The American Journal of Human Genetics, Volume *103*

Supplemental Data

Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data

Michael H. Guo, Lacey Plummer, Yee-Ming Chan, Joel N. Hirschhorn, and Margaret F. Lippincott

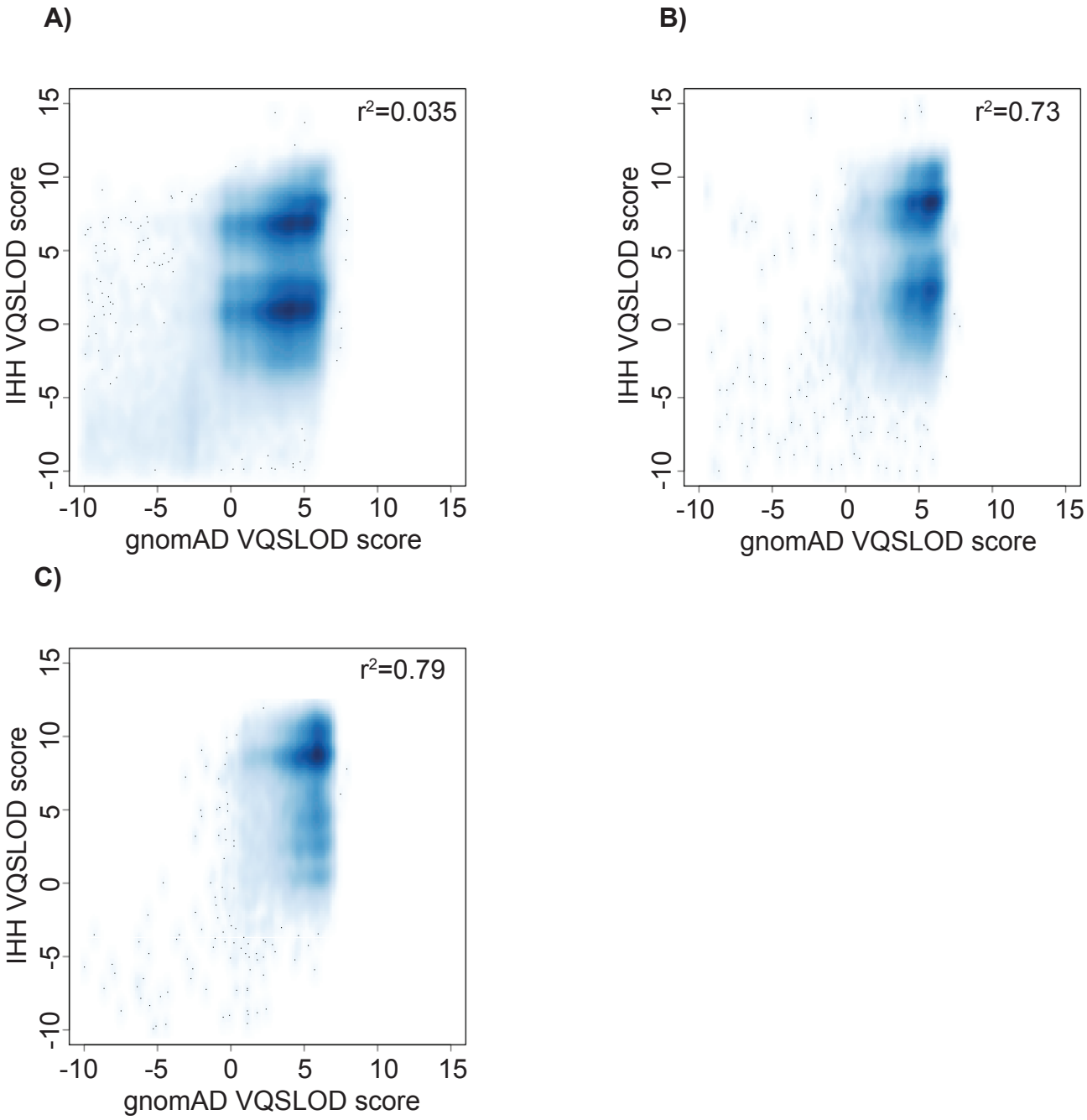

Figure S1: Correlation of VQSLOD scores. Shown is the correlation of VQSLOD scores between shared SNV sites in gnomAD (x-axis) and case cohort sequencing (y-axis) at different minor allele frequency cutoffs: MAF < 0.01 (A), 0.01≤MAF<0.05 (B), and 0.05≤MAF<0.50 (C).

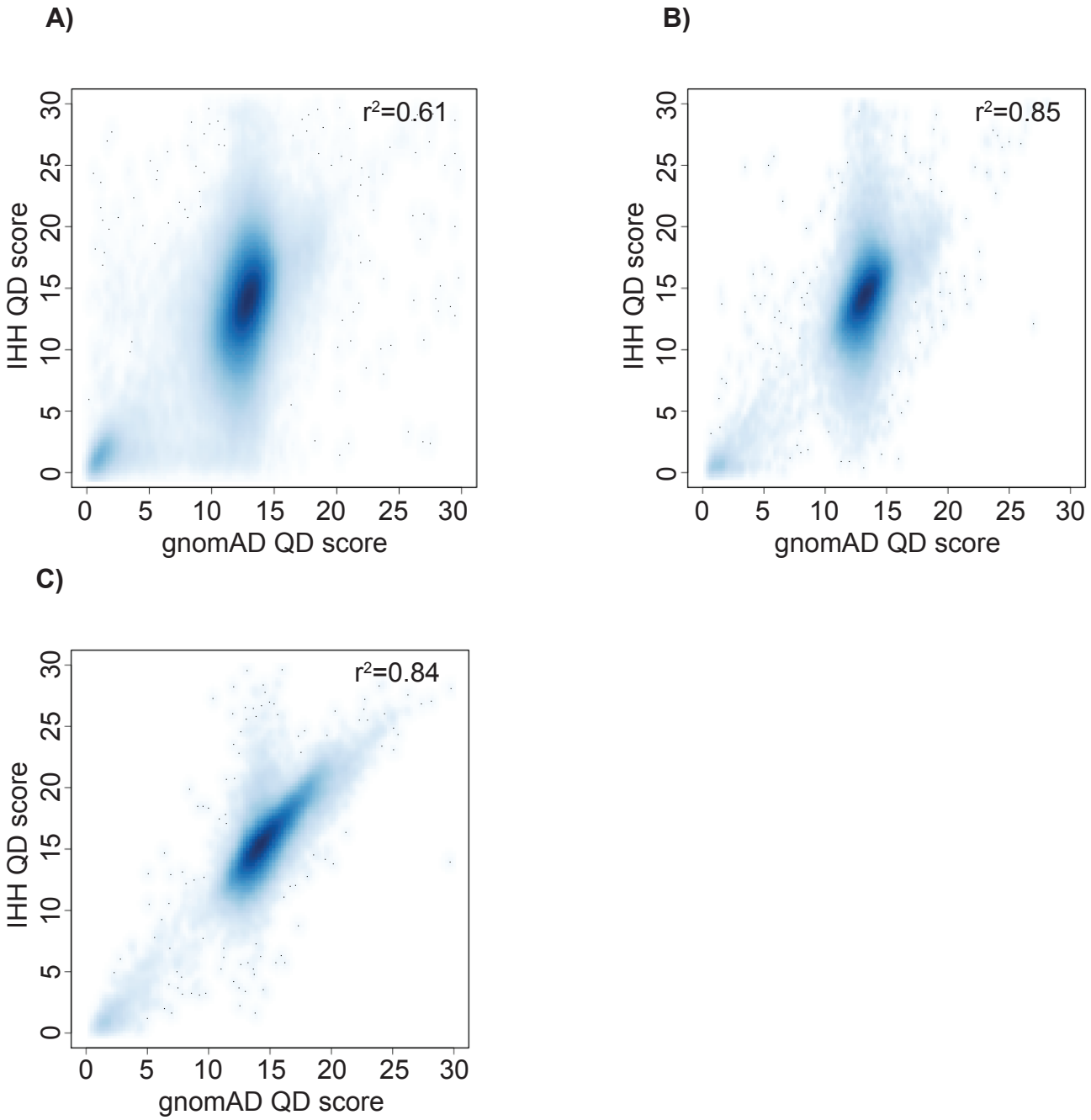

Figure S2: Correlation of QD scores. Shown is the correlation of QD scores between shared SNV sites in gnomAD (x-axis) and case cohort sequencing (y-axis) at different minor allele frequency cutoffs: MAF < 0.01 (A), 0.01≤MAF<0.05 (B), and 0.05≤MAF<0.50 (C).

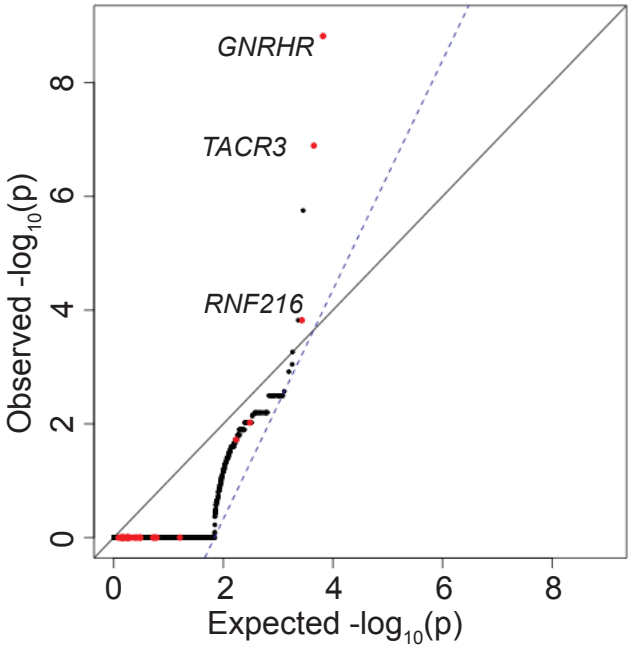

Figure S3: Burden testing under recessive model. The same QD filters were used as in Figure 5. Variants with MAF < 0.1% were used, and the QQ plot shows results using PTVs plus missense variants computationally predicted to be damaging.

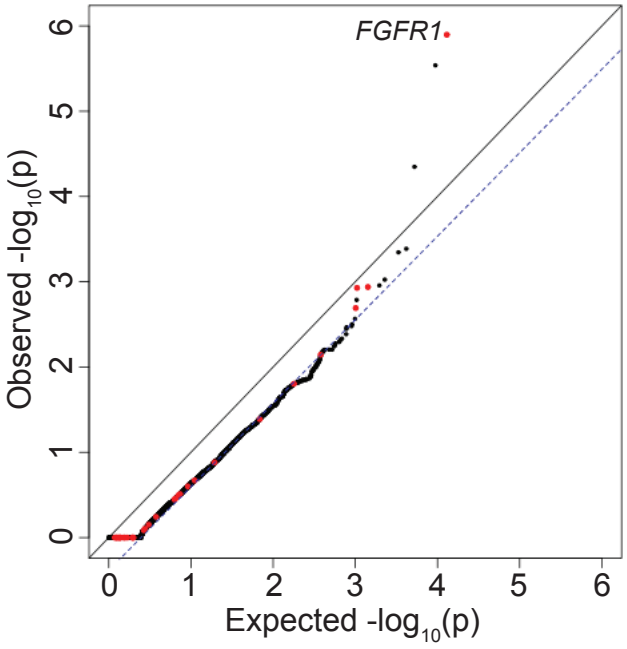

Figure S4: Burden testing results with individuals of European ancestry only. For the case sequencing cohort, only individuals of European ancestry as determined by PCA were used (n= 263). For controls, only non-Finnish European individuals in gnomAD were used (n=55,860). The same QD filters were used as in Figure 5. Variants with MAF < 0.1% were used, and the QQ plot shows results using PTVs plus missense variants computationally predicted to be damaging.

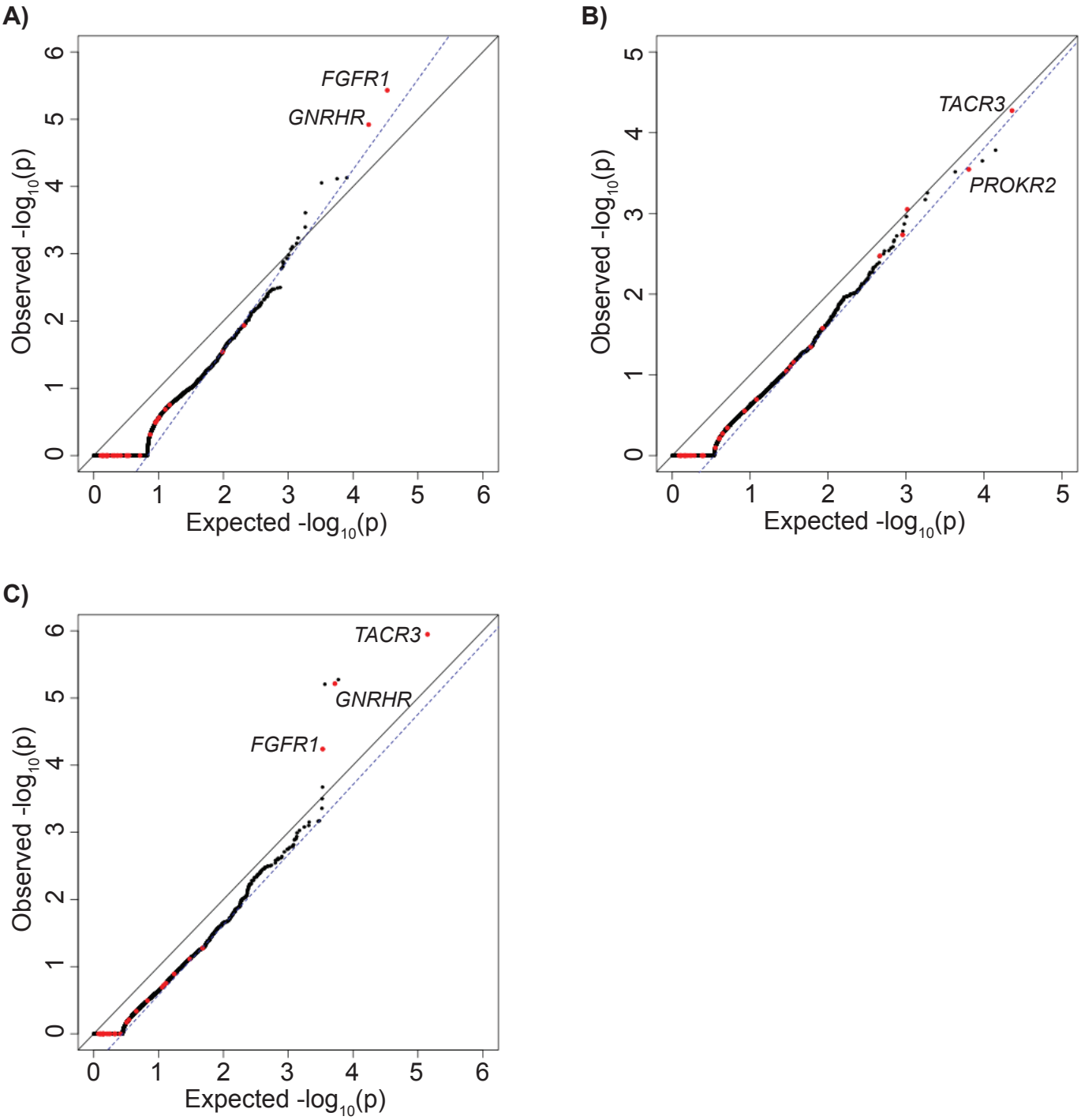

Figure S5: Burden testing by sequencing batch. Burden testing results for each case sequencing batch as compared to gnomAD controls: A) Batch 1 sequenced at Yale, B) Batch 2 sequenced at the Broad Institute with Agilent capture and with some selection for *PROKR2* heterozygotes and negative screening as described in Subjects and Methods, and C) Batch 3 sequenced at the Broad Institute using ICE capture. The same QD filters were used as in Figure 5. Variants with MAF < 0.1% were used, and the QQ plot shows results using PTVs plus missense variants.

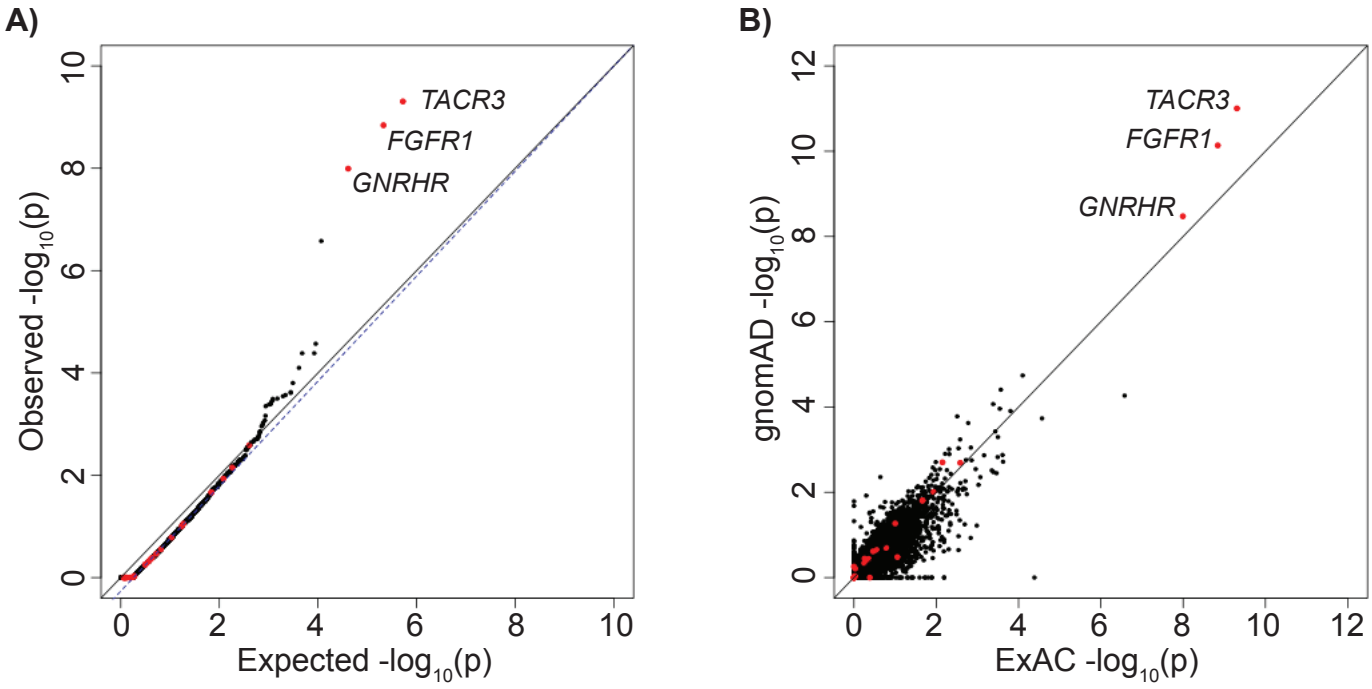

Figure S6: Burden testing with ExAC as control database. A) Burden testing when comparing the IHH case cohort (n=393) to ExAC (n=60,706) as a control cohort. For IHH case cohort sequencing, SNVs in the top 95% of QD scores and indels in the top 95% were considered. For ExAC control cohort, SNVs in the top 80% of QD scores and indels in the top 80% were considered. Variants with MAF < 0.1% were used, and the QQ plot shows results using PTVs plus missense variants computationally predicted to be damaging. B) Comparison of p-values for burden testing when using ExAC (x-axis) or gnomAD (y-axis) as the control cohort. Shown are the $-\log_{10}$(p-value) when testing variants with MAF < 0.1% and which are either PTVs or missense variants computationally predicted to be damaging.

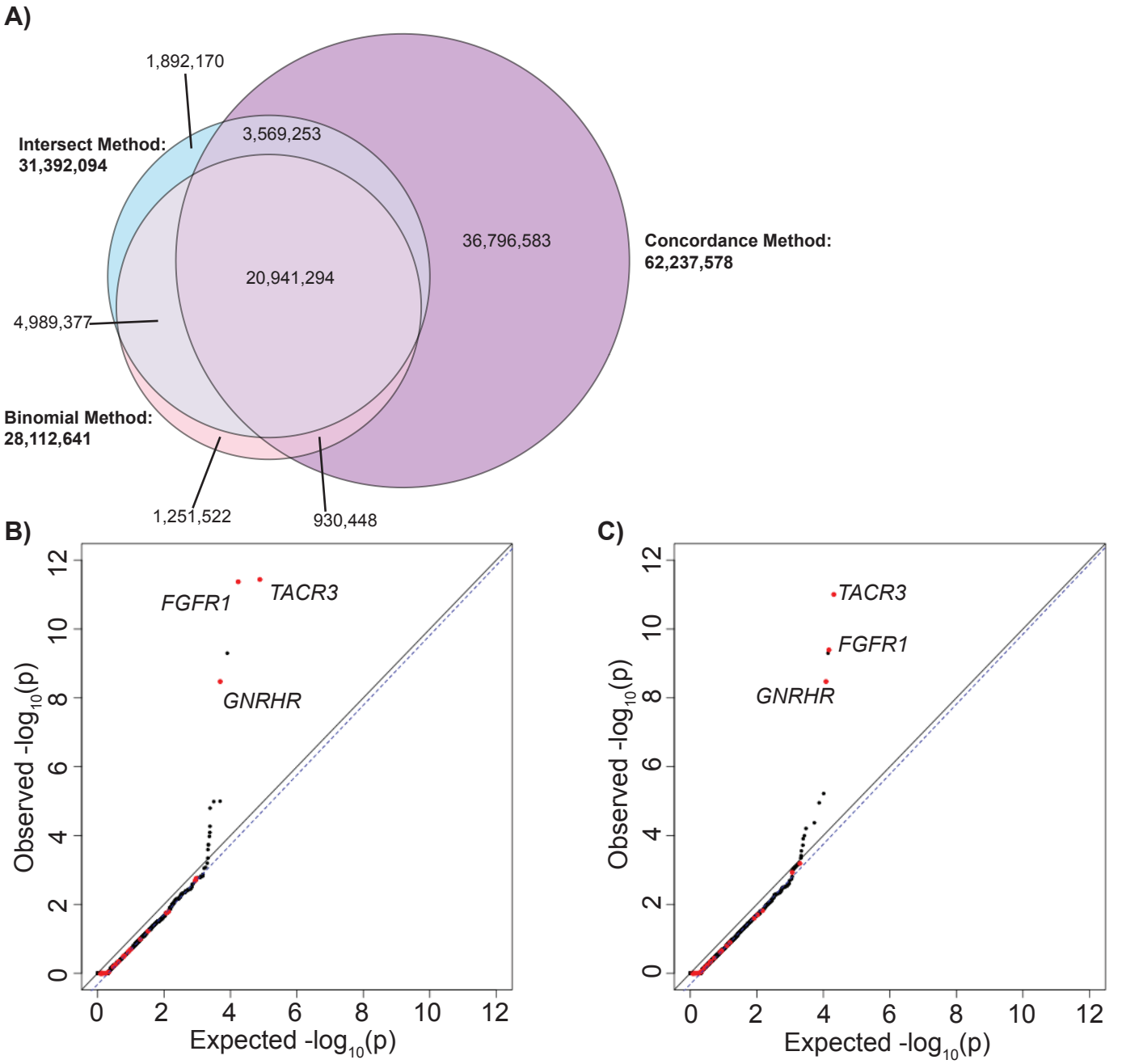

Figure S7: Comparison of approaches for adjusting for read depth. A) Venn diagram comparing the number of coding base pairs analyzed when using three approaches to adjust for read depth. The "Intersect Method" is the approach used in this paper, where only bases with >10X coverage in >90% of samples in both the case and control cohorts are analyzed. The "Binomial Method" is the approach used in Raghavan et al., where only bases that are not significantly different (p>0.001) in number of individuals covered at >10X in cases versus controls are analyzed. The "Concordance Method" is the approach used in Cirulli et al., where only exons that are >90% concordant in the number of bases covered at >10X in >90% of samples in case as compared to control sequencing are analyzed. B) Burden testing QQ plot when the approach used in Raghavan et al. of only considering bases that are not significantly different in number of individuals covered is applied to adjust for read depth ("Binomial Method"). C) Burden testing QQ plot when the approach used in Cirulli et al. of only considering exons with high concordance in coverage between case and control sequencing is applied to adjust for read depth ("Concordance Method"). Compare panels B and C with Figure 5B, which uses the approach utilized in this paper of only analyzing sites with sufficient coverage in both case and control sequencing ("Intersect Method"). As in Figure 5B, only qualifying variants with MAF < 0.1% and which are either PTVs or missense variants computationally predicted to be damaging are considered.