ONETOOL Manual

Release 0.1.0

Yeunjoo E. Song **Sungyoung Lee** **Kyungtaek Park**
Robert C. Elston **Sungho Won**

February 14, 2018

1 Getting Started **3**
1.1 Getting ONETOOL binary . . . 3
1.2 Building ONETOOL from source . . . 4
1.3 Running ONETOOL . . . 4

2 Overview **5**
2.1 Input . . . 6
2.2 Features . . . 7
2.3 Output . . . 8

3 Tutorials **9**
3.1 Input . . . 9
3.2 InfoQC Ananlysis . . . 14
3.3 Trait Ananlysis . . . 17
3.4 Linkage Ananlysis . . . 19
3.5 Association Ananlysis . . . 20

4 Other Resources **23**

5 Indices and tables **25**

Contents:

CHAPTER

ONE

GETTING STARTED

ONETOOL is freely available. However, we ask you to add an appropriate statement (including the NIH grant number) under "acknowledgments" in any publication of results obtained by using this program.

Suggested wording is:

"(Some of)The results of this paper were obtained by using the software package ONETOOL, which was supported by the National Research Foundation of Korea Grant funded by Korean Government (NRF-2014S1A2A2028559)."

License:

- Copyright 2017 ONETOOL team

1.1 Getting ONETOOL binary

The pre-built ONETOOL binaries can be downloaded from the homepage for free.

ONETOOL has been compiled and tested on the platforms indicated below.

Platform	Operating system type	Operating system version
64-bit (x86_64)	Linux	Kernel 2.6+ RHEL5, Ubuntu
64-bit (x86_64)	Windows	Vista, 7, 10

Two different versions are provided, static and dynamic, for each platform.

- **Static** version does not include R library, so will run under most system environments, but pedigree plot option (`--plot`) is not available.
- **Dynamic** version requires R system library in the right place when it is executed. This version supports pedigree plot function (`--plot`).

Depending on the version, you will need:

- Intel Fortran Compiler Redistributable is required to run Windows version of ONETOOL.
- R package kinship2 is required to use `--plot` function in ONETOOL for both dynamic versions.
- For Windows, ONETOOL expect R is installed on the initial installation directory (`C:\Program Files\R\R-X.X.X`).

Unzip the distribution after you download it.

For Linix version, **un-tar** ONETOOL package.

```
tar -xzf ONETOOL.v2.0_Linux_x86_64_dynamic_withR.tgz
```

For Windows version, double-click to **unzip** ONETOOL distribution.

Once you've set up ONETOOL, we recommend that you download the example files. for some test runs. Most of examples in this documentation are using the same example files.

1.2 Building ONETOOL from source

The ONETOOL source code. To clone the ONETOOL repository using Git, run

```
$ git clone https://github.com/SWonlab/onetool.git
$ cd onetool
```

You can also download the source code directly from Github.

ONETOOL uses several external libraries. These should load automatically. If not, these must be explicitly installed.

- boost
- blas and lapack
- R 3.3.1 with packages `kinship`.

1.3 Running ONETOOL

To run ONETTOL binary, go to the directory where ONETOOL is installed.

```
$ ./onetool --fam test_miss0.fam
```

Several ONETOOL tests are available on homepage.

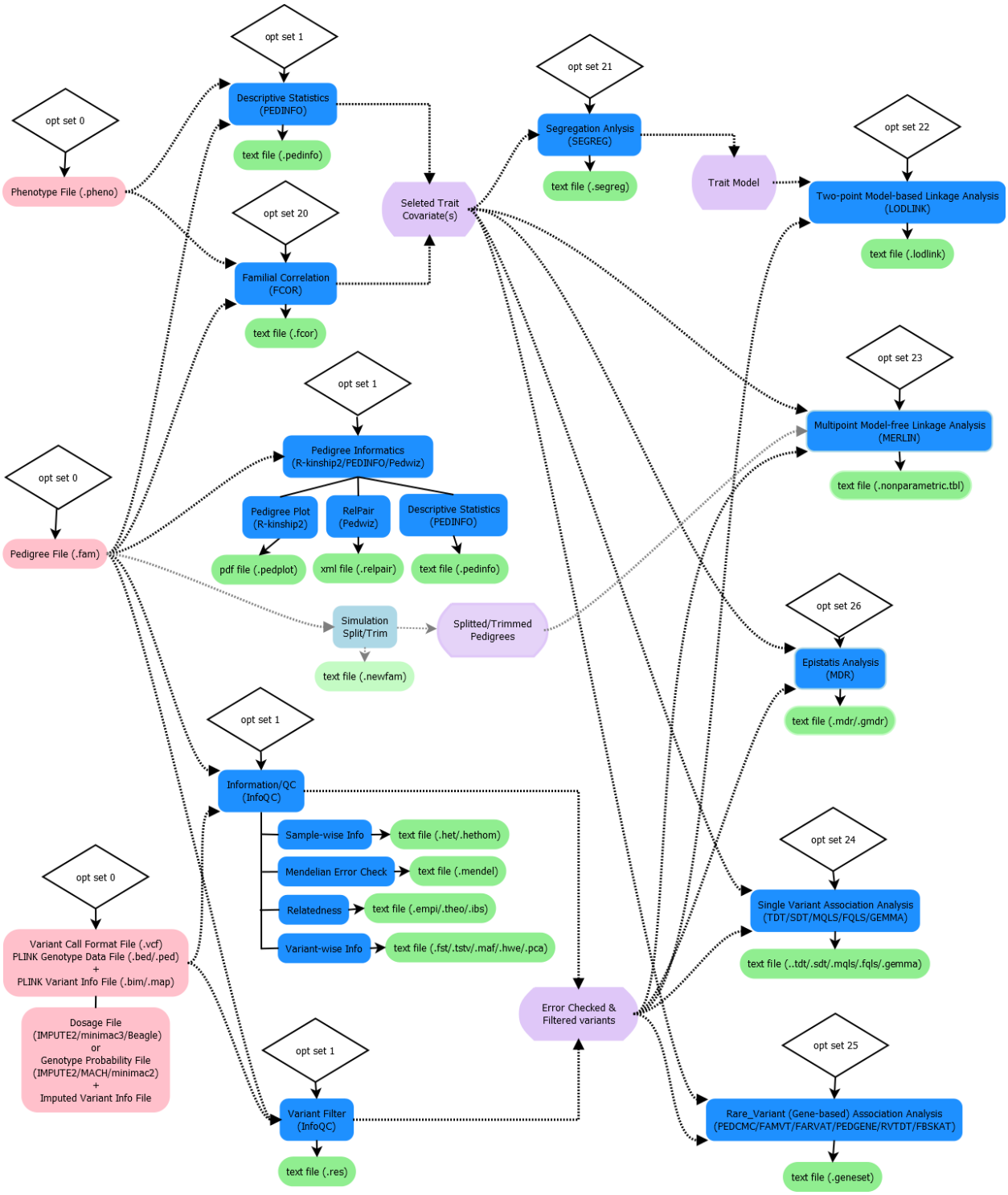
CHAPTER

TWO

OVERVIEW

ONETOOL is a tool for family-based big data analyses. It implements the properties of well-known existing family data analysis tools and recently developed methods in a computationally efficient manners, so is suitable for analyzing the vast amount of variant data availble from sequencing family members, providing a rich choice of analysis methods for big data from families.

Overall workflow in ONETOOL is:



2.1 Input

ONETOOL supports two different sets of input files, **PLINK set** and **VCF set**. The PLINK set consists of three files (i.e., .fam, .bed, and .bim) that are used to run PLINK, and the VCF set con-sist of a plink format family file (.fam) and a Variant Call Format (.vcf). The additional phenotypes and covariates are supported through an optional input

file (.pheno) for both sets of input files.

ONETOOL also support two different ways to specify the desired analysis options, through a command line and a script file.

The full list of data types and input file formats supported can be found in *Input*.

2.2 Features

The main features in ONETOOL are:

- **InfoQC** analysis
    1. Variant information
    2. Sample information
    3. Pedigree information
    4. Mendelian error detection
    5. Relatedness matrix
    6. Pedigree plot
- **Trait** analysis
    1. Familial Aggregation
    2. Heritability
    3. Segregation Analysis
- **Linkage** analysis
    1. Model-based
    2. Model-free
- **Association** analysis
    1. Single variant - common variant
    2. Gene-based - rare variant
    3. Dosage data
- **Epistasis** analysis
    1. Multidimensional Dimensionality Reduction (MDR) analysis
    2. Generalized MDR analysis
- **Imputation** of missing genotype
- **Data management**
    1. Variant filtering
    2. Sample filtering
    3. LD-based prunning

2.3 Output

Each method in ONETOOL outputs the result file with the appropriate extension, so that the user can recognize it easily. It has the familiar user interface and the same or similar analysis option names as the existing tools, so no, or only a minimal, learning curve is needed.

CHAPTER

THREE

TUTORIALS

ONETOOL is a tool for family-based big data analyses. It implements the properties of well-known existing family data analysis tools and recently developed methods in a computationally efficient manners, so is suitable for analyzing the vast amount of variant data availble from sequencing family members, providing a rich choice of analysis methods for big data from families.

In this tutorial, we provide a guided tour through the main features of ONETOOL. You can read it in HTML or PDF versions.

3.1 Input

ONETOOL supports many different types of data in many different file formats. For a run, however, it only requires the input file(s) that are relevant to the specified analyses in that run.

3.1.1 Input Data Types

Though the main purpose of ONETOOL is for family-based big data analyses, it can analyze unrelated individual data as well.

Types of data supported:

- **Sample**
    1. Family samples (related individuals)
    2. Independent samples (unrelated individuals)
- **Phenotype**
    1. Binary
    2. Continuous
- **Variant**
    1. genotype - common and rare SNPs
    2. genotype probability/dosage - imputed variants

Note: The terms ‘sample’, ‘subject’ and ‘individual’ are used interchangeably.

Note: The terms ‘variant’, ‘SNP’ and ‘genotype’ are used interchangeably.

Note: The terms ‘trait’ and ‘phenotype’ are used interchangeably.

3.1.2 Input File Types

The types of input file (with the expected extension in parenthesis) that can be used for an ONETOOL run are listed for different data types.

Data Type	File Type	Extension
sample	*PLINK FAM file*	.fam
phenotype	*PLINK Phenotype file*	.pheno
variant	*Variant Call Format (VCF) file*	.vcf
variant	*PLINK BED/BIM file*	.bed/.bim
variant (dosage)	*IMPUTE2 file*	.impute2/.impute2_info
sample + variant	*PLINK PED/MAP file*	.ped/.map

Additinal files for a specific analysis:

- *S.A.G.E. Parameter file (.par)*
- *S.A.G.E. Trait genotype probability file (.typ)*
- *MERLIN MAP file (.map)*
- *Gene SET file (.set)*
- *Script file (.script)*

Two main input file sets are 'VCF set' and 'PLINK set'.

The **VCF set** consist of a PLINK format family file (.fam) and a Variant Call Format file(.vcf).

The **PLINK set** consists of three files (i.e., .fam, .bed, and .bim) that are used to run PLINK.

The additional phenotypes and covariates are supported through an optional input file (.pheno) for both sets of input files.

3.1.3 SCRIPT file

ONETOOL also support two different ways to run the program, through a command line and a script file (.script).

A script file includes the input file name(s) and all command-line options selected for a ONETOOL run.

```
$ onetool --script test.txt
```

3.1.4 Main Input Data File

FAM file

This is the FAM file in PLINK.

Each lines of the FAM file describes an individual. It contains a white-space (space or tab) delimited records for each individual including fields for identifier, sex, two parents and an optional trait field.

1. Family ID
2. Individual ID
3. Father ID
4. Mother ID

5. Sex (1=male; 2=female; other=unknown)
6. Phenotype (optional)

Note: Unlike in PLINK, the last column with dummy data can be omitted in the FAM file when the main phenotype is in the phenotype file.

Note: It is the required input file and can be used alone for certain types of analyses in ONETOOL.

```
$ onetool --fam test_miss00.fam
```

PHENO file

This is the alternate phenotype file in PLINK to specify an alternate phenotype for analysis, i.e. other than the one in the PED (or FAM) file.

Each lines of the PHENO file contains the phenotype data for an individual. The first two columns contain the identifiers and the rest of columns are phenotype data.

1. Family ID
2. Individual ID
3. Phenotype1
4. Phenotype2
5. Phenotype3
6. Phenotype4
7. ...

The detailed description of the PHENO file can be found in PLINK.

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --pheno test_miss00.pheno --pname sbp
```

VCF file

This is the Variant Call Format(VCF) file used in 1000 Genomes Project. Files in both plain text format (.vcf) or gzipped format (.bcf) are supported. The meta information lines (starting with ##) are ignored.

The first 9 columns in header and data lines are:

1. CHROM
2. POS
3. ID
4. REF
5. ALT
6. QUAL
7. FILTER
8. INFO
9. FORMAT

Note: The sample IDs in the header line (starting with #CHROM) have to match with the individual IDs in FAM file uniquely.

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf
```

BED/BIM file

BED file is the PLINK binary PED file and it is used together with BIM file (extended MAP file: two extra cols = allele names).

The detailed description of the BED/BIM files can be found in PLINK.

```
$ onetool --fam test_miss00.fam --bed test_miss00.bed --bim test_miss0.bim
```

IMPUTE2 file

IMPUTE2 file is the output files from IMPUTE2 program and both .impute2 and .impute2_info files are required.

The genotype file stores (.impute2) data on a one-line-per-SNP format. The first 5 entries of each line should be:

1. SNP ID
2. RS ID of the SNP
3. base-pair position of the SNP
4. the allele coded A
5. the allele coded B.

The next three numbers on the line should be the probabilities of the three genotypes AA, AB and BB at the SNP for the first individual in the sample and the next three numbers for the second individual and so on.

The SNP-wise information file (.impute2_info) contains the following columns (header shown in parentheses):

1. SNP identifier from -g file (snp_id)
2. rsID (rs_id)
3. base pair position (position)
4. expected frequency of allele coded '1' in the -o file (exp_freq_a1)
5. measure of the observed statistical information associated with the allele frequency estimate (info)
6. average certainty of best-guess genotypes (certainty)
7. internal "type" assigned to SNP (type)

The detailed description of the IMPUTE2 files can be found in IMPUTE2.

```
$ onetool --fam test_miss00.fam --dosage test_miss00.impute2 --mqls
```

PED/MAP file

PED file is the default a white-space (space or tab) delimited text file format used in PLINK. The same first six columns as in FAM file are mandatory.

When a PED file is used, it has to be used with PLINK MAP file which contains the following 4 columns.

1. chromosome (1-22, X, Y or 0 if unplaced)
2. rs# or snp identifier
3. Genetic distance (morgans)
4. Base-pair position (bp units)

The detailed description of the PED/MAP files can be found in PLINK.

```
$ onetool --ped test_miss00.ped --map test_miss00.map
```

3.1.5 Additinal Input files

PAR file

This is S.A.G.E. parameter file which is an optional file to run a segregation analysis in ONETOOL. It is a text file containing a list of S.A.G.E. SEGREG instructions written according to a specific syntax (i.e., instruction blocks).

For the detailed description of the PAR file, see the S.A.G.E. user manual.

Note: When a `par` file is not specified, ONETOOL runs the default analysis in SEGREG which is the comingling analysis.

```
$ onetool --fam test_miss00.fam --pheno test_miss00.pheno --pname t2d --segreg --par segreg.par
```

TYP file

This is S.A.G.E. trait genotype probability file which is required to run a model-based linkage analysis in ONETOOL. It is a text file containing the individual specific type probabilities conditional on the model and all pedigree information available, and individual specific penetrance information. This file is prodiced by a segregation analysis.

For the detailed description of the PAR file, see the S.A.G.E. user manual.

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf -lodlink --typ test_segreg.typ
```

MERLIN MAP file

This is MERLIN genetic map file which is required to run a model-free linkage analysis in ONETOOL.

For the detailed description of this MAP file, go to MERLIN.

Note: The PLINK MAP file contains 4 columns while the MERLIN MAP file contains 3 columns (sex-average) or 5 columns (sex-specific). The PLINK MAP file with the correct genetic map information in 3rd column can be used for the model-free linkage analysis in ONETOOL.

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --merlin --map test_miss00.map
```

SET file

This is a required file to run gene-based association analyses for rare variants. It contains the SNP clustering information in genes in the following format:

1. Gene ID

2. SNP ID

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --genetest --set test_gene.txt
```

3.2 InfoQC Ananlysis

3.2.1 Pedigree Information

For the informatics of family data, ONETOOL utilizes PEDINFO in the S.A.G.E. package that provides many useful descriptive statistics on pedigree data, including means, standard deviations; family, sibship and pedigree sizes; and counts of each type of relative pair. This is done for ONETOOL run by default.

```
$ onetool --fam test_miss00.fam
$ onetool --fam test_miss00.fam --pheno test_miss0_phen.txt --pname sbp
```

The output file is a text file containing a formatted table with several sub-tables. Each sub-table contains the pedigree-wise/pair-wise/sibship-wise/individual-wise descriptive statistics as shown below:

```
                    General Statistics: All Pedigrees

                |    Count| Mean Size +/- Std. Dev. (     Min.,      Max.)
Pedigrees       |       10|     10.00 +/-      ---  (       10,        10)

Generation Statistics |  Nuc  Family Statistics  |  Inh  Vector Bit Stats
# of Gens | # of Peds |  # of Nuc Fams |# of Peds|   # of Bits   |# of Peds
        3|        10|      3 -     4|       10|      5 -     8|        10

                |    Count| Mean Size +/- Std. Dev. (     Min.,      Max.)
Sibships        |       30|      2.00 +/-     1.41 (       1,         4)

Constituent     |         |  Marriage       |         |                 |
Pedigrees       |       10|  Rings          |        0|  Loops          |        0

Pairs                   |      Count|  Individuals            |      Count
Parent/Off              |        120|  Male                   |         50
Sib/Sib                 |         60|  Female                 |         50
Sis/Sis                 |         10|  Unknown                |          0
Bro/Bro                 |         10|        Total            |        100
Bro/Sis                 |         40|                         |
Grandp.                 |        160|  Founder                |         40
Avunc.                  |          0|  Non-founder            |         60
Half Sib                |          0|  Singleton              |          0
Cousin                  |          0|        Total            |        100

                    Individuals With Multiple Mates
none

                    Consanguineous Mating Pairs
none
```

1. pedigree count, mean size with standard deviation, range
2. pedigree count by generation and numclear family count and inheritance vector bit size
3. sibship count, mean size with standard deviation, range
4. number of constituent pedigrees (i.e., disjoint sub-pedigrees), marriage rings and loops (if found)
5. relative pair count by type and individual count by sex and founder/non-founder/singleton status
6. individuals with multiple mates (if found)
7. consanguineous mating pairs (if found)

Note: The colored-blocks are added to show the different sub-tables.

3.2.2 Pedigree Plot

ONETOOL dynamic version with R plugin provides visualization of family data utilizing the R package kinship2 to generate a plot (`--plot`).

```
$ onetool --fam test_miss00.fam --plot
```

3.2.3 Mendelian Error Check

Two types of errors are there in family genetic data, pedigree errors and genotyping error, which both can lead to either increased false negatives or false positives in both linkage and association studies.

Pedigree errors are the misspecified relationships among individuals in family data.

Mendelian inconsistencies can be used to identify relationship errors and genotyping error.

Mendelian inconsistencies are usually identified by comparing the genotypes of one or both parents to the genotypes of their offspring. It involves checking for each locus whether one of the two alleles that a offspring has could have been inherited from one of its parents.

ONETOOL detects Mendelian inconsistencies in pedigree data. Each marker is individually checked for inconsistencies in every pedigree. To check for Mendelian errors in given genotype data, simply add `--mendel` option.

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --mendel
```

ONETOOL reports the results by family, by sample and by variant.

- **By family (.family.res)**

Column	Description
FID	Family ID
ERROR	Total number of alleles with Mendelian error in the family
AVAIL	Total number of called alleles in the family
PROP	Proportion of alleles with Mendelian error in the family

- **By sample (.sample.res)**

Column	Description
FID	Family ID of the sample
IID	Individual ID of the sample
ERROR	Total number of alleles with Mendelian error in the sample
AVAIL	Total number of called alleles in the sample
PROP	Proportion of alleles with Mendelian error in the sample

- **By variant (.variant.res)**

Column	Description
CHR	Chromosome name
VARIANT	Variant ID
POS	Physical position of the variant
ALT	An alternative allele
ERROR	Total number of alleles with Mendelian error
AVAIL	Total number of called alleles
PROP	Proportion of alleles with Mendelian error

3.2.4 Relatedness Matrix

One of the key factors in heritability estimation and a family-based association test is how to model the relatedness among the pairs of family members. Therefore, testing with different options for relatedness is crucial to overcome the limitation of one method over another. ONETOOL provides 3 different options:

- traditional pedigree-based kinship matrix (`--kinship`)
- identity-by-state (IBS) matrix using genome-wide variant data (`--ibs`)
- genomic relationship matrix (GRM) using genome-wide variant data (**default**)

To generate the out file contaning the symmetric relatedness matrix, simply add `--makecore` option. The extensions of the output files are (.theo.cor), (.emp.cor), and (.ibs.cor) for kinship matrix, GRM matrix and IBS matrix respectively.

```
$ onetool --fam test_miss00.fam --makecor --kinship
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --makecor
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --makecor --ibs
```

Note: Kinship matrix (`--kinship`) is pedigree-based, i.e., no variant data are needed. IBS matrix (`--ibs`) and GRM matrix are estimated from the given variant data.

3.2.5 Sample Information

The sample-wise information from variant data helps to better understand the genetic background of the individuals in family in population level. ONETOOL calculates the following information on each sample.

- **Heterozygosity**

  ```
  $ onetool --fam test_miss00.fam --vcf test_miss00.vcf --het
  ```

- **Ratio of Heterozygote and homozygote**

  ```
  $ onetool --fam test_miss00.fam --vcf test_miss00.vcf --hethom
  ```

3.2.6 Variant Information

ONETOOL's options for the variant QC and informatics are similar with those in PLINK, but they are implemented in a computationally optimized way providing more speed and efficiency.

Types of variant-wise information analysis supported:

- **F-statistics** - Wright's fixation index to describe population structure

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --pheno test_miss00.pheno --pname ethnicit
```

- **Allele frequency** - Minor allele frequency

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --freq
```

- **HWE** - Hardy-Weinberg Equlibrium

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --hwe
```

- **PCA** - Principle component analysis

```
$ onetool --fam test_miss00.fam --vcf test_miss00.vcf --pca --npc 5
```

Note: To specify the number of principal components to compute, use `--npc`.

3.3 Trait Ananlysis

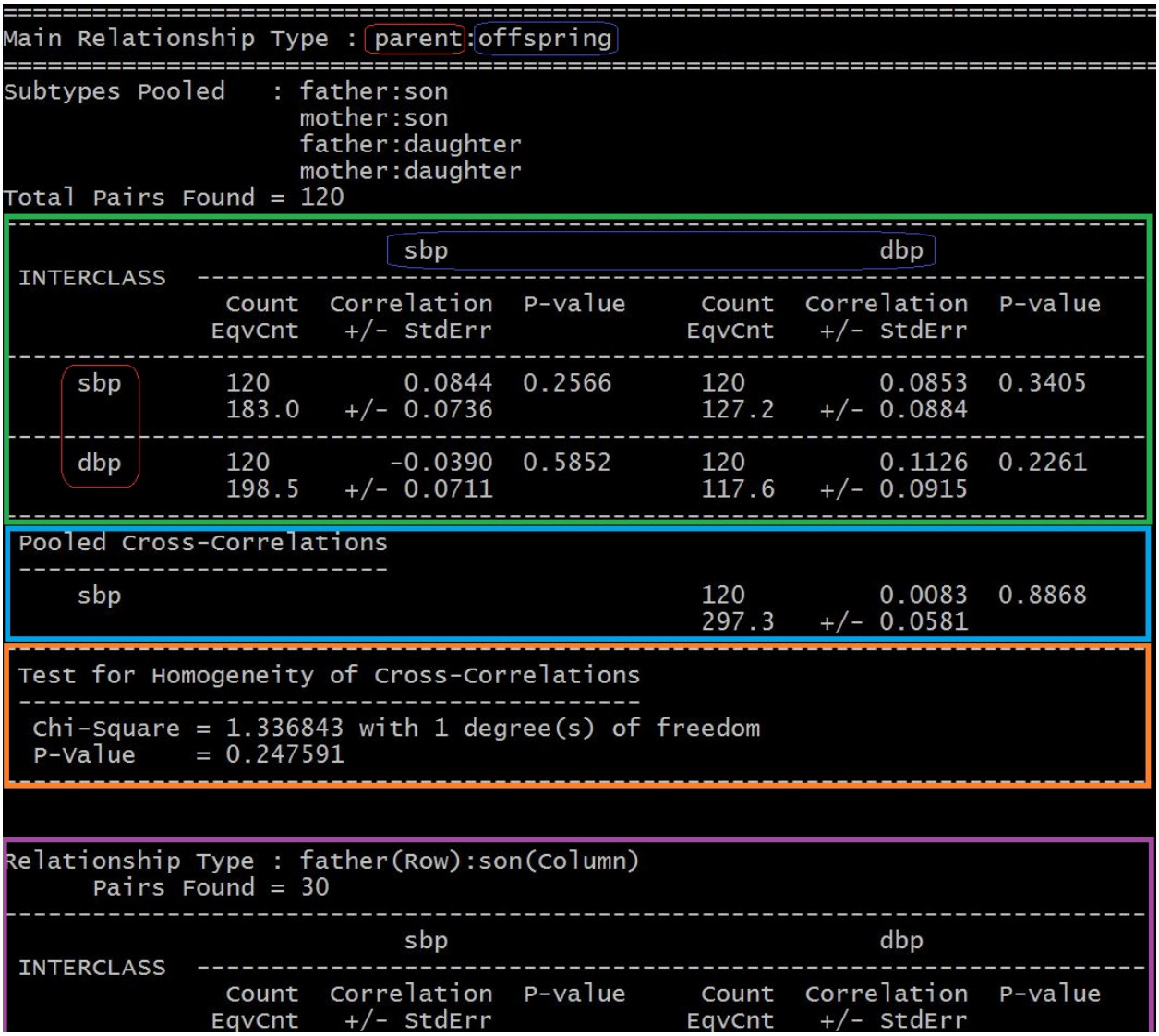
Trait analysis is the first step in any genetic analysis to establish the causal relationship between genetic variants and a trait of interest. This includes studies of familial aggregation (correlation), narrow-sense heritability estimation, and segregation analysis.

3.3.1 Familial Aggregation (Correlation)

Familial aggregation studies determine whether a disease of interest is observed in families more than would be expected by chance alone. The FCOR module in ONETOOL calculates multivariate familial correlations with their asymptotic standard errors.

```
$ onetool --fam test_miss00.fam --fcor
$ onetool --fam test_miss00.fam --pheno test_miss0_phen.txt --pname sbp --fcor
```

The output file is a text file containing tables of correlations, their standard errors, used pair counts and equivalent pair counts for each pair of traits for each subtype and/or main type of relative up to 2nd generation.



The header containes the information on main relation type (parent:offspring) with subtypes pooled and total pairs count. The following blocks contain:

1. **main result table for parent:offspring type**
    - row (rounded box in red) - trait(s) for parent
    - column (rounded box in blue) - trait(s) for offspring
    - correlations, their standard errors and P-values
    - used pair counts and equivalent pair counts
2. pooled cross-correlations of interclass relative types
3. homogeneity test result when there are more than 1 trait
4. followed by the main result tables for each subtypes pooled

Note: The colored-blocks and boxes are added to show the different parts of tables in the example output above.

3.3.2 Segregation analysis

The SEGREG module fits and tests Mendelian segregation models in the presence of residual familial correlations.

```
$ onetool --fam test_miss00.fam --segreg
$ onetool --fam test_miss00.fam --pheno test_miss0_phen.txt --pname sbp --segreg
```

It generates 3 output files:

- Summary output file (.segreg.sum) - contains the table of final estimates of the parameters and their standard errors, and other results.
- Detailed output file (.segreg.det) - contains the table of final estimates and variance-covariance matrix of the parameter estimates.
- Trait genotype probability and penetrance function output file (.segreg.typ) - contains the individual specific type probabilities conditional on the model and all the pedigree information available, and individual specific penetrance information which is an input file into model-based linkage module (`--lodlink --typ onetool.segreg.typ`).

At the end of the detailed output contains the likelihood comparison table from comingling analysis (by default).

```
================================================================================
                    Likelihoods for Commingling Analysis

Option                       LN(Likelihood)  -2 LN(Likelihood)  Akaike's AIC score
--------------------------------------------------------------------------------
one mean                          -221.8950          443.7901            449.7901
two means                         -220.1383          440.2766            452.2766
three means                       -217.7625          435.5251            449.5251
================================================================================
```

3.3.3 Heritability Estimation

Heritability estimation can be done using different relatedness options:

- traditional pedigree-based kinship matrix (`--kinship`)
- identity-by-state (IBS) matrix using genome-wide variant data (`--ibs`)
- genomic relationship matrix (GRM) using genome-wide variant data (**default**)

```
$ onetool --fam test_miss00.fam --heritability --kinship
$ onetool --fam test_miss00.fam --vcf test_miss0.vcf --heritability --ibs
$ onetool --fam test_miss00.fam --vcf test_miss0.vcf --heritability
```

Note: Kinship matrix (`--kinship`) is pedigree-based, i.e., no variant data are used. IBS matrix (`--ibs`) and GRM matrix are estimated from the given variant data.

3.4 Linkage Ananlysis

Linkage analysis in ONETOOL is composed of two parts, two-point model-based linkage and two-point model-free linkage accounting for linkage disequilibrium (LD).

3.4.1 Model-based

LODLINK module performs model-based lod score calculations for two-point linkage between a main trait (that follows Mendelian transmission and has either two or three types) and each of the variants in the input genotype file. The output from SEGREG model (.typ file) is used as input to specify the main trait model.

LODLINK in S.A.G.E. is used which uses the genotype/phase elimination algorithms proposed by Lange and Boehnke (1983) and Lange and Goradia (1987), together with other enhancements, to perform relatively fast exact linkage calculations.

```
$ onetool --fam test_miss00.fam --vcf test_miss0.vcf --lodlink --typ test_segreg.typ
```

It generates 2 output files:

- Summary output file (.lodlink.sum) - contains lod scores and results of linkage and linkage homogeneity tests. Results in this file are based on calculations done on the pedigree data file as a whole.
- Detailed output file (.lodlink.det) - contains results on per family.

3.4.2 Model-free

ONETOOL uses Merlin for multipoint model-free linkage analysis accounting for linkage disequilibrium (LD), providing the Kong and Cox (1997) NPL statistics.

```
$ onetool --fam test_miss00.fam --vcf test_miss0.vcf --merlin --map test_miss0.map
```

It generates a output file with the extention .nonparametric.tbl.

Note: The same restriction applies to the size of family as in MERLIN.

3.5 Association Ananlysis

Association analysis module provides many different family-based association tests between trait(s) and both single-variant (common variant) and gene-based (rare variant).

3.5.1 Single variant

Summary of single variant analyses available in ONETOOL:

		Trait type		Covariate	Family data structure	Note
		binary	continuous			
Single variant	scoretest	N	Y	Y	general pedigree	usually efficient for randomly selected samples
	tdt	Y	N	N	trio	parental genotype need to be known but not used
	sdt	Y	N	N	nuclear family	need the genotype data of unaffected sibs
	mqls	Y	N	N	general pedigree	efficient for ascertained families
	fqls	Y	Y	N	general pedigree	efficient for ascertained families
	gemma	N	Y	Y	general pedigree	usually efficient for randomly selected samples
	multifqls	Y	Y	Y	general pedigree	efficient for ascertained families

3.5.2 Gene-based (Rare variant)

Summary of gene-based (rare variant) analyses available in ONETOOL:

		Trait type		Covariate	Family data structure	Note
		binary	continuous			
Rare variant	collapsing	Y	Y	N	general pedigree	efficient when effects of rare variants are homogeneous
	pedcmc	Y	N	N	general pedigree	efficient when effects of rare variants are homogeneous
	famvt	N	Y	N	general pedigree	efficient if rarer variants have stronger effect on disease
	farvat	Y	Y	Y	general pedigree	robust to the heterogeneity of effects of rare variants
	pedgene	Y	Y	Y	general pedigree	conditioning on phenotypes, treating the genotype data random, for pedigrees sampled because of multiple affected members
	fbskat	Y	N	N	general pedigree	efficient if rare variants with both positive and negative effect on disease are grouped to a single set
	rvtdt	Y	N	N	trio	efficient if rare variants with both positive and negative effect on disease are grouped to a single set

3.5.3 Dosage/genotype probability data

ONETOOL can take the dosage and genotype probability files from several popular imputation tools available for population data.

CHAPTER

FOUR

OTHER RESOURCES

- S.A.G.E.
- MERLIN
- PLINK
- Variant calling data files

CHAPTER

FIVE

INDICES AND TABLES

- *genindex*