

**SUPPLEMENTAL MATERIAL****This file includes the following:**

<b>Supplemental data</b>	<b>Page</b>
Table of Contents	1
<b><i>Supplemental Tables:</i></b>	
Table S1: Summary of Affymetrix 6.0 microarray CNV data for 420 schizophrenia cases of European ancestry	2
Table S2: Summary of Affymetrix 6.0 microarray CNV data for 416 OPGP control individuals of European ancestry	3
Table S3: Frequency of typical large gains/losses at chromosome 2q13 in 23,838 population-based controls	4
Table S4: All stringent rare CNVs in schizophrenia cases and OPGP control individuals	See Excel attachment
Table S5: Gene-set association results for all gene-sets tested, involving very rare CNVs restricted to exonic losses	See Excel attachment
Table S6: Additional gene-set information for the 70 significant gene-sets showing significant enrichment for schizophrenia	See Excel attachment
Table S7: Gene-set association results for 25 neurological gene-sets tested, involving very rare CNVs restricted to exonic losses	See Excel attachment
<b><i>Supplemental Figures:</i></b>	
Figure S1: Overview of study design and CNV analysis workflow	5
Figure S2: Large rare pathogenic CNVs (see Table 1) and smaller rare CNVs overlapping putative candidate genes for schizophrenia	6
Figure S3: Large rare CNV at 7q22.3-q31.2 overlapping <i>FOXP2</i> in an adult with schizophrenia	7
Figure S4: Overlapping large rare CNVs at three novel loci for schizophrenia	8
Figure S5: Detailed gene-set association results and clusters	9
Figure S6: Detailed gene-set association results, clusters and genes	10
Figure S7: Overlap of gene-set association results with ASD genes	11
Supplemental References	12

**Table S1. Summary of Affymetrix 6.0 microarray CNV data for 420 schizophrenia cases of European ancestry**

	All (n=420) Caucasian probands						
	Non-genic			Genic			All <sup>1</sup>
	Gain	Loss	Mixed	Gain	Loss	Mixed	
Total number of stringent CNVs for all samples with results	2,485	5,400	19	3,759	5,088	173	16,924
% of total stringent CNVs	14.69	31.91	0.11	16.30	30.06	1.02	100.00
Average number of CNVs per genome <sup>2</sup> (N=420)	5.92	12.86	0.05	8.59	12.11	0.41	40.30
Median size (range)	31,810 (10,086-1,620,721)	24,086 (10,022-1,156,392)	111,225 (45,159-251,551)	83,941 (10,072-6,498,447)	52,393 (10,033-3,687,431)	83,941 (26,147-900,917)	39,394 (10,022-6,498,447)
Median size (range) for rare <sup>3</sup> CNVs	45,840 (11,231-1,073,381)	37,363.5 (10,115-626,127)	0	98,494 (10,904-6,498,447)	45,417 (10,033-3,687,431)	0	49657.5 (10,033-6,498,447)
Number of >100kb CNVs	339	649	15	1,805	1,632	83	4,523
% of total CNVs, of same type, that are >100kb	13.64	12.02	78.95	65.42	32.08	47.98	26.73
Number of rare <sup>3</sup> CNVs	188	318	0	316	299	0	1121
% of total CNVs, of same type, that are rare <sup>3</sup>	7.57	5.89	0.00	11.45	5.88	0.00	6.62
Average number of rare <sup>3</sup> CNVs per genome <sup>4</sup>	0.45	0.76	0.00	0.75	0.71	0.00	2.67
Average number of rare <sup>3</sup> CNVs per genome <sup>5</sup>	0.49	0.83	0.00	0.83	0.78	0.00	2.94
Number of subjects with one or more rare <sup>3</sup> CNVs	144	213	0	211	214	0	382
% of subjects with one or more rare <sup>3</sup> CNVs	34.29	50.71	0.00	50.24	50.95	0.00	90.95
Number of subjects with one or more rare <sup>3</sup> CNVs <sup>6</sup>	131	205	0	199	210	0	376
% of subjects with one or more rare <sup>3</sup> CNVs <sup>6</sup>	31.19	48.81	0.00	47.38	50.00	0.00	89.52

This table accounts for stringent CNVs 10kb-6.5Mb; no restriction on chromosome (i.e., includes X chromosome except where indicated)

<sup>1</sup> “All” category comprises genic and non-genic CNVs

<sup>2</sup> Average number of CNVs per genome= total number of CNVs in each column divided by total number of subjects

<sup>3</sup> Rare CNVs defined as CNVs present in <0.1% of population controls

<sup>4</sup> Average number of rare CNVs per genome= total number of rare CNVs in each column divided by the total number of subjects (include subjects with no rare CNV)

<sup>5</sup> Average number of rare CNVs per genome = total number of rare CNVs in each column divided by the total number of subjects (only include subjects with 1 or more rare CNVs)

<sup>6</sup> Excluded CNVs on X chromosome

**Table S2. Summary of Affymetrix 6.0 microarray CNV data for 416 OPGP control individuals of European ancestry\***

	All (n=416) Caucasian OPGP controls						
	Non-genic			Genic			All <sup>1</sup>
	Gain	Loss	Mixed	Gain	Loss	Mixed	
Total number of stringent CNVs for all samples with results	2,129	5,774	64	3,472	5,056	225	16,720
% of total stringent CNVs	12.73	34.53	0.38	20.77	30.24	1.35	100.00
Average number of CNVs per genome <sup>2</sup> (N=416)	5.117788462	13.87980769	0.153846154	8.346153846	12.15384615	0.540865385	40.19230769
Median size (range)	29,851 (10,103-1,654,909)	22,754 (10,022-2,041,019)	58,168 (12,803-409,887)	76,846 (10,022-2,605,726)	46,286 (10,016-1,855,636)	83,936 (26,147-1,354,407)	39,394 (10,016-2,605,726)
Median size (range) for rare <sup>3</sup> CNVs	39,821.5 (10,178-672,304)	37,586 (10,134-2,041,019)	0	92,828 (10,037-2,605,726)	39,536 (10,016-1,430,916)	0	47,466.5 (10,016-2,605,726)
Number of >100kb CNVs	366	646	25	1,633	1,481	106	4,254
% of total CNVs, of same type, that are >100kb	17.19	11.19	39.06	47.03	29.29	47.11	25.44
Number of rare <sup>3</sup> CNVs	168	325	0	273	256	0	1022
% of total CNVs, of same type, that are rare <sup>3</sup>	7.89	5.63	0	7.86	5.06	0	6.11
Average number of rare <sup>3</sup> CNVs per genome <sup>3</sup>	0.40	0.78	0	0.66	0.62	0	2.46
Average number of rare <sup>3</sup> CNVs per genome <sup>4</sup>	0.44	0.85	0	0.71	0.67	0	2.68
Number of subjects with one or more rare <sup>3</sup> CNVs	138	220	0	186	205	0	382
% of subjects with one or more rare <sup>3</sup> CNVs	33.17	52.88	0	44.71	49.28	0	91.83
Number of subjects with one or more rare <sup>3</sup> CNVs <sup>6</sup>	118	211	0	181	203	0	372
% of subjects with one or more rare <sup>3</sup> CNVs <sup>6</sup>	28.37	50.72	0	43.51	48.80	0	89.42

This table accounts for stringent CNVs 10kb-6.5Mb; no restriction on chromosome (i.e., includes X chromosome except where indicated)

<sup>1</sup> "All" category comprises genic and non-genic CNVs

<sup>2</sup> Average number of CNVs per genome= total number of CNVs in each column divided by total number of subjects in each category

<sup>3</sup> Rare CNVs defined as CNVs present in <0.1% of population controls

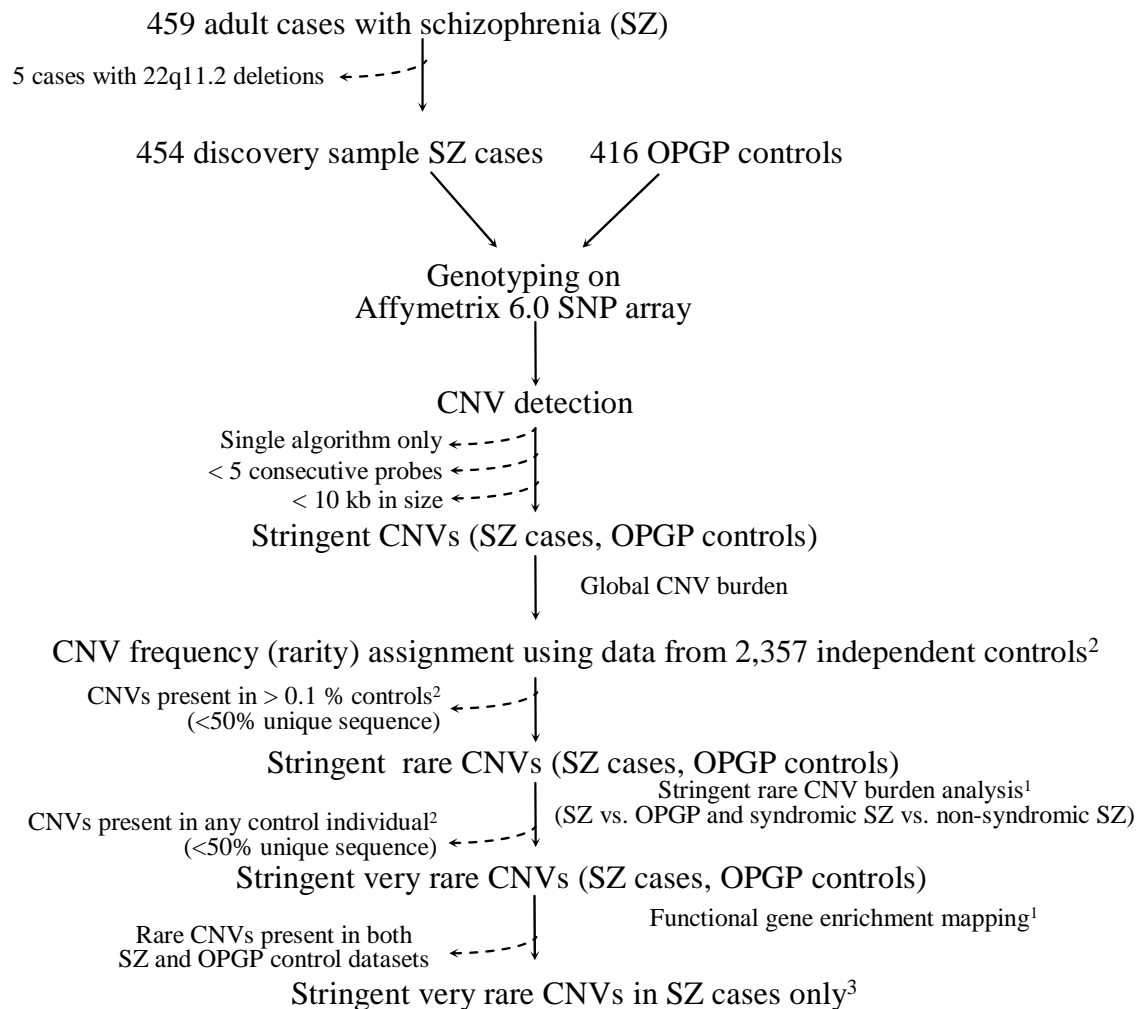
<sup>4</sup> Average number of rare CNVs per genome= total number of rare CNVs in each column divided by the total number of subjects (include subjects with no rare CNV)

<sup>5</sup> Average number of rare CNVs per genome = total number of rare CNVs in each column divided by the total number of subjects (only include subjects with 1 or more rare CNVs)

<sup>6</sup> Excluded CNVs on X chromosome

**Table S3. Frequency of typical large gains/losses at chromosome 2q13 in 23,838 population-based controls.**

Dataset	N	Array type	Geographic origin	Reference	Typical large gains/losses at chromosome 2q13 (chr2:111,105,101-112,832,463)
POPGEN controls	1,123	Affymetrix 6.0	Germany	(1)	0
Ottawa Heart Institute controls	1,234	Affymetrix 6.0	Ontario, Canada	(2)	0
HapMap (Phase 3) controls	1,056	Affymetrix 6.0	HapMap populations across the world	(3)	0
Wellcome Trust Case Control Consortium (WTCCC) controls	4,783	Affymetrix 6.0	UK	(4)	0
Ontario Population Genomics Platform (OPGP) controls	416	Affymetrix 6.0	Ontario, Canada	(5)	0
Starr County Diabetes cases & controls	1,794	Affymetrix 6.0	USA	(6)	2 (gains; chr2:111,105,101-112,832,463)
Geneva NHS/HPFS Diabetes cases & controls	5,966	Affymetrix 6.0	USA	(7)	0
SAGE consortium controls	1,287	Illumina 1M	USA	(8)	1 (loss; chr2:111,078,231-112,816,485)
Health, Aging, and Body Composition (Health ABC) Study controls	2,566	Illumina 1M-Duo	USA	(9)	1 (gain; chr2:111,078,231-112,816,485)
Ontario ARCTIC controls	1,120	Affymetrix 500k	Ontario, Canada	(10)	0
Itsara et al. controls	2,493	Illumina Hap550, Hap650Y and Hap300	Various	(11)	0
<b>Total</b>	23,838	-	-	-	4

**Figure S1. Overview of study design and CNV analysis workflow**

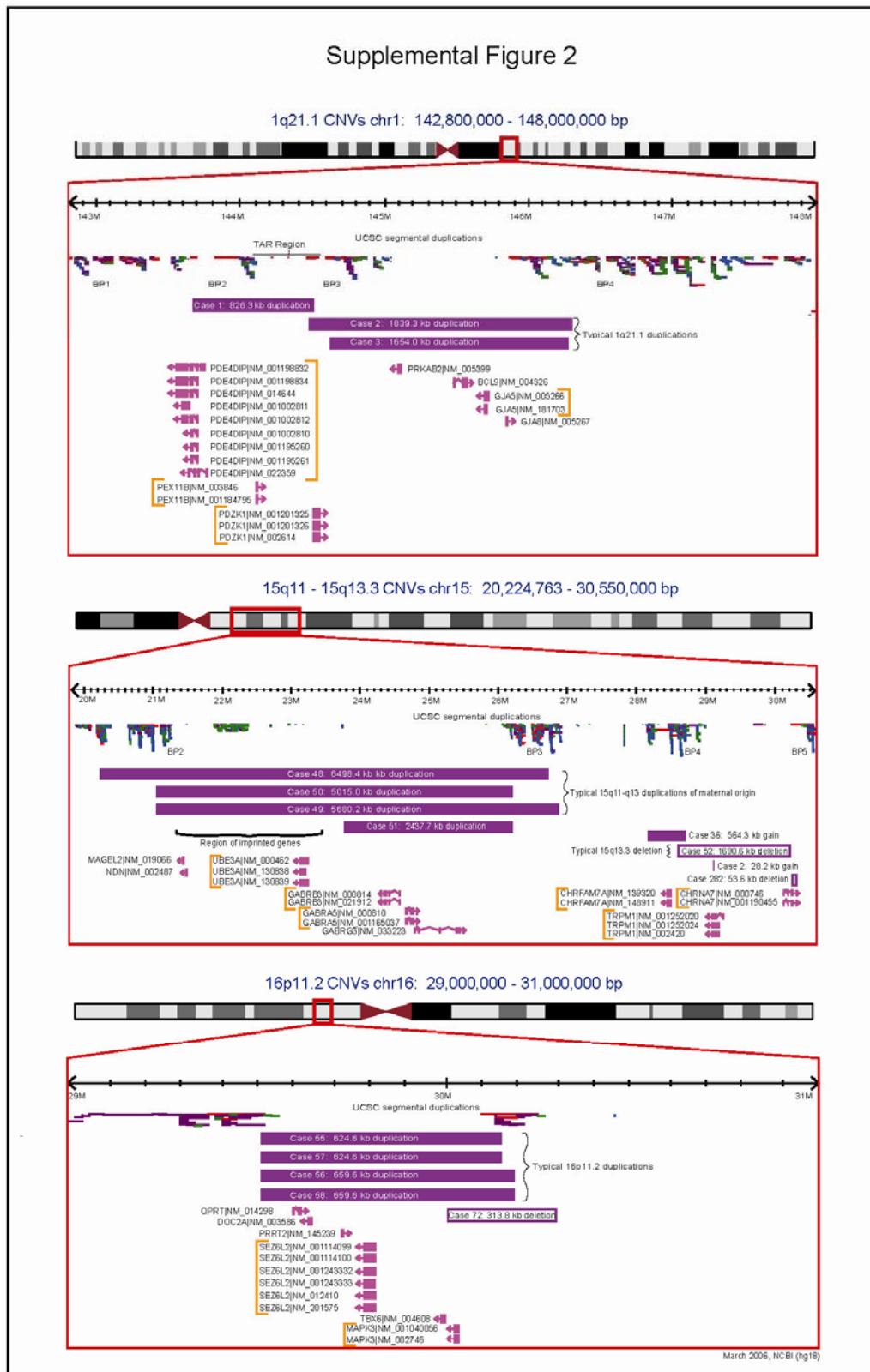
- Comprehensive examination of genes for evidence of relevant expression in human and model animal (mouse, zebrafish) systems
- Overlap with previous genetic findings in SZ
- Recurrent and overlapping rare CNVs identifying familiar and novel SZ risk genes
- Experimental validation of CNVs (quantitative PCR or clinical microarray)

<sup>1</sup>CNV burden analysis and functional gene enrichment mapping focused on comparing the subset of unrelated SZ cases of European ancestry (n = 420) with ancestry-matched unrelated adult OPGP controls (n=416).

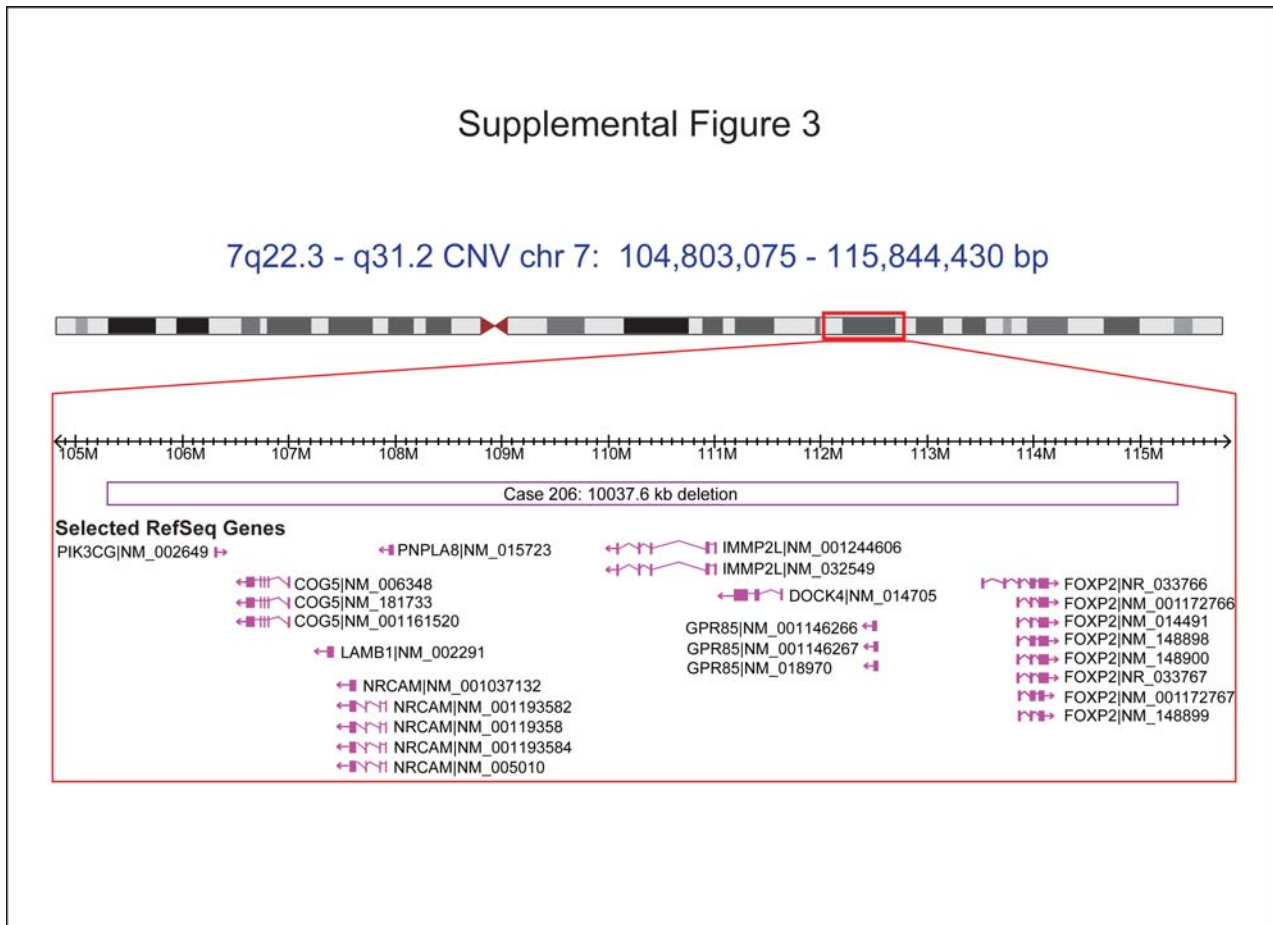
<sup>2</sup>Population-based adult controls of European ancestry (n=2,357), used in previous studies of autism and attention deficit disorder and independent to OPGP controls, were genotyped on the Affymetrix 6.0 platform and analyzed for CNVs in an identical manner to the SZ cases and OPGP controls.

<sup>3</sup>Included CNVs 4–10 kb in size (n=7) that overlapped CNVs > 10kb in size overlapping SZ candidate genes

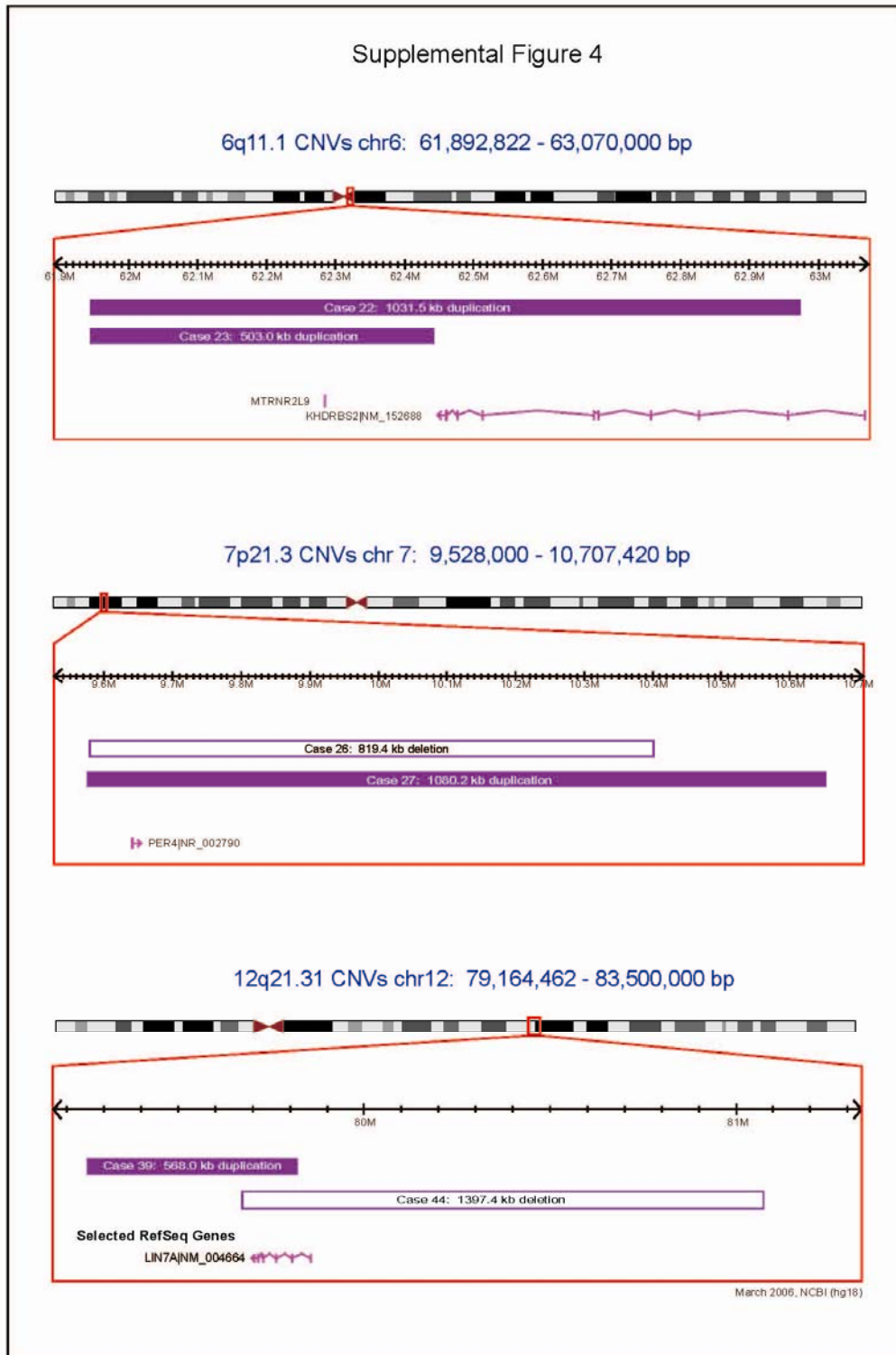
**Figure S2. Large rare pathogenic CNVs (see Table 1) and smaller rare CNVs overlapping putative candidate genes for schizophrenia.**



**Figure S3. Large rare CNV at 7q22.3-q31.2 overlapping *FOXP2* in an adult with schizophrenia.**



**Figure S4. Overlapping large rare CNVs at three novel loci for schizophrenia.**





**Figure S5. Detailed gene-set association results and clusters.**

Supplementary Figure 5

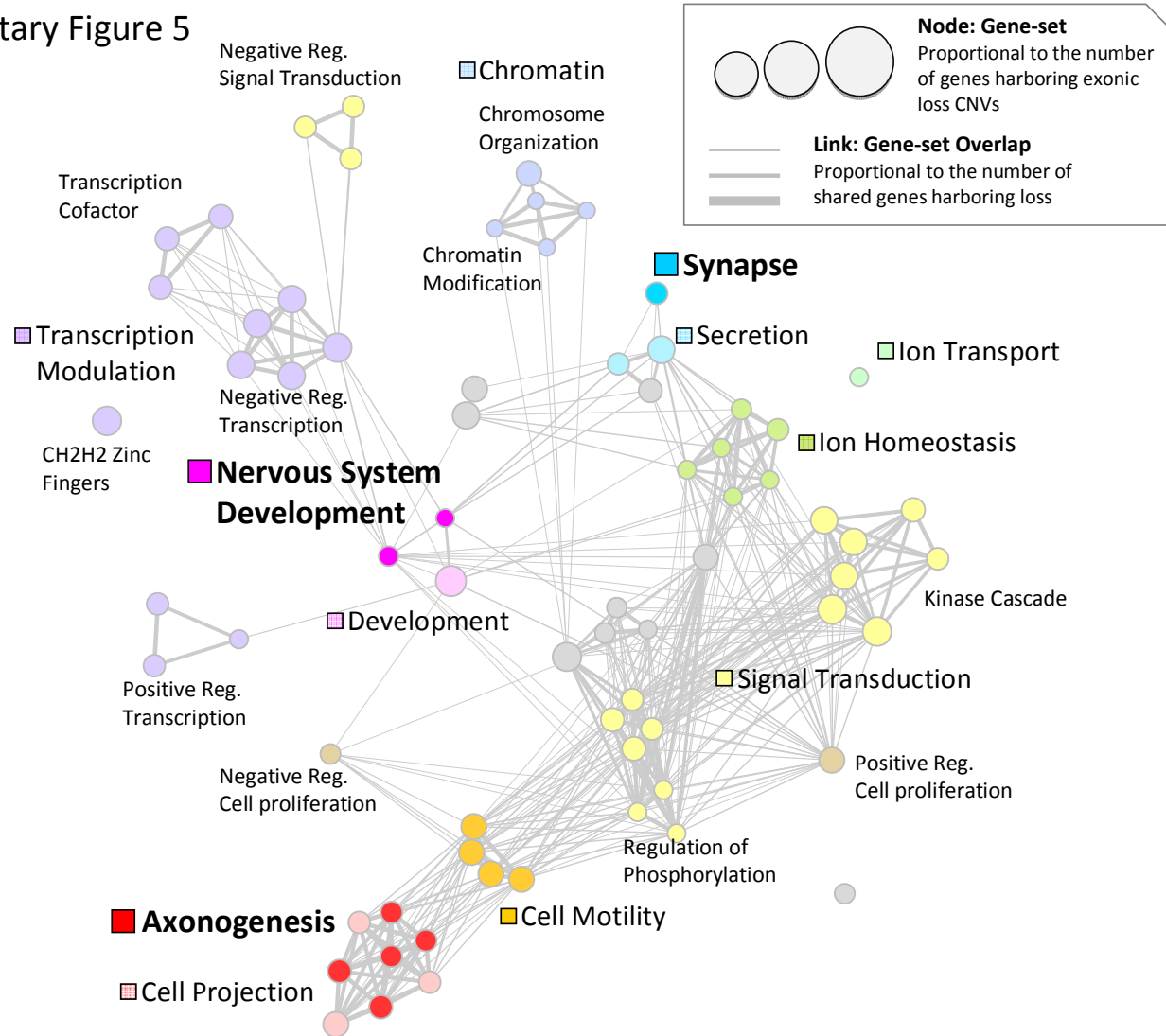
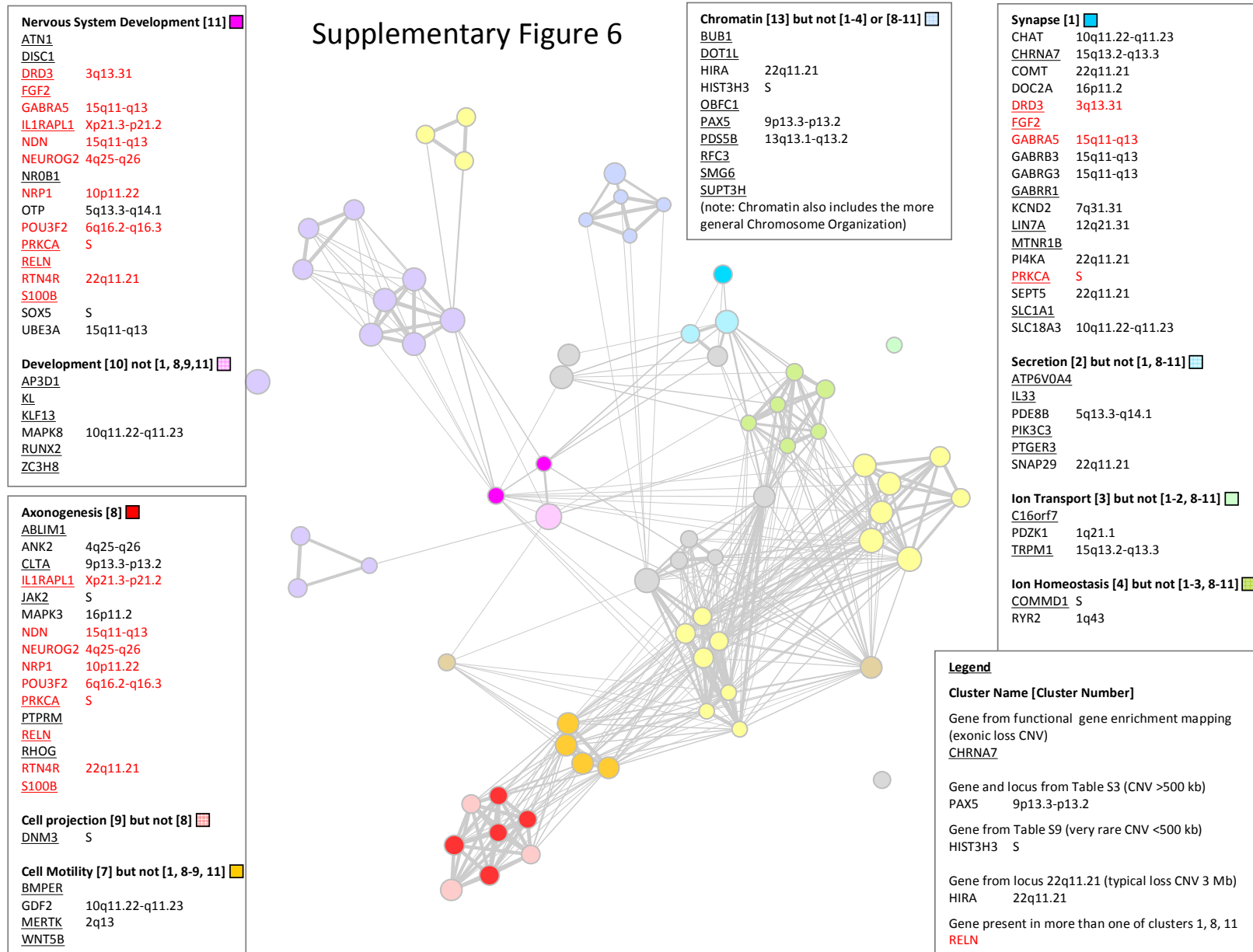
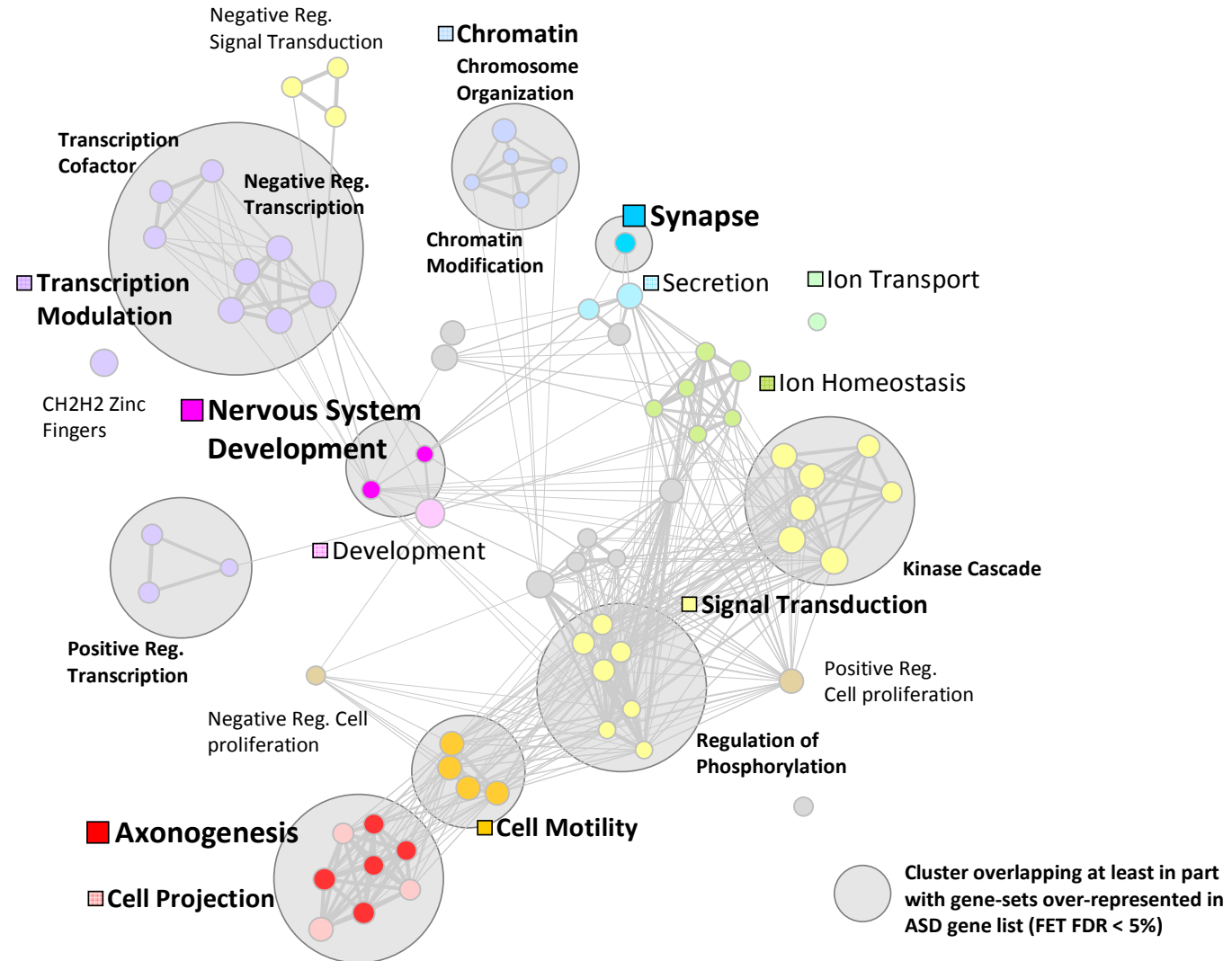


Figure S6. Detailed gene-set association results, clusters and genes.



**Figure S7. Overlap of gene-set association results with ASD genes.** Functional map of schizophrenia (see Figure 2 for details), highlighting the clusters of the gene-sets significantly associated to schizophrenia for which at least half the gene-sets were also significantly over-represented in known ASD genes. (12) Over-representation was tested using Fisher's Exact Test (FET) and Benjamini-Hochberg FDR for multiple test correction, with a significance threshold of 5% FDR.

Supplementary Figure 7



## Supplemental References

- 1 Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P.J., El Mokhtari, N.E. and Schreiber, S. (2006) PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.*, **9**, 55-61.
- 2 Stewart, A.F., Dandona, S., Chen, L., Assogba, O., Belanger, M., Ewart, G., LaRose, R., Doelle, H., Williams, K., Wells, G.A. *et al.* (2009) Kinesin family member 6 variant Trp719Arg does not associate with angiographically defined coronary artery disease in the Ottawa Heart Genomics Study. *J. Am. Coll. Cardiol.*, **53**, 1471-1472.
- 3 International HapMap, C., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F. *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52-58.
- 4 Wellcome Trust Case Control, C., Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F. *et al.* (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, **464**, 713-720.
- 5 Silversides, C.K., Lionel, A.C., Costain, G., Merico, D., Migita, O., Liu, B., Yuen, Y., Rickaby, J., Thiruvahindrapuram, B., Marshall, C.R. *et al.* (2012) Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. *PLoS Genet.*, **8**, e1002843.
- 6 Below, J.E., Gamazon, E.R., Morrison, J.V., Konkashbaev, A., Pluzhnikov, A., McKeigue, P.M., Parra, E.J., Elbein, S.C., Hallman, D.M., Nicolae, D.L. *et al.* (2011) Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals. *Diab. Tologia*, **54**, 2047-2055.
- 7 Qi, L., Cornelis, M.C., Kraft, P., Stanya, K.J., Linda Kao, W.H., Pankow, J.S., Dupuis, J., Florez, J.C., Fox, C.S., Pare, G. *et al.* (2010) Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum. Mol. Genet.*, **19**, 2706-2715.
- 8 Bierut, L.J., Agrawal, A., Bucholz, K.K., Doheny, K.F., Laurie, C., Pugh, E., Fisher, S., Fox, L., Howells, W., Bertelsen, S. *et al.* (2010) A genome-wide association study of alcohol dependence. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 5082-5087.
- 9 Coviello, A.D., Haring, R., Wellons, M., Vaidya, D., Lehtimaki, T., Keildson, S., Lunetta, K.L., He, C., Fornage, M., Lagou, V. *et al.* (2012) A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. *PLoS Genet.*, **8**, e1002805.
- 10 Zogopoulos, G., Ha, K.C., Naqib, F., Moore, S., Kim, H., Montpetit, A., Robidoux, F., Laflamme, P., Cotterchio, M., Greenwood, C. *et al.* (2007) Germ-line DNA copy number variation frequencies in a large North American population. *Hum. Genet.*, **122**, 345-353.
- 11 Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I. *et al.* (2009) Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.*, **84**, 148-161.
- 12 Betancur, C. (2011) Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.*, **1380**, 42-77.