American Journal of Human Genetics, Volume *93*

Supplemental Information

Identification of Small Exonic CNV from Whole-Exome Sequence Data and Application to Autism Spectrum Disorder

Christopher S. Poultney, Arthur P. Goldberg, Elodie Drapeau, Yan Kou, Hala Harony-Nicolas, Yuji Kajiwara, Silvia De Rubeis, Simon Durand, Christine Stevens, Karola Rehnström, Aarno Palotie, Mark J. Daly, Avi Ma'ayan, Menachem Fromer, and Joseph D. Buxbaum

Supplemental Inventory

Supplemental Figures and Tables

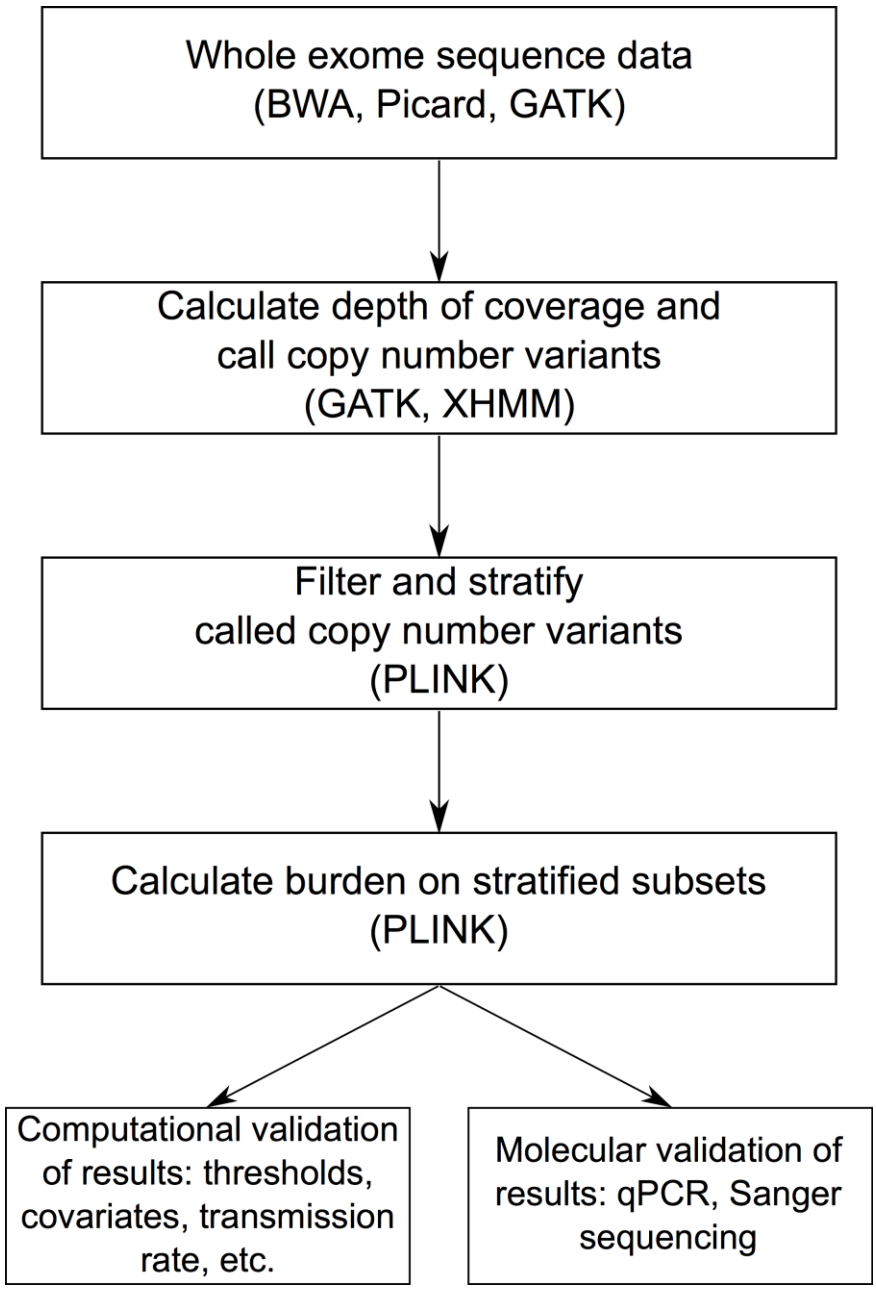

Figure S1. Processing and Validation Pipeline for Calling Small (1-30 kb) CNV from Whole Exome Sequencing Data

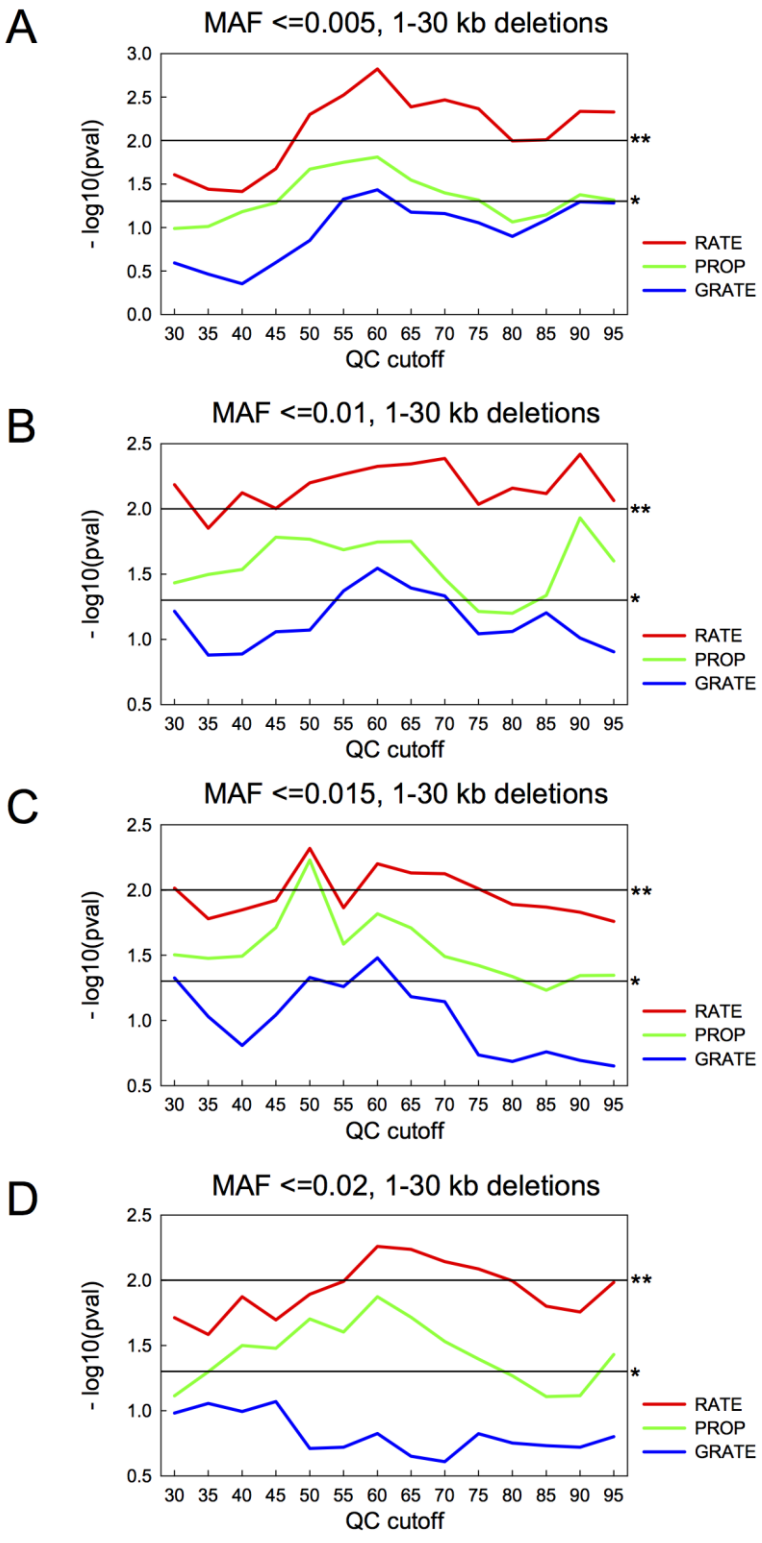

Figure S2. Computational Validation of SQ and MAF Cutoffs

Each panel shows P-values for the three burden tests across SQ values for MAF <= 0.005 (A), 0.01 (B), 0.015 (C), and 0.02 (D).

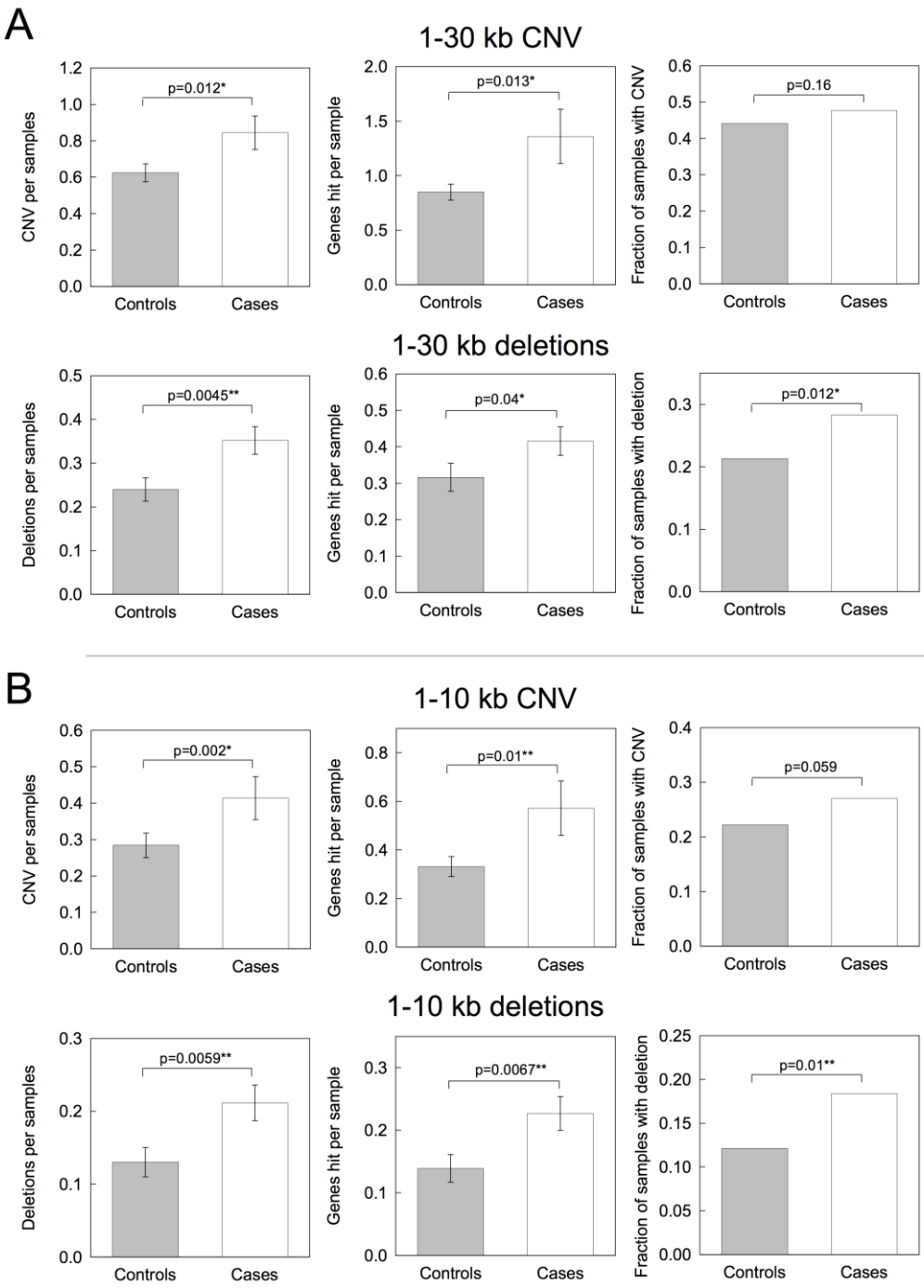

Figure S3. Enrichment for Small Deletions in ASD, with Explicit Handling of Read Depth and Processing Batch

To account for the bimodal distribution of sample mean read depth, clusters of samples with similar mean read depth were determined based on processing batch and read depth distribution. Burden tests were then performed, generating p-values derived using within-cluster permutation. As in Figure 2, panel A shows burden test results for all CNV (top) and deletions (bottom) in the 1-30 kb range, and panel B shows the same for the 1-10 kb range. Burden test results using within-cluster permutation are very similar to results derived without clustering.

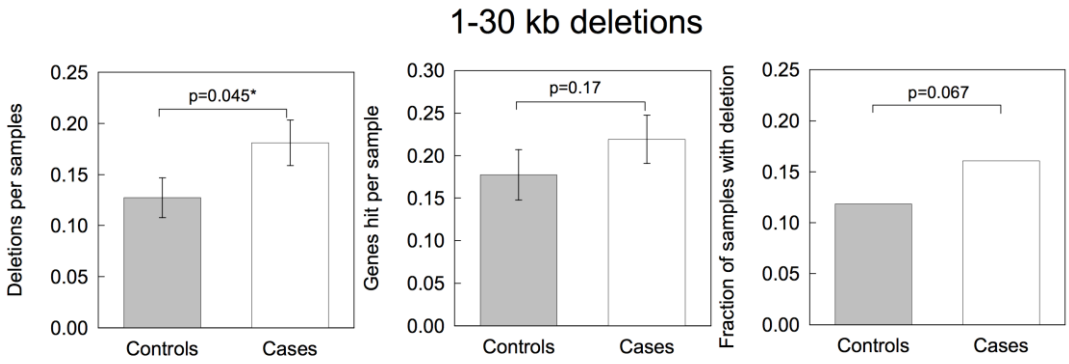

Figure S4. Enrichment for Small (1-30 kb) Deletions in Singleton CNV in ASD

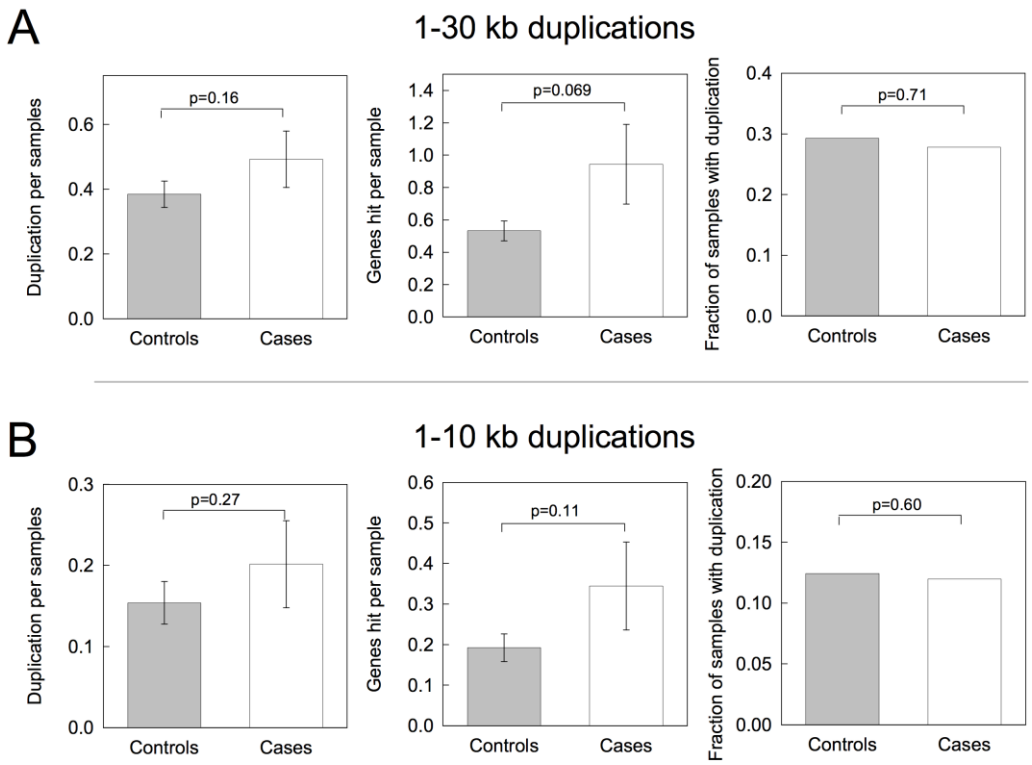

Figure S5. Lack of Significant Enrichment in Small (1-30 kb) Duplications in ASD

While there are nominal increases in duplications per sample and genes hit per sample, these changes were not significant. This is, however, possibly a power issue.

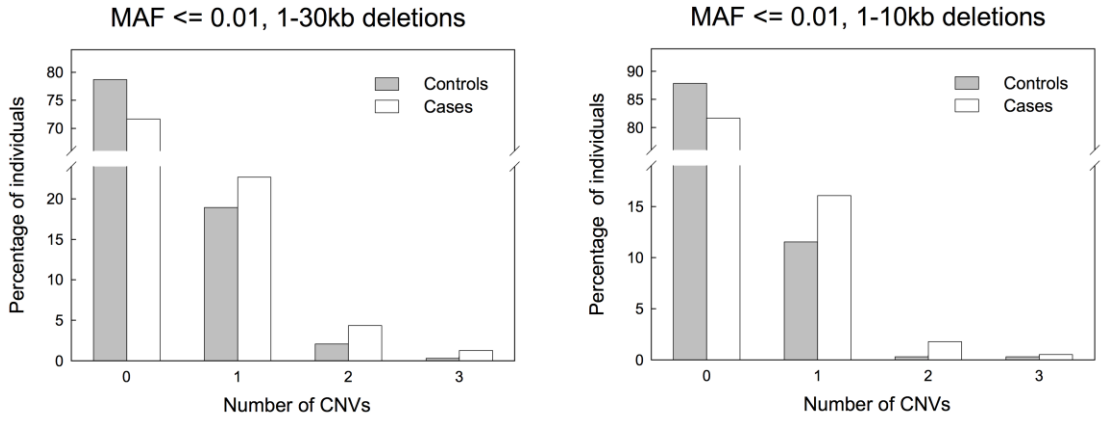

Figure S6. Distribution of Small Deletions

Percent of individuals is shown as a function of number of events for 1-30 kb deletions (left) and 1-10 kb deletions (right). The ratio of the rates between case and control differs significantly from one using a two-sided test at P=0.0066 (1-30 kb deletions) and P=0.0096 (1-10 kb deletions).

Table S1. XHMM Runtime Arguments

Argument	Value
minTargetSize	10
maxTargetSize	10000
minMeanTargetRD	10
maxMeanTargetRD	500
minMeanSampleRD	25
maxMeanSampleRD	250
maxSdSampleRD	168
maxSdTargetRD	30
PVE_mean_factor	0.7

Table S2. XHMM Parameter File Values Used for Genotyping Stage

Parameter	Value
Pr(start DEL) = Pr(start DUP)	1e-08
Mean number of targets in CNV	6
Mean distance between targets in CNV	70 kb
DEL read depth distribution	~N(mean=-3, var=1)
DIP read depth distribution	~N(mean=0, var=1)
DUP read depth distribution	~N(mean=3, var=1)