PopPAnTe: Population and Pedigree Association Testing of Quantitative Data

Users' Guide

1 Background

PopPAnTe is a user-friendly Java program that enables association studies of quantitative data in related samples. Relationships between individuals can be either described by known family structures of any size and complexity, or by genetic similarity matrices (GSM) inferred from genome-wide genetic data. This approach is particularly useful when some degree of hidden relatedness (including population stratification) is expected, but extensive genealogical information is missing or incomplete, thus facilitating the usage of biobank collections from population isolates.

PopPAnTe analyses quantitative data in a variance component framework in order to model the resemblance among individuals. In fact, when dealing with related samples it is necessary to model the resemblance among family members that could be determined by polygenic effects. Therefore, the association between variables is calculated using a mixed model where the polygenic relationship between individuals is modelled as a random component. When genealogical information is available, PopPAnTe evaluates the relatedness matrix from the known pedigree relationships. When such information is not available, as in population-based analyses, a GSM can be inferred by using genome wide genotype data. PopPAnTe implements an inheritance and an association model. Let the association model be described as

$$r_i = \mu + \beta p_i + g_i + e_i$$

where r_i represents the values of the dependent quantitative variable (*response*, for instance transcript levels or quantitative phenotype) of the i-th individual, μ is the response mean, β is the estimate of the independent quantitative variable (*predictor*, for instance DNA methylation or metabolite level) p_i effects and g_i and e_i are the polygenic and environment effect, respectively. This linear mixed model can be extended to include covariates as fixed effects, as

$$r_i = \mu + \beta p_i + \Sigma_j \beta_j F_{ij} + g_i + e_i$$

where β_j is the estimate of the j-th fixed effect F.
Let the inheritance model be described as

$$p_i = \mu + g_i + e_i$$

where p_i is the value of the dependent variable for the i-th individual, μ is its mean and g_i and e_i are the polygenic and environment effect, respectively. This linear mixed model can also be extended to include covariates as fixed effects, and data can be described as

$$p_i = \mu + \Sigma_j \beta_j F_{ij} + g_i + e_i$$

In both models, the variance covariance matrix can be calculated as

$$\omega = 2\Phi\sigma_g^2 + I\sigma_e^2$$

where 2Φ is the matrix of the expected proportion of alleles shared IBD over the genome between each pair of individuals, I is the identity matrix, and σ_a^2 and σ_e^2 are the additive genetic and environmental variance, respectively. When the family structure is known PopPAnTe evaluates the kinship matrix Φ internally, on the basis of the theoretical kinships extracted from the know pedigree relationships, and assuming pedigree founders as unrelated [6]. When the family structure is not available, the kinship matrix can be estimated from genome-wide genetic data with any of several well-established tools (such as PLINK [1], GCTA [2], or LDAK [3]) and used as input for PopPAnTe.
The random effects are assumed to follow a multivariate normal distribution with zero mean and numerical procedures can be used to estimate the variance component parameters σ_a^2 and σ_e^2 and the likelihood of the data.

The heritability value is calculated as

$$h^2 = \frac{\sigma_a^2}{\sigma_e^2 + \sigma_a^2}$$

When the relatedness matrix is evaluated by PopPAnTe using the known pedigree relationships, it is ensured that this results in a variance-covariance matrix that is usually both symmetric and semi-positive definite. Therefore, PopPAnTe assess the maximum likelihood estimates of the variance components through efficient Cholesky decomposition. However, when the kinship matrix is provided in input, the property of positive-definiteness may not hold. In this case, a bending procedure [7] is used by default to transform the matrix when it is not positive semi-definite. The user has also the option to use a LU decomposition instead. Additionally, PopPAnTe implements the QR decomposition to solve the rare cases where the variance-covariance matrix is not invertible and neither the Cholesky nor the LU decompositions can be used.
When predictors can be ordered in space, as it is for instance for gene expression or epigenetic markers, PopPAnTe allows the computation of region-based association tests by gathering information from flanking predictors included in a sliding window of user-defined size, whose values are replaced with their first principal component.

PopPAnTe assesses the significance of the results both through formal likelihood testing and through an empirical procedure. In fact, a likelihood ratio test between a model where the intensity signal is modelled in the fixed effect $L(\mu, \sigma_a^2, \sigma_e^2, \beta|r, P)$ – where P is a matrix of the independent values – and a model where its effects are assumed to be equal to 0, namely $L(\mu, \sigma_a^2, \sigma_e^2, \beta = 0|p, P)$ can be than used to asses the significance of the association between the intensity signal and the quantitative value.
PopPAnTe also allows the evaluation of empirical p-values by randomly permuting the independent values (or the window-specific principal component) among subjects and reassessing the association under the null hypothesis. When family structure is provided as input, values are randomly permuted among family members. To speed up the performance PopPAnTe implements an adaptive permutation approach [4], stopping the generation of randomly permuted samples either when *a)* a number of successes s is reached, s being chosen such that desired level of significance α will be sampled with a 68% CI (about 1 standard error) and contained within a specified level of precision π, or (to ensure the termination in a finite number of iterations) when *b)* a number of iterations i is reached, i being chosen as a function of α and π.

Quantile normalisation can be automatically applied to each variable to improve normality of the response variables and of the predictors. Moreover, PopPAnTe implements two approaches to correct the association test for unwanted biological (*e.g.*, sex) and technical variability (*e.g.*, batch effects). When the source of the confounders is known, it can be directly included in the association model. To deal with unknown source of biological and technical co-variation, PopPAnTe can integrate in the association model the principal components that are required to explain a user-specified percentage of variation.
PopPAnTe implements the Benjamini-Hochberg procedure (BH step-up procedure) to control the false discovery rate [5]. To aid in results interpretation and further analyses, PopPAnTe generates also basic Q-Q and manhattan plots.

References

[1] Purcell, S., et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses, Am J Hum Genet, 81, 559-575.

[2] Yang, J., Lee, S.H., Goddard, M.E., & Visscher P.M. (2011) GCTA: a tool for Genome-wide Complex Trait Analysis, Am J Hum Genet, 88, 76-82.

[3] Speed, D., Hemani, G., Johnson, M. R., & Balding, D. J. (2012). Improved heritability estimation from genome-wide SNPs. The American Journal of Human Genetics, 91(6), 1011-1021.

[4] Che, R., Jack, J. R., Motsinger-Reif, A. A., & Brown, C. C. (2014). An adaptive permutation approach for genome-wide association study: evaluation and recommendations for use. BioData Mining, 7(9).

[5] Benjamini, Y., & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) : 289-300.

[6] Thompson, E. A. (1986). Pedigree analysis in human genetics, Johns Hopkins University Press

[7] Hayes, JF and Hill, WG (1981). Modification of estimates of parameters in the construction of genetic selection indices ('bending'), Biometrics, 483–493

2 Version and Copyright

PopPAnTe is a command-line java jar executable. It is currently at version 1.0.2, released on October 20th, 2018. It requires java version 1.7+.

For any bugs or problems found, or to require additional features, please contact us at:

- Dr. Alessia Visconti alessia.visconti@kcl.ac.uk.
- Dr. Mario Falchi mario.falchi@kcl.ac.uk;

3 Acknowledgement

PopPAnTe includes and modifies procedures developed by Hariklia Eleftherohorinou (variance component core) Sam Halliday (mtj libraries), Carlos Morcillo Suarez (Q-Q and Manhattan plots), Ivan Akimov & *fanweixiao* (Hashids). We would like to thank them for allowing the free download, modification, and distribution of their work accordingly to the terms of public licenses.
We would also like to thank Massimo Mangino for being our beta-tester. His feedback has been really important in the editing of this users' guide.

4 Input file format

PopPAnTe requires text input files having a PLINK-like format (PED/MAP). **None of the input files has header, and files are all white-space (space or tab) delimited.** Users are advised to check the validity of their files before running the software.

4.1 PED file

The pedigree (PED) file has seven mandatory columns representing:

```
Family ID
Individual ID
Paternal ID
Maternal ID
Sex
Affection status
Twin status
```

Response values (column 8 onwards) must be numeric, and missing values can be represented with any of the following values: `X, x, NA, na, NaN, NAN, nan`. Individuals having a missing value for a dependent variable will be excluded from the analysis of that variable.
Any quantitative variable, such as transcript or metabolite levels, or quantitative phenotypes, can be used as dependent variable (or *response*).
When the family structure is known, PopPAnTE evaluates the pedigree-based relatedness matrix from the known pedigree relationships. To allow the correct reconstruction of the relatedness matrix, the structure of each family included in the PED file must be complete, that is, each individual must have their parents included in the PED file. Hence, if families are not complete, mock parents should be added to the PED file and their response values set to *missing* (see above). Mock individuals' paternal and maternal IDs should be set to zero, and their affection status set to -9. Each individual with affection value equal to -9 will be considered as mock and ignored during the testing. When a genetic similarity matrix is provided as input, the creation of mock parents is not required.

Please also note that:

- the IDs are alphanumeric and the combination of family and individual ID should uniquely identify a person;
- sex should be represented as: 1 (male), 2 (female) or 0 (unknown);
- affection status should be represented as: 1 (unaffected), 2 (affected), 0 (missing) or -9 (mock);
- twin status should be represented as: MZ or 1 (monozygotic twin), DZ or 2 (dizygotic twin) or 0 (not twin or unknown);
- the number of phenotypes should be the same as the number of rows of the RESPONSE files (see Section 4.2);
- PopPAnTe does not consider sex and affection status in the main analyses. If the users would like to include sex as covariate, they should use the COVARIATE file (see Section 4.5).

To check the validity of the PED file one can use the **pedcheck** mode available within PopPAnTe (See Section 5 for details) by running the following command[1]:

```
java -jar poppante.jar -ped mydata.ped -mode pedcheck
```

[1] PopPAnTe is a command-line java jar executable and must be run from a command prompt (Windows) or terminal window (Linux, Unix, Mac OS X). The examples in this users' guide suppose that executable and data are located in the same folder.

4.2 RESPONSE file

The RESPONSE file describes the dependent quantitative variables (*e.g.,* phenotypes). It has as many rows as the columns included in the PED file, and consists of only one mandatory column:

```
Response Name
```

Please note that:

- the rows order must follow the columns order of the response values in the PED file;
- if the number of columns is greater than one, subsequent columns will be ignored.

4.3 PREDICTOR file

This PREDICTOR describes the independent variable. It has two mandatory columns representing:

```
Family ID
Individual ID
```

Predictor values (column 3 onwards) must be numeric, and missing values can be represented with any of the following values: `X, x, NA, na, NaN, NAN, nan`. Individuals having a missing value for a variable will be excluded from the analysis of that variable. When the heritability mode is chosen, the variable to test should be represented using the PREDICTOR file.

Please note that:

- the IDs should correspond to those listed in the PED file. Data belonging to individuals not listed in the PED file will be ignored;
- the number of columns should be the same as the number of rows of the MAP files (see Section 4.4);

4.4 MAP file

The MAP file describes the predictors. We are aware that often predictors are values measured in genome-wide experiment, as it is, for instance, epigenetic markers in epigenome-wide association studies. For this reason we allows the users to include in the MAP file the position of the markers. When specified, the markers position will be loaded in the system and printed in the results file, facilitating the localisation of signals in genomic segments. The MAP file has either one or three mandatory columns:

```
Predictor Name
```

or

```
Marker Name
Chromosome
Position
```

according to the variable under study.

Please note that:

- the rows order of the predictors follows the columns order of the independent values in the PREDICTOR file;
- if the number of columns is smaller than three, only the first column will be loaded;
- if the number of columns is greater than three, the subsequent columns will be ignored.

4.5 COVARIATE file

PopPAnTe supports the inclusion of one or more **quantitative** covariates in the analysis. Covariates are given in a separate file and are optional. The COVARIATE file should be formatted as the PREDICTOR file, having two mandatory columns:

```
Family ID
Individual ID
```

Covariate values (column 3 onwards) must be numeric, and missing values can be represented with any of the following values: `X, x, NA, na, NaN, NAN, nan`. Individuals having a missing value for a covariate will be excluded from all the analyses.
PopPAnTe will include all covariates that are present in the covariates file. If the users would like to perform different analyses using different sets of covariates, they should create different COVARIATE files and run separate analyses.

Please note that:

- the IDs should correspond to those listed in the PED file. Data belonging to individuals not listed in the PED file will be ignored;

4.6 INCLUDE file

PopPAnTe supports the analysis of only a subset of predictors, and these predictors should be listed using the INCLUDE files, consisting of only one mandatory column:

```
Marker/Predictor Name
```

4.7 FILTER file

PopPAnTe supports the analysis of only a subset of responses, and these responses should be listed using the FILTER files, consisting of only one mandatory column:

```
Response Name
```

4.8 CORRECTION file

PopPAnTe supports the inclusion of one or more covariates for correcting the predictor values. These covariate will be regressed out of the data to correct for these unwanted variations. Another implemented approach is to regress out the first N principal components (see Section 5 for details).
These covariates are given in a separate file and are optional. The covariate file should be formatted as the PREDICTOR file, having two mandatory columns:

```
Family ID
Individual ID
```

Covariate values (column 3 onwards) must be numeric, and no missing values are allowed.

4.9 KINSHIP file

When the family structure is not available, PopPAnTe supports the usage of a genetic similarity matrix estimated from genome-wide genetic data. The kinship file has five mandatory columns:

```
Family ID of the first individual
Individual ID of the first individual
Family ID of the second individual
Individual ID of the second individual
Coefficient of relationship
```

Please note that:

- the IDs should correspond to those listed in the PED file. Data belonging to individuals not listed in the PED file will be ignored;
- if not specified in the kinship file the coefficient of self-relationship is set to one;
- if not specified in the kinship file the coefficient of relationship between individuals is set to zero;
- PopPAnTe allow the user to set a minimum coefficient value (see the option `-mink` below) under which individuals are considered unrelated, *i.e.*, their coefficient of relationship is set to zero. We suggest to use this option to speed up the analyses.

5 Running PopPAnTe

PopPAnTe is a command-line java jar executable and must be run from a command prompt (Windows) or terminal window (Linux, Unix, Mac OS X). The examples in this users' guide suppose that executable and data are located in the same folder.

PopPAnTe performs three different analyses, as specified by the `-mode` parameter:

- **pedcheck:** verifies the validity of the pedigree (PED) file (see Section 4.1);
- **association:** performs the association between the dependent and the independent variables; and
- **heritability:** evaluates the heritability of the independent variables.

To run the **pedcheck** mode only two parameter are mandatory, namely `-mode` and `-ped`:

```
java -jar poppante.jar -mode pedcheck -ped mydata.ped
```

`-mode` specifies the analysis mode, while `-ped` specifies the path of the PED file. The output will be a console message either confirming the correctness of the PED file or listing the problem found during the assessment.

To run the **association** mode five parameters are mandatory, namely `-mode`, `-ped`, `-response`, `-predictor`, and `-map`:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
```

while the `-response` option is not necessary when the **heritability** mode is request. In this case, indeed, the dependent variables are not used and thus not loaded.
Again, `-mode` specifies the analysis mode, while the other parameters specify the path to the input files.
These commands provide a minimal output. For a more detailed log of the analysis one should use the `-verbose` parameter:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -verbose true
```

Using this parameter a more detailed log will be shown, including the parameter in use, the data loaded, and the time used to perform each step of the analysis.

If no `-output` parameter is set, as in the previous example, the results are printed on the terminal window. To redirect the output to a file, let's say `results`, one should run:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -output results
```

and the results will be saved in a tab-separated file called `results.tsv`. If the `-plot` parameter is set, as in the following:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -output results -plot true
```

two extra files (`results_Manhattan-plot.png` and `results_QQ-plot.png`), representing Manhattan and Q-Q plot, will be saved. If the evaluation of the empirical p-value is set (with the `-alpha` and `-c` parameters, see below), the Manhattan and Q-Q plot will be also produced (with names: `results_Manhattan-plot_epvalue.png` and `results_QQ-plot_epvalue.png`). When the parameter `-output` is not set, the plots are saved in the current directory as `QQ-plot.png` and `Manhattan-plot.png` (or `QQ-plot_epvalue.png` and `Manhattan-plot_epvalue.png`)

In the **association** and in the **heritability** mode, and when predictors can be ordered in space, PopPAnTe can analyse both one predictor at a time or multiple predictors at the same time (region-based testing). In the latter case, the users should specify the size of the region (in bp) using the `-region` parameter:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -region 250
```

When the `-region` parameter is not set a *single* analysis is performed, meaning that each predictors is analysed separately.
When a region size is specified all the predictors from a given marker and within the given window are collapsed together by means of a principal component analysis, where the first principal component (PC1) is used instead of the single predictor values. When only one predictor is available within the given window, that predictor is used.

For the sake of clarity let us consider an example where a heritability test has been performed with a region included in a 250 bp window. The markers used are the following (MAP file):

```
foo1234567 1 1001000
foo1234568 1 1001100
foo1234569 1 1001150
foo1234570 1 1002000
```

The first test will include all the values from chr1:1001000 to chr1:1001240 (that is from the first probe until the end of the chosen window): the values of the three probes in this window (foo1234567, foo1234568 and foo1234569) will be collapsed (by substituting

their values with the PC1), and the result will be labelled with the the probe that is at the start of the window (that is foo1234567). The second test will include all the values from chr1:1001100 to chr1:1001340 (that are foo1234568 and foo1234569), and so forth. Summarising, PopPAnTe scans the marker list and collapses all the markers that are included from a marker until the end of the window. Then, it labels the results with the fist marker.

To use a COVARIATE file the `-covariate` parameter should be used:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -covariate covariate.txt
```

To test only a subset of predictors, the INCLUDE file should be specified using the `-include` parameter:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -include mysites.txt
```

To use a kinship matrix estimated with external tools the `-kinship` parameter should be used:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -kinship kinship.txt
```

To evaluate an empirical p-value by means of an adaptive procedure the `-alpha` and the `-c` parameter should be used:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -alpha 0.01 -c 0.1
```

This setting will guarantee that the desired level of significance (`alpha=0.01`) will be sampled with a standard error within the required precision (`c=0.1`), that is `alpha*c=0.001`. The two parameters are both mandatory, that is, if the users specify one, they must specify both.
Please note that the p-value resulting from the log likelihood ratio test is always computed.

To correct the predictor values for a set of known **quantitative** covariates the `-correct` parameter, followed by the covariates file path, should be used:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -correct correction_covariate.txt
```

If the correction covariates are unknown, the predictor values can be corrected by means of principal components. In this case the the `-correct` parameter should be followed by a number (in [0,1]) describing the proportion of overall variability represented by the first principal components that the users would like to regress out:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -correct 0.2
```

In this case the predictors will be corrected using as many principal components as necessary to account for the 20% of overall variability.

To normalise the response variable and the predictors values the `-normalise` parameter should be used:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -normalise both
```

This command normalises both dependent and independent variables, while:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -normalise predictor
```

normalises only the independent one. To normalise only the dependent variables one should use:

```
java -jar poppante.jar -mode association -ped mydata.ped
  -response response.txt -predictor predictors.txt -map map.txt
  -normalise response
```

When using large dataset, PopPAnTe could runs out of memory. This is because of the way that Java runs on a computer - what is actually run is a program called a virtual machine (the JVM) which executes the java instructions. The JVM has limits on the memory that can be allocated to the java program - and you might need to increase them if you are working with particularly large amount of data. In order to increase the amount of memory for PopPAnTe, the program should be run from the command line by writing for example:

```
java -jar -Xms64M -Xmx256M poppante.jar  -mode heritability
-ped mydata.ped -response response.txt -predictor predictors.txt
-map map.txt
```

This sets the initial and maximum memory size to 64MB and 256MB. The M suffix can be changed with G to represent gigabyte.

6 Output

PopPAnTe produces a tab-delimited output, that can be read with several software application such as spreadsheet (*e.g.*, Microsoft Excel, LibreOffice Calc), statistical tools (*e.g.*, R, Matlab) or text editors (*e.g.*, Notepad, Textmate).
When the **heritability** mode is selected the output file has 10 or 12 columns representing:

```
Predictors/Marker Name
[Chromosome]
[Position]
Nobs
Log likelihod of the null model (Lnlk_Null)
Log likelihod of the full model (Lnlk_Full)
Degree of freedom of the null model (dfnull)
Degree of freedom of the full model (dffull)
chi^2
pvalue
adj_pvalue
heritability
```

the chromosome and the position are included only if available in the MAP file.
When the **association** mode is selected the output file has 13 or 15 columns representing:

```
Response Name
Predictors/Marker Name
[Chromosome]
[Position]
Nobs
Log likelihod of the null model (Lnlk_Null)
Log likelihod of the full model (Lnlk_Full)
Degree of freedom of the null model (dfnull)
Degree of freedom of the full model (dffull)
chi^2
pvalue
adj_pvalue
beta
standard error
percentage of variance explained
```

again, the chromosome and the position are included only if available in the MAP file.
Please note that:

- the results are printed on the standard output (console) unless the parameter `-output` is set;
- the output has a header unless the parameter `-header` is set to false;
- when the `-relc` parameter is set two extra columns are added to the results:

  ```
  posF
  giniC
  ```

  representing the percentage of families showing a positive contribution and the Gini coefficient assessed on their contribution to the chi-square statistics, respectively;
- when the `-variance` parameter is set two extra columns are added to the results:

  ```
  varNull
  varFull
  ```

  representing the estimate of the variance component in the null and full model, respectively. In the inheritance model the null model has a single variance estimate, that is the environmental effect, while the full model shows the variance estimates for both the environmental and the genetic effect;
- when the parameter `-region` is set and thus the region-based analysis is performed the predictor in each result line corresponds to the marker representing the start of the window.
- when the parameter `-alpha` and the parameter `-c` are set two extra columns are added to the results:

  ```
  epvalue
  adj_epvalue
  ```

  representing the empirical p-value evaluated in the adaptive procedure.
- when the parameter `-plot` is set the output includes also Manhattan and Q-Q plot.

7 Example

Example files can be found on the PopPAnTe website.

`mydata.ped` contains data for 159 individuals organised in 10 families; it also contains quantitative levels for one phenotype, the individuals body mass index (BMI), which is described in the `response.txt` file. CpG values for 12 sites are used as predictors and given in `predictors.txt` and described in `map.txt`. A gene similarity matrix for these individuals is given in `kinship.txt` and covariates (age and sex) are given in `covariates.txt`. `correction_covariates.txt` provides the covariates for the correction of the DNA methylation values (cell-type proportions as estimated using the Houseman method). `mysites.txt` list a subset of the methylation sites.

To run the program for the association analysis use:

```
java -jar poppante.jar -mode association -ped mydata.ped
-response response.txt -predictor predictors.txt -map map.txt
```

specifying the size of the region if region-based test is to be performed:

```
java -jar poppante.jar -mode association -ped mydata.ped
-response response.txt -predictor predictors.txt -map map.txt
-region 200
```

Additional covariates can be included from the `covariates.txt` file. Try:

```
java -jar poppante.jar -mode association -ped mydata.ped
-response response.txt -predictor predictors.txt -map map.txt
-covariate covariate.txt
```

To run the program with an external kinship matrix use the `-kinship` parameter

```
java -jar poppante.jar -mode association -ped mydata.ped
-response response.txt -predictor predictors.txt -map map.txt
-kinship kinship.txt
```

To run the program for the heritability analysis use the mode `-mode heritability` instead of `-mode association`.

8 PopPAnTe parameters

A complete list of the parameter (in alphabetical order) is listed in the following.

Mandatory parameters:

`-mode <pedcheck |heritability| association>`	analysis to perform
`-map file path`	predictor information (MAP file) – *non mandatory if mode is pedcheck*
`-ped file path`	pedigree file (PED file)
`-predictor file path`	independent values (PREDICTOR file) – *non mandatory if mode is pedcheck*
`-response file path`	response information (RESPONSE file)– *non mandatory if mode is pedcheck or heritability*

Optional parameters:

`-alpha p-value`	p-value that controls the experiment-wise error rate (EWER), used in the adaptive permutation procedure – *default: null*
`-c precision`	desired precision in the adaptive permutation procedure – *default: null*
`-correct <file path|double>`	file containing the covariate values used to correct the independent values or the threshold of overall variability accounted by the principal component one would like to regress out from the data – *default: null*
`-covariate file path`	file containing the covariate values – *default: null*
`-decomposition <QR|LU>`	applies the QR/LU decomposition when the genetic relationship matrix is provided as input *(default: apply a bending procedure and use the Cholesky decomposition)*
`-include file path`	file containing the predictors to include in the analysis – *default: null*
`-filter file path`	file containing the responses to include in the analysis – *default: null*
`-header <true|false>`	whether the output has a header – *default: true*
`-help`	Print a help message and exit
`-kinship file path`	genetic similarity matrix file. The matrix should be estimated with an external tool – *default: null*
`-mink <threshold|c2|c3>`	minimum genomic relationship coefficient, all the kinship value smaller that threshold are set to 0. When set to c2 (c3) the minimum genomic relationship coefficient is set to 0.0315 (second cousins, 0.0078; third cousins) – *default: 0*
`-normalise <response|predictor|both>`	whether the values of responses, predictors or both should be transformed to their corresponding quantile in a standard normal transformation – *default null*
`-output file path`	output file – *default: standard output*

`-plot <true|false>`	whether to plot the Manhattan and Q-Q plot – *default: false*
`-region bp`	window size for the region-based testing. If not set one predictor is analysed at a time – *default: no region size set*
`-relc threshold`	whether the contribution of the sample to the final statistics must be evaluated. It allows one to verify whether the positive signal has been generated by a uniform contribution of the families within the sample or by a strong contribution of a small number of families. This option will generate two additional columns, one reporting the percentage of families showing a positive contribution and the second one the Gini coefficient assessed on their contribution to the chi-square statistics – *default: false*
`-threads n`	number of threads to use – *default: 1*
`-variance <true|false>`	whether the variance is printed. This option will generate two additional columns – *default: false*
`-verbose <true|false>`	whether verbose – *default: false*