**Analysis Plan: Genome Wide Search for Parent of Origin Effects on Birth Weight**

**Phenotype Preparation**

Exclusions:

* Gestational Age < 37 weeks (keep >= 37 weeks)
* Multiple Births
* Congenital Anomalies, where known
* Stillbirths
* Outliers with BW Z-score >4 SD or <-4SD

**Transformation of outcome variable:**

BW Z-score = (value−mean)/standard deviation

Please note in a README file accompanying your results if you were unable to make one or more of the exclusions (e.g. if gestational age was unavailable).

**Please send the boxplot of BW Z-score along with your results**

**Genotype Preparation**

Please ensure relatedness between mother and child in each pair is as expected (i.e. if you have not already done so, please use the GWAS genotypes to check that the mother-child pairs are indeed compatible and not mismatched samples). The program, “King” is useful for this. (URL: http://people.virginia.edu/~wc9c/KING/). Please email us (addresses below) if you need any advice.

Please use 1000Genomes imputed dosages. From these generate the maternally and paternally inherited alleles. Perl scripts are provided to generate maternally and paternally inherited alleles from minimac- or impute2-imputed genotypes. Currently these scripts assume mother and child have identical identifiers. If your study has different identifiers for mother and child please either modify the script or alternatively if you provide the format of mother and child identifiers we can modify the scripts for you (email addresses below). For more information on the scripts please see enclosed README file. Alternatively generate these alleles manually (details below).

**For minimac imputed dosages**

The included perl script will generate a Phenotype file from phenotypes provided.

To make allele dosages and phenotype file:

./poe\_generator\_v0.1 -child chr1\_child.mldose -child\_sample chr1\_child.info \

-mother chr1\_mother.mldose -mother\_sample chr1\_mother.info -chr 1 \

-minimac -pheno child\_BW.ped -out chr1

where child\_BW.ped has a header line and 2 columns; child ID and phenotype.

**For impute2 imputed dosages**

Phenotype files will need to be generated manually.

To make allele dosages:

./poe\_generator\_v0.1 -child chr1\_child.gen.gz -child\_sample chr1\_child.sample \

-mother chr1\_mother.gen.gz -mother\_sample chr1\_mother.sample -chr 1 \

-pheno child\_BW.ped -out chr1

***NB If you have phased imputed data (output using the option -phase) please add the option -phased.***

**Non-minimac/impute2 users - to manually generate the maternal and paternal alleles**

We suggest using SNPTEST to run the association tests below since there will be significantly different numbers of individuals for each SNP and mach2qtl doesn’t report N for each individual SNP/INDEL. This is because in the case of both mother and fetus being heterozygous for a SNP it is not possible to be to determine which allele is inherited from which parent. Also it is possible to have missing values due to incompatible genotypes between mother and fetus (ie. mother heterozygous for “A” and fetus heterozygous for “a”).

Please create three sets of files, one containing maternal allele for all individuals, one containing paternal allele for all individuals and one containing a variable coded as 0 for those individuals in row 2 of the table below, 1 for those individuals in row 8 of the table and missing otherwise.

To generate the maternally and paternally inherited alleles:

• convert imputed probabilities to best-guess genotypes (you can do this using eg. fcGENE)

• assign maternal and paternal alleles according to the following table. Code these as “A” where the reference allele is inherited and “a” where the alternate allele is inherited. To use SNPTEST to run these analyses we recommend coding A as “1 0 0” and a as “0 1 0”. The SNPTEST results can then be interpreted as usual gwas results with the exception of maf which is half the actual minor allele frequency.

|  |  |  |  |
| --- | --- | --- | --- |
| Mother’s Genotype | Child’s Genotype | Maternal Allele | Paternal Allele |
| AA | AA | A | A |
| AA | Aa | A | a |
| AA | aa | Missing | Missing |
| Aa | AA | A | A |
| Aa | Aa | Missing | Missing |
| Aa | aa | a | a |
| aa | AA | Missing | Missing |
| aa | Aa | a | A |
| aa | aa | a | a |

Table 1: Mother’s genotype is the best guess genotype generated from imputed data from the mother, Child’s genotype is the best guess genotype generated from imputed data from the fetus, Maternal allele is the maternally inherited allele of the fetus, Paternal allele is the paternally inherited allele of the fetus. A represents allele 1, a represents allele 2. **SNPTEST uses allele 2 (labeled as alleleB in SNPTEST output) as the effect allele.**

NB Technically, there should not be mother = “AA” and child = “aa” (or vice versa) because this would be incompatible with a parent-offspring relationship. However, in an imputed dataset, there will inevitably be a small number of poorly imputed genotypes, which will lead to the above scenarios. These should be coded as “missing”, as shown in the table.

**Analysis**

To run the analysis we suggest using SNPTEST to allow easy reporting of N of tested genotypes for each SNP/INDEL. The provided scripts for minimac or impute2 imputed genotypes will generate files in the correct format.

Run analyses separately for maternally and paternally inherited alleles against Birth Weight.

Please use the option -use\_raw\_phenotypes when running SNPTEST and set the -lower\_sample\_limit to 10 due to missing genotypes of double heterozygous pairs.

Example SNPTEST command:

snptest\_v2.5 -data chr1.father.gz BW.pheno -o chr1.father.out -frequentist 1 \\

-method expected -use\_raw\_phenotypes -cov\_all -lower\_sample\_limit 10 -pheno BW

where chr1.father.gz is the genotype file, BW.pheno is the phenotype file, chr1.father.out is the output filename and BW is the name of the phenotype in the phenotype file.

There are 3 analyses to be run using the regression model

BW Z-score ~ SNP or INDEL + SEX + GEST\_AGE

Where

SEX = sex of the offspring

GEST\_AGE = gestational age at birth

SNP or INDEL is:

1. The allele inherited from the father
2. The allele inherited from the mother
3. 0 for row 2 of the above table, 1 for row 8 of the above table, missing otherwise (testing if there is a difference in phenotype between those inheriting allele a from mother and those inheriting a from father).

All use the same regression model described above:

BW Z-score ~ SNP or INDEL + SEX + GEST\_AGE

Analysis 1:

Testing alleles inherited from father in the regression model. For genotypes generated using the provided scripts please use the file ending .father.txt. For manually generated genotypes use genotypes in column 4 of the table coding allele A as homozygous for reference allele (1 0 0 in SNPTEST) and allele a as heterozygous (0 1 0 in SNPTEST).

Analysis 2:

Testing alleles inherited from mother in the regression model. For genotypes generated using the provided scripts please use the file ending .mother.txt. For manually generated genotypes use genotypes in column 3 of the table coding allele A as homozygous for reference allele (1 0 0 in SNPTEST) and allele a as heterozygous (0 1 0 in SNPTEST).

Analysis 3:

Testing for differences in phenotype between the two groups of children with genotype Aa; comparing those inheriting allele a from mother and allele A from the father with those inheriting allele A from the mother and allele a from the father. For genotypes generated using the provided scripts please use the file ending .hets.txt. For manually generated genotypes code those in row 2 of the table as homozygous for reference allele (1 0 0 in SNPTEST) and those in row 8 of the table as heterozygous (0 1 0 in SNPTEST).

**Required format of results**

The submitted data should be formatted as tab-delimited text files. **SNP positions should be provided in hg19**. Please code missing values in any column with “NA”.SNP names should be provided in the format chr:pos. **Please note in particular the note below about reporting “N” and “EAF”.**

Filenames should be formatted STUDY\_ANALYSIS\_IMPUTATION\_DATE.gz where STUDY is the study name, ANALYSIS indicates which of analyses 1-3 above were carried out (FATHER, or MOTHER, or HETS), IMPUTATION is the imputation method used, DATE is the date on which the file was prepared. eg. ALSPAC\_MOTHER\_IMPUTE2\_281015.gz

Please also send the boxplot of BW z-score and any accompanying information in a README file, labelled with the study name and date, e.g.

ALSPAC\_BOXPLOT\_281015.pdf

ALSPAC\_README\_281015.txt

 

**Column headings and descriptions**

|  |  |
| --- | --- |
| CHR | Chromosome |
| POS | Base pair position |
| SNP | Markername of SNP or INDEL |
| STRAND | The strand on which the alleles are reported: we request SNPs be aligned to the forward (+) strand. State the strand as a single character: + or - |
| N | Sample size for the SNP. In SNPTEST this is the sum of the 3 columns all AA, all AB and all BB, **NOT the column all total** |
| EFFECT\_ALLELE | Allele for which the effect (BETA) is reported (alleleB in SNPTEST) |
| NON\_EFFECT\_ALLELE | Second allele at the SNP |
| EFFECT\_ALLELE\_FREQUENCY | Allele frequency of the allele for which the effect (BETA) is reported. Note: Effect allele frequency can be calculated as (all\_AB)/(all\_AB + all\_AA) |
| BETA | Effect size |
| SE | Standard error of the estimate of the effect size |
| P\_VAL | P Value associated with the effect size |
| QUALITY | Imputation Quality |

**Contacts and deadlines**

Analysis query should be directed via e-mail to the following:

Robin Beaumont: r.beaumont@exeter.ac.uk

Rachel Freathy: r.freathy@exeter.ac.uk