www.sciencemag.org/cgi/content/full/science.1227764/DC1

Supplementary Materials for

Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders

Brian J. O'Roak, Laura Vives, Wenqing Fu, Jarrett D. Egertson, Ian B. Stanaway, Ian G. Phelps, Gemma Carvill, Akash Kumar, Choli Lee, Katy Ankenman, Jeff Munson, Joseph B. Hiatt, Emily H. Turner, Roie Levy, Diana R. O'Day, Niklas Krumm, Bradley P. Coe, Beth K. Martin, Elhanan Borenstein, Deborah A. Nickerson, Heather C. Mefford, Dan Doherty, Joshua M. Akey, Raphael Bernier, Evan E. Eichler,* Jay Shendure*

*To whom correspondence should be addressed. E-mail: shendure@uw.edu (J.S.); eee@gs.washington.edu (E.E.E.)

Published 15 November 2012 on *Science* Express
DOI: 10.1126/science.1227764

This PDF file includes

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.1227764/DC1)

Table S4. Sequences of primers, ASD1, and ASD2 MIP probes (as an Excel file)

Supplementary Materials:

Materials and Methods:

Design of molecular inversion probes. The general format for the molecular inversion probes (MIPs) used here is a common 30 bp linker flanked by an extension arm of 16 to 20 bp and a ligation arm of 20 to 24 bp (total MIP length always equals 70 bp). The unique arms of each MIP target a specific 112 bp genomic region for a gap-fill and circularization. A post-capture PCR amplification with primers corresponding to the common linker is used to append the Illumina sequencing adaptors (Illumina, San Diego, CA) and sample-specific barcodes. Analysis of a 55,000 MIP capture reaction previously revealed significant variation in individual probe capture efficiency (*7*). A predictor of capture efficiency based on the theoretical melting temperature (*24*) of probe targeting arms was trained on these empirical data and used to stratify potential MIP designs into one of three predicted performance categories (high, medium, and low) based on their extension and ligation arm sequences (fig. S1). MIPs for a given target region were chosen iteratively from the 5' to 3' end of the target region optimizing for predicted capture efficiency while minimizing the number of probes. First, a predicted high-capture efficiency MIP overlapping the 5' end of the target region with at least 50% of its gap-fill sequence is chosen. Each successive MIP is chosen to satisfy the following criteria in order of priority: 1) resides on the opposite strand as the previous MIP, 2) has <50% insert overlap with the previously selected MIP, 3) is a predicted high performer, 4) has minimal overlap with the previous MIP, 5) avoids single nucleotide variants (SNVs) from dbSNP in the targeting arms, and 6) avoids regions of high copy number (>10) in the reference genome. If no suitable MIPs can be picked due to SNVs, the MIPs were designed against both alleles. Read depth results from a 300,000 MIP experiment designed to cover the exome showed that the tetra-nucleotide sequence immediately 5' of the ligation junction had ~15-fold effect on probe capture efficiency. Therefore, MIPs containing a 4-mer in the bottom quartile of mean depth performance were excluded from future designs. Relevant scripts are available at http://krishna.gs.washington.edu/mip_pipeline/.

ASD1 MIPs were designed against the hg18 human genome reference using dbSNP129 to identity polymorphisms that might interfere with capture. All other MIP sets were designed using hg19 and dbSNP132.

Probes pooling and 5' phosphorylation. Individual MIPs were column synthesized as 70-mer oligonucleotides at the 25 nanomole scale and hydrated to 100 µM in IDTE (1X TE Buffer), pH 8.0 (Integrated DNA Technologies, Coralville, IA). For initial testing, probes were pooled at equimolar concentrations. Phosphorylations were performed in 1X T4 DNA ligase buffer with 10mM ATP (New England Biolabs (NEB), Ipswich, MA) and 100-200 units of T4 Polynucleotide Kinase (PNK) (NEB), incubating the reaction at 37°C for 45 min and then at 65°C for 20 min.

Multiplex capture of targeted sequences. Hybridization of MIPs to genomic DNA, gap filling, and ligation were performed in one 25 µl reaction of 1X Ampligase buffer (Epicentre, Madison, WI), with 50-120 ng of genomic DNA, the corresponding amount of MIPs to obey the ratio of 200 or 800 MIP copies to one haploid genome copy, 0.16-0.32 µM dNTPs, 0.4 units of Taq Stoffel Fragment (Applied Biosystems, Carlsbad, CA), and one unit of Ampligase (Epicentre). Reactions were incubated at 95°C for 10 min and at 60°C for 22-48 h. To degrade linear DNA, we added 2 µl of exonuclease mix (containing 10 units of exonuclease I (NEB) and 50 units of

exonuclease III (NEB) in 1X Ampligase buffer) and incubated the reaction at 37°C for 30-45 min and then at 95°C for 2 min.

Amplification of capture circles and library barcoding. We amplified the captured DNA by performing PCR in a 25-50 µl reaction using 5 µl of capture reaction, 0.5 µM SLXA_PE_MIPBC_FOR common primer (table S4), 0.5 µM different barcoded primer for each sample, and 1X iProof HF Master Mix (Bio-Rad, Hercules, CA) at 98°C for 30 s, X cycles of 98°C for 10 s, 60°C for 30 s, 72°C for 30 s, and finally 72°C for 2 min, where X was defined using a real-time PCR to determine the point at which the reaction plateaus.

Clean-up and pooling for Illumina sequencing. We pooled 5 µl of ~96 different libraries together and purified the pools with 1.8X AMPure XP beads (Beckman Coulter, Brea, CA) using the standard protocol. Libraries with excessive off-target captures were purified with an alternative protocol using 0.8X AMPure XP beads. Libraries were resuspended in 100 µl of 1X EB (Qiagen, Valencia, CA). We visualized the libraries on a 6% nondenaturing polyacrylamide gel as a control checkpoint and quantified each pool using a Nanodrop 1000. Multiple libraries were combined to create final megapools of 192-384 individual capture reactions. One lane of 101 bp paired-end reads was generated for each mega-pool on an Illumina HiSeq 2000 according to manufacturer's instructions. During confirmations, some parent captures were run on an Illumina MiSeq using the same protocol.

SNV calling, sensitivity, and positive predictive value (PPV). We evaluated MIP performance on a set of DNA samples from Joubert syndrome (JS) patients that had previously been Sanger sequenced at selected loci. We designed MIPs targeting 26 genes (1,429 probes, 86.4 kb target), including 16 known loci. We tested the uniformity of this set and then rebalanced by increasing the relative concentration of the bottom 16% of performers (235/1,429) by 10-fold. We then captured a set of 384 samples, including 350 JS affected samples (341 individuals with 9 technical replicates), 10 JS-like/Meckel-Gruber syndrome (MKS) samples, and 24 controls. Of these, 121 had bi-directional Sanger sequencing for at least a subset of the targeted genes. Our comparison of Sanger-based variant calls and MIP-based variant calls focused specifically on sites of rare variation, i.e., those present at a minor allele frequency (MAF) of <5% in the cohort, as the discovery of rare variation is the most likely application of resequencing (furthermore, common sites were not tracked in the available Sanger data).

Variants were called as previously described (*3*), but at relaxed coverage and quality thresholds. We evaluated MIP-based variant calling with a range of heuristic-based thresholds (table S6). Discordant and ambiguous Sanger-based variant calls were re-evaluated by performing an additional Sanger-based validation to establish a firm "gold-standard" truth set of 205 variant calls for which we calculated sensitivity (recall) and PPV (precision) with each heuristic. We then evaluated overlap performance using two F-measure formulas for either equal weighting of sensitivity (F1) and PPV or twice the weight for sensitivity (F2).

Copy number variant (CNV) calling, sensitivity, and PPV. We sought to assess whether relative read depth could be used to identify CNVs as has previously been demonstrated for MIP protocols directed at genotyping rather than exon capture (*25*). To this end, we designed MIPs targeting 32 epilepsy genes (1,325 probes, 86.5 kb target). The bottom 11% of the probes were rebalanced at a 10-fold greater concentration. After rebalancing, we captured and sequenced 380 samples, of which a subset (n = 15) were captured in duplicate and contained previously detected

CNVs. We normalized the read depth of each MIP to overall read depth and converted these values to standardized Z-scores. Low performing (median coverage <20X) or excessively noisy probes (relative standard deviation >25%) were removed. Only autosomal probes were considered. Samples with call rates greater than 10 standard deviations from the mean were removed, leaving 357 samples (including 14 duplicates). Similar to the CoNIFER method developed for exomes (*26*), we used singular value decomposition (SVD) to remove systematic noise. However, because the matrix was overrepresented for true signal events, only a single noise component was removed.

We used a sliding window caller to find regions with mean SVD adjusted Z-scores >=2.5 or <=-2.5. We then adjusted the window and minimum average SVD adjusted Z-score parameters to generate sensitivity and PPV estimates (fig. S6 and table S9).

All captures from ASD2 were called as above removing 20 SVD components. Calls with seven or more calls were considered. All proband calls were examined in independent array data, finding support for three out of four events (figs. S11-12).

Samples. For MIP capture, all ASD probands (n = 2,494) and related family member samples came from the Simons Simplex Collection (SSC). Detailed information on available phenotypes and inclusion criteria are available (https://base.sfari.org). The primary DNA source was whole blood; however, when not available, cell line pellet or spit-derived DNA was used. European-American ancestry healthy (non-ASD) individuals (n = 768) were obtained from the National Institute for Mental Health (NIMH) (https://www.nimhgenetics.org/available_data/controls/). The DNA source for all NIMH samples was cell line pellet derived.

Variant filtering annotation and confirmation. Initial 101 bp paired-end reads were trimmed to 76 bp for SNV calling. This avoided double counting of overlapping sequence and use of lower quality data from the end of reads. The full 101 bp reads were used for additional indel calling, as the read overlap can resolve insertions or deletions at the read junction and complex events, which might otherwise not align. Reads were mapped and consensus sequences called as in (*3*). ASD1 was mapped to hg18, while ASD2 and ASD1/2 captures were mapped to hg19. Missense variants were stringently filtered (>=25X coverage; >=Q30 quality; allele balance =<0.75). Due to their reduced number and potential high impact, we used a more relaxed filter for nonsense and splice variants (>=8X coverage; >=Q30 quality; allele balance =<0.75). All candidate truncation sites in probands and NIMH samples were confirmed by Sanger sequencing. Indel variants were called and filtered to those with at least eight reads and 25% support of the variant allele from both 76 and 101 bp reads.

To identify potential *de novo* events in probands, we filtered this call set against 1,779 exomes sequenced at the University of Washington and also removed recurrent variant sites present in the MIP call set. Rare disruptive sites (missense, nonsense, splice-site, and indel) were tested for inherited versus *de novo* status by MIP-based resequencing of their parents using either the full probe set or subsets of probes targeting rare variant sites. Variants that appeared *de novo* or failed parental capture were further validated with PCR and Sanger sequencing. Overall, we validated 97.7% (606/620) of the candidate sites as *bona fide* inherited variants (n = 579; validated by MIP- and/or Sanger-based sequencing of the parent-child trio) or *de novo* mutations (n = 27; validated by Sanger-based sequencing of the parent-child trio). Paternity was confirmed by tabulating Mendelian violations using very high-quality variant sites identified in the proband

(50X, Q100, allele balance <=0.7) across sequence data from all probes. Positions of domains (*23*) and mutations in the context of the protein structure were generated using DOG 1.0 (*27*).

Uniformity and target coverage. Uniformity and target coverage were calculated using trimmed 76 bp paired-end reads. Uniformity plots were generated from six sample captures. Read counts for each MIP were normalized by total reads mapped, averaged, and sorted in descending order. Log10 plots of the ranked uniformity were examined to identify poor performers, i.e., those at ~2 orders of magnitude lower abundance. For each probe set, the "target" was defined as the collapsed exon coding sequence for all RefSeq isoforms for each gene symbol and an additional 2 bp flanking the start and stop of each exon.

Protein-protein interaction (PPI) network reconstruction and null model estimation. PPI interactions were evaluated as in (*3*) using the severe proband *de novo* variants identified in the University of Washington (209 trios) (*3*) and Yale (225 trios) (*4*) SSC data sets. In total, 201 gene products were evaluated: 121 unique to University of Washington, 75 unique to Yale, and 5 in common (figs. S8-9).

Recurrent mutation analysis. We developed a probabilistic model that incorporates the overall rate of mutation in coding sequences, estimates of relative locus-specific rates based on human-chimpanzee fixed differences in each gene's coding and splice sequences, and other factors that may influence the distribution of mutation classes, e.g., codon structure. We used this model to define a matrix of weights reflecting the relative probability of observing mutations of each type (i.e., missense, nonsense, canonical splice-site, or indel) in each gene. In each of 500 million simulations, we applied these weights by randomly placing 2,172 simulated mutations into 19,008 genes across the exome using the multinomial distribution. We then estimated (from the 500 million simulations) the probability of observing X or more total events of which Y or more were nonsense, splice-site, or indels ("*trunc*") in any given gene (Fig. 2A). Importantly, the initial exome mutations (*2, 3*) were *excluded* from the observed counts.

The number of mutations in each simulation (n = 2,172) was based on the rate of protein-disrupting *de novo* events that we previously observed exome-wide (192 events in 209 probands, or 0.9187 events per proband) (*3*) and the number of probands screened exclusively by MIPs (n = 2,364).

To define the overall rate and mutation class weights for each gene, we first defined weights for each base position. All possible amino acid changes and the location of splice-sites for each RefSeq (hg 19) isoform were annotated for the 19,008 genes. Next, we calculated specific weights for observing different types of mutation events in each gene. These initial weights consisted of three components:

1) Locus-specific substitution rate, w_g, calculated from human and chimpanzee alignments as in (*3*). Only coding sequences (including splice-sites) were considered. For genes without an estimate or with an estimate of 0, we assumed a background rate of 1×10^{-12} per generation.
2) Relative substitution rate for point mutation, w_s. Time-reversibility was assumed in the substitution models (*28*). Factors, such as nucleotide composition (w_n) and transition/transversion ratios (*ti/tv)* for different mutation classes ($w_{F,\,ti/tv}$) were considered, i.e., $w_s = w_n \times w_{F,ti/tv}$.

The nucleotide-specific substitution rates were defined for the reference allele either AT or GC bases from Lynch (*15*):

$w_{n=AT} = 0.884$

$w_{n=CG} = 0.942$

For different mutation classes, relative transition substitutions were defined using the values from Tennessen et al. (*29*):

$w_{synonymous,ti} = 5.60$, $w_{missense,ti} = 2.31$, $w_{nonsense,ti} = 2.13$ and $w_{splice,ti} = 1.69$

For transversion substitutions:

$w_{synonymous,tv} = w_{missense,tv} = w_{nonsense,tv} = w_{splice,tv} = 1$

3) Relative rate for indel mutation, w_f. The relative indel rate compared with the substitution rate in coding regions is estimated as 0.0258 between human and chimpanzee (*30*). We defined $w_f = 0.0258 w_s$ for each base without consideration of insertion or deletion sequence length.

Thus, for each single base *b*, where the reference allele can be replaced by the other three nucleotides, the weights for different mutation classes can be calculated as:

$$w_{b,synonymous} = \sum_{synonymous} w_{g,b} w_{s,b} = \sum_{synonymous} w_{g,b} w_{synonymous,ti/tv,b} w_{n,b}$$

$$w_{b,missense} = \sum_{missense} w_{g,b} w_{s,b} = \sum_{missense} w_{g,b} w_{missense,ti/tv,b} w_{n,b}$$

$$w_{b,nonsense} = \sum_{nonsense} w_{g,b} w_{s,b} = \sum_{nonsense} w_{g,b} w_{nonsense,ti/tv,b} w_{n,b}$$

For indels, the weight for indel is:

$$w_{b,indel} = 0.0258 w_s = 0.0258(w_{b,synonymous} + w_{b,missense} + w_{b,nonsense})$$

For the special case, where a base is part of the canonical splice-site (e.g., 5'-GT…AG-3'), the weight for a splice change is:

$$w_{b,splice} = \sum_{splice} w_{g,b} w_{s,b} = \sum_{splice} w_{g,b} w_{splice,ti/tv,b} w_{n,b}$$

Otherwise, $w_{b,splice} = 0$.

If a base intersected with more than one isoform of a gene, we averaged the weights across isoforms for different types of mutation events.

With the weights defined for each base position in a gene, the overall gene weights for different classes of mutation events (*i*) (i.e., $w_{i,synonymous}$, $w_{i,missense}$, $w_{i,nonsense}$, $w_{l,splice}$ and $w_{i,indel}$) can be obtained by summing the individual base weights. In this study, we focused on the protein-altering events, so we defined weight for protein-altering events in gene *i* as: $w_{i,protein-altering} = w_{i,missense} + w_{i,nonsense} + w_{i,splice} + w_{i,indel}$.

Finally, we normalized the weights across genes such that they sum to 1. $w^*_{i,protein-altering} = \frac{w_{i,protein-altering}}{\sum_i w_{i,protein-altering}}$

After these normalized protein-altering weights for each gene were obtained, 500 million simulations were performed to evaluate the significance of the observed number of protein-altering events in the genes of interest. Specifically, in each simulation, 2,172 simulated protein-altering events (0.9187 x 2,364) were randomly introduced into 19,008 genes according to the multinomial distribution, using the normalized protein-altering weights for each gene ($w^*_{i,protein-altering}$) as the priors of the multinomial distribution. As we were interested in distinguishing between events falling into two classes (namely, a missense class and a more severe *trunc* class (nonsense, splice-site disruption and indels)), after a simulated event was assigned to one gene (e.g., gene *i*), we used the binomial distribution with the probability of:

$$\frac{w_{i,nonsense} + w_{i,splice} + w_{i,indel}}{w_{i,protein-altering}}$$

to assign this event to one of the two classes. Through simulation, we learned the distribution of 2,172 simulated events across the genes and compared this to the observed number of protein-altering (or severe) events. P-values were calculated for genes of interest based on 500 million simulations (note: for all robustness analyses, 1 million simulations were conducted). Specifically, let $X_i^{obs}, X_i^{sim}, Y_i^{obs}, Y_i^{sim}$ denote as the observed and simulated numbers of any protein-altering (*X*) or *trunc* (*Y*) events for gene *i*. The p-value for observing X_i^{obs} or more protein-altering events can be calculated as:

$$P_{i,all} = \frac{\#(X_i^{sim} \geq X_i^{obs})}{500{,}000{,}000}$$

Likewise, the p-value for observing X_i^{obs} or more protein-altering events, and among them Y_i^{obs} or more severe (*trunc*) events, can be calculated as:

$$P_{i,trunc} = \frac{\#(X_i^{sim} \geq X_i^{obs} \quad and \quad Y_i^{sim} \geq Y_i^{obs})}{500{,}000{,}000}$$

Analysis of in- versus out-of-network mutations. We calculated the relative expectation of in- versus out-of-network genes using weights defined for the simulation (0.51). This proportion effectively takes into account both differences in gene length and locus-specific mutation rates. We used this value to calculate the binomial probability of observing X or more events in Y trials. A similar expectation is obtained from only the transcript length (0.52).

Analysis of *de novo* mutations in unaffected siblings. We intersected the 44-gene candidate list with *de novo* calls from 629 unaffected siblings from the SSC (*3, 4, 6*). This included O'Roak et al. (n = 50), previously unpublished siblings from families in O'Roak et al. (n = 39), Sanders et al. (n = 197, note: three duplicate families removed), and Iossifov et al. (n = 343). This analysis identified a missense variant (*CNOT4*) and a synonymous variant (*SESN2*), both reported in

Iossifov et al. but not experimentally validated (table S14). We estimated the expected number of *de novo* events in the 44 genes by adjusting the parameters of the simulation to consider sampling 629 individuals.

Phenotype analysis. In order to examine the phenotypic presentation in individuals with identified mutations, summaries of patient characteristics—including cognitive ability, presence of comorbid medical and psychiatric disorders, presence of frank dysmorphology, and raw physical measurements (e.g., head circumference)—were culled from the SSC phenotype data distributions (SFARI.org) and presented in narrative form. Standardized head circumference scores (Z-scores) were calculated using norms established by Roche (*31*) to account for age and gender. Because there are no generally accepted normative statistics for head circumference in adults, we extrapolated upward for parents from normative scores at 18 years of age in the Roche sample. Standardized head circumference of those individuals with mutations of interest (*CHD8*, *DYRK1A*, and *PTEN*) was compared to the complete sample distribution as well as to other family members. To account for the small sample sizes of the mutations of interest and the unequal distributions between groups, a two-sample permutation test with Monte-Carlo approximation (replications = 10,000, alternative hypothesis = two-sided) of the exact conditional distribution was calculated following Hothorn (*32*). Analysis of variance was calculated to examine differences in head circumference between probands and other family members. This was calculated using all family members and calculated only comparing probands to same-sex family members.

Supplementary Online Text:

Capture method, optimization, and performance. We previously demonstrated the feasibility of "library-free" MIP-based targeted resequencing of ~55,000 exons using capture oligonucleotides that were cost-effectively synthesized on and released from a DNA microarray (*7*). However, the method suffered from several critical limitations. (a) Microarray-derived MIPs are difficult to produce at a scale that would support their use in thousands of samples. (b) The non-uniform synthesis and amplification of microarray-derived MIPs negatively impacted the performance of targeted capture. (c) High-quality microarray-derived oligonucleotide libraries are not yet broadly accessible. We developed a modified workflow that enables cost-effective capture and sequencing of a limited number of candidate genes across thousands of individuals (table S1). This method utilizes novel algorithms for MIP design against arbitrary targets (fig. S1), column-synthesized oligonucleotides, optimized capture steps and conditions, and massive multiplexing (192 or 384 samples per HiSeq 2000 lane) (Fig. 1A).

We evaluated this workflow by designing 330 MIPs tiling the coding regions of six genes (ASD1) (330 probes; ~24 kb target sequence) identified as candidates for ASD from the exome sequencing of 20 parent-child trios (*2*) (table S4) with an algorithm that exploited the relative performance of previously evaluated MIP probes (*7*) to guide selection (*12*). Optimization established a one-step capture protocol requiring 50 ng of genomic DNA as input. After "rebalancing" the pool of column-synthesized MIPs on the basis of empirically measured capture efficiency (*33*), we performed targeted capture and sequencing on 16 parent-child trios that had previously been exome sequenced (fig. S2). Following PCR amplification with sample-specific

barcodes, samples were pooled and sequenced on a fraction of an Illumina GAIIx lane (PE76 + 8 bp index).

Rebalancing brought the relative capture efficiencies for 90% of the MIPs from a 90-fold range into a 15-fold range (fig. S3A) with an average 896-fold median target coverage and, on average, 96% of the target covered at >25-fold (fig. S2). Overall, 725/732 (99%) SNVs detected by exome sequencing were also detected by MIP-based resequencing. The data were also consistent with a higher sensitivity for the MIP assay (table S5). Additional probes (n = 25) targeted to poorly captured regions were added to the ASD1 set, before capturing additional samples.

To further assess sensitivity and PPV, we evaluated performance on a set of DNA samples from the Joubert syndrome (JS) patients that had previously been Sanger sequenced at selected loci. We designed MIPs targeting 26 genes (1,429 probes; ~86 kb target), including 16 known JS loci. After rebalancing, we captured and sequenced 384 samples, of which 121 had high-quality bi-directional Sanger sequencing data available for at least a subset of the targeted genes, in two Illumina HiSeq lanes (PE101 + 8 bp index) (table S2). We evaluated variant calling under heuristics focused specifically on sites of low frequency or rare variation (MAF < 5%) (table S6). Considering all such sites, including those supported by <10 reads, the F1 values (a measure of accuracy) maximized at 93.7% sensitivity and 98.0% PPV (fig. S4). At sites supported by ≥10 reads, the F1 values maximized at 99.5% sensitivity and 98.0% PPV (fig. S4).

As the capture efficiencies of individual MIPs are highly reproducible (*7, 8*), we also assessed whether relative read depth could be used to identify CNVs (*25*) with MIPs targeting 32 epilepsy genes (1,325 probes; ~87 kb target) (table S2). After rebalancing, we captured and sequenced 380 samples, of which a subset (n = 15) were captured in duplicate and contained previously detected CNVs. After normalizing the read depth of each MIP and applying SVD (*26*), we used a sliding window caller and minimum calling thresholds. A total of 357 samples (including 14 duplicates) remained after removing samples with excessive noise (*12*), we confirmed 10 of 14 known sites (table S7), with false negative sites corresponding to regions without a sufficient number of well-performing probes covering the event (fig. S5, table S8). In addition, there was strong concordance between replicates (table S7). Although limited by the number of sites, we applied an approach analogous to SNVs to estimate sensitivity and PPV under different calling thresholds (fig. S6, table S9) (*12*). Using a seven-probe window with a mean SVD adjusted absolute Z-score greater than or equal to 2.7 and considering all sites in the MIP target, the sensitivity was 71% and PPV 83% (fig. S6). Excluding the four regions without a sufficient number of well-performing probes the sensitivity was 100% and PPV 83%. Overall, these data demonstrate that we can sensitively detect large single exon as well as multi-exon copy number alterations with a reasonably high sensitivity and PPV (fig. S7).

Supporting Data. We find strong evidence of mutation burden and a significant skew of *de novo* mutation events toward severe class mutations. Six genes (*CHD8*, *GRIN2B*, *DYRK1A*, *PTEN*, *TBR1*, and *TBL1XR1*) showed evidence of mutation burden at an alpha of 0.05 after applying a Holm-Bonferroni correction for multiple testing (Fig. 2A). Notably, three additional in-network genes—*ADNP*, *ARID1B*, and *CTNNB1* (*β-catenin*)—each have a single severe class mutation, providing suggestive evidence for these genes but insufficient data to account for multiple testing (uncorrected $p < 0.05$).

Several additional analyses support these conclusions. First, although we conservatively used the highest available empirical estimate of the overall mutation rate in coding sequences (*3*), the significance of five of the six implicated genes (all except *TBL1XR1*) is robust to a doubling of the estimated mutation rate (as is the significance of the overall burden). Second, the results are similarly robust to reasonable variations in the estimated locus-specific mutation rates. Specifically, we performed a bootstrap analysis (n = 1,000) and calculated 95% confidence intervals (CI) for the locus-specific mutation rate of each of the 44 genes. The upper-limit value of the 95% CI is, on average, ~1.9-fold higher than the estimated rate (3-fold for the 99th percentile). We recalculated the probabilities for each of the genes sequenced here using the 95% CI rates by substituting this rate for a single gene in the simulation, i.e., changing the rate for just the gene in question. These ranged from 1.5- to 3-fold of the estimated rate (table S13). Results for five of the six implicated genes were again significant (the exception once more being the borderline locus *TBL1XR1*). Similarly, increasing the locus-specific rate by 2- or 3-fold (regardless of the CI estimates) showed the same five of six as remaining significant. At 4-fold (outside of all CI estimates for these genes), four genes are significant (*TBR1* and *TBL1XR* fail). At 6-fold, only *CHD8* and *DYRK1A* are significant. Above 8-fold, only *CHD8* remains significant. Finally, we further tested this model using mutation type specific Ti/Tv values from the published SSC unaffected siblings (synonymous: 4.04, missense: 2.09, nonsense: 2.33, splice: 2.00) (*3, 4, 6*) rather than the site-based values from standing variation (*29*). The results were equivalent with the same six implicated genes.

We further evaluated the robustness of these results under an alternative model using empirical estimates of sequence composition-dependent *de novo* rates estimated from recent whole-genome data (*16*), rather than Ti/Tv rates for different mutation classes. Specifically, we modified the simulation (item 2, w_s.) to use the observed dinucleotide transition and transversion mutation rates. The CpG dinucleotides were defined based on reference human sequence (hg19).

Rates for non-CpG:

$$w_{non-CpG,A\leftrightarrow G/C\leftrightarrow T} = 6.18\times10^{-9}$$
$$w_{non-CpG,A\leftrightarrow C/G\leftrightarrow T} = 1.90\times10^{-9}$$
$$w_{non-CpG,A\leftrightarrow T} = 8.24\times10^{-10}$$
$$w_{non-CpG,C\leftrightarrow G} = 1.03\times10^{-9}$$

Rates for CpG:

$$w_{CpG,A\leftrightarrow G/C\leftrightarrow T} = 1.12\times10^{-7}$$
$$w_{CpG,A\leftrightarrow C/G\leftrightarrow T} = 4.85\times10^{-9}$$
$$w_{CpG,A\leftrightarrow T} = 0$$
$$w_{CpG,C\leftrightarrow G} = 4.72\times10^{-9}.$$

Using these rates and the previous framework, we calculated the analogous weights (i.e., missense, nonsense, canonical splice-site, or indel) for each base and then each gene (*12*). P-values were calculated for genes of interest based on one million simulations (*12*). Again, we obtained the same six implicated genes, with similar probabilities to the original model (max 2-

fold change in any direction for any gene). Of note, the results for *TBL1XR1*, were very slightly strengthened under this model (p=0.000836 with the dinucleotide model versus p=0.001173 with original model).

In addition to the simulation data, exome sequencing of 629 unaffected siblings of ASD probands (including 39 previously unpublished exomes) identified only one nonsynonymous *de novo* mutation across all 44 genes (a missense mutation in *CNOT4*; table S14) (*3, 4, 6*), consistent with expectation based on our estimated locus-specific mutation rates (mean expected missense n = 1.3; mean expected severe class n = 0.16). Finally, in the six implicated genes, no loss-of-function variants were observed in any of 629 exome-sequenced unaffected siblings, 762 MIP-sequenced non-ASD individuals, or 6,500 exome-sequenced individuals from non-ASD cohorts (*29*).

Genotype-phenotype correlations. We examined head size using age and sex normalized head circumference (HC) Z-scores in individuals with protein-truncation or splice-site mutations (*12*) (Fig. 2B). For *CHD8* we observed significantly larger head sizes relative to those individuals screened without *CHD8* mutations (two-sample permutation test, Z = 3.46, two-sided p = 0.0007, 99% CI for p = 0.0002-0.0017; mean *CHD8* probands HC = 2.24; SD = 0.61; mean overall proband HC = 0.70; SD = 1.32). Of note, *de novo CHD8* mutations are present in ~2% of macrocephalic (HC > 2.0) SSC probands (n = 366), suggesting a useful phenotype for patient subclassification. We also found that individuals with *de novo DYRK1A* mutations (n = 3) have significantly smaller head sizes relative to those individuals without *DYRK1A* mutations (two-sample permutation test, Z = -3.65, two-sided p = 0.0005, 99% CI for p = 0.0001-0.0014; mean *DYRK1A* probands HC = -2.72; SD = 1.07). Comparison of head size in the context of the families (Fig. 2C-D, table S17) provides further support for this reciprocal trend with *CHD8* disruptive mutations associating with macrocephaly (ANOVA: F(3,25) = 2.0, p = 0.1) and *DYRK1A* mutations with microcephaly (ANOVA: F(3,8) = 10.4, p = 0.004). These findings are consistent with case reports of patients with structural rearrangements and mouse transgenic models of *DYRK1A* and *CHD8*, which implicate these genes as regulators of brain growth (*18-21*). Additionally, macrocephaly was also observed in individuals with *de novo* and inherited *PTEN* mutations (HC range = 1.96-6.13) (*22*).

Connected biological functions for implicated genes. Five of the six genes with strong support fall within the previously reported (or expanded) β-catenin/chromatin remodeling protein network. The exception, *PTEN*, has been shown to be mutated in children with macrocephaly, ID, or ASD (*22*). A recent report identified a novel role for *PTEN* in regulating neurogenesis in the adult mouse hippocampus (*34*). There have now been several reports linking *CHD8* to β-catenin/Wnt regulation and neurodevelopment (*35-37*). A recent report of a *CHD8* translocation patient noted a phenotype consisting of macrocephaly, prominent forehead, and shallow supraorbital ridges in addition to ASD, consistent with the macrocephaly and autistic features observed in individuals with point mutations (*21*). Three patients with large overlapping *de novo* deletions at 14q11.2 (including *CHD8*) have been reported with shared dysmorphic features, developmental delay, and cognitive impairment, but apparently no autistic features (*38*). Macrocephaly was present in the individual with the smallest deletion encompassing *SUPT16H* and *CHD8*. TBR1 is a transcription factor reported to regulate several ASD candidate genes in mouse, including *GRIN2B*, through interactions with CASK (*39*). TBR1 plays a fundamental

role in specifying the identity of postmitotic cortical neurons (*39*), while GRIN2B is a subunit of the N-Methyl-D-Aspartate (NMDA) ionotropic glutamate receptor and has also been associated with developmental delay, autism, and schizophrenia (*40, 41*). *DYRK1A* has been a strong candidate gene underlying some of the effects of Down syndrome, plays a still poorly understood role in neurodevelopment and brain/head size (*42*), and is under positive selection in the human linage (*43*). Recently, DYRK1A has also been shown to be a positive regulator of p120-catenin signaling (*44*). Finally, it has been demonstrated that TBL1XR1, also known as TBLR1, is required for β-catenin/Wnt -mediated transcription (*45, 46*).

Clinical Information:
Family 11654. Proband (female) is the eldest of two children with a younger sister (11654.s1).
Patient ID: 11654.p1 (proband)
Event: *CHD8 de novo* splice mutation

Patient is a 96-month-old non-Hispanic white female diagnosed with autism. Mother experienced edema during pregnancy. Labor was induced and a C-section was performed. Extremely low VIQ (47) and NVIQ (41) with clinical range deficits in social responsiveness and adaptive skills. She experienced a delay in phrase speech. Possible loss of skills reported during development. Placed on a gluten free, casein free diet for one year at age 4.5. Elevated attention problems, affective problems, somatic complaints, and thought problems with a report of current sleep apnea. History of past antiepileptic and antibiotic medication use. No history of seizures; MRI/CT and EEG at age 5 apparently normal. Currently taking antihypertensive medication. Family history of migraines (paternal grandfather) and reading disorder (paternal cousin, paternal aunt/uncle). Head circumference of 55.2 cm (z =2.3) and normal BMI.

Patient ID: 11654.fa (father)
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 29. Normative range of social responsiveness; no presence of broader autism phenotype. He has a B.A. degree and reports an annual household income of $101-130k. No medication use endorsed for current or past. No comorbid diagnoses endorsed. Father has head circumference of 58.8 cm (z=1.2) and BMI suggestive of being overweight.

Patient ID: 11654.mo (mother)
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 29. Normative range of social responsiveness; no presence of broader autism phenotype. She has a B.A. degree and reports an annual household income of $101-130k. No medication use reported for mother before, during, or after pregnancy except for epidural during labor. No comorbid diagnoses endorsed. Mother has head circumference of 53.85 cm (z=-0.5) and normative BMI.

Patient ID: 11654.s1 (sibling)
Summary:

Sibling is a non-Hispanic white 6.5-year old female. Normative adaptive scores and social responsiveness from parent and teacher noted. No behavioral problems reported. No medication use endorsed for current or past. No comorbid diagnoses endorsed. Head circumference of 51.4 cm (z=0.2) and normative BMI.

Family 12991. Proband (male) is an only child.
Patient ID: 12991.p1 (proband)
Event: *CHD8 de novo* frameshift mutation

Patient is a 151-month-old non-Hispanic white male diagnosed with autism. Mother took oral fertility medication, which was successful. Labor was augmented with Pitocin and a C-section was performed due to failure to progress. Low VIQ (60), NVIQ (67), and adaptive (73) scores.

Clinical range deficits in social responsiveness and elevation in anxious/depressed mood with no comorbid diagnoses. Hernia surgery at 10 months. Diagnosed as excessively clumsy/uncoordinated at 3. Chronic diarrhea from age 5 to 8. Abnormal EEG and seizures at 12, concurrent with head injury. Currently taking antiepileptic medication. Past antidepressant use, antibiotic use, and gluten free casein free diet. Proband has endorsement of mixed receptive-expressive language disorder and pragmatic language disorder. Head circumference of 58 cm (z =2.7) and BMI suggestive of being overweight. Family history of speech delay (proband and maternal cousin), migraine (maternal grandparent), and stroke and cancer (specific family members unknown).

Patient ID: 12991.fa (father)
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 23. Normative range of social responsiveness; no presence of broader autism phenotype. Father is a high school graduate with an annual household income of $51-65k. Past antibiotic and painkiller use. Past tobacco and marijuana use reported. Current alcohol use. No comorbid diagnoses endorsed. Father has head circumference of 58.2cm (z=0.8) and BMI suggestive of being obese.

Patient ID: 12991.mo (mother)
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 22. Normative range of social responsiveness; no presence of broader autism phenotype. Mother is a high school graduate with an annual household income of $51-65k. Currently taking birth control and antidepressants. Past tobacco and marijuana use reported. Current alcohol use. No comorbid diagnoses endorsed. Mother has head circumference of 55.7 cm (z=0.9) and BMI suggestive of being overweight.

Family 14016. Proband (male) has one older sister (14016.s1) and two younger sisters (not participating).
Patient ID: 14016.p1 (proband)
Event: *CHD8* nonsense mutation

Patient is a 63-month-old non-Hispanic white male diagnosed with autism. He is one of four children born to the same parents, and the third of six pregnancies (mother's first and fourth pregnancies resulted in miscarriage within the first 13 weeks). Mother had an unspecified viral illness during pregnancy. Proband was induced at birth with Pitocin due to failure to progress. Meconium staining at birth. Low VIQ (79), average NVIQ (92), and low adaptive (58) scores. Clinical range deficits in social responsiveness, externalizing behaviors (oppositional and aggressive behaviors) and internalizing behaviors (affective problems, anxiety problems). History of speech delay, febrile seizures, and allergies (strawberries, sunscreen and soaps/lotion). History of intermittent unusual stool, diarrhea, and constipation during early childhood. Diagnosed as excessively clumsy/uncoordinated but EEG and MRI/CT within normal range. Currently taking antibiotics and a sedative. Took anti-inflammatory meds in the past. Family

history of asthma (father, sibling) and stuttering (maternal aunt/uncle). One maternal cousin suspected of ASD. Head circumference of 55.4 cm (z =2.5) and normal BMI.

Patient ID: 14016.fa (father)
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 23. Normative range of social responsiveness; no presence of broader autism phenotype. Some college with annual household income of $66-80K. Past antibiotic use but no other medications or comorbid diagnoses endorsed. Father has head circumference of 60 cm (z=1.8) and BMI suggestive of obesity.

Patient ID: 14016.mo (mother)
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 22. Normative range of social responsiveness; no presence of broader autism phenotype. Some college with annual household income of $66-80K. No past medication use or comorbid diagnoses endorsed. Mother has head circumference of 59.4 cm (z=3.6) and BMI suggestive of obesity.

Patient ID: 14016.s1 (sibling)
Summary:

Sibling is a non-Hispanic white 8-year old female. Normative internalizing behaviors and borderline range of externalizing behaviors (oppositional and aggressive behaviors) reported. Borderline impaired social responsiveness and above average adaptive skills per parent report. Currently takes medication for asthma with history of antibiotic and anti-inflammatory medication use in the past. No comorbid diagnoses endorsed. Head circumference of 53.3 cm (z=0.9) and normative BMI.

Family 13986. Proband is the third of five children born to the same mother with a younger sister (13986.s1).

Patient ID: 13986.p1 (proband)

Event: *CHD8* frameshift mutation

Patient is a 67-month-old non-Hispanic white male diagnosed with autism. Mother's first pregnancy resulted in miscarriage (2nd trimester). Mother smoked cigarettes during the first trimester. She experienced an upper respiratory infection during pregnancy (trimester unknown). Labor was induced/augmented with a cervical suppository and Pitocin due to failure to progress. Hyperbilirubinemia at birth but no treatment. Proband is nonverbal with extremely low VIQ (25), NVIQ (38), and adaptive (57) scores. Elevated internalizing behaviors (withdrawn). Clinical range deficits in social responsiveness per parent and teacher report. Currently on a gluten free casein free diet. Allergic to nuts. Has taken antibiotics in past. Currently taking clonidine and cyproheptadine. Respiratory problems (retraction in lungs) diagnosed at 12 months. Diagnosed with cerebral palsy and excessively clumsy/uncoordinated at 18 months. MRI/CT normal at 18 months; EEG normal at 24 months. Family history of diabetes Type 2 (paternal grandfather),

migraines (maternal aunt/uncle), depression (maternal aunt/uncle, grandparent), and anxiety (paternal grandparent). Head circumference of 55 cm (z =2.0) and normal BMI.

Patient ID: 13986.fa (father)
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 29. Normative range of social responsiveness; no presence of broader autism phenotype. Associates degree with annual household income of $36-50K. Past tobacco and marijuana use reported. Current alcohol use. Has taken antibiotics in the past. History of high cholesterol. No comorbid diagnoses endorsed. Father has head circumference of 58 cm (z=0.7) and BMI suggestive of being obese.

Patient ID: 13986.mo (mother)
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 25. Clinical range of social responsiveness, but no presence of broader autism phenotype. Some college with annual household income of $36-50K. No medication use and no comorbid diagnoses endorsed. Current alcohol use. Mother has head circumference of 56.25 cm (z=1.3) and BMI suggestive of being overweight.

Patient ID: 13986.s1 (sibling)
Summary:

Sibling is a non-Hispanic white 4-year old female. Normative adaptive scores and social responsiveness from parent noted. No behavioral problems. Has taken antibiotics in the past. No comorbid diagnoses endorsed. Head circumference of 51.5 cm (z=1.1) and normative BMI.

Family 12714. Proband is an only child.
Patient ID: 12714.p1 (proband)
Event: *CHD8 de novo* nonsense mutation

Patient is a 55-month-old non-Hispanic white male diagnosed with autism. Mother consumed artificial sweetener during all three trimesters; no medications or other substances consumed during pregnancy. Testing during pregnancy revealed some abnormalities (unknown). Mother had x-ray/radiation exposure during 2nd trimester. Labor was augmented by maternal request with Pitocin. Hyperbilirubinemia at birth but no treatment. Low VIQ (75), NVIQ (78), and below average adaptive (80) scores. Clinical range deficits in social responsiveness per parent and teacher report. Speech delay with possible loss of language skills during development. Elevated internalizing behaviors endorsed (social withdrawal and somatic complaints) with no comorbid diagnoses. Diagnosed with chronic constipation at 1.5 years, which was resolved at 2.5 years. History of chronic loose stool. On a gluten free casein free diet. Past antibiotic use. Family history of diabetes Type 2 (maternal grandparent), heart disease and cancer (maternal grandparent, paternal grandparent). Head circumference of 53 cm (z =1.0) and normal BMI.

Patient ID: 12714.fa (father)
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 28. Normative range of social responsiveness; no presence of broader autism phenotype. Father has a BA degree with annual household income of 101-130K. Past use of antibiotics and acne medication. Past use of tobacco and current use of alcohol and marijuana. Possible problem with alcohol (endorsed trying to cut down). No comorbid diagnoses endorsed. Father has head circumference of 57.6 cm (z=0.5) and normal BMI.

Patient ID: 12714.mo (mother)
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 28. Normative range of social responsiveness; no presence of broader autism phenotype. Mother has an Associate's degree with annual household income of 101-130K. Past use of tobacco, hallucinogens, and inhalants. Current use of alcohol and marijuana. Possible problem with alcohol (endorsed "trying to cut down"). No comorbid diagnoses endorsed. Mother has head circumference of 58 cm (z=2.6) and normal BMI.

Family 14233. Proband is the third of four children born to the same mother with a younger brother (**14233.s1**).
Patient ID: 14233.p1 (proband)
Event: *CHD8* frameshift mutation

Patient is a 201-month-old non-Hispanic white male diagnosed with autism. Mother diagnosed with anemia during second and third trimesters and took iron supplements. She had an unknown viral illness during the second trimester. She also reports edema during pregnancy. Proband had hyperbilirubinemia at birth but no treatment. Extremely low VIQ (6), NVIQ (19), and adaptive (39) scores. Clinical range deficits in social responsiveness per parent and teacher report. Loss of language skills (words) during early development and elevated attention problems. History of suspected meningitis (per doctors in Mexico). MRI/CT and EEG normal range. History of suspected head injury (by parent at age 1). History of sleep problems (disordered breathing, sleep apnea, daytime sleepiness). Past use of antibiotics and antiepileptic medication; current use of mood stabilizers, dipentum and loratidine. Allergic to pollen. Adenoidectomy, tonsillectomy, and hernia surgery. Family history of migraines and reduced articulation (maternal cousin), cancer (paternal grandparent), ADHD (sibling), congenital heart defect (maternal aunt/uncle). Family history of diabetes Type 2 (father and paternal grandparent). Maternal grandparent has history of hypothyroidism, bowel disorder, anxiety, depression, and psychiatric hospitalization. Head circumference of 60.8 cm (z =3.0) and normal BMI.

Patient ID: 14233.fa (father)
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 37. Normative range of social responsiveness; no presence of broader autism phenotype. Father has an Associate's degree with annual household income of 101-130K. Past use of antibiotics, pain

killers, and medication for acne. Currently takes medication for diabetes, high cholesterol, and allergies. Father has head circumference of 61 cm (z=2.4) and BMI suggestive of obesity.

Patient ID: 14233.mo (mother)
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 29. Normative range of social responsiveness; no presence of broader autism phenotype. Mother has a graduate degree with annual household income of 101-130K. No current medication use and no comorbid diagnoses endorsed. Mother has head circumference of 58 cm (z=2.6) and BMI suggestive of being overweight.

Patient ID: 14233.s1 (sibling)
Summary:

Sibling is a non-Hispanic white 14-year old male. Normative adaptive scores and social responsiveness per parent report. No behavioral problems. Currently takes medication for asthma; past antibiotic use. No comorbid diagnoses endorsed. Head circumference of 60.2 cm (z=2.9) and normal BMI.

Family 13844. Proband is second of three children with an older sister (13844.s1) and younger brother (13844.s2).
Patient ID: 13844.p1 (proband)
Event: *CHD8 de novo* nonsense, *CUBN de novo* nonsense, 2X inherited CNV

Summary: Patient is a 99-month-old non-Hispanic white male diagnosed with autism. Extremely low VIQ (20), NVIQ (34), and adaptive (59) scores. Clinical range deficits in social responsiveness (120). Possible loss of language skills during development and elevated social withdrawal behaviors with no comorbid diagnoses. Large head (z = 2.62) and normal BMI. Food allergies (gluten and casein). Gastrointestinal constipation diagnosis with bloating and abdominal pain. Roseola diagnosed at 2.5 years and Epstein Barr virus contracted at 8 years. Respiratory problems diagnosed at 11 months and kidney problems diagnosed at 9 months. No diagnosis of cardiac or metabolic syndromes noted. No report of congenital anomalies. Family history of Down syndrome (maternal cousin). NICU admission shortly after birth with oxygen treatment. Meconium aspiration at birth. Family history among several members for migraines. Currently on GFCF diet. Took asthma medication in the past but not currently.

Patient ID: 13844.fa (father)
Summary: Father is an adult non-Hispanic white male. Age at conception of proband is 40. Normative range of social responsiveness, but elevated score for rigidity on broader autism phenotype. Some signs of alcoholism (use, attempting to cut down, annoyed by criticism about drinking, feeling bad about drinking, eye opening experience). No medication use endorsed for current or past. Some college education. Annual household income = $101–130K. Father has head circumference of 58.5 cm (z = 1.57) and normative BMI. No comorbid diagnoses endorsed.
Patient ID: 13844.mo (mother)
Summary: Mother is an adult non-Hispanic white female. Age at conception of proband is 35. Normative range of social responsiveness. No evidence of broader autism phenotype. Antibiotics

taken during second trimester of pregnancy with proband. Currently taking thyroid medication and antidepressant (not taken during pregnancy). Endorsement of current tobacco use and past marijuana use. Some college education. Annual household income = $101–130K. Mother has head circumference of 54 cm (z = -.41) and normative BMI. No comorbid diagnoses endorsed.

Patient ID: 13844.s1 (sibling)

Summary: Sibling is a non-Hispanic white 10-year-old female. Normative adaptive scores and social responsiveness from parent and teacher noted. Behavioral elevations for somatic problems and complaints. Mother was prescribed an unspecified hormone treatment to aid with growth in past (not currently taking). No other endorsement of medication use. Head circumference of 54 cm (z = 0.96) and normative BMI. No comorbid diagnoses endorsed. Cognitive decline following Epstein-Barr virus reported by parents.

Patient ID: 13844.s2 (sibling)

Summary: Sibling is a non-Hispanic white 5-year-old male. Adaptive scores not available. Normative social responsiveness from parent and teacher. No behavioral elevations across any domain. No endorsement of medication use. Head circumference of 52 cm (z = 0.15) and BMI suggestive of being underweight. No comorbid diagnoses endorsed.

Family 12752. Proband is an only child.

Patient ID: 12752.p1

Event: *CHD8 de novo* frameshift mutation, *ETFB de novo* nonsense, *IQGAP2 de novo* nonsense

Summary: Patient is a 55-month-old non-Hispanic white female diagnosed with autism. Normative range VIQ (90) and NVIQ (93) with low adaptive behavior skills (59). Clinical range deficits in social responsiveness (90). Clinical elevations in attention problems, internalizing problems, and affective problems with no comorbid diagnoses. Large head (z = 2.40) and BMI indications of being underweight. No loss or regression of language skills. Diagnosis of chronic constipation, ongoing from 3.5 months with intermittent episodes of abnormal stool. Coordination problems noted since 3.5 months. No cardiac or metabolic syndromes noted. No report of congenital anomalies. Hyperbilirubinemia diagnosis with phototherapy shortly after birth, no complications after treatment.

Patient ID: 12752.fa (father)

Summary: Patient is an adult non-Hispanic white male. Age at conception of proband is 38. Normative range social responsiveness. Elevated score for aloofness and pragmatic social skills. Diagnosis of diabetes. Current tobacco and alcohol use endorsed. Current and past use of antihypertensive meds and medication for high cholesterol. Past use of sedatives and pain killers. Some college education. Annual household income = $36–50K. Father has a head circumference of 59.5 cm (z = 1.56). BMI information unavailable.

Patient ID: 12752.mo (mother)

Summary: Patient is an adult non-Hispanic white female. Age at conception of proband is 36. Normative range of social responsiveness. No presence of broader autism phenotype. No endorsement of medications currently or during pregnancy with proband. Current tobacco and alcohol use endorsed. Some college education. Annual household income = $36–50K. Mother has head circumference of 54 cm (z = -.41). BMI information unavailable. Mother has been diagnosed with heart disease.

Family 14406. Proband is the eldest of two children with a younger brother (14406.s1).

Patient ID: 14406.p1

Event: *CHD8 de novo* 3 bp deletion

Patient is a 163-month-old non-Hispanic white male diagnosed with autism. Mother had herpes 1 virus during first, second, and third trimesters of pregnancy; UTI during 2nd trimester; and gestational diabetes during pregnancy. No medications taken by mother during pregnancy. Pitocin used to augment labor due to failure to progress. Proband has history of speech delay and a hearing accommodation at school. Below average VIQ (84), average NVIQ (98), and very low adaptive (66) scores. Clinical range deficits in social responsiveness per parent report. Elevated social problems, attention problems, affective problems, and social withdrawal behaviors with no comorbid diagnoses. Past gluten free, casein free diet and vitamins. Past antibiotics. Currently has sleep problems (frequent awakenings at night and difficulty sleeping). Head circumference of 57 cm (z =1.6) and BMI suggestive of obesity. Family history of heart disease (maternal grandparent, paternal grandparent, maternal aunt/uncle), cancer (maternal grandparent, paternal grandparent, paternal cousin), ADHD (paternal cousin), and diabetes type 2 (maternal grandparent).

Patient ID: 14406.fa

Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 25. Some evidence of broader autism phenotype (elevated pragmatic language impairment and aloof behaviors). Associates degree with annual household income of $66-80K. Past antibiotic and anti-inflammatory use. No comorbid diagnoses endorsed. Father has head circumference of 58 cm (z=0.7) and BMI suggestive of being overweight.

Patient ID: 14406.mo

Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 26. Normative range of social responsiveness; no presence of broader autism phenotype. Associates degree with annual household income of $66-80K. No current or past medication use reported. No comorbid diagnoses endorsed. Mother has head circumference of 54.5 cm (z=-0.04) and BMI suggestive of obesity.

Patient ID: 14406.s1

Summary:

Sibling is a non-Hispanic white 139-year old male. Normative adaptive scores and social responsiveness per parent report. Slight behavioral elevations for somatic problems and social withdrawal behaviors per parent report. Teacher report indicates borderline elevation in social responsiveness. Currently taking antibiotics and Valtrex (herpes 1). No comorbid diagnoses endorsed. Head circumference of 55 cm (z=0.7) and normative BMI.

Family 12099. Proband is the younger of two children with an older brother (12099.s1).
Patient ID: 12099.p1
Event: *DYRK1A de novo* frameshift mutation

Patient is a 96-month-old non-Hispanic white male diagnosed with autism. Mother had herpes 2 virus during third trimester and took Valtrex. Mother reported drinking during third trimester (unknown amount). Born via planned C-section. Meconium staining at birth. Polydactyly. Mother reports he was floppy and lethargic as an infant. Feeding difficulties (poor suck). Very low VIQ (63), NVIQ (55), and low adaptive (74) scores. Clinical range deficits in social responsiveness and elevated attention problems per parent report. History of febrile seizures and speech delay. Comorbid diagnosis of Mild Mental Retardation. Excessively clumsy, uncoordinated. Past use of antibiotics. Head circumference of 47.6 cm (z =-3.8) and normative BMI. Family history of seizures (paternal half sibling, maternal aunt/uncle), migraines (maternal grandparent), heart disease (maternal grandparent), cancer (maternal grandparent, paternal grandparent), mental retardation (paternal aunt/uncle), bipolar disorder (maternal aunt/uncle), depression (maternal aunt/uncle), dysthymic disorder (paternal half sibling), psychiatric hospitalization (paternal half sibling), and hyperthyroidism (maternal grandparent).

Patient ID: 12099.fa
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 55. Normative range of social responsiveness; no presence of broader autism phenotype. BA degree with annual household income of $81-100K. Past use of antibiotic medication; current use of antihypertensive medication and medication for high cholesterol. Current use of alcohol and reports possible problem (felt bad about drinking behavior). No comorbid diagnoses endorsed. Father has head circumference of 58.5 cm (z=1.0) and BMI suggestive of being obese.

Patient ID: 12099.mo
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 39. Normative range of social responsiveness; no presence of broader autism phenotype. BA degree with annual household income of $81-100K. No comorbid diagnoses endorsed. History of past drug use (marijuana, cocaine, amphetamines, ecstasy, hallucinogens, inhalants); currently drinks alcohol. Mother has head circumference of 55.3 cm (z=0.6) and BMI suggestive of being overweight.

Patient ID: 12099.s1
Summary:

Sibling is a non-Hispanic white 13-year-old male. Normative adaptive scores and social responsiveness from parent and teacher noted. No comorbid diagnoses endorsed. Head circumference of 57 cm (z=1.6) and normative BMI.

Family 13552. Proband is the younger of two children born to the same parents with an older brother (13552.s1).
Patient ID: 13552.p1
Event: *DYRK1A de novo* frameshift mutation

Patient is a 71-month-old non-Hispanic white male diagnosed with autism. Mother reports smoking during the first trimester and having an upper respiratory illness during the first, second, and third trimesters. She also reports having Group B strep. Meconium staining at birth. Average VIQ (91), very low NVIQ (66) and adaptive (68) scores. Clinical range deficits in social responsiveness per parent and teacher report. History of speech delay and seizures (febrile and non-febrile). Elevated ADHD behaviors and attention problems with no comorbid diagnoses. History of recurrent otitis media with myringotomy, urinary problems, and strep throat. Tonsils and adenoids removed. Currently incontinent at night. Currently takes antiepileptic medication. Past use of mood stabilizers and antibiotic medication. Allergic to pollen. Head circumference of 48 cm (z =-2.7) and normative BMI. Family history of cancer (maternal grandparent), bipolar disorder (maternal aunt/uncle, 2 cases), congenital heart defect (paternal grandparent)

Patient ID: 13552.fa
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 37. Normative range of social responsiveness; no presence of broader autism phenotype. Graduate degree with annual household income of $131-160K. Has used pain killers, antidepressants, and antibiotics in past. Currently drinks alcohol and reports possible problem with drinking (has tried to cut down; has felt bad about drinking behavior.) No comorbid diagnoses endorsed. Father has head circumference of 58 cm (z=0.7) and BMI suggestive of being overweight.

Patient ID: 13552.mo
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 36. Normative range of social responsiveness; no presence of broader autism phenotype. Graduate degree with annual household income of $131-160K. Only over-the-counter medication use endorsed. No comorbid diagnoses endorsed. Mother has head circumference of 55 cm (z=0.3) and BMI suggestive of being overweight.

Patient ID: 13552.s1
Summary:

Sibling is a non-Hispanic white 92-month-old male. Normative adaptive scores and social responsiveness from parent and teacher noted. No comorbid diagnoses endorsed. Current antibiotic use. Head circumference of 53.5 cm (z=0.7) and normative BMI.

Family 13890. Proband is the younger of two children with an older sister (13890.s1).
Patient ID: 13890.p1
Event: *DYRK1A de novo* splice mutation

Patient is a 164-month-old non-Hispanic white female diagnosed with autism. Mother reports having a skin rash and upper respiratory virus during pregnancy. She also reports Rh incompatibility and use of RhoGAM as well as oligohydramnios during pregnancy. Mother reports drinking during first trimester (unknown amount) and taking over-the-counter medication during the third trimester (Sudafed). Labor was augmented with Pitocin due to failure to progress and an emergent C-section was performed. Proband was irritable and stiff as an infant. Extremely low VIQ (26), NVIQ (42), and adaptive (41) scores. Clinical range deficits in social

responsiveness per parent and teacher report. History of significant speech delay (no phrase speech; single words only) with comorbid diagnosis of Severe Mental Retardation. MRI/CT in normal range; EEG results unclear. History of otitis media, roseola, gastrointestinal problems, and recurrent respiratory problems. History of sleep difficulty and daytime sleepiness. Past use of mood stabilizers, antibiotics, and over-the-counter medications. Head circumference of 51.5 cm (z =-1.6) and BMI suggestive of obesity. Family history of heart disease (both maternal grandparents, paternal grandparent), cancer (maternal grandparent), bipolar disorder with psychiatric hospitalization (maternal grandparent), depression (maternal aunt/uncle, 2 cases), and anxiety disorder (maternal cousin).

Patient ID: 13890.fa
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 37. Some evidence of broader autism phenotype (elevated rigid and aloof behaviors). Associates degree with annual household income of $81-100K. Currently takes medication for high cholesterol, antihypertensives, and pantoprazole (GERD?); past use of antibiotics and asthma medication. No comorbid diagnoses endorsed. Father has head circumference of 57.3 cm (z=0.3) and BMI suggestive of obesity.

Patient ID: 13890.mo
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 35. Normative range of social responsiveness; no presence of broader autism phenotype. BA degree with annual household income of $81-100K. She currently takes medication for high cholesterol. No comorbid diagnoses endorsed. Mother has head circumference of 57 cm (z=1.8) and BMI suggestive of obesity.

Patient ID: 13890.s1
Summary:

Sibling is a non-Hispanic white 16-year old female. Normative adaptive scores and social responsiveness from parent noted. Past use of antibiotics and over-the-counter medications. Has migraines. Head circumference of 53.3 cm (z=-0.8) and BMI suggestive of obesity.

Family 11390. Proband is the first of two children with a younger sister (11390.s1).
Patient ID: 11390.p1
Event: *PTEN de novo* missense mutation

Patient is a 99-month-old non-Hispanic white female diagnosed with autism. Mother reports anemia, migraines, edema, hypertension, preeclampsia, and an upper respiratory infection during pregnancy. Proband was induced with Pitocin due to failure to progress and born via C-section (cervix too small for head). Very low VIQ (57), low NVIQ (77) and low adaptive (79) scores. Clinical range deficits in social responsiveness per parent report. History of speech delay with loss of language skills during development (words). Elevated attention problems, ADHD symptoms, affective problems, and social withdrawal behaviors with comorbid diagnosis of Pica. History of otitis media with myringotomy, strep throat, and gastrointestinal problems. Currently has difficulty going to bed and exhibits symptoms of sleep disordered breathing. Head

circumference of 56 cm (z =2.8) and normative BMI. Family history of migraines (maternal aunt/uncle, maternal cousin), speech delay (paternal half sibling), articulation difficulty (paternal half sibling), heart disease (maternal grandparent, both paternal grandparents), stroke and death under 50 years old (maternal grandparent), cancer (maternal aunt/uncle, 2 cases), depression (maternal cousin, maternal grandparent, maternal aunt/uncle), personality disorder (maternal cousin), psoriasis (maternal aunt/uncle), and diabetes type 2 (paternal grandfather).

Patient ID: 11390.fa
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 34. No evidence of broader autism phenotype. Associates degree with annual household income of $51-65K. No comorbid diagnoses endorsed. Father has head circumference of 58.5 cm (z=0.9); BMI unavailable.

Patient ID: 11390.mo
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 32. Some evidence of broader autism phenotype (elevated pragmatic language impairment and rigid behaviors). Associates degree with annual household income of $51-65K. Currently taking antidepressants but no comorbid diagnoses endorsed. Mother has head circumference of 54 cm (z=-0.4); BMI unavailable.

Patient ID: 11390.s1
Summary:

Sibling is a non-Hispanic white 81-month-old female. Normative social responsiveness from parent noted. Behavioral elevations for attention problems, ADHD symptoms, and somatic problems. No comorbid diagnoses endorsed. Head circumference of 53 cm (z=1.4); BMI unavailable.

Family 14433. Proband is the second of two children with an older brother (14433.s1).
Patient ID: 14433.p1
Event: *PTEN de novo* missense mutation

Patient is a 49-month-old non-Hispanic white male diagnosed with autism. Mother had a flu shot during pregnancy and tested positive for Rh incompatibility (RhoGAM taken). She also experienced edema, dizzy spells, hyperemesis, and blurry vision. Mother reports drinking during first trimester (unknown amount) and taking asthma medication during second trimester. Proband had hyperbilirubinemia after birth but no treatment required. Mother reports feeding difficulty after birth. Very low VIQ (55) and NVIQ (50) with low adaptive (73) scores. Clinical range deficits in social responsiveness per parent report. Elevated affective problems and social withdrawal behaviors with comorbid diagnosis of Mild Mental Retardation. History of speech delay and possible non-febrile seizures. Mother suspects movement abnormalities but no diagnosis. EEG normal. Proband has sleep difficulties (irregular bedtime, frequent awakenings, difficulty waking up in the morning, and daytime sleepiness). Currently takes medication for asthma. Head circumference of 57.8 cm (z =4.7) and BMI suggestive of obesity. Family history of migraines (maternal aunt/uncle, maternal grandparent), heart disease (both maternal

grandparents), cancer (maternal grandparent), reading disorder (maternal aunt/uncle), ADHD (maternal cousin), birth defect (maternal grandparent has webbed toes and fingers)

Patient ID: 14433.fa
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 38. Normative range of social responsiveness; no presence of broader autism phenotype. Graduate degree with annual household income over $161K. No comorbid diagnoses endorsed. Father has head circumference of 63 cm (z=3.5) and BMI suggestive of being overweight.

Patient ID: 14433.mo
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 32. Normative range of social responsiveness; no presence of broader autism phenotype. Graduate degree with annual household income over $161K. No comorbid diagnoses endorsed. Currently taking thyroid medication and birth control. Mother has head circumference of 58.1 cm (z=2.7) and normative BMI.

Patient ID: 14433.s1
Summary:

Sibling is a non-Hispanic, white 67-month-old male. Normative social responsiveness with slightly below average adaptive scores per parent report. Mild hearing loss. No comorbid diagnoses endorsed. Head circumference of 53.9 cm (z=1.2) and BMI suggestive of being overweight.

Family 14611. Proband is the youngest of three children born to the same parents with an older sister (14611.s1) and brother.
Patient ID: 14611.p1
Event: *PTEN de novo* frameshift mutation

Patient is a 111-month-old non-Hispanic white male diagnosed with autism. Mother experienced migraines, a vaginal infection, edema, hyperemesis, and an upper respiratory illness during pregnancy. Proband was born vaginally with the use of forceps. He received phototherapy for hyperbilirubinemia after birth. Mother reports he had feeding difficulties, was irritable/inconsolable, and reacted poorly to MMR vaccination (lethargic and sick). Extremely low VIQ (19), NVIQ (33), and adaptive (57) scores. Clinical range deficits in social responsiveness per parent and teacher report. History of speech delay and loss of words during development; currently nonverbal. Comorbid diagnoses of Severe Mental Retardation, ADHD, OCD, and Anxiety Disorder NOS. History of otitis media. Currently has sleep difficulties (difficulty going to sleep; mother must lay down with him.) Head circumference of 56 cm (z =2.0) and normative BMI. Family history of stroke (paternal grandparent) and dyslexia (paternal cousin).

Patient ID: 14611.fa
Summary:

Father is an adult non-Hispanic white male. Age at conception of proband is 39. Some evidence of broader autism phenotype (elevated rigid and aloof behaviors). BA degree with annual household income of $81-100K. Past use of pain killers, antihypertensives, and sedatives. Currently takes thyroid medication, myfortic, and sensipar. No comorbid diagnoses endorsed. Father has head circumference of 59 cm (z=1.3) and normative BMI.

Patient ID: 14611.mo
Summary:

Mother is an adult non-Hispanic white female. Age at conception of proband is 39. Normative range of social responsiveness; no presence of broader autism phenotype. BA degree with annual household income of $81-100K. Takes medication for asthma (not taken during pregnancy). No comorbid diagnoses endorsed. Mother has head circumference of 55 cm (z=0.3) and normative BMI.

Patient ID: 14611.s1
Summary:

Sibling is a non-Hispanic white 177-month-old female. Normative adaptive scores and social responsiveness from parent and teacher noted. History of asthma. No comorbid diagnoses endorsed. Head circumference of 54 cm (z=-0.1) and normative BMI.

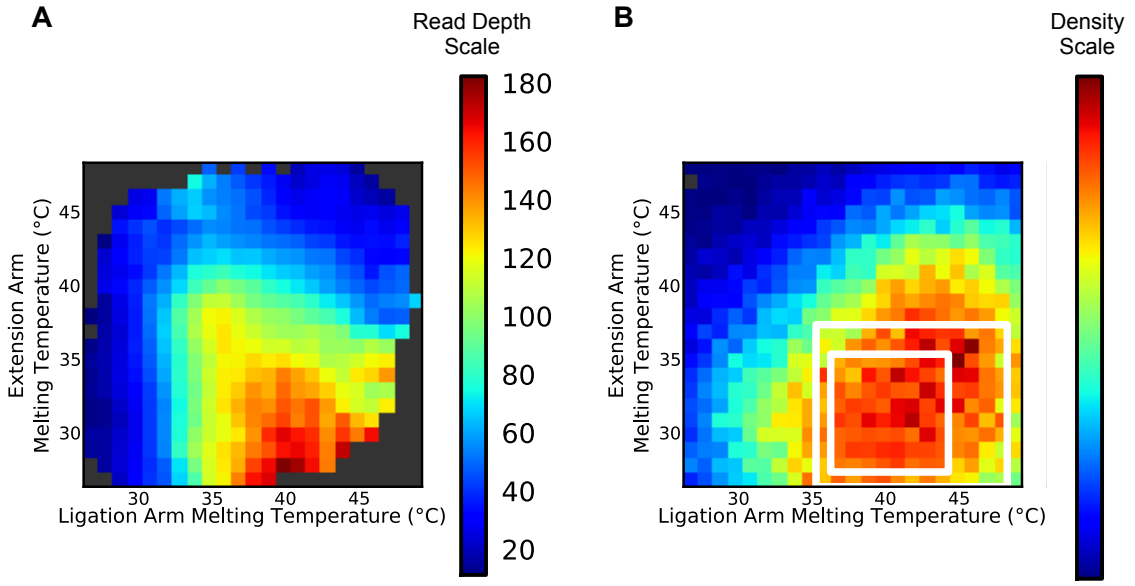

Figure S1. MIP design optimization from 55k MIP test set.
(**A**) MIP capture efficiency depends on ligation arm (x-axis) and extension arm (y-axis) melting temperatures. The average performance (i.e., read depth on heat map scale) is plotted for various pairings of ligation arm and extension arm melting temperatures (*24*) for successful captures in a 55,000 MIP capture reaction (*7*). The read depth signal is smoothed by averaging the read depth of each bin with the read depth from each of its neighboring bins. These data indicate that MIPs perform best when the melting temperature of the ligation arm is greater than that of the extension arm. (**B**) The "selection space" of potential MIP designs is visualized by plotting the frequency distribution of ligation arm (x-axis) and extension arm (y-axis) melting temperatures for every possible MIP that can be designed against the human exome under certain constraints. Specifically, the design constraints mandate that each potential MIP targeted a region of fixed length (112 bp) but allow extension/ligation arm lengths to vary between 16/24, 17/23, 18/22, 19/21, and 20/20 bp. Designing MIPs with a ligation arm length greater than or equal to the extension arm length creates more MIPs with a higher ligation arm melting temperature, which tend to perform better (see **A**). The heat map scale corresponds to the number of potential MIP designs with a given pair of ligation arm and extension arm melting temperatures. Under these design constraints, a substantial fraction of potential MIPs are predicted high performers (inner white box) or medium performers (outer white box). Note that for this plot, the distribution was approximated by randomly sampling ~100,000 MIPs from the full distribution.

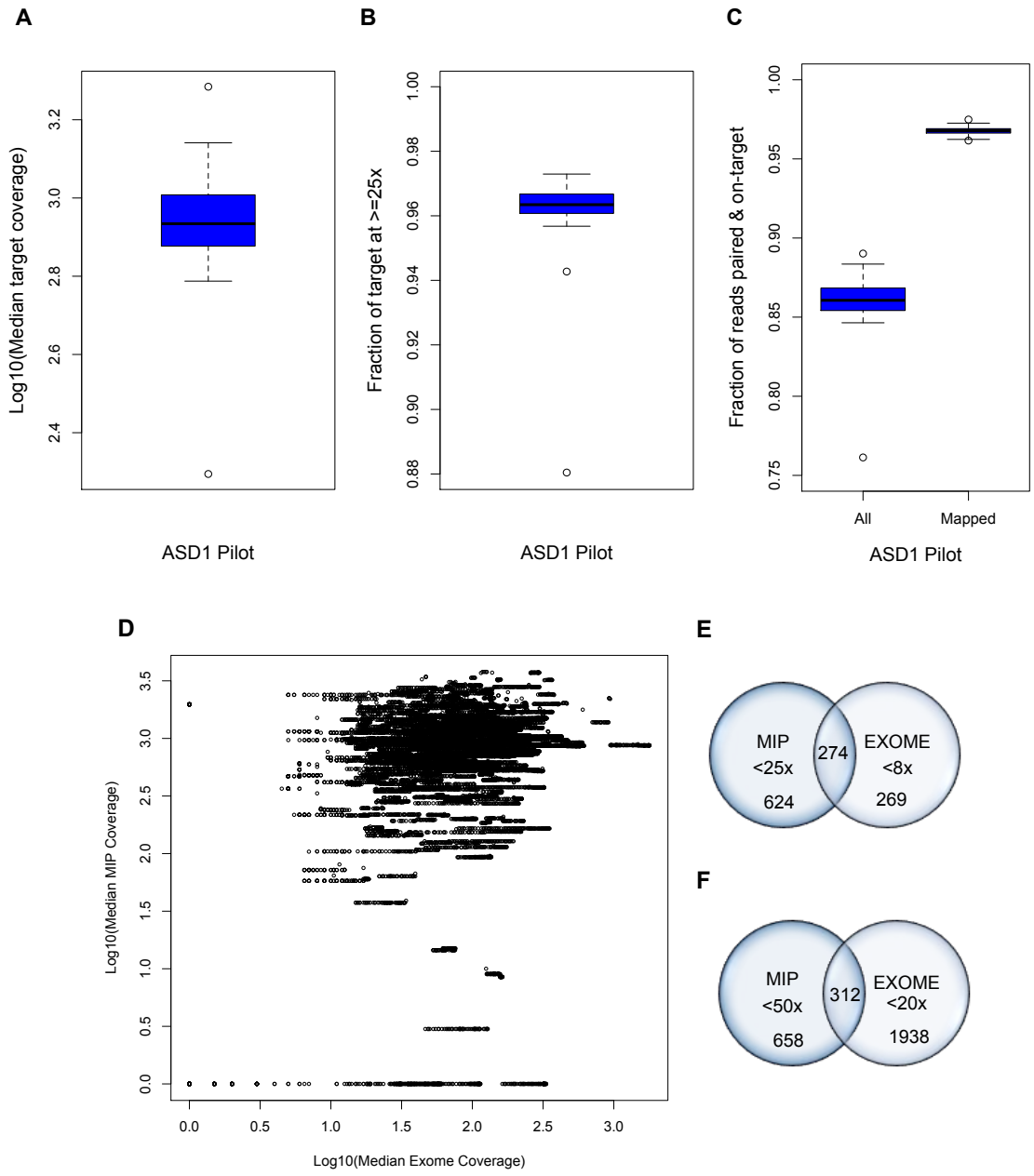

Figure S2. Performance of ASD1 MIPs on 48 samples (16 trios) that were previously exome sequenced.

(**A-C**) Box and whisker plots. (**A**) Log10 median target coverage. (**B**) Fraction of targeted bases at 25-fold or greater coverage. (**C**) Fraction of read pairs that were correctly paired and matching expected MIP capture locations. "All" is the fraction of all raw reads from the sequencer with assignable barcode sequences, correctly paired reads that map to targeted locations. "Mapped" is the fraction of the mapped, correctly paired reads that map to targeted locations. (**D**) Scatter plot of the log10 median coverage of 48 samples across all 24,566 target positions for MIP capture versus exome capture. Spearman correlation = 0.154. (**E-F**) Overlap of poorly captured positions by MIP capture and exome capture at various minimum coverage thresholds.

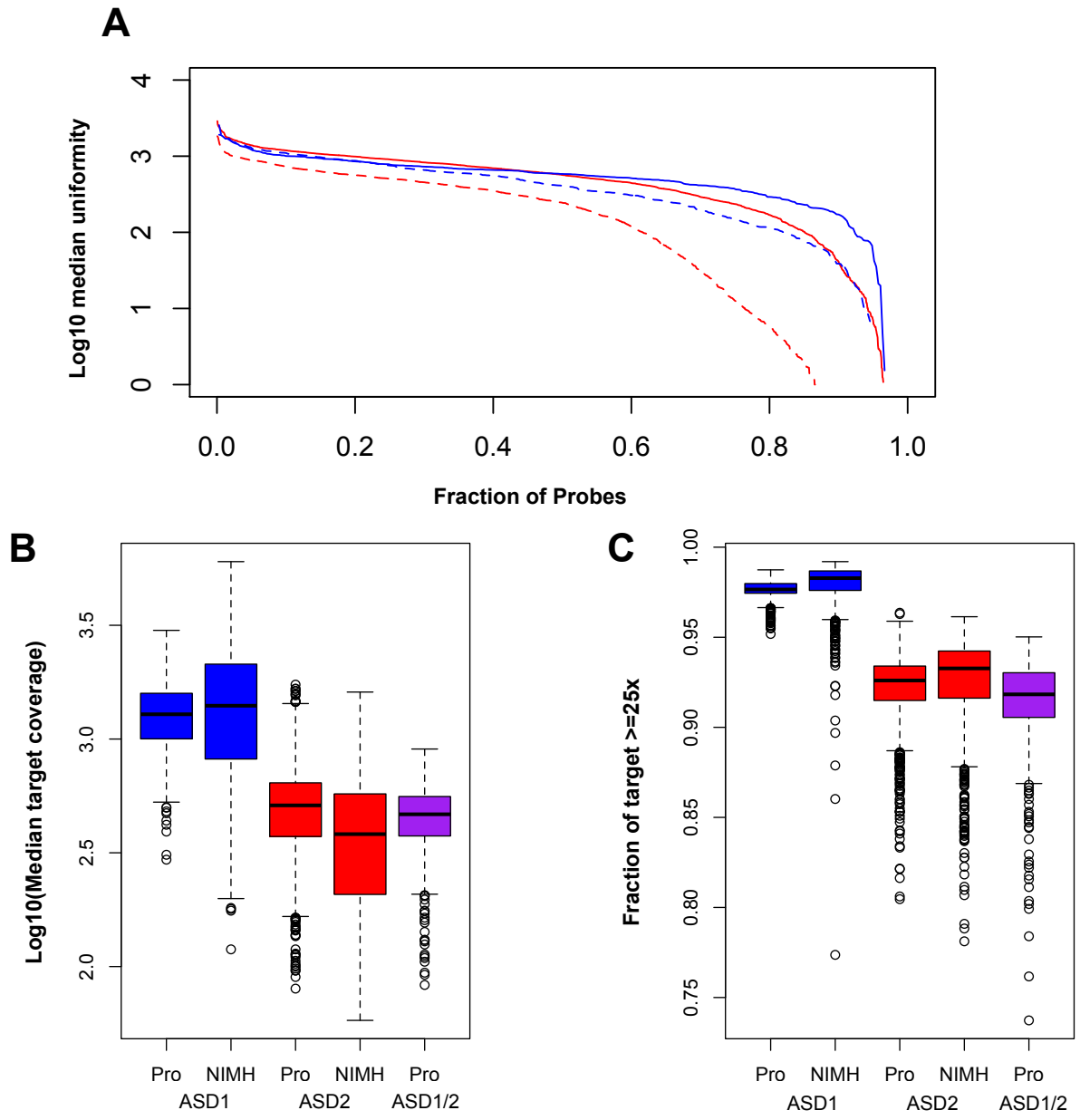

Figure S3. Performance of ASD1 and ASD2 probe sets.
(**A**) Uniformity plots for pre-normalization (dashed) and post-normalization (solid) MIP pools (ASD1 = blue, ASD2 = red). (**B-C**) Box and whisker plots of MIP-captured samples. (**B**) Median coverage across the coding target. (**C**) The fraction of coding target bases sequenced to 25-fold coverage or greater. Abbreviations: Pro-ASD probands, NIMH-non-ASD samples from the National Institute of Mental Health.

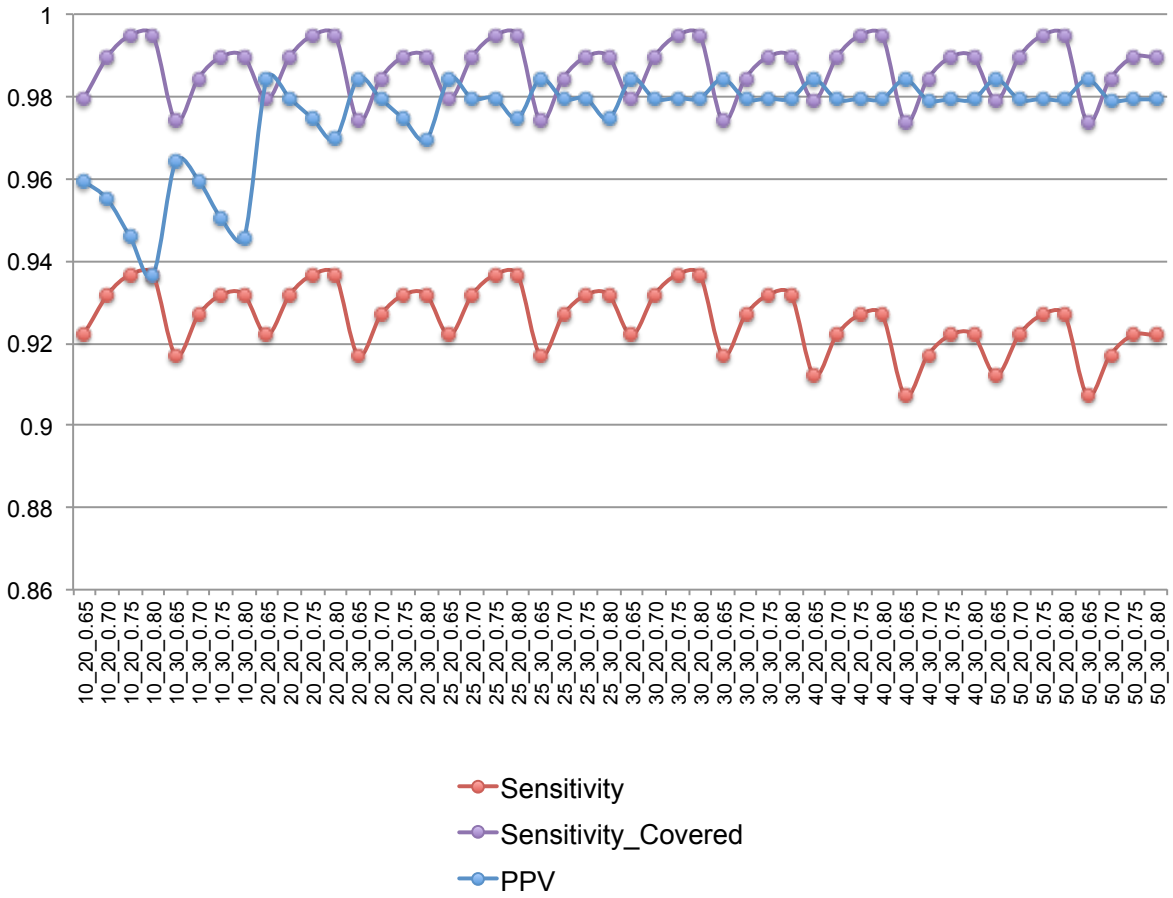

Figure S4. Sensitivity and PPV plots for SNV calling heuristics.
The estimated sensitivity for SNVs at all (red) or sufficiently covered (purple) targeted bases, as well as the estimated PPV (blue), is shown for various filtering thresholds. The filtering thresholds (x-axis) are expressed as X_Y_Z, where X is minimum coverage, Y is minimum consensus or SNP quality, and Z is allele balance cutoff (maximum fraction of Q20 reference bases).

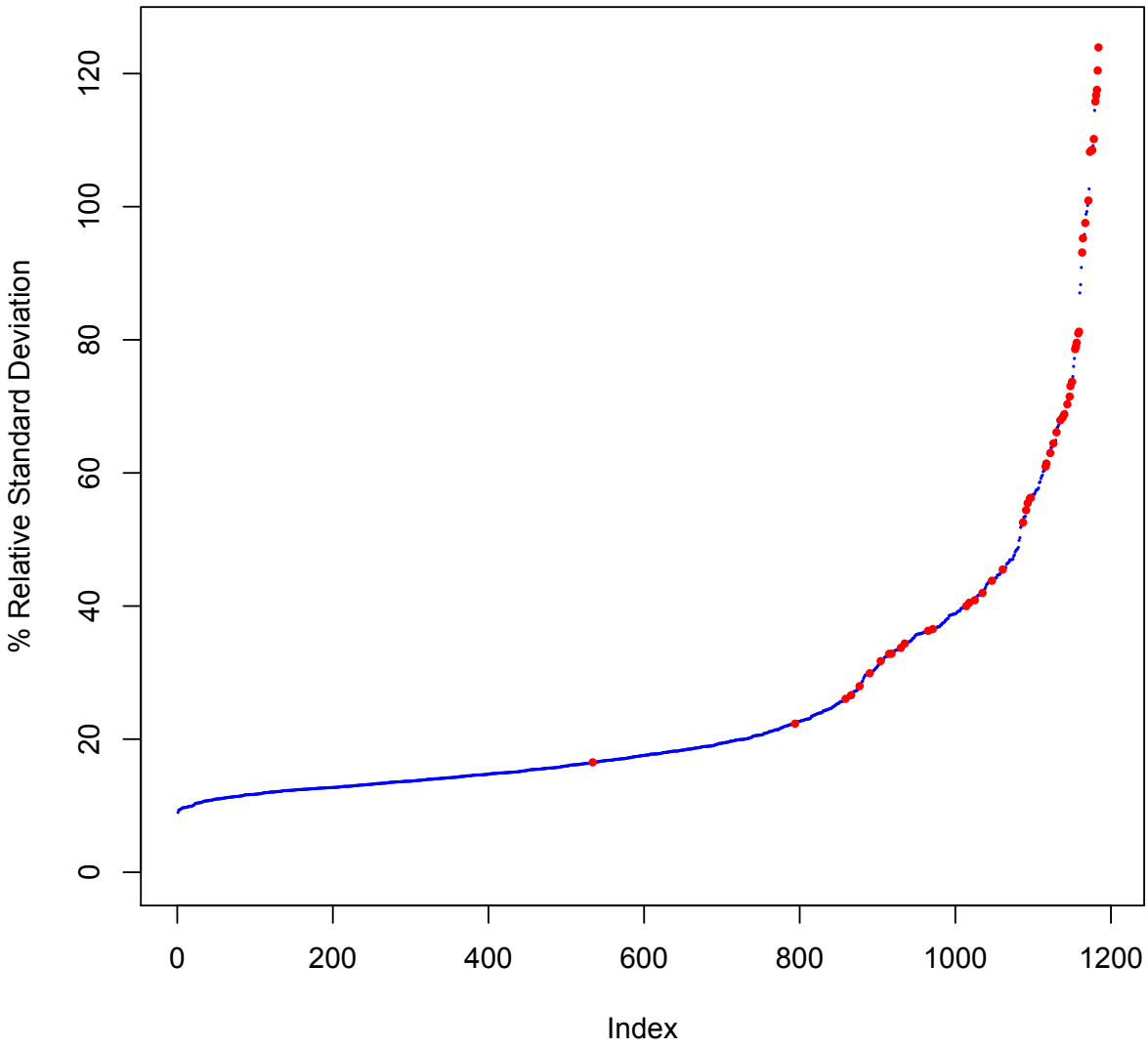

Figure S5. Initial probe performance strongly predicts utility of probe read depth for CNV calling.
The ranked-ordered percent relative standard deviations of each MIP's (EP set) relative abundance (# probe mapped reads/# total mapped reads) are shown. Blue points are well-performing probes that were not rebalanced (n = 1,130). Red points are poorly performing probes that were rebalanced at 10-fold excess (n = 65). The poorly performing probes exhibit highly variable capture efficiency relative to well-performing probes and are not useful for copy number analysis. Probes with effectively no coverage, median reads <20, or overlapping SNV sites are not shown (n = 130).

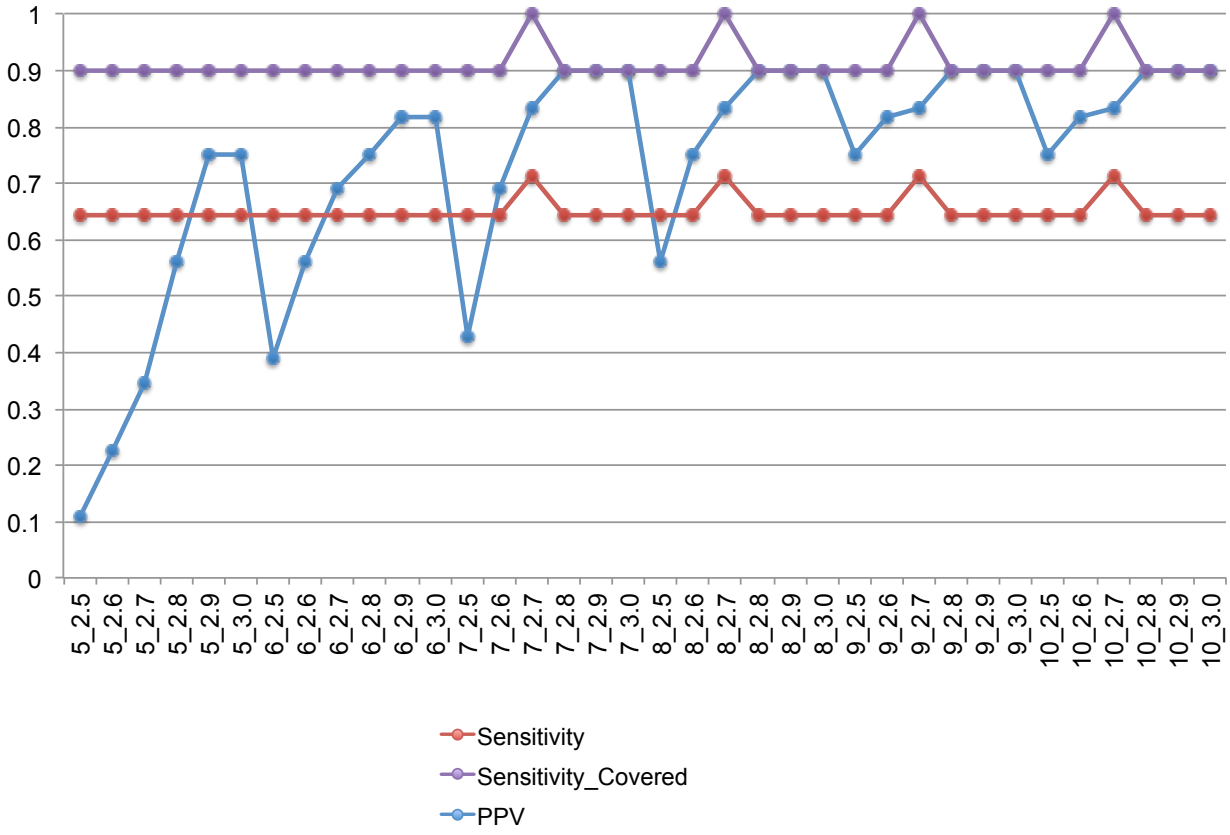

Figure S6. PPV and sensitivity curves generated for CNV events from 357 epilepsy samples. After removing samples with very high call rates (>10 standard deviations from the mean number of CNV calls), 357 samples were left for analysis (including 14 duplicates) and only autosomal probes were considered. We required true positive sites to be called in both replicates and counted these as only a single observation. At all sites the number of expected events was 14. At covered sites the number of expected events is 10. The estimated sensitivity for CNVs at all (red) or sufficiently covered (purple) regions, as well as the estimated PPV (blue), is shown for various filtering thresholds. The filtering thresholds (x-axis) are expressed as A_B, where A is minimum number of probes and B is average SVD adjusted Z-score of the calling window.

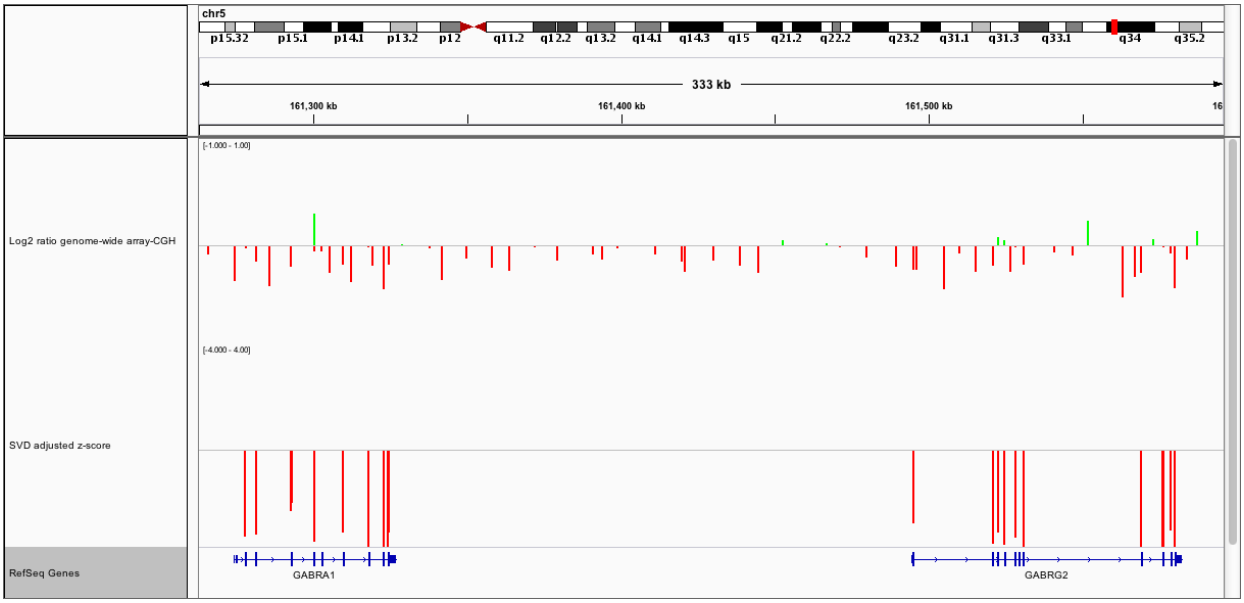

Figure S7. Example of a known CNV called using MIP read depth.
Integrative Genome Viewer screenshot of *GABRA1/2* heterozygous deletion region. Top track shows log2 array CGH signal. The middle track shows adjusted Z-score values for individual MIP probes (red are less than -2.5). The bottom track shows RefSeq gene models.

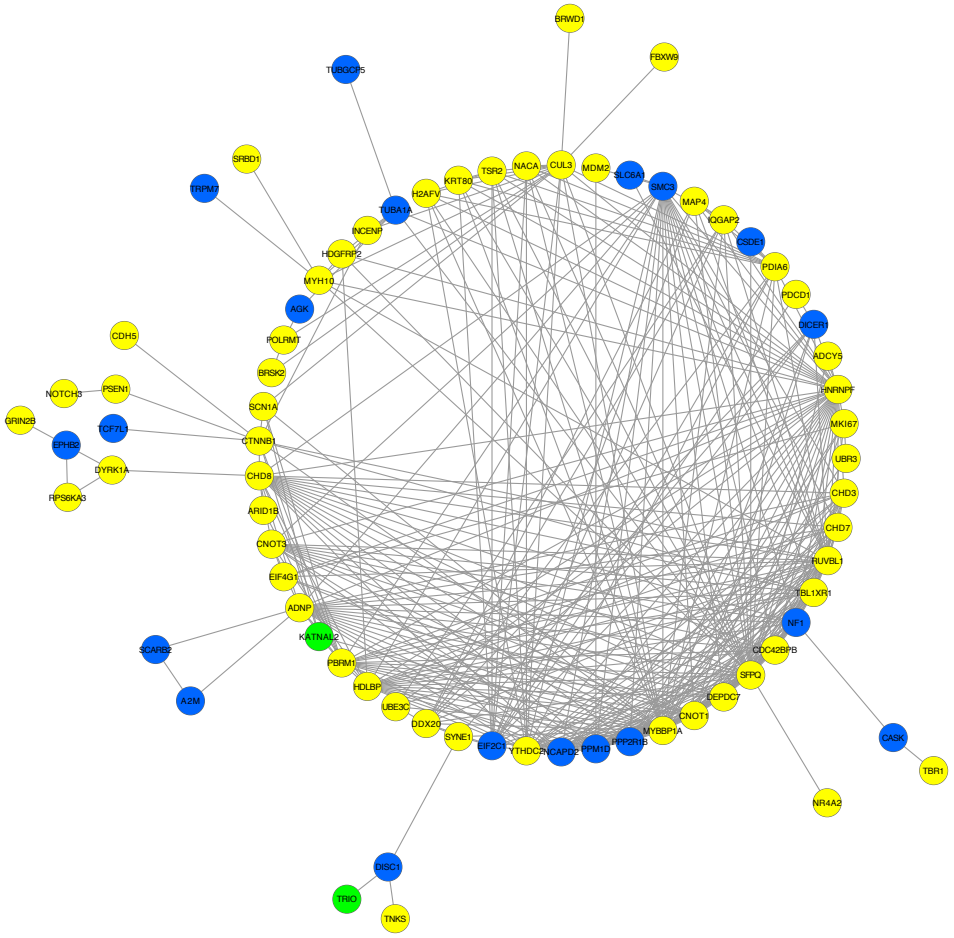

Figure S8. Joint PPI network analysis of severe mutation events identified in O'Roak et al. and Sanders et al. identifies 74-member component.
Gene products are nodes, while gray lines are direct PPI. Yellow nodes (n = 53): O'Roak et al. (*2, 3*); blue nodes (n = 19): Sanders et al. (*4*); green nodes (n = 2): both studies.

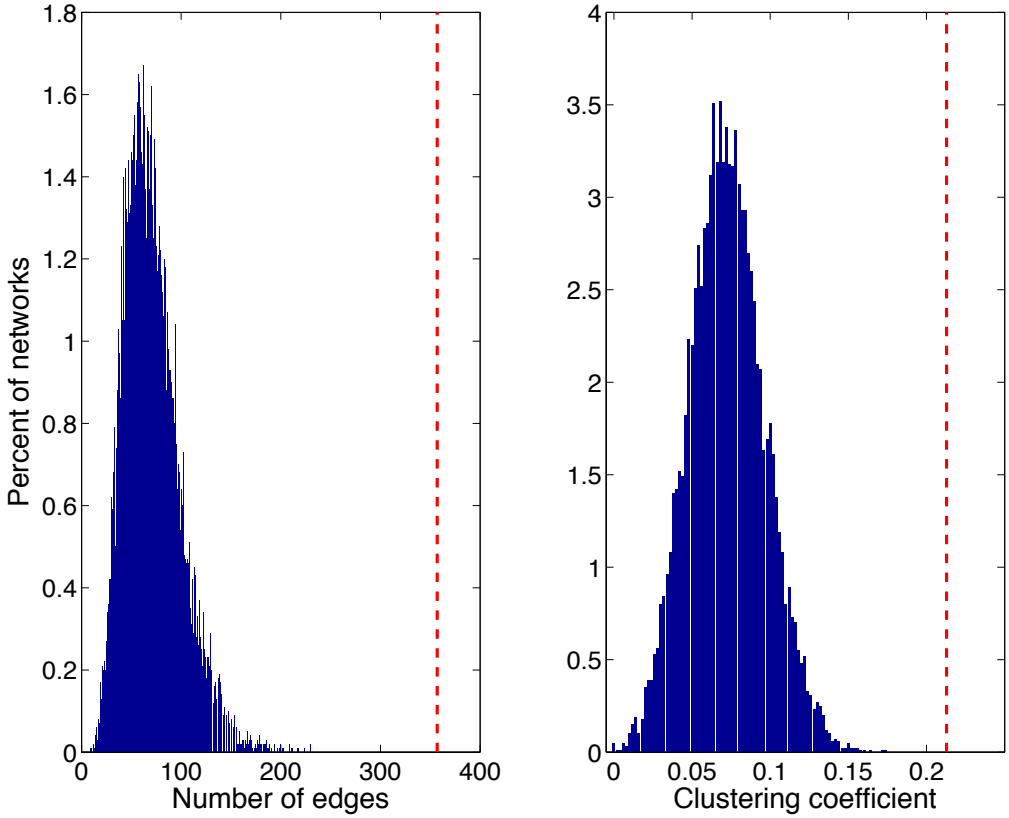

Figure S9. Histograms of network statistics for 10,000 simulated null networks.
Left: Number of edges in the networks. Right: Global clustering coefficient of the networks. In both figures, the red dotted line indicates the corresponding value in the experimentally determined network.

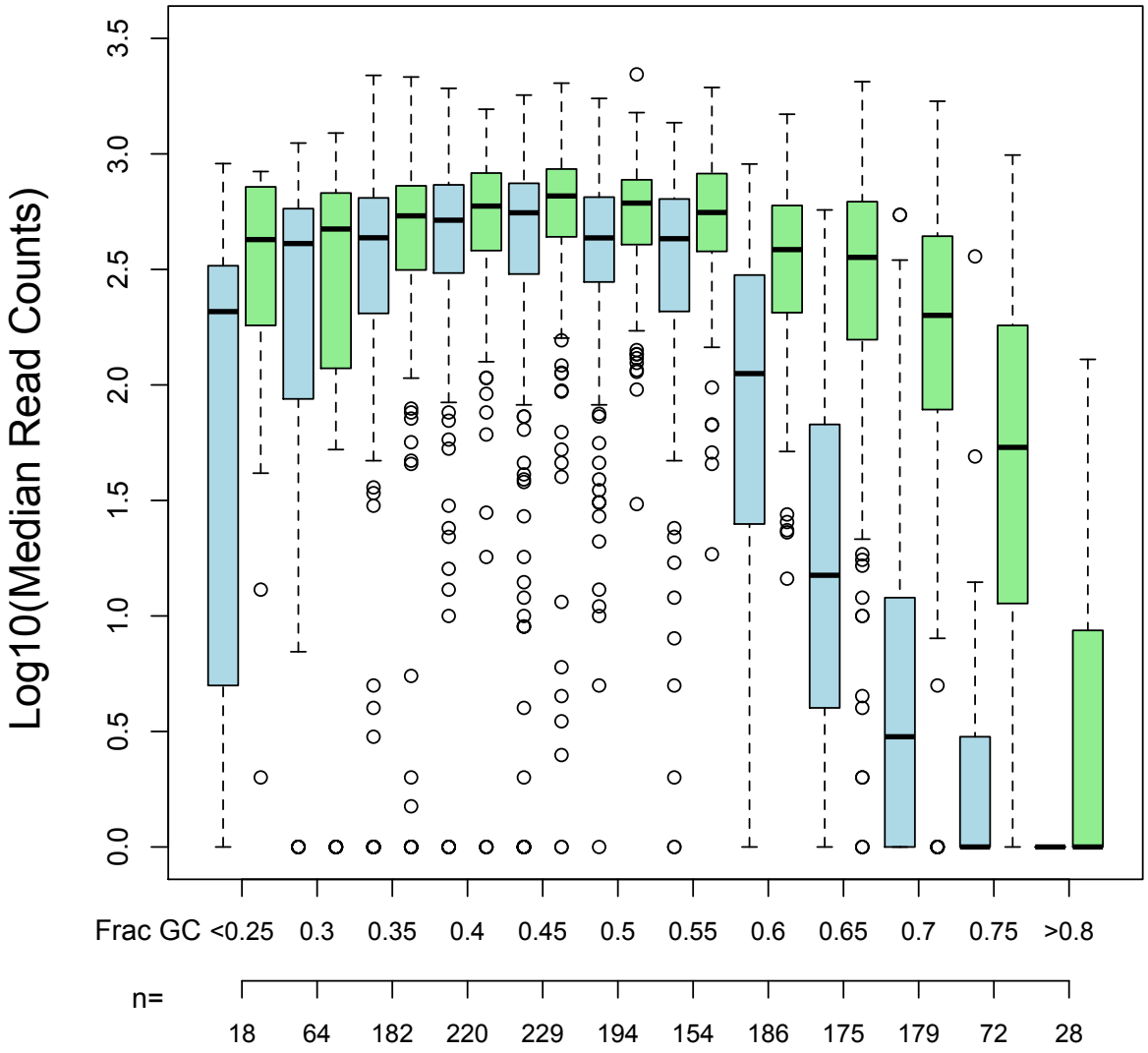

Figure S10. Performance of ASD2 MIPs as a function of the GC content of the targeted gap-fill region.

Box and whisker plots of the log10 median read counts for MIPs targeting gap-fill regions with various GC content bins are shown. Capture efficiency exhibits reduced and more variable performance at the GC extremes. Normalizing probes based on initial performance dramatically improves capture efficacy. Median calculated from seven successful captures and all sites with zero reads were set to one to allow for plotting on a log10 scale. Unnormalized probes are blue. Normalized probes are green. First x-axis shows the GC content bins. Second x-axis shows the number of probes in each bin. Note: Probes overlapping common SNPs for which two probes exist (one for each SNP allele) are collapsed (13 of 1,714 total probes).

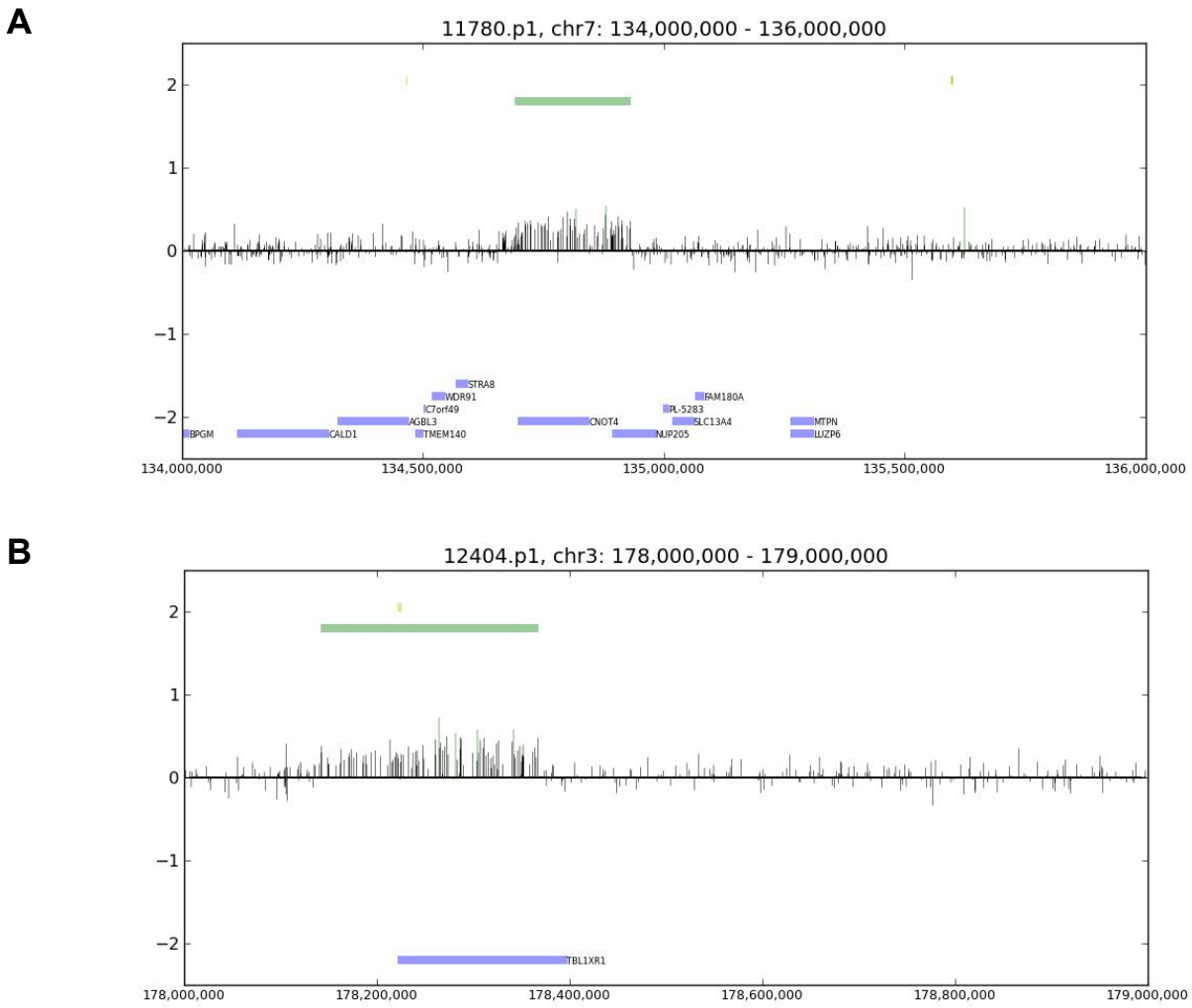

Figure S11. Proband MIP read depth CNV calls with Illumina 1M genotyping data. X-axis is hg18-defined chromosomal position. Y-axis is the log (base2) R ratio values plotted from Illumina 1M genotyping data (*47*) for individuals with a MIP-based copy number call. (**A**) Proband 11780 has a maternally inherited duplication overlapping the *CNOT4* locus supported by the array data, indicated by green horizontal bar. (**B**) Proband 12404 has a paternally inherited duplication overlapping the *TBL1XR1* locus supported by the array data, indicated by green horizontal bar. Green vertical lines >=+0.5, Red vertical lines <=-0.5. Blue bars show gene transcripts.

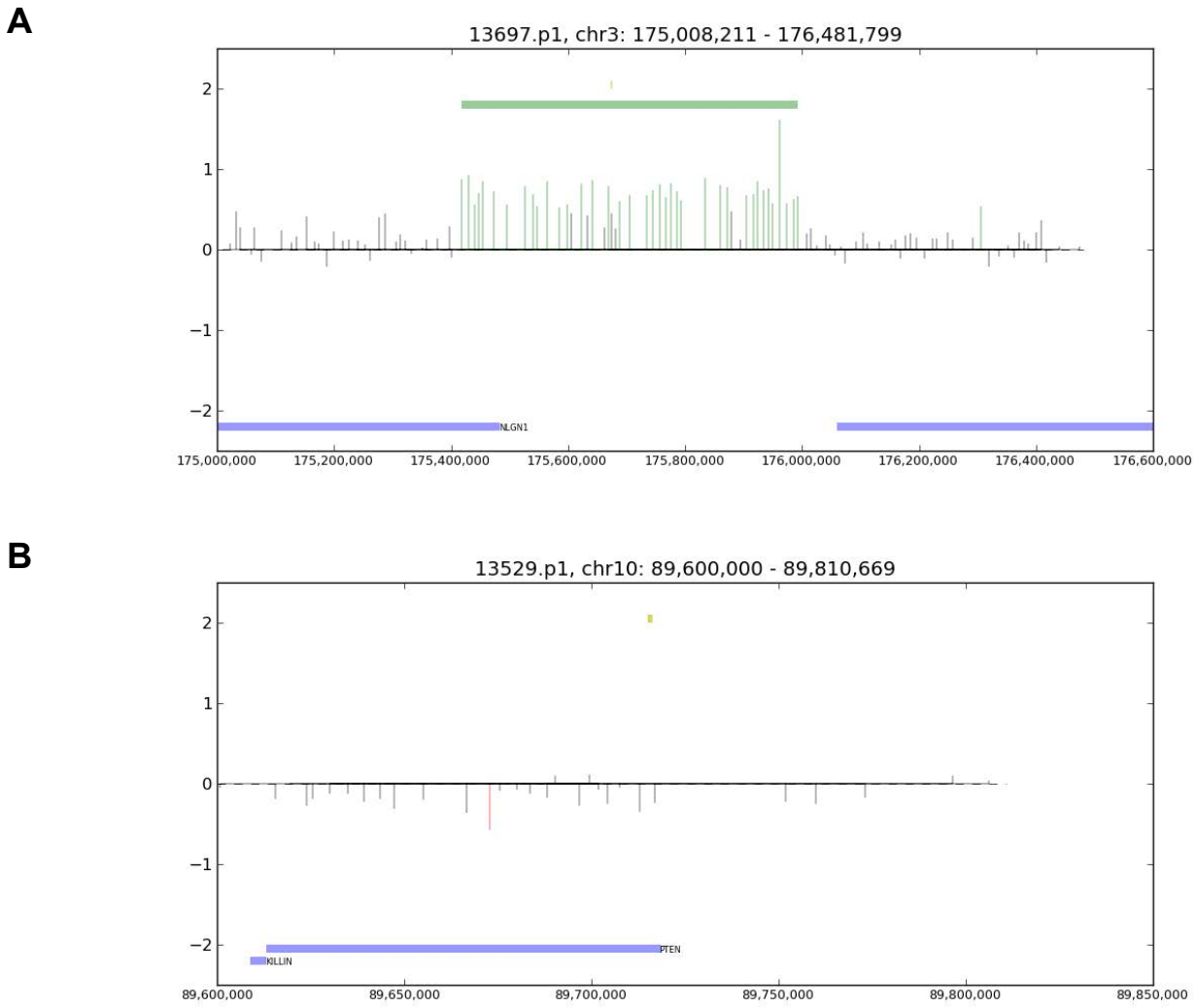

Figure S12. Proband MIP read depth CNV calls with array CGH data.
X-axis is hg18-defined chromosomal position. Y-axis is the log (base 2) relative intensity values plotted from Agilent custom array CGH (*3*). (**A**) Proband 13697 has a maternally inherited duplication overlapping the *NLGN1* locus supported by the array data, indicated by green horizontal bar. (**B**) Proband 13529 has a predicted *PTEN* duplication not locus supported by the array data. Green vertical lines >=+0.5, Red vertical lines <=-0.5. Light blue bars show gene transcripts.

A

Assay	Chrom	Pos(hg18)	Pos(hg19)	Coverage	Allele Balance	Sample
MIP	chr2	166556252	166848006	1965	0.53	12340.p1
Exome	chr2	166556252	166848006	17	0.53	12340.p1

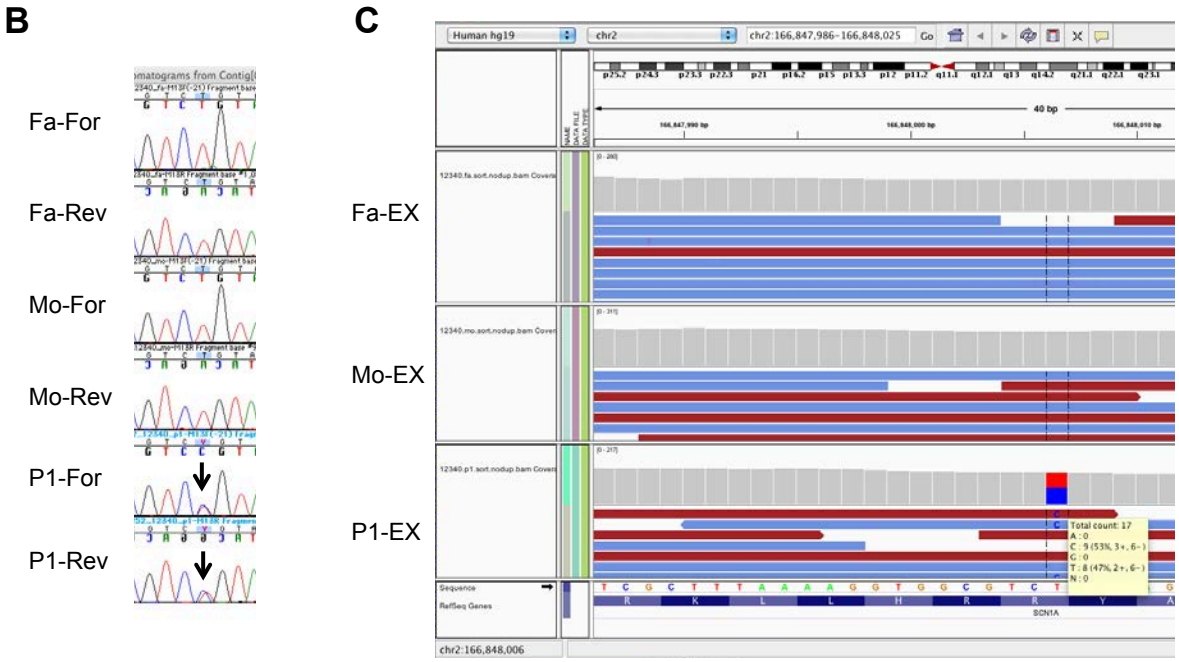

Figure S13. Example of *de novo* variant not reported in exome sequencing studies. (**A**) Summary of coverage and allele balance stats for MIP and exome sequence (*4*) of 12340.p1 at site of *SCN1A de novo* T->C substitution. (**B**) Sanger traces for family 12340. (**C**) Integrative Genomics Viewer screenshot of mapped exome sequence. In this case, the variant failed to reach minimum coverage criterion in exome study. Yellow box shows fraction of reads supporting variant and their strand. For-forward, Rev-reverse, Fa-father, Mo-mother, P1-proband, EX-exome.

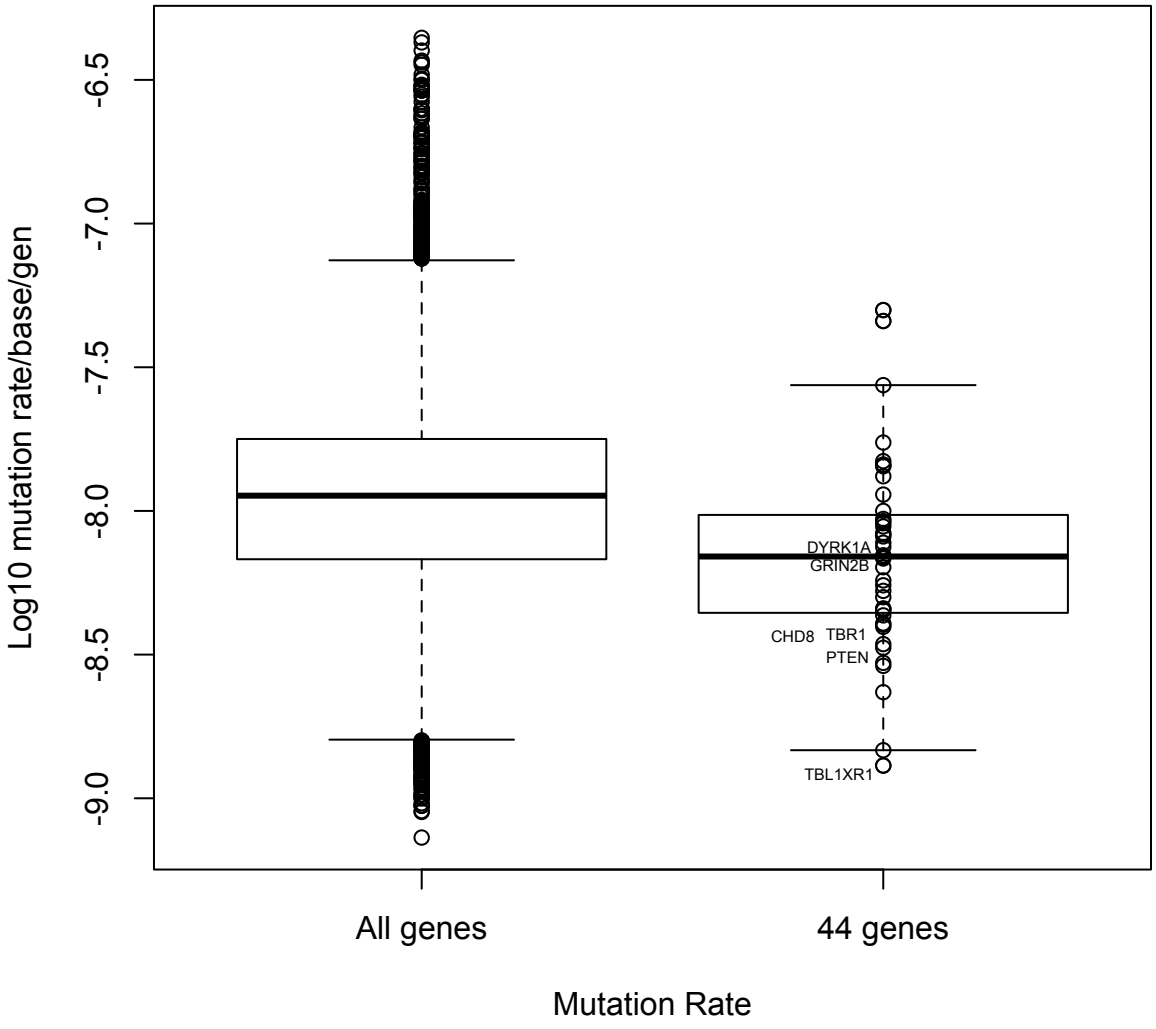

Figure S14. Distribution of locus-specific mutation rates based on human-chimpanzee comparisons.
Box and whisker plots of the log10 mutation rates (per base-pair per generation) shown for all genes (left) and the 44 MIP target genes (right). Locations of the six implicated genes are noted (see table S13).

Table S1. Reagent cost estimates for MIP capture and sequencing.

Amortizing costs	Total pool cost	Per sample (assuming 3k reactions)	Per sample per gene
44 genes (2,069 oligos, and oligo preparation (pooling and phosphorylation))	$14,397.83	$4.80	$0.11
Fixed costs	**Per 192 samples**	**Per sample**	**Per sample per gene**
Reagents and plasticware to capture 192 samples	$493.83	$2.57	$0.06
Reagents for 1 lane HiSeq sequencing (PE101)	$1,388.00	$7.23	$0.16
Capture and sequencing cost	$1,881.83	$9.80	$0.22
Total cost summary*	**$16,279.66**	**$14.60**	**$0.33**

*The cost estimates shown are specifically for reagents and do not include other costs, e.g., technician labor, amortized machine purchase costs, machine service contracts, machine purchases, computation, etc.

Table S2. Parameters for MIP pools.

	ASD1	ASD2	ASD1+ASD2	Epilepsy (EP)	Joubert syndrome (JS)
Genes	6	38	44	32	26
Total target bp	24,566 bp	120,702 bp	145,268 bp	86,494 bp	86,462 bp
MIPs	355	1714	2069	1325	1423
DNA amount	50 ng	100 ng	120ng	100 ng	100 ng
MIP: target ratio	200:1	800:1	800:1	200:1	200:1
dNTPs	0.16 µM	0.32 µM	0.32 µM	0.32 µM	0.32 µM
Incubation time	48 hours	23 hours	23 hours	48 hours	48 hours
PCR volume	50 µl	25 µl	25 µl	25 µl	25 µl
~#samples/HiSeq2000 lane	384	192	192	192	192
MIPs rebalance	By individual probe	50X increase of poor performers	50X increase of poor performers	10X increase of poor performers	10X increase of poor performers

Table S3. ASD candidate loci targeted by MIPs.

Gene (RefSeq)	Pool	# MIPs	Initial Mut	Nomination Category	OMIM	Notes	CNV Reports/ Implicated Regions	References
FOXP1	asd1	40	fs	ID, ASD, CNV, OMIM	613670	CNV and point mutations: ID, ASD, and language impairments	Y	(*48, 49*)
FOXP2	asd1	38	none	Gene similarity (FOXP1), CNV, OMIM	602081	CNV and point mutations: Speech – language disorders and ASD	Y	(*50, 51*)
GRIN2A	asd1	57	none	Gene similarity (GRIN2B), ID, EP, CNV, OMIM	623971	CNV and point mutations: ID, EP, neurodevelopmental defects; point mutations: melanoma	Y	(*40, 52*)
GRIN2B	asd1	54	sp	ID, EP, ASD, SCZ, CNV, OMIM	613970	CNV and point mutations: ID, EP, neurodevelopmental defects; rare missense variants: ASD and SCZ; GWAS: SCZ	Y	(*21, 40, 41, 53-55*)
LAMC3	asd1	77	ms	Novel, OMIM	614115	Point mutations: cortical malformations (recessive)	N	(*56*)
SCN1A	asd1	89	ms	Syndromic, OMIM	607208	CNV and point mutations; Dravet syndrome, EP, ASD	Y	(*57*)
Total	**asd1**	**355**						
ADCY5	asd2	58	ms	Novel, animal model brain development		Mouse KO: associated with lower birth weight/glucose levels, highly concentrated in the dorsal striatum and nucleus accumbens	N	(*58*)
ADNP	asd2	35	fs	Novel, animal model ASD-like		Mouse KO & antagonist: neuronal/glial pathology and reduced cognitive functions and ASD-like behavior	N	(*59, 60*)
AP3B2	asd2	56	ms	CNV, novel		CNV: congenital diaphragmatic hernia, cognitive deficits; brain specific, neurotransmitter release	15q25.2:del	(*61*)
ARID1B	asd2	81	fs	CNV, OMIM	614562	CNV and point mutations: ID, ASD, Coffin-Siris	Y	(*62-65*)
BRSK2	asd2	43	aa	CNV, novel		Mouse KO: neuronal polarization in the cerebral cortex	11p15.5:del	(*14, 66*)
CHD7	asd2	117	ms	Syndromic, OMIM	214800	CHARGE syndrome w/ ASD features; Interacts with CHD8	N	(*13, 67, 68*)
CHD8	asd2	105	fs/ns	CNV, novel		Interacts with CHD7; Case report: large 14q11.2 deletions and DD, ID, and dysmorphic features; Case report of balanced translation ASD, and dysmorphic features.	Y	(*21, 38, 67*)
CNOT4	asd2	33	ms	Novel, ubiquitin		E3 ubiquitin ligase	N	
CTNNB1	asd2	35	ms	Novel		Beta-catenin	N	

Gene (RefSeq)	Pool	# MIPs	Initial Mut	Nomination Category	OMIM	Notes	CNV Reports/ Implicated Regions	References
CUL3	asd2	37	ns	Novel, ubiquitin	614496	E3 ubiquitin ligase; Pseudohypoaldosteronism, type IIE (renal disease);	N	
CUL5	asd2	41	ms	Novel, ubiquitin		In vitro: cortical development, ubiquitin pathway	N	(*69*)
DYRK1A	asd2	38	sp	Syndromic, animal model brain development, CNV, OMIM	614014	CNV and translocations: ID, microcephaly, and dysmorphic features; Mouse KO: developmental delay and abnormal brains; positive selection in human linage	Y	(*18-20, 42, 70*)
HDGFRP2	asd2	34	ms	CNV, novel		Rat expression: throughout brain	19p13.3:del/ dup	(*71*)
HDLBP	asd2	62	ms	CNV, novel		Shown to be reduced in patient with ASD 2q37 deletion syndrome	2q37:del	(*72*)
MBD5	asd2	54	fs	CNV, OMIM	156200	The only gene deleted in all subjects with the 2q23.1 microdeletion syndrome	2q23.1:del	(*73, 74*)
MDM2	asd2	30	ms	Novel		E3 ubiquitin ligase; cancer related	N	
NLGN1	asd2	28	ms	CNV, known		CNV mutations: ID, ASD, EP,	Y	(*75, 76*)
NOTCH3	asd2	95	ms	CNV, novel	125310	CNV mutation: ID, obesity, and hypertrichosis; CADASIL syndrome	Y	(*77*)
NR4A2	asd2	22	ms	Novel		Point mutations: SCZ and Parkinsons; Involved in survival of dopamine neurons	N	(*78, 79*)
NTNG1	asd2	25	ms	Syndromic		Balanced translocation: Rett syndrome	N	(*80*)
OPRL1	asd2	14	ms	CNV, novel		Association with opioid addiction and alcohol dependence	20q13.33:del	(*81*)
PDCD1	asd2	16	fs	CNV, novel	126200	In Williams et al.'s mapping of the critical region for BDMR, the only child without HDAC4 deleted has ASD	2q37:del	(*82*)
PSEN1	asd2	21	ms	Novel	607822	Point mutation: Alzheimer's disease and EP; Notch signaling cleavage, synaptic homeostasis	N	(*83*)
PTEN	asd2	21	ms	Syndromic (overgrowth syndromes), OMIM	153480, 605309	Point mutations: autism, macrocephaly, ID; point mutations and CNV: Bannayan-Riley-Ruvalcaba syndrome, Cowden syndrome	N	(*13, 84-86*)
RPS6KA3	asd2	45	ns	Syndromic, OMIM	303600, 300844	CNV and point mutations: Coffin-Lowry syndrome and nonsyndromic X-linked mental retardation.	Y	(*87-89*)
RUVBL1	asd2	22	ms	Novel		Interacts with beta-catenin, p53 pathway; cancer related	N	(*90, 91*)
SBF1	asd2	97	ms	CNV, SET domain		Within the 22q13 Deletion Syndrome	22q13:del/dup	(*14*)

Gene (RefSeq)	Pool	# MIPs	Initial Mut	Nomination Category	OMIM	Notes	CNV Reports/ Implicated Regions	References
						region.		
SESN2	asd2	23	ms	Novel		Linked to p53, mTOR,-TSC1/2 pathway and NMDA receptors	N	(*92*)
SETBP1	asd2	55	fs	CNV, novel, OMIM	269150	CNV: del(18) Syndrome: hypotonia, expressive language delay, short stature, and behavioral problems ; recurrent missense (gain-of-function): Schinzel-Giedion midface retraction syndrome	del(18)(q12.2q21.1) Syndrome	(*93, 94*)
SETD2	asd2	95	fs	Novel, SET domain		Belongs to the histone-lysine methyltransferase family; Cancer related	N	(*95*)
SGSM3	asd2	43	ms	Novel		RAP and RAB-mediated neuronal signal transduction; Associated with neuronal function	N	(*96*)
TBL1XR1	asd2	28	ms	Novel, beta-catenin, ubiquitin		Links to both wnt and ubiquitin signaling; Lymphoma related	N	(*45, 46*)
TBR1	asd2	25	fs	Novel		Mouse KO: molecular and functional defects in early-born cortical neurons, cortical malformation	N	(*39, 97*)
TSPAN17	asd2	17	ns	CNV, novel		Within Cooper et al. defined novel pathogenic genomic region	5q35.2:del	(*14*)
UBE3C	asd2	55	ms	Novel		Association with cocaine dependence and major depressive episode	N	(*98*)
USP15	asd2	46	fs	Novel, beta-catenin		Functions in the USP regulation of beta-catenin	N	(*99*)
ZBTB41	asd2	37	ms	Novel		Little known, brain expressed zinc finger	N	
ZNF420	asd2	25	ms	Novel		Binds to p53 and regulates (Apak)	N	(*100*)
Total		1714						

Abbreviations: ASD-autism spectrum disorders, DD-developmental delay, ID-intellectual disability, SCZ-schizophrenia, EP-epilepsy, OMIM-Online Mendelian Inheritance in Man, CNV-copy number variant, Mut-mutation type, fs-frameshifting indel, ns-nonsense, sp-splice-site, aa-single amino acid deletion, ms-missense.

Table S4. Sequences of primers, ASD1, and ASD2 MIP probes.
See accompanying Excel spreadsheet.

Table S5. Comparison of ASD1 MIP variant calls to exome variant calls for 48 samples (16 trios).

	Total	in dbSNP	% in dbSNP	Not in dbSNP	% Novel
Called by both	725	699	96.4	26*	3.6
Called only by MIP	72	63	87.5	9	12.5
Called only by exome	7	7	100	0	0

*All *de novo* events called by exome sequencing were also called by MIP-based resequencing.

Table S6. Sensitivity and PPV for MIP variant calls (JS set) compared with bi-directional Sanger sequence under various calling heuristics.

Condition	PPV (Precision)	Sensitivity (Recall)	F1	F2	Sensitivity (Recall) @ Covered Sites (CS)	F1 @ CS	F2 @ CS	True +	False +	False -	Coverage Below Threshold
10_20_0.65	0.9594	0.9220	0.9403	0.9292	0.9793	0.9692	0.9752	189	8	4	12
10_20_0.70	0.9550	0.9317	0.9432	0.9363	0.9896	0.9720	0.9825	191	9	2	12
10_20_0.75	0.9458	**0.9366**	0.9412	0.9384	**0.9948**	0.9697	0.9846	192	11	1	12
10_20_0.80	0.9366	**0.9366**	0.9366	0.9366	**0.9948**	0.9648	0.9826	192	13	1	12
10_30_0.65	0.9641	0.9171	0.9400	0.9261	0.9741	0.9691	0.9721	188	7	5	12
10_30_0.70	0.9596	0.9268	0.9429	0.9332	0.9845	0.9719	0.9794	190	8	3	12
10_30_0.75	0.9502	0.9317	0.9409	0.9354	0.9896	0.9695	0.9815	191	10	2	12
10_30_0.80	0.9455	0.9317	0.9386	0.9344	0.9896	0.9671	0.9805	191	11	2	12
20_20_0.65	**0.9844**	0.9220	0.9521	0.9338	0.9793	0.9818	0.9803	189	3	4	12
20_20_0.70	0.9795	0.9317	0.9550	0.9409	0.9896	0.9845	0.9876	191	4	2	12
20_20_0.75	0.9746	**0.9366**	0.9552	0.9440	**0.9948**	0.9846	0.9907	192	5	1	12
20_20_0.80	0.9697	**0.9366**	0.9529	0.9430	**0.9948**	0.9821	0.9897	192	6	1	12
20_30_0.65	0.9843	0.9171	0.9495	0.9298	0.9741	0.9792	0.9761	188	3	5	12
20_30_0.70	0.9794	0.9268	0.9524	0.9369	0.9845	0.9819	0.9834	190	4	3	12
20_30_0.75	0.9745	0.9317	0.9526	0.9400	0.9896	0.9820	0.9866	191	5	2	12
20_30_0.80	0.9695	0.9317	0.9502	0.9390	0.9896	0.9795	0.9856	191	6	2	12
25_20_0.65	**0.9844**	0.9220	0.9521	0.9338	0.9793	0.9818	0.9803	189	3	4	12
25_20_0.70	0.9795	0.9317	0.9550	0.9409	0.9896	0.9845	0.9876	191	4	2	12
25_20_0.75	0.9796	**0.9366**	**0.9576**	**0.9449**	**0.9948**	**0.9871**	**0.9917**	192	4	1	12
25_20_0.80	0.9746	**0.9366**	0.9552	0.9440	**0.9948**	0.9846	0.9907	192	5	1	12
25_30_0.65	0.9843	0.9171	0.9495	0.9298	0.9741	0.9792	0.9761	188	3	5	12
25_30_0.70	0.9794	0.9268	0.9524	0.9369	0.9845	0.9819	0.9834	190	4	3	12
25_30_0.75	0.9795	0.9317	0.9550	0.9409	0.9896	0.9845	0.9876	191	4	2	12
25_30_0.80	0.9745	0.9317	0.9526	0.9400	0.9896	0.9820	0.9866	191	5	2	12
30_20_0.65	**0.9844**	0.9220	0.9521	0.9338	0.9793	0.9818	0.9803	189	3	4	12
30_20_0.70	0.9795	0.9317	0.9550	0.9409	0.9896	0.9845	0.9876	191	4	2	12
30_20_0.75	0.9796	**0.9366**	**0.9576**	**0.9449**	**0.9948**	**0.9871**	**0.9917**	192	4	1	12
30_20_0.80	0.9796	**0.9366**	**0.9576**	**0.9449**	**0.9948**	**0.9871**	**0.9917**	192	4	1	12
30_30_0.65	0.9843	0.9171	0.9495	0.9298	0.9741	0.9792	0.9761	188	3	5	12
30_30_0.70	0.9794	0.9268	0.9524	0.9369	0.9845	0.9819	0.9834	190	4	3	12
30_30_0.75	0.9795	0.9317	0.9550	0.9409	0.9896	0.9845	0.9876	191	4	2	12
30_30_0.80	0.9795	0.9317	0.9550	0.9409	0.9896	0.9845	0.9876	191	4	2	12
40_20_0.65	0.9842	0.9122	0.9468	0.9257	0.9791	0.9816	0.9801	187	3	4	14
40_20_0.70	0.9793	0.9220	0.9497	0.9329	0.9895	0.9844	0.9875	189	4	2	14
40_20_0.75	0.9794	0.9268	0.9524	0.9369	**0.9948**	0.9870	0.9916	190	4	1	14
40_20_0.80	0.9794	0.9268	0.9524	0.9369	**0.9948**	0.9870	0.9916	190	4	1	14
40_30_0.65	0.9841	0.9073	0.9442	0.9217	0.9738	0.9789	0.9759	186	3	5	14
40_30_0.70	0.9792	0.9171	0.9471	0.9289	0.9843	0.9817	0.9833	188	4	3	14

Condition	PPV (Precision)	Sensitivity (Recall)	F1	F2	Sensitivity (Recall) @ Covered Sites (CS)	F1 @ CS	F2 @ CS	True +	False +	False -	Coverage Below Threshold
40_30_0.75	0.9793	0.9220	0.9497	0.9329	0.9895	0.9844	0.9875	189	4	2	14
40_30_0.80	0.9793	0.9220	0.9497	0.9329	0.9895	0.9844	0.9875	189	4	2	14
50_20_0.65	0.9842	0.9122	0.9468	0.9257	0.9791	0.9816	0.9801	187	3	4	14
50_20_0.70	0.9793	0.9220	0.9497	0.9329	0.9895	0.9844	0.9875	189	4	2	14
50_20_0.75	0.9794	0.9268	0.9524	0.9369	**0.9948**	0.9870	0.9916	190	4	1	14
50_20_0.80	0.9794	0.9268	0.9524	0.9369	**0.9948**	0.9870	0.9916	190	4	1	14
50_30_0.65	0.9841	0.9073	0.9442	0.9217	0.9738	0.9789	0.9759	186	3	5	14
50_30_0.70	0.9792	0.9171	0.9471	0.9289	0.9843	0.9817	0.9833	188	4	3	14
50_30_0.75	0.9793	0.9220	0.9497	0.9329	0.9895	0.9844	0.9875	189	4	2	14
50_30_0.80	0.9793	0.9220	0.9497	0.9329	0.9895	0.9844	0.9875	189	4	2	14

Condition nomenclature is X_Y_Z, where X is minimum coverage, Y is minimum consensus or SNP quality, and Z is allele balance cutoff (maximum fraction of Q20 reference bases). Maximum values are bolded.

Table S7. Copy number calls from known deletion and duplication carriers.

True+

Sample	CNV	Gene	Chr	Start	Stop	Size (kb)	Z-score (A)	Z-score (B)	SD	#Probes
8893	5q34del	*GABRA1, GABRG2*	5	161,277,798	161,580,177	302	-3.71	-3.73	0.01	78
T1158	1p34.3dup	*DLGAP3*	1	35,350,603	35,371,014	20	3.29	3.22	0.05	16
T18349	5p13.2del	*SLC1A3*	5	36,608,549	36,686,391	78	-3.11	-3.13	0.01	42
T2217	9p22.2del	*SH3GL2*	9	17,747,040	17,795,624	48	-3.81	-3.45	0.26	30
T2948	15q26.1del	*SV2B*	15	91,769,498	91,835,683	66	-3.40	-3.35	0.03	54
T3729	9p24.1-22.2del	*SMARCA2, SLC1A1*	9	2,029,046	4,585,563	256	-3.70	-3.66	0.03	160
T3729	9p22.2dup	*SH3GL2*	9	17,747,040	17,795,624	49	2.72	3.11	0.27	30
T3810	1q32.1dup	*SYT2*	1	202,565,899	202,574,761	89	6.33	5.45	0.62	22
T892	3q11.2dup	*EPHA6*	3	96,706,202	97,467,516	761	3.54	3.16	0.27	72
T964	7q21.11del	*GNAI1, CACNA2D1*	7	79,818,435	81,579,719	176	-3.69	-3.90	0.15	126

False-

Sample	CNV	Gene	Chr	Start	Stop	size(kb)	Notes
K5003-3	20q13.33del	*KCNQ2*	20	62,037,996	62,103,816	66	insufficient coverage for reliable CNV calling
T20847	7q21.11del	*SEMA3A*	7	83,823,790	83,823,902	1	insufficient coverage for reliable CNV calling
T3472	2q33.3del	*ADAM23*	2	207,308,590	207,310,248	17	insufficient coverage for reliable CNV calling
T1962	4p16del	*CPLX1*	4	780,288	818,810	38	insufficient coverage for reliable CNV calling
T439/438	7q21.11del	*CACNA2D, GNAI1*	7	79,764,470	82,072,774	2308	sample too noisy for reliable CNV calling*

Footnotes: Coordinates are those detected by MIP-based CNV calling, or in the case of false negative samples, the expected aberration coordinates obtained from array CGH. All coordinates are hg19 and the presented data excludes the X chromosome. SD = standard deviation. *The number of called events in the sample exceeded 10 SDs from mean calculated from the entire cohort.

Table S8. Summary of epilepsy gene set and probes usable for copy number analysis.

Gene	Usable Probes[1]	Total Probes[2]	% Usable
ADAM22	63	65	97
ADAM23	41	53	77
ARX†	11	20	55
CACNA2D1	53	79	67
CDKL5†	49	50	98
CPLX1	0	7	0
DLGAP1	31	40	78
DLGAP2	17	39	44
DLGAP3	8	40	20
EPHA6	36	53	68
GABRA1	20	22	91
GABRG2	19	26	73
GNAI1	10	17	59
GRIN2A	40	56	71
KCNQ2	0	44	0
MAPK8	18	23	78
OTX1	0	13	0
PCDH19†	35	39	90
SCN1A	75	91	82
SCN1B	0	17	0
SCN2A	70	81	86
SEMA3A	27	38	71
SH3GL2	15	18	83
SLC1A1	23	26	88
SLC1A3	21	24	88
SLC2A1	16	31	52
SMARCA2	57	76	75
SPTAN1	87	120	73
STXBP1	35	40	88
SV2A	24	32	75
SV2B	27	27	100
SYT2	11	18	61
Total	**939**	**1325**	**71**

[1]The number of probes that passed copy number quality control filtering.
[2]Includes 17 probes that overlap SNP locations and cannot be distinguished.
†X-linked loci were not included in final analysis.

Table S9. Sensitivity and PPV for 357 epilepsy samples compared with array CGH under various calling heuristics.

Condition	PPV (Precision)	Sensitivity (Recall)	Sensitivity (Recall) @ Covered Sites	True +	False -	False +	Coverage Below Probe Threshold
5_2.5	0.11	0.64	0.90	9	5	73	4
5_2.6	0.23	0.64	0.90	9	5	31	4
5_2.7	0.35	0.64	0.90	9	5	17	4
5_2.8	0.56	0.64	0.90	9	5	7	4
5_2.9	0.75	0.64	0.90	9	5	3	4
5_3.0	0.75	0.64	0.90	9	5	3	4
6_2.5	0.39	0.64	0.90	9	5	14	4
6_2.6	0.56	0.64	0.90	9	5	7	4
6_2.7	0.69	0.64	0.90	9	5	4	4
6_2.8	0.75	0.64	0.90	9	5	3	4
6_2.9	0.82	0.64	0.90	9	5	2	4
6_3.0	0.82	0.64	0.90	9	5	2	4
7_2.5	0.43	0.64	0.90	9	5	12	4
7_2.6	0.69	0.64	0.90	9	5	4	4
7_2.7	0.83	**0.71**	**1.00**	10	4	2	4
7_2.8	**0.90**	0.64	0.90	9	5	1	4
7_2.9	**0.90**	0.64	0.90	9	5	1	4
7_3.0	**0.90**	0.64	0.90	9	5	1	4
8_2.5	0.56	0.64	0.90	9	5	7	4
8_2.6	0.75	0.64	0.90	9	5	3	4
8_2.7	0.83	**0.71**	**1.00**	10	4	2	4
8_2.8	**0.90**	0.64	0.90	9	5	1	4
8_2.9	**0.90**	0.64	0.90	9	5	1	4
8_3.0	**0.90**	0.64	0.90	9	5	1	4
9_2.5	0.75	0.64	0.90	9	5	3	4
9_2.6	0.82	0.64	0.90	9	5	2	4
9_2.7	0.83	**0.71**	**1.00**	10	4	2	4
9_2.8	**0.90**	0.64	0.90	9	5	1	4
9_2.9	**0.90**	0.64	0.90	9	5	1	4
9_3.0	**0.90**	0.64	0.90	9	5	1	4
10_2.5	0.75	0.64	0.90	9	5	3	4
10_2.6	0.82	0.64	0.90	9	5	2	4
10_2.7	0.83	**0.71**	**1.00**	10	4	2	4
10_2.8	**0.90**	0.64	0.90	9	5	1	4
10_2.9	**0.90**	0.64	0.90	9	5	1	4
10_3.0	**0.90**	0.64	0.90	9	5	1	4

Footnotes: Only autosomal probes were considered. We required true positive sites to be called in both replicates, however, each only counted as a single site. At all sites the number of expected events was 14. At covered sites the number of expected events is 10. The filtering thresholds are expressed as A_B, where A is minimum number of probes and B is average SVD adjusted Z-score of the calling window. Maximum values are bolded.

Table S10. Copy number calls from MIP read depth from ASD2 probe set.

ID	Type	Array Avail-able	Gene	Chr	Call Start	Call Stop	Size*	del/ dup ?	Mean Adjust-ed Z-score	#Probes	In DGV?	Notes
150-12639	NIMH	N	*ARID1B*	6	157,150,408	157,528,997	378589	del	-3.43	64	Y	
150-10729	NIMH	N	*CNOT4*	7	135,047,867	135122985	75118	dup	3.37	26	N	
150-12639	NIMH	N	*MDM2*	12	69,202,218	69233663	31445	dup	3.11	20	N	
150-12491	NIMH	N	*PSEN1*	14	73,614,699	73,685,914	71215	dup	3.34	17	N	
150-12639	NIMH	N	*USP15*	12	62,654,223	62,798,044	143821	dup	2.71	30	N	
11780.p1	Proband	Y	*CNOT4*	7	135,047,867	135,122,985	75118	dup	3.83	26	N	confirmed by array
13697.p1	Proband	Y	*NLGN1*	3	173,525,517	173,999,105	473588	dup	3.23	21	Y	confirmed by array
13529.p1	Proband	Y	*PTEN*	10	89,690,831	89,720,689	29858	dup	4.06	7	N	not supported by array, possible pseudo-gene issue
12404.p1	Proband	Y	*TBL1XR1*	3	176,744,236	176,771,614	27378	dup	3.11	21	N	confirmed by array

*Size estimated only from coding probes.

Table S11. Genes with recurrent *de novo* mutation in ASD probands.

Proband	Sex	Gene	Accession	Mut	Assay[1]	Chr	Pos[2] (hg19)	Ref	Geno[3]	HGVS	Codons[4]	Diagnosis	NVIQ	NIMH[5] /ESP[6]	Notes[7]
12688.p1	M	*ADCY5**	NM_183357	ms	MIP	3	123049782	C	Y	p.Ala534Thr	1262	autism	111	0/0	R (*6*)
11653.p1	M	*ADCY5**	NM_183357	ms	EX	3	123046605	G	R	p.Arg603Cys	1262	autism	44	0/0	
12130.p1	F	*ADNP**	NM_015339	fs	EX	20	49510027	*	-TT/*	p.Lys408ValfsX31	1103	autism	55	0/1	
13545.p1	M	*ADNP**	NM_015339	fs	MIP	20	49509094	*	+T/*	p.Tyr719X	1103	autism	38	0/1	
14393.p1	F	*ARID1B**	NM_020732	fs	MIP	6	157510805	*	+C/*	p.Gln1196ProfsX14	2250	autism	71	0/0	
13447.p1	F	*ARID1B**	NM_020732	fs	EX	6	157527664	*	-TGTT/*	p.Phe1798LeufsX52	2250	autism	51	0/0	
12714.p1	M	*CHD8**	NM_001170629	ns	MIP	14	21899618	G	S	p.Ser62X	2582	autism	78	0/0	
13986.p1	M	*CHD8**	NM_001170629	fs	MIP	14	21878133	*	+T/*	p.Tyr747X	2582	autism	38	0/0	
11654.p1	F	*CHD8**	NM_001170629	sp	MIP	14	21871373	T	Y	c.3519-2A>G	2582	autism	41	0/0	NR (*4*)
13844.p1	M	*CHD8**	NM_001170629	ns	EX	14	21871178	G	R	p.Gln1238X	2582	autism	34	0/0	
14016.p1	M	*CHD8**	NM_001170629	ns	MIP	14	21870169	G	R	p.Arg1337X	2582	autism	92	0/0	
12991.p1	M	*CHD8**	NM_001170629	fs	MIP	14	21861643	*	-CTTC/*	p.Glu2103ArgfsX3	2582	autism	67	0/0	
12752.p1	F	*CHD8**	NM_001170629	fs	EX	14	21861376	*	-CT/*	p.Leu2120ProfsX13	2582	autism	93	0/0	
14233.p1	M	*CHD8**	NM_001170629	fs	MIP	14	21859175	*	+T/*	p.Asn2371LysfsX2	2582	autism	19	0/0	
14406.p1	M	*CHD8**	NM_001170629	aa	MIP	14	21854022	*	-GGT/*	p.His2498del	2582	autism	98	0/0	
12211.p1	F	*CTNNB1**	NM_001098209	ns	MIP	3	41275346	G	R	p.Trp504X	782	autism	57	0/0	
12703.p1	M	*CTNNB1**	NM_001098209	ms	EX	3	41275757	C	Y	p.Thr551Met	782	autism	58	0/0	
12099.p1	M	*DYRK1A**	NM_001396	fs	MIP	21	38845116	*	-AT/*	p.Ile48LysfsX2	764	autism	55	0/0	NR (*4*)
13890.p1	F	*DYRK1A**	NM_001396	sp	EX	21	38865466	G	R	c.1098+1G>A	764	autism	42	0/0	
13552.p1	M	*DYRK1A**	NM_001396	fs	MIP	21	38877833	*	-C/*	p.Ala498ProfsX94	764	autism	66	0/0	R (*6*)
11691.p1	M	*GRIN2B*[†]	NM_000834	fs	MIP[&]	12	14019043	*	+G/*	p.Ser34GlnfsX25	1485	autism	62	0/0	NR (*3*)
13932.p1	M	*GRIN2B*[†]	NM_000834	ms	MIP	12	13768560	C	Y	p.Cys456Tyr	1485	autism	55	0/0	
12547.p1	M	*GRIN2B*[†]	NM_000834	ns	MIP[&]	12	13764762	C	Y	p.Trp559X	1485	autism	65	0/0	
12681.p1	F	*GRIN2B*[†]	NM_000834	sp	EX	12	13722953	T	Y	c.2172-2A>G	1485	autism	65	0/0	
11666.p1	M	*LAMC3*	NM_006059	ms	EX	9	133914290	A	R	p.Asp339Gly	1576	autism	51	1/4	
11704.p1	M	*LAMC3*	NM_006059	ms	MIP[&]	9	133952690	A	R	p.Tyr1249Cys	1576	autism	82	1/4	
12532.p1	M	*NTNG1*	NM_001113226	ms	EX	1	107691283	A	R	p.Tyr23Cys	540	autism	110	0/0	
11660.p1	F	*NTNG1*	NM_001113226	ms	EX	1	107867061	C	Y	p.Thr135Ile	540	autism	60	0/0	
14433.p1	M	*PTEN*	NM_000314	ms	MIP	10	89692908	C	Y	p.Thr131Ile	404	autism	50	0/0	
14611.p1	M	*PTEN*	NM_000314	fs	MIP	10	89692920	*	+A/*	p.Cys136MetfsX44	404	autism	33	0/0	
11390.p1	F	*PTEN*	NM_000314	ms	EX	10	89711882	C	M	p.Thr167Asn	404	autism	77	0/0	
12930.p1	M	*SBF1*	NM_002972	ms	MIP	22	50901032	G	R	p.Arg695Trp	1894	autism	76	0/1	NR (*9*)
13793.p1	M	*SBF1*	NM_002972	ms	EX	22	50899030	T	Y	p.Thr1027Ala	1894	autism	56	0/1	
12499.p1	M	*SCN1A**	NM_001165963	ms	EX	2	166848071	G	R	p.Pro1905Leu	2010	autism	67	1/0	
12340.p1	F	*SCN1A**	NM_001165963	ms	MIP[&]	2	166848006	T	Y	p.Arg1927Gly	2010	autism	29	1/0	NR (*4*)
14569.p1	M	*SETD2*	NM_014159	ms	MIP	3	47166005	T	W	p.Ile41Phe	2565	asd	144	0/1	

Proband	Sex	Gene	Accession	Mut	Assay[1]	Chr	Pos[2] (hg19)	Ref	Geno[3]	HGVS	Codons[4]	Diagnosis	NVIQ	NIMH[5] /ESP[6]	Notes[7]
12565.p1	F	*SETD2*	NM_014159	fs	EX	3	47098932	*	-T/*	p.Asn2114IlefsX33	2565	autism	79	0/1	
11523.p1	M	*SGSM3*	NM_015705	ms	EX	22	40802589	C	Y	p.Arg370Cys	750	autism	85	0/5	
12791.p1	M	*SGSM3*	NM_015705	ms	MIP	22	40804965	A	R	p.Asp644Gly	750	asd	127	0/5	
12335.p1	F	*TBL1XR1**	NM_024665	ms	EX	3	176765107	A	R	p.Leu282Pro	515	autism	47	0/0	
14612.p1	M	*TBL1XR1**	NM_024665	fs	MIP	3	176752047	*	-A/*	p.Ile397SerfsX19	515	autism	41	0/0	
11480.p1	M	*TBR1*[†]	NM_006593	fs	EX	2	162273322	*	-C/*	p.Ala136ProfsX80	683	autism	41	0/0	
13814.p1	M	*TBR1*[†]	NM_006593	ms	MIP	2	162273603	A	R	p.Lys228Glu	683	autism	78	0/0	
13796.p1	F	*TBR1*[†]	NM_006593	fs	MIP	2	162275481	*	+C/*	p.Ser351X	683	autism	63	0/0	NR (*4*)
11006.p1	M	*UBE3C**	NM_014671	ms	EX	7	157041114	C	Y	p.Ser845Phe	1084	autism	125	0/0	
12851.p1	M	*UBE3C**	NM_014671	ms	MIP	7	157049644	T	K	p.Phe996Cys	1084	autism	49	0/0	R (*6*)

[1]Primary assay that identified the variant. [2]Chromosomal position in the human genome hg19 assembly. For indels, the position listed follows the SAMtools/VCF convention of listing the position prior to the event. [3]Genotypes listed are heterozygous using IUPAC codes for SNVs. For indels, "*" indicates a copy of the reference allele, while the +/- indicates the sequence inserted or deleted. [4]Number of protein codons including stop. [5]Number of observed nonsense, splice-site, or indel events in MIP-screened NIMH samples. [6]Number of observed nonsense and splice-site events in the ESP6500 data release. [7]If the proband was previously exome sequenced. *Part of 49-member connected component reported in (*3*). [†]Part of expanded 74-member connected component. [&]Variant reported in MIP screen from (*3*). Abbreviations: M-male, F-female, Mut-mutation type, fs-frameshifting indel, ns-nonsense, sp-splice-site, aa-single amino acid deletion, ms-missense, NVIQ-nonverbal intellectual quotient, R-reported in cited exome study, NR-not reported in cited exome study.

Table S12. Fraction of nonsynonymous variants by class from SSC proband sequencing data.

	missense	nonsense/ splice/indel	nonsense	splice	indel
SSC209 Exome rare[1]	0.93	0.07	0.03	0.01	0.04
SSC209 44 genes rare	0.96	0.04	0.01	0.00	0.03
SSCMIP 44 genes rare[2]	0.97	0.03	0.01	0.01	0.01
SSC209 Exome *de novo*	0.80	0.20	0.09	0.02	0.09
SSC209 44 genes *de novo*	0.61	0.39	0.09	0.05	0.25
SSCMIP 44 genes *de novo*	0.37	0.63	0.15	0.04	0.44
Random (*15*)	0.84	0.16	0.04	0.04	0.08

[1]Exome sites from the 209 SSC probands previously published (*2, 3*). Rare defined as not present in 1,779 other exomes (*12*). [2]MIP rare sites were not present in 1,779 other exomes and were also singletons in the MIP data set (i.e., not present in another proband or the 762 NIMH samples).

Table S13. Locus-specific mutation rate estimates for 44 genes.

Gene	Chimp diffs	Length mapped to chimp	Total length of sequence	Mut rate per site (diffs / mapped seq)	Mutation rate/base/gen*	Lower 95% CI†	Upper 95% CI†
ADCY5	15	3834	3958	3.91E-03	8.15E-09	3.80E-09	1.31E-08
ADNP	11	3321	3321	3.31E-03	6.90E-09	3.13E-09	1.13E-08
AP3B2	11	3353	3353	3.28E-03	6.83E-09	3.10E-09	1.12E-08
ARID1B	21	5284	6830	3.97E-03	8.28E-09	5.12E-09	1.18E-08
BRSK2	10	1398	2087	7.15E-03	1.49E-08	7.39E-09	2.39E-08
CHD7	19	9141	9142	2.08E-03	4.33E-09	2.50E-09	6.39E-09
CHD8	15	7904	7904	1.90E-03	3.95E-09	2.11E-09	6.07E-09
CNOT4	5	2288	2291	2.19E-03	4.55E-09	9.09E-10	9.13E-09
CTNNB1	8	2402	2402	3.33E-03	6.94E-09	2.60E-09	1.30E-08
CUL3	6	2371	2371	2.53E-03	5.27E-09	1.75E-09	9.69E-09
CUL5	4	2419	2419	1.65E-03	3.44E-09	8.60E-10	6.90E-09
DYRK1A	9	2420	2420	3.72E-03	7.75E-09	3.44E-09	1.38E-08
FOXP1	6	2267	2267	2.65E-03	5.51E-09	1.83E-09	1.01E-08
FOXP2	8	2379	2379	3.36E-03	7.01E-09	2.62E-09	1.23E-08
GRIN2A	20	4443	4443	4.50E-03	9.38E-09	5.62E-09	1.36E-08
GRIN2B	15	4503	4503	3.33E-03	6.94E-09	3.70E-09	1.07E-08
HDGFRP2	20	1522	2084	1.31E-02	2.74E-08	1.63E-08	3.99E-08
HDLBP	17	3911	3911	4.35E-03	9.06E-09	4.78E-09	1.33E-08
LAMC3	40	4815	4840	8.31E-03	1.73E-08	1.21E-08	2.30E-08
MBD5	20	4525	4525	4.42E-03	9.21E-09	5.52E-09	1.34E-08
MDM2	3	1538	1538	1.95E-03	4.06E-09	0	8.14E-09
NLGN1	12	2492	2492	4.82E-03	1.00E-08	5.00E-09	1.59E-08
NOTCH3	42	6049	7098	6.94E-03	1.45E-08	1.03E-08	1.90E-08
NR4A2	4	1821	1821	2.20E-03	4.58E-09	1.14E-09	1.03E-08
NTNG1	2	1782	1782	1.12E-03	2.34E-09	0	5.86E-09
OPRL1	12	500	1125	2.40E-02	5.00E-08	2.47E-08	8.03E-08
PDCD1	13	591	887	2.20E-02	4.58E-08	2.09E-08	7.13E-08
PSEN1	2	1444	1444	1.39E-03	2.89E-09	0	7.23E-09
PTEN	2	1248	1248	1.60E-03	3.34E-09	0	8.37E-09
RPS6KA3	3	2108	2311	1.42E-03	2.96E-09	0	6.93E-09
RUVBL1	1	1415	1415	7.07E-04	1.47E-09	0	4.42E-09
SBF1	40	5843	5846	6.85E-03	1.43E-08	1.03E-08	1.89E-08
SCN1A	26	6134	6134	4.24E-03	8.83E-09	5.43E-09	1.22E-08
SESN2	10	1443	1483	6.93E-03	1.44E-08	5.75E-09	2.47E-08
SETBP1	18	5004	5004	3.60E-03	7.49E-09	4.16E-09	1.13E-08
SETD2	31	7704	7779	4.02E-03	8.38E-09	5.67E-09	1.14E-08
SGSM3	12	2190	2334	5.48E-03	1.14E-08	5.69E-09	1.81E-08
TBL1XR1	1	1601	1601	6.25E-04	1.30E-09	0	3.91E-09
TBR1	4	2073	2073	1.93E-03	4.02E-09	1.00E-09	8.06E-09

Gene	Chimp diffs	Length mapped to chimp	Total length of sequence	Mut rate per site (diffs / mapped seq)	Mutation rate/base/gen*	Lower 95% CI†	Upper 95% CI†
TSPAN17	6	944	1035	6.36E-03	1.32E-08	4.40E-09	2.44E-08
UBE3C	9	3274	3344	2.75E-03	5.73E-09	2.54E-09	1.02E-08
USP15	9	2943	2943	3.06E-03	6.37E-09	2.83E-09	1.06E-08
ZBTB41	6	2770	2770	2.17E-03	4.51E-09	1.50E-09	8.29E-09
ZNF420	5	2079	2079	2.41E-03	5.01E-09	1.00E-09	1.00E-08

*We assumed a divergence time between human and chimpanzee of 12 million years ago and an average generation time of 25 years. †95% confidence interval based on 1,000 bootstraps.

Table S14. *De novo* variants identified in SSC unaffected sibling exome data and rare, severe variants identified in NIMH cohort and Exome Sequencing Project (ESP) in the 44 targeted genes.

Sample	Type	Gene	Mut Type	Chr	Pos (hg19)[1]	Ref	Allele	Validation Status	ESP[2] #Het	ESP[2] #Homo Ref	Study
13094.s1	sibling	*CNOT4*	ms	7	135078802	G	A	Not Attempted	NA	NA	Iossifov
12926.s1	sibling	*SESN2*	syn	1	28601365	G	A	Not Attempted	NA	NA	Iossifov
150-12552	NIMH	*FOXP2*	sp	7	114271582	G	C	Validated	NA	NA	
150-10273	NIMH	*HDGFRP2*	fs	19	4491807	*	+C	Validated	NA	NA	
150-10140	NIMH	*LAMC3*	fs	9	133914337	*	-G	Validated	NA	NA	
150-13048	NIMH	*SCN1A*	ns	2	166901591	G	A	Validated	0	6503	
150-12952	NIMH	*USP15*	ns	12	62798045	C	T	Validated	1	6502	
150-13699	NIMH	*ZNF420*	ns	19	37619669	T	A	Validated	1	6502	
unknown	ESP	*ADNP*	ns	20	49510170	G	A	Not Attempted	1	6502	
unknown	ESP	*CUL5*	ns	11	107920788	C	T	Not Attempted	1	6496	
unknown	ESP	*HDLBP*	ns	2	242195646	G	A	Not Attempted	1	6502	
unknown	ESP	*LAMC3*	ns	9	133914634	A	T	Not Attempted	1	6502	
unknown	ESP	*LAMC3*	sp	9	133942593	G	A	Not Attempted	2	6501	
unknown	ESP	*LAMC3*	ns	9	133954629	C	T	Not Attempted	1	6502	
unknown	ESP	*SBF1*	sp	22	50904578	C	T	Not Attempted	1	6328	
unknown	ESP	*SESN2*	sp	1	28599305	G	C	Not Attempted	1	6500	
unknown	ESP	*SETBP1*	sp	18	42449194	G	A	Not Attempted	2	6501	
unknown	ESP	*SETD2*	ns	3	47079170	C	A	Not Attempted	1	6502	
unknown	ESP	*SGSM3*	ns	22	40803844	C	T	Not Attempted	1	6502	
unknown	ESP	*SGSM3*	sp	22	40804279	A	T	Not Attempted	1	6501	
unknown	ESP	*SGSM3*	ns	22	40805375	C	T	Not Attempted	2	6501	
unknown	ESP	*SGSM3*	sp	22	40805530	G	A	Not Attempted	1	6502	
unknown	ESP	*TSPAN17*	ns	5	176074632	C	T	Not Attempted	1	6495	
unknown	ESP	*TSPAN17*	ns	5	176078811	G	A	Not Attempted	1	6502	
unknown	ESP	*TSPAN17*	ns	5	176081914	C	T	Not Attempted	1	6502	
unknown	ESP	*USP15*	ns	12	62798045	C	T	Not Attempted	1	6502	
unknown	ESP	*ZBTB41*	ns	1	197128671	G	A	Not Attempted	1	6502	
unknown	ESP	*ZNF420*	ns	19	37581953	C	A	Not Attempted	1	6502	
unknown	ESP	*ZNF420*	ns	19	37618182	G	T	Not Attempted	1	6502	
unknown	ESP	*ZNF420*	ns	19	37619669	T	A	Not Attempted	1	6502	

[1]Chromosomal position in the human genome hg19 assembly. Indel positions follow the SAMtools/VCF convention of listing the position before the event. [2]ESP call set does not include indels. Number of heterozygous and homozygous reference individuals from the ESP6500 release (http://evs.gs.washington.edu/EVS/). Abbreviations: syn-synonymous, fs-frameshifting indel, ns-nonsense, sp-splice-site, ms-missense.

Table S15. Inherited truncation/splice events identified in ASD probands.

Proband	Sex	Gene	Accession	Mut	Origin[1]	Chr	Pos[2] (hg19)	Ref	Geno[3]	HGVS	Codons[4]	Diagnosis	NVIQ	NIMH[5]/ ESP[6]	Sib[7]
12652.p1	M	*ADCY5**	NM_183357	fs	fa	3	123166430	*	-T/*	p.Gln321ArgfsX56	1262	autism	79	0/0	s1-NT
12707.p1	M	*ADNP**	NM_015339	fs	mo	20	49507971	*	+GG/*	p.Gly1094ProfsX5	1103	autism	90	0/1	s2-T
11255.p1	M	*AP3B2*	NM_004644	fs	mo	15	83357565	*	-AT/*	p.Tyr94CysfsX8	1083	autism	108	0/0	s2-NT
13704.p1	M	*CHD7**	NM_017780	fs	mo	8	61777890	*	-T/*	p.Leu2798ArgfsX19	2998	autism	93	0/0	NA
14023.p1	M	*CHD7**	NM_017780	fs	fa	8	61778440	*	+A/*	p.Asp2982ArgfsX4	2998	autism	69	0/0	s1-NT
12389.p1	M	*CNOT4*	NM_001190850	sp	fa	7	135095264	C	Y	c.821+1G>A	714	autism	94	0/0	NA
13680.p1	M	*DYRK1A**	NM_101395[†]	sp-3'UTR	fa	21	38878659	T	K	c.*47+2T>G	585	autism	76	0/0	s1-NT
14595.p1	M	*FOXP2*	NM_148900[†]	fs	fa	7	114269648	*	+T/*	p.Leu135PhefsX148	733	autism	79	1/0	s1-NT, s2-T
14316.p1	M	*LAMC3*	NM_006059	sp	fa	9	133911720	G	S	c.976+1G>C	1576	autism	72	1/4	s1-T
14404.p1	M	*LAMC3*	NM_006059	ns	mo	9	133932503	T	W	p.Cys709X	1576	autism	20	1/4	s1-T
14144.p1	F	*LAMC3*	NM_006059	fs	mo	9	133942353	*	-C/*	p.Cys786ValfsX98	1576	autism	54	1/4	NA
11298.p1	M	*LAMC3*	NM_006059	ns	mo	9	133954629	C	Y	p.Arg1291X	1576	autism	132	1/4	s1-NT
11797.p1	M	*MBD5*	NM_018328	sp	mo	2	149243521	T	W	c.3054+2T>A	1495	autism	117	0/0	s1-T
12517.p1	M	*PTEN*	NM_000314	ns	mo	10	89717615	C	Y	p.Gln214X	404	autism	41	0/0	s1-NT
13811.p1	M	*SBF1*	NM_002972	ns	mo	22	50904837	G	R	p.Gln244X	1894	aspergers	100	0/1	NA
14012.p1	M	*SETBP1*	NM_015559	ns	mo	18	42643634	C	Y	p.Arg1588X	1597	autism	94	0/2	s1-NT
12736.p1	M	*SETD2*	NM_014159	ns	mo	3	47205396	G	R	p.Gln7X	2565	autism	111	0/1	s1-NT
13783.p1	M	*SETD2*	NM_014159	ns	fa	3	47164944	A	W	p.Cys394X	2565	asd	79	0/1	NA
13625.p1	M	*SGSM3*	NM_015705	sp	fa	22	40802668	T	Y	c.1185+2T>C	750	autism	100	0/5	s1-T, s2-NT
12637.p1	M	*SGSM3*	NM_015705	sp	fa	22	40804702	G	R	c.1853+1G>A	750	asd	110	0/5	s1-NT
11494.p1	M	*TSPAN17*	NM_012171	ns	mo	5	176074644	G	K	p.Glu10X	333	autism	107	0/3	s1-NT
14096.p1	M	*UBE3C**	NM_014671	fs	mo	7	156979599	*	+A/*	p.Asn406LysfsX47	1084	autism	75	0/0	s1-NT
14557.p1	M	*ZNF420*	NM_144689	ns	fa	19	37619946	C	Y	p.Gln685X	689	autism	104	1/3	s1-NT

[1]Parent that carried the variant. [2]Chromosomal position in the human genome hg19 assembly. For indels, the position listed follows the SAMtools/VCF convention of listing the position prior to the event. [3]Genotypes listed are heterozygous using IUPAC codes for SNVs. For indels, "*" indicates a copy of the reference allele, while the +/- indicates the sequence inserted or deleted. [4]Number of protein codons including stop. [5]Number of observed nonsense, splice-site, or indel events in MIP-screened NIMH samples. [6]Number of observed nonsense and splice-site events in the ESP6500 data release. [7]Unaffected sibling carrier status. *Part of 49-member connected component reported in (*3*). [†]Variant only affects an alternative protein isoform. Abbreviations: Mut-mutation type, fs-frameshifting indel, ns-nonsense, sp-splice-site, sp-3'UTR-splice-site of a 3' translated region exon, ms-missense, fa-father, mo-mother, s1-unaffected sibling 1, s2-unaffected sibling 2, NVIQ-nonverbal intellectual quotient, NT-not transmitted, T-transmitted, NA-not available.

Table S16. Rare (<0.1%) copy number calls intersecting with Tables S11 and S15 probands.

MIP detected variants				Rare and *de novo* CNVs								
Sample	**Gene**	**Mut Type**	**Mut Origin**	**Study**	**Chr**	**Start (hg18)**	**Stop (hg18)**	**Size (kb)**	**CNV Type**	**Probes/ Exons**	**CNV Origin**	**Notes**
11653.p1	*ADCY5*	ms	DN	Sanders	3	147,530,802	147,649,939	119	Dup	36	Pat	Rare
11653.p1	*ADCY5*	ms	DN	Sanders	22	39,907,910	39,957,019	49	Dup	24	Pat	Rare
12130.p1	*ADNP*	fs	DN	Sanders	5	112,939,064	112,974,949	36	Dup	23	Mat	Rare
12130.p1	*ADNP*	fs	DN	Exome_ CNV	10	132,997,405	133,611,285	614	Dup	7	Mat	Rare
12991.p1	*CHD8*	fs	DN	HS3.1_ CGH	5	2,427,405	3,219,216	792	Dup	NA	Mat	Rare
14016.p1	*CHD8*	ns	DN	HS3.1_ CGH	6	189,784	2,673,615	2,484	Dup	NA	DN	Duplication of ATR-16 Deletion Locus.
13844.p1	*CHD8*	ns	DN	Exome_ CNV	10	3,114,579	3,811,782	697	Del	49	Mat	Associated with nominally significant genes in (*14*).
12752.p1	*CHD8*	fs	DN	Exome_ CNV	11	99,684,300	100,309,201	625	Dup	12	Mat	Rare
13844.p1	*CHD8*	ns	DN	Exome_ CNV	13	22,675,833	22,722,856	47	Dup	3	Mat	Rare
13844.p1	*CHD8*	ns	DN	Exome_ CNV	17	4,157,064	4,336,627	180	Del	12	Pat	Rare
12211.p1	*CTNNB1*	ns	DN	Levy	8	53,573,143	53,983,978	411	Dup	NA	Mat	Rare
12211.p1	*CTNNB1*	ns	DN	HS3.1_ CGH	11	107,733,766	107,757,732	24	Dup	NA	Pat	Rare
12211.p1	*CTNNB1*	ns	DN	Sanders	14	20,885,048	20,945,327	60	Del	28	UNK	Rare
13890.p1	*DYRK1A*	sp	DN	Exome_ CNV	1	217,642,150	218,168,405	526	Dup	5	Pat	Rare
13890.p1	*DYRK1A*	sp	DN	HS3.1_ CGH	6	88,898,857	88,910,225	11	Dup	NA	Mat	Rare
12099.p1	*DYRK1A*	fs	DN	Sanders	10	134,924,912	134,950,336	25	Del	28	Mat	Associated with nominally significant genes in (*14*).
12681.p1	*GRIN2B*	sp	DN	Sanders	16	75,420,309	75,878,106	458	Dup	233	Mat	Rare
11704.p1	*LAMC3*	ms	DN	Sanders	X	49,068,067	49,256,601	189	Dup	25	UNK	No Case-Control Data
11704.p1	*LAMC3*	ms	DN	Sanders	X	148,650,868	148,859,781	209	Dup	38	UNK	No Case-Control Data
11704.p1	*LAMC3*	ms	DN	Sanders	X	151,979,017	152,184,470	205	Dup	38	UNK	No Case-Control Data
11704.p1	*LAMC3*	ms	DN	Sanders	X	153,071,710	153,149,289	78	Dup	25	UNK	No Case-Control Data
11660.p1	*NTNG1*	ms	DN	Sanders	2	110,210,164	110,583,308	373	Del	82	Mat	***NPHP1* Het Deletion**
11390.p1	*PTEN*	ms	DN	Sanders	2	142,554,362	142,715,864	162	Del	64	UNK	Rare
11390.p1	*PTEN*	ms	DN	Sanders	2	158,975,179	159,111,588	136	Del	38	UNK	Rare
11390.p1	*PTEN*	ms	DN	Sanders	4	62,108,499	62,215,442	107	Del	24	UNK	Rare
11390.p1	*PTEN*	ms	DN	Sanders	4	163,051,156	163,127,283	76	Del	28	UNK	Rare
11390.p1	*PTEN*	ms	DN	Sanders	5	19,495,583	19,597,571	102	Del	72	UNK	Rare
11390.p1	*PTEN*	ms	DN	Sanders	12	85,203,291	85,463,148	260	Del	59	UNK	Rare
11390.p1	*PTEN*	ms	DN	Sanders	X	134,560,640	134,758,278	198	Dup	37	Pat	No Case-Control Data
12930.p1	*SBF1*	ms	DN	Sanders	2	68,612,975	68,707,845	95	Del	30	UNK	Rare
12930.p1	*SBF1*	ms	DN	Sanders	X	49,079,323	49,246,642	167	Dup	23	UNK	No Case-Control Data

MIP detected variants				Rare and *de novo* CNVs								
Sample	**Gene**	**Mut Type**	**Mut Origin**	**Study**	**Chr**	**Start (hg18)**	**Stop (hg18)**	**Size (kb)**	**CNV Type**	**Probes/ Exons**	**CNV Origin**	**Notes**
12499.p1	*SCN1A*	ms	DN	Levy	3	81,499,246	81,978,405	479	Dup	NA	Mat	Rare
12499.p1	*SCN1A*	ms	DN	Sanders	5	1,783,637	1,925,329	142	Dup	104	Mat	Rare
12340.p1	*SCN1A*	ms	DN	Sanders	X	134,443,514	134,675,472	232	Dup	59	Mat	No Case-Control Data
12565.p1	*SETD2*	fs	DN	Sanders	X	49,068,067	49,246,642	179	Dup	24	Both	No Case-Control Data
12791.p1	*SGSM3*	ms	DN	Levy	8	156,591	362,890	206	Dup	NA	Pat	Rare
11480.p1	*TBR1*	fs	DN	Sanders	6	26,077,937	26,375,779	298	Dup	237	Pat	Rare
13796.p1	*TBR1*	fs	DN	HS3.1_ CGH	7	65,180,000	65,320,000	140	Del	NA	Mat	Rare
12851.p1	*UBE3C*	ms	DN	Sanders	8	68,839,122	69,073,722	235	Del	96	Pat	Rare
12851.p1	*UBE3C*	ms	DN	Sanders	11	133,669,828	133,851,329	182	Dup	120	Pat	Rare
11255.p1	*AP3B2*	fs	Mat	Sanders	2	86,140,891	86,363,012	222	Dup	128	Mat	Rare
11255.p1	*AP3B2*	fs	Mat	Sanders	6	88,898,873	88,923,095	24	Dup	21	Pat	Rare
11255.p1	*AP3B2*	fs	Mat	Sanders	13	19,372,771	19,443,150	70	Dup	20	Pat	Rare
14023.p1	*CHD7*	fs	Pat	HS3.1_ CGH	16	29,425,000	30,275,000	850	Dup	NA	DN	**16p11.2 Duplication**
12517.p1	*PTEN*	ns	Mat	Sanders	X	132,098,972	132,515,319	416	Dup	130	UNK	No Case-Control Data
13783.p1	*SETD2*	ns	Pat	HS3.1_ CGH	3	1,069,303	1,447,670	378	Dup	NA	UNK	Rare
12736.p1	*SETD2*	ns	Mat	Sanders	16	29,554,843	30,170,811	616	Dup	324	DN	**16p11.2 Duplication**
12637.p1	*SGSM3*	sp	Pat	Sanders	11	48,431,847	48,588,099	156	Dup	48	Pat	Rare
11494.p1	*TSPAN17*	ns	Mat	Levy	9	114,756,013	114,826,318	70	Del	NA	Mat	Rare
11494.p1	*TSPAN17*	ns	Mat	Sanders	13	40,188,779	40,223,018	34	Dup	21	Pat	Rare
11494.p1	*TSPAN17*	ns	Mat	Sanders	18	28,823,161	28,938,002	115	Del	29	UNK	Rare
11494.p1	*TSPAN17*	ns	Mat	Sanders	19	54,088,549	54,130,175	42	Del	32	Pat	Rare

Calls from Sanders et al., Levy et al., O'Roak et al. (Exome_CNV) and unpublished array CGH data (HS3.1_CGH) (*3, 47, 70*). Regions associated with genomic disorders are bold in the notes column. Regions found to intersect nominally significant genes from the developmental delay morbidity map are indicated (*14*). Abbreviations: DN-*de novo*, Mat-maternal, Pat-paternal, UNK=unknown, syn-synonymous, fs-frameshifting indel, ns-nonsense, sp-splice-site, ms-missense.

Table S17. Head circumference Z-score means and standard deviations for the SSC cohort.

Full Sample	Mean[1]	Std. Deviation
Proband	0.70	1.32
Siblings	0.68	1.21
Mothers	0.97	1.41
Fathers	0.82	1.06
CHD8 **trunc/splice (8 families)**		
Proband	2.24	0.61
Siblings	1.14	0.96
Mothers	1.07	1.56
Fathers	1.19	0.63
DYRK1A **trunc/splice (3 families)**		
Proband	-2.72	1.07
Siblings	0.53	1.22
Mothers	0.91	0.81
Fathers	0.67	0.34

[1]Values are standardized head circumference Z-scores calculated using norms established by Roche (*31*) to account for age and gender. As such, the reference population mean is 0 and standard deviation is 1.

Table S18. Other *de novo* variants identified in published SSC exome sequencing studies intersecting with Tables S11 and S15 probands.

MIP detected variants				Other exome *de novo* mutations	
Proband	**Gene**	**Mut Type**	**Mut Origin**	**Exome Study**	**DNV**
12688.p1	*ADCY5*	ms	DN	Iossifov	none
11653.p1	*ADCY5*	ms	DN	O'Roak	*ST3GAL3*:ms
12130.p1	*ADNP*	fs	DN	O'Roak	none
13447.p1	*ARID1B*	fs	DN	O'Roak	none
11654.p1	*CHD8*	sp	DN	Sanders	none
13844.p1	*CHD8*	ns	DN	O'Roak	*CUBN*:ns, *TRAPPC8*:syn
12752.p1	*CHD8*	fs	DN	O'Roak	*ETFB*:ns, *IQGAP2*:ns
12703.p1	*CTNNB1*	ms	DN	O'Roak	*INPP5B*:ms
12099.p1	*DYRK1A*	fs	DN	Sanders	none
13890.p1	*DYRK1A*	sp	DN	O'Roak	*MSH6*:ms
13552.p1	*DYRK1A*	fs	DN	Iossifov	*KIF18A*:ms, *ALMS1*:ms
11691.p1	*GRIN2B*	fs	DN	O'Roak	*AMY2B*:ms
12681.p1	*GRIN2B*	sp	DN	O'Roak	*EPHB6*:syn
11666.p1	*LAMC3*	ms	DN	O'Roak	none
12532.p1	*NTNG1*	ms	DN	O'Roak	*NAA40*:ms, *FRYL*:ms, *C9orf30*:sym
11660.p1	*NTNG1*	ms	DN	O'Roak	*CACNA1A*:syn
11390.p1	*PTEN*	ms	DN	O'Roak	none
12930.p1	*SBF1*	ms	DN	Iossifov	*STK11*:ms
13793.p1	*SBF1*	ms	DN	O'Roak	*PCDHB4*:ms
12499.p1	*SCN1A*	ms	DN	O'Roak	*SYNE1*:ms, *PLCD1*:syn
12340.p1	*SCN1A*	ms	DN	Sanders	*PPM1D*:ns, *BCORL1*:ms
12565.p1	*SETD2*	fs	DN	O'Roak	none
11523.p1	*SGSM3*	ms	DN	O'Roak	none
12335.p1	*TBL1XR1*	ms	DN	O'Roak	*STK36*:ms
11480.p1	*TBR1*	fs	DN	O'Roak	*IRF2BPL*:ms, *SLC40A1*:syn
13796.p1	*TBR1*	fs	DN	Sanders	none
11006.p1	*UBE3C*	ms	DN	O'Roak	*PKD1L3*:syn, *NDST4*:syn
12851.p1	*UBE3C*	ms	DN	Iossifov	*SLC25A29*:ms, *APOB*:syn, *HMGXB3*:ms
12652.p1	*ADCY5*	fs	Pat	Iossifov	*SPATA13*:fs, *SWAP70*:ms, *CDC34*:syn
11298.p1	*LAMC3*	ns	Mat	Sanders	*SLC6A13*:ms
11797.p1	*MBD5*	sp	Mat	Sanders	*C2orf42*:ms
12736.p1	*SETD2*	ns	Mat	Sanders	*RPRD1A*:syn
13625.p1	*SGSM3*	sp	Pat	Sanders	none
12637.p1	*SGSM3*	sp	Pat	Iossifov	*RHOT2*:ms, *RNPEPL1*:ms

Calls from O'Roak et al., Sanders et al., and Iossifov et al. (*3, 4, 6*).
Abbreviations: DNV-*de novo* variant, DN-*de novo*, Mat-maternal, Pat-paternal, syn-synonymous, fs-frameshifting indel, ns-nonsense, sp-splice-site, ms-missense.

References and Notes

1. G. V. Kryukov, A. Shpunt, J. A. Stamatoyannopoulos, S. R. Sunyaev, Power of deep, all-exon resequencing for discovery of human trait genes. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3871 (2009). doi:10.1073/pnas.0812824106 Medline

2. B. J. O'Roak *et al*., Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* **43**, 585 (2011). doi:10.1038/ng.835 Medline

3. B. J. O'Roak *et al*., Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246 (2012). doi:10.1038/nature10989 Medline

4. S. J. Sanders *et al*., De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237 (2012). doi:10.1038/nature10945 Medline

5. B. M. Neale *et al*., Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242 (2012). doi:10.1038/nature11011 Medline

6. I. Iossifov *et al*., De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285 (2012). doi:10.1016/j.neuron.2012.04.009 Medline

7. E. H. Turner, C. Lee, S. B. Ng, D. A. Nickerson, J. Shendure, Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat. Methods* **6**, 315 (2009). doi:10.1038/nmeth.f.248 Medline

8. G. J. Porreca *et al*., Multiplex amplification of large sets of human exons. *Nat. Methods* **4**, 931 (2007). doi:10.1038/nmeth1110 Medline

9. S. Krishnakumar *et al*., A comprehensive assay for targeted multiplex amplification of human DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9296 (2008). doi:10.1073/pnas.0803240105 Medline

10. See supplementary text on *Science* Online.

11. G. D. Fischbach, C. Lord, The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron* **68**, 192 (2010). doi:10.1016/j.neuron.2010.10.006 Medline

12. Materials and methods are available as supplementary material on *Science* Online.

13. C. Betancur, Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42 (2011). doi:10.1016/j.brainres.2010.11.078 Medline

14. G. M. Cooper *et al*., A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43**, 838 (2011). doi:10.1038/ng.909 Medline

15. M. Lynch, Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 961 (2010). doi:10.1073/pnas.0912629107 Medline

16. A. Kong *et al*., Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471 (2012). doi:10.1038/nature11396 Medline

17. C. A. Williams, A. Dagli, A. Battaglia, Genetic disorders associated with macrocephaly. *Am. J. Med. Genet. A.* **146A**, 2023 (2008). doi:10.1002/ajmg.a.32434 Medline

18. R. S. Møller *et al*., Truncation of the Down syndrome candidate gene *DYRK1A* in two unrelated patients with microcephaly. *Am. J. Hum. Genet.* **82**, 1165 (2008). doi:10.1016/j.ajhg.2008.03.001 Medline

19. B. W. van Bon *et al*., Intragenic deletion in *DYRK1A* leads to mental retardation and primary microcephaly. *Clin. Genet.* **79**, 296 (2011). doi:10.1111/j.1399-0004.2010.01544.x Medline

20. F. Guedj *et al*., DYRK1A: A master regulatory protein controlling brain growth. *Neurobiol. Dis.* **46**, 190 (2012). doi:10.1016/j.nbd.2012.01.007 Medline

21. M. E. Talkowski *et al*., Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* **149**, 525 (2012). doi:10.1016/j.cell.2012.03.028 Medline

22. J. Zhou, L. F. Parada, PTEN signaling in autism spectrum disorders. *Curr. Opin. Neurobiol.* **22**, 873 (2012). doi:10.1016/j.conb.2012.05.004 Medline

23. I. Letunic, T. Doerks, P. Bork, SMART 7: Recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40**(Database issue), D302 (2012). doi:10.1093/nar/gkr931 Medline

24. N. Sugimoto, S. Nakano, M. Yoneyama, K. Honda, Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res.* **24**, 4501 (1996). doi:10.1093/nar/24.22.4501 Medline

25. Y. Wang *et al*., Analysis of molecular inversion probe performance for allele copy number determination. *Genome Biol.* **8**, R246 (2007). doi:10.1186/gb-2007-8-11-r246 Medline

26. N. Krumm *et al*., Copy number variation detection and genotyping from exome sequence data. *Genome Res.* **22**, 1525 (2012). doi:10.1101/gr.138115.112 Medline

27. J. Ren *et al*., DOG 1.0: Illustrator of protein domain structures. *Cell Res.* **19**, 271 (2009). doi:10.1038/cr.2009.6 Medline

28. W.-H. Li, *Molecular Evolution* (Sinauer Associates, Sunderland, MA, 1997).

29. J. A. Tennessen *et al*., Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64 (2012). doi:10.1126/science.1219240 Medline

30. J. Q. Chen *et al*., Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. *Mol. Biol. Evol.* **26**, 1523 (2009). doi:10.1093/molbev/msp063 Medline

31. A. F. Roche, D. Mukherjee, S. M. Guo, W. M. Moore, Head circumference reference data: Birth to 18 years. *Pediatrics* **79**, 706 (1987). Medline

32. T. Hothorn, K. Hornik, M. A. V. van de Wiel, A. Zeileis, Implementing a class of permutation tests: The coin package. *J. Stat. Softw.* **28**, 1 (2008).

33. J. Deng *et al*., Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* **27**, 353 (2009). doi:10.1038/nbt.1530 Medline

34. A. Amiri *et al*., Pten deletion in adult hippocampal neural stem/progenitor cells causes cellular abnormalities and alters neurogenesis. *J. Neurosci.* **32**, 5880 (2012). doi:10.1523/JNEUROSCI.5462-11.2012 Medline

35. M. Nishiyama, A. I. Skoultchi, K. I. Nakayama, Histone H1 recruitment by CHD8 is essential for suppression of the Wnt-β-catenin signaling pathway. *Mol. Cell. Biol.* **32**, 501 (2012). doi:10.1128/MCB.06409-11 Medline

36. M. Nishiyama *et al*., CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. *Nat. Cell Biol.* **11**, 172 (2009). doi:10.1038/ncb1831 Medline

37. B. A. Thompson, V. Tremblay, G. Lin, D. A. Bochar, CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. *Mol. Cell. Biol.* **28**, 3894 (2008). doi:10.1128/MCB.00322-08 Medline

38. F. Zahir *et al*., Novel deletions of 14q11.2 associated with developmental delay, cognitive impairment and similar minor anomalies in three children. *J. Med. Genet.* **44**, 556 (2007). doi:10.1136/jmg.2007.050823 Medline

39. F. Bedogni *et al*., Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13129 (2010). doi:10.1073/pnas.1002285107 Medline

40. S. Endele *et al*., Mutations in *GRIN2A* and *GRIN2B* encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat. Genet.* **42**, 1021 (2010). doi:10.1038/ng.677 Medline

41. R. A. Myers *et al*., A population genetic approach to mapping neurological disorder genes using deep resequencing. *PLoS Genet.* **7**, e1001318 (2011). doi:10.1371/journal.pgen.1001318 Medline

42. V. Fotaki *et al*., Dyrk1A haploinsufficiency affects viability and causes developmental delay and abnormal brain morphology in mice. *Mol. Cell. Biol.* **22**, 6636 (2002). doi:10.1128/MCB.22.18.6636-6647.2002 Medline

43. R. E. Green *et al*., A draft sequence of the Neandertal genome. *Science* **328**, 710 (2010). doi:10.1126/science.1188021 Medline

44. J. Y. Hong *et al*., Down's-syndrome-related kinase Dyrk1A modulates the p120-catenin-Kaiso trajectory of the Wnt signaling pathway. *J. Cell Sci.* **125**, 561 (2012). doi:10.1242/jcs.086173 Medline

45. J. Li, C. Y. Wang, TBL1-TBLR1 and beta-catenin recruit each other to Wnt target-gene promoter for transcription activation and oncogenesis. *Nat. Cell Biol.* **10**, 160 (2008). doi:10.1038/ncb1684 Medline

46. H. K. Choi *et al*., Reversible SUMOylation of TBL1-TBLR1 regulates β-catenin-mediated Wnt signaling. *Mol. Cell* **43**, 203 (2011). doi:10.1016/j.molcel.2011.05.027 Medline

47. S. J. Sanders *et al*., Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863 (2011). doi:10.1016/j.neuron.2011.05.002 Medline

48. D. Horn *et al*., Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. *Hum. Mutat.* **31**, E1851 (2010). doi:10.1002/humu.21362 Medline

49. F. F. Hamdan *et al*., De novo mutations in *FOXP1* in cases with intellectual disability, autism, and language impairment. *Am. J. Hum. Genet.* **87**, 671 (2010). doi:10.1016/j.ajhg.2010.09.017 Medline

50. L. Feuk *et al*., Absence of a paternally inherited *FOXP2* gene in developmental verbal dyspraxia. *Am. J. Hum. Genet.* **79**, 965 (2006). doi:10.1086/508902 Medline

51. C. S. Lai *et al*., The SPCH1 region on human 7q31: Genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *Am. J. Hum. Genet.* **67**, 357 (2000). doi:10.1086/303011 Medline

52. X. Wei *et al*., Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat. Genet.* **43**, 442 (2011). doi:10.1038/ng.810 Medline

53. P. Awadalla *et al*., Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am. J. Hum. Genet.* **87**, 316 (2010). doi:10.1016/j.ajhg.2010.07.019 Medline

54. P. Jia *et al*., Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia. *PLOS Comput. Biol.* **8**, e1002587 (2012). doi:10.1371/journal.pcbi.1002587 Medline

55. J. Tarabeux *et al*., Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. *Transl. Psychiatry* **1**, e55 (2011). doi:10.1038/tp.2011.52 Medline

56. T. Barak *et al*., Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat. Genet.* **43**, 590 (2011). doi:10.1038/ng.836 Medline

57. C. Lossin, A catalog of SCN1A variants. *Brain Dev.* **31**, 114 (2009). doi:10.1016/j.braindev.2008.07.011 Medline

58. K. S. Kim *et al*., Adenylyl cyclase type 5 (AC5) is an essential mediator of morphine action. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3908 (2006). doi:10.1073/pnas.0508812103 Medline

59. A. Pinhasov *et al*., Activity-dependent neuroprotective protein: A novel gene essential for brain formation. *Brain Res. Dev. Brain Res.* **144**, 83 (2003). doi:10.1016/S0165-3806(03)00162-7 Medline

60. J. M. Hill *et al*., Blockage of VIP during mouse embryogenesis modifies adult behavior and results in permanent changes in brain chemistry. *J. Mol. Neurosci.* **31**, 183 (2007). Medline

61. M. J. Wat *et al*., Recurrent microdeletions of 15q25.2 are associated with increased risk of congenital diaphragmatic hernia, cognitive deficits and possibly Diamond–Blackfan anaemia. *J. Med. Genet.* **47**, 777 (2010). doi:10.1136/jmg.2009.075903 Medline

62. G. W. Santen *et al*., Mutations in SWI/SNF chromatin remodeling complex gene *ARID1B* cause Coffin-Siris syndrome. *Nat. Genet.* **44**, 379 (2012). doi:10.1038/ng.2217 Medline

63. J. Hoyer *et al*., Haploinsufficiency of *ARID1B*, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am. J. Hum. Genet.* **90**, 565 (2012). doi:10.1016/j.ajhg.2012.02.007 Medline

64. C. Halgren *et al*., Corpus callosum abnormalities, intellectual disability, speech impairment, and autism in patients with haploinsufficiency of *ARID1B*. *Clin. Genet.* **82**, 248 (2012). doi:10.1111/j.1399-0004.2011.01755.x Medline

65. A. S. Nord *et al*., Reduced transcript expression of genes affected by inherited and de novo CNVs in autism. *Eur. J. Hum. Genet.* **19**, 727 (2011). doi:10.1038/ejhg.2011.24 Medline

66. M. Kishi, Y. A. Pan, J. G. Crump, J. R. Sanes, Mammalian SAD kinases are required for neuronal polarization. *Science* **307**, 929 (2005). doi:10.1126/science.1107403 Medline

67. T. Batsukh *et al*., CHD8 interacts with CHD7, a protein which is mutated in CHARGE syndrome. *Hum. Mol. Genet.* **19**, 2858 (2010). doi:10.1093/hmg/ddq189 Medline

68. M. C. Jongmans *et al*., CHARGE syndrome: The phenotypic spectrum of mutations in the CHD7 gene. *J. Med. Genet.* **43**, 306 (2006). doi:10.1136/jmg.2005.036061 Medline

69. L. Feng, N. S. Allen, S. Simo, J. A. Cooper, Cullin 5 regulates Dab1 protein levels and neuron positioning during cortical development. *Genes Dev.* **21**, 2717 (2007). doi:10.1101/gad.1604207 Medline

70. D. Levy *et al*., Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70**, 886 (2011). doi:10.1016/j.neuron.2011.05.015 Medline

71. H. M. El-Tahir *et al*., Expression of hepatoma-derived growth factor family members in the adult central nervous system. *BMC Neurosci.* **7**, 6 (2006). doi:10.1186/1471-2202-7-6 Medline

72. B. Felder *et al*., *FARP2*, *HDLBP* and *PASK* are downregulated in a patient with autism and 2q37.3 deletion syndrome. *Am. J. Med. Genet. A.* **149A**, 952 (2009). doi:10.1002/ajmg.a.32779 Medline

73. M. E. Talkowski *et al*., Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. *Am. J. Hum. Genet.* **89**, 551 (2011). doi:10.1016/j.ajhg.2011.09.011 Medline

74. S. R. Williams *et al*., Haploinsufficiency of *MBD5* associated with a syndrome involving microcephaly, intellectual disabilities, severe speech impairment, and seizures. *Eur. J. Hum. Genet.* **18**, 436 (2010). doi:10.1038/ejhg.2009.199 Medline

75. A. Millson *et al*., Chromosomal loss of 3q26.3-3q26.32, involving a partial neuroligin 1 deletion, identified by genomic microarray in a child with microcephaly, seizure disorder, and severe intellectual disability. *Am. J. Med. Genet. A.* (2011). 10.1002/ajmg.a.34349 Medline

76. J. T. Glessner *et al*., Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569 (2009). doi:10.1038/nature07953 Medline

77. N. Van der Aa, G. Vandeweyer, R. F. Kooy, A boy with mental retardation, obesity and hypertrichosis caused by a microdeletion of 19p13.12. *Eur. J. Med. Genet.* **53**, 291 (2010). doi:10.1016/j.ejmg.2010.05.006 Medline

78. Y. H. Chen, M. T. Tsai, C. K. Shaw, C. H. Chen, Mutation analysis of the human NR4A2 gene, an essential gene for midbrain dopaminergic neurogenesis, in schizophrenic patients. *Am. J. Med. Genet.* **105**, 753 (2001). doi:10.1002/ajmg.10036 Medline

79. W. D. Le *et al*., Mutations in *NR4A2* associated with familial Parkinson disease. *Nat. Genet.* **33**, 85 (2003). doi:10.1038/ng1066 Medline

80. I. Borg *et al*., Disruption of *Netrin G1* by a balanced chromosome translocation in a girl with Rett syndrome. *Eur. J. Hum. Genet.* **13**, 921 (2005). doi:10.1038/sj.ejhg.5201429 Medline

81. J. A. Briant *et al*., Evidence for association of two variants of the nociceptin/orphanin FQ receptor gene *OPRL1* with vulnerability to develop opiate addiction in Caucasians. *Psychiatr. Genet.* **20**, 65 (2010). doi:10.1097/YPG.0b013e32833511f6 Medline

82. S. R. Williams *et al*., Haploinsufficiency of *HDAC4* causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *Am. J. Hum. Genet.* **87**, 219 (2010). doi:10.1016/j.ajhg.2010.07.011 Medline

83. B. Borroni *et al*., Atypical presentation of a novel *Presenilin 1 R377W* mutation: Sporadic, late-onset Alzheimer disease with epilepsy and frontotemporal atrophy. *Neurol. Sci.* **33**, 375 (2012). doi:10.1007/s10072-011-0714-1 Medline

84. A. Goffin, L. H. Hoefsloot, E. Bosgoed, A. Swillen, J. P. Fryns, PTEN mutation in a family with Cowden syndrome and autism. *Am. J. Med. Genet.* **105**, 521 (2001). doi:10.1002/ajmg.1477 Medline

85. G. E. Herman *et al*., Increasing knowledge of PTEN germline mutations: Two additional patients with autism and macrocephaly. *Am. J. Med. Genet. A.* **143**, 589 (2007). doi:10.1002/ajmg.a.31619 Medline

86. E. M. Arch *et al*., Deletion of PTEN in a patient with Bannayan-Riley-Ruvalcaba syndrome suggests allelism with Cowden disease. *Am. J. Med. Genet.* **71**, 489 (1997). doi:10.1002/(SICI)1096-8628(19970905)71:4<489::AID-AJMG24>3.0.CO;2-B Medline

87. E. Trivier *et al*., Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome. *Nature* **384**, 567 (1996). doi:10.1038/384567a0 Medline

88. K. Merienne *et al*., A missense mutation in *RPS6KA3* (*RSK2*) responsible for non-specific mental retardation. *Nat. Genet.* **22**, 13 (1999). doi:10.1038/8719 Medline

89. M. Field *et al*., Mutations in the *RSK2(RPS6KA3)* gene cause Coffin-Lowry syndrome and nonsyndromic X-linked mental retardation. *Clin. Genet.* **70**, 509 (2006). doi:10.1111/j.1399-0004.2006.00723.x Medline

90. K. Taniue, T. Oda, T. Hayashi, M. Okuno, T. Akiyama, A member of the ETS family, EHF, and the ATPase RUVBL1 inhibit p53-mediated apoptosis. *EMBO Rep.* **12**, 682 (2011). doi:10.1038/embor.2011.81 Medline

91. Y. Feng, N. Lee, E. R. Fearon, TIP49 regulates beta-catenin-mediated neoplastic transformation and T-cell factor target gene induction via effects on chromatin remodeling. *Cancer Res.* **63**, 8726 (2003). Medline

92. A. V. Budanov, M. Karin, p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mTOR signaling. *Cell* **134**, 451 (2008). doi:10.1016/j.cell.2008.06.028 Medline

93. K. Buysse *et al*., Delineation of a critical region on chromosome 18 for the del(18)(q12.2q21.1) syndrome. *Am. J. Med. Genet. A.* **146A**, 1330 (2008). doi:10.1002/ajmg.a.32267 Medline

94. A. Hoischen *et al*., De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.* **42**, 483 (2010). doi:10.1038/ng.581 Medline

95. P. Xie *et al*., Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. *Cell. Signal.* **20**, 1671 (2008). doi:10.1016/j.cellsig.2008.05.012 Medline

96. H. Yang, T. Sasaki, S. Minoshima, N. Shimizu, Identification of three novel proteins (SGSM1, 2, 3) which modulate small G protein (RAP and RAB)-mediated signaling pathway. *Genomics* **90**, 249 (2007). doi:10.1016/j.ygeno.2007.03.013 Medline

97. R. F. Hevner *et al*., Tbr1 regulates differentiation of the preplate and layer 6. *Neuron* **29**, 353 (2001). doi:10.1016/S0896-6273(01)00211-2 Medline

98. B. Z. Yang, S. Han, H. R. Kranzler, L. A. Farrer, J. Gelernter, A genomewide linkage scan of cocaine dependence and major depressive episode in two populations. *Neuropsychopharmacology* **36**, 2422 (2011). doi:10.1038/npp.2011.122 Medline

99. X. Huang, C. Langelotz, B. K. Hetfeld-Pechoc, W. Schwenk, W. Dubiel, The COP9 signalosome mediates beta-catenin degradation by deneddylation and blocks adenomatous polyposis coli destruction via USP15. *J. Mol. Biol.* **391**, 691 (2009). doi:10.1016/j.jmb.2009.06.066 Medline

100. C. Tian *et al*., KRAB-type zinc-finger protein Apak specifically regulates p53-dependent apoptosis. *Nat. Cell Biol.* **11**, 580 (2009). doi:10.1038/ncb1864 Medline