ONLINE METHODS

1. Data Collection

For this study, data were collected on 8,800 schizophrenia, bipolar disorder and autism cases, and 6391 controls analyzed on one of four microarray platforms (ROMA, NimbleGen HD2, Affymetrix500K, Affymetrix 6.0). Ascertainment of samples in the Primary (1906 schizophrenia cases and 3971 controls), Replication (2645 cases and 2420 controls), Autism (934 cases) and Bipolar Disorder (3315 cases) datasets is provided in the supplementary note. A break down of samples-by-array is provided in the supplementary table 1.

2. Intensity Data Processing

Processing of microarray data was performed at three different sites. Affymetrix 500K data from the NIMH (83 cases) were processed at NIMH using published described methods [4]. Affymetrix 500K data from Cardiff University and the WTCCC (471 schizophrenia cases, 1697 bipolar disorder cases and 2792 controls), were processed at Cardiff University using the same software and initial QC parameters for CNV calls and arrays as published previously [6] however, for these analyses we did not remove CNVs that were detected by <10 probes on both the Nsp and Sty arrays.

All other data were processed at Cold Spring Harbor Laboratory using different methods for dual color intensity data (Array-CGH Platforms: ROMA 85K and NimbleGen HD2) and single color intensity data (single nucleotide (SNP) genotyping arrays: Affymetrix 500K, Affymetrix 5.0 and Affymetrix 6.0). Processing of dual and single color intensity data is described in more detail in the supplementary note. Briefly,

a) Array-CGH Intensity Data

1. *ROMA*

Normalization of ROMA intensity data by *Lo*cally *We*ighted *S*catterplot *S*moothing (*LOWESS*) and geometric mean estimation of $\log_2$ ratios has been described previously [35].

2. *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 preserving the variability in the data. The $\log_2$ ratio for each probe is then estimated using the geometric mean of normalized and raw intensity data.

b) *SNP Genotyping Data*

1. *Affymetrix 500K, Affymetrix 5.0 & Affymetrix 6.0*

To analyze Affymetrix SNP Array single color intensity data we developed a two-step process that first, normalizes all arrays by invariant set normalization to a single reference array and second calculates the ratio of intensities for each experiment in comparison to a sex matched “virtual reference genome” (VRG).

c) *GC Correction of $\log_2$ Ratios*

The final step of data processing involves the correction of genomic waves effects in $\log_2$ ratios due to regional correlations with GC content based on the fitted linear regression model proposed by Diskin et al. [36]

3. Chr16p11.2 Detection by HMM Segmentation.

To detect 16p11.2 rearrangements in our ROMA and GC-corrected Affymetrix log_2 ratio data we implemented the seven-state HMM algorithm described previously [35]. We used a modified version of this HMM algorithm to identify CNVs in our higher resolution Affymetrix 5.0, 6.0 and NimbleGen HD2 GC-corrected data sets [37]. The results of segmentation were examined for the presence of CNVs overlapping at least 50% of the 16p11.2 region (chr16:29557498-30107355 of the UCSC Human Genome version HG18 (NCBI Build 36.1)).

4. 16p11.2 Genotyping: Rare CNV detection by Outlier Clustering

a. *Principles*

As an alternative method for genotyping rare CNVs, we developed an algorithm called *Median Z-Score Outlier Detection* (MeZOD) to detect rare variants based on the probe intensity data across the population of experiments. The principles of this method are similar to other approaches that genotype common CNVs by probe intensity clustering [38,39]; however, in most cases very few individuals carry the rare genotype. Therefore, rather than using cluster analysis to identify variants in the population, our method detects rare outliers of the standardized probe intensity distribution.

b. *Selection of Target probes and Flanking probes*

The 16p11.2 rearrangements are genotyped using probes selected from within the target region (chr16:29564890-30100063). Two unique sequences: one proximal (chr16:27388307-28952358) and one distal (chr16:30304580-31870683)) to the 16p11.2 target region, are combined into a single set of “invariant” probes. The results are displayed

as a scatterplot. Median Z-scores of target probes are shown on the X-axis, and Median Z-scores of the invariant probes are shown on the Y-axis.

To avoid patterns of common copy number polymorphism, probes were excluded if the positive or negative Pearson correlations with neighboring probes exceeded conservative maximum or minimum thresholds, respectively. Probes not exceeding these thresholds were used for genotyping. The accompanying bed file contains all platform specific probes within the target and invariant regions. The selected genotyping probes in the target and invariant regions are represented in red and green, respectively in the UCSC human genome browser.

c. Median Z-Score Calculation and Outlier Detection

Calculation of the median Z-Scores is a three stage process first involving experiment-wise $\log_2$ ratio standardization, second, probe-wise standardization of the genotyping probe Z-scores and finally median Z-score determination for the target and invariant region. For each probe *m* of experiment *N*, the standardized $\log_2$ ratio Z-Score *Z* is simply calculated by:

$$Z(m_N) = \frac{m_N - \mu_N}{\sigma_N}\text{, where}$$

μ_N and σ_N are the mean and standard deviation of probe ratios for experiment *N*, respectively. The *Z*-Score for each genotyping probe *G* in experiment *N* was then standardized probe-wise within the population of experiments for a given platform by:

$$Z(G_N) = \frac{G_N - \mu_G}{\sigma_G}\text{, where}$$

μ_G is the mean and σ_G is standard deviation of genotyping probe *G*. Finally the median for experiment *N* is calculated for the target genotyping probes and the combined proximal and distal invariant genotyping probes.

To detect rearrangements of 16p11.2, outliers of the target median Z-Score distribution were analyzed. Thresholds were set for microduplications at target median Z-Scores >2 for ROMA and >1 for Affymetrix500K, Affymetrix 6.0 and NimbleGen HD2 while the outlier threshold for microduplications on all platforms was below a target median Z-Score of -2. As noted earlier, with the exception of Affymetrix 500K data analyzed locally by the NIMH, Cardiff University and the WTCCC, all intensity data was analyzed by MeZOD at CSHL.

5. Validations *of 16p11.2* Rearrangements

All rearrangements of 16p11.2 detected in the primary and replication samples were validated using an additional microarray platform. Microduplications detected on the NimbleGen HD2 platform were confirmed on Agilent 244K array. CNVs detected on other platforms including the Cardiff schizophrenia cases were validated on the NimbleGen HD2 array. Rearrangements detected in the Welcome Trust Case Control Consortium (WTCCC) were detected independently on both Affymetrix *Nsp* and *Sty* arrays (Supplementary Figure 2). Additional DNA was not available on WTCCC controls to perform additional fine-mapping of events detected in these samples. Of fifteen CNVs detected in additional cohorts of autism and bipolar disorder, genomic DNA was available for twelve (Supplementary Table 2), and all of these were validated.

6. **Meta-Analysis and Strength of 16p11.2 Associations in multiple psychiatric disorders**

Data from this study were combined with data from three independent published studies [1,3,5] to obtain a combined sample of 8590 schizophrenia , 4822 bipolar disorder and 2172 autism or developmental delay cases, and a combined sample of 30,492 controls. Controlling for study, 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. Additional information on each study included in the meta-analysis is provided in the in the supplementary note.

7. **Statistical Analysis**

   a. *Association of 16p11.2 microduplication with schizophrenia*

The primary sample consisted of data from multiple microarray platforms that vary in probe density. All have good sensitivity to detect CNVs that are 500 Kb in size. However, subtle differences in sensitivity could influence the overall frequency of 16p11.2 microduplications when all platforms are combined into a single dataset. Therefore, we used array type as a stratifying variable when testing for association using the Cochran-Mantel-Haenszel (CMH) exact test. Logistic regression was used to estimate the combined odds ratio (OR) and 95% confidence intervals (95%CI) and to measure the effect of array type based on the deviance P-value. The Fishers Exact Test was used to test the association of the 16p11.2 microduplications in the replication data set (single array platform). The Breslow-Day-Tarone test was used to assess the homogeneity of the ORs between the primary and replication data sets. We also examined whether sex had an effect on the

association using gender as a covariate. Results of these analyses are discussed further in the supplementary note.

b. *Meta-analysis of duplications and deletions of 16p11.2 in multiple psychiatric disorders*

The association of the microduplication and microdeletion was examined independently in each disorder using the CMH exact test with source as a stratifying variable. The Breslow-Day-Tarone was used to assess the homogeneity of the partial OR between the studies of each disorder used in the meta-analysis. A common *P*-value and OR were reported from the CMH exact test and from the logistic regression respectively only if there was homogeneity in the ORs between the studies in the meta-analysis. Given the very small number of deletion observations in the GAIN schizophrenia study and in each of the BD studies, approximate ORs were calculated by replacing the number of deletions n with $\{n + 0.5\}$. Results of these analyses are discussed further in the supplementary note.

8. **Analysis of Quantitative Clinical Features with 16p11.2 Rearrangements**

Quantitative clinical data on height, weight, and head circumference/ Occipital-Frontal Circumference was collected from records on 16p11.2 carriers in this study, in previously published studies (Weiss et al and Ghebranious et al) and from unpublished carriers ascertained by referral for global developmental delay (Tamim Shaikh. personal communication). We excluded from our analysis subjects with known Hispanic, Polynesian and African American ethnicity and any subjects with documented cytogenetic abnormalities.

OFC and 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 (URL 1). OFC percentile rankings were further verified using the online tool developed by Simulconsult, which is based on the same reference database (URL 2). Height and OFC percentiles were converted to Z-scores using online resources (URL 3). Z-scores were contrasted among 16p11.2 microdeletions versus microduplications carriers, using the Wilcoxon two-sample Rank Sum test. We repeated the analysis using the craniofacial normative database from Farkas et al.[40,41] (European ancestry.based). The above analysis was also performed within subsets of samples defined by their diagnoses schizophrenia (SCZ), developmental delay (DD) or autism spectrum disorders (ASD), to further examine if the observed effect was present within each individual group. The results of these analyses are discussed further in the supplementary note.

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 microdeletions and duplications were not examined.