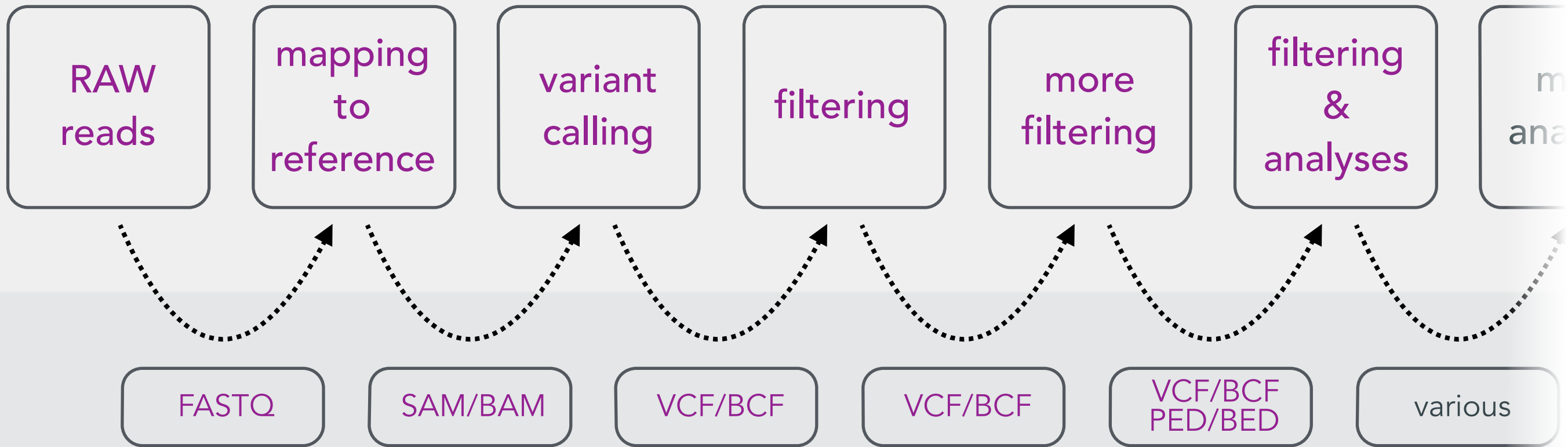


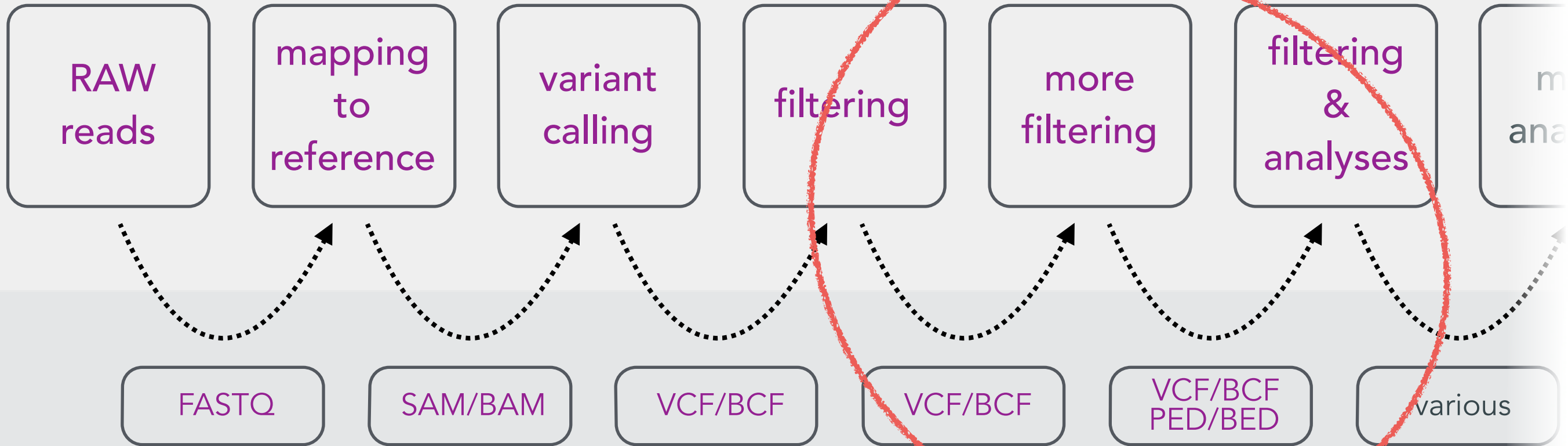
FileFormats.VCFtools.PLINK

# Pipeline

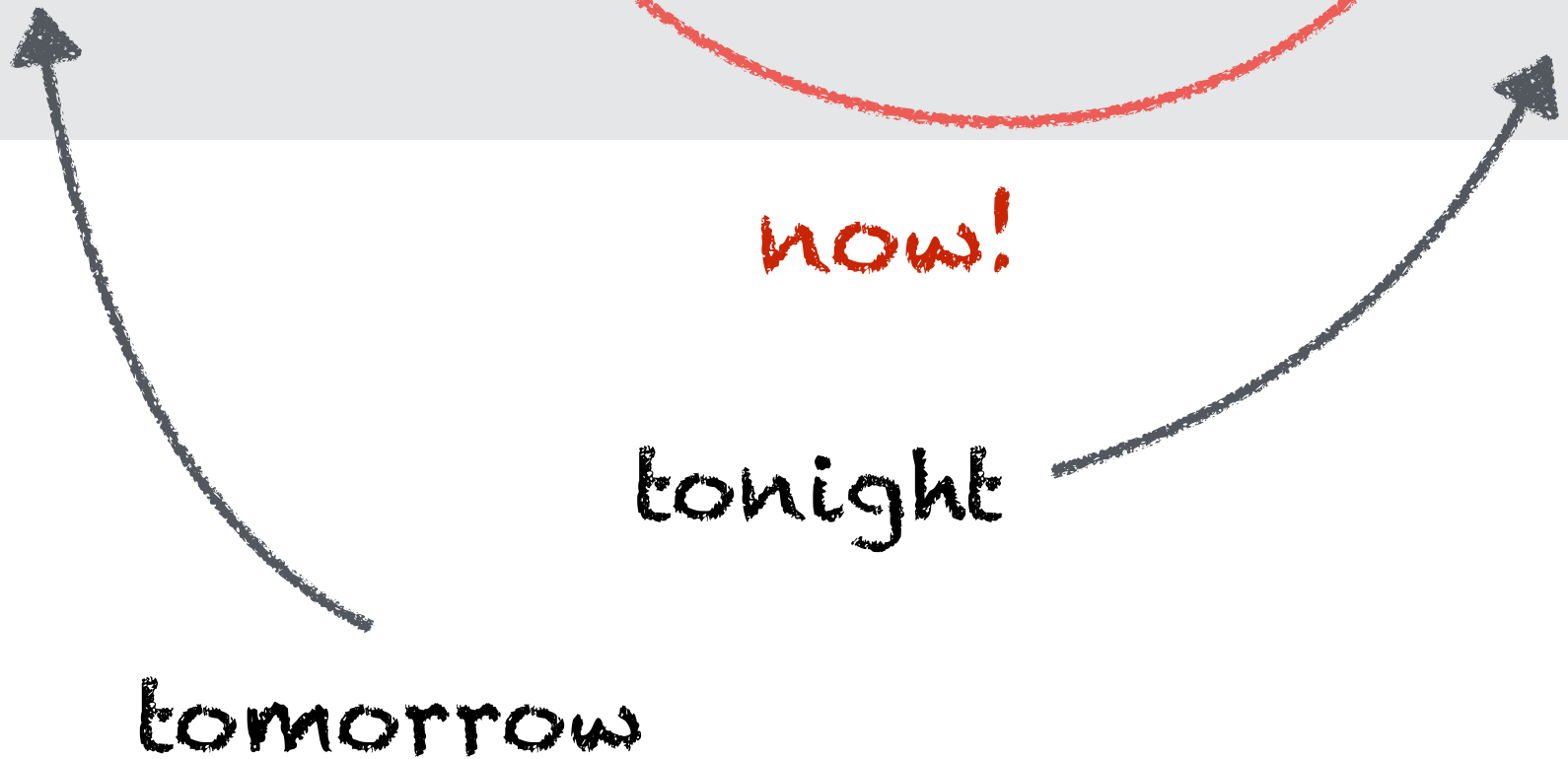


# File format

# Pipeline



# File format



## Name

bcftools — utilities for variant calling and manipulating VCFs and BCFs.

## Synopsis

**bcftools** [--version|--version-only] [--help] [COMMAND] [OPTIONS]

## DESCRIPTION

BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF. All commands work transparently with both VCFs and BCFs, both uncompressed and BGZF-compressed.

Most commands accept VCF, bgzipped VCF and BCF with filetype detected automatically even when streaming from a pipe. Indexed VCF and BCF will work in all situations. Un-indexed VCF and BCF and streams will work in most, but not all situations. In general, whenever multiple VCFs are read simultaneously, they must be indexed and therefore also compressed.

BCFtools is designed to work on a stream. It regards an input file "-" as the standard input (stdin) and outputs to the standard output (stdout). Several commands can thus be combined with Unix pipes.

## VERSION

This manual page was last updated **2015-12-15 14:02 GMT** and refers to bcftools git version **1.2-191-g6737c5c+**.

## BCF1

The BCF1 format output by versions of samtools <= 0.1.19 is **not** compatible with this version of bcftools. To read BCF1 files one can use the view command from old versions of bcftools packaged with samtools versions <= 0.1.19 to convert to VCF, which can then be read by this version of bcftools.

```
samtools-0.1.19/bcftools/bcftools view file.bcf1 | bcftools view
```

## VARIANT CALLING

See *bcftools call* for variant calling from the output of the *samtools mpileup* command. In versions of samtools <= 0.1.19 calling was done with *bcftools view*. Users are now required to choose between the old samtools calling model (*-c/--consensus-caller*) and



# VCFTools

A set of tools written in Perl and C++ for working with VCF files.

Home

Documentation

↓ Download ZIP

↓ Download TAR

🐱 View On GitHub

## Welcome to VCFTools

---

**VCFTools** is a program package designed for working with VCF files, such as those generated by the [1000 Genomes Project](#). The aim of VCFTools is to provide easily accessible methods for working with complex genetic variation data in the form of VCF files.

This toolset can be used to perform the following operations on VCF files:

- Filter out specific variants
- Compare files
- Summarize variants
- Convert to different file types
- Validate and merge files
- Create intersections and subsets of variants

VCFTools consists of two parts, a **perl module** and a **binary executable**. The perl module is a general Perl API for manipulating VCF files, whereas the binary executable provides general analysis routines.

## Download

---

To obtain VCFTools, please visit the [downloads](#) page.

## Variant call format specification

---

VCFTools is compatible with VCF versions 4.0, 4.1 and 4.2. For more information regarding the VCF format, please visit the [VCF specification page](#).

## Contact

---

Hosted on [GitHub Pages](#)

Copyright 2015 ©  
VCFTools



## Whole genome association analysis toolset

[Introduction](#) | [Basics](#) | [Download](#) | [Reference](#) | [Formats](#) | [Data management](#) | [Summary stats](#) | [Filters](#) | [Stratification](#) | [IBS/IBD](#) | [Association](#) | [Family-based](#) | [Permutation](#) | [LD calculations](#) | [Haplotypes](#) | [Conditional tests](#) | [Proxy association](#) | [Imputation](#) | [Dosage data](#) | [Meta-analysis](#) | [Result annotation](#) | [Clumping](#) | [Gene Report](#) | [Epistasis](#) | [Rare CNVs](#) | [Common CNPs](#) | [R-plugins](#) | [SNP annotation](#) | [Simulation](#) | [Profiles](#) | [ID helper](#) | [Resources](#) | [Flow chart](#) | [Misc.](#) | [FAQ](#) | [gPLINK](#)

### 1. Introduction

### 2. Basic information

- [Citing PLINK](#)
- [Reporting problems](#)
- [What's new?](#)
- [PDF documentation](#)

### 3. Download and general notes

- [Stable download](#)
- [Development code](#)
- [General notes](#)
- [MS-DOS notes](#)
- [Unix/Linux notes](#)
- [Compilation](#)
- [Using the command line](#)
- [Viewing output files](#)
- [Version history](#)

### 4. Command reference table

- [List of options](#)
- [List of output files](#)
- [Under development](#)

### 5. Basic usage/data formats

- [Running PLINK](#)
- [PED files](#)
- [MAP files](#)
- [Transposed filesets](#)
- [Long-format filesets](#)
- [Binary PED files](#)
- [Alternate phenotypes](#)
- [Covariate files](#)
- [Cluster files](#)
- [Set files](#)

### 6. Data management

- [Recode](#)
- [Reorder](#)
- [Write SNP list](#)
- [Update SNP map](#)

### New (15-May-2014): PLINK 1.9 is now available for beta-testing!

**PLINK** is a free, open-source whole genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner.

The focus of **PLINK** is purely on *analysis* of genotype/phenotype data, so there is no support for steps prior to this (e.g. study design and planning, generating genotype or CNV calls from raw data). Through integration with [gPLINK](#) and [Haploview](#), there is some support for the subsequent visualization, annotation and storage of results.

**PLINK** (one syllable) is being developed by Shaun Purcell at the Center for Human Genetic Research (CHGR), Massachusetts General Hospital (MGH), and the [Broad Institute](#) of Harvard & MIT, with the [support of others](#).

**New in 1.07:** [meta-analysis](#), [result annotation](#) and analysis of [dosage data](#).

### Data management

- Read data in a variety of formats
- Recode and reorder files
- Merge two or more files
- Extracts subsets (SNPs or individuals)
- Flip strand of SNPs
- Compress data in a binary file format

### Summary statistics for quality control

#### Quick links

- [PLINK tutorial](#)
- [gPLINK](#)
- [Join e-mail list](#)
- [Resources](#)
- [FAQs](#) | [PDF](#)
- [Citing PLINK](#)
- [Bugs, questions?](#)

[Software >>](#)[Statistical tests](#)[PLINK 1.9 home](#)[plink2-users](#)[File formats](#)[PLINK 1.9 index](#)**Introduction, downloads**

S: 24 Dec 2015 (b3.29)

D: 23 Jan 2016

[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](#)

# PLINK 1.90 beta

This is a comprehensive update to Shaun Purcell's [PLINK](#) command-line program, developed by [Christopher Chang](#) with support from the [NIH-NIDDK](#)'s Laboratory of Biological Modeling, the [Purcell Lab](#) at Mount Sinai School of Medicine, and others. ([What's new?](#)) ([Credits.](#)) ([Methods paper.](#))

## Binary downloads

Operating system <sup>1</sup>	Build		
	Stable (beta 3.29, 24 Dec)	Development (23 Jan)	Old <sup>2</sup> (v1.07)
Linux 64-bit	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>
Linux 32-bit	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>
OS X (64-bit)	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>
Windows 64-bit	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>
Windows 32-bit	<a href="#">download</a>	<a href="#">download</a>	<a href="#">download</a>

1: Solaris is no longer explicitly supported, but it should be able to run the Linux binaries.

2: These are just mirrors of the binaries posted at <http://pngu.mgh.harvard.edu/~purcell/plink/download.shtml>.

Source code, compilation instructions, and the like are on the [developer page](#).

**The following documented PLINK 1.07 flags are not supported by 1.90 beta 3:**

- [--qual-geno-scores](#)<sup>3</sup>
- [--segment](#)<sup>4</sup>
- [--dfam](#)
- [--p2](#), [--genedrop](#)
- [--hap](#), [--hap-window](#), [--hap-snp](#)<sup>5</sup>
- [--proxy-assoc](#), [--proxy-impute](#)<sup>5</sup>
- [--cnv-list](#), [--cfile](#), [--gfile](#)



<http://evomics.org/learning/population-and-speciation-genomics/fileformats-vcftools-plink/>