**Background and Introduction**

High blood pressure, which affects millions of people worldwide, is a major risk factor for myocardial infarction, stroke and chronic kidney disease. Approximately 9 million deaths each year are attributable to high blood pressure, including >50% of deaths from coronary heart disease and stroke [[1](#_ENREF_1), [2](#_ENREF_2)]. Genome-wide association studies (GWAS) have identified over 50 genetic loci influencing blood pressure in predominantly European populations [[3](#_ENREF_3)]. Albeit a lot of studies have been carried out in adults, only one study has focused children (<18 yrs. Age) where they have used the HapMap imputed GWAS [4]. The current study in children will focus on the 1000Genmomes or HRC imputed panels comprising greater genome coverage while also including another novel trait (heart rate). This may help us identify the underlying genetic architecture of the vascular system better compared to adults as there are very few/no confounders in terms of medication or other lifestyle factors.

We hereby propose and invite all the studies to participate in the 1000 genomes/HRC1.1 imputation based GWAS on Blood Pressure and related traits (Systolic BP, Diastolic BP, Pulse Pressure, Mean Arterial Pressure, and Heart Rate) among children.

**AIM:** This analysis plan aims to coordinate collection of meta-data for 1000 Genomes/HRC1.1 Michigan Reference Panel (HRC 1.1 Michigan preffered) imputed genome wide association studies (GWAS) for “***Childhood Blood Pressure (BP) and Related Traits***” (Systolic BP (SBP), Diastolic BP (DBP), Pulse Pressure (PP), Mean Arterial Pressure (MAP); in *mmHg*), and Heart Rate (resting; beats per minute or bpm).

We aim to carry out:

* A meta-analysis on single-marker association meta-data for all/available measures from above

Please contact **Tian Xie (t.xie@umcg.nl)** or **Tarun veer S. Ahluwalia (tarun.veer.singh.ahluwalia@regionh.dk)** if you have any questions regarding trait definitions or analyses implementation.

**General guidelines for analyses**

Each individual study will perform data quality control (QC) and analysis and provide summary results for meta-analysis.

*Outcomes*

* **Systolic blood pressure (SBP)** in *mmHg*
* **Diastolic blood pressure (DBP) in** *mmHg*
* **Pulse Pressure (PP) = SBP – DBP**
* **Mean Arterial Pressure (MAP) = (1/3 × SBP) + (2/3 × DBP)**
* **Heart Rate (HR) in beats per minute, or** bpm

*Exclusions*

* All individuals >= 18 years of age.
* Reported cases of hypertension.

*Stratified analyses*

* Ethnicity: Perform all analyses stratified by self-reported race/ethnicity (Caucasians/Europeans, African Americans, etc.).
* Gender based: Perform analyses stratifying for sex, separately in males and females as well as a combined analyses, pooling all. We may need a case-stratified analyses depending on the study design.

*Pre-analysis Sample Quality Control (QC), including but not limited to:*

* Exclude individuals with more than > 5% missing markers
* Exclude ethnicity outliers based on population clustering
* Exclude gender mismatches
* For duplicate and twin pairs, keep the individual with less missing data

**NOTES:**

* For twin studies, we recommend that data from one individual from each twin pair (preferably with a higher genotyping call rate and if the sample is a case) be used. Family based studies can run their analyses adjusting for the genetic relatedness matrix/PCs.
* We will refer to principal components (PCs), which are used to control for population stratification and other confounders. Here, we don’t distinguish between PCs generated using software like smartPCA/EIGENSOFT or components calculated using multidimensional scaling (MDS) as implemented in PLINK. For your analyses, you should include either PCs or MDS components as covariates in the models. **Principal components in studies with related individuals to be calculated in founders.**
* Limit your dataset to individuals with valid data for phenotype, genotype, covariates, and PCs.

**Analyses**

*Trait Transformation*

Transform the traits as follows:

* Calculate residuals by running this regression model for each trait separately in males and females:

Trait (SBP/DBP/MAP/PP) ~ age + height + PCs (if available) + (study specific covariates if required)

* Perform rank-based Inverse Normal transformation of these residuals as follows using the tool of choice:
* **in SAS:** proc rank data=mydata out=inv normal=blom

var &trait;

run;

* **in R:** #if you have missing data

y<-qnorm((rank(x,na.last="keep")-0.5)/sum(!is.na(x)))

* **in STATA:** pctile pvariable = variable, nquantiles(N+1) genp(percent\_variable)

gen inv\_normal\_variable=invnormal(percent\_variable/100)

(Where N is number of observations)

* Use the rank based inverse normal transformed data to run the association testing with the SNPs for each individual trait, as below:

*Models to be run:*

1. **All individuals (here you can pool the inverse normal values calculated for males and females separately)**

*Outcome (inv\_SBP/inv\_DBP/inv\_PP/inv\_MAP/inv\_HR) ~ SNP*

1. **In Males**

*Outcome (inv\_SBP/inv\_DBP/inv\_PP/inv\_MAP/ inv\_HR) ~ SNP*

1. **In Females**

*Outcome (inv\_SBP/inv\_DBP/inv\_PP/inv\_MAP/ inv\_HR) ~ SNP*

**Meta Data**

An Excel Sheet asking for meta-data summary will be provided to all participating cohorts where information on data summary, genotyping summary, author list, a brief study description and acknowledgements for each study would be required. This file can then be completed by each study group and uploaded on the details provided for data upload as “1000G.descriptive.STUDYNAME.xls”.

**Data upload info:**

Contact **Tian Xie (t.xie@umcg.nl)** when you are ready for uploading your results, then you will receive a Sharefile link for results upload.

Contact **Tian Xie (t.xie@umcg.nl)** for result upload details or issues related to it.

**Data upload format for single variant analyses:**

|  |  |
| --- | --- |
| **MARKER** | Variant identifier of marker analyzed in the format of CHR:POSITION:SNP|INDEL, eg. 1:12345:SNP or 10:54321:INDEL. Position should be reported in build hg19. Do not code chromosome with prefix “chr”. |
| **RS** | rs number if available (eg. rs884808). “-“ if not available. |
| **STRAND** | A single character ‘+’ or ‘-‘ ; strand on which the alleles are reported (preferably ‘+’). |
| **CHR** | Chromosome (autosomes: 1-22). Do not code code chromosome with prefix “chr”. |
| **BUILD** | 37.1 proffered |
| **POS** | Base position of the variant in build hg19. |
| **N** | Positive integer; the effective number of subjects analyzed. |
| **EFFECT\_ALLELE** | The upper case character string of the effect allele, e.g. “A”, “C”, “T”, “G” for single nucleotide variants, and the 1000g allele string for INDELs (e.g. “ATCCG”). |
| **OTHER\_ALLELE** | The upper case character string of the other allele, e.g. “A”, “C”, “T”, “G” for single nucleotide variants, and the 1000g allele string for INDELs (e.g. “ATCCG”). |
| **EAF** | Effect allele frequency (range 0-1). |
| **BETA** | Estimates of effect size. |
| **SE** | Estimated standard error on the estimate of the effect size. |
| **P** | Significance of the variant association, uncorrected for genomic control. |
| **P\_HWE** | Only for genotyped data; Exact HWE *P* value for the samples analyzed. |
| **CALLRATE** | Only for genotyped data; Call rate of this marker across all subjects. Perfectly genotyped (100%) data will have call rate=1.00. |
| **INFO\_TYPE** | Only for imputed data; code indicating the type of data in the ‘Information’ column (i.e. type of imputation and analyses software used)   1. For directly genotyped markers and if the marker was not tested by using imputation or genotyping uncertainity/doses 2. For MACH imputed marker 3. For SNPTEST imputed marker 4. For a PLINK imputed marker 5. For QuickTest based ‘Information’ column |
| **INFO** | Only for Imputed data; Numeric; Corresponds to a value (0-1) where PLINK generated values can exceed 1. The following are the info to be put under the ‘Information’ column and correspond to the type of imputation software used (stated under ‘Information\_type’)   * ‘*r2\_Hat*’ from MACH2DAT/MACH2QTL (if ‘Information\_type’ was 1) * ‘*proper-info*’ from SNPTEST (if ‘Information\_type’ was 2) * ‘*INFO*’ from PLINK (if ‘Information\_type’ was 3) * ‘*r2hat*’ from QUICKTEST (if ‘Information\_type’ was 4) |

**File naming convention**

Please use the following naming scheme:

STUDY.TRAIT.ANALYSES.ETHNICITY.DATE.INITIALS.txt

* STUDY is a short (14 characters or less) identifier for the population studied.
* TRAIT pertains to the acronym for the trait analyzed: **SBP**, **DBP**, **PP**, **MAP and HR**
* ANALYSES pertains to models: “POOLED”, “MALES”, or ”FEMALES”.
* ETHNICITY: European (EA), European American (EA), African American (AA).
* DATE in format ddmmyy.
* INITIALS (analyst).

Eg. COPSAC\_SBP\_POOLED\_EA\_240617\_TSA.txt

COPSAC\_PP\_CASES\_EA\_240617\_TSA.txt

COPSAC\_MAP\_MALES\_EA\_240617\_TSA.txt

**Note:** If applicable, please compress your file with gzip before uploading.

CONTACT **Tian Xie ([t.xie@umcg.nl](mailto:t.xie@umcg.nl)) or Tarun Ahluwalia (**[**tarun.veer.singh.ahluwalia@regionh.dk**](mailto:tarun.veer.singh.ahluwalia@regionh.dk)**)**, if you have any questions regarding data analyses or methodology.

**References**

1. Imano H, Kitamura A, Sato S, Kiyama M, Ohira T, Yamagishi K, Noda H, Tanigawa T, Iso H, Shimamoto T: **Trends for blood pressure and its contribution to stroke incidence in the middle-aged Japanese population: the Circulatory Risk in Communities Study (CIRCS)**. *Stroke; a journal of cerebral circulation* 2009, **40**(5):1571-1577.

2. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, Amann M, Anderson HR, Andrews KG, Aryee M *et al*: **A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010**. *Lancet* 2012, **380**(9859):2224-2260.

3. Kato N, Loh M, Takeuchi F, Verweij N, Wang X, Zhang W, Kelly TN, Saleheen D, Lehne B, Mateo Leach I *et al*: **Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation**. *Nature genetics* 2015, **47**(11):1282-1293.

4. Parmar PG, Taal HR, Timpson NJ, et al. **International Genome-Wide Association Study Consortium Identifies Novel Loci Associated With Blood Pressure in Children and Adolescents.** Circ Cardiovasc Genet. 2016 Jun;9(3):266-78.