3833	0.84	6	6*	1,928	0.93
3,4	6	6	1042	0.72	6	10	622	0.78

* If lod >5 but < 6, then minumum probes = 9

Shown are the Narrow and Broad CNV calling criteria for each copy number (CN). Narrow criteria were used in the primary analyses. Probes = the minimum number of probes within the CNV call; LOD = the minimum Birdseye LOD score required to make the call. N = the number of calls made in 151 specimens used in the analysis of duplicate concordance used to validate the criteria. Concordance = the proportion of calls in DNA specimen "1" for a given subject (meeting these criteria) that overlapped by at least 50% with a call in the same direction (deletion or duplication) in specimen "2" for that subject (with no minimum call criteria for specimen 2).

Table S3: Regions excluded from CNV analyses (HG18)

Immunoglobulin gene regions

chr2:88937989-89411302
chr14:105065301-106352275
chr14:21159897-22090937
chr22:20715572-21595082

Centromic regions

chr1:121100001-128000000
chr2:91000001-95700000
chr3:89400001-93200000
chr4:48700001-52400000
chr5:45800001-50500000
chr6:58400001-63400000
chr7:57400001-61100000
chr8:43200001-48100000
chr9:46700001-60300000
chr10:38800001-42100000
chr11:51400001-56400000
chr12:33200001-36500000
chr13:13500001-18400000
chr14:13600001-19100000
chr15:14100001-18400000
chr16:34400001-40700000
chr17:22100001-23200000
chr18:15400001-17300000
chr19:26700001-30200000
chr20:25700001-28400000
chr21:10000001-13200000
chr22:9600001-16300000
chrX:56600001-65000000
chrY:11200001-12500000

Telomeric regions (100kb from pter or qter)

chr1:1-100000
chr2:1-100000
chr3:1-100000
chr4:1-100000
chr5:1-100000
chr6:1-100000
chr7:1-100000
chr8:1-100000
chr9:1-100000
chr10:1-100000
chr11:1-100000
chr12:1-100000
chr13:1-100000
chr14:1-100000
chr15:1-100000
chr16:1-100000
chr17:1-100000
chr18:1-100000
chr19:1-100000
chr20:1-100000
chr21:1-100000
chr22:1-100000
chrX:1-100000
chrY:1-100000
chr1:247149719-247249719
chr2:242851149-242951149
chr3:199401827-199501827
chr4:191173063-191273063
chr5:180757866-180857866
chr6:170799992-170899992
chr7:158721424-158821424
chr8:146174826-146274826
chr9:140173252-140273252
chr10:135274737-135374737
chr11:134352384-134452384
chr12:132249534-132349534
chr13:114042980-114142980
chr14:106268585-106368585
chr15:100238915-100338915
chr16:88727254-88827254
chr17:78674742-78774742
chr18:76017153-76117153
chr19:63711651-63811651
chr20:62335964-62435964
chr21:46844323-46944323
chr22:49591432-49691432
chrX:154813754-154913754
chrY:57672954-57772954

Table S4: MGS dataset CNV counts (Narrow criteria)

	Total CNVs	CNVs/ subject	Large (>100kb)	CNVs/ subject
ALL				
Deletions	452,007	62.05	45,791	6.29
Duplications	68,916	9.46	31,415	4.31
RARE (<1%)				
Deletions	37,561	5.16	4,243	0.58
Duplications	13,746	1.89	5,044	0.69

Shown are the total number of deletions or of duplications in the dataset of 7,285 subjects, and the number per subject, separately for CNVs of all sizes (Total) and those larger than 100,000 bp, and for CNVs of all frequencies (ALL) or <1% frequency (RARE -- filtered by PLINK as discussed in the text). for CNVs meeting the Narrow call criteria (Table S2). These N's were observed after data cleaning, i.e., after exclusion of samples, merger of segments as described in the main text, exclusion of CNVs in the regions listed in Table S3, and exclusion of CNVs with apparent plate effects (most CNVs in the region from specimens on 1 or 2 plates).

Table S4b: Genome-wide CNVs for case DNAs extracted from LCLs vs. from blood

	NARROW LARGE DELS			NARROW LARGE DUPS			NARROW SMALL DELS			NARROW SMALL DUPS		
	EMP-P	LCL	BLOOD	EMP1	LCL	BLOOD	EMP-P	LCL	BLOOD	EMP1	LCL	BLOOD
N CNVs		1613	496		1944	650		12986	3902		3437	1106
RATE	0.3590	0.537	0.526	0.9017	0.648	0.689	***0.0263***	4.326	4.138	0.7630	1.145	1.173
PROP	0.2000	0.405	0.389	0.9122	0.463	0.488	0.5454	0.962	0.962	0.8192	0.671	0.686
KBTOT	0.5274	373.7	375.3	0.9886	454.8	517.0	0.5310	81.7	81.9	0.0207	67.9	63.3
KBAVG	0.3986	277.3	272.6	0.9800	328.4	366.7	0.9620	19.0	19.9	0.0018	40.2	37.2
GRATE	0.4849	0.630	0.621	0.5124	1.346	1.349	0.4564	1.383	1.375	0.0242	0.754	0.668
GPROP	0.2204	0.228	0.215	0.6869	0.348	0.356	0.5643	0.690	0.691	0.1869	0.404	0.387
GRICH	0.6083	0.0048	0.0049	0.1220	0.0076	0.0071	0.6048	0.026	0.025	0.2865	0.021	0.019
	BROAD LARGE DELS			**BROAD LARGE DUPS**			**BROAD SMALL DELS**			**BROAD SMALL DUPS**		
N CNVs		1619	500		2051	729		22774	6553		4873	1587
RATE	0.4299	0.540	0.534	0.998	0.684	0.779	***0.0001***	***7.599***	7.001	0.907	1.626	1.696
PROP	0.2491	0.412	0.399	0.997	0.481	0.531	0.3843	0.996	0.995	0.832	0.778	0.792
KBTOT	0.6972	365	377.1	0.996	457.2	526.8	0.1197	98.71	95.86	0.391	72.82	72.12
KBAVG	0.5970	273.8	277.6	0.987	322.1	359.8	0.9535	13.71	14.28	***0.010***	35.33	33.29
GRATE	0.4062	0.644	0.620	0.553	1.393	1.404	***0.0375***	***2.51***	2.377	0.003	0.976	0.847
GPROP	0.2490	0.226	0.215	0.897	0.358	0.380	0.1310	0.881	0.867	0.112	0.496	0.471
GRICH	0.3626	0.0045	0.0044	0.027	0.0076	0.0067	0.4174	0.036	0.035	0.523	0.023	0.023

Shown are PLINK analyses of genome-wide rare (<1%) CNV counts for ***cases*** with LCL- vs. Blood-derived DNA after QC, for Narrow and Broad call criteria, deletions and duplications, and large (>100kb) or small (<100kb) CNVs -- after excluding CNVs > 4 Mb and CNVs in the established association regions shown in Table 1 of the main text. Ns are total CNVs for all subjects (3,002 with LCL and 943 with blood DNA for Narrow; 2997 and 936 for Broad, where a few additional subjects were excluded from analyses) . Nominal empirical (permutation-based) P-values are shown. LCL and Blood columns show mean values (per subject) for each variable. RATE=CNVs/subject; PROP=proportion of subjects with at least 1 CNV; KBTOT = kb of CNV/subject; KBAVG = mean CNV size; GRATE = number of genes spanned per CNV; GPROP = proportion of CNVs with at least one gene; GRICH = number of genes per total CNV kb.

After QC (excluding subjects and CNVs with possible LCL effects such as large numbers of CNVs or low intensity variance), there were no LCL-blood differences for large CNVs. For small CNVs, LCL DNAs had 4.5% more deletions (by the Narrow call criteria used in our analyses), but not more duplications. The effect was more pronounced for Broad small deletions. Thus the effect was from small deletion calls, and particularly those with lower lod scores, possibly due to lower-variance specimens which did not meet the exclusion threshold. The ratios of LCL:Blood CNV rates were 1.045 for Narrow and 1.085 for Broad small deletions.

As shown in Table S9 below, the excess of genic CNVs in cases was limited to large deletions, for which no LCL-blood difference was observed. However, because more controls than cases had LCL specimens, the small excess of small deletions in LCL specimens would make the case:control analyses of small deletions slightly conservative. Examination of pointwise data for LCL specimens did not reveal additional associations than were reported here.

Table S5: qPCR results

		-------N(CNVs) -------		----N probes in CNVs----		Probes tested in controls	
CNV regions	**N probes‡**	**Predicted**	**Confirmed**	**Predicted**	**Inconsistent**	**Total**	**Del/Dup calls**
1q21	6	12	12	239	11	239	11
NRXN1 (exonic dels)	3	10	10	14	0	51	0
15q13.2 dup and del*	17	12	12	164	13	204	1
16p11.2 dup and del**	12	18	18	191	20	328	7
22q11.21 del	4	18	18	76	0	159	22
VIPR2	2	12	12	24	1	42	0
3q29	3	5	5	15	0	30	0
3p26.1 (intergenic)#	2	13	13	26	0	42	1
3q26.1 (intergenic)	2	5	5	9	0	42	5
NEDD4L	2	7	7	14	0	42	2
CGNL1#	2	13	13	26	5	42	0
DLG2_del#	2	4	4	4	0	42	1
WWOX#	2	6	6	12	0	42	5
Totals		135	135	814	50	1305	55
Rate			100.0%		6.1%		4.2%

‡ - Number of probes tested within each CNV region. Not all CNVs were predicted to contain all tested probes. Three 22q dels and one NRXN1 exonic del were tested with only 1 probe, and one 16p dup with 2 probes.
* qPCR and Birdseye boundaries differed in 11/12 subjects .
** 13 of the inconsistent results in CNV cases were for the leftmost probe, indicating that Birdseye called a wider CNV than was confirmed.
# - CNVs that were subsequently dropped from analysis after further examination including comparison with other datasets.

CNVs in 13 regions were tested with qPCR as summarized below. Shown are: the number of CNVs predicted and confirmed (with at least one probe, typically with all or almost all probes) in each region (100% confirmation rate); the number of probes predicted to lie within CNVs and the number of qPCR assays that were inconsistent with prediction (6.1%); and the number of individual probe assays tested in controls (20-40 per region) where predicted CN was 2, and the number in which deletion or duplication calls were made by qPCR (4.2% inconsistent with prediction).

Methods summary: In selected regions, copy number was assayed for case and control CNVs by quantitative PCR. In brief, each assay was run as a duplex real-time PCR reaction (10ul), with a FAM-labeled assay for the target sequence and VIC-labeled endogenous control assay (human RNase P). Real-time PCR was performed in 384-well plates on an ABI 7900 instrument. Each sample was assayed with 4 replicates. The relative quantity of target sequence vs. reference probe (VIC-RNase P) in each sample was determined by ΔCt (FAM Ct-VIC Ct). Using median ΔCt value of > 10 samples with predicted CN=2 as two-copy reference genome (or calibrator), we calculated the relative DNA quantity between a sample and the reference (calibrator) as $\Delta\Delta$Ct = [(FAM Ct-VIC Ct) sample] – [(FAM Ct- VIC Ct) calibrator], from which the copy number was estimated with a formula 2 x 2 ^-$\Delta\Delta$Ct. (Sharp et al., 2008; Xu et al., 2008) We defined a duplication or deletion when qPCR estimated the CN as >2.5 or <1.5 respectively, based on the distribution of estimated copy numbers from 1224 qPCR data points for reference subjects with predicted CN=2.

Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R, Broomer A, Casuga I, Wang Y, Xiao C, Barbacioru C, Gimelli G, Bernardina BD, Torniero C, Giorda R, Regan R, Murday V, Mansour S, Fichera M, Castiglia L, Failla P, Ventura M, Jiang Z, Cooper GM, Knight SJ, Romano C, Zuffardi O, Chen C, Schwartz CE, Eichler EE. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. Nat Genet. 2008 Mar;40(3):322-8. Epub 2008 Feb 17. PubMed PMID: 18278044

Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet. 2008 Jul;40(7):880-5. Epub 2008 May 30. PubMed PMID: 18511947.

Table S6: Suggestive point-wise results and comparison with ISC dataset

Gene/region	Chr	Location (bp)	Effect	Best MGS Case	Best MGS Cont	Best ISC Case	Best ISC Cont	Comments
MAST2/PIK3R3	1	46,041,871-46,371,295	dups	5	0	0	2	Case:control ratio (dups) 8:1 in MAST2, 4:0 in PIK3R3
3q26.1 intergenic	3	165,606,061-165,655,524	all dels	5	0	-	-	Multiple TFBS including POU2F1 (6-66kb; 1 large ISC case del spans region)
3q29 (DLG1/BDH1)	3	197,203,162-198,825,243	large dels	5	0	2	0	~1.6 Mb deletions affecting 21 genes; CHOP: 3 smaller dels in DLG1
ZNF595/ZNF718	4	~60,000-192,000	all dels	5	0	6	0	No evidence in ISC for association of exonic CNVs
PARK2	6	161,688,579-163,068,824	large dels all dups	8 5	1 0	6 6	1 1	Exonic dups and dels have low ORs in both datasets
IMMP2L	7	110,090,345-110,989,583	large dels	9	2	4	0	Exonic dels have low ORs in both datasets
CNTNAP2	7	145,444,385-147,749,019	all dels	5	0	2	0	7:3 for exonic deletions (MGS+ISC)
VIPR2	7	158,513,626-158,630,410	large dups	7	1	4	0	MGS and ISC best signals at same location
DDX10	11	108,041,025-108,316,858	all dups	5	0	0	0	No signal for exonic dups (2:1)
RNASE3	14	20,429,401-20,430,347	all dups	9	1	0	0	Common dels and dups also affect this gene with ORs < 2
CGNL1	15	55,455,996-55,630,213	large dups	11	2	8	2	Signals in different locations; lower OR for exonic dups in ISC
C16orf72	16	9,093,037-9,121,056	all dups	10	0	1	0	All MGS and ISC dups affect exons
NOMO3	16	16,233,889-16,296,168	large dels	11	3	3	0	SDs throughout region; reduced del freq in LCLs
16p intergenic	16	~18,150,000-18,500,00	large dels	9	1	0	3	Consistency of CNV calls across methods is weaker in this region
WWOX	16	76,691,051-77,804,065	all dups	6	0	0	0	Downstream of gene; multiple TFBS
NEDD4L	18	53,862,777-54,216,369	all dups	5	0	2	0	MGS and ISC best signals in different locations
TEX101	19	48,584,602-48,614,607	all dups	5	0	0	0	All MGS dups affect exons

TBFS - transcription factor binding sites; OR - odds ratio

Shown are pointwise results that achieved empirical suggestive genome-wide significance (expected less than once per genome-wide study) in the MGS dataset as computed by PLINK. The same analysis was carried out in the ISC dataset and the best result in each gene or region is shown, but note that (a) only CNVs > 100kb were publicly available for ISC while some of the MGS CNVs were smaller; and (b) the best point was usually not the same in the two datasets. For chromosomal regions, numbers of CHOP subjects with a CNV in that region are shown, whereas for individual genes, CHOP data were included in the analyses of exonic CNVs in the most promising regions (see main text, Table 3). Some of these regions represent additional candidates that deserve exploration in meta-analyses of larger combined datasets.

Table S7: HG18 locations and descriptions of genes in schizophrenia-associated multigenic CNVs

Gene Symbol	Chr	StartBP	EndBP	Description
MGS CNV candidate genes				
VIPR2	7	158513626	158630410	vasoactive intestinal peptide receptor 2
CSMD3	8	113304332	114518418	CUB and Sushi multiple domains 3
AGTPBP1	9	87351273	87546764	ATP/GTP binding protein 1
GLB1L3	11	133651484	133694668	galactosidase, beta 1-like 3
GLB1L2	11	133707018	133751428	galactosidase, beta 1-like 2
C16orf72	16	9093037	9121056	chromosome 16 open reading frame 72
NEDD4L	18	53862777	54216369	neural precursor cell expressed, developmentally down-regulated 4-like
3q29 large deletion candidate region (3q29 microdeletion syndrome)				
TFRC	3	197260551	197293429	transferrin receptor (p90, CD71)
ZDHHC19	3	197408719	197422697	zinc finger, DHHC-type containing 19
OSTalpha	3	197427779	197444698	organic solute transporter alpha
PCYT1A	3	197449649	197498981	phosphate cytidylyltransferase 1, choline, alpha
TCTEX1D2	3	197502494	197529542	Tctex1 domain containing 2
TM4SF19	3	197534815	197549655	transmembrane 4 L six family member 19
UBXD7	3	197564765	197643742	UBX domain protein 7
RNF168	3	197683024	197714979	ring finger protein 168
C3orf43	3	197718146	197726634	chromosome 3 open reading frame 43
WDR53	3	197765455	197779810	WD repeat domain 53
FBXO45	3	197780121	197800327	F-box protein 45
LRRC33	3	197851052	197873271	leucine rich repeat containing 33
C3orf34	3	197917544	197923520	chromosome 3 open reading frame 34
PIGX	3	197923642	197947273	phosphatidylinositol glycan anchor biosynthesis, class X
PAK2	3	197951124	198043915	p21 protein (Cdc42/Rac)-activated kinase 2
SENP5	3	198079123	198145981	SUMO1/sentrin specific peptidase 5
NCBP2	3	198146669	198153861	nuclear cap binding protein subunit 2, 20kDa
PIGZ	3	198157610	198180101	phosphatidylinositol glycan anchor biosynthesis, class Z
MFI2	3	198214552	198241083	antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5
DLG1	3	198253827	198509844	discs, large homolog 1 (Drosophila)
BDH1	3	198721050	198784591	3-hydroxybutyrate dehydrogenase, type 1
1q21.1 schizophrenia-associated deletion region				
FAM108A3	1	144786832	144791590	family with sequence similarity 108, member A3
PRKAB2	1	145093308	145110753	protein kinase, AMP-activated, beta 2 non-catalytic subunit
FMO5	1	145124461	145163546	flavin containing monooxygenase 5
CHD1L	1	145180914	145234067	chromodomain helicase DNA binding protein 1-like
BCL9	1	145479805	145564639	B-cell CLL/lymphoma 9
ACP6	1	145585791	145609258	acid phosphatase 6, lysophosphatidic
GJA5	1	145694955	145712108	gap junction protein, alpha 5, 40kDa
GJA8	1	145841559	145848017	gap junction protein, alpha 8, 50kDa
GPR89B	1	145867129	145932377	G protein-coupled receptor 89B
GPR89C	1	145892190	145932379	G protein-coupled receptor 89C
NBPF11	1	146040946	146076705	neuroblastoma breakpoint family, member 11
15q13.3 schizophrenia-associated deletion region				
CHRFAM7A	15	28440734	28473156	CHRNA7 (cholinergic receptor, nicotinic, alpha 7, exons 5-10) and FAM7A (family with sequence similarity 7A, exons A-E) fusion
ARHGAP11B	15	28706170	28718305	Rho GTPase activating protein 11B
MTMR15	15	28983420	29022600	myotubularin related protein 15
MTMR10	15	29018435	29071099	myotubularin related protein 10
TRPM1	15	29080842	29181216	transient receptor potential cation channel, subfamily M, member 1
KLF13	15	29406374	29457394	Kruppel-like factor 13
OTUD7A	15	29562620	29734834	OTU domain containing 7A
CHRNA7	15	30110017	30248527	cholinergic receptor, nicotinic, alpha 7

16p11.2 schizophrenia-associated duplication region (note - 3 genes shaded in gray are duplicated in each flanking SD region)				
BOLA2	16	29,362,071	29,373,786	BolA-like protein 2
GIYD2	16	29,373,376	29,377,041	GIY-YIG domain containing 2
SULT1A4	16	29,373,902	29,383,783	sulfotransferase family, cytosolic, 1A, phenol-preferring, member 4
SPN	16	29,581,801	29,589,329	sialophorin (leukosialin, CD43)
QPRT	16	29,597,942	29,616,816	nicotinate-nucleotide pyrophosphorylase (carboxylating)
C16orf54	16	29,661,285	29,664,841	chromosome 16 open reading frame 54
KIF22	16	29,709,559	29,724,207	kinesin family member 22
MAZ	16	29,725,356	29,730,005	MYC-associated zinc finger protein (purine-binding transcription factor)
PRRT2	16	29,730,910	29,734,703	proline-rich transmembrane protein 2
MVP	16	29,731,591	29,766,842	major vault protein
C16orf53	16	29,735,029	29,741,317	chromosome 16 open reading frame 53
CDIPT	16	29,777,179	29,782,079	CDP-diacylglycerol--inositol 3-phosphatidyltransferase (phosphatidylinositol synthase)
SEZ6L2	16	29,790,329	29,818,074	seizure related 6 homolog (mouse)-like 2
ASPHD1	16	29,819,648	29,824,878	aspartate beta-hydroxylase domain containing 1
KCTD13	16	29,825,164	29,845,046	potassium channel tetramerisation domain containing 13
TMEM219	16	29,880,852	29,891,874	transmembrane protein 219
TAOK2	16	29,892,723	29,911,082	TAO kinase 2
HIRIP3	16	29,911,817	29,914,888	HIRA interacting protein 3
INO80E	16	29,915,032	29,924,613	coiled-coil domain containing 95
DOC2A	16	29,924,336	29,929,902	double C2-like domains, alpha
C16orf92	16	29,942,156	29,943,524	chromosome 16 open reading frame 92
FAM57B	16	29,943,249	29,949,687	family with sequence similarity 57, member B
ALDOA	16	29,971,992	29,989,236	aldolase A, fructose-bisphosphate
PPP4C	16	29,994,886	30,004,196	protein phosphatase 4 (formerly X), catalytic subunit
TBX6	16	30,004,583	30,010,709	T-box 6
YPEL3	16	30,011,136	30,015,022	yippee-like 3 (Drosophila)
GDPD3	16	30,023,632	30,032,379	glycerophosphodiester phosphodiesterase domain containing 3
MAPK3	16	30,032,927	30,042,131	mitogen-activated protein kinase 3
CORO1A	16	30,102,427	30,107,898	coronin, actin binding protein, 1A
BOLA2B	16	30,111,757	30,113,128	bolA homolog 2B (E. coli)
GIYD1	16	30,112,718	30,116,383	GIY-YIG domain containing 1
SULT1A3	16	30,113,244	30,123,151	sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3
22q11.21 schizophrenia-associated deletion region				
DGCR6	22	17273735	17279601	DiGeorge syndrome critical region gene 6
PRODH	22	17280293	17303814	proline dehydrogenase (oxidase) 1
DGCR2	22	17403794	17489967	DiGeorge syndrome critical region gene 2
DGCR14	22	17497791	17512190	DiGeorge syndrome critical region gene 14
TSSK2	22	17498320	17500136	testis-specific serine kinase 2
GSC2	22	17516503	17517796	goosecoid homeobox 2
SLC25A1	22	17543094	17546285	solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1
CLTCL1	22	17546986	17659239	clathrin, heavy chain-like 1
HIRA	22	17698223	17799219	HIR histone cell cycle regulation defective homolog A (S. cerevisiae)
MRPL40	22	17800035	17803596	mitochondrial ribosomal protein L40
C22orf39	22	17810894	17815220	chromosome 22 open reading frame 39
UFD1L	22	17817700	17846726	ubiquitin fusion degradation 1 like (yeast)
CDC45L	22	17847415	17888135	CDC45 cell division cycle 45-like (S. cerevisiae)
CLDN5	22	17890549	17892860	claudin 5
SEPT5	22	18081986	18092297	septin 5
GP1BB	22	18091065	18092297	glycoprotein Ib (platelet), beta polypeptide
TBX1	22	18124225	18151112	T-box 1
GNB1L	22	18155933	18222462	guanine nucleotide binding protein (G protein), beta polypeptide 1-like
C22orf29	22	18213660	18222371	chromosome 22 open reading frame 29
TXNRD2	22	18243039	18309359	thioredoxin reductase 2
COMT	22	18309308	18336530	catechol-O-methyltransferase
ARVCF	22	18337418	18384309	armadillo repeat gene deleted in velocardiofacial syndrome
C22orf25	22	18388630	18433447	chromosome 22 open reading frame 25

DGCR8	22	18447833	18479400	DiGeorge syndrome critical region gene 8
HTF9C	22	18479397	18484768	TRM2 tRNA methyltransferase 2 homolog A (S. cerevisiae) (formerly HTF9C)
RANBP1	22	18485023	18494704	RAN binding protein 1
ZDHHC8	22	18499364	18513974	zinc finger, DHHC-type containing 8
RTN4R	22	18608937	18635816	reticulon 4 receptor
DGCR6L	22	18681799	18687608	DiGeorge syndrome critical region gene 6-like
RIMBP3	22	18835993	18841786	RIMS binding protein 3
ZNF74	22	19078479	19092752	zinc finger protein 74
SCARF2	22	19108874	19122146	scavenger receptor class F, member 2
KLHL22	22	19125805	19180122	kelch-like 22 (Drosophila)
MED15	22	19191885	19271919	mediator complex subunit 15
PI4KA	22	19391978	19543070	phosphatidylinositol 4-kinase, catalytic, alpha
SERPIND1	22	19458382	19472008	serpin peptidase inhibitor, clade D (heparin cofactor), member 1
SNAP29	22	19543291	19574109	synaptosomal-associated protein, 29kDa
CRKL	22	19601713	19637890	v-crk sarcoma virus CT10 oncogene homolog (avian)-like
AIFM3	22	19649433	19665649	apoptosis-inducing factor, mitochondrion-associated, 3
LZTR1	22	19666557	19683326	leucine-zipper-like transcription regulator 1
THAP7	22	19684060	19686404	THAP domain containing 7
P2RX6	22	19699463	19712302	purinergic receptor P2X, ligand-gated ion channel, 6
SLC7A4	22	19713006	19716847	solute carrier family 7 (cationic amino acid transporter, y+ system), member 4
RIMBP3C	22	20067662	20073455	RIMS binding protein 3C
RIMBP3B	22	20068039	20073455	RIMS binding protein 3B
HIC2	22	20101692	20135750	hypermethylated in cancer 2

Table S8: Case-control analysis of exonic CNVs: Genes with genome-wide suggestive significance in the MGS sample

DELETIONS								DUPLICATIONS							
CHR	GENE	BP1	BP2	AFF	UN	EMP1	EMP2	CHR	GENE	BP1	BP2	AFF	UN	EMP1	EMP2
11	GLB1L3	133651484	133694668	15	3	0.0058	0.6486	19	ZNF600	57960559	57981846	5	0	0.0384	0.9990
2	NRXN1	50000991	51113178	10	1	0.0089	0.7137	9	ZNF658B	41578832	41582207	5	0	0.0402	0.9990
11	GLB1L2	133707018	133751428	14	3	0.0103	0.8517	1	C1orf25*	183,353,840	183,392,739	15	3	0.0057	0.6875
3	BDH1	198721050	198784591	8	1	0.0259	0.9878	1	C1orf26*	183,392,913	183,527,536	15	3	0.0057	0.6875
2	VWA3B	98070026	98295842	16	5	0.0222	0.9889	1	OR2T12	246,524,540	246,525,503	10	1	0.0093	0.7507
3	TFRC	197260551	197293429	5	0	0.0380	0.9990	1	OR2M7	246,553,554	246,554,493	10	1	0.0093	0.7507
3	ZDHHC19	197408719	197422697	5	0	0.0380	0.9990	2	RBM44	238,372,126	238,408,253	6	0	0.0208	0.9828
3	OSTalpha	197427779	197444698	5	0	0.0380	0.9990	2	RAMP1	238,432,925	238,485,494	6	0	0.0208	0.9828
3	PCYT1A	197449649	197498981	5	0	0.0380	0.9990	15	CGNL1	55,455,996	55,630,213	11	2	0.0156	0.9851
3	TCTEX1D2	197502494	197529542	5	0	0.0380	0.9990	1	MAST2	46,041,871	46,274,383	8	1	0.0256	0.9955
3	TM4SF19	197534815	197549655	5	0	0.0380	0.9990	9	GLDC	6,522,463	6,635,692	8	1	0.0258	0.9955
3	UBXD7	197564765	197643742	5	0	0.0380	0.9990	7	VIPR2	158,513,626	158,630,410	10	2	0.0260	0.9977
3	RNF168	197683024	197714979	5	0	0.0380	0.9990	21	BAGE2	10,042,712	10,120,796	10	2	0.0273	0.9977
3	C3orf43	197718146	197726634	5	0	0.0380	0.9990	1	OR2T33	246,502,776	246,503,739	10	2	0.0275	0.9977
3	WDR53	197765455	197779810	5	0	0.0380	0.9990	5	BTNL3	180,348,506	180,366,333	12	3	0.0271	0.9980
3	FBXO45	197780121	197800327	5	0	0.0380	0.9990	9	AGTPBP1	87,351,273	87,546,764	5	0	0.0390	1.0000
3	LRRC33	197851052	197873271	5	0	0.0380	0.9990	8	CSMD3**	113,304,332	114,518,418	5	0	0.0390	1.0000
3	C3orf34	197917544	197923520	5	0	0.0380	0.9990								
3	PIGX	197923642	197947273	5	0	0.0380	0.9990								
3	PAK2	197951124	198043915	5	0	0.0380	0.9990								
3	SENP5	198079123	198145981	5	0	0.0380	0.9990								
3	NCBP2	198146669	198153861	5	0	0.0380	0.9990								
3	PIGZ	198157610	198180101	5	0	0.0380	0.9990								
3	MFI2	198230220	198241083	5	0	0.0380	0.9990								

* 23:11 for Broad CNV calls.
** 9:3 for Broad CNV calls.

Shown are the results of case-control association analysis of CNVs affecting exons in RefSeq genes. Locations of genes and of exons within genes were based on HG18 download files. The PennCNV utility script “scan_region.pl” was used to identify MGS CNVs (from the post-QC file of Narrow-criteria CNVs of all frequencies) that overlapped with any exon.

After excluding non-exonic CNVs, gene-wise case-control analysis was then carried out with PLINK, which computed two empirical p-values for each gene based on 10,000 permutations of case-control status. Genes with deletions or duplications in > 0.5% of all subjects were excluded. EMP1 is a pointwise p-value (how often would this case-control ratio be observed at this location), and EMP2 is a genome-wide p-value (how often would a case-control ratio this extreme be observed anywhere in the genome).

All genes with EMP2<1 (expected less than once per genome-wide analysis) are shown in the table. The shaded rows are the 21 genes in the 3q29 microdeletion region. Genes whose association with schizophrenia was supported after inclusion of ISC and CHOP data are shown in Table 3 in the main text. Some loci were excluded from consideration because of much weaker results using Broad calls or significant differences between cell line vs. blood DNA in cases.

Table S9: Analyses of number of genome-wide CNVs per subject ("global burden")

a. Deletions

Type of Deletion	Type of CNV	Tested effect	All (<1%)		Cell line DNA only		Singleton CNVs		>500kb CNVs		>1Mb CNVs	
			Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Large (>100kb or larger as shown)	**All**	CNVs/subject	0.535	0.512	0.537	0.511	0.118	0.107	0.061	0.069	0.020	0.019
		Subjects w. CNV	0.402	0.397	0.405	0.397	0.109	0.101	0.058	0.065	0.020	0.019
	Gene-containing	Genes/CNV	***0.628****	0.531	***0.630****	0.524	0.028	0.024	0.221	0.180	***0.137****	0.070
		Prop genic CNVs	***0.225****	0.20	***0.228****	0.205	0.027	0.024	0.038	0.041	***0.0124****	0.0083
		Genes/CNV kb	***0.0048****	0.0042	***0.0048****	0.0042	0.0015	0.0014	0.0033	0.0031	***0.0043****	0.002
	Exon-containing	CNVs/subject	***0.304****	0.282	0.303	0.281	***0.075****	0.063	0.049	0.054	0.013	0.009
		Subject w. CNV	0.255	0.243	0.259	0.242	0.069	0.061	0.046	0.052	***0.0129****	0.0086
Small (<100kb)	**All**	CNVs/subject	4.281	4.215	***4.326****	4.216	0.907	0.883				
		Subject w. CNV	0.962	0.959	0.962	0.959	0.577	0.574				
	Gene-containing	Genes/CNV	1.381	1.367	1.383	1.365	0.273	0.278				
		Prop genic CNVs	0.690	0.688	0.690	0.688	0.235	0.235				
		Genes/CNV kb	0.026	0.030	0.026	0.030	0.054	0.048				
	Exon-containing	CNVs/subject	0.630	0.630	0.625	0.627	0.190	0.188				
		Subjects w. CNV	0.467	0.457	0.462	0.456	0.171	0.168				

b. Duplications

Type of Duplication	Type of CNV	Tested effect	All (<1%)		Cell line DNA only		Singleton CNVs		>500kb CNVs		>1Mb CNVs	
			Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Large (>100kb or larger as shown)	**All**	CNVs/subj	0.658	0.626	0.648	0.626	0.155	0.146	0.110	0.102	0.027	0.032
		Subjs w CNV	0.469	0.444	0.463	0.444	0.137	0.129	0.103	0.096	0.026	0.030
	Exon-containing	CNVs/subj	0.505	0.479	0.499	0.479	0.117	0.113	0.091	0.088	0.022	0.028
		Subjs w CNV	0.387	0.369	0.381	0.369	0.108	0.103	0.086	0.084	0.022	0.027
Small (<100kb)	**All**	CNVs/subj	1.152	1.135	1.145	1.134	0.338	0.325				
		Subjs w CNV	0.675	0.664	0.671	0.663	0.282	0.270				
	Exon-containing	CNVs/subj	0.467	0.475	0.478	0.474	0.147	0.140				
		Subjs w CNV	0.365	0.378	0.369	0.377	0.134	0.132				

* empirical P < 0.05; ** empirical P < 0.01 (uncorrected p-values)

Shown are results of analyses of the overall frequency of deletions and of duplications in cases vs. controls, separately for larger and smaller (<100kb) CNVs; and for all CNVs, those overlapping with at least one gene (for deletions), and those overlapping with at least one exon. CNVs observed in > 1% of subjects were excluded. Separate analyses considered all deletions; and deletions greater than 500 kb or 1 Mb in length. Because lymphoblastic cell line transformation can create structural variants, we also separately analyzed cases vs. controls with DNA extracted from cell lines. All analyses excluded the five associated CNVs shown in Table 2, and CNVs longer than 4 Mb (which were observed primarily in specimens from cell lines, and may represent artifacts). Note that cases did not have significantly larger CNVs than controls. ***Case-control differences were for large deletions, for which no difference was observed between LCL and blood DNAs (Table S4b).***

Table S9c: Global analyses of Narrow large deletions in EA and AA subsamples

	EA narrow large dels			AA narrow large dels		
	EMP1	AFF	UNAFF	EMP1	AFF	UNAFF
N CNVs		1401	1367		708	480
RATE	0.3544	0.525	0.516	0.0429	0.556	0.499
PROP	0.7543	0.395	0.403	0.0468	0.416	0.380
KBTOT	0.5869	383.8	388	0.5878	354.8	360.4
KBAVG	0.8530	284.3	299.4	0.8016	260	273.9
GRATE	*0.0520*	0.656	0.568	*0.0601*	0.569	0.430
GPROP	*0.0881*	0.233	0.217	***0.0299***	0.209	0.176
GRICH	***0.0167***	0.0051	0.0044	***0.0395***	0.0043	0.0036

RATE=CNVs/subject; PROP=proportion of subjects with at least 1 CNV; KBTOT = kb of CNV/subject; KBAVG = mean CNV size; GRATE = number of genes spanned per CNV; GPROP = proportion of CNVs with at least one gene; GRICH = number of genes per total CNV kb.

Separate analysis of EA and AA subsamples show that the direction of effects for genic CNVs was similar in the EA and AA subsamples, although AA cases showed an overall increase in CNVs while for EA cases the effect was limited to genic CNVs. All of these effects are small and some could be due to chance. The increase in the proportion of CNVs spanning at least one gene was 1.07 for EA and 1.19 for AA cases.

Supplementary Methods:
Estimating genome-wide significant and suggestive thresholds for rare CNVs

There is no generally-accepted threshold of statistical significance for genome-wide CNV studies. Reasonable estimates of genome-wide thresholds have been made for genome-wide linkage studies based on the structure of recombination in the genome (1) and for genome-wide association studies of ***common*** SNPs based on their linkage disequilibrium structure.(2-4) But for rare CNVs, individual events cannot be easily collapsed into a set of discrete categories: in some regions there are multiple individuals with CNVs with nearly identical boundaries, but in other regions there are multiple subsets of partially-overlapping and non-overlapping CNVs with diverse lengths and boundaries (whether in the same gene or affecting different sets of genes) which makes counting the "number of tests" difficult; and there is no straightforward correlational structure for estimating the burden of multiple testing. Further, GWAS arrays imperfectly measure the boundaries of rare CNVs.

Within a single dataset such as MGS, some estimates can be made empirically, by re-computing association statistics for all events after randomly permuting case-control status many times. PLINK does this for the pointwise and gene-based analyses used here. The *pointwise* statistic is based on the numbers of cases and controls whose deletions or duplications overlap each of a set of points that include the start and end positions of all CNVs in the dataset, and 1 bp beyond each stop position. The pointwise statistic is useful for identifying regions of interest, but does not identify "CNVs" as such (i.e., categorial classes of events), because CNVs with diverse lengths and which span diverse genomic features can overlap with the same point. The *gene-based* analysis is useful for regions (like NRXN1) where all or most CNVs affect a specific gene or set of genes. But in some regions with long, multigenic CNVs, shorter CNVs affect some of the genes, and thus a set of gene-based tests does not include a specific test of, for example, "15q13.3 deletions that span a specific 1.5 Mb region", which turns out to be the class of interest. Such events have been identified essentially by inspection of complex data, with subsequent confirmation of the hypothesis. Another problem with empirical p-values is that when one combines data from multiple studies for specific CNVs of interest, one generally lacks comparable genome-wide data across studies (i.e., genome-wide results are often not available, and when they are, there are still many methodological differences between the CNV datasets).

While we have no ideal solution to this problem, we have applied a rather tentative but pragmatic approach to the estimation of thresholds for genome-wide significant association and for suggestive evidence for association (expected once per genome-wide study (1)), as guidelines for interpretation of our results.

To review the procedure for which we wish to define thresholds:

We initially searched for regions of interest with PLINK's pointwise tests, using uncorrected empirical genome-wide p-values to select regions with P<1 (suggestive association). This is a liberal criterion, because we did this for deletions and duplications separately, and for all detected CNVs and then for CNVs >100kb ("large") within each type. However, single datasets cannot provide definitive results for most rare CNVs, so we err on the side of more liberal "suggestive" criteria to identify the best-supported candidate CNVs for testing in other datasets.

We then examined the regions with suggestive findings to identify (i) sets of CNVs that fit the criteria established by previous findings (i.e., the 5 regions shown in Table 3); and (ii) any class of CNVs (such as long, multigenic CNVs) that were present in regions with suggestive

results for one or more points. We computed pooled Fisher's exact tests and stratified CMH exact for these sets of CNVs in the MGS dataset, and then after adding any additional available data.

Finally, we selected exonic CNVs with a frequency of <0.5% (with no size restriction) and carried out PLINK's gene-based analysis, separately for deletions and for duplications, for "CNVs that overlap to any degree with any exon in gene 1, 2, 3..." We then used uncorrected genome-wide empirical p<1 to select genes of interest, and computed Fisher's exact tests in all available data for exonic CNVs in those genes (Table 2). Again, the uncorrected empirical p-values were used to avoid prematurely rejecting hypotheses that deserve testing in larger datasets.

What are reasonable significant and suggestive thresholds for the Fisher's exact tests of data from several datasets? To obtain a rough estimate, we applied the "--segment-group" command in PLINK to our files of rare, exonic deletions and duplications. This command "takes all segments in a given region (whole genome unless otherwise specified) and forms 'pools' of overlapping segments. Several pools of overlapping segments will be created; these will be listed in order of decreasing size (number of segments); note that the same segment can appear in multiple pools (e.g. if A overlaps with C, and B overlaps with C, but A and B do not overlap)." (PLINK manual, http://pngu.mgh.harvard.edu/~purcell/plink/cnv.shtml) We reasoned that this count would roughly correspond to the number of groupings that might attract our attention when analyzing and reviewing data as described above, although it will overestimate the count in one respect (i.e., smaller pools cannot yield genome-wide suggestive results even if only cases are affected) and undestimate it in other respects (i.e., for longer genes, CNVs of interest, such as exonic CNVs, will not all overlap; and the groupings do not always define the precise set of CNVs that prove to be of interest, such as "1.5 Mb 15q13.3 deletions", so that there is additional "testing" involved in manually reviewing the data to look for those subsets).

For CNVs of all sizes (by our Narrow criteria), we observed 2,317 "pools" (1,006 for deletions and 1,311 for duplications), or 846 after omitting those with less than 5 CNVs in the pool (which is the lowest frequency that we analyzed). For large CNVs, there were 974 pools (280 for deletions and 694 for duplications), or 374 with 5 or more CNVs. Thus if we consider these analyses to be completely independent (which they are not), we would count 1220 pools. In gene-wise analyses of exonic CNVs with <0.5% frequency, but at least 5 carriers, there were 395 genes that met these criteria for deletions and 665 for duplications, for a total of 1060 analyses, but on average each CNV in this set affected 2.1 genes, for an estimate of 500 independent gene- based tests. Thus there were approximately 1720 analyses, which we round up to 2000 to account for the less formal process of reviewing results for subsets of interest. Dividing 1 (the expectation of observing a finding only once per genome scan) by 2000 gives a ***rough threshold of p=0.0005 for suggestive evidence for association***.

For a more stringent threshold for ***genome-wide significant association***, we divide p=0.05 by 2000 as computed above, which gives a threshold of 2.5×10^{-5}. Whereas we prefer not to over-correct the suggestive threshold, here it might be preferable to be more conservative by assuming that additional correction is required for the informal process of searching for subsets of CNVs within interesting regions. We therefore apply a ***threshold of roughly 10^{-5} for significant association***.

We consider these thresholds as guidelines for the present study (note that the number of pools will differ according to the criteria for calling CNVs and the method used to group them), and we do not apply them strictly, but we comment in the text that p-values for the bolded results shown in Table 3 for deletions in 1q21.1, 15q13.3 and 22q11.21, exonic deletions in NRXN1 and

duplications in 16p11.2 are multiple orders of magnitude lower than this estimated threshold, providing some degree of confidence that they should be considered genome-wide significant.

1. Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat. Genet. 1995; 11(3):241-247
2. Dudbridge F, Gusnanto A: Estimation of significance thresholds for genomewide association scans. Genet Epidemiol 2008; 32(3):227-34
3. Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ: Genome-wide significance for dense SNP and resequencing data. Genet Epidemiol 2008; 32(2):179-85
4. Pe'er I, Yelensky R, Altshuler D, Daly MJ: Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 2008; 32(4):381-5

Figure S1
Plots of CNVs (1q21.1, NRXN1, 15q13.3, 16p11.2, 22q11.21 and VIPR2)

Shown below are plots of CNVs of MGS subjects for CNVs described in Table 3 plus VIPR2 (Figure 1 in the main text shows 3q29 CNVs). In each plot, the location of the CNV boundaries as called by Birdseye, the Birdseye lod score and copy number are shown at the top.

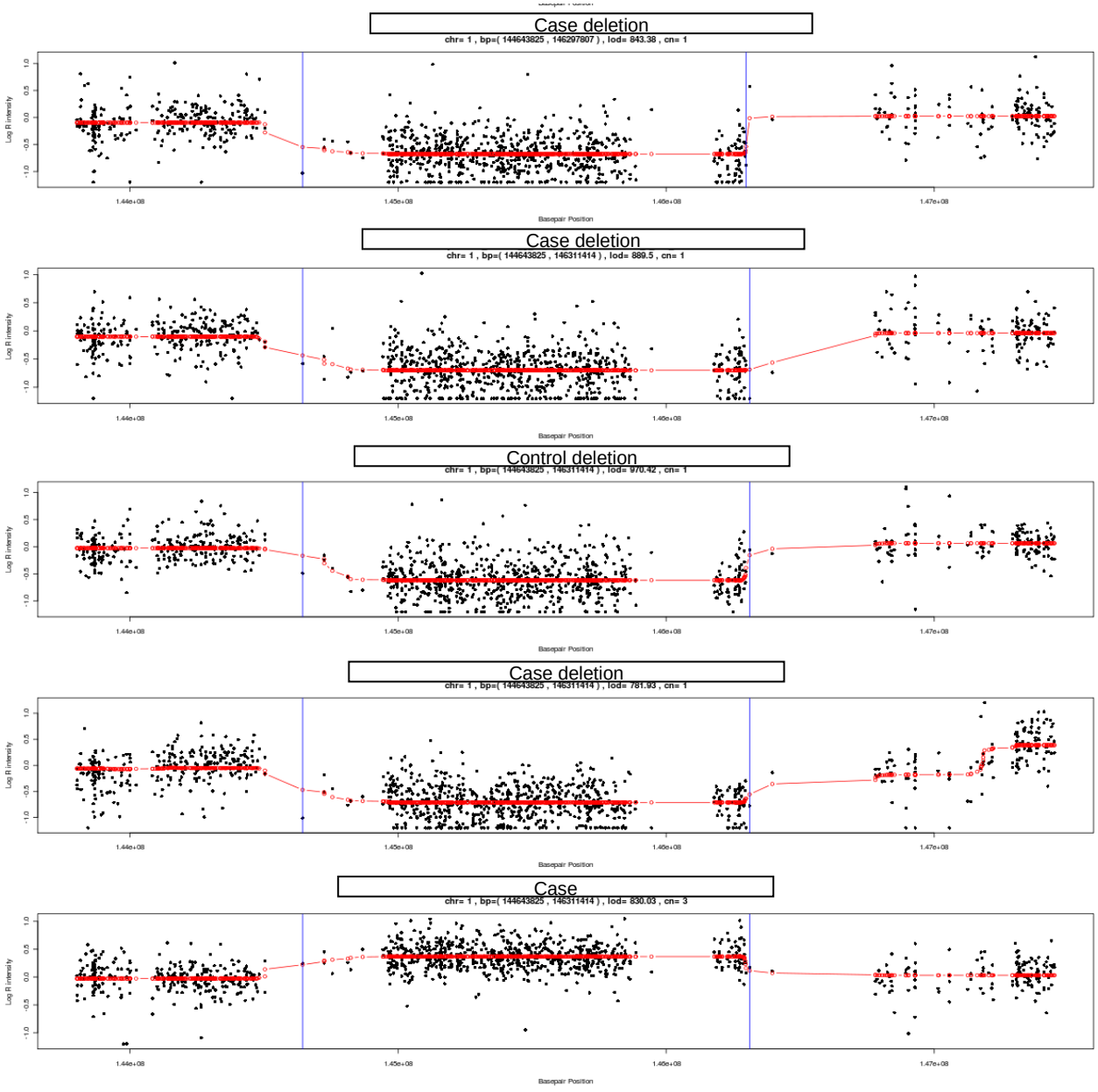
X-axis values are bp locations. Y-axis values are the log [R] ratio, which is the normalized intensity ratio described below -- values of -1.0, -0.5, 0, 0.5 and 1 represent copy numbers of 0, 2, 2, 3 and 4 respectively.

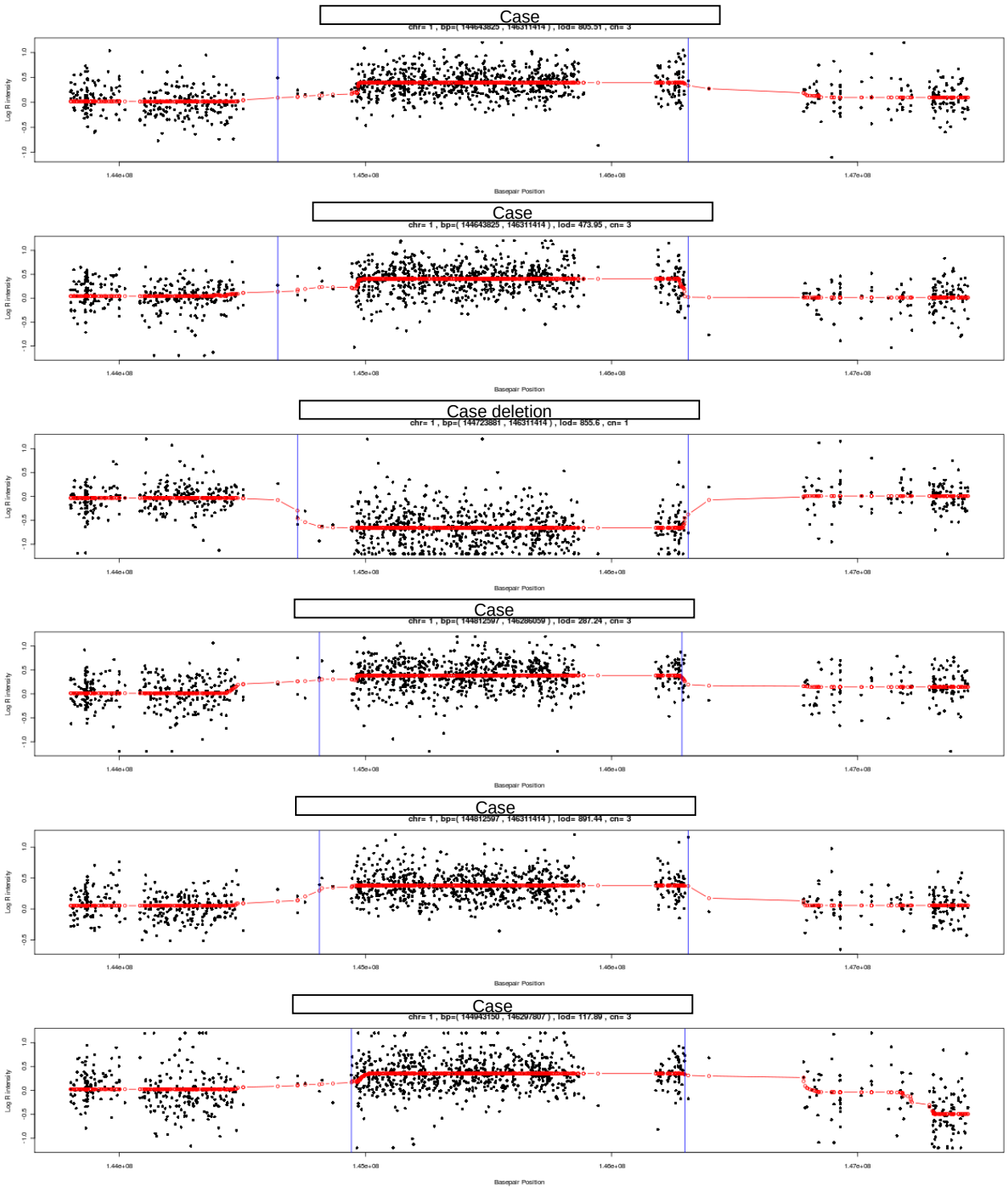
Each individual black dot represents the mean of the fluorescent intensity of all probes for a chromosomal location (8 probes for each SNP - 4 for each allele; and 4 for each monomorphic copy number probe on the Affymetrix 6.0 array), divided by the mean intensity for all specimens on the same DNA plate (which are typically assayed on the same day and thus share small technical variations which influence intensity values).

Blue vertical lines represent the CNV boundaries assigned by Birdseye. The red dots (which look like red lines) are the point-by-point copy number estimates made by an alternative algorithm (Lai et al., ref. 15 in the paper). Usually the two algorithms agree, but in some cases (particularly in segments with few SNPs, generally segmental duplication regions where SNPs are difficult to design) they differ in placement of boundaries.

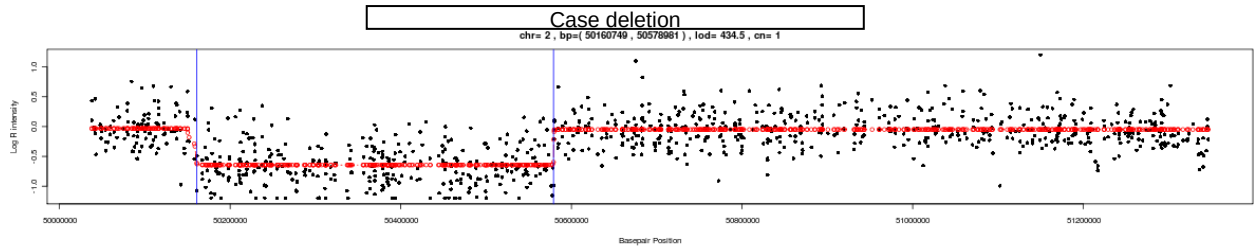
The plots illustrate that these CNVs are well-supported by the intensity data. They also illustrate that the background variation in intensity (noise) is greater than the typical shift in intensity for CN=1 (signal), which is likely to account for much of the differences between calling algorithms as well as for discordancies between duplicate assays (see text).

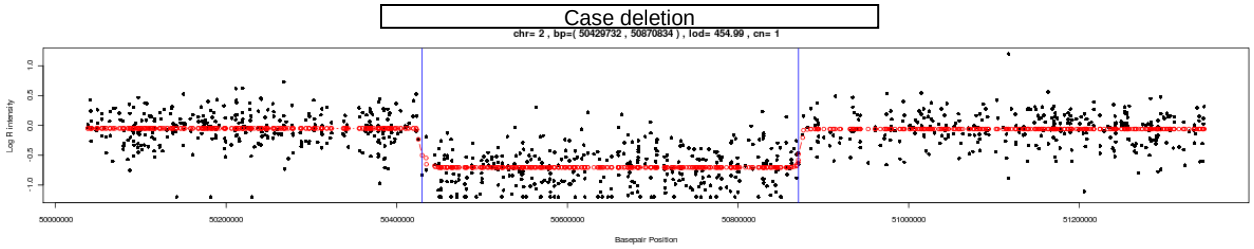
(1) Long deletions and duplications in chromosome 1q21.1

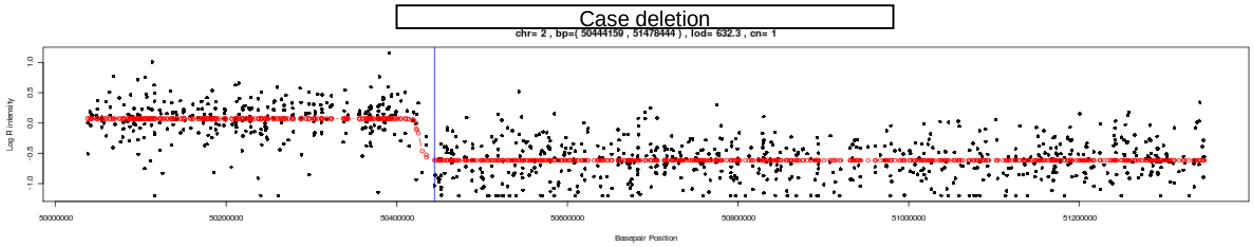


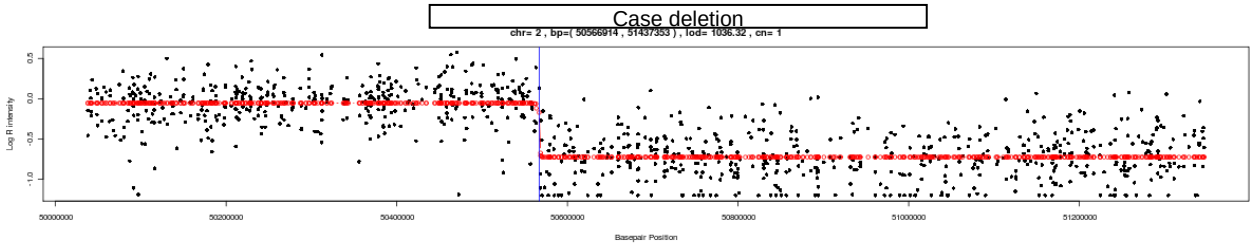

Case
chr= 1 , bp=(144643825 , 146311414) , lod= 805.51 , cn= 3
Case
chr= 1 , bp=(144643825 , 146311414) , lod= 473.95 , cn= 3
Case deletion
chr= 1 , bp=(144723881 , 146311414) , lod= 855.6 , cn= 1
Case
chr= 1 , bp=(144812597 , 146286059) , lod= 287.24 , cn= 3
Case
chr= 1 , bp=(144812597 , 146311414) , lod= 891.44 , cn= 3
Case
chr= 1 , bp=(144943150 , 146297807) , lod= 117.89 , cn= 3
Log R intensity
Basepair Position
1.44e+08
1.45e+08
1.46e+08
1.47e+08

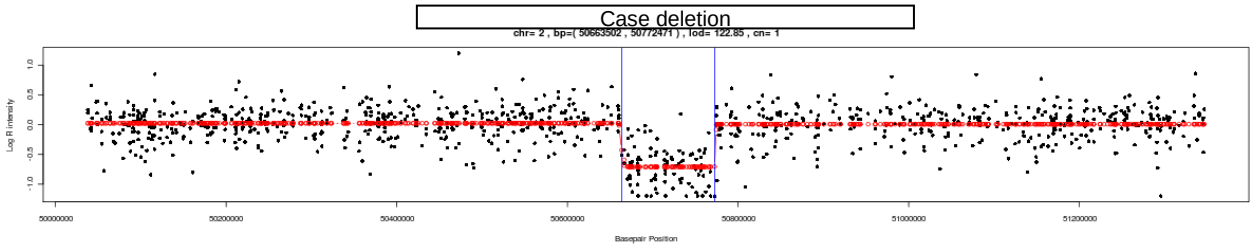
(2) Exonic deletions in NRXN1

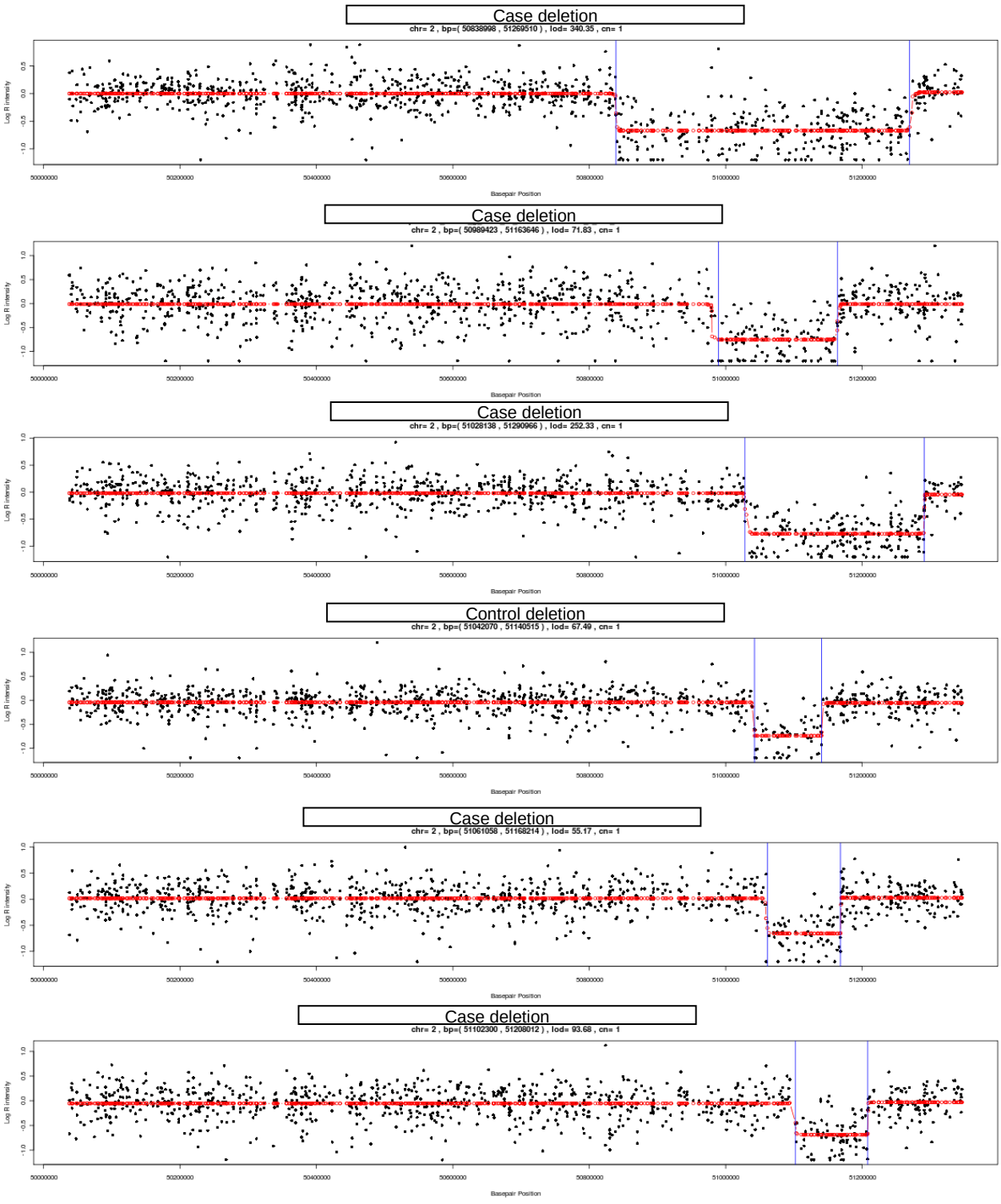




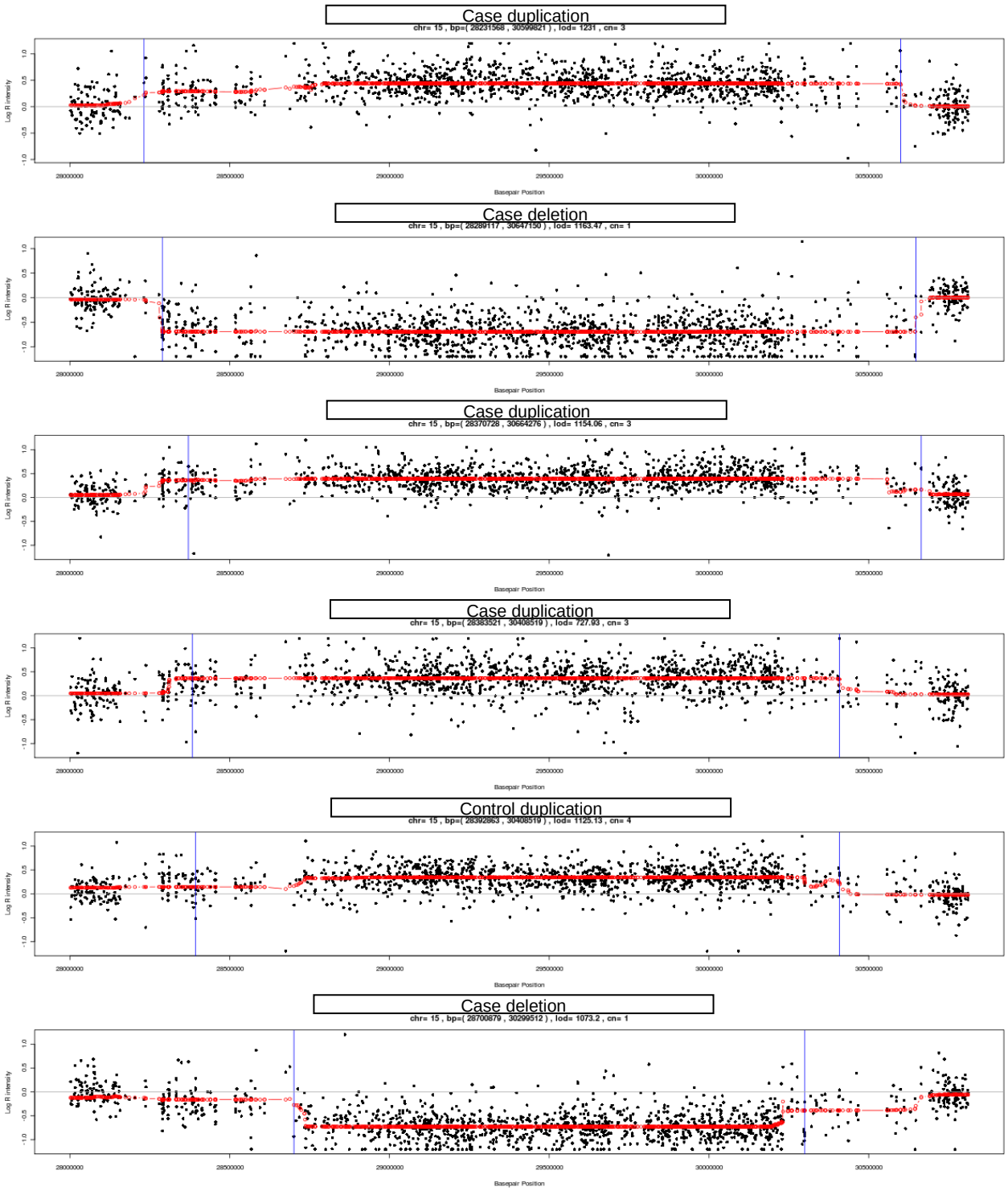


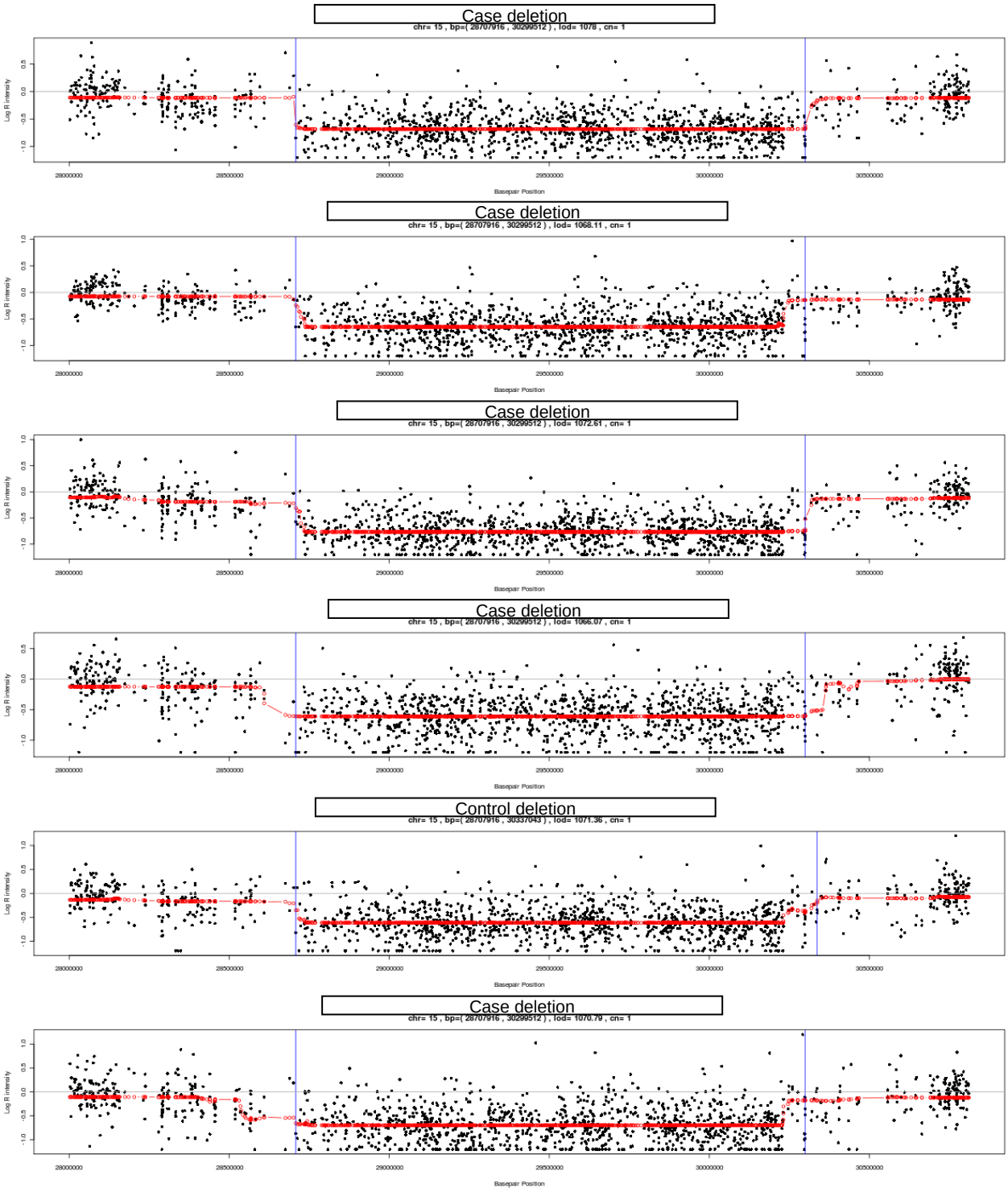




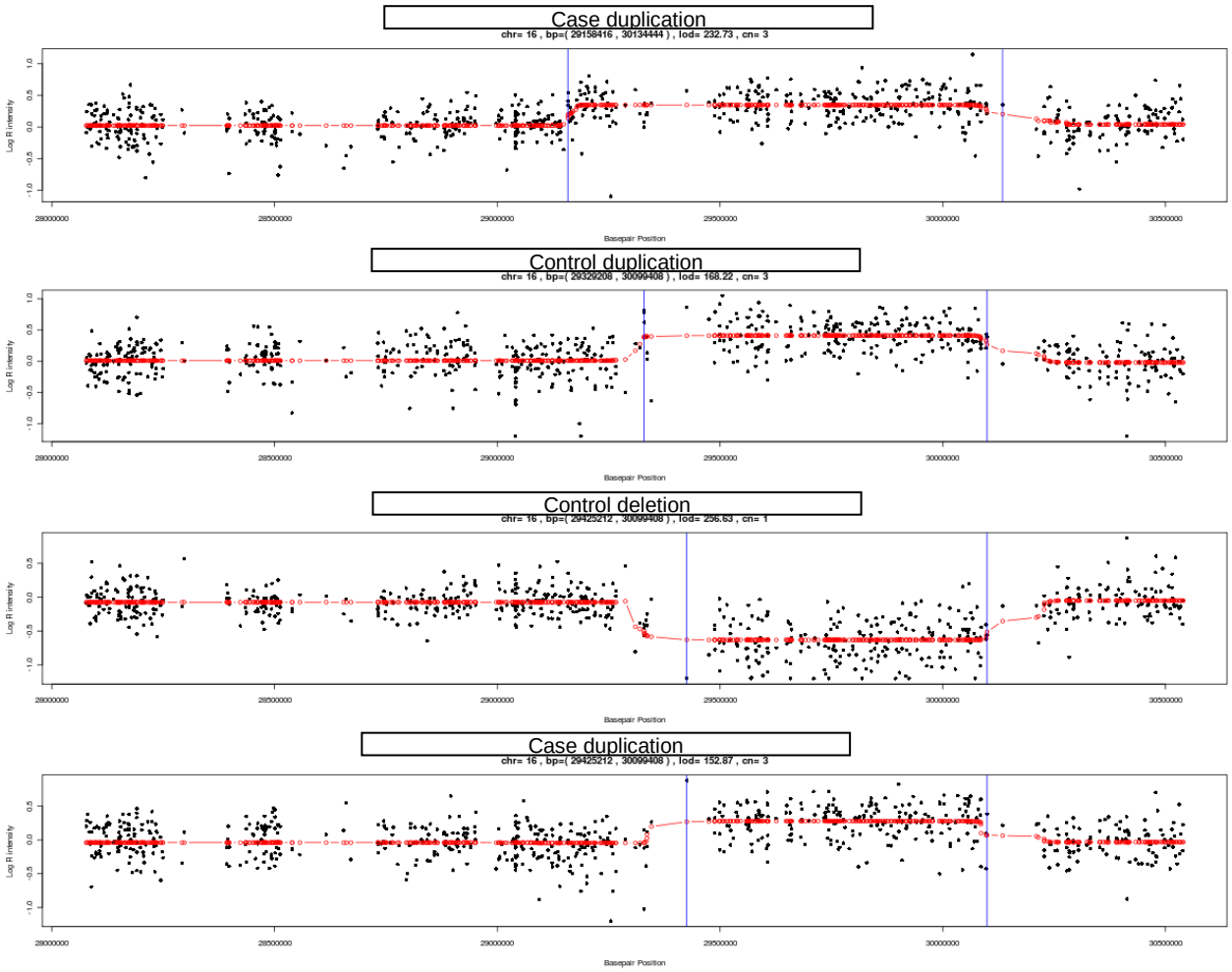

Case deletion
chr= 2 , bp=(50838998 , 51269510) , lod= 340.35 , cn= 1
Case deletion
chr= 2 , bp=(50989423 , 51163646) , lod= 71.83 , cn= 1
Case deletion
chr= 2 , bp=(51028138 , 51290966) , lod= 252.33 , cn= 1
Control deletion
chr= 2 , bp=(51042070 , 51140515) , lod= 67.49 , cn= 1
Case deletion
chr= 2 , bp=(51061058 , 51168214) , lod= 55.17 , cn= 1
Case deletion
chr= 2 , bp=(51102300 , 51208012) , lod= 93.68 , cn= 1
Log R intensity
Basepair Position

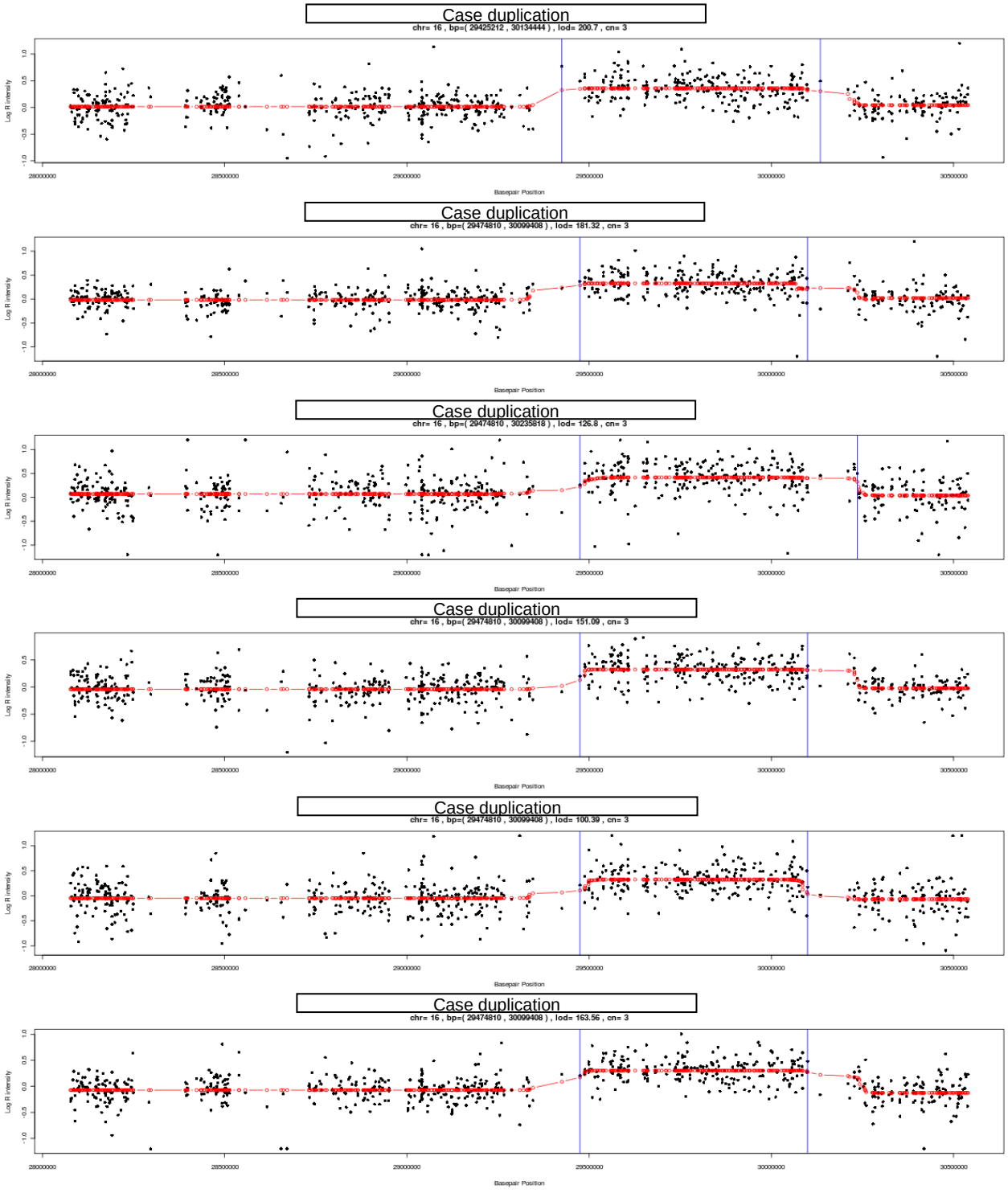
(3) Long deletions and duplications in 15q13.3

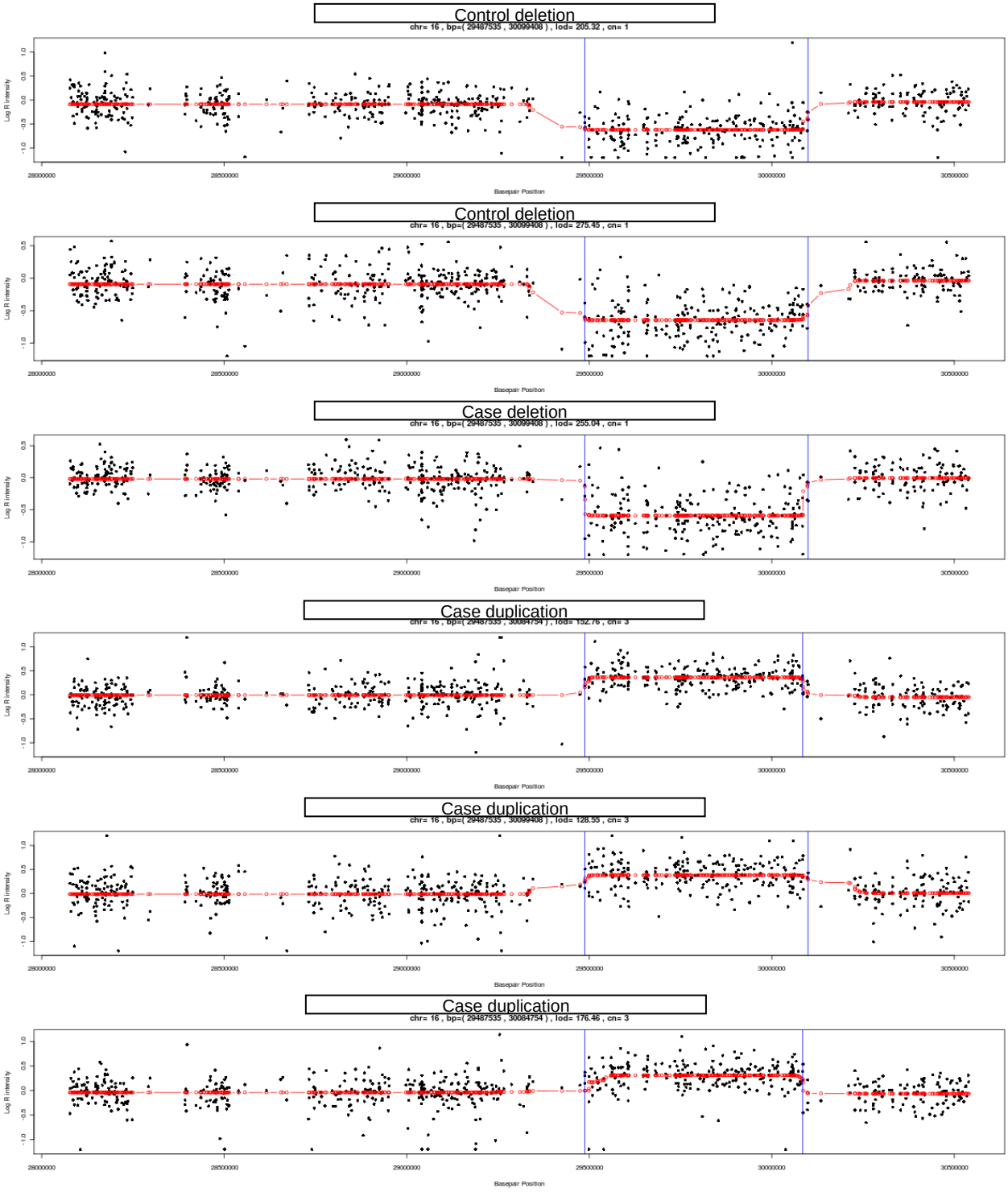


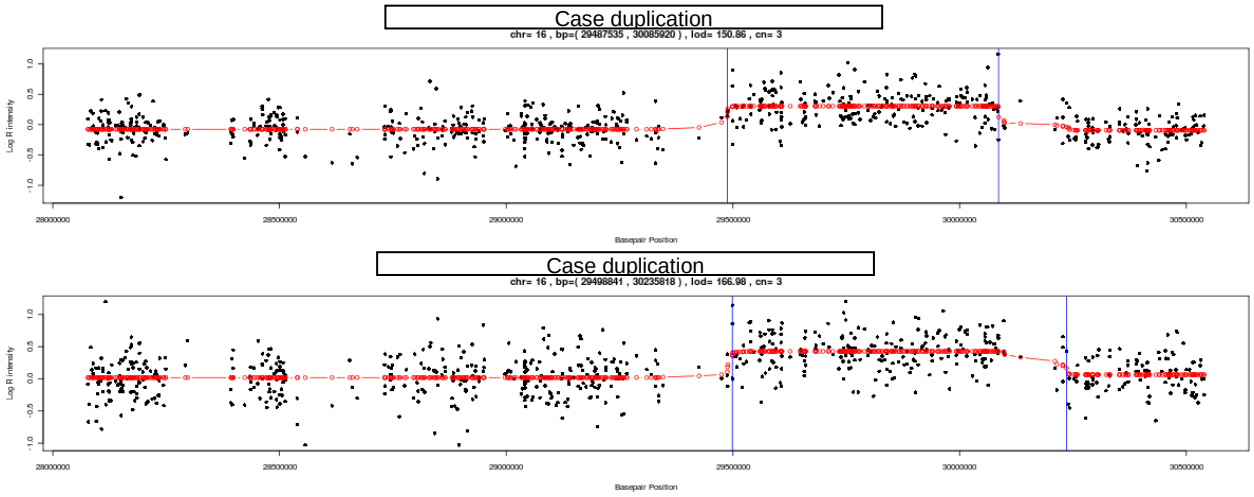

Case deletion
chr= 15 , bp=(28707916 , 30299512) , lod= 1078 , cn= 1
Case deletion
chr= 15 , bp=(28707916 , 30299512) , lod= 1068.11 , cn= 1
Case deletion
chr= 15 , bp=(28707916 , 30299512) , lod= 1072.61 , cn= 1
Case deletion
chr= 15 , bp=(28707916 , 30299512) , lod= 1066.07 , cn= 1
Control deletion
chr= 15 , bp=(28707916 , 30337043) , lod= 1071.36 , cn= 1
Case deletion
chr= 15 , bp=(28707916 , 30299512) , lod= 1070.79 , cn= 1
Log R intensity
Basepair Position

(4) 16p11.2 duplications

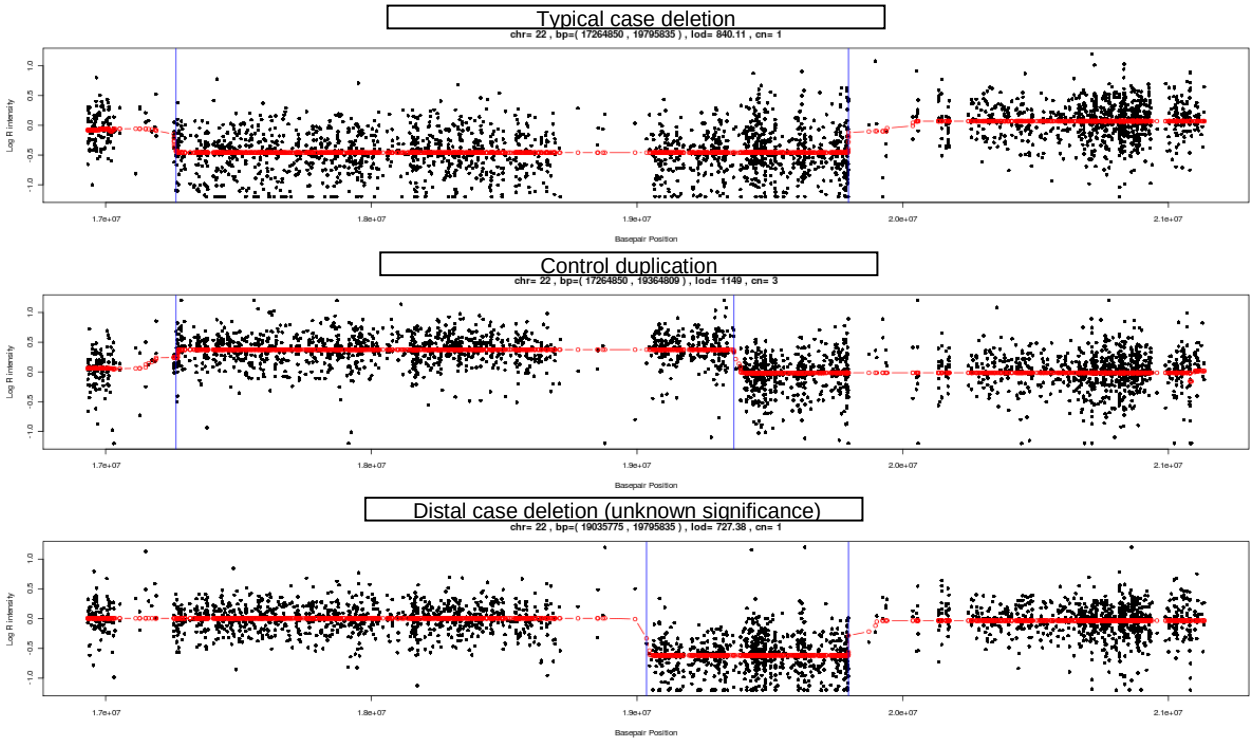



Case duplication
chr= 16 , bp=(29425212 , 30134444) , lod= 200.7 , cn= 3
Case duplication
chr= 16 , bp=(29474810 , 30099408) , lod= 181.32 , cn= 3
Case duplication
chr= 16 , bp=(29474810 , 30235818) , lod= 126.8 , cn= 3
Case duplication
chr= 16 , bp=(29474810 , 30099408) , lod= 151.09 , cn= 3
Case duplication
chr= 16 , bp=(29474810 , 30099408) , lod= 100.39 , cn= 3
Case duplication
chr= 16 , bp=(29474810 , 30099408) , lod= 163.56 , cn= 3
Log R intensity
Basepair Position

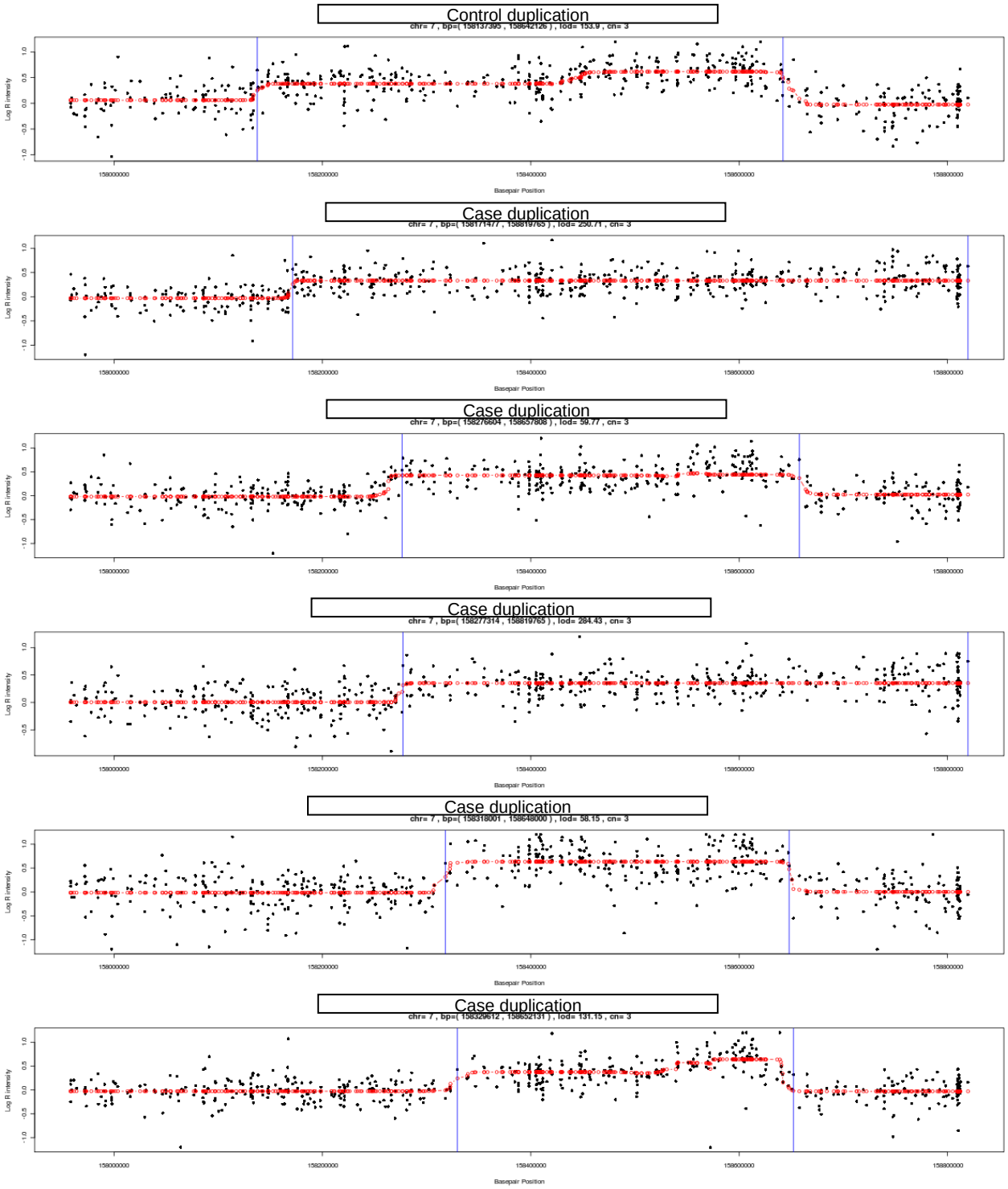

Control deletion
chr= 16 , bp=(29487535 , 30099408) , lod= 205.32 , cn= 1
Log R intensity
Basepair Position
Control deletion
chr= 16 , bp=(29487535 , 30099408) , lod= 275.45 , cn= 1
Log R intensity
Basepair Position
Case deletion
chr= 16 , bp=(29487535 , 30099408) , lod= 255.04 , cn= 1
Log R intensity
Basepair Position
Case duplication
chr= 16 , bp=(29487535 , 30084754) , lod= 152.76 , cn= 3
Log R intensity
Basepair Position
Case duplication
chr= 16 , bp=(29487535 , 30099408) , lod= 128.55 , cn= 3
Log R intensity
Basepair Position
Case duplication
chr= 16 , bp=(29487535 , 30084754) , lod= 176.46 , cn= 3
Log R intensity
Basepair Position

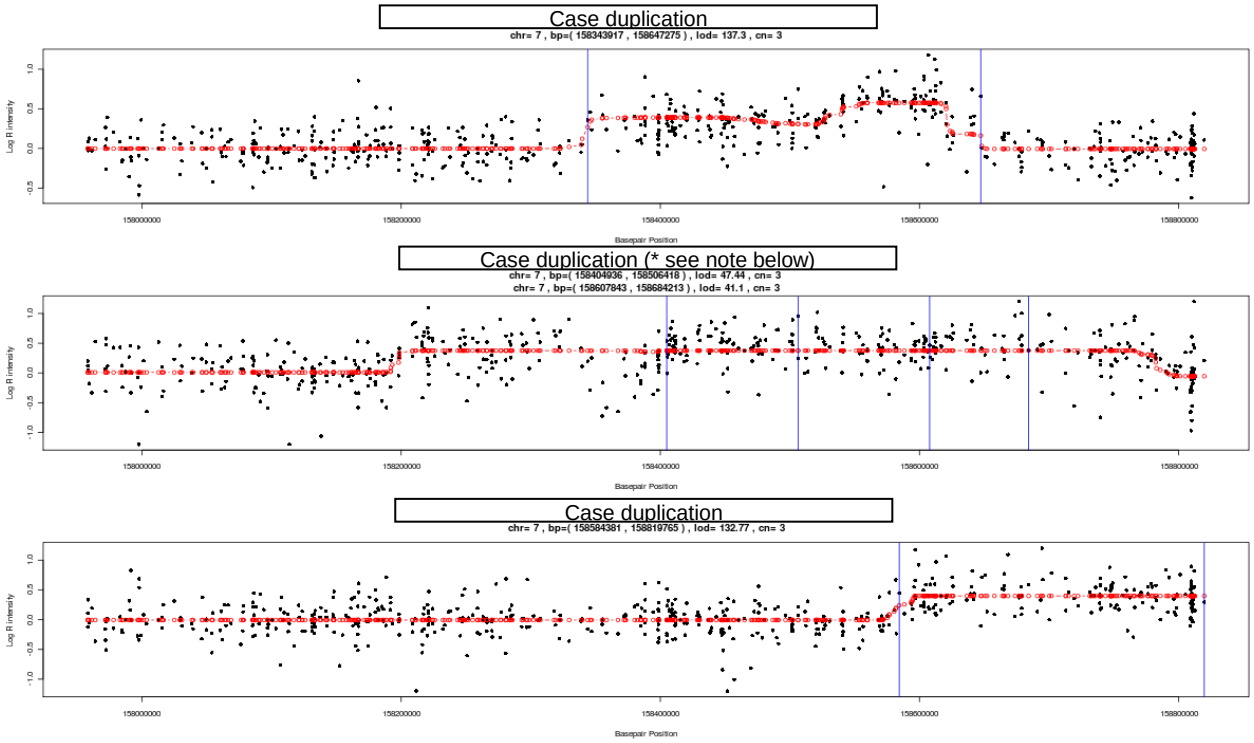


(5) 22q11.21 CNVs. The 21 22q11.21 deletions counted in the analysis in Table 3 are typical of a well-documented CNV as described in the text, with clearcut evidence of the deletion. Therefore we show here only illustrative examples of a case deletion with typical boundaries and a control duplication (6 controls and 2 cases had long 22q11.21duplications). The third plot is a highly atypical distal case deletion which has not been reported in schizophrenia and whose pathological significance is not known -- it was not counted in the MGS results that included 19 typical and 2 proximal 22q11.21 deletions (all in cases), consistent with previous reports. It does not overlap with the proximal “atypical” deletions that are found in all VCFS studies.)



(6) Duplications in VIPR2





*Note: Birdseye called two separate duplications in this interval, whereas the algorithm of Lai et al. estimated CN=3 in one continuous and longer interval, which seems more consistent with intensity data. This is typical of the issue of when to "merge" nearby CNVs.