The challenge with mixed up chromosomes

- The problem was with the phenotype dataset.
- Apparently, the order of individual IDs was not the same as the one in the genotype file.
- So, after sorting the order to fit the one on the genotype file, the plots became same as the ones in GN2
- The figure below shows the resulting plot for trait 01 (heat shock) after sorting the order of individual ids in the phenotype dataset.

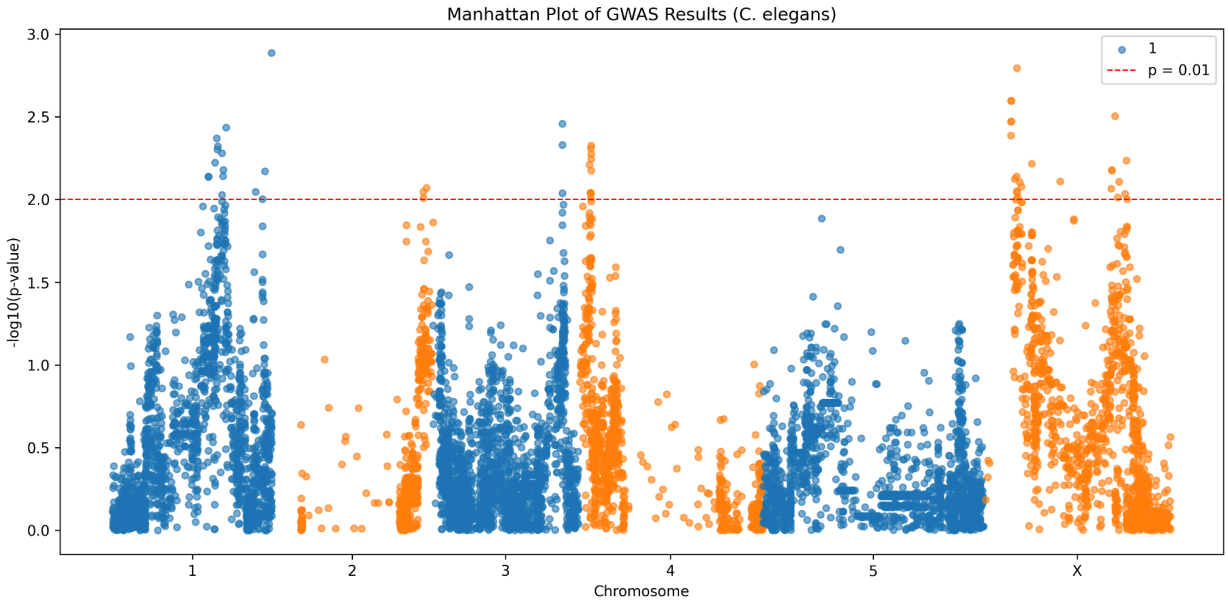


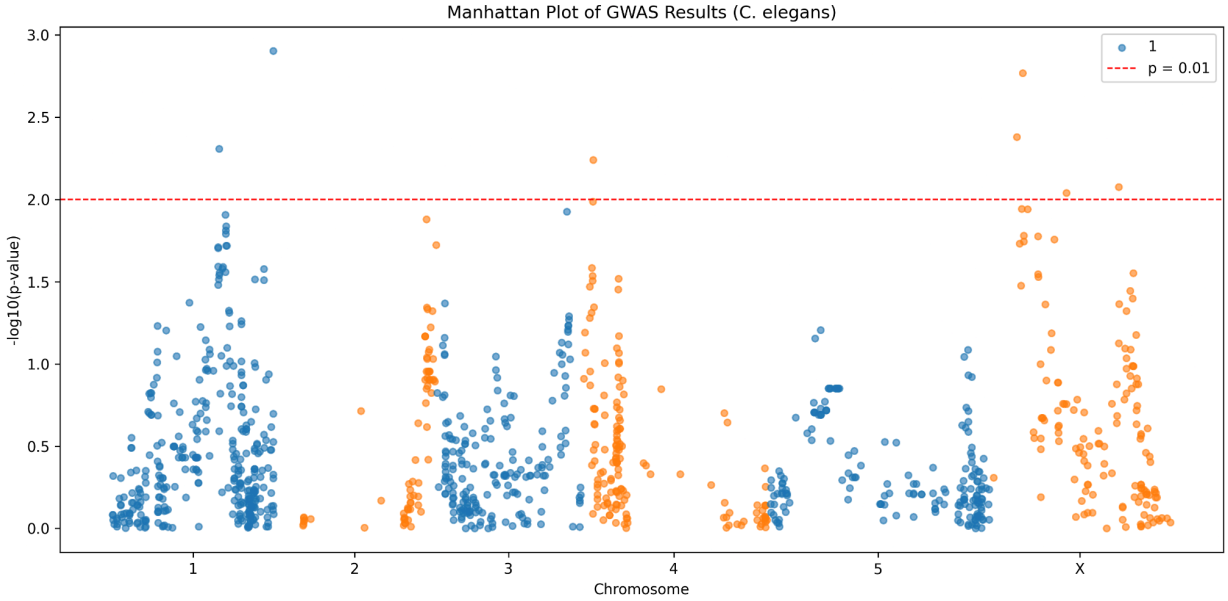
Figure01; Resulting manhattan plot with proper chromosome order as expected in GN2

Results for Genotype smoothing after adapting plink filtering algorithm

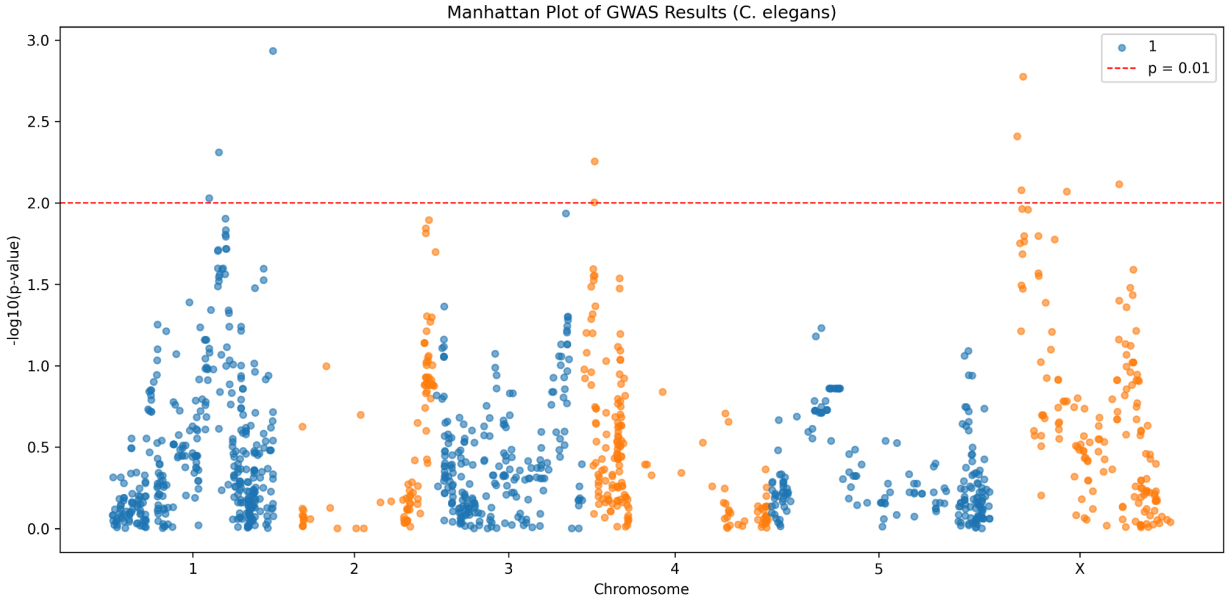
- Plink's algorithm involved the QC steps (remove markers with high missing data, remove low MAF markers, remove monomorphic markers), imputation of missing values, and then LD pruning.
- For more information, visit the following github link: C.elegans_genotype_smoothing

- The following show plots generated after smoothing the genotype file for different set of parameters (focusing on the r^2 threshold and sliding window in base pairs)

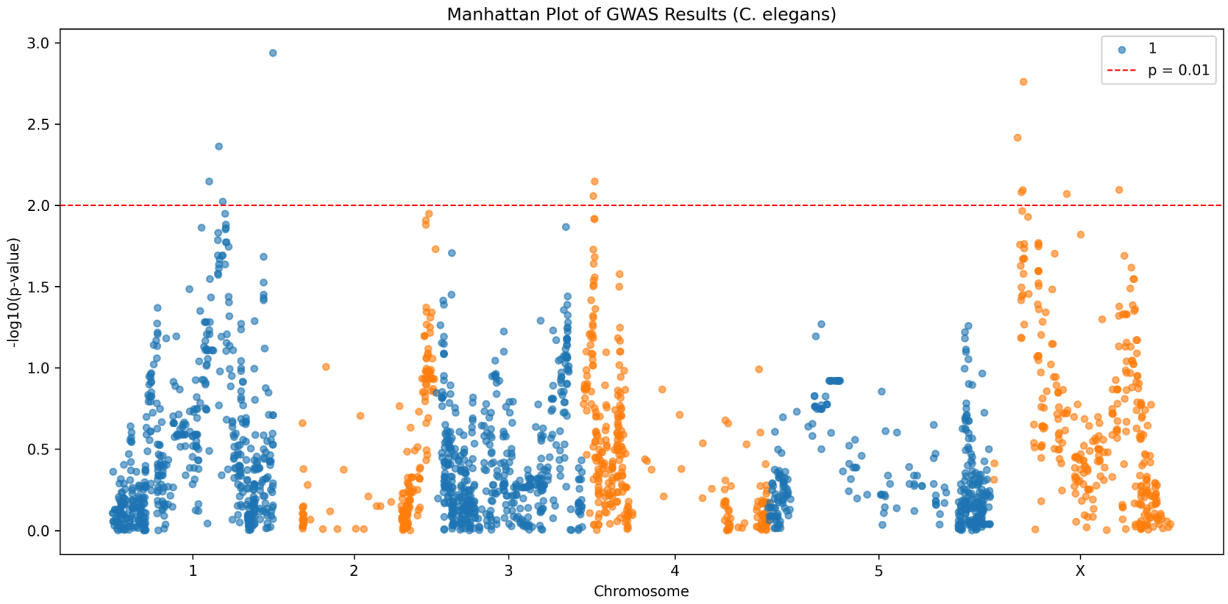
01. Figure02 (R2 = 0.8, window = 100, p value = 0.01)



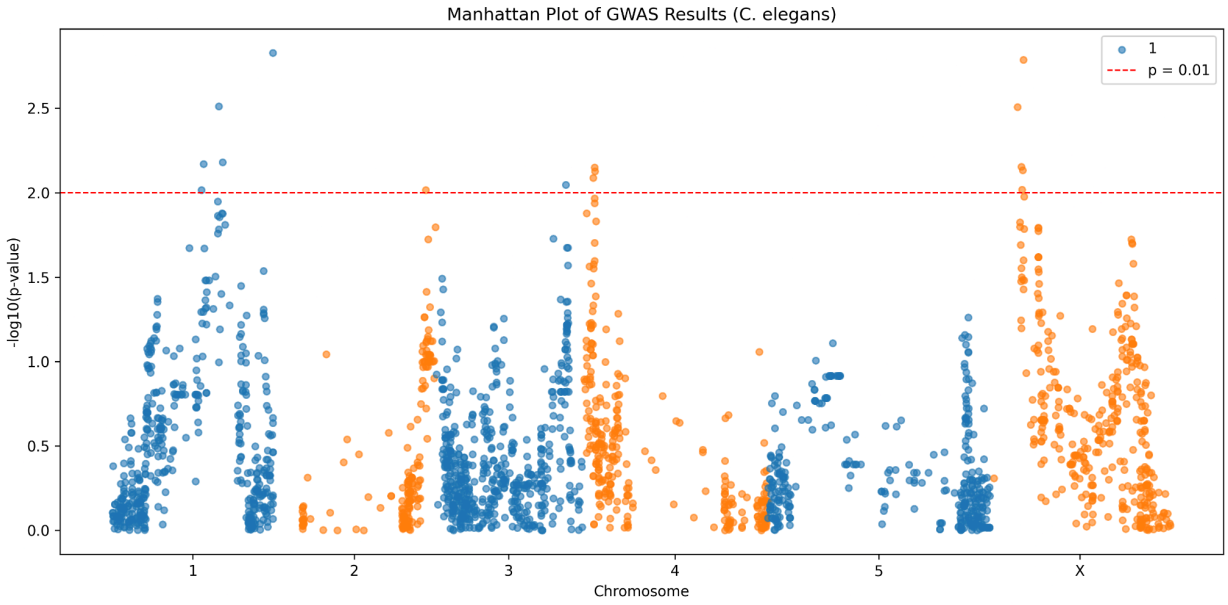
02. Figure03 (R2 = 0.85, window = 100, p value = 0.01)



03. Figure04 (R2 = 0.95, window = 100, p value = 0.01)



04. Figure05 (R2 = 0.95, window = 50, p value = 0.01)



Interpretation:

- Chromosomes 1, 4, and X have significant hits across the four smoothed plots (Figures 02 to 05)
- There's a significant decrease in the level of noise for the 4 plots as compared to the unsmoothed plot (Figure01 above), this suggests then the plink algorithm works
- Yet, more research is needed to make sure we do not lose important biological signals from the data, as we clean them.