

**Introduction, downloads**

S: 21 May 2017 (b4.4)  
D: 2 Jul 2017  
Recent version history  
What's new?  
Future development  
Limitations  
Note to testers

**[Jump to search box]**

**General usage**

Citation instructions

**Standard data input**

PLINK 1 binary (.bed)  
Autoconversion behavior  
PLINK text (.ped, .tped...)  
VCF (.vcf{.gz}, .bcf)  
Oxford (.gen{.gz}, .bgen)  
23andMe text  
Generate random  
Unusual chromosome  
IDs  
Recombination map  
Phenotypes  
Covariates  
Clusters of samples  
Variant sets  
Binary distance matrix  
IBD report (.genome)

**Input filtering**

Sample ID file  
Variant ID file  
Cluster membership  
Set membership  
Attribute-based  
Chromosomes  
SNPs only  
Simple variant window  
Multiple variant ranges  
Sample/variant thinning  
Covariates (--filter)  
Missing genotypes  
Missing phenotypes  
Minor allele frequencies  
Hardy-Weinberg  
Mendel errors  
Quality scores  
Relationships

**Main functions**

Data management  
--make-bed  
--recode  
--output-chr  
--zero-cluster  
--split-x/--merge-x  
--set-me-missing  
--fill-missing-a2  
--set-missing-var-ids  
--update-map...  
--update-ids...  
--flip  
--flip-scan  
--keep-allele-order...  
--indiv-sort  
--write-covar...  
--{b}merge...  
Merge failures  
VCF reference merge  
--merge-list  
--write-snp-list  
--list-duplicate-vars  
Basic statistics  
--freq{x}  
--missing  
--test-mishap  
--hardy  
--mendel  
--het/--ibc  
--check-sex/--impute-sex  
--fst  
Linkage disequilibrium  
--indep...  
--r/--r2  
--show-tags  
--blocks  
Distance matrices  
Identity-by-

# File format reference

This page describes specialized PLINK input and [output](#) file formats which are identifiable by file extension. (Most extensions not listed here have very simple one-entry-per-line text formats.)

**Jump to:** [.adjusted](#) | [.allele.no.snp](#) | [.assoc](#) | [.assoc.dosage](#) | [.assoc.fisher](#) | [.assoc.linear](#) | [.assoc.logistic](#) | [.auto.R](#) | [.bcf](#) | [.beagle.dat](#) | [.bed](#) | [.bim](#) | [.blocks\\*](#) | [.chr-\\*.dat](#) | [.chr-\\*.map](#) | [.clst](#) | [.clumped\\*](#) | [.cluster\\*](#) | [.cmh](#) | [.cmh2](#) | [.cnv](#) | [.cnv.indiv](#) | [.cnv.overlap](#) | [.cnv.summary](#) | [.cov](#) | [.dfam](#) | [.diff](#) | [.dist](#) | [.dupvar](#) | [.eigenvec\\*](#) | [.epi.\\*](#) | [.fam](#) | [.flipscan](#) | [.frq](#) | [.frq.cc](#) | [.frq.count](#) | [.frq.strat](#) | [.frqx](#) | [.fst](#) | [.gen](#) | [.genome](#) | [.grm](#) | [.grm.N.bin](#) | [.grm.bin](#) | [.gvar](#) | [.het](#) | [.hh](#) | [.hom](#) | [.hom.indiv](#) | [.hom.overlap\\*](#) | [.hom.summary](#) | [.homog](#) | [.hwe](#) | [.ibc](#) | [.imiss](#) | [.info](#) | [.lasso](#) | [.ld](#) | [.ldset](#) | [.lgen](#) | [.list](#) | [.lmiss](#) | [.map](#) | [.mdist](#) | [.mdist.missing](#) | [.mds](#) | [.\\*mendel](#) | [.meta](#) | [.mibs](#) | [.missing](#) | [.missing.hap](#) | [.model](#) | [.mperm](#) | [.nearest](#) | [.occur.dosage](#) | [.out.dosage](#) | [.ped](#) | [.perm](#) | [.pphe](#) | [.prob](#) | [.profile](#) | [.qassoc](#) | [.qassoc.gxe](#) | [.qassoc.means](#) | [.qfam.\\*](#) | [.range.report](#) | [.raw](#) | [.recode.\\*.txt](#) | [.recode.phase.inp](#) | [.recode.strct.in](#) | [.ref](#) | [.rel](#) | [.rlist](#) | [.sample](#) | [.set](#) | [.set.{m}perm](#) | [.set.table](#) | [.sexcheck](#) | [.simfreq](#) | [.tags.list](#) | [.tdt](#) | [.tdt.poo](#) | [.tfam](#) | [.tped](#) | [.traw](#) | [.twolocus](#) | [.var.ranges](#) | [.vcf](#)

**.\*.adjusted (basic multiple-testing corrections)**

Produced by [--adjust](#).

A text file with a header line, and then one line per set or polymorphic variant with the following 8-11 fields:

CHR	Chromosome code. <i>Not present with set tests.</i>
'SNP'/'SET'	Variant/set identifier
UNADJ	Unadjusted p-value
GC	Devlin & Roeder (1999) genomic control corrected p-value. <i>Requires an additive model.</i>
QQ	P-value quantile. <i>Only present with 'qq-plot' modifier.</i>
BONF	Bonferroni correction
HOLM	Holm-Bonferroni (1979) adjusted p-value
SIDAK_SS	Šidák single-step adjusted p-value
SIDAK_SD	Šidák step-down adjusted p-value
FDR_BH	Benjamini & Hochberg (1995) step-up false discovery control
FDR_BY	Benjamini & Yekutieli (2001) step-up false discovery control

Variants/sets are sorted in p-value order. (As a result, if the QQ field is present, its values just increase linearly.)

**.allele.no.snp (allele mismatch report)**

Produced by [--update-alleles](#) when there is a mismatch between the loaded alleles for a variant and columns 2-3 of the [--update-alleles](#) input file.

A text file with no header line, and one line per mismatching variant with the following three fields:

1. Variant identifier
2. Expected allele #1 (from [--update-alleles](#) input file)
3. Expected allele #2

**.assoc, .assoc.fisher (case/control association allelic test report)**

Produced by [--assoc](#) acting on a case/control phenotype.

A text file with a header line, and then one line per variant typically with the following 9-10 fields:

CHR	Chromosome code
SNP	Variant identifier
BP	Base-pair coordinate
A1	Allele 1 (usually minor)
F_A	Allele 1 frequency among cases

state/Hamming  
(--distance...)

Relationship/covariance  
(--make-grm-bin...)

--rel-cutoff

Distance-pheno. analysis  
(--ibs-test...)

Identity-by-descent

--genome

--homozyg...

Population stratification

--cluster

--pca

--mds-plot

--neighbour

Association analysis

Basic case/control  
(--assoc, --model)

Stratified case/control  
(--mh, --mh2, --homog)

Quantitative trait  
(--assoc, --gxe)

Regression w/ covariates  
(--linear, --logistic)

--dosage

--lasso

--test-missing

Monte Carlo permutation

Set-based tests

REML additive heritability

Family-based association

--tdt

--dfam

--qfam...

--tucc

Report postprocessing

--annotate

--clump

--gene-report

--meta-analysis

Epistasis

--fast-epistasis

--epistasis

--twolocus

Allelic scoring (--score)

R plugins (--R)

**Secondary input**

GCTA matrix  
(.grm.bin...)

**Distributed computation**

**Command-line help**

**Miscellaneous**

Tabs vs. spaces

Flag/parameter reuse

System resource usage

Pseudorandom numbers

**Resources**

1000 Genomes phase 1

Teaching materials

Gene range lists

Functional SNP attributes

**Errors and warnings**

**Output file list**

**Order of operations**

**For developers**

GitHub repository

Compilation

**Core algorithms**

Partial sum lookup

Bit population count

Ternary dot product

Vertical population count

Exact statistical tests

Multithreaded gzip

Adding new functionality

**Google groups**

plink2-users

plink2-dev

**Credits**

**File formats**

Quick [index](#) search

**A<sub>2</sub> Allele 2** frequency among controls

**CHISQ** Allelic test chi-square statistic. *Not present with 'fisher'/'fisher-midp' modifier.*

**P** Allelic test p-value

**OR** odds(allele 1 | case) / odds(allele 1 | control)

If the 'counts' modifier is present, the 5th and 6th fields are replaced with:

**C\_A** Allele 1 count among cases

**C\_U** Allele 1 count among controls

If **--ci 0.xy** has also been specified, there are three additional fields at the end:

**SE** Standard error of odds ratio estimate

**Lxy** Bottom of xy% symmetric approx. confidence interval for odds ratio

**Hxy** Top of xy% approx. confidence interval for odds ratio

### **.assoc.dosage (dosage association analysis report)**

Produced by **--dosage**.

A text file with a header line, and then usually one line per variant with the following 8-10 fields:

**CHR** Chromosome code. *Requires --map.*

**SNP** Variant identifier.

**BP** Base-pair coordinate. *Requires --map.*

**A1** Allele 1 (usually minor)

**A2** Allele 2 (usually major)

**FRQ** Allele 1 frequency

**INFO** R-squared quality metric/information content

**'BETA'/'OR'** Regression coefficient (for quantitative traits) or odds ratio

**SE** Standard error of effect (not odds ratio) estimate

**P** Association test p-value

If the 'case-control-freqs' modifier is present, the FRQ column is replaced with FRQ\_A and FRQ\_U columns reporting case and control frequencies, respectively, and NCHROBS will not include missing-phenotype samples. (Unless the phenotype is quantitative instead of case/control; then phenotypes are ignored and FRQ\_A and FRQ\_U are both equal to the overall FRQ value.)

### **.assoc.linear, .assoc.logistic (multi-covariate association analysis report)**

Produced by **--linear/--logistic**.

A text file with a header line, and **T** lines per variant typically with the following nine fields (where **T** is normally the number of terms, but the 'genotypic' and 'hethom' modifiers and the --tests flag can change this):

**CHR** Chromosome code. *Not present with 'no-snp' modifier.*

**SNP** Variant identifier. *Not present with 'no-snp'.*

**BP** Base-pair coordinate. *Not present with 'no-snp'.*

**A1** Allele 1 (usually minor). *Not present with 'no-snp'.*

**TEST** Test identifier

**NMISS** Number of observations (nonmissing genotype, phenotype, and covariates)

**'BETA'/'OR'** Regression coefficient (--linear, '--logistic beta') or odds ratio (--logistic without 'beta')

**STAT** T-statistic

**P** Asymptotic p-value for t-statistic

If **--ci 0.xy** has also been specified, the following three fields are inserted before 'STAT':

**SE** Standard error of beta/odds ratio estimate

**Lxy** Bottom of xy% symmetric approx. confidence interval

**Hxy** Top of xy% approx. confidence interval

Refer to the [PLINK 1.07 documentation](#) for more details.

### **.auto.R (R plugin function results)**

Produced by **--R**.

A text file with no header line, and one line per variant, each with at least four fields. The first

four are:

1. Chromosome code
2. Variant identifier
3. Base-pair coordinate
4. Allele 1 (corresponding to allele counts in GENO matrix; usually minor)

Subsequent fields are defined by the plugin function. Lines are permitted to contain different numbers of fields.

### **.bcf (1000 Genomes Project binary Variant Call Format, version 2)**

Variant information + sample ID + genotype call binary file, loaded with `--bcf`. **Cannot currently be generated by PLINK**; use `'--recode vcf{-fid/-iid}'` to produce a **VCF file** for now.

The specification for this format is at <http://www.1000genomes.org/wiki/analysis/variant-call-format/bcf-binary-vcf-version-2>.

### **.beagle.dat, .chr-\*.dat, .chr-\*.map (BEAGLE unphased genotype and variant information files)**

Produced by `'--recode beagle{-nomap}'`, for use by **BEAGLE**. In 'beagle' mode, one file pair is generated per autosome, while in 'beagle-nomap' mode, a single .beagle.dat file is generated containing all autosomes. **This format cannot be loaded by PLINK.**

Each .dat file produced by PLINK is a text file with three header lines, followed by one line per variant with  $2N+2$  fields where **N** is the number of samples:

<u>1st header line</u>	<u>2nd header line</u>	<u>3rd header line</u>	<u>Subsequent contents</u>
'P'	'I'	'A' for C/C pheno., 'T' for scalar	'M'
'FID'	'IID'	'PHE'	Variant identifier
FIDs, 2x per sample...	IIDs, 2x per sample	Phenotypes, 2x per sample	Allele calls (unphased)

Each .chr-\*.map file produced by PLINK is a text file with no header line, and one line per variant with the following four fields:

1. Variant identifier
2. Base-pair coordinate
3. Allele 1 (usually minor), 'X' if absent
4. Allele 2 (usually major), 'X' if absent

### **.bed (PLINK binary biallelic genotype table)**

Primary representation of genotype calls at biallelic variants. Must be accompanied by **.bim** and **.fam** files. Loaded with `--bfile`; generated in many situations, most notably when the `--make-bed` command is used. **Do not confuse this with the UCSC Genome Browser's BED format, which is totally different.**

The first three bytes should be **0x6c**, **0x1b**, and **0x01** in that order. (There are old versions of the .bed format which start with a different "magic number"; PLINK 1.9 recognizes them, but will convert sample-major files to the current variant-major format on sight. See the bottom of the [original .bed definition page](#) for details; that page also contains a more verbose version of the discussion below.)

The rest of the file is a sequence of **V** blocks of  $N/4$  (rounded up) bytes each, where **V** is the number of variants and **N** is the number of samples. The first block corresponds to the first marker in the .bim file, etc.

The low-order two bits of a block's first byte store the first sample's genotype code. ("First sample" here means the first sample listed in the accompanying .fam file.) The next two bits store the second sample's genotype code, and so on for the 3rd and 4th samples. The second byte stores genotype codes for the 5th-8th samples, the third byte stores codes for the 9th-12th, etc.

The two-bit genotype codes have the following meanings:

- 00 Homozygous for first allele in .bim file
- 01 Missing genotype
- 10 Heterozygous
- 11 Homozygous for second allele in .bim file

If **N** is not divisible by four, the extra high-order bits in the last byte of each block are always zero.

For example, consider the following text fileset:

```
test.ped:
1 1 0 0 1 0 G G 2 2 C C
1 2 0 0 2 0 A A 0 0 A C
1 3 1 2 1 2 0 0 1 2 A C
2 1 0 0 1 0 A A 2 2 0 0
2 2 0 0 2 2 A A 2 2 0 0
2 3 1 2 1 2 A A 2 2 A A
```

```
test.map:
1 snp1 0 1
1 snp2 0 2
1 snp3 0 3
```

If you load it in PLINK 1.9, a .bed file containing the following sequence of bytes will be autogenerated (you can view it with e.g. Unix `xxd`):

```
0x6c 0x1b 0x01 0xdc 0x0f 0xe7 0x0f 0x6b 0x01
```

and the following .bim file will accompany it:

```
1 snp1 0 1 G A
1 snp2 0 2 1 2
1 snp3 0 3 A C
```

(For brevity, we don't reproduce the .fam here.) We can decompose the .bed file as follows:

- The first three bytes are the magic number.
- Since there are six samples, each marker block has size 2 bytes (six divided by four, rounded up). Thus genotype data for the first marker ('snp1') is stored in the 4th and 5th bytes.
- The 4th byte value of 0xdc is **11011100** in binary. Since the low-order two bits are '00', the first sample is homozygous for the first allele for this marker listed in the .bim file, which is 'G'. The second sample has genotype code '**11**', which means she's homozygous for the second allele ('A'). The third sample's code of '**01**' designates a missing genotype call, and the fourth code of '**11**' indicates another AA.
- The 5th byte value of 0x0f is 0000**1111** in binary. This indicates that the fifth and sixth samples also have the AA genotype at snp1. There is no sample #7 or #8, so the high-order 4 bits of this byte are zero.
- The 6th and 7th bytes store genotype data for the second marker ('snp2'). The 6th byte value of 0xe7 is **11100111** in binary. The '**11**' code for the first sample means that he's homozygous for the second snp2 allele ('2'), the '**01**' code for the second sample indicates a missing call, the '**10**' code for the third indicates a heterozygous genotype, and '**11**' for the fourth indicates another homozygous '2'. The 7th byte value of 0x0f indicates that the fifth and sixth samples also have homozygous '2' genotypes.
- Finally, the 8th and 9th bytes store genotype data for the third marker ('snp3'). You can test your understanding of the file format by interpreting this by hand and then comparing to the .ped file above.

### **.bim (PLINK extended MAP file)**

Extended variant information file accompanying a .bed binary genotype table. (`--make-just-bim` can be used to update just this file.)

A text file with no header line, and one line per variant with the following six fields:

1. Chromosome code (either an integer, or 'X'/'Y'/'XY'/'MT'; '0' indicates unknown) or name
2. Variant identifier
3. Position in morgans or centimorgans (safe to use dummy value of '0')
4. Base-pair coordinate (normally 1-based, but 0 ok; limited to  $2^{31}-2$ )
5. Allele 1 (corresponding to clear bits in .bed; usually minor)
6. Allele 2 (corresponding to set bits in .bed; usually major)

Allele codes can contain more than one character. Variants with negative bp coordinates are ignored by PLINK.

### **.blocks, .blocks.det (haplotype blocks, estimated using Haploview's default algorithm)**

Produced by `--blocks`.

`.blocks` files contain one line per block, each with an asterisk followed by variant IDs.

`.blocks.det` files have a header line, followed by one line per block with the following six fields:

CHR	Chromosome code
BP1	First base-pair coordinate
BP2	Last base-pair coordinate
KB	Block length in kbs
NSNPS	Number of variants in block
SNPS	' '-delimited variant IDs

### **.clst (cluster membership file)**

Produced by `--write-cluster`. Valid input for `--within`.

A text file with no header line, and one line per sample with the following three fields:

1. Family ID
2. Within-family ID
3. Cluster name

Samples may not appear more than once.

### **.clumped, .clumped.best, .clumped.ranges (reprocessed LD-clumped reports)**

Produced by `--clump`.

The `.clumped` file normally has one header line, followed by one line per index variant (lowest p-values first) with the following 11-12 fields:

CHR	Chromosome code
F	1-based file number
SNP	Index variant identifier
BP	Base-pair coordinate
P	Index variant p-value
TOTAL	Number of other variants in clump
NSIG	Number of clumped variants with $p \geq .05$
S05	Number of clumped variants with $.01 \leq p < .05$
S01	Number of clumped variants with $.001 \leq p < .01$
S001	Number of clumped variants with $.0001 \leq p < .001$
S0001	Number of clumped variants with $p < .0001$
SP2	Comma-delimited other member IDs and file numbers. <i>Not present with --clump-verbose.</i>

With `--clump-verbose`, the header line above is repeated for every clump, instead of just appearing once, and dashed line dividers are present between clumps. Also, each nonempty clump has its own subsection, with the different header line below, one line corresponding to the index variant (with '(INDEX)' before the variant ID), a blank line, and then one line for each other clump member with the following 6-7 fields:

(blank)	Variant identifier
KB	[current variant bp coordinate] - [index bp coordinate], signed
RSQ	Squared correlation coefficient with index variant
ALLELES	Minor allele for index variant, more-common-than-expected haplotypes otherwise
F	1-based file number
P	P-value
ANNOT	Comma-delimited extra fields. <i>Requires --clump-annotate.</i>

Each nonempty clump also has the following 2-3 footer lines:

1. 'RANGE:', followed by 'chr[#]:[bp1]..[bp2]' (including `--clump-range-border` padding)
2. 'SPAN:', followed by range length in kbs
3. 'GENES w/SNPs:', followed by names of regions containing at least one variant in the clump (only present with `--clump-range`)

Finally, with `--clump-range + --clump-verbose`, there is a final footer line starting with 'GENES:', followed by names of regions physically overlapping the clump. (This is reported even for empty clumps.)

If `--clump-range` is used without `--clump-verbose`, region overlaps are reported in a separate `.clumped.ranges` file instead. This has a header line, followed by one line per clump with the following seven fields:

CHR	Chromosome code
SNP	Index variant identifier
P	Index variant p-value
N	Number of variants in clump (including index variant)
POS	Base-pair range, as 'chr[#]:[bp1]..[bp2]' (including <code>--clump-range-border</code> padding)
KB	Range length in kbs (i.e. $([bp2] - [bp1] + 1) / 1000$ )
RANGES	Comma-delimited names of overlapped <code>--clump-range</code> regions, in brackets

Finally, if `--clump-best` is specified, a `.clumped.best` file is generated. This has a header line, followed by one line per clump with the following 7-8 fields:

INDEX	Index variant identifier
PSNP	ID of best proxy (maximum r-squared), or 'NA' if there is none
RSQ	Squared correlation coefficient between index and proxy
KB	[proxy bp coordinate] - [index bp coordinate], signed
P	Proxy p-value
ALLELES	More-common-than-expected haplotypes
F	Proxy file number
(blank)	Comma-delimited extra fields for proxy variant. <i>Requires <code>--clump-annotate</code>.</i>

### **.cluster1, .cluster2, .cluster3, .cluster3.missing (hierarchical clustering reports)**

`--cluster` normally generates three files, with the extensions `.cluster1`, `.cluster2`, and `.cluster3{.missing}`. The `.cluster2` file shares the `.clst` format, so it is valid input for `--within`. The other two files are also text files with no header line.

`.cluster1` files contain one line per cluster, with a cluster name in front ('SOL-0', 'SOL-1', ...), followed by IDs of the cluster's members (formatted as FID + '\_' + IID + possibly case/control status in parentheses).

`.cluster3{.missing}` files contain one line per sample, with their FID and IID as the first two fields (*not* merged with an underscore here), followed by a sequence of nonnegative integers representing the sample's cluster assignment at each stage of the clustering process.

### **.cmh (Cochran-Mantel-Haenszel 2x2xK test report)**

Produced by `--mh/--bd`.

A text file with a header line, and then one line per variant with the following 12-14 fields (where 0.xy is the `--ci` parameter, or 0.95 if none was specified):

CHR	Chromosome code
SNP	Variant identifier
BP	Base-pair coordinate
A1	Allele 1 (usually minor)
MAF	Allele 1 frequency
A2	Allele 2 (usually major)
CHISQ	Cochran-Mantel-Haenszel statistic (1df)
P	Asymptotic p-value for CMH test statistic
OR	CMH odds ratio
SE	Standard error of odds ratio estimate
Lxy	Bottom of xy% symmetric approx. confidence interval
Hxy	Top of xy% approx. confidence interval
CHISQ_BD	Breslow-Day test statistic. <i>Requires <code>--bd</code>.</i>
P_BD	Asymptotic p-value for Breslow-Day test statistic. <i>Requires <code>--bd</code>.</i>

### **.cmh2 (Cochran-Mantel-Haenszel 1xjxK test report)**

Produced by `--mh2`.

A text file with a header line, and then one line per variant with the following five fields:

CHR	Chromosome code
SNP	Variant identifier
CHISQ	Cochran-Mantel-Haenszel 1xjxK test statistic

DF Chi-square degrees of freedom  
P Asymptotic p-value

(DF was not directly reported by PLINK 1.07.)

### **.cnv (segmental copy number variant data)**

Produced by postprocessing the output of [Birdsuite](#) or a similar package. Loaded with `--cnv-list/--cfile`. Must be accompanied by a `.fam` file.

A text file with an optional header line, and one line per segmental call with the following eight fields:

FID Family ID  
IID Within-family ID  
CHR Chromosome code  
BP1 First base-pair coordinate  
BP2 Last base-pair coordinate  
TYPE Number of copies of variant  
SCORE Confidence score associated with variant (safe to use dummy value of '0')  
SITES Number of probes in the variant (safe to use dummy value of '0')

### **.cnv.indiv (per-sample segment summary)**

Produced whenever `--cfile/--cnv-list` loading completes.

A text file with a header line, and one line per sample with the following 6-7 fields:

FID Family ID  
IID Within-family ID  
PHE Phenotype  
NSEG Number of segments that sample has  
KB Total kilobase distance spanned by segments  
KBAVG Average segment size  
COUNT (Only present with `--cnv-count`, which is not yet implemented.)

### **.cnv.overlap (overlapping CNV segment report)**

Produced by `--cnv-check-no-overlap`.

A text file with a header line, and one line per overlap with the following five fields:

FID Family ID  
IID Within-family ID  
CHR Chromosome code  
BP1 Segment start (base-pair units)  
BP2 Segment end

### **.cnv.summary (per-variant CNV summary)**

Produced whenever `--cfile/--cnv-list` loading completes.

A text file with a header line, and one line per variant with the following five fields:

CHR Chromosome code  
SNP Variant identifier  
BP Base-pair coordinate  
AFF CNV count at variant, all cases  
UNAFF CNV count at variant, all controls

### **.cov (covariate table)**

Produced by `--write-covar`, `--make-bed`, and `--recode` when an input covariate table has been named with `--covar`. Valid input for `--covar`.

A text file with a header line, and one line per sample with the following 2+**C** or 6+**C** fields (where **C** is the number of covariates):

FID Family ID  
IID Within-family ID  
PAT Paternal within-family ID. Requires `--with-phenotype` without 'no-parents'.  
MAT Maternal within-family ID. Requires `--with-phenotype` without 'no-parents'.  
SEX Sex. Requires `--with-phenotype` without 'no-sex'.

<i>PHENOTYPE</i>	Main phenotype value. <i>Only present with --with-phenotype.</i>
Covariate IDs...	Covariate values

Note that `--covar` can also be used with files lacking a header row.

### **.dfam (sib-TDT association report)**

Produced by `--dfam`.

A text file with a header line, and then one line per variant with the following eight fields:

CHR	Chromosome code
SNP	Variant identifier
A1	Allele 1 (usually minor)
A2	Allele 2 (usually major)
OBS	Number of observed A1 alleles
EXP	Expected number of A1 alleles
CHISQ	Sib-TDT test statistic
P	Asymptotic p-value for sib-TDT test statistic

### **.diff (merge conflict report)**

Produced by `--merge/--bmerge` + `--merge-mode` 6 or 7.

A text file with a header line, and then one line per conflict with the following five fields:

SNP	Variant identifier
FID	Family ID
IID	Within-family ID
NEW	Genotype in merge fileset (named in <code>--merge-list file/--merge/--bmerge</code> )
OLD	Genotype in reference fileset (loaded with e.g. <code>--bfile</code> )

### **.dist (genomic Hamming distance matrix)**

Produced by `--distance`.

A tab-delimited text file that is either lower-triangular (first line has only one entry containing the {genome 1-genome 2} Hamming distance, second line has two entries containing the {genome 1-genome 3} and {genome 2-genome 3} Hamming distances in that order, etc.) or square. If square, the upper-right triangle may be either zeroed out or the mirror-image of the lower-left triangle, depending on whether the 'square0' or 'square' modifier was used.

When missing values are present, the affected raw Hamming distances are rescaled to be comparable to pairwise distances unaffected by missing data.

### **.dupvar (duplicate-position-and-alleles variant report)**

Produced by `--list-duplicate-vars`.

Normally a tab-delimited text file with a header line, followed by one line per duplicate variant group with the following 4 columns:

CHR	Chromosome code
POS	Base-pair coordinate
ALLELES	Comma-separated allele codes
IDS	<b>Space</b> -separated variant IDs

With the 'ids-only' modifier, the header and the position/allele columns are omitted; only space-delimited lists of variant IDs remain. (This form is directly usable with `--extract/--exclude`.)

With 'require-same-ref' (and without 'ids-only'), the ALLELES column is replaced with the following two columns:

REF	A2 allele
ALT	A1 allele (will become a comma-separated list in PLINK 2.0)

### **.eigenvec, .eigenvec.var (principal components)**

Produced by `--pca`. Accompanied by an `.eigenval` file, which contains one eigenvalue per line.

The `.eigenvec` file is, by default, a space-delimited text file with no header line and `2+V` columns per sample, where `V` is the number of requested principal components. The `--pca 'header'` modifier causes a header line to be written, and the 'tabs' modifier makes this file tab-delimited. The first two columns are the sample's FID/IID, and the rest are principal

component weights in the same order as the .eigenval values (if the header line is present, these columns are titled 'PC1', 'PC2', ...).

With the 'var-wts' modifier, an .eigenvec.var file is also generated. It replaces the FID/IID columns with 'CHR', 'VAR', 'A1', and 'A2' columns containing chromosome codes, variant IDs, A1 alleles, and A2 alleles, respectively; otherwise the formats are identical.

### **.epi.[cc/co/qt], .epi.[cc/co/qt].summary (epistatic interaction scan reports)**

Produced by `--epistasis` and `--fast-epistasis`. 'cc' secondary extension indicates a case/control test, 'co' indicates '--fast-epistasis case-only', and 'qt' indicates --epistasis linear regression on a quantitative trait.

The main report is normally a text file with a header line, followed by one line per variant pair clearing the --epi1 threshold with the following 5-7 fields:

CHR1	Variant 1 chromosome code
SNP1	Variant 1 identifier
CHR2	Variant 2 chromosome code
SNP2	Variant 2 identifier
'OR_INT'/BETA_INT'	Odds ratio (case/control) or regression coefficient (QT). <i>Requires --epistasis.</i>
STAT	Chi-square statistic
DF	Chi-square degrees of freedom. <i>Only present with 'boost'.</i>
P	Chi-square p-value. <i>Not present with --fast-epistasis 'nop' modifier.</i>

The .summary file is a text file with a header line, followed by one line per variant (or just one line per variant in set #1, if 'set-by-set' or 'set-by-all' was specified) with the following 7-8 fields:

CHR	Chromosome code
SNP	Variant identifier
N_SIG	Number of 'significant' (based on --epi2 value) epistatic test results
N_TOT	Total number of valid test results
PROP	Proportion significant. <i>Not always present in intermediate --parallel files.</i>
BEST_CHISQ	Largest chi-square statistic (approximate when 'boost' test and $\leq$ --epi1 threshold)
BEST_CHR	Chromosome of largest-statistic variant
BEST_SNP	ID of largest-statistic variant

For the 'boost' test, the BEST\_CHISQ/BEST\_CHR/BEST\_SNP entry occasionally doesn't correspond to lowest p-value, since DF is variable.

For two-set tests, if variant  $v_1$  is in both sets but  $v_2$  is only in set #1, the  $v_1$ - $v_2$  test is only counted in the  $v_2$  summary row. (This is a change from PLINK 1.07.)

### **.fam (PLINK sample information file)**

Sample information file accompanying a .bed binary genotype table. (`--make-just-fam` can be used to update just this file.) Also generated by '`--recode lgen`' and '`--recode rlist`'.

A text file with no header line, and one line per sample with the following six fields:

1. Family ID ('FID')
2. Within-family ID ('IID'; cannot be '0')
3. Within-family ID of father ('0' if father isn't in dataset)
4. Within-family ID of mother ('0' if mother isn't in dataset)
5. Sex code ('1' = male, '2' = female, '0' = unknown)
6. Phenotype value ('1' = control, '2' = case, '-9'/'0'/non-numeric = missing data if case/control)

With the use of `additional loading flag(s)`, PLINK can also correctly interpret some .fam files missing one or more of these fields.

If there are any numeric phenotype values other than  $\{-9, 0, 1, 2\}$ , the phenotype is interpreted as a quantitative trait instead of case/control status. In this case, -9 normally still designates a missing phenotype; use `--missing-phenotype` if this is problematic.

Several PLINK commands (e.g. `--cluster`) merge the FID and IID with an underscore in their reports; for example, a sample with FID = 'Chang' and IID = 'Christopher' would be referenced as 'Chang\_Christopher'. We preserve this behavior for backwards compatibility, so you should avoid using underscores in FIDs and IIDs (consider '~' instead).

If your case/control phenotype is encoded as '0' = control and '1' = case, you'll need to specify `--1` to load it properly.

### **.flipscan, .flipscan.verbose (case/control strand inconsistency report)**

Produced by `--flip-scan`.

The .flipscan file is a text file with a header line, and one line per variant with the following 11 fields:

CHR	Chromosome code
SNP	Variant identifier
BP	Base-pair coordinate
A1	Allele 1 (usually minor)
A2	Allele 2 (usually major)
F	Allele 1 frequency
POS	Number of positive LD matches
R_POS	Positive LD match average correlation
NEG	Number of negative LD matches
R_NEG	Negative LD match average correlation
NEGSNPS	Negative LD match ID(s), ' ' delimited

If the 'verbose' modifier is present, a .flipscan.verbose file is also generated. This is a text file with a header line, and one line per relevant variant pair (i.e. index variant has at least one negative LD match, and case and/or control correlation has sufficient absolute value) with the following nine fields:

CHR_INDX	Chromosome code
SNP_INDX	Index variant identifier
BP_INDX	Index variant base-pair coordinate
A1_INDX	Index variant allele 1
SNP_PAIR	Second variant identifier
BP_PAIR	Second variant base-pair coordinate
A1_PAIR	Second variant allele 1
R_A	Case-only correlation
R_U	Control-only correlation

### **.frq (basic allele frequency report)**

Produced by `--freq`. Valid input for `--read-freq`.

A text file with a header line, and then one line per variant with the following six fields:

CHR	Chromosome code
SNP	Variant identifier
A1	Allele 1 (usually minor)
A2	Allele 2 (usually major)
MAF	Allele 1 frequency
NCHROBS	Number of allele observations

### **.frq.cc (case/control phenotype-stratified allele frequency report)**

Produced by `'--freq case-control'`. *Not* valid input for `--read-freq`.

A text file with a header line, and then one line per variant with the following eight fields:

CHR	Chromosome code
SNP	Variant identifier
A1	Allele 1 (usually minor)
A2	Allele 2 (usually major)
MAF_A	Allele 1 frequency in cases
MAF_U	Allele 1 frequency in controls
NCHROBS_A	Number of case allele observations
NCHROBS_U	Number of control allele observations

### **.frq.count (basic allele count report)**

Produced by `'--freq counts'`. Valid input for `--read-freq`.

A text file with a header line, and then one line per variant with the following seven fields:

CHR	Chromosome code
-----	-----------------

SNP Variant identifier  
 A1 Allele 1 (usually minor)  
 A2 Allele 2 (usually major)  
 C1 Allele 1 count  
 C2 Allele 2 count  
 G0 Missing genotype count (so  $C1 + C2 + 2 * G0$  is constant on autosomal variants)

### **.frq.strat (cluster-stratified allele frequency report)**

Produced by `--freq` when used with `--within/--family`. Not valid input for `--read-freq`.

A text file with a header line, and then **C** lines per variant (where **C** is the number of clusters) with the following 8-9 lines:

CHR Chromosome code  
 SNP Variant identifier  
 CLST Cluster identifier  
 A1 Allele 1 (usually minor)  
 A2 Allele 2 (usually major)  
 MAF Allele 1 frequency in cluster  
 MAC Allele 1 count in cluster  
 NCHROBS Number of allele observations in cluster

### **.frqx (genotype count report)**

Produced by `--freqx`. Valid input for `--read-freq`.

A text file with a header line, and then one line per variant with the following ten fields:

CHR Chromosome code  
 SNP Variant identifier  
 A1 Allele 1 (usually minor)  
 A2 Allele 2 (usually major)  
 C(HOM A1) A1 homozygote count  
 C(HET) Heterozygote count  
 C(HOM A2) A2 homozygote count  
 C(HAP A1) Haploid A1 count (includes male X chromosome)  
 C(HAP A2) Haploid A2 count  
 C(MISSING) Missing genotype count

### **.fst (fixation index report)**

Produced by `--fst`.

A text file with a header line, and then one line per autosomal diploid variant with the following five fields:

CHR Chromosome code  
 SNP Variant identifier  
 POS Base-pair coordinate  
 NMISS Number of genotype calls considered  
 FST Wright's  $F_{ST}$  estimate, via Weir and Cockerham's method

### **.gen (Oxford genotype file format)**

Native text genotype file format for Oxford statistical genetics tools, such as `IMPUTE2` and `SNPTEST`. Should always be accompanied by a `.sample` file. Loaded with `--data/--gen`, and produced by `'--recode oxford'`.

A text file with no header line, and one line per variant with  $3N+5$  fields where **N** is the number of samples. Each line stores information for a single SNP. The first five fields are:

1. Chromosome code (or ignorable value; see `--oxford-single-chr`)
2. RS ID of SNP
3. Base-pair coordinate
4. Allele 1 (usually minor)
5. Allele 2 (usually major)

Each subsequent triplet of values then indicate likelihoods of homozygote A1, heterozygote, and homozygote A2 genotypes at this SNP, respectively, for one sample. If they add up to less than one, the remainder is a no-call probability weight.

Since the PLINK 1 binary format cannot represent genotype probabilities, calls with uncertainty greater than `0.1` are currently treated as missing, and the rest are treated as hard

calls. (This behavior can be changed with `--hard-call-threshold`.) We plan to remove this limitation in the future.

### **.genome (identity-by-descent report)**

Produced by `--genome`. Valid input for `--read-genome`.

A text file with a header line, and one line per pair of distinct samples typically with the following 14 fields:

FID1	First sample's family ID
IID1	First sample's within-family ID
FID2	Second sample's family ID
IID2	Second sample's within-family ID
RT	Relationship type inferred from .fam/.ped file
EZ	IBD sharing expected value, based on just .fam/.ped relationship
Z0	P(IBD=0)
Z1	P(IBD=1)
Z2	P(IBD=2)
PI_HAT	Proportion IBD, i.e. $P(\text{IBD}=2) + 0.5 \cdot P(\text{IBD}=1)$
PHE	Pairwise phenotypic code (1, 0, -1 = case-case, case-ctrl, and ctrl-ctrl pairs, respectively)
DST	IBS distance, i.e. $(\text{IBS2} + 0.5 \cdot \text{IBS1}) / (\text{IBS0} + \text{IBS1} + \text{IBS2})$
PPC	IBS binomial test
RATIO	HETHET : IBS0 SNP ratio (expected value 2)

With the 'full' modifier, there are five additional fields at the end:

IBS0	Number of IBS 0 nonmissing variants
IBS1	Number of IBS 1 nonmissing variants
IBS2	Number of IBS 2 nonmissing variants
HOMHOM	Number of IBS 0 SNP pairs used in PPC test
HETHET	Number of IBS 2 het/het SNP pairs used in PPC test

### **.grm (GCTA text relationship matrix)**

Produced by `--make-grm-gz`. Readable by `--grm-gz`.

A text file with no header line, and one line per pair of samples (not necessarily distinct) with the following four fields:

1. 1-based index of first sample in .grm.id file
2. 1-based index of second sample in .grm.id file
3. Number of observations (variants where neither sample has a missing call)
4. Relationship value

### **.grm.N.bin, .grm.bin (GCTA 1.1+ triangular binary relationship matrix)**

Produced by `--make-grm-bin`. Readable by `--grm-bin`.

These files contain single-precision (4-byte) floating point values. Using 1-based matrix indices, the first value in each file is the (1, 1) relationship value (.grm.bin) or observation count (.grm.N.bin); the second and third values are the (2, 1) and (2, 2) relationships/counts; the fourth through sixth values are the (3, 1), (3, 2) and (3, 3) relationships/counts in that order; and so on.

Note that .grm.bin files generated by GCTA versions before 1.1 have a different format.

### **.gvar (genetic variant format)**

Produced by packages such as `Birdsuite`. Loaded with `--gfile`. Must be accompanied by .fam and .map files.

A text file with no header line, and one line per variant call with the following seven fields:

1. Family ID
2. Within-family ID
3. Variant name
4. Code for allele from first parent
5. Copy number for first allele (can be non-integer)
6. Code for allele from second parent
7. Copy number for second allele

### **.het (method-of-moments F coefficient estimates)**

Produced by `--het`.

A text file with a header line, and one line per sample with the following six fields:

FID	Family ID
IID	Within-family ID
O(HOM)	Observed number of homozygotes
E(HOM)	Expected number of homozygotes
N(NM)	Number of non-missing autosomal genotypes
F	Method-of-moments F coefficient estimate

### **.hh (heterozygous haploid and nonmale Y chromosome call list)**

Produced automatically when the input data contains heterozygous calls where they shouldn't be possible (haploid chromosomes, male X/Y), or there are nonmissing calls for nonmales on the Y chromosome.

A text file with one line per error (sorted primarily by variant ID, secondarily by sample ID) with the following three fields:

1. Family ID
2. Within-family ID
3. Variant ID

### **.hom (run-of-homozygosity list)**

Produced when a flag in the `--homozyg` family is present. Accompanied by at least a `.hom.indiv` and a `.hom.summary` file.

A text file with a header line, and one line per run with the following thirteen fields:

FID	Family ID
IID	Within-family ID
PHE	Phenotype value
CHR	Chromosome code
SNP1	ID of first SNP in run
SNP2	ID of last SNP in run
POS1	Base-pair coordinate of SNP1
POS2	Base-pair coordinate of SNP2
KB	Length of region in kb
NSNP	Number of SNPs in run
DENSITY	Inverse SNP density in kb/SNP
PHOM	Proportion of calls homozygous
PHET	Proportion of calls heterozygous

Note that PHOM + PHET can be less than 1 when missing calls are present.

### **.hom.indiv (sample-based runs-of-homozygosity report)**

Produced when a flag in the `--homozyg` family is present.

A text file with a header line, and one line per sample with the following six fields:

FID	Family ID
IID	Within-family ID
PHE	Phenotype value
NSEG	Number of runs of homozygosity
KB	Total length of runs (kb)
KBAVG	Average length of runs (kb)

### **.hom.overlap (run-of-homozygosity pool list)**

Produced by `'--homozyg group{-verbose}'`.

`.hom.overlap` files contain a header line, and **P**+2 lines per segment pool (where **P** is the number of segments in the pool) with the following 13 fields:

<u>Header</u>	<u>First <b>P</b> lines</u>	<u>Last two lines</u>
POOL	Pool ID	(same)
FID	Family ID	'CON'/'UNION'
IID	Within-family ID	<b>P</b>
PHE	Phenotype value	[case ct]:[noncase ct]

CHB	Chromosome code	(same)
SNP1	ID of first SNP in segment	(same)
SNP2	ID of last SNP in segment	(same)
BP1	Base-pair coordinate of SNP1	(same)
BP2	Base-pair coordinate of SNP2	(same)
KB	Length of region in kb	(same)
NSNP	Number of SNPs in run	(same)
NSIM	Number of matching segments in pool	'NA'
GRP	Allelic-match group (see <a href="#">--homozyg-match</a> )	'NA'

The second-to-last line for each pool describes the consensus match segment, while the last line describes the union of all segments in the pool. Pools are separated by blank lines, and sorted primarily by pool size (largest first) and secondarily by physical position. The first pool in the file has ID 'S1', the second pool has ID 'S2', etc.

PLINK 1.07's production of this file has a minor bug and a few quirks (pairwise allelic matches are judged from  $([\# \text{ mismatches on joint-homozygous overlapping variants}] / [\# \text{ of overlapping variants}])$  instead of  $([\# \text{ mismatches on joint-homozygous overlapping variants}] / [\# \text{ of joint-homozygous overlapping variants}])$ , contrary to the documentation; pools are sorted by *reverse* physical position; some ID numbers are skipped; samples within an allelic-match group written in an unsorted order) which are not replicated by PLINK 1.9.

### **.hom.overlap.S\*.verbose (single ROH pool report)**

'[--homozyg group-verbose](#)' also produces one `.hom.overlap.[pool ID].verbose` file *per pool*. (Be careful with this, lest you inadvertently fill up your entire hard drive.) These files each contain **G**+3 sections, where **G** is the number of allelic-match groups. (Note that this format was not really intended to be machine-readable; if there is sufficient interest, we may clean it up in the future.)

The first section has a header line, followed by one line per sample in the pool with the following four fields:

```
(blank) '1', '2', etc.
FID      Family ID
IID      Within-family ID
GRP      Allelic-match group (without trailing '*')s
```

It ends with a single blank line.

The second section has a header line, followed by a blank line, followed by one line per variant in the segment union with the following **P**+1 fields:

```
SNP      Variant identifier
'1', '2', etc.  '/'-separated genotype call, [bracketed] when it's part of a ROH
```

There are single blank lines marking the beginning and end of the consensus match segment, and two consecutive blank lines at the end of this section.

The next **G** sections each start with the following **S**+6 header lines (where **g** is the 1-based allelic-match group index, **S** is the size of the group, and **p** is the 1-based index assigned to the sample in the first field of the first section):

```
1. 'Group g'
2. (blank line)
3-(S+2). 4 fields: 'p', FID, IID, phenotype value
S+3. (blank line)
S+4. (blank line)
S+5. S+1 fields: 'SNP', p1, ..., pS
S+6. (blank line)
```

This is followed by one line per variant with the following **S**+2 fields:

```
1. Variant identifier
2. Consensus haplotype, or '?' if there isn't one
3-(S+2). Genotype call from section 2 (including brackets)
```

Single blank lines mark the beginning and end of the consensus match segment, as well as the end of the section.

The final section starts with two additional blank lines, followed by one line per variant with the following **G**+1 fields:

1. Variant identifier
- 2-(**G**+1). Consensus haplotype for allelic-match group

### **.hom.summary (SNP-based runs-of-homozygosity report)**

Produced when a flag in the `--homozyg` family is present.

A text file with a header line, and one line per SNP with the following five fields:

CHR	Chromosome code
SNP	Variant identifier
BP	Base-pair coordinate
AFF	Number of cases with a run-of-homozygosity including this SNP
UNAFF	Number of non-cases with a ROH including this SNP

Note that samples with missing phenotypes are counted in the 'UNAFF' column. If the phenotype is quantitative, everyone will be counted in 'UNAFF'.

### **.homog (chi-square partitioning odds ratio homogeneity test report)**

Produced by `--homog`.

A text file with a header line, followed by **K**+3 lines per variant with the following 13 fields (where **K** > 1 is the number of clusters):

CHR	Chromosome code
SNP	Variant identifier
A1	Allele 1 (usually minor)
A2	Allele 2 (usually major)
F_A	Case A1 frequency
F_U	Control A1 frequency
N_A	Case allele count
N_U	Control allele count
TEST	Type of test: one of {'TOTAL', 'ASSOC', 'HOMOG', cluster names}
CHISQ	Chi-square association statistic
DF	Degrees of freedom
P	Asymptotic p-value
OR	Odds ratio

### **.hwe (Hardy-Weinberg equilibrium exact test statistic report)**

Produced by `--hardy`.

A text file with a header line, and one line per marker with the following nine fields:

CHR	Chromosome code
SNP	Variant identifier
TEST	Type of test: one of {'ALL', 'AFF', 'UNAFF', 'ALL(QT)', 'ALL(NP)'}
A1	Allele 1 (usually minor)
A2	Allele 2 (usually major)
GENO	'/'-separated genotype counts (A1 hom, het, A2 hom)
O(HET)	Observed heterozygote frequency
E(HET)	Expected heterozygote frequency
P	Hardy-Weinberg equilibrium exact test p-value

### **.ibc (GCTA inbreeding coefficient report)**

Produced by `--ibc`.

A text file with a header line, and one line per sample with the following six fields:

FID	Family ID
IID	Within-family ID
NOMISS	Number of nonmissing genotype calls
Fhat1	Variance-standardized relationship minus 1
Fhat2	Excess homozygosity-based inbreeding estimate (same as PLINK <code>--het</code> )
Fhat3	Estimate based on correlation between uniting gametes

### **.imiss (sample-based missing data report)**

Produced by `--missing`, with a companion `.imiss` file.

A text file with a header line, and one line per sample with the following six fields:

FID	Family ID
IID	Within-family ID
MISS_PHENO	Phenotype missing? (Y/N)
N_MISS	Number of missing genotype call(s), not including obligatory missings
N_GENO	Number of potentially valid call(s)
F_MISS	Missing call rate

### **.info (Haploview map file)**

Produced by '--recode HV{-1chr}', for use by [Haploview](#). Accompanies a [.ped](#) file. With '--recode HV', one [.ped](#) + [.info](#) fileset is generated per chromosome, and the full file extensions are of the form [.chr-\[chromosome number\].info](#). **This format cannot be loaded by PLINK.**

A text file with no header line, and one line per variant with the following two fields:

1. Variant identifier
2. Base-pair coordinate

### **.lasso (LASSO variant effect size estimates)**

Produced by [--lasso](#). Valid input for [--score](#).

A text file with a header line, and one line per variant with the following four fields:

CHR	Chromosome code (or 'COV' for covariates)
SNP	Variant/covariate identifier
A1	Allele 1 (usually minor; 'NA' for covariates)
EFFECT	A1 effect size estimate on <i>normalized</i> phenotype ('NA' on monomorphic variants)

### **.ld (inter-variant correlation table or matrix)**

Produced by [--r/--r2](#).

If a matrix format was requested, the output is structured like a [.dist](#) file. (With the 'spaces' modifier, it's space-delimited instead of tab-delimited.)

If a table report was requested instead, the file contains a header line, followed by one line per filtered variant pair with the following 7-11 fields:

CHR_A	Chromosome code for first variant
BP_A	Base-pair coordinate of first variant
SNP_A	ID of first variant
MAF_A	Allele 1 frequency for first variant. <i>Requires 'with-freqs'</i> .
CHR_B	Chromosome code for second variant
BP_B	Base-pair coordinate of second variant
SNP_B	ID of second variant
PHASE	In-phase allele pairs. <i>Requires 'in-phase'</i> .
MAF_B	Allele 1 frequency for second variant. <i>Requires 'with-freqs'</i> .
'R'/'R2'	Correlation coefficient (squared if <a href="#">--r2</a> ).
'D'/'DP'	Linkage disequilibrium D, or Lewontin's D-prime. <i>Requires 'd'/'dprime'/'dprime-signed'</i> .

### **.ldset (high-LD same-set variant pair report)**

Produced by [--set-r2](#) when the 'write' modifier is present.

A text file with no header line, and one section per set. A section has one line for each variant in the set, starting with the following two fields:

1. Set name
2. Variant ID

These are followed by a (space-delimited) list of ID(s) of other same-set variants which have pairwise  $r^2 \geq 0.5$  with the current variant.

Note that sets containing no significant variants are not present in this report; this is a change from PLINK 1.07's [--write-set-r2](#)'s behavior. (Use '[--set-p 1](#)' if this is a problem.)

### **.lgen (PLINK long-format genotype file)**

Produced by '--recode lgen' and '--recode lgen-ref'. Accompanied by a [.fam](#), [.map](#), and possibly a [.ref](#) file. Loaded with [--lfile](#).

A text file with no header line, and one line per genotype call (or just not-homozygous-major calls if 'lgen-ref' was invoked) usually with the following five fields:

1. Family ID
2. Within-family ID
3. Variant identifier
4. Allele call 1 ('0' for missing)
5. Allele call 2

There are several variations which are also handled by PLINK; see the [original discussion](#) for details.

### **.list (genotype list file)**

Produced by '[--recode list](#)'. **This format cannot be loaded by PLINK.**

A text file with no header line, and four lines per variant. Each line starts with the following three fields:

1. Chromosome code
2. Variant identifier
3. Genotype ('00' for missing)

This is followed by two additional fields (FID, then IID) for each sample with the specified genotype call at the variant.

### **.imiss (variant-based missing data report)**

Produced by [--missing](#), with a companion [.imiss](#) file.

A text file with a header line, and **K** line(s) per variant with the following 5-7 fields (where **K** is the number of cluster(s) if [--within/--family](#) was specified, or 1 if it wasn't):

CHR	Chromosome code
SNP	Variant identifier
CLST	Cluster identifier. <i>Only present with --within/--family.</i>
N_MISS	Number of missing genotype call(s), not counting obligatory missings
N_CLST	Cluster size (does not include nonmales on Ychr). <i>Only present with --within/--family.</i>
N_GENO	Number of potentially valid call(s)
F_MISS	Missing call rate

### **.map (PLINK text fileset variant information file)**

Variant information file accompanying a [.ped](#) text pedigree + genotype table. Also generated by '[--recode rlist](#)'.

A text file with no header file, and one line per variant with the following 3-4 fields:

1. Chromosome code. PLINK 1.9 also permits contig names here, but most older programs do not.
2. Variant identifier
3. Position in morgans or centimorgans (optional; also safe to use dummy value of '0')
4. Base-pair coordinate

All lines must have the same number of columns (so either no lines contain the morgans/centimorgans column, or all of them do).

### **.mdist (genomic distance proportion matrix)**

Produced by '[--distance 1-ibs](#)' and [--distance-matrix](#).

A text file that is space-delimited if produced with [--distance-matrix](#) and tab-delimited otherwise. Shape and contents are identical to that of [.dist](#) files, except that all values are divided by twice the total variant count to convert them from Hamming distances to fractions between 0 and 1.

### **.mdist.missing (identity-by-missingness matrix)**

Produced by '[--cluster missing](#)'.

A triangular space-delimited text file with identity-by-missingness coefficients.

### **.mds (Haploview-friendly multidimensional scaling report)**

Produced by [--mds-plot](#).

A text file with a header line with the following **D**+3 fields (where **D** is the number of

requested dimensions), and one line per sample with the same fields:

FID Family ID  
IID Within-family ID  
SOL Cluster index (0-based)  
Cx... Position on dimension x (1-based dimension indices)

### **.mendel, .imendel, .fmendel, .Imendel (Mendel error reports)**

Produced by [--mendel](#).

The .mendel file is a text file with a header line, and one line per error with the following six columns:

FID Family ID  
KID Child within-family ID  
CHR Chromosome code  
SNP Variant identifier  
CODE Numeric error code  
ERROR Description of error

Note that '\*'/\*' in the error description does not (necessarily) refer to a missing genotype call; instead, it means a Mendel error is present regardless of what that parent's genotype is.

The .Imendel file has a header line, and one line per variant with the following three columns:

CHR Chromosome code  
SNP Variant identifier  
N Number of Mendel errors

The .imendel file has a header line, and one subsection per nuclear family. Each subsection contains one line per family member with the following three columns:

FID Family ID  
IID Within-family ID  
N Number of errors implicating this sample (only considering nuclear family)

Samples may appear more than once in this file.

Finally, the .fmendel file has a header line, and one line per nuclear family with the following five columns:

FID Family ID  
PAT Paternal within-family ID (0 if missing)  
MAT Maternal within-family ID (0 if missing)  
CHLD Number of offspring in nuclear family  
N Number of Mendel errors in nuclear family

### **.meta (meta-analysis)**

Produced by [--meta-analysis](#).

A text file with a header line, and then one line per analyzed variant with the following 8-(**F**+14) fields (where **F** is the number of input files):

CHR Chromosome code. *Not present with 'no-map' modifier.*  
BP Base-pair coordinate. *Not present with 'no-map' modifier.*  
SNP Variant identifier  
A1 Allele 1. *Not present with 'no-map' or 'no-allele' modifier.*  
A2 Allele 2. *Not present with 'no-map' or 'no-allele' modifier.*  
N Number of valid studies for variant  
P Fixed-effects meta-analysis p-value  
P(R) Random-effects meta-analysis p-value  
'BETA'/'OR' Fixed-effects BETA/OR estimate  
'BETA(R)'/OR(R)' Random-effects BETA/OR estimate  
Q p-value for Cochran's Q statistic  
I I<sup>2</sup> heterogeneity index (0-100 scale)  
WEIGHTED\_Z Weighted Z-score, as computed by METAL. *Requires 'weighted-z' modifier.*  
P(WZ) p-value for weighted Z-score. *Requires 'weighted-z' modifier.*  
Study x (0-based input file indices) effect estimate. *Requires 'study'*

*F[x]... modifier.*

### **.mibs (identity-by-state matrix)**

Produced by '`--distance ibs`' and `--matrix`.

A text file that is space-delimited if produced with `--distance-matrix` and tab-delimited otherwise. Possible shapes are the same as for `.dist` and `.mdist` files. Each identity-by-state value is just equal to one minus the corresponding `.mdist` value.

### **.missing (case/control nonrandom missingness test report)**

Produced by `--test-missing`.

A text file with a header line, and then one line per nondegenerate variant with the following 5 fields:

CHR	Chromosome code
SNP	Variant identifier
F_MISS_A	Missing call frequency, cases
F_MISS_U	Missing call frequency, controls
P	Fisher's exact test p-value

### **.missing.hap (adjacent variant-based nonrandom missingness test report)**

Produced by `--test-mishap`.

A text file with a header line, and then one section per autosomal diploid variant with 5+ missing calls. Each section contains one line per considered flanking haplotype, followed by a 'HETERO' line covering flanking heterozygosity (just one flanking call needs to be heterozygous), with the following 9 fields:

SNP	Central variant identifier
HAPLOTYPE	Haplotype allele(s), or 'HETERO'
F_0	Haplotype frequency, central call missing
F_1	Haplotype frequency, central call nonmissing
M_H1	#(central call missing, this hap.) / #(central call nonmissing, this hap.)
M_H2	#(central call missing, other hap.) / #(central call nonmissing, other hap.)
CHISQ	Chi-square statistic
P	Chi-square p-value
FLANKING	Flanking variant ID(s), ' ' delimited

Haplotype frequencies are estimated via the EM algorithm.

### **.model (case/control full model association report)**

Produced by `--model`.

A text file with a header line, and then 1-5 lines per variant with the following 8-10 fields:

CHR	Chromosome code
SNP	Variant identifier
A1	A1 allele (usually minor)
A2	A2 allele (usually major)
TEST	Type of test: one of {'GENO', 'TREND', 'ALLELIC', 'DOM', 'REC'}
AFF	'/'-separated genotype or allele counts among cases
UNAFF	'/'-separated genotype or allele counts among controls
CHISQ	Chi-square statistic. <i>Not present with 'fisher'/'fisher-midp' modifier.</i>
DF	Chi-square degrees of freedom. <i>Not present with 'fisher'/'fisher-midp'.</i>
P	P-value

### **.\*.mperm (max(T) permutation test report)**

Produced by several association analysis commands when the '`mperm=[value]`' modifier is used.

A text file with a header line, and then typically one line per variant with the following four fields:

CHR	Chromosome code
SNP	Variant identifier
EMP1	Empirical p-value (pointwise), or lower-p-value permutation count
EMP2	Corrected empirical p-value (max(T) familywise) or permutation count

In the `--linear/--logistic no-snp` case, there is instead one line per variable with the following three fields:

TEST	Test identifier
EMP1	Empirical p-value, or lower-p-value permutation count
NP	Number of permutations performed

### **.nearest (nearest neighbor distance report)**

Produced by `--neighbour`.

A text file with a header line, and  $n_2 - n_1 + 1$  lines per sample with the following 7-8 fields:

FID	Family ID
IID	Within-family ID
NN	Nearest neighbor level
MIN_DST	IBS distance of NNth nearest neighbor
Z	Z score of MIN_DST
FID2	FID of NNth nearest neighbor
IID2	IID of NNth nearest neighbor
PROP_DIFF	Proportion of neighbors below <code>--ppc</code> threshold. <i>Not present without <code>--ppc</code>.</i>

### **.occur.dosage (dosage data variant occurrence report)**

Produced by `'--dosage occur'`.

A text file with no header line, and one line per variant with the following 2 fields:

1. Variant ID
2. Number of input files the variant appears in

### **.out.dosage (merged dosage data file)**

Produced by `--write-dosage`.

A text file with a header line, and one line per variant with the following 3 initial fields:

SNP	Variant ID
A1	Allele 1 (usually minor)
A2	Allele 2 (usually major)

This is followed by  $N$  2-field blocks in the header line (with FID/IIDs), and  $N$  blocks of  $m$  dosage data fields in subsequent lines (where  $m$  is the `--dosage 'format'` parameter).

### **.ped (PLINK/MERLIN/Haploview text pedigree + genotype table)**

Original standard text format for sample pedigree information and genotype calls. Normally must be accompanied by a `.map` file; Haploview requires an accompanying `.info` file instead. Loaded with `--file`, and produced by `--recode`.

Contains no header line, and one line per sample with  $2V + 6$  fields where  $V$  is the number of variants. The first six fields are the same as those in a `.fam` file. The seventh and eighth fields are allele calls for the first variant in the `.map` file ('0' = no call); the 9th and 10th are allele calls for the second variant; and so on.

If all alleles are single-character, PLINK 1.9 will correctly parse the more compact 'compound genotype' variant of this format, where each genotype call is represented as a single two-character string. This does not require the use of an additional loading flag. You can produce such a file with `'--recode compound-genotypes'`.

It is also possible to [load .ped files missing some initial fields](#).

### **.\*.perm (adaptive permutation test report)**

Produced by several association analysis commands when the 'perm' modifier is used.

A text file with a header line, and then one line per variant with the following 4-7 fields:

CHR	Chromosome code
SNP	Variant identifier
BETA	Regression slope for real data. <i>Only present with <code>'--qfam emp-se'</code>.</i>
EMP_BETA	Sample mean of permutation regression slopes. <i>Only present with <code>'--qfam emp-se'</code>.</i>
EMP_SE	Sample stdev of permutation regression slopes. <i>Only present with <code>'--qfam emp-se'</code>.</i>
EMP1	Empirical p-value (pointwise), or lower-p-value permutation count

NP Number of permutations performed for this variant

### **.pphe (phenotype permutations)**

Produced by `--make-perm-pheno`. Valid input for `--pheno`.

A text file with no header line, and one line per sample with the following **P+2** fields (where **P** is the requested number of permutations):

1. Family ID
2. Within-family ID
- 3-(**P+2**). Permuted phenotypes

Missing phenotypes are always represented by the `--{output-}missing-phenotype` value (this is a very minor change from PLINK 1.07).

### **.prob (meta-analysis rejected variant list)**

Produced by `--meta-analysis`, when at least one variant is rejected.

A text file with no header line, and then one line per problem with the following 3 fields:

1. Filename
2. Variant ID
3. Problem code (one of {'BAD\_[CHR/BP/ES/SE/P/ESS]', 'MISSING\_[A1/A2]', 'ALLELE\_MISMATCH', 'DUPLICATE'})

Multiple problems may be reported for a single (filename, variant ID) pair.

### **.profile (allelic scoring results)**

Produced by `--score`.

A text file with a header line, and then one line per sample with the following 4-6 fields:

FID	Family ID
IID	Within-family ID
PHENO	Phenotype value
CNT	# of nonmissing alleles used for scoring. <i>May require 'include-cnt'.</i>
CNT2	Sum of named allele counts. <i>Not present with --dosage.</i>
'SCORE'/'SCORESUM'	Score (normally an allele-based average, unless 'sum' modifier used)

### **.qassoc (quantitative trait association test report)**

Produced by `--assoc` acting on a quantitative phenotype.

A text file with a header line, and then one line per variant with the following 9-11 fields:

CHR	Chromosome code
SNP	Variant identifier
BP	Base-pair coordinate
NMISS	Number of nonmissing genotype calls
BETA	Regression coefficient
SE	Standard error
R2	Regression r-squared
T	Wald test (based on t-distribution)
P	Wald test asymptotic p-value
LIN	Lin statistic. <i>Only present with 'lin' modifier.</i>
LIN_P	Lin test p-value. <i>Only present with 'lin'.</i>

### **.qassoc.gxe (quantitative trait interaction test report)**

Produced by `--gxe`.

A text file with a header line, and then one line per variant with the following 10 fields:

CHR	Chromosome code
SNP	Variant identifier
NMISS1	Nonmissing genotype calls in first group
BETA1	Regression coefficient for first group
SE1	Regression coefficient standard error for first group
NMISS2	Nonmissing genotype calls in second group
BETA2	Regression coefficient for second group
SE2	Regression coefficient standard error for second group

Z\_GXE Z score, test for interaction  
P\_GXE Asymptotic p-value

### **.qassoc.means (quantitative trait association genotype-stratified mean report)**

Produced by '[--assoc qt-means](#)'.

A text file with a header line, and then five lines per variant with the following six fields:

CHR Chromosome code  
SNP Variant identifier  
VALUE Type of value: one of {'GENO', 'COUNTS', 'FREQ', 'MEAN', 'SD'}  
G11 Value for homozygous A1 genotype  
G12 Value for heterozygous genotype  
G22 Value for homozygous A2 genotype

### **.qfam.\* (family-based quantitative trait association report)**

Produced by the [--qfam](#) family of commands.

The .qfam.[within/parents/between/total] file has a header line, and one line per variant with the following nine fields:

CHR Chromosome code  
SNP Variant identifier  
BP Base-pair coordinate  
A1 Allele 1 (usually minor)  
TEST Test type ('TOT', 'BET', or 'WITH')  
NIND Number of samples in linear regression  
BETA Regression coefficient  
STAT T-statistic (just for permutation test; don't use it directly)  
RAW\_P Uncorrected p-value

A [.qfam.\[within/parents/between/total\].perm file](#) is also generated.

### **.range.report (reprocessed gene-based report)**

Produced by [--gene-report](#).

The .range.report file has one subsection per nonempty gene. Each subsection contains a header line of the form "[gene name] -- [start/end coordinate pairs, comma-separated if necessary] ( [kb length] ) [border description, if necessary]"; this is followed by a blank line, the original report's header line with 'DIST' inserted in front, and the lines in the original report which concerned SNPs in the gene (preceded by [current pos] - [gene start coordinate] DIST values). Subsections are separated by two blank lines.

There are four small changes from PLINK 1.07:

- Genes now appear in natural-sorted instead of ASCII-sorted order (e.g. ABCA1 < ABCA3 < ABCA10, instead of the old ABCA1 < ABCA10 < ABCA3).
- kb lengths are larger by 0.001, since intervals in gene region files are fully closed instead of half-open.
- If [--gene-list-border](#) was specified, intervals and lengths in header lines do not include the additional padding.
- When a gene contains several disjoint regions on the same chromosome, they are now reported in a single subsection.

### **.raw (additive + dominant component file)**

Produced by '[--recode A](#)' and '[--recode AD](#)', for use with R. **This format cannot be loaded by PLINK.**

A text file with a header line, and then one line per sample with **V+6** (for '[--recode A](#)') or **2V+6** (for '[--recode AD](#)') fields, where **V** is the number of variants. The first six fields are:

FID Family ID  
IID Within-family ID  
PAT Paternal within-family ID  
MAT Maternal within-family ID  
SEX Sex (1 = male, 2 = female, 0 = unknown)  
PHENOTYPE Main phenotype value

This is followed by one or two fields per variant:

[Variant ID]_[counted allele]	Allelic dosage (0/1/2/'NA' for diploid variants, 0/2/'NA' for haploid)
[Variant ID]_HET	Dominant component (1 = het, 0 otherwise). <i>Requires '--recode AD'.</i>

If 'include-alt' was specified, the header line also names alternate allele codes in parentheses, e.g. 'rs5939319\_G(/A)'.

### **.recode.{geno,pheno,pos}.txt (BIMBAM genotype, phenotype, and variant position file)**

Produced by '--recode bimbam', for use by [BIMBAM](#). **This format cannot be loaded by PLINK.**

The .recode.geno.txt file produced by PLINK is a *comma-delimited* text file. It starts with two short header lines: **N** on its own line (where **N** is the number of samples), followed by number of variants on its own line. The third header line starts with 'IND', and is followed by the IIDs of all samples.

The main body of the file has one line per variant with **N+1** fields: the variant ID, followed by compound genotypes (with missing genotypes denoted by '??').

The .recode.pheno.txt file produced by PLINK is just a sequence of sample phenotype values, one per line.

The .recode.pos.txt file produced by PLINK is a text file with no header line, and one line per variant with the following 2-3 (space-delimited) fields:

1. Variant identifier
2. Base-pair coordinate
3. *Chromosome code (not present with 'bimbam-1chr')*

### **.recode.phase.inp (fastPHASE format)**

Produced by '--recode fastphase{-1chr}', for use by [fastPHASE](#). With '--recode fastphase', one file is generated per chromosome, and the full file extensions are of the form .chr-[chromosome number].recode.phase.inp. **This format cannot be loaded by PLINK.**

Each .phase.inp file produced by PLINK starts with two short header lines: number of samples on its own line, followed by **V** on its own line (where **V** is the number of variants). The third header line starts with 'P', and is followed by the base-pair coordinates of all variants.

The main body of the file has three lines per sample. The first line in each triplet is:

1. '#'
2. 'ID'
3. Within-family ID

The second and third lines each have a single **M**-character string, with one character per allele call. Missing calls are coded as '?'.

### **.recode.strct\_in (Structure format)**

Produced by '--recode structure', for use by [Structure](#). **This format cannot be loaded by PLINK.**

A text file with two header lines: the first header line lists all **V** variant IDs, while each entry in the second line is the *difference* between the current variant's base-pair coordinate and the previous variant's bp coordinate (or -1 when the current variant starts a new chromosome). This is followed by one line per sample with the following **2V+2** fields:

1. Within-family ID
2. Positive integer, unique for each FID
- 3-(**2V+2**). Genotype calls, with the A1 allele coded as '1', A2 = '2', and missing = '0'

### **.ref (long-format reference allele file)**

Reference allele file which accompanies a .lgen file when it's generated with '--recode lgen-ref'. Loaded with --lfile + --reference.

A text file with no header line, and one line per polymorphic variant with the following 2-3 fields:

1. Variant identifier
2. Major allele
3. Minor allele (*not present if there is no minor allele*)

### **.rel (text relationship matrix)**

Produced by `--make-rel`.

Contents are identical to that of a `.grm/.grm.bin` file. Possible shapes are essentially the same as for `.dist` files; the only difference is that `.dist` files have an omitted or zero diagonal while `.rel` files do not.

### **.rlist (rare genotype list file)**

Produced by `'--recode rlist'`. Accompanied by `.fam` and `.map` files. **This format cannot be loaded by PLINK.**

A text file with no header line, and 0-3 lines per variant. Each line starts with the following four fields:

1. Variant identifier
2. Genotype class ('HOM' = homozygous minor, 'HET' = heterozygous, 'NIL' = missing call)
3. Allele 1 ('0' for missing)
4. Allele 2

This is followed by two additional fields (FID, then IID) for each sample with the specified genotype call at the variant. If there are no such samples, the entire line is omitted from the file. (As a result, any variants with nothing but homozygous major genotypes are not mentioned at all.)

### **.sample (Oxford sample information file)**

Sample information file accompanying a `.gen` genotype dosage file. Loaded with `--data/--sample`, and produced by `'--recode oxford'`.

The `.sample` space-delimited files emitted by `--recode` have two header lines, and then one line per sample with 3-5 relevant fields:

<u>First header line</u>	<u>Second header line</u>	<u>Subsequent contents</u>
ID_1	0	Family ID
ID_2	0	Within-family ID
missing	0	Missing call frequency
sex	D	Sex code ('1' = male, '2' = female, '0' = unknown)
phenotype	'B'/'P'	Binary ('0' = control, '1' = case) or continuous phenotype

The full specification for this format is on the [Oxford statistical genetics website](#).

### **.set ('END'-terminated variant set membership list file)**

Produced by `--write-set`, and loaded with `--set`.

A text file with a sequence of variant set definitions. Each set definition starts with the set ID, followed by IDs of all variants in the set, followed by 'END'. Spaces, tabs, and newlines are acceptable and equivalent token delimiters; the files emitted by `--write-set` have a single token per line and a blank line between sets, but you can e.g. describe an entire set per line instead, and `--set` will still read the file correctly.

For example, the `.set` file

```
GENE1
rs123456
rs10912
rs66222
END

GENE2 rs66222 rs929292
rs288222 END
```

assigns variants `rs123456` and `rs10912` to set 'GENE1', `rs929292` and `rs288222` to 'GENE2', and `rs66222` to both sets.

When multiple set definitions share the same set ID, that currently results in an error rather than a merge.

### **.set.{m}perm (set association permutation test report)**

Produced by `--assoc/--model/--linear/--logistic/--tdt/--mh/--bd` when run with the `'set-test'` modifier.

A text file with a header line, and then one line per set with the following 6-7 fields:

SET	Set ID
NSNP	Set size
NSIG	Raw number of significant variants
ISIG	Final size of most-significant-variants subset (after --set-r2 and --set-max thresholds)
EMP1	Empirical set p-value, or lower-p-value permutation count
NP	Number of permutations performed. <i>Requires 'perm-count'.</i>
SNPS	' '-delimited IDs for most-significant-variants subset ('NA' if empty)

Calculation of NSIG is no longer cut short when the --set-max value is hit.

### **.set.table (variant set membership table)**

Produced by `--set-table`.

A tab-delimited text file with a header line, and then one line per variant with the following 3+S columns (where S is the number of sets):

SNP	Variant identifier
CHR	Chromosome code
BP	Base-pair coordinate
Set IDs...	1 = member, 0 = nonmember

Variants which aren't a member of any set still appear in the table.

PLINK 1.07 wrote double-tabs on most lines between the 3rd and 4th columns; this no longer occurs.

### **.sexcheck (X chromosome-based sex validity report)**

Produced by `--check-sex/--impute-sex`.

A text file with a header line, and then one line per sample with the following 6-7 fields:

FID	Family ID
IID	Within-family ID
PEDSEX	Sex code in input file
SNPSEX	Imputed sex code (1 = male, 2 = female, 0 = unknown)
STATUS	'OK' if PEDSEX and SNPSEX match and are nonzero, 'PROBLEM' otherwise
F	Inbreeding coefficient, considering only X chromosome. <i>Not present with 'y-only'.</i>
YCOUNT	Number of nonmissing genotype calls on Y chromosome. <i>Requires 'ycount'/'y-only'.</i>

### **.simfreq (simulation parameter file)**

Produced by `--simulate{-qt}`, and can be reread by them.

If generated by `--simulate` without the 'tags' or 'haps' modifier, it is a text file with no header line, and one line per SNP set with the following 6 fields:

1. Number of SNPs in set (always 1 in autogenerated file)
2. Label of this set of SNPs
3. Reference allele frequency lower bound
4. Reference allele frequency upper bound (equal to lower bound in autogenerated file)
5. odds(case | heterozygote) / odds(case | homozygous for alternate allele)
6. odds(case | homozygous for ref. allele) / odds(case | homozygous for alt. allele)

With 'tags' or 'haps', each line has the following 9 fields instead:

1. Number of SNPs in set (always 1 in autogenerated file)
2. Label of this set of SNPs
3. Reference allele frequency lower bound, causal variant
4. Reference allele frequency upper bound, causal variant
5. Reference allele frequency lower bound, marker
6. Reference allele frequency upper bound, marker
7. Marker-causal variant LD
8. odds(case | heterozygote) / odds(case | homozygous for alternate allele)
9. odds(case | homozygous for ref. allele) / odds(case | homozygous for alt. allele)

With `--simulate-qt`, in both subcases the last two fields are replaced with:

- Additive genetic variance for each SNP
- Dominance deviation

### **.tags.list (tagging variant report)**

Produced by `--show-tags`, when used in 'all' mode or with the `--list-all` flag.

A text file with a header line, and then one line per target variant with the following eight fields:

SNP	Variant identifier
CHR	Chromosome code
BP	Base-pair coordinate
NTAG	Number of other variants tagging this
LEFT	Base-pair coordinate of earliest tag variant, <i>including</i> this
RIGHT	Base-pair coordinate of latest tag variant, including this
KBSPAN	$(RIGHT - LEFT + 1) / 1000$
TAGS	' '-delimited list of IDs of other variants tagging this (or 'NONE')

### **.tdt (transmission disequilibrium test report)**

Produced by `--tdt` (unless parent-of-origin analysis was requested).

A text file with a header line, and then one line per autosomal/Xchr variant typically with the following 14-15 fields:

CHR	Chromosome code
SNP	Variant identifier
BP	Base-pair coordinate
A1	Allele 1 (usually minor)
A2	Allele 2 (usually major)
T	Transmitted A1 allele count
U	Untransmitted A1 allele count
OR	TDT odds ratio
CHISQ	TDT chi-square statistic. <i>Not present with 'exact'/'exact-midp'.</i>
p	Chi-square (default) or binomial test (if 'exact'/'exact-midp' specified) p-value
A:U_PAR	Parental affected A2 excess:unaffected A2 excess
CHISQ_PAR	Parental discordance chi-square statistic
P_PAR	Parental discordance chi-square p-value
CHISQ_COM	Combined test chi-square statistic
P_COM	Combined test chi-square p-value

The last five fields do not appear if no considered trio has parents with discordant phenotypes.

If `--ci 0.xy` has also been specified, the following two fields are inserted after 'OR':

Lxy	Bottom of xy% symmetric approx. confidence interval for TDT odds ratio
Uxy	Top of xy% approx. confidence interval for TDT odds ratio

### **.tdt.poo (parent-of-origin analysis)**

Produced by `'--tdt poo'`.

A text file with a header line, and then one line per autosomal/Xchr variant with the following 11 fields:

CHR	Chromosome code
SNP	Variant identifier
A1:A2	Allele 1 code:allele 2 code
T:U_PAT	Paternal A1:A2 transmission counts
CHISQ_PAT	Paternal chi-square statistic
P_PAT	Paternal chi-square p-value
T:U_MAT	Maternal A1:A2 transmission counts
CHISQ_MAT	Maternal chi-square statistic
P_MAT	Maternal chi-square p-value
Z_POO	Z score for paternal/maternal odds ratio difference
P_POO	Asymptotic parent-of-origin test p-value

### **.tfam (PLINK sample information file)**

Sample information file accompanying a `.tped` file; identical format to `.fam` files.

### **.tped (PLINK transposed text genotype table)**

Variant information + genotype call text file. Must be accompanied by a `.tfam` file. Loaded with `--tfile`, and produced by `'--recode transpose'`.

Contains no header line, and one line per variant with  $2N+4$  fields where  $N$  is the number of samples. The first four fields are the same as those in a `.map` file. The fifth and sixth fields are allele calls for the first sample in the `.tfam` file ('0' = no call); the 7th and 8th are allele calls for the second individual; and so on.

### **.traw (variant-major additive component file)**

Produced by `'--recode A-transpose'`, for use with R. **This format cannot be loaded by PLINK.**

A text file with a header line, and then one line per variant with the following  $N+6$  fields (where  $N$  is the number of samples):

CHR	Chromosome code
SNP	Variant identifier
(C)M	Position in morgans or centimorgans
POS	Base-pair coordinate
COUNTED	Counted allele (defaults to A1)
ALT	Other allele(s), comma-separated
[FID]_[IID]...	Allelic dosages (0/1/2/'NA' for diploid variants, 0/2/'NA' for haploid)

Since this format is new to PLINK 1.9, it is tab-delimited by default; use the `'spacex'` modifier to force spaces.

### **.twolocus (4x4 joint genotype count table, single variant pair)**

Produced by `--twolocus`.

A text file with 1-3 sections, depending on whether cases and/or controls are present. The first section starts with two header lines:

1. 'All individuals'
2. (underline)

This is followed by two tables. Each table has two header lines of its own:

1. Second variant ID
2. Five column headers: [A1 allele code]/[A1 allele code], [A1]/[A2], [A2]/[A2], '0/0', '\*/\*'

then rows corresponding to A1/A1, A1/A2, A2/A2, and missing first variant genotypes, then a fifth row with (sub)totals. The first table contains raw counts, while the second table contains proportions of the grand total.

This is followed by a 'Cases' section if there is at least one case, and finally a 'Controls' section if there is at least one control.

### **.var.ranges (equal-size variant ranges)**

Produced by `--write-var-ranges`.

A text file with a header line, and then one line per range with the following two fields:

FIRST	First variant ID
LAST	Last variant ID

### **.vcf (1000 Genomes Project text Variant Call Format)**

Variant information + sample ID + genotype call text file. Loaded with `--vcf`, and produced by `'--recode vcf'` (or `vcf-fid/vcf-iid`). **Do not use PLINK for general-purpose VCF handling: all information in VCF files which cannot be represented by the PLINK 1 binary format is ignored.**

The VCFv4.2 files emitted by `--recode` normally start with  $5+C$  header lines, where  $C$  is the number of chromosomes:

1. `##fileformat=VCFv4.2`
2. `##fileDate=[yyyymmdd date]`
3. `##source=PLINKv1.90`
- 4- $(C+3)$ . `##contig=<ID={chromosome code}>,length={last bp coordinate value + 1, or  $2^{31} - 3$  if unknown}>`
- $C+4$ . `##INFO=<ID=PR,Number=0,Type=Flag,Description="Provisional reference allele, may not be based on real reference genome">`

**C+5. ##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">**

(The INFO line is omitted when [--real-ref-alleles](#) is specified.)

This is followed by a tab-delimited header line with the following **N**+9 fields (where **N** is the number of samples), and one tab-delimited line per variant with the same fields:

#CHROM	Chromosome code/name
POS	Base-pair coordinate
ID	Variant identifier
REF	Allele 2 code (missing = 'N')
ALT	Allele 1 code (missing = '.')
QUAL	Left blank ('.')
FILTER	Left blank ('.')
INFO	Normally 'PR'; '.' when <a href="#">--real-ref-alleles</a> specified
FORMAT	'GT' (signaling the presence of genotype calls)
Sample IDs...	Genotype calls ('/'-separated if diploid, 0=ref, 1=alt, '.'=missing)

Allele codes are supposed to either start with '<', only contain characters in the set {A,C,G,T,N,a,c,g,t,n}, or represent a breakend. [--recode](#) issues a warning if an allele code does not satisfy this restriction.

The full VCFv4.2 specification is in the [hts-specs GitHub repository](#).

[Complete flag index >>](#)