GWAS ANALYSIS REPORT

The goal of this study was to perform a Genome-Wide Association Study (GWAS) to identify associations between genetic variants single nucleotide polymorphisms (SNPs) and a specific phenotype (fen4). The analysis was conducted using both covariate and non-covariate models to evaluate the genetic contributions to the phenotype. The covariate used in this study was BMI, calculated from height and weight data. The phenotype fen4 (unknown) represents a behavioral trait measured across different individuals.

The dataset for this analysis was prepared by calculating BMI as a covariate and filtering out missing values for both the phenotype (fen4) and the covariate (BMI). This filtered dataset was then used for the subsequent GWAS analysis.

Statistical Description of Phenotype and Covariate

The initial dataset included several phenotypes (fen1 to fen6), along with height, weight, and a calculated BMI. Summary statistics, including mean, standard deviation, minimum, and maximum values, were calculated for both fen4 and BMI. Histograms were generated to evaluate the distribution of the phenotype and covariate.

- **Phenotype (fen4)**: The histogram of fen4 indicated a non-normal, multimodal distribution with several peaks, suggesting heterogeneity in the data. The statistical summary for fen4 was as follows:
    - Mean: 42.78
    - Standard Deviation: 25.53
    - Minimum: 0.0
    - Maximum: 130.0
- **Covariate (BMI)**: The calculated BMI showed a normal-like distribution after initial data cleaning. The statistical summary for BMI was as follows:
    - Mean: 21.65
    - Standard Deviation: 3.90
    - Minimum: 14.50
    - Maximum: 40.39

GWAS Analysis

Two GWAS models were conducted:

1. **Without Covariate**: In the first model, the phenotype fen4 was analyzed without incorporating any covariate.
2. **With Covariate (BMI)**: The second model included BMI as a covariate to adjust for potential confounding effects on the association between SNPs and fen4, **if any**.

Manhattan and Q-Q Plots

- **Manhattan Plots**: The Manhattan plots for both models were generated. The plots without and with BMI as a covariate showed similar significant peaks across chromosomes, suggesting potential associations. This indicates that the associations found in both models were largely consistent, and the inclusion of BMI did not greatly affect the significance of these associations.

    - **Without Covariate**: The Manhattan plot showed multiple peaks with p-values reaching above 6 on the -log10 scale, suggesting many significant associations.

    - **With Covariate (BMI)**: The plot showed a similar significant peaks with p-values reaching above 6 on the –long10 scale, indicating a reduction in false-positive associations.

- **Q-Q Plots**: The Q-Q plots were used to evaluate the quality of the models. The model without covariates showed a slightly lower deviation from the expected line, indicating no suspicious confounder effect. The non-covariate model showed a closer fit to the expected distribution, suggesting a better model fit.

    - **Without Covariate**: The Q-Q plot does not show significant deviation from the expected line, with a genomic inflation factor (λ GC) of **1.02**, indicating very slight inflation.

    - **With Covariate (BMI)**: The Q-Q plot showed less deviation, with a genomic inflation factor (λ GC) of **1.01**, suggesting a similar fit and similar inflation compared to the non-covariate model.

Note: In both Q-Q Plots, the majority of observed SNPs were arranged along the expected line, indicating no significant confounder effect.

Choice of the Best Model

The quality of the analysis is very high for both models and does not show confounding effects. The genomic inflation factor (λ GC) is approximately 1 for both models. The covariate-adjusted model showed similar inflation, which indicates similar reliable identification of SNPs associated with fen4. For this analysis, I proceeded with the model with BMI as a covariate because it showed a slightly lesser λ GC of 0.01 when compared with the model without covariate.

Note that there was no need to perform eigen or PCA analysis for confounder effects because BMI was explicitly provided as the covariate for this task.

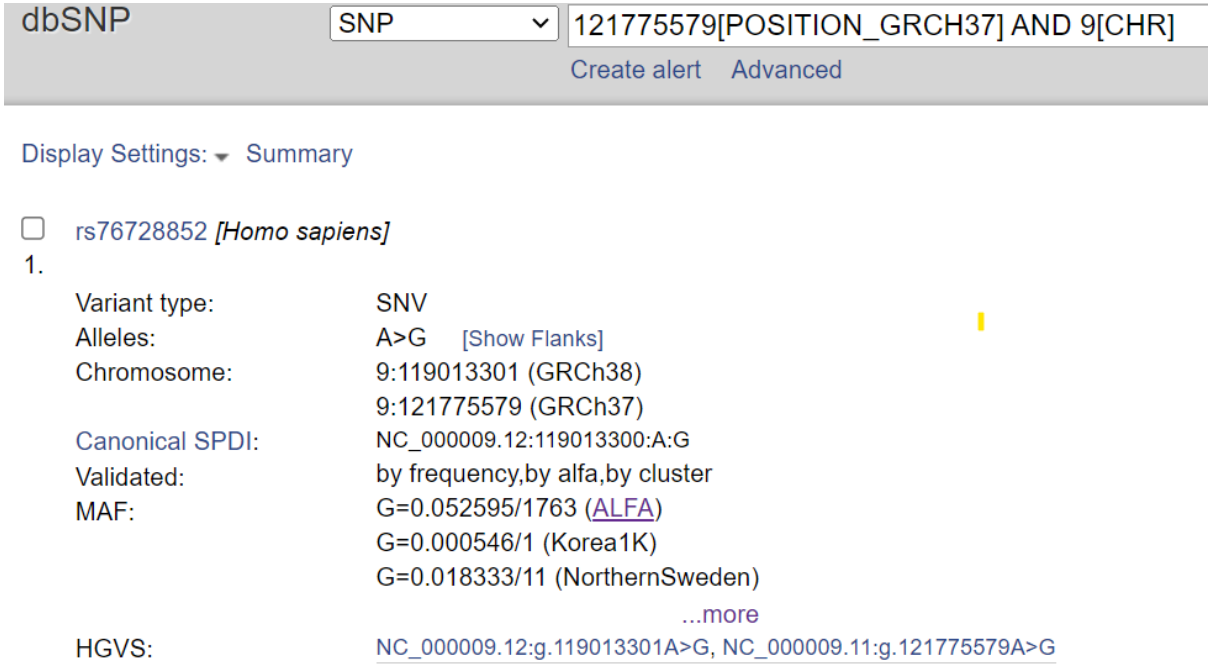
Significant SNPs and Genotype Analysis

- **Top 3 SNPs**: The three SNPs with the lowest p-values were identified from the covariate-adjusted GWAS model.

    1. **SNP1 (locus Chr9:121775579, rs76728852) shown in Fig 1.**

        - **Alleles**: [A, G]: The significant p-value of approximately 0 indicates this SNP locus is not by chance.
        - **Beta**: 30.87 (positive association): This indicates that Adenine which is the reference alleles is negatively associated with fen4 phenotype and the alternative allele (Guanine) is positively associated with fen4.
        - **Genotypes**: AG showed a higher association with fen4 compared to AA. The mean value for AA was **41.09**, and for AG it was **72.05**. GG was not available. A T-test result showed a significant difference between AA and AG homozygote with a p-value of **1.7e-07**.
        - **Literature**: No significant literature was found for this SNP.

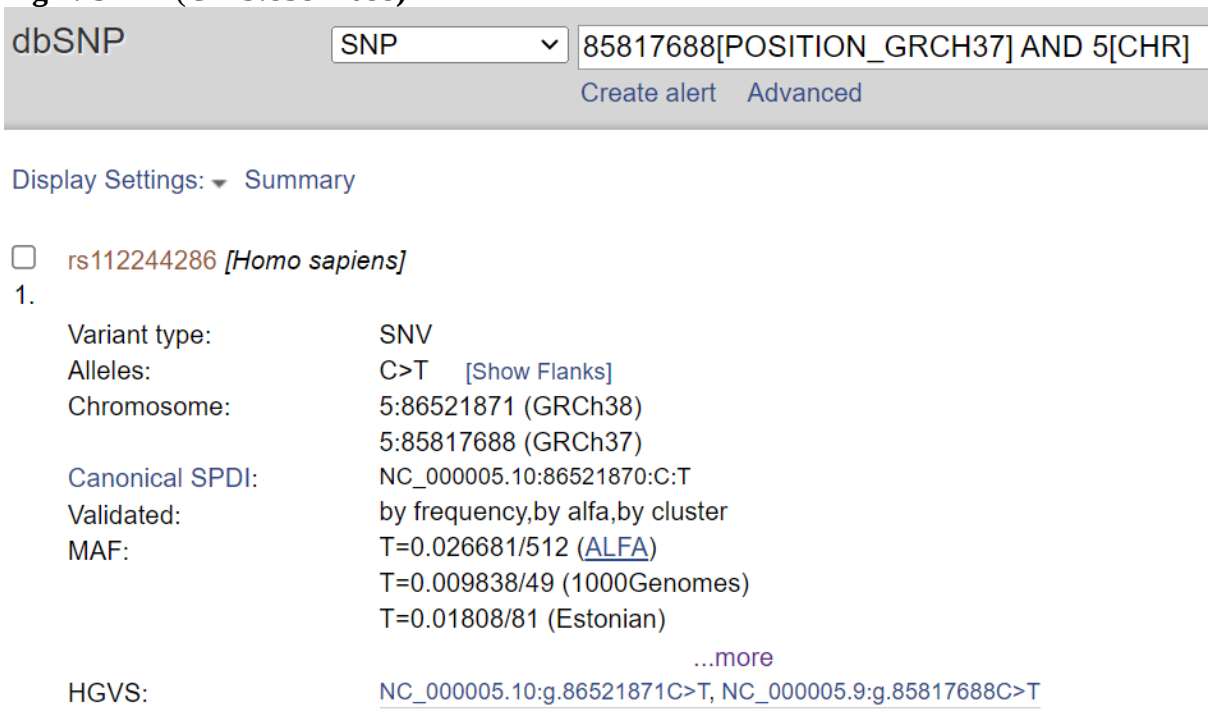
Fig 1: SNP1 (Chr9:121775579)



2. **SNP2 (locus Chr5:85817688, rs112244286) shown in Fig 2**

   - **Alleles**: [C, T]: The p-value of 1e-6 shows the SNP locus is significant and not by chance.
   - **Beta**: 34.16 (positive association): This indicates that the reference allele (cytosine) has negative association which the alternative alleles (thymine) has a positive association with fen4 phenotype represented as 0 and 1 respectively in the genotype analysis.
   - **Genotypes**: CT showed a higher mean value (**75.64**) compared to CC (**41.44**). A T-test showed a significant difference between these genotypes (p-value **5.4e-07**).
   - **Literature**: No significant literature was found for this SNP.

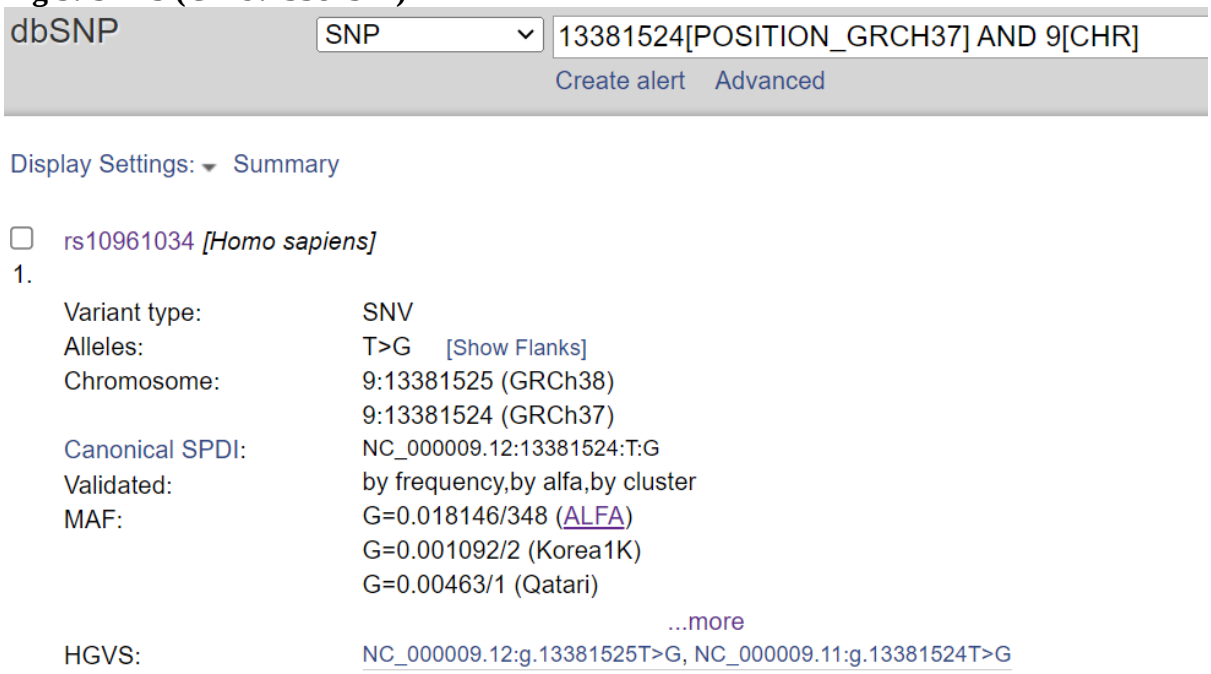
Fig 2: SNP2 (Chr5:85817688)



3. **SNP3 (locus Chr9:13381524, rs10961034) shown in Fig 3**

    - **Alleles**: [T, G]: The p-value of 1e-6 indicates that the SNP is alleles is significant.
    - **Beta**: 43.70 (positive association): This indicates that the reference allele (thymine) has a negative association while the alternative allele has a positive association with fen4 for this SNP.
    - **Genotypes**: GT showed a higher mean value (**85.63**) compared to TT (**41.76**). A T-test showed a significant difference between TT and GT with a p-value of **9.95e-07**.
    - **Literature**: No significant literature was found for this SNP.

Fig 3: SNP3 (Chr9:13381524)



Literature Review for SNPs

A literature search was conducted on the NCBI database for the top 3 SNPs. These searches revealed that although these SNPs are well-known and documented, there are no publication citations for them. However, there is other meaningful information in the NCBI database for these SNPs. For example, a click on the ALFA Allele Frequency of our third SNP (chr9:13381524) showed how this allele is distributed and researched among populations (Fig 4).

Fig 4: ALFA Allele Frequency of chr9:13381524

Release Version: 20231103111315

Search:

Population	Group	Sample Size	Ref Allele	Alt Allele
Total	Global	19178	T=0.98185	G=0.01815
European	Sub	14286	T=0.97746	G=0.02254
African	Sub	2970	T=0.9976	G=0.0024
African Others	Sub	114	T=1.000	G=0.000
African American	Sub	2856	T=0.9975	G=0.0025
Asian	Sub	116	T=1.000	G=0.000
East Asian	Sub	88	T=1.00	G=0.00
Other Asian	Sub	28	T=1.00	G=0.00
Latin American 1	Sub	154	T=1.000	G=0.000
Latin American 2	Sub	616	T=0.990	G=0.010
South Asian	Sub	98	T=1.00	G=0.00

Feedback

Summary

This GWAS analysis aimed to identify SNPs associated with the phenotype `fen4`, using both covariate and non-covariate models. The covariate-adjusted model, which included `BMI` as a covariate, was used since both models had approximately the same λ GC. I Calculated and plot mean phenotype for three different genotypes with the lowest significant locus. The three of them had no significant literature available. Future studies should further investigate the regulatory impact of these SNPs and their association with the `fen4` phenotype.