**Selection of T2D Phenotypes**

Pishoy et al. (2021) performed a review of recent finding of phenotypes associated with type 2 diabetes (T2D) and categorised them into five groups. The major phenotypes identified were severe insulin resistance, severe insulin deficiency, mild obesity, high BMI and mild age related diabetes.

In a study by Shoily et al. (2021) the phenotypes of diabetes, cardiovascular disease, inflammation, hypertension and kidney diseases were examined, and four SNPs were found to associated: rs5186, rs1800795, rs1799983 and rs1800629.

Another study by Willer et al. (2007) examined 134 SNPs from a literature review of 120 published studies looking at SNPs associated with type 2 diabetes. They reported significant associations between type 2 diabetes and the following phenotypes: rs5210, rs5219, rs5400, rs1044498, rs1800610, rs1800795, rs1801262, rs1801282, rs2071023, rs2701175, rs3856806.

**Construction of the Database**

The database compromises of four tables as follows.

A screenshot of a computer

AI-generated content may be incorrect.

*Table SNP\_Associations*

The SNP associations table data was built using data from the GWAS catalogue. The All Associations v1.0 file (gwas\_catalog\_v1.0-associations\_e113\_r2025-01-08.tsv) was downloaded and inserted raw into a SQL database. The table was then filtered on the column DISEASE\_TRAIT for the phenotypes identified by Pishoy et al. (2021) and also for the specific SNPs identified by Shoily et al. (2021) and Willer et al. (2007). Further filtering on the column INITIAL\_SAMPLE\_SIZE for Pakistani (Punjabi), Bangladesh (Bangali) or South Asian (Telugu). Extracted just the data for the phenotypes of interest associated with type 2 diabetes and for three South Asian populations, a total of 1192 rows of data.

*Table Gene\_Associations*

Studies by Gupta et al. (2011) and Nair et al. (2010) were cross referenced to compile a list of gene associations. Particular attention was given to ensuring relevance for South Asian populations while also incorporating globally validated T2D-associated loci. A total of 40 rows of data were added to the table.

*Table Population*

Population data was downloaded from the International Genome Sample Resource website (<https://www.internationalgenome.org/home>). This data can be filtered by population (Bengali, Gugarati, Telugu, Tamil and Punjabi) to produce 661 samples, however it was decided to download all 4978 samples for all populations as this would make the application easier to expand to populations beyond those in South Asia.

**References**

Gouda, P., Zheng, S., Peters, T., Fudim, M., Randhawa, V.K., Ezekowitz, J., Mavrakanas, T.A., Giannetti, N., Tsoukas, M., Lopes, R. and Sharma, A., 2021. Clinical phenotypes in patients with type 2 diabetes mellitus: characteristics, cardiovascular outcomes and treatment strategies. Current Heart Failure Reports, 18(5), pp.253-263.

Gupta, V., Vinay, D.G., Rafiq, S., Kranthikumar, M.V., Janipalli, C.S., Giambartolomei, C., Evans, D.M., Mani, K.R., Sandeep, M.N., Taylor, A.E. and Kinra, S., 2012. Association analysis of 31 common polymorphisms with type 2 diabetes and its related traits in Indian sib pairs. *Diabetologia*, *55*, pp.349-357.

Nair, A.K., Sugunan, D., Kumar, H. and Anilkumar, G., 2010. Case-control analysis of SNPs in GLUT4, RBP4 and STRA6: Association of SNPs in STRA6 with type 2 diabetes in a South Indian population. *PloS one*, *5*(7), p.e11444.

Shoily, S.S., Ahsan, T., Fatema, K. and Sajib, A.A., 2021. Common genetic variants and pathways in diabetes and associated complications and vulnerability of populations with different ethnic origins. Scientific Reports, 11(1), p.7504.

Willer, C.J., Bonnycastle, L.L., Conneely, K.N., Duren, W.L., Jackson, A.U., Scott, L.J., Narisu, N., Chines, P.S., Skol, A., Stringham, H.M. and Petrie, J., 2007. Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. Diabetes, 56(1), pp.256-264.