**MetaboBTT Database**

**What is the MetaboBTT Database?**

It is a MySQL database that stores project-specific data related to a current study in the field of genetics. The database is easily accessible and queryable by all members of the research group via a user-friendly web interface.

**Where can the database be accessed?**

The database can be accessed via a web interface available at <http://www.bioinf.wits.ac.za/software/metabobtt> and must be accessed with a username and password.

**What data does the database contain?**

The MetaboBTT Database houses phenotype, SNP annotation and association analysis data from an ongoing project focused on identifying risk factors for cardiometabolic disease in South Africans. The data is from participants and their female caregivers from the Birth to Twenty (Bt20) cohort and DNA samples were genotyped using the Metabochip.

Each individual recorded in the database has a unique Individual ID (with a suffix ‘C’ or ‘CG’ for the Bt20 participants and female caregivers, respectively).

The Bt20 cohort is a longitudinal cohort consisting of data collected at multiple time points since its inception. **PHENOTYPE DATA** currently present in the database is from the year 17/18 data collection time point for the Bt20 participants and the year 13 data collection time point for the female caregivers. Tables have been constructed for other data collection time points for the Bt20 participants (year 5, year 7, year 9/10, year 11/12, year 13, year 14, year 15, year 16, year 19, year 20) and can be populated with the data when available. Phenotype data only exists in the database for individuals with available genotype data and includes:

* Gender (1 = males, 2 = females)
* Ethnicity (2 = Africans)
* Relationship to Child (1 = Mother , 2 = Aunt, 3 = Grandmother, 4 = Sister, 5 = Other) *[only relevant to the female caregivers]*
* Caregiver’s ID *[only relevant to the Bt20 participants]*
* Age (in years)
* Height (in metres)
* Weight (in kilograms)
* Body Mass Index (BMI) (in kg/m2)
* Hip Circumference (in metres)
* Waist Circumference (in metres)
* Waist to Hip Ratio
* Average Systolic Blood Pressure (SBP) (in millimetres of mercury) *[average of the 2nd and 3rd of three readings taken]*
* Average Diastolic Blood Pressure (DBP) (in millimetres of mercury) *[average of the 2nd and 3rd of three readings taken]*
* Subtotal Body Fat (in grams)
* Subtotal Lean Mass (in grams)
* Percentage Body Fat (%)

**SNP ANNOTATION/METABOCHIP** **DATA** exists for all 196725 SNPs on the Metabochip and includes:

* Build 36 SNPID
* Build 37 SNPID
* Chromosome
* Build 36 Base Pair Position
* Build 37 Base Pair Position
* Nearest Gene *[refers to the gene(s) in which the SNP lies or is intergenic to]*
* Location Within Gene (CODING, COMPLEX, INTERGENIC, INTRON or UTR)
* Allele 1
* Allele 2
* After QC (BOTH = remained in both the Bt20 participant and female caregiver datasets after QC, BATCH1 = remained in only the female caregiver dataset after QC, BATCH2 = remained in only the Bt20 participant dataset after QC)

All **ASSOCIATION ANALYSIS DATA** for the available phenotypes under investigation are recorded and includes:

* Build 36 SNPID
* Dataset\* (BATCH1\_ALL, BATCH2\_ALL, BATCH2\_FEMALES, BATCH2\_MALES, MERGED\_ALL, MERGED\_FEMALES)
* Phenotype
* Uncorrected P-value
* Uncorrected Beta/OR
* Corrected P-value *[corrected for the covariates listed]*
* Corrected Beta/OR *[corrected for the covariates listed]*
* Covariate 1
* Covariate 2
* Covariate 3

\**The datasets available can be analysed as individual or merged datasets – results for all possible scenarios are recorded.*

*BATCH1\_ALL = yr13 female caregivers*

*BATCH2\_ALL = all yr 17/18 Bt20 participants*

*BATCH2\_FEMALES = female yr 17/18 Bt20 participants*

*BATCH2\_MALES = male yr 17/18 Bt20 participants*

*MERGED\_ALL = yr13 female caregivers and Bt20 participants merged*

*MERGED\_FEMALES = yr13 female caregivers and female yr17/18 Bt20 participants merged*

The **GENOTYPE DATA** has undergone an extensive quality control (QC) process. The data exists as cleaned/QC’ed binary PLINK format files *(.bed/.bim/.fam*) for both the Bt20 participants and the female caregivers and these can be accessed on request.

**What can users do?**

Users can access the database from the user interface to generate **summary statistics** (basic and complex counts and average/minimum/maximum) on the phenotype data, **download** relevant phenotype, Metabochip and association analysis data that match certain user-supplied criteria and get information on how to work with the genotype files in PLINK.

1. File uploads

When specifying a list of individuals/SNPs etc., the uploaded file must be a text file containing a list of Individual IDs/SNPIDs etc. each on a separate line.

1. Specifying criteria

When phenotype criteria can be specified, up to three criteria can be added (in the form <PHENOTYPE> <OPERATOR> <VALUE>).

e.g. Height (m) >= 1.6

1. Output

Summary statistics will always be printed to the screen. Phenotype, Metabochip and association analysis data download outputs can be printed to the screen (‘Print to screen’) or downloaded as a CSV file (‘Save to File’).

**Additional features**

* A reset button is present on each page to clear all previously selected fields.
* Each page has a footer with links to useful websites (NCBI and Ensembl) and the home and parent pages.

**Contact**

To request access to the database or for any queries pertaining to the database or data contained in the database can be directed to Liesl Hendry (Sydney Brenner Institute for Molecular Bioscience/University of the Witwatersrand) at lieslmaryhendry@gmail.com.