Supplementary Methods

The sample. Genotypes from 6709 samples from 10 sites were returned from the Translational Genomics Research Institute or TGEN. In the received diagnosis file, there was a total of 1496 nominal families (Table 1, manuscript), genotypes from 1491 families were returned by TGEN, and when intersected there were 1317 families with both (useful) genotypes and diagnostic status. Nominal family structures were simple and most consisted of nuclear families (≈86%). The average number of marriages per family was roughly 1.21.

Counts of genotyped DNAs returned from TGEN.

AGRE	IMGSAC	CPEA	Canada	Stanford	Duke	Vanderbilt	Paris	UNC	Mt. Sinai	Total
2213	1357	937	530	469	344	295	240	236	88	6709

Structure and counts of families.

Diagnostic group	No. marriages	No. families	No. nuclear	marriage/family
Narrow	659	522	450 (0.86)	1.26
Broad	904	731	636 (0.87)	1.24
hASD	507	437	393 (0.90)	1.16

The distribution of affected pairs is given below as counts of pairs of a specific kind per family, although we note that some of the different relative pairs fall in the same family. Suppose for example, in a participating family there was an affected sibling pair and an affected cousin; this family would contribute 1 to the full sibling pairs (FS) and 2 to cousins (CC) count. In addition, if there were four affected siblings in a single nuclear family, there would be 6 pairs and the family would contribute a count of 1 to the FS column in row 6.

Counts of relative pairs per family.

pairs	FS	PO	HS	NN	CC	125	OTHER
1	1017	1	17	6	36	5	6
2	3	0	4	1	6	1	0
3	85	0	0	0	2	0	0
4	0	0	0	0	2	0	2
5	0	0	0	0	0	0	0
6	2	0	0	0	0	0	0

FS = Full Sibs; PO = Parent/Offspring; HS = Half Sibs; NN = Uncle/Aunt with Nephew/Niece; CC = First Cousins; 125 = pairs with Kinship Coefficient of .125; and OTH = Pairs with Kinship Coefficient < .125.

Power. Our power estimates integrate results from various sources from the literature (Risch 1990, Krawczak 2001). Roughly, we should have a power of 80% to detect linkage of loci with a stringent LOD score of 3 threshold contributing a locus-specific risk of ≥ 1.29 for a sample size of 1,181 ASPs. (Our sample is slightly larger, in the sense that some families are extended.) For a sample of roughly 700 ASP, locus-specific risk of ≥ 1.54 is required have a power of 80% to detect linkage; and, for a sample of roughly 400 ASP, locus-specific risk of ≥ 1.88 is required.

Discussion of merging CNV information with linkage data. Our approach to merging CNV information with linkage data is necessarily limited by the uncertainty inherent in judging CNV-

conferred risk. One would like to partition CNVs into known CNAs, those likely to be CNAs, those unlikely to be CNAs, and those CNVs that are simply polymorphisms unrelated to ASD. In fact, a statistically-efficient linkage analysis would use a continuous scoring system for CNVs, scoring them 0 if they are certain to be a CNA and 1 if they are certain to be only a polymorphism, with intermediate scores reflecting the degree of ASD risk conferred. It would be natural to place the scoring system in a Bayesian framework, in which the prior (a function of the score and confidence in the CNV call) was established by expert opinion. Because the relationship between CNV and CNA is so unclear at the moment, and the grid of markers too sparse for high confidence in CNV calls, we decided to forego such analyses. Nonetheless our rudimentary approach to linkage analysis with data on CNV yields some intriguing results. Linkage statistics in key genomic regions drop for the filtered and plate methods of CNV-calling, but surprisingly often they increase for the batch method. The most conservative explanation for this observation is that the effect of batch is due to chance, while the effect of filtered and plate is due to diminishing sample size. It is also possible, however, that the increased linkage seen with batch results could be meaningful. If they were, then some of the CNVs dropped in batch, but not the other CNV-calling methods, must be particularly salient for risk to autism.

Supplementary References

Risch, N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am. J. Hum. Genet. 46: 229-241.

Krawczak, M (2001) ASP – a simulation-based power calculator for genetic linkage studies of qualitative traits, using sib-pairs. Hum. Genet. 109:675-677.