

# Comparing SUGEN and GENESIS

Results from both Imputed and Genotyped data  
for 7 traits

R. Tao<sup>1;2</sup>, J. Haessler, D.Y. Lin, T. Thornton, S. Buyske, Y. Patel, C.R. Gignoux, E. Stahl, T. Matisse, G.L. Wojcik, K.E. North, S. Bien, C.D. Bustamante, L. Hindorf, R.J.F. Loos, E.E. Kenny, C. Carlson, C. Kooperberg, D.O. Stram.

1. Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599

2. Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN 37203

# Background

- All study comparisons were made in PAGE HCHS/SOL data (Population Architecture using Genomics and Epidemiology study (<https://www.pagestudy.org/>)).
- 53,424 participants in the PAGE study were genotyped for 1.7 million SNPs on the Illumina MEGA array.
- Genome-wide association studies of the PAGE data require statistical methods to control for hidden population stratification, admixture, and relatedness.
- Two adjustment methods, one based on linear mixed models (LMM) and another on generalized estimating equations (GEE) were compared.

# Methods

LMM is implemented in the R package GENESIS.

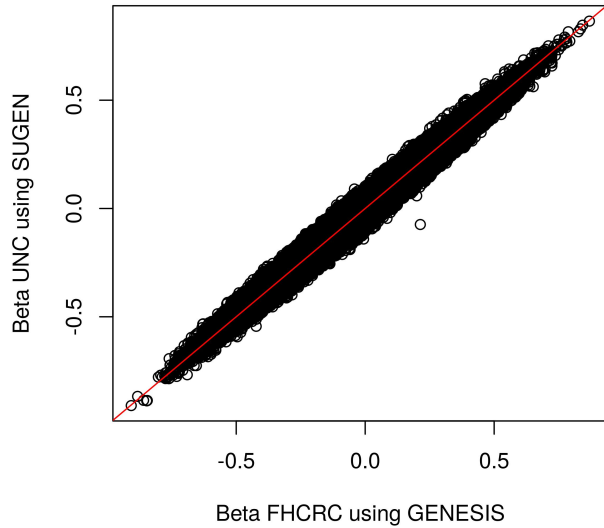
GEE is implemented in the standalone software program SUGEN.

Markers with  $MAF > \frac{1}{2} \%$  were used in the analysis.

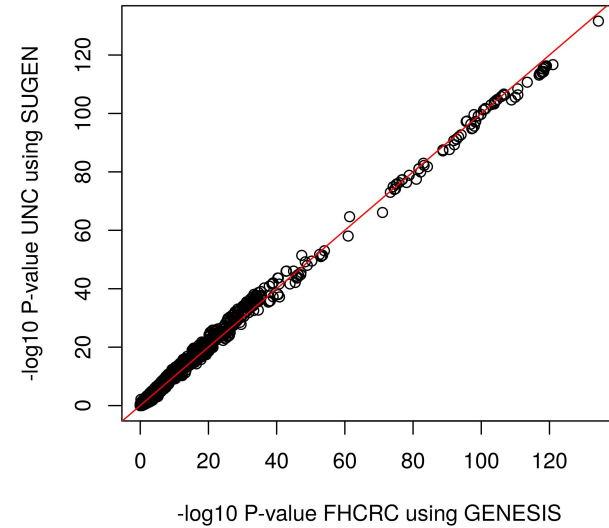
All analyses included principal components, age, sex, ancestry group, and study as adjustment variables.

# CRP Imputed Data

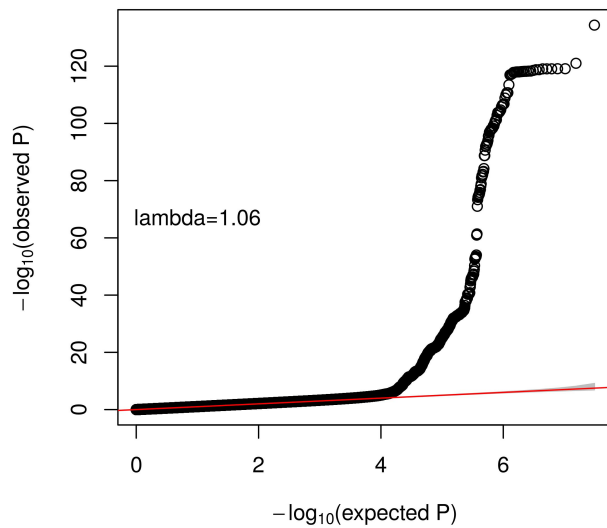
Betas for GENESIS and SUGEN  
for 30872231 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



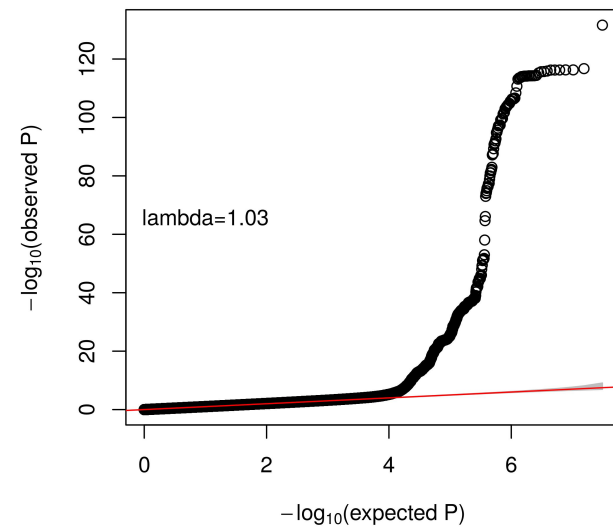
P-values for GENESIS and SUGEN  
for 30872231 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



GENESIS Results QQ-Plot

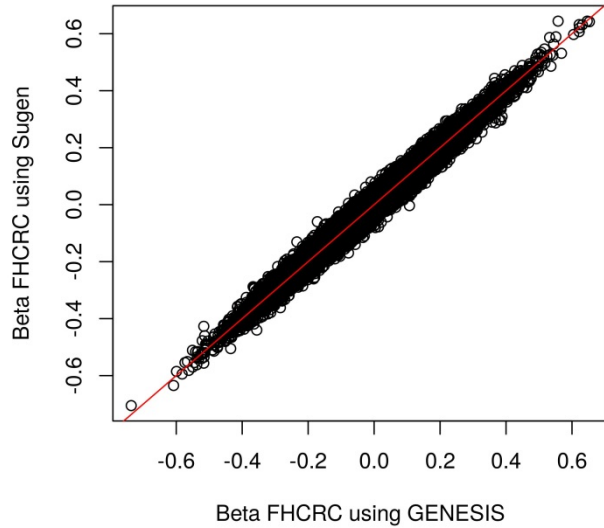


SUGEN Results QQ-Plot

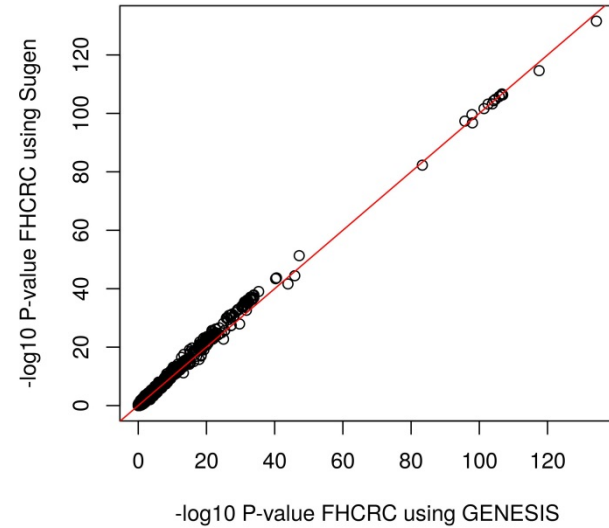


# CRP Genotyped Data

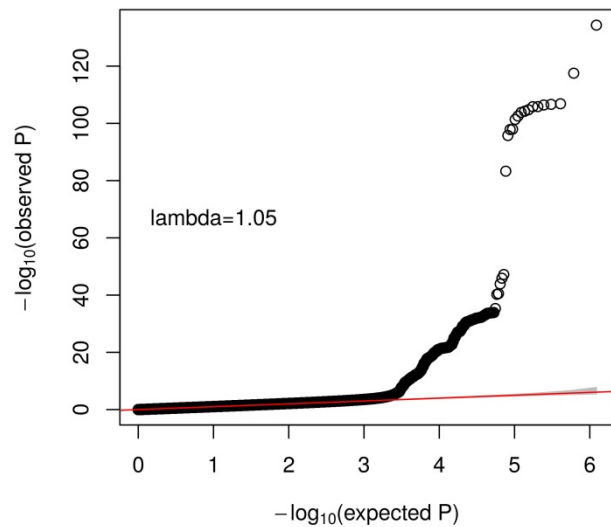
Betas for GENESIS and SUGEN done at FHCRC for 1224145 Genotyped SNPs with  $\text{effN} \geq 30$



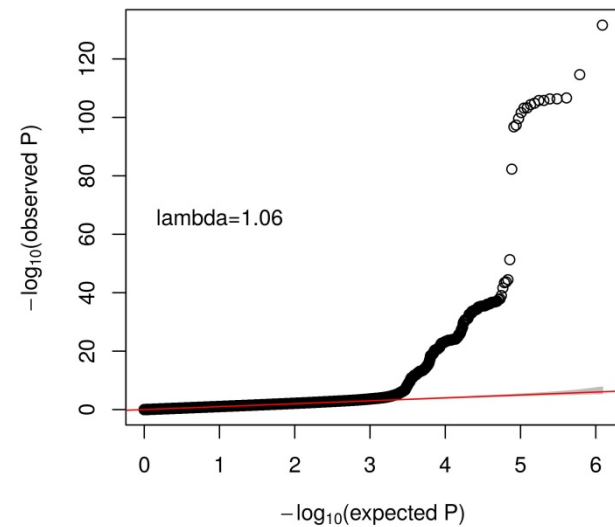
$-\log_{10}$  P-values for GENESIS and SUGEN done at FHCRC for 1224145 Genotyped SNPs with  $\text{effN} \geq 30$



GENESIS Results QQ-Plot

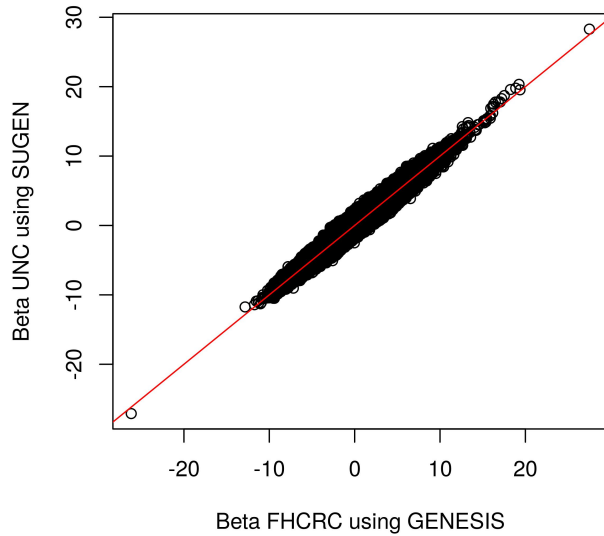


SUGEN Results QQ-Plot

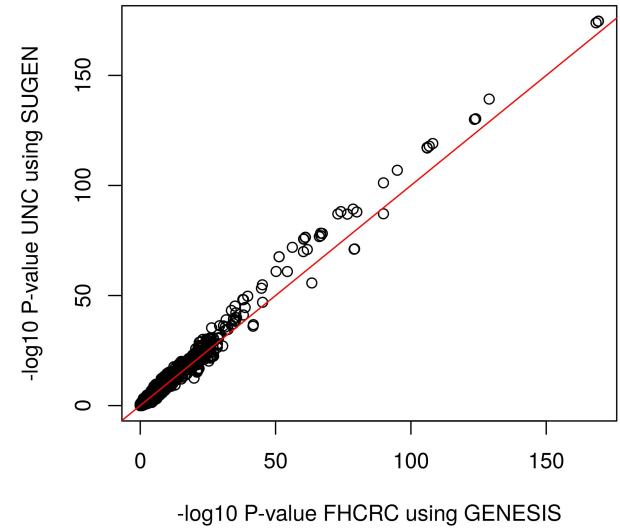


# HDL Imputed Data

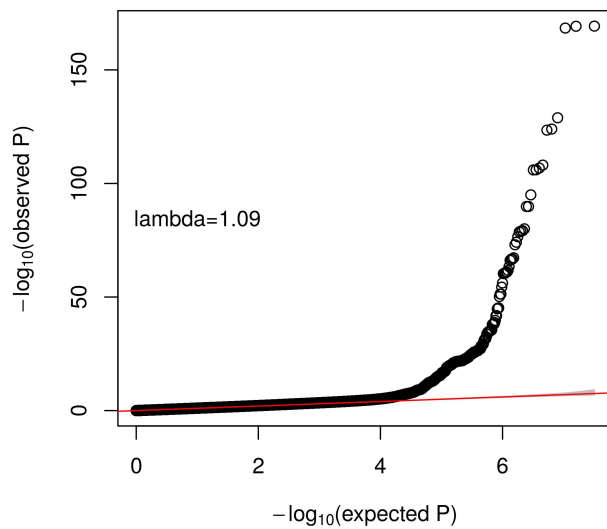
Betas for GENESIS and SUGEN  
for 31978361 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



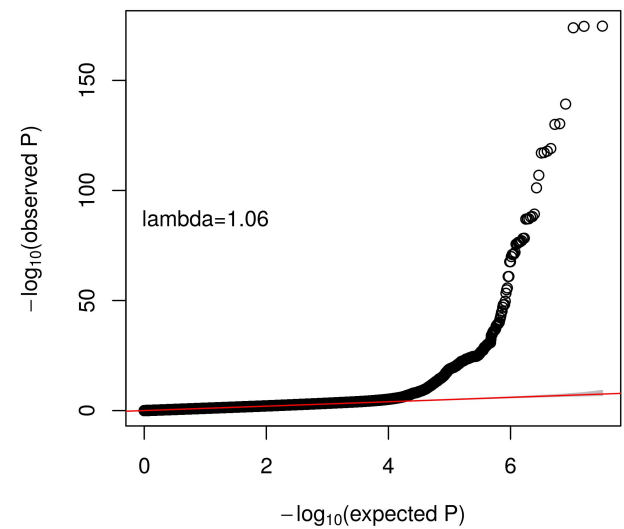
P-values for GENESIS and SUGEN  
for 31978361 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



GENESIS Results QQ-Plot

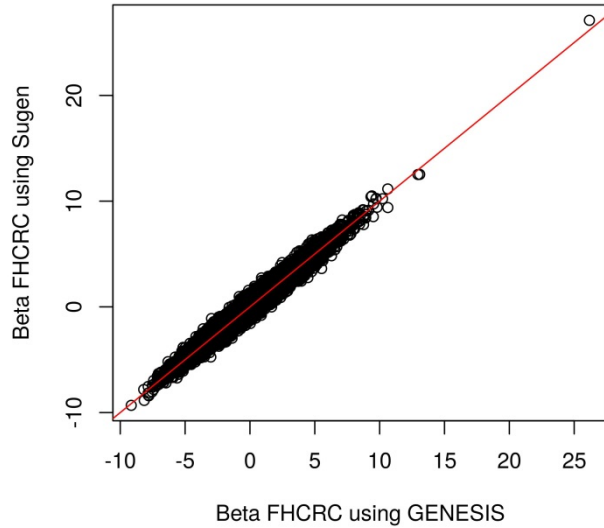


SUGEN Results QQ-Plot

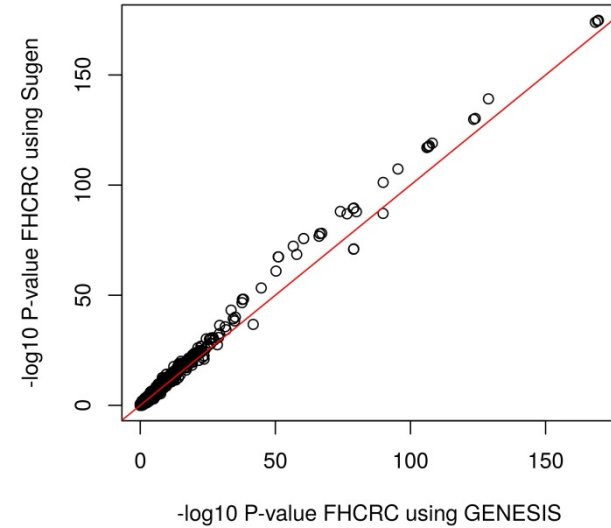


# HDL Genotyped Data

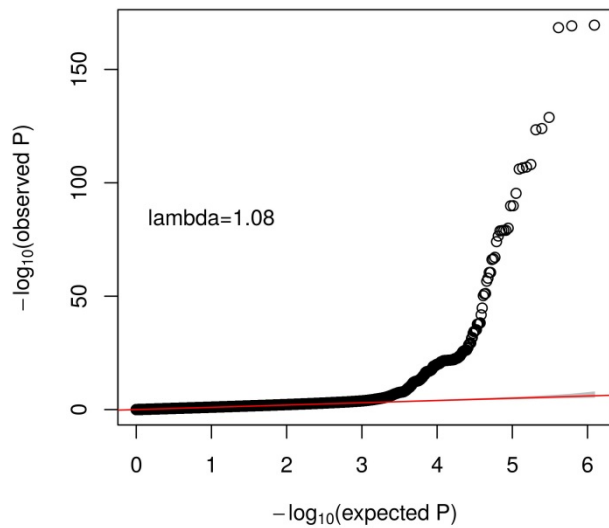
Betas for GENESIS and SUGEN done at FHCRC for 1232708 Genotyped SNPs with  $\text{effN} \geq 30$



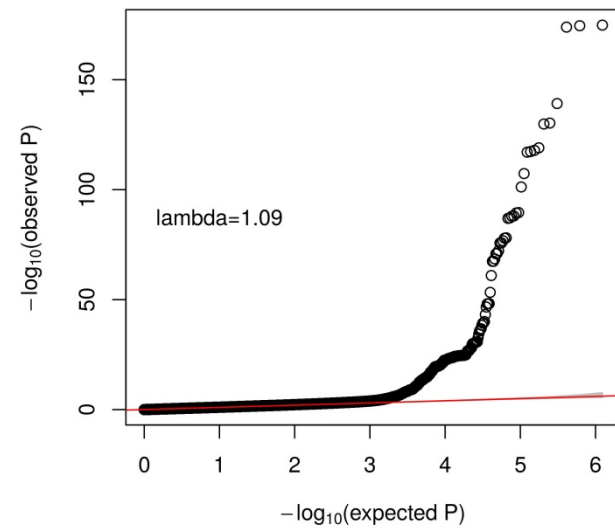
$-\log_{10}$  P-values for GENESIS and SUGEN done at FHCRC for 1232708 Genotyped SNPs with  $\text{effN} \geq 30$



GENESIS Results QQ-Plot

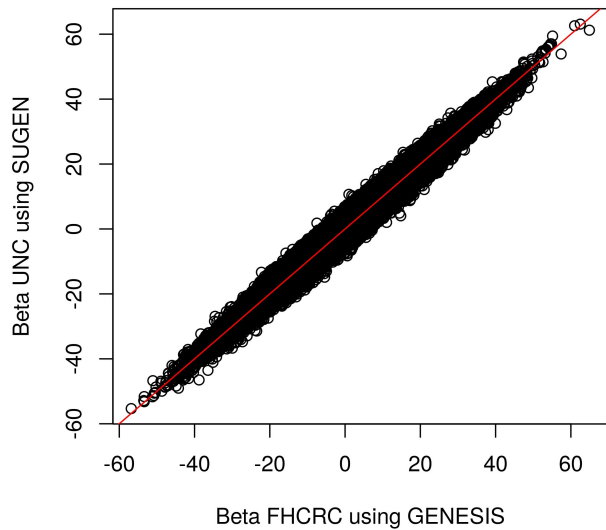


SUGEN Results QQ-Plot

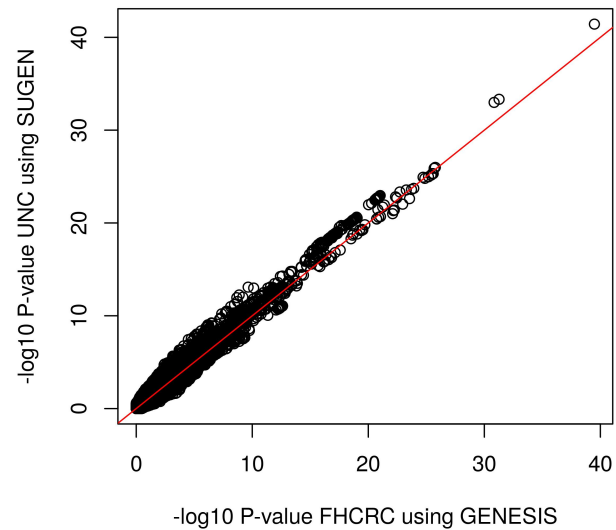


# Platelets Imputed Data

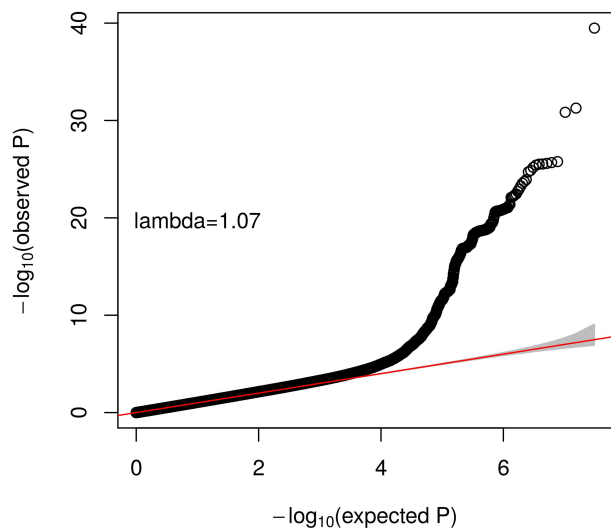
Betas for GENESIS and SUGEN  
for 30743395 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



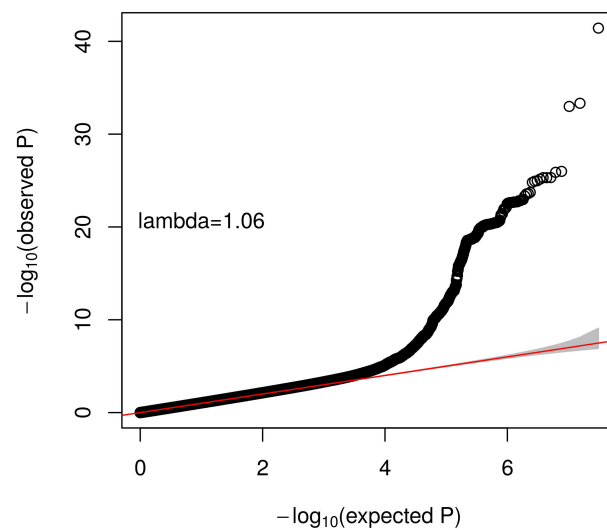
P-values for GENESIS and SUGEN  
for 30743395 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



GENESIS Results QQ-Plot



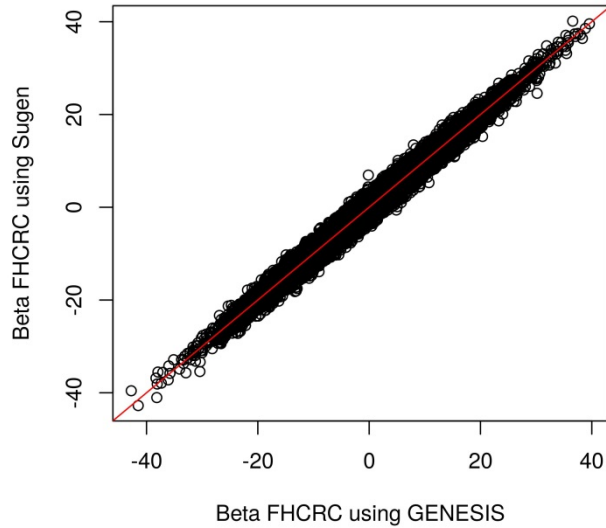
SUGEN Results QQ-Plot



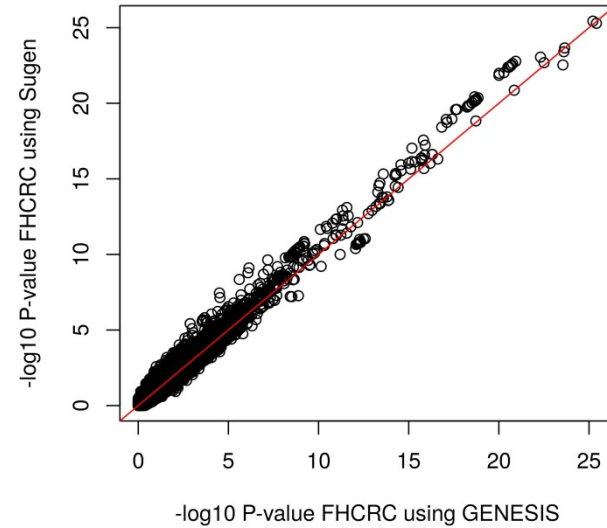


# Platelets Genotyped Data

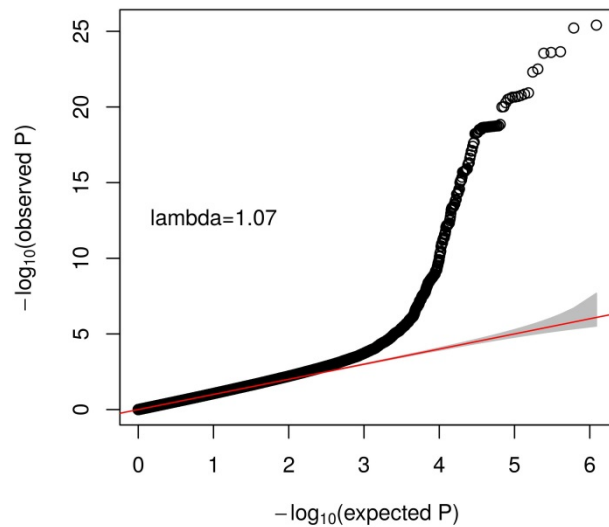
Betas for GENESIS and SUGEN done at FHCRC for 1224120 Genotyped SNPs with  $\text{effN} \geq 30$



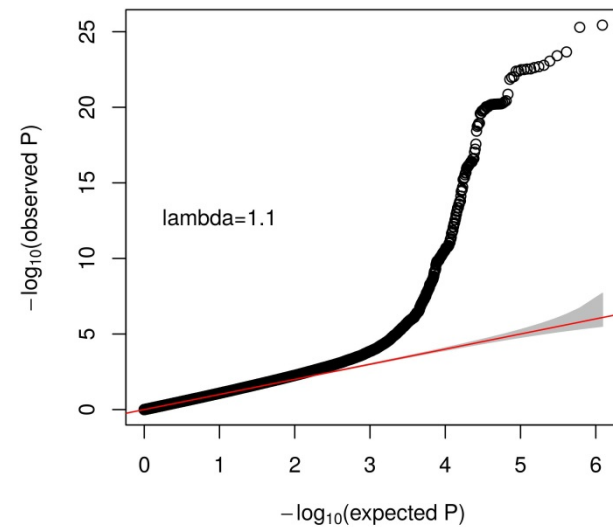
$-\log_{10}$  P-values for GENESIS and SUGEN done at FHCRC for 1224120 Genotyped SNPs with  $\text{effN} \geq 30$



GENESIS Results QQ-Plot

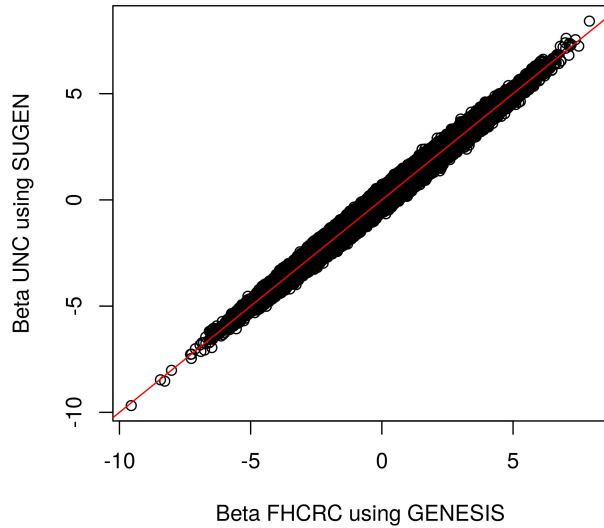


SUGEN Results QQ-Plot

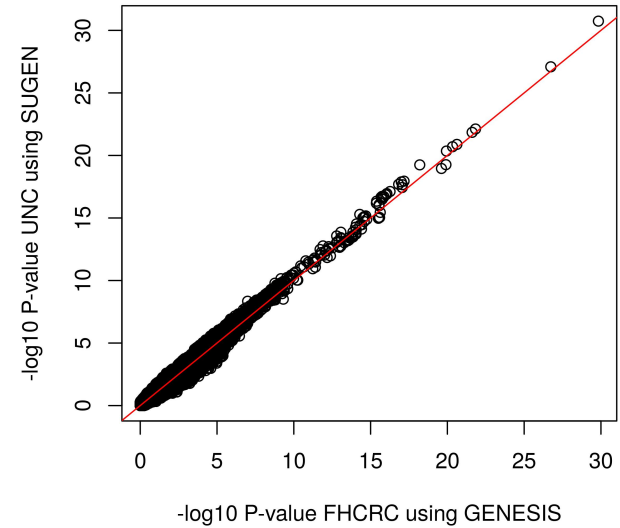


# QRS Imputed Data

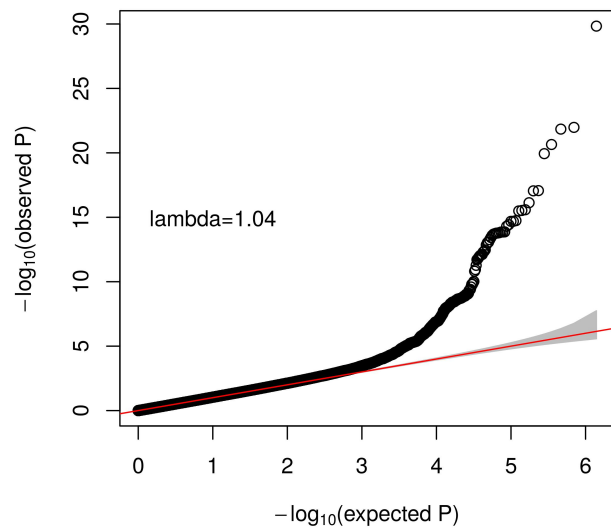
Betas for GENESIS and SUGEN  
for 26567875 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



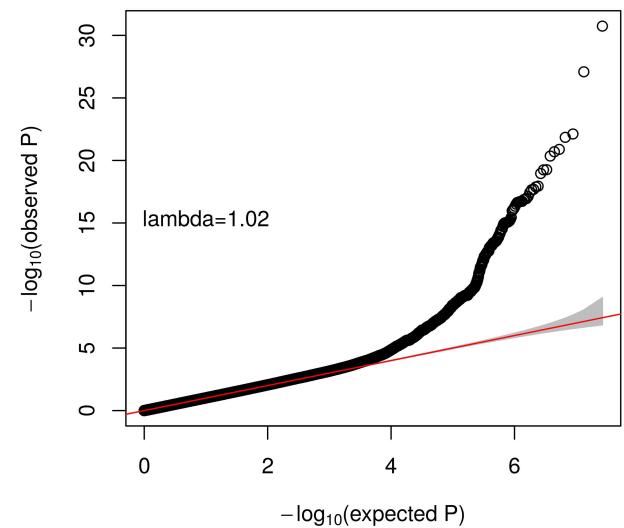
P-values for GENESIS and SUGEN  
for 26567875 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



GENESIS Results QQ-Plot

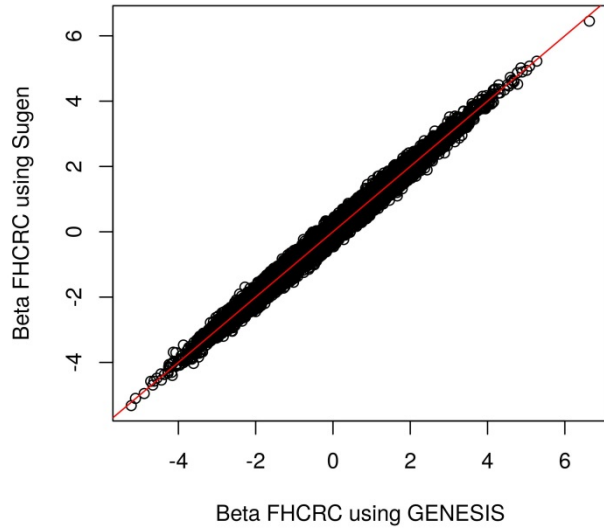


SUGEN Results QQ-Plot

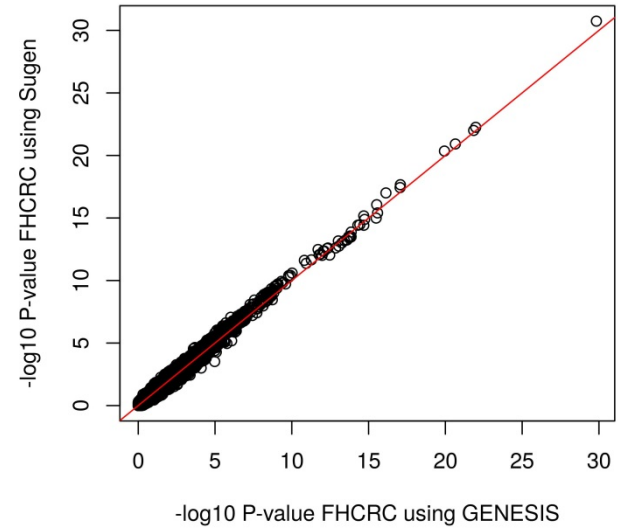


# QRS Genotyped Data

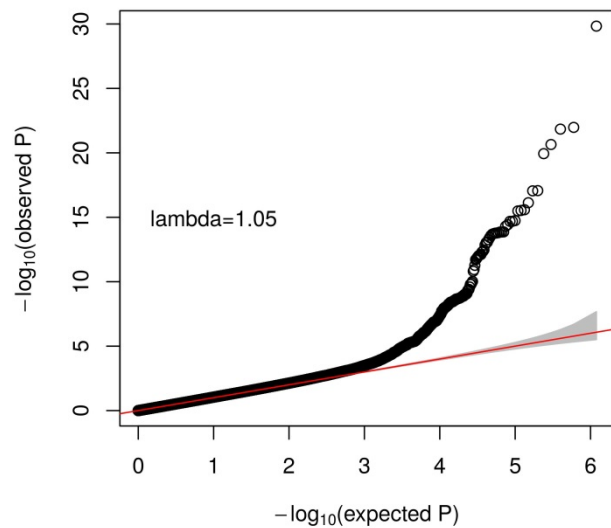
Betas for GENESIS and SUGEN done at FHCRC for 1195830 Genotyped SNPs with  $effN \geq 30$



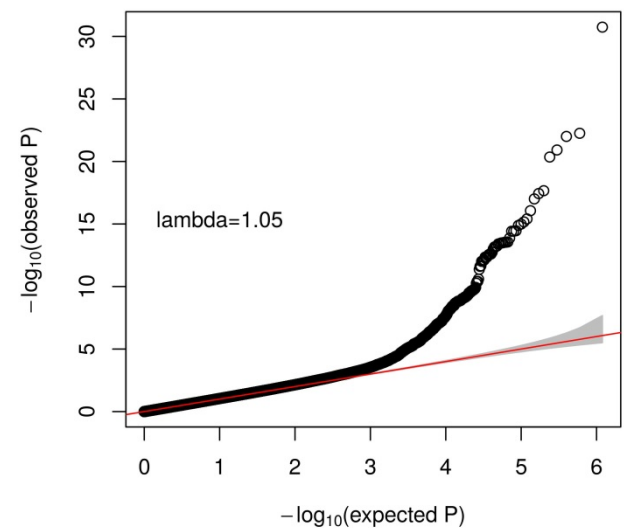
$-\log_{10}$  P-values for GENESIS and SUGEN done at FHCRC for 1195830 Genotyped SNPs with  $effN \geq 30$



GENESIS Results QQ-Plot

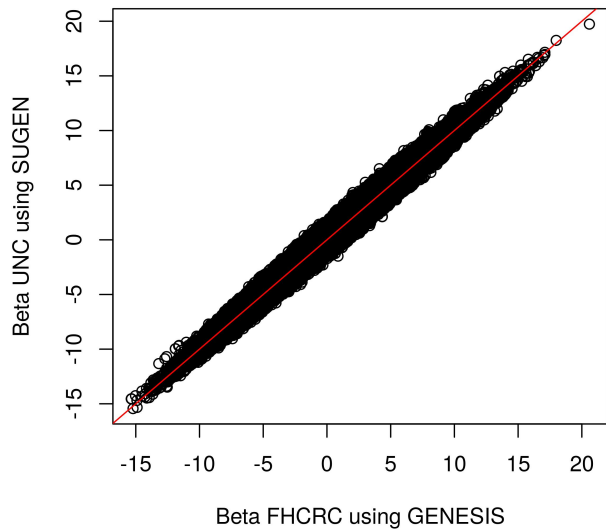


SUGEN Results QQ-Plot

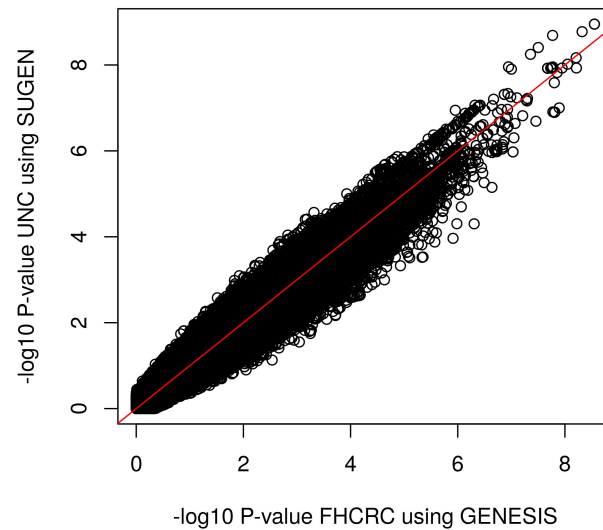


# SBP Imputed Data

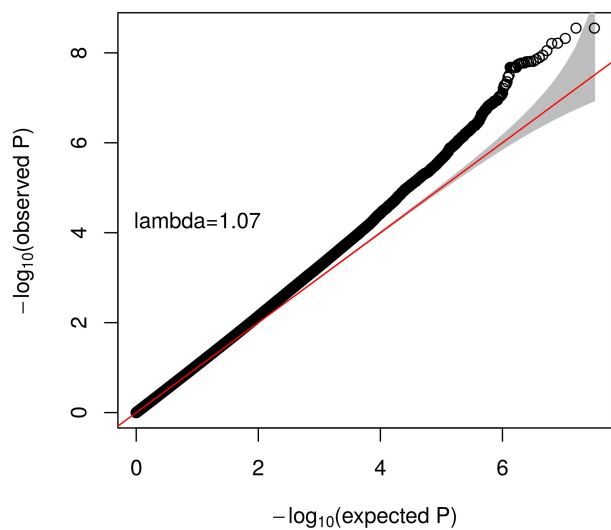
Betas for GENESIS and SUGEN  
for 32101933 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



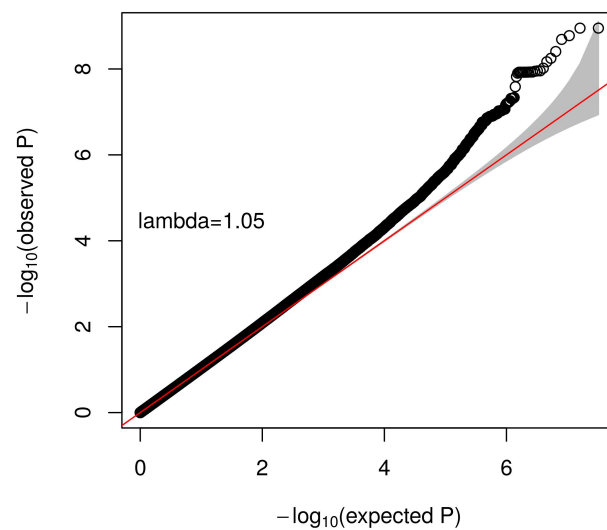
P-values for GENESIS and SUGEN  
for 32101933 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



GENESIS Results QQ-Plot

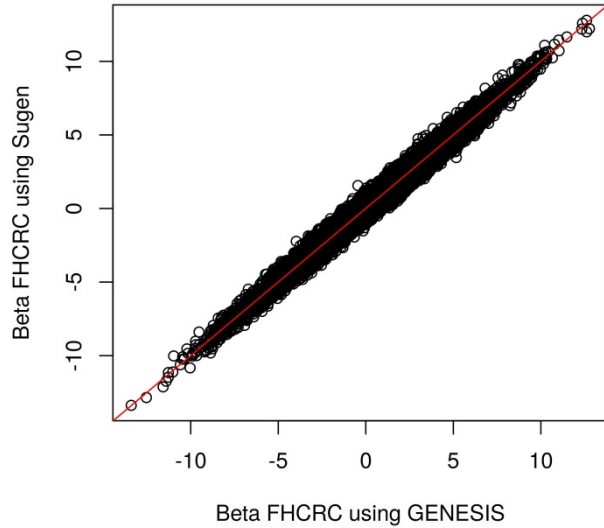


SUGEN Results QQ-Plot

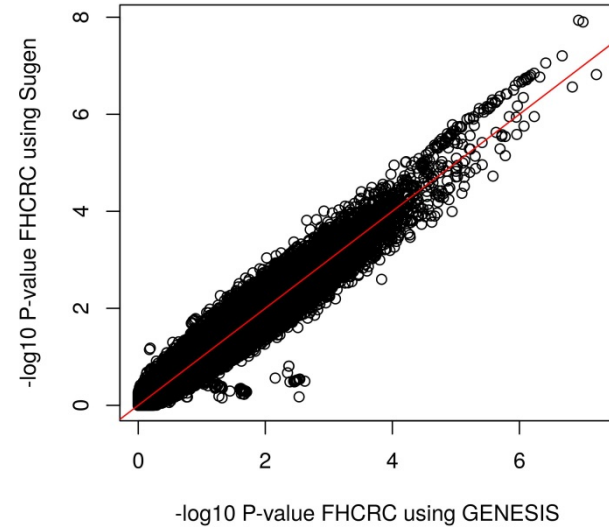


# SBP Genotype Data

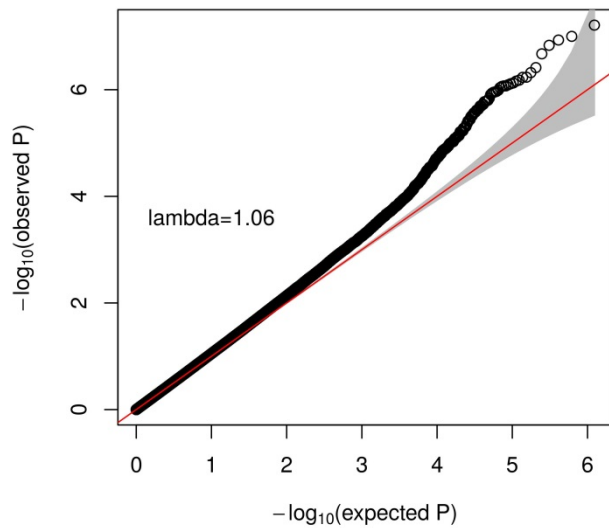
Betas for GENESIS and SUGEN done at FHCRC for 1234762 Genotyped SNPs with  $\text{effN} \geq 30$



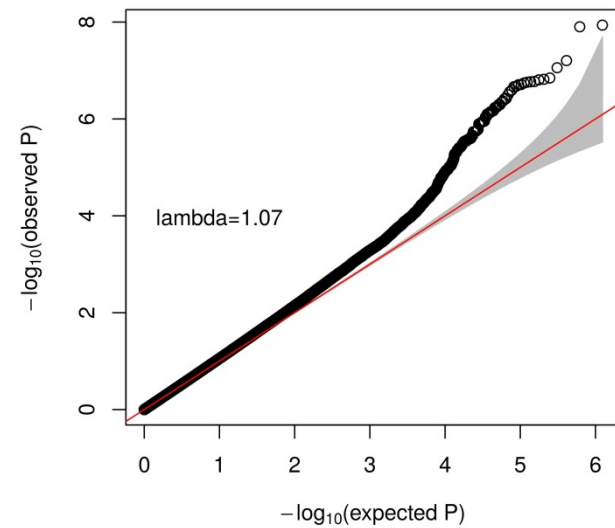
$-\log_{10}$  P-values for GENESIS and SUGEN done at FHCRC for 1234762 Genotyped SNPs with  $\text{effN} \geq 30$



GENESIS Results QQ-Plot

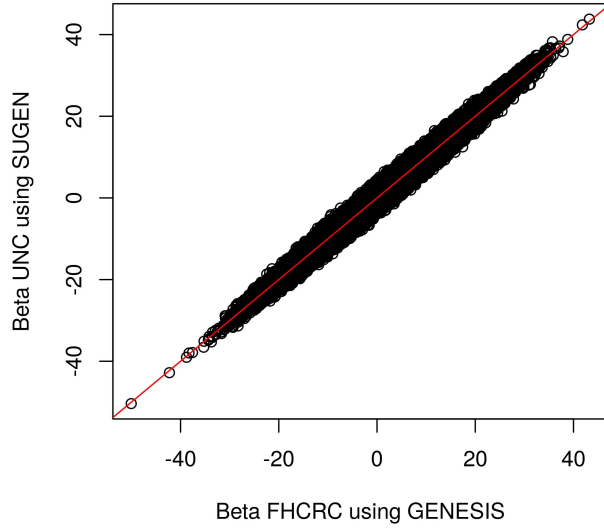


SUGEN Results QQ-Plot

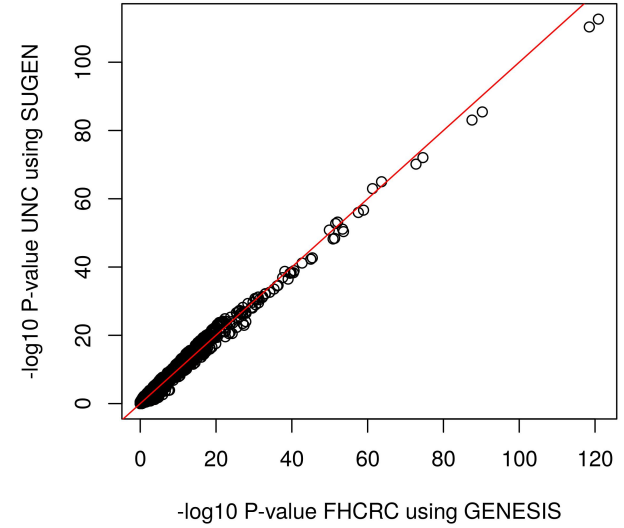


# Total Cholesterol Imputed Data

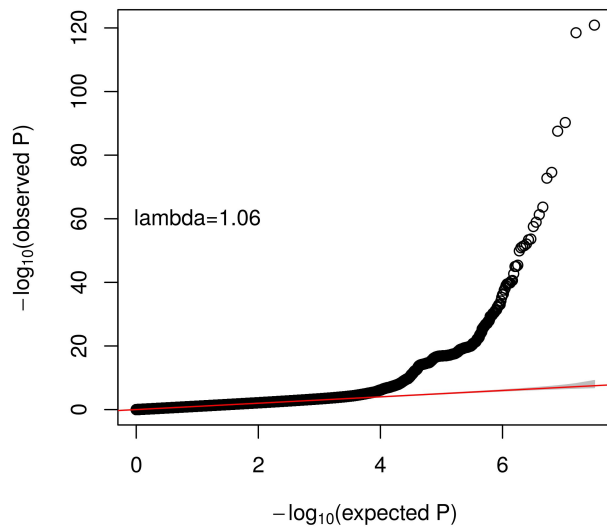
Betas for GENESIS and SUGEN  
for 32003504 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



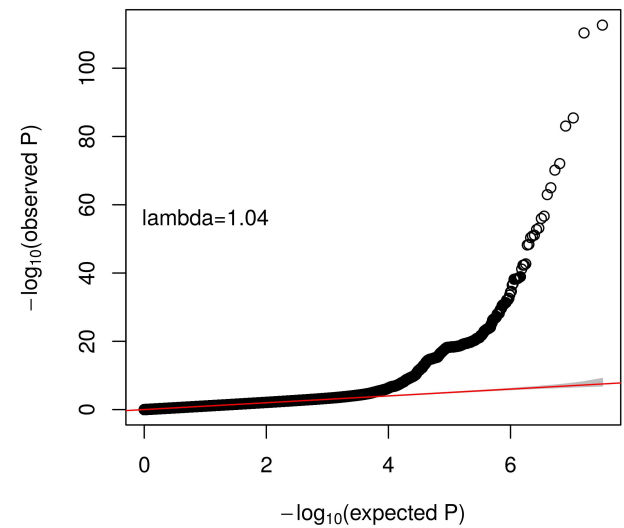
P-values for GENESIS and SUGEN  
for 32003504 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



GENESIS Results QQ-Plot

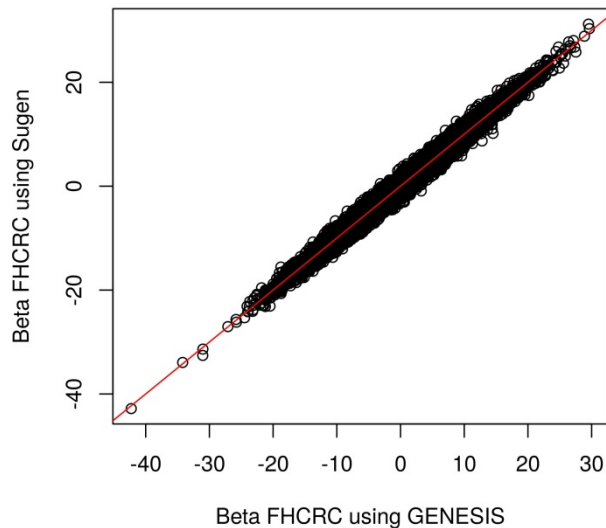


SUGEN Results QQ-Plot

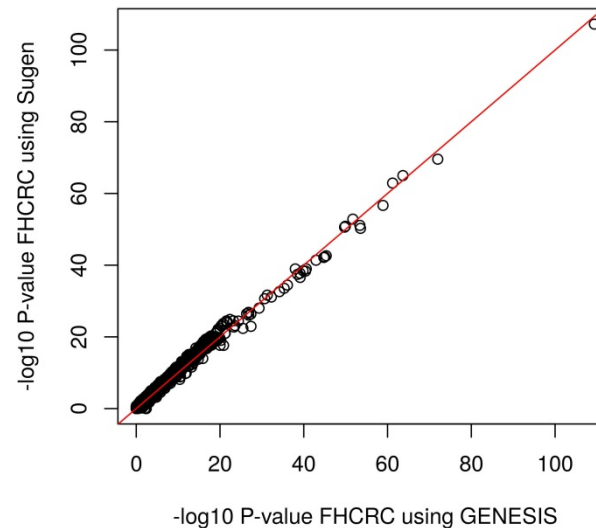


# Total Cholesterol Genotyped Data

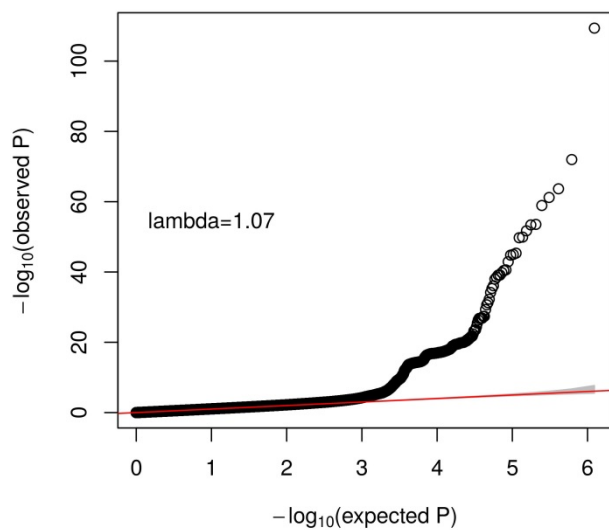
Betas for GENESIS and SUGEN done at FHCRC for 1232791 Genotyped SNPs with  $effN \geq 30$



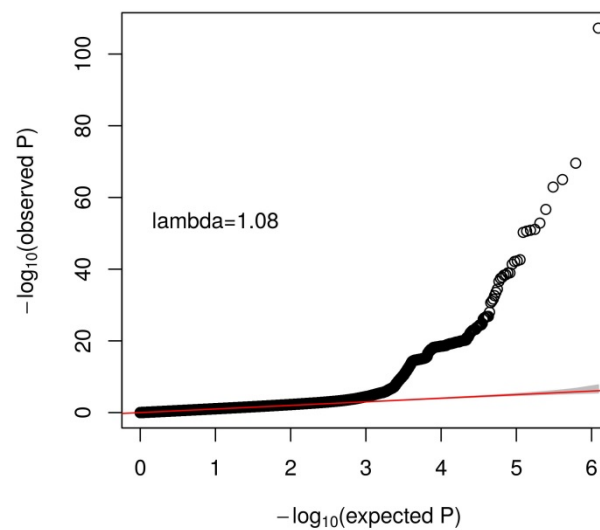
$-\log_{10}$  P-values for GENESIS and SUGEN done at FHCRC for 1232791 Genotyped SNPs with  $effN \geq 30$



**GENESIS Results QQ-Plot**

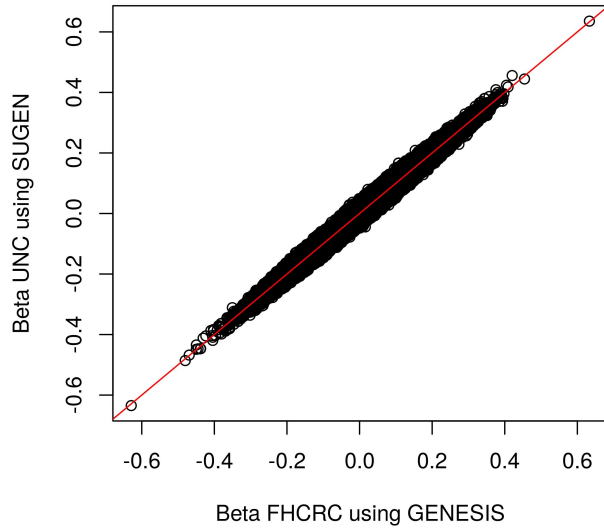


**SUGEN Results QQ-Plot**

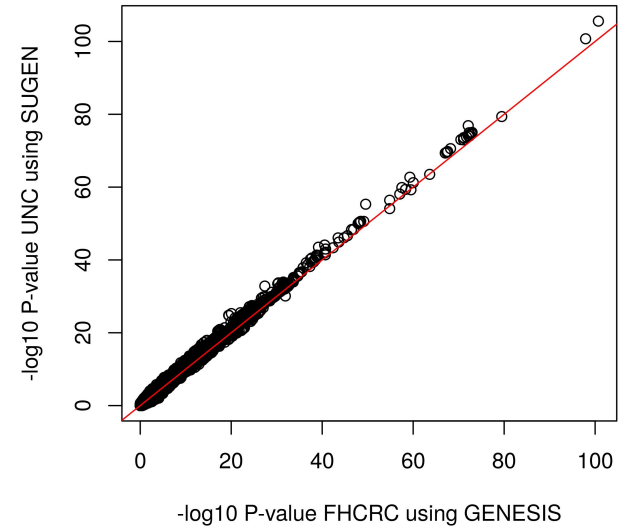


# Triglycerides Imputed Data

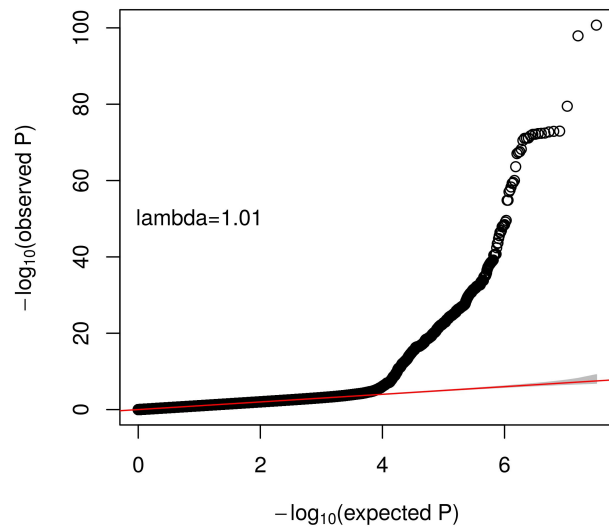
Betas for GENESIS and SUGEN  
for 31978265 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



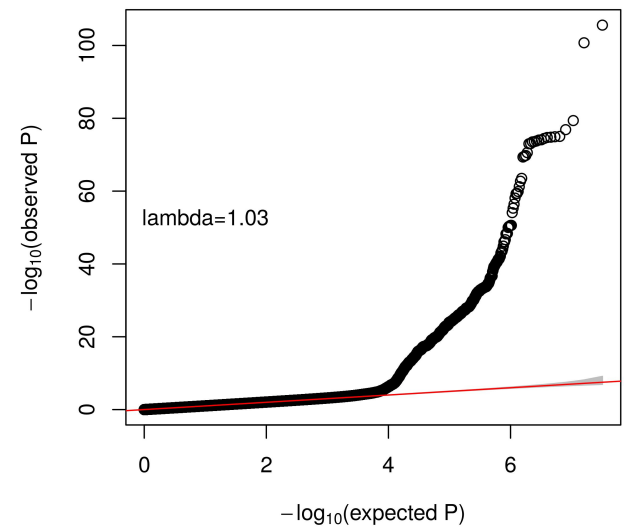
P-values for GENESIS and SUGEN  
for 31978265 Imputed SNPs with info  $\geq 0.40$  and  $\text{effN} \geq 30$



GENESIS Results QQ-Plot



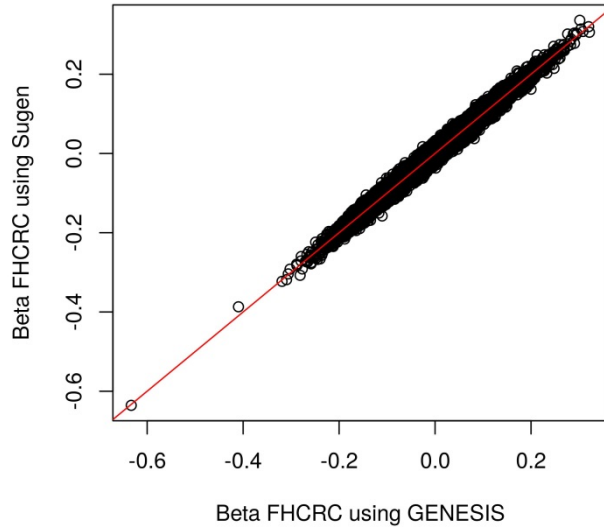
SUGEN Results QQ-Plot



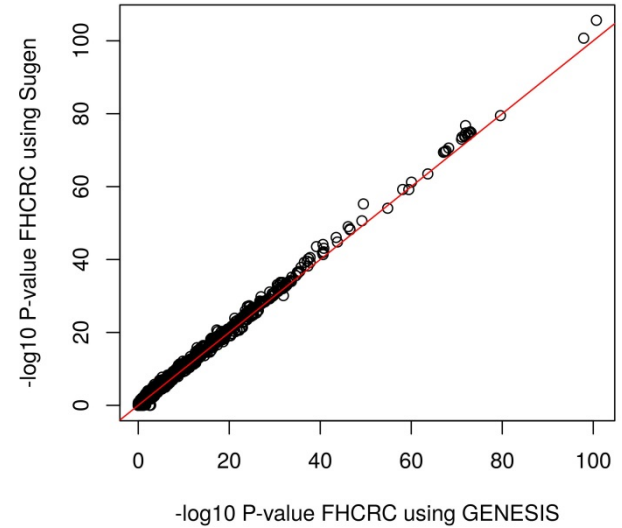


# Triglycerides Genotyped Data

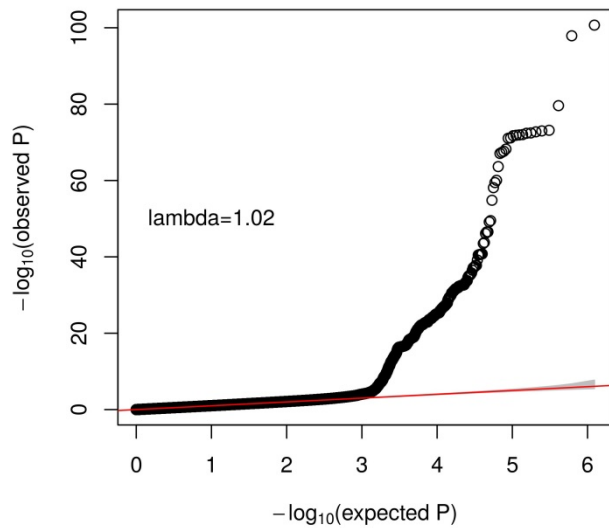
Betas for GENESIS and SUGEN done at FHCRC for 1232795 Genotyped SNPs with  $effN \geq 30$



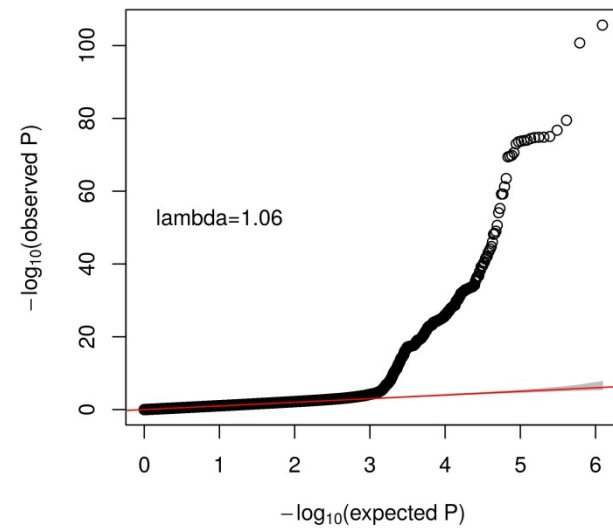
$-\log_{10}$  P-values for GENESIS and SUGEN done at FHCRC for 1232795 Genotyped SNPs with  $effN \geq 30$



GENESIS Results QQ-Plot



SUGEN Results QQ-Plot



# Conclusions

Proper control for hidden population stratification, admixture, and relatedness is paramount in the association analysis of the PAGE data.

The two adjustment methods, LMM and GEE, yield similar results.