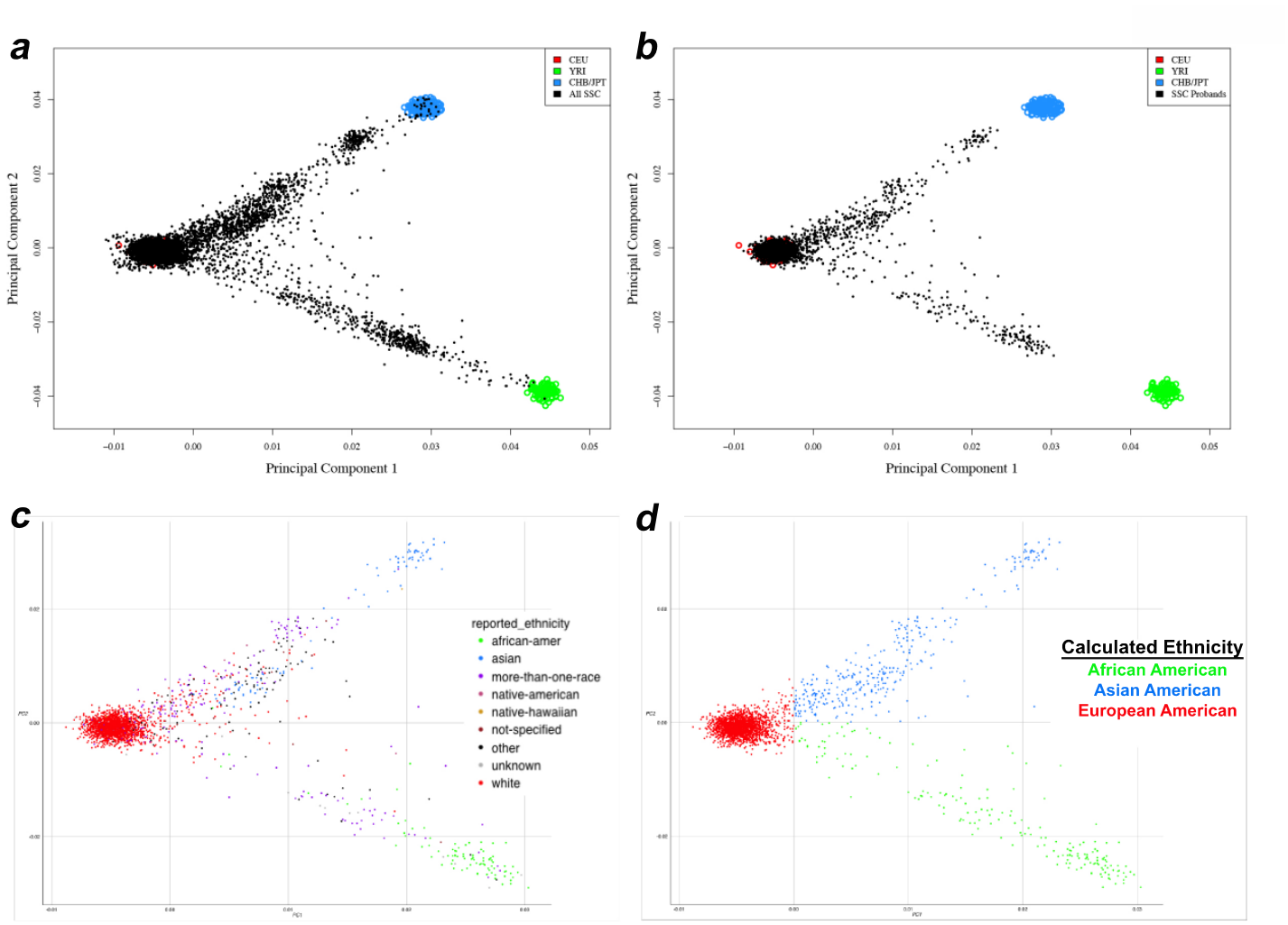
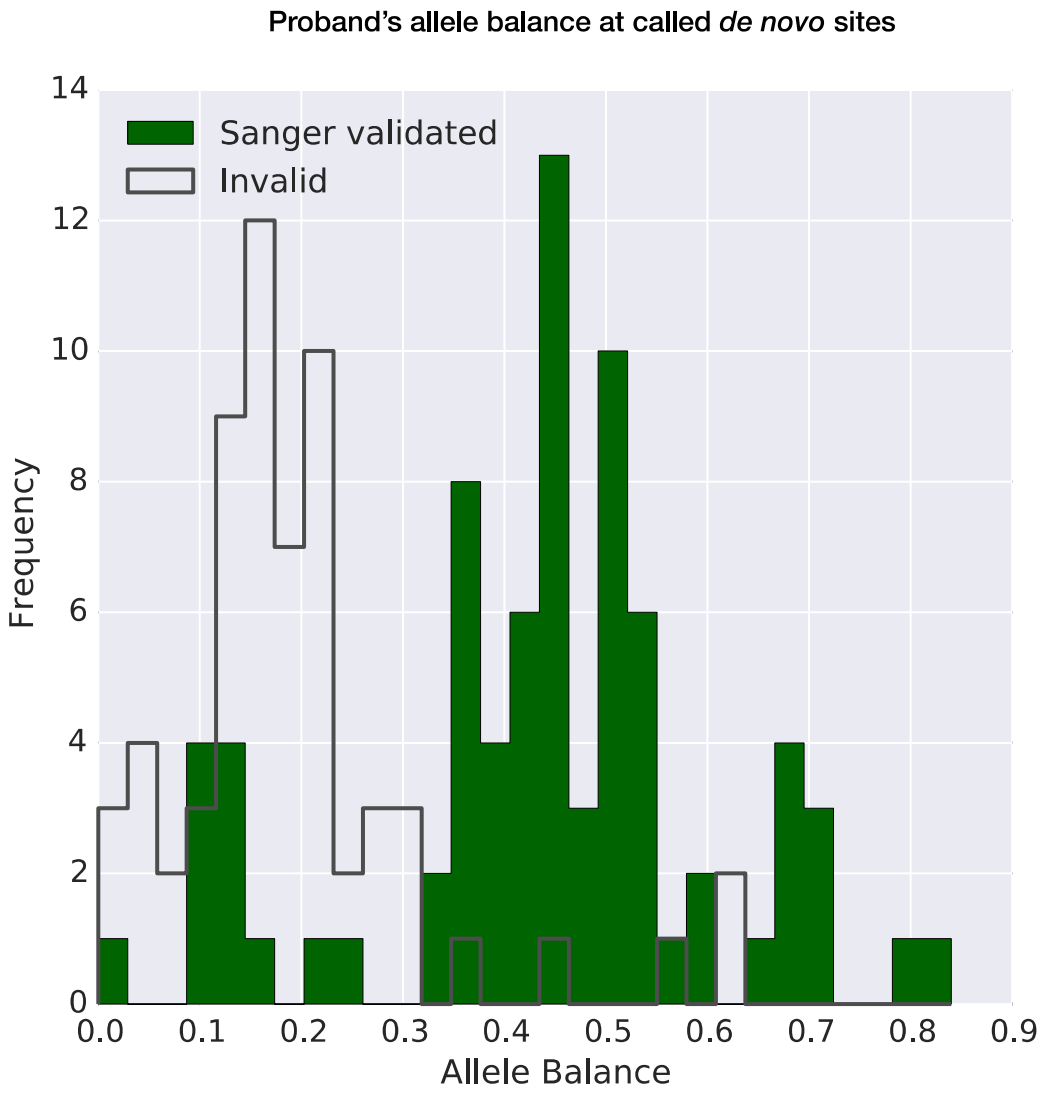


Supplementary Figure 1

PCA for ancestry in SNV data.

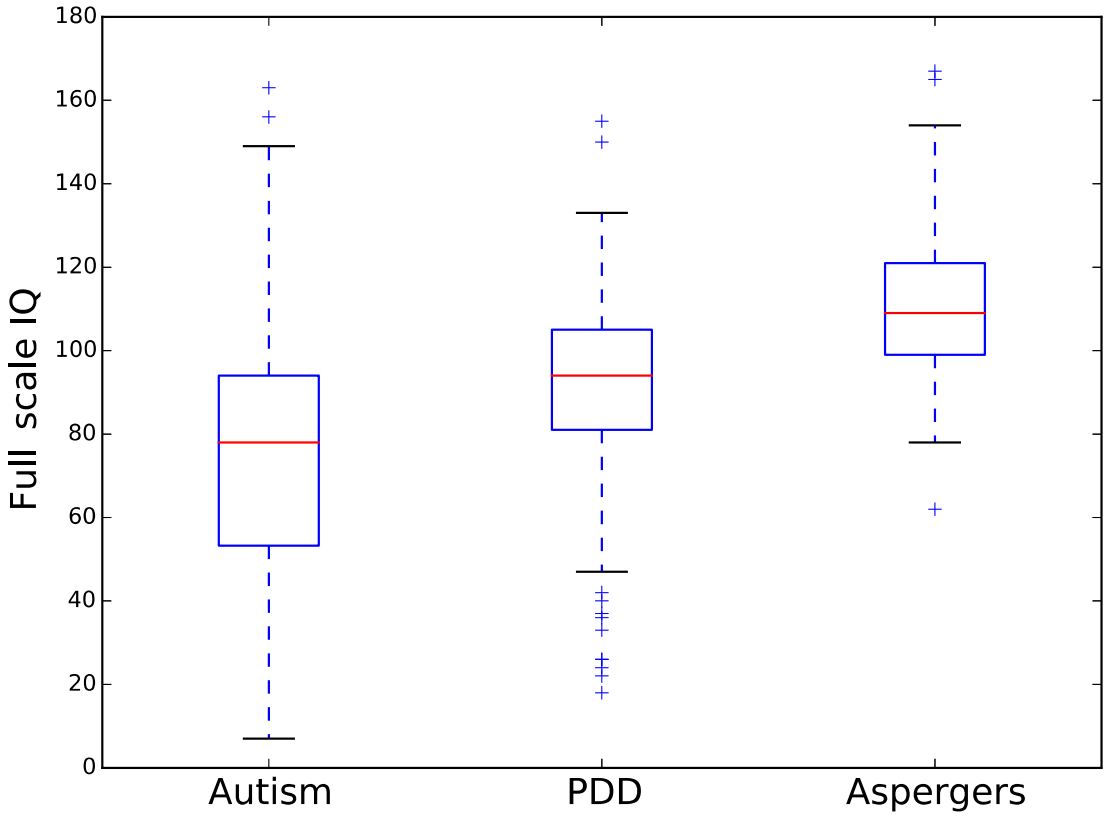
(**a**) EIGENSTRAT principal-component analysis (PCA) of SNV genotype data on all samples. (**b**) PCA of only proband SNV genotype data. (**c**) PCA of SNV genotype data with colors indicating ethnicity specified in the Simons Foundation Autism Research Initiative (SFARI) Base. (**d**) PCA of SNV genotype data with colors indicating calculated ethnicity.



Supplementary Figure 2

Allele balance in probands at identified *de novo* variant sites.

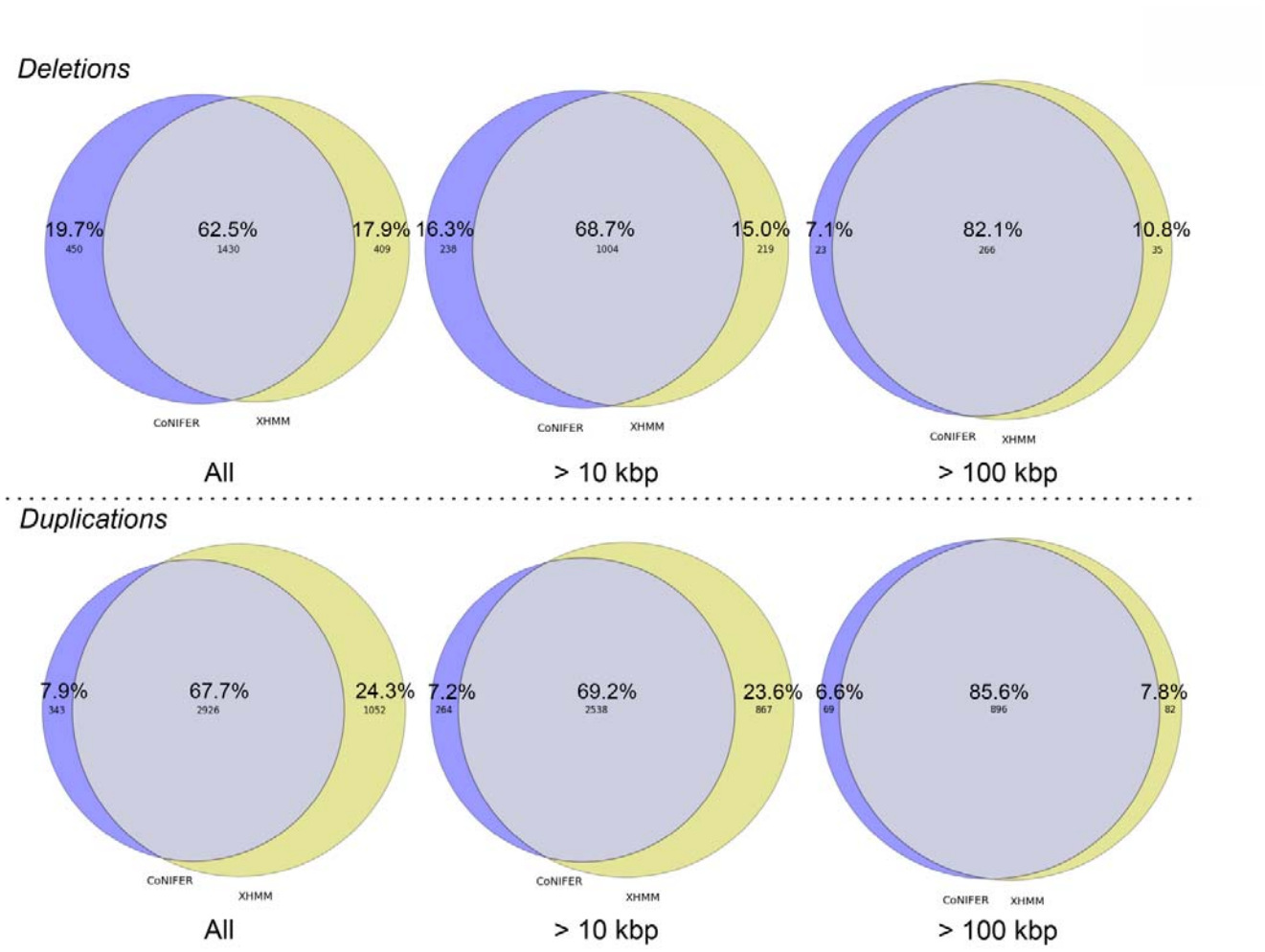
Histogram of proband's allele balance at 141 sites that we attempted to validate. Green, Sanger validation was successful (the site is a true *de novo* variant); black outline, the site failed to validate as a *de novo* variant.



Supplementary Figure 3

Distribution of IQ for autism, PDD and Asperger's clinical impressions.

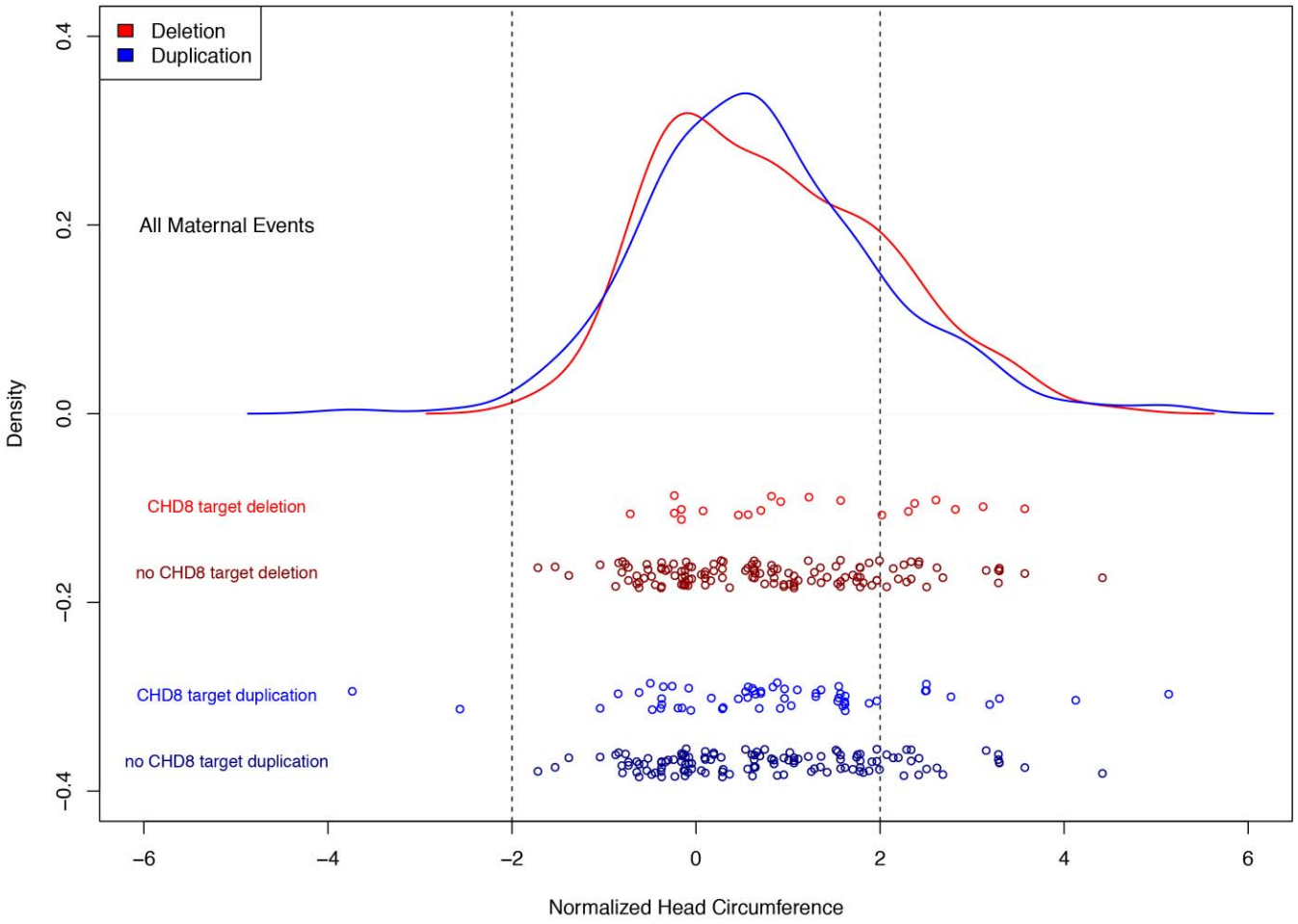
Pearson correlation of IQ and all three categories: $r^2 = 0.18$, $P < 1 \times 10^{-10}$.



Supplementary Figure 4

CNV validation rates for CoNIFER and XHMM.

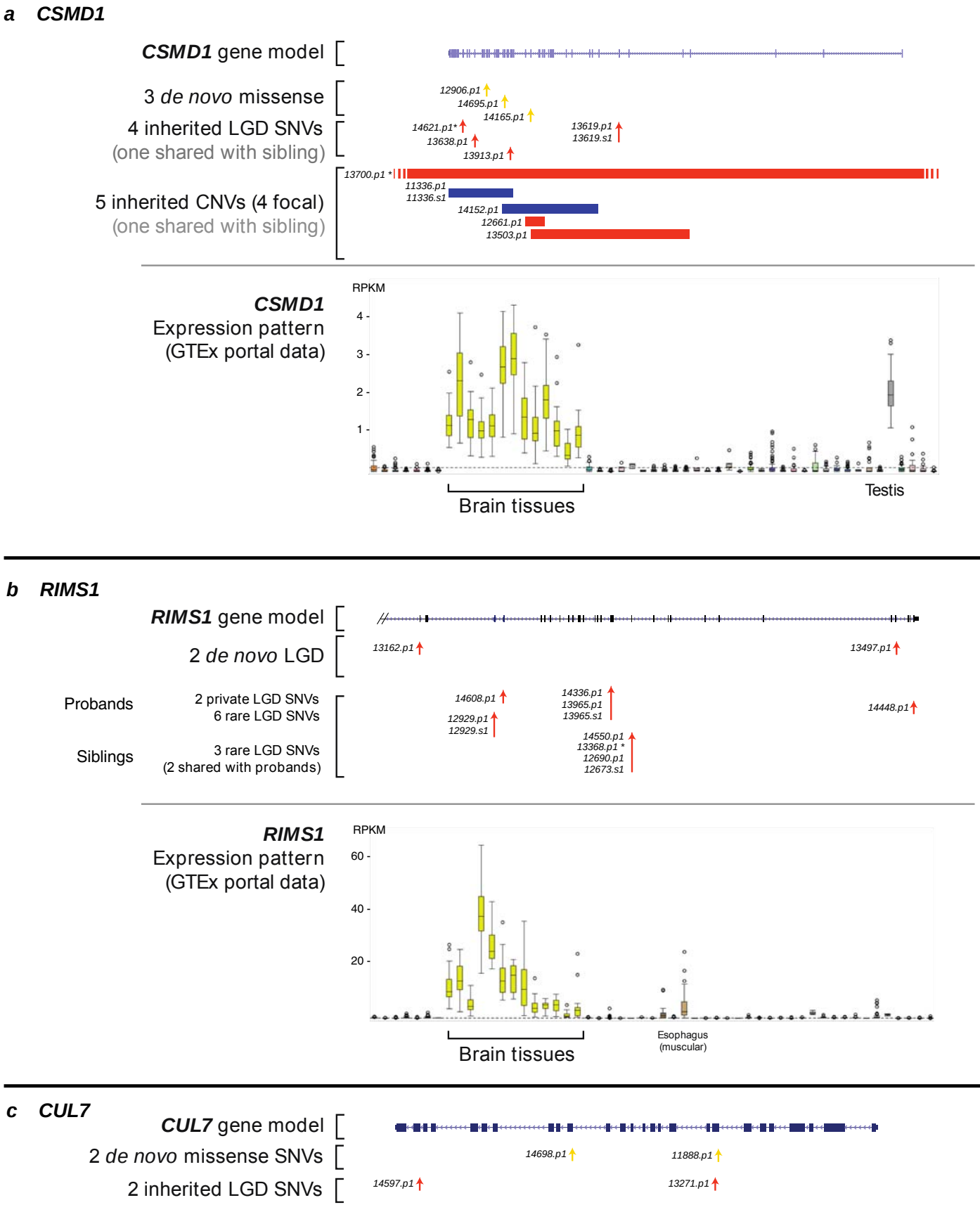
Shown are Venn diagrams with blue indicating events (deletions/duplications), at varying size ranges, called by CoNIFER and yellow indicating events called by XHMM. These are events validated by SNP microarray data.



Supplementary Figure 5

Head size distribution for maternal CNV carriers.

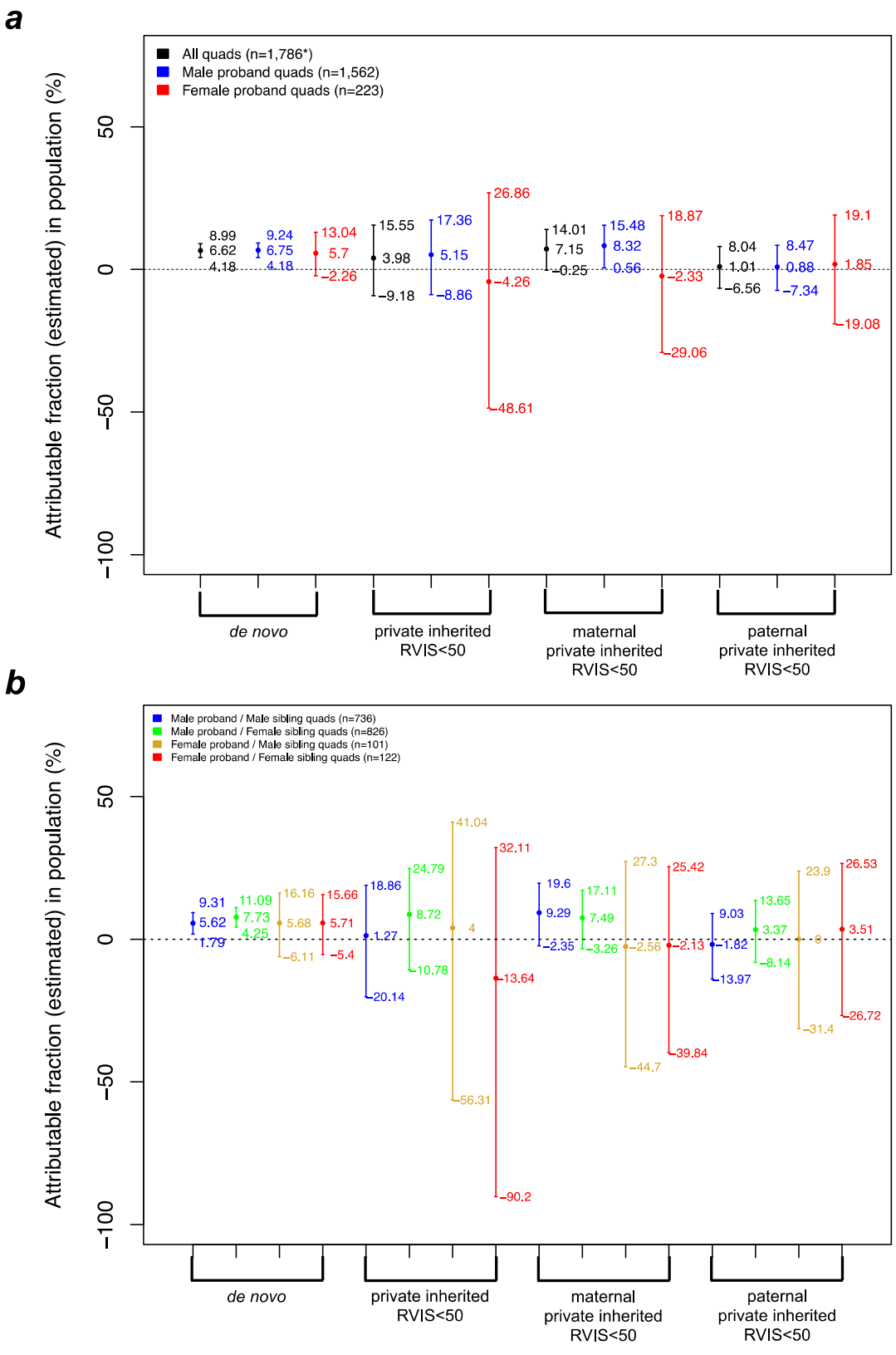
The top portion of the diagram shows density distributions of normalized head size per individual for maternally inherited deletions (red) and duplications (blue). Below are jitter plots with each point indicating an individual sample and the respective normalized head size. These are shown for individuals with deletions and duplications carrying a CHD8 target gene and also those who have deletions and duplications that do not contain a CHD8 target gene. A normalized head size of less than or equal to –2 is considered microcephaly and greater than or equal to 2 is macrocephaly.


a CSMD1
CSMD1 gene model
3 de novo missense
4 inherited LGD SNVs
(one shared with sibling)
5 inherited CNVs (4 focal)
(one shared with sibling)
12906.p1
14695.p1
14165.p1
14621.p1*
13638.p1
13913.p1
13619.p1
13619.s1
13700.p1 *
11336.p1
11336.s1
14152.p1
12661.p1
13503.p1
CSMD1
Expression pattern
(GTEx portal data)
RPKM
4
3
2
1
Brain tissues
Testis
b RIMS1
RIMS1 gene model
2 de novo LGD
Probands
2 private LGD SNVs
6 rare LGD SNVs
Siblings
3 rare LGD SNVs
(2 shared with probands)
13162.p1
13497.p1
14608.p1
12929.p1
12929.s1
14336.p1
13965.p1
13965.s1
14448.p1
14550.p1
13368.p1 *
12690.p1
12673.s1
RIMS1
Expression pattern
(GTEx portal data)
RPKM
60
40
20
Brain tissues
Esophagus
(muscular)
c CUL7
CUL7 gene model
2 de novo missense SNVs
2 inherited LGD SNVs
14698.p1
11888.p1
14597.p1
13271.p1

Supplementary Figure 6

Convergence of *de novo* and inherited mutations.

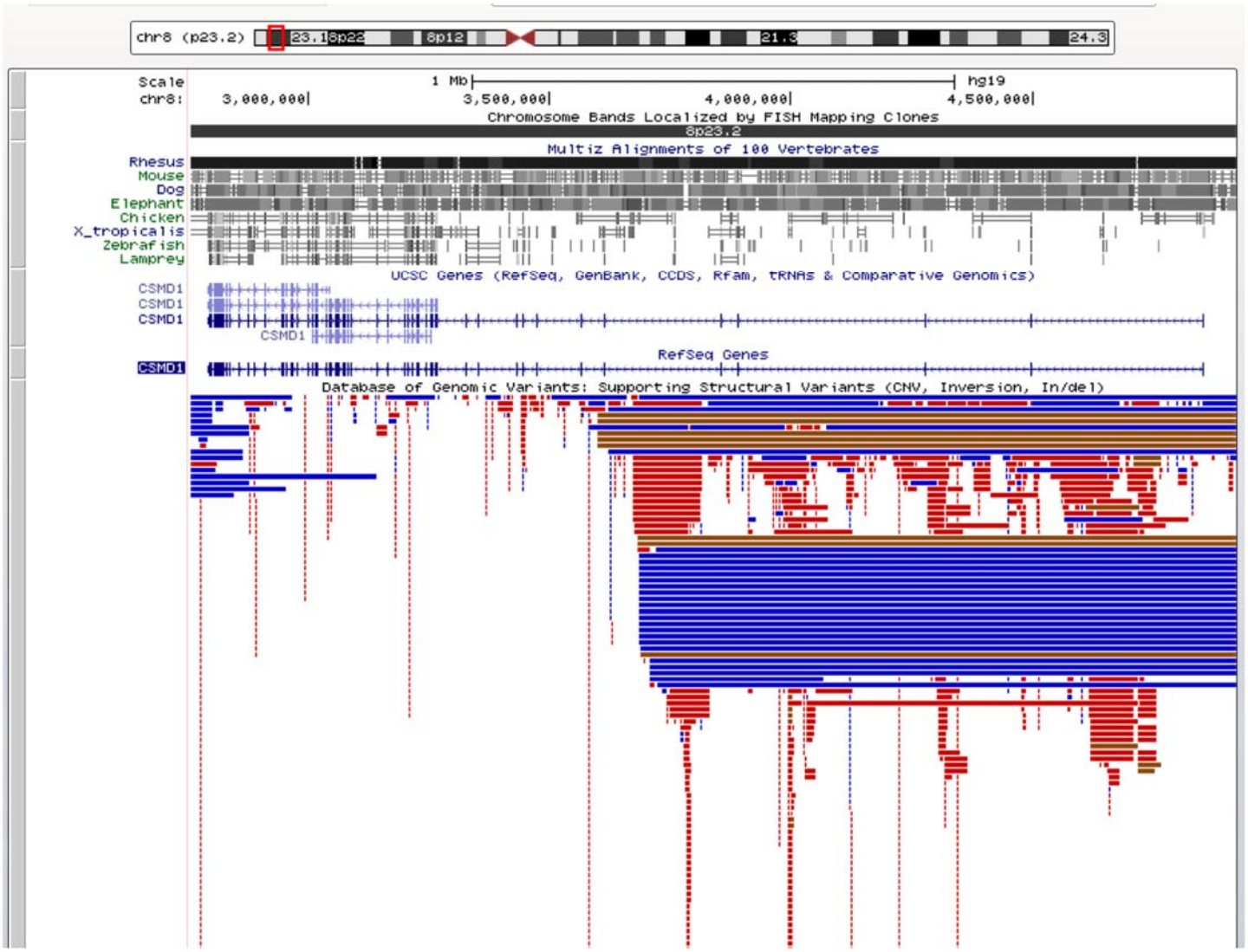
Examples of genes with an excess of disruptive mutations in probands include (**a**) *CSMD1*, (**b**) *RIMS1* and (**c**) *CUL7*. Displayed for each gene is a RefSeq gene model (larger ticks are exons), validated *de novo* and private LGD mutations (red arrows) and missense (yellow arrows); disruptive CNV deletions (red) and duplications (blue) are compared for proband (p1) and sibling (s1). *CSMD1* and *RIMS1* show highly brain-specific expression patterns (adapted from the GTEx Consortium online portal: http://gtexportal.org/). Proband IDs with asterisks are part of trios, not quads.



Supplementary Figure 7

Population attributable risk estimates.

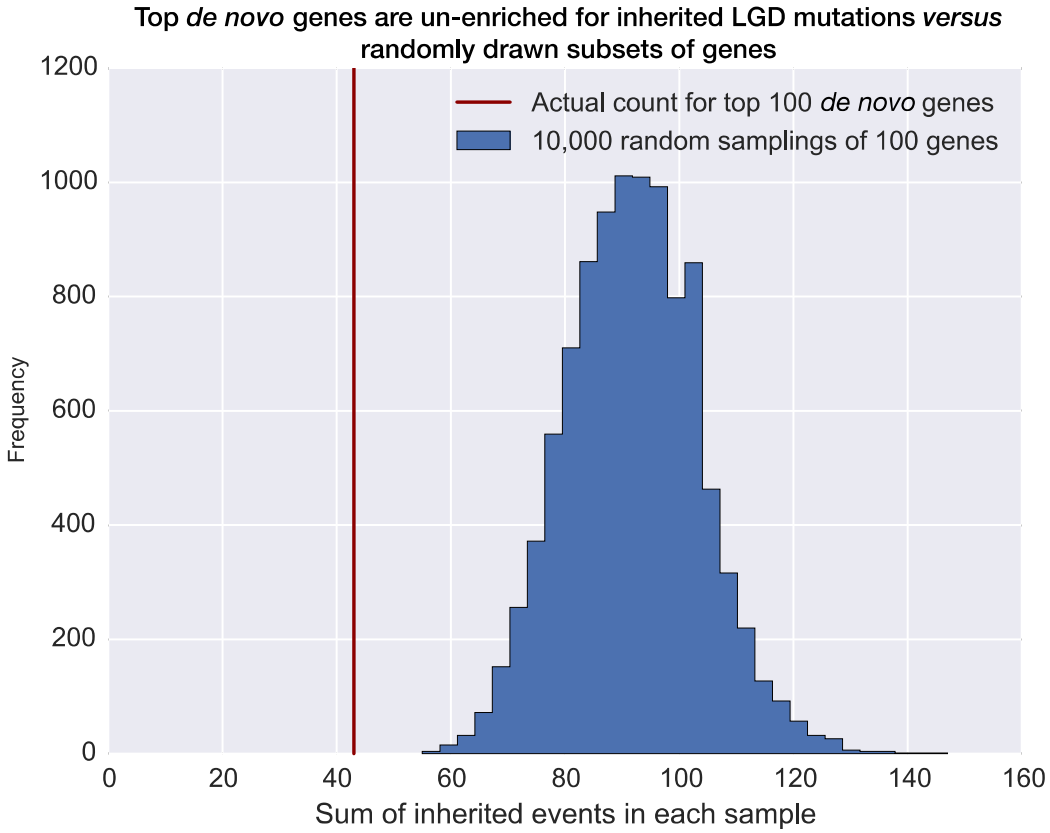
Epidemiological features of variant data were assessed for SNVs in 1,786 quad families (parents plus 1 affected proband and 1 unaffected sibling). We subsequently looked at subsets of these families (1,785 families where both proband and sibling sex were known) stratified by (**a**) proband sex and ultimately by (**b**) proband-sibling pairing by sex (male proband/male sibling, male proband/female sibling, female proband/male sibling, female proband/female sibling) in families where the sex was known for both the proband and the sibling.



Supplementary Figure 8

Detailed view of known CNVs affecting the *CSMD1* locus.

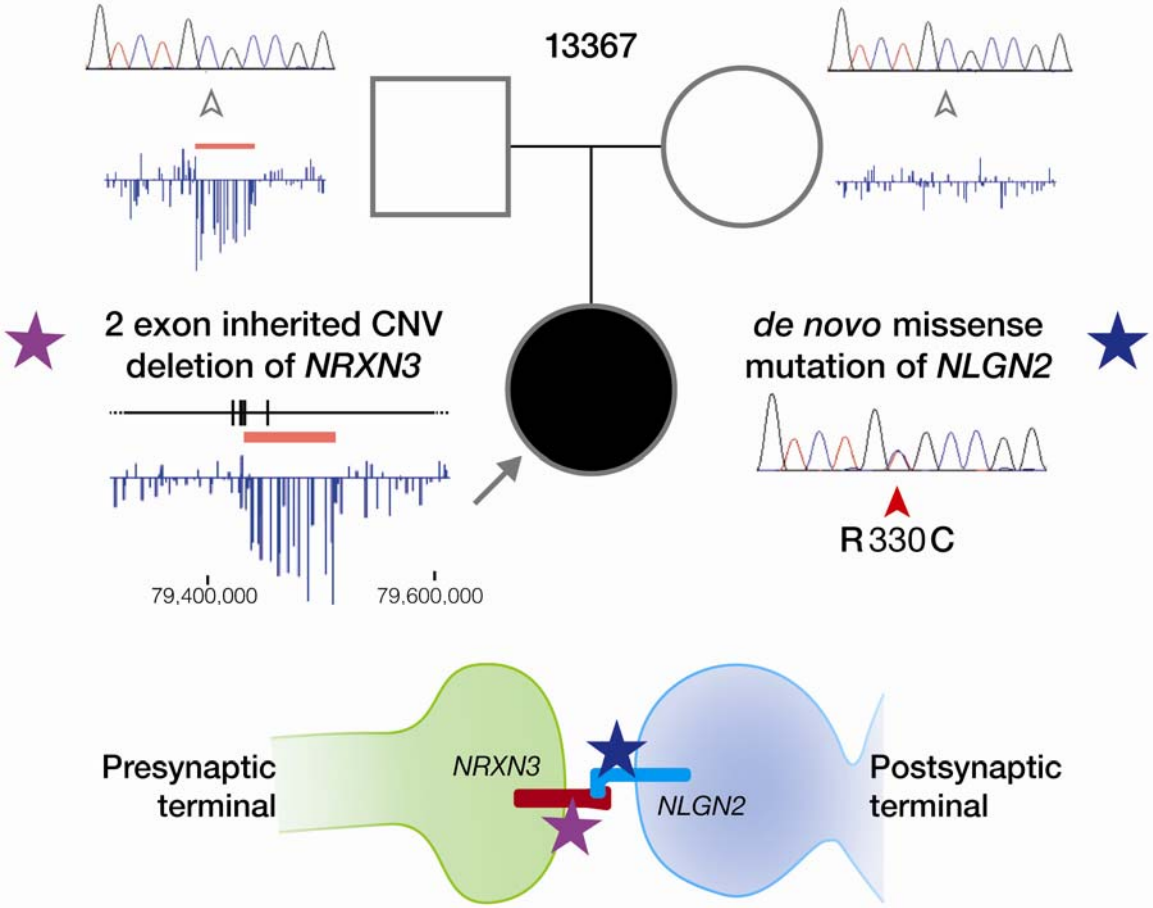
Of note is the depletion of CNV events at the 3′ end of the gene in the Database of Genomic Variants as compared to the 5′ portion of the gene.



Supplementary Figure 9

Sampling results from overlap analysis of genes enriched for *de novo* and inherited events.

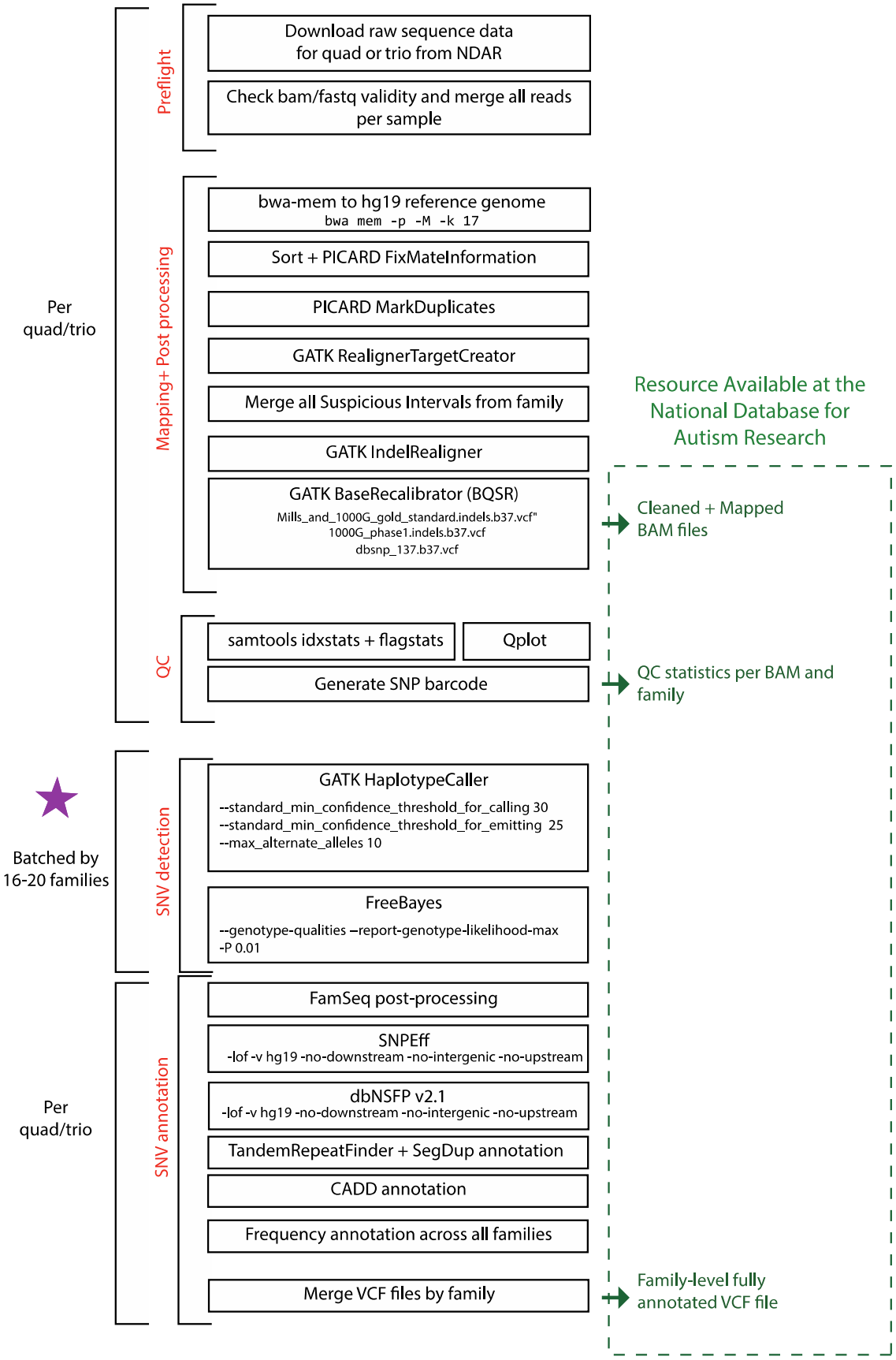
The blue distribution shows the sum of inherited CNVs and SNVs in probands for random 100-gene samplings from **Supplementary Table 8** (10,000 iterations). The red vertical line represents the actual (observed) count of CNVs and SNVs in probands for the 100 top enriched genes by *de novo* events.



Supplementary Figure 10

Detail of *de novo* and inherited mutations in interacting proteins found in a single family.

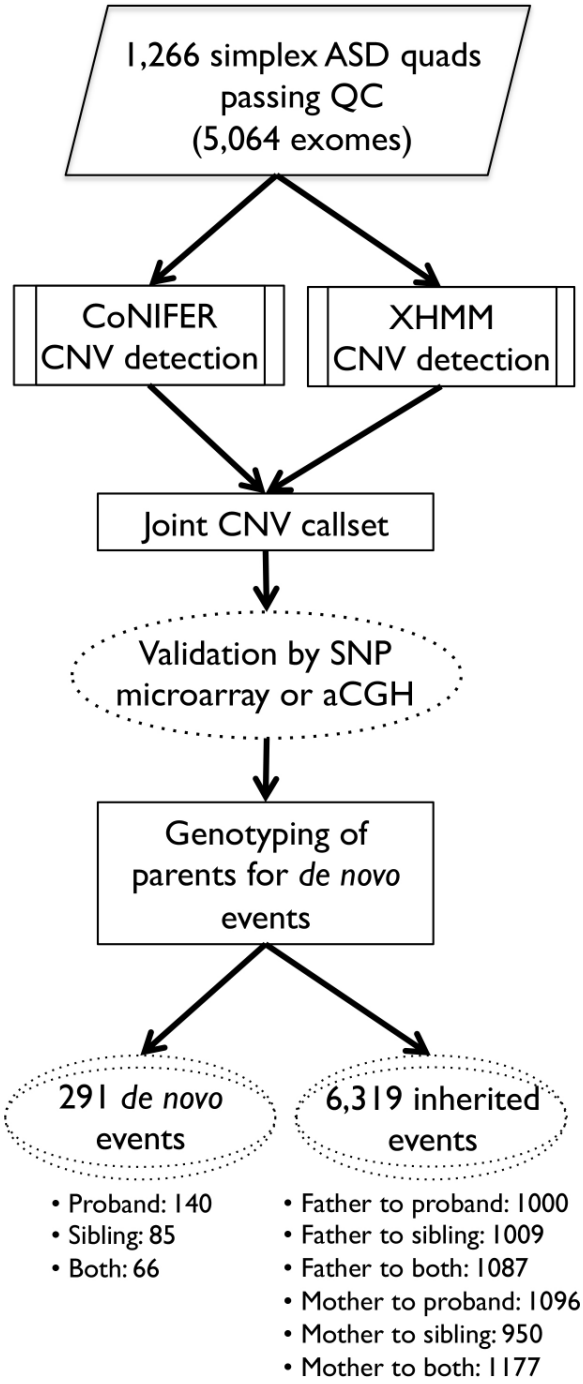
A proband with a compound mutation in synaptic receptor–ligand pair; namely, a *de novo* non-conservative mutation in neuroligin and an inherited two-exon deletion in neurexin. Phenotypic details of this proband include nonverbal IQ = 57, verbal IQ = 46, adaptive IQ = 77, mixed expressive-receptive language disorder, reported autism severity = 8 (of 10), elevated externalizing symptoms, delays in phrase speech, diagnosed heart murmurs, abnormal EEG, elevated BMI (z = 3.69) and macrocephaly (z = 3.08).



Nature Genetics: doi:10.1038/ng.3303

Supplementary Figure 11

Sequence file workflow and SNV identification.

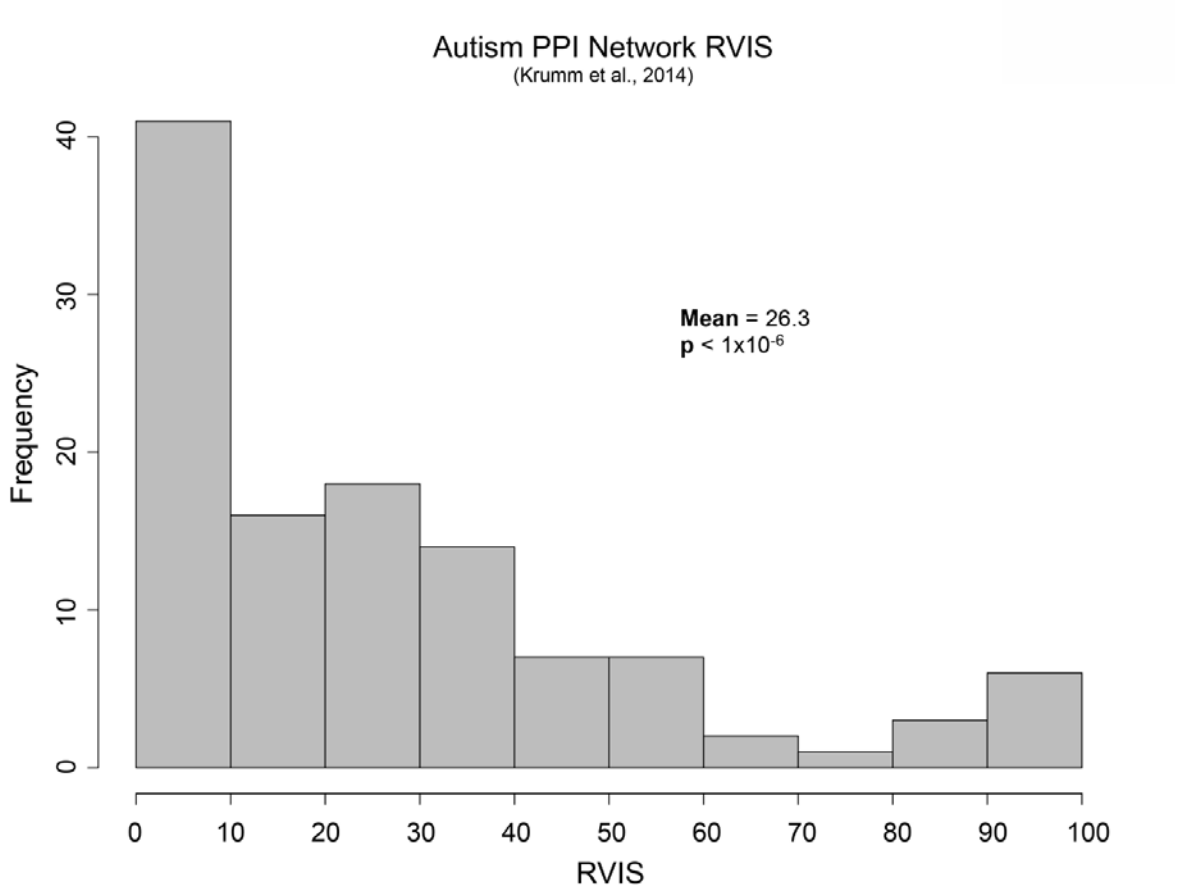


Supplementary Figure 12

CNV calling workflow.

1. BAM data were assembled from all previous SSC studies. 2. All reads were realigned using mrsFAST for CoNIFER and using BWA-mem for XHMM. For calling CNVs as part of CoNIFER, a pipeline using DNAcopy and CGHcall was used for segmentation. 3. Post-calling, CNVs were combined into one data set. 4. To validate exome calls, we utilized both available SNP microarray and array comparative genomic hybridization (aCGH) data. For SNP microarray analysis, probe-level copy number estimates were determined by

CRLMM, and validation of events was performed by permutation testing. 5. To identify true *de novo* events, parents were re-genotyped using CRLMM. 6. Post-validation, there were 291 *de novo* and 6,319 inherited events.



Supplementary Figure 13

RVIS values in the known autism network.

Histogram of RVIS values for all genes in the autism network described in Krumm *et al*.[53]. Also indicated is *P* value based on 1,000,000 permutations of a randomly selected gene set of the same size from the whole genome.