**ABOUT**

With increasing understanding of genome research and the increasing statistics associated with diabetes, the need for a comprehensive diabetes genome platform is in need. “DIABETES” is a fast, efficient, and user-friendly website allowing individuals to easily extract data associated with genetic variants related to type 1 diabetes. The “DIABETES” website was developed by 4 students at Queen Mary University of London, under the supervision of Matteo and Conrad, both lecturers in bioinformatics at the establishment.

The search features of the website are linked to a database that contains broad information associated with the variants seen in type 1 diabetes across all chromosomes including the following: P-value, Relative Allele Frequency (RAF), Genome ontology, and finally Linkage disequilibrium values.

Being user friendly the website contains multiple pages that enable extraction of the specific values mentioned depending on either variant ID, chromosome location, or gene name. The website allows further comparison between up to 5 variants and extraction of a linkage disequilibrium for each variant search as well as a linkage disequilibrium heat map plot.

**SOFTWARE SCHEMATICS:**

Diagram

Description automatically generated

Figure 1: DIABETES software schematics and component integration.

Firgure 1 shows the main component and their integration with one another to produce the final functioning website. The databased was used collating data from a variety of database websites including Genome Wide Association Study (GWAS), Ensembl-BioMart, and NIH (National institute of healthcare. The main database containing information from GWAS was produced using SQLIite3.

Due to a variety of data types, each data table was connected to Flask using SQL Alchemy, this is a python toolkit designed for database accessing that considers a familiar relationship within varying databases, as opposed to a collection of tables, simply put; all tables have a consistent relationship, which in this case is the variant ID or Rs values.

Finally the website routes are defined using the HTML language,

**DATA COLLECTION.**

**Variant ID, Chromosome number, Chromosome Location, P-value:**

The following values: Variant ID, chromosome number, Chromosome Location, P-value were obtained using the website containing information on the Genome Wide Association Study (GWAS). This was easily obtained using the search icon provided, whereby Type 1 Diabetes can be entered and variants associated with the disease including the values mentioned and more can be downloaded into a Comma Separated Value (CSV). The data collected from GWAS is genome wide and so is not limited to population base, however obtaining the information within a CSV, we were able to organize the data in a chromosomal order and thus the variant IDs associated with. The organisation helped in then searching other databases for relevant information including the Relative Allele Frequency (RAF), and the Linkage Disequilibrium (LD)data.

**Relative Allele Frequency (RAF):**

Obtaining genome wide data from GWAS aided in then searching for the necessary RAF data. Obtaining the RAF data was done manually. BioMart provides community-based data which includes RAF for certain populations of continents including European sub populations, Asian sub populations, and finally African Sub populations, respectively those of Great Britain, Han Chinese, and Yoruba. The GWAS data obtained indicated alleles associated with the disease (A, T, C, G); this aided in two things, the first being an indication if the relative allele frequency was even known for that variant, and if it was what the relative frequency, because of the the RAF data collected for all chromosomes is slightly less than that of the variant IDs provided.

**Linkage Disequilibrium:**

The National Institute of Healthcare.