In the last 10 years there has been a tremendous amount of software released for the analysis of genetic data.

There are thousands of software packages available for genetic analysis.

Finding the software that suits your needs can be challenging.

Fortunately there are resources available that can be used to aide in this process.

This guide is designed to introduce you to some of the available resources for genetic analysis software as well as highlight the usefulness for a few of them.
"An Alphabetical List of Genetic Analysis Software" website is a very good resource for available genetic software http://linkage.rockefeller.edu/soft/

Software on the following topics:
- population genetics
- genetic linkage analysis for human pedigree data
- QTL analysis for animal/plant breeding data
- genetic marker ordering
- genetic association analysis
- haplotype construction
- pedigree drawing
"An Alphabetical List of Genetic Analysis Software" website was created in 1995 and is updated quite regularly
- More than 240 programs listed in December 2004
- More than 350 programs by August 2005
- Close to 400 programs by December 2006
- Close to 480 programs by November 2008
- 520 programs in August 2010
R is a programming language and software environment for statistical computing and graphics.

There are a number of R packages implementing statistical methods and algorithms for the analysis of genetic data and for related population genetics studies.

A brief description of these packages are given at the following website:
http://cran.r-project.org/web/views/Genetics.html
R packages for the following topics:

- Population Genetics
- Phylogenetics
- Linkage
- QTL mapping
- Association
- Linkage Disequilibrium and haplotype mapping
- Genome-Wide Association Studies (GWAS)
- Multiple testing
- Sequence Data
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: pngu.mgh.harvard.edu/~purcell/plink/

PLINK has numerous useful features for managing and analyzing genetic 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

- Allele, genotypes frequencies, HWE tests
- Missing genotype rates
- Inbreeding, IBS and IBD statistics for individuals and pairs of individuals
- non-Mendelian transmission in family data
- Sex checks based on X chromosome SNPs
- Tests of non-random genotyping failure
Basic association testing
  - Case/control
    - Standard allelic test
    - Fisher’s exact test
    - Cochran-Armitage trend test
    - Mantel-Haenszel and Breslow-Day tests for stratified samples
    - Dominant/recessive and general models
    - Model comparison tests (e.g. general versus multiplicative)
- Family-based association (TDT, sibship tests)
- Quantitative traits, association and interaction
- Association conditional on one or more SNPs
- Asymptotic and empirical p-values
- Flexible clustered permutation scheme
- Analysis of genotype probability data and fractional allele counts (post-imputation)
Plink: Statistical and Software Resources for Genetic Epidemiology Studies

- Multimarker predictors, haplotypic tests
  - Suite of flexible, conditional haplotype tests
  - Case/control and TDT association on the probabilistic haplotype phase
  - A set of proxy association” methods to study single SNP associations in their local haplotypic context
  - Imputation heuristic, to test untyped SNPs given a reference panel

- Copy number variant analysis
  - Joint SNP and CNV tests for common copy number variants
  - Filtering and summary procedures for segmental (rare) CNV data
  - Case/control comparison tests for global CNV properties
  - Permutation-based association procedure for identifying specific loci
• Gene-based tests of association
• Screen for epistasis
• Gene-environment interaction with continuous and dichotomous environments
• Meta-analysis
  • Automatically combine several generically-formatted summary files, for millions of SNPs
EIGENSOFT/EIGENSTRAT:
http://genepath.med.harvard.edu/~reich/Software.htm
- Identify and adjust for population structure by using principal components analysis (PCA).
- The top principal components are viewed as continuous axes of variation that reflect subpopulation genetic variation in the sample.
- Computationally feasible for high density genome scans.
STRUCTURE: http://pritch.bsd.uchicago.edu/software.html

- This method uses MCMC to cluster genetically similar individuals into "subpopulations."
- Not computationally feasible for GWAS
- Assignments of individuals are highly sensitive to the number of subpopulations in the sample, which is often unknown when there is cryptic structure

frappe:
http://med.stanford.edu/tanglab/software/frappe.html

- *frappe* uses a frequentist approach for estimating individual ancestry admixture proportions
Software for Structured Samples with Related Individuals

- Many genetic studies have samples with related individuals
- ROADTRIPS: RObust Association Detection Test for Related Individuals and Population Structure
  
  http://galton.uchicago.edu/~mcpeek/software/index.html

- Software for partially or completely unknown population and pedigree structure.
- Uses an empirical covariance matrix to capture both population and pedigree structure.

- EMMAX: Efficient Mixed-Model Association eXpedited
  
  http://genetics.cs.ucla.edu/emmax/

  - Uses a Variance Components model to account for structure in GWAS
Quanto is a program that computes sample size and/or power for association studies of genes, environmental factors, gene-environment interaction, or gene-gene interaction: http://hydra.usc.edu/GxE

Available study designs for a disease (binary) outcome include the unmatched case-control, matched case-control, case-sibling, case-parent, and case-only designs.

Study designs for a quantitative trait include independent individuals and case parent designs.
Haploview is designed to simplify and expedite the process of haplotype analysis by providing a common interface to several tasks relating to such analyses.

http://www.broad.mit.edu/personal/jcbarret/haploview/

- Haploview currently allows users to examine block structures, generate haplotypes in these blocks, run association tests, and save the data in a number of formats.

**Online Tools**

- LocusZoom: a tool to plot regional association results from genome-wide association scans or candidate gene studies:
  http://csg.sph.umich.edu/locuszoom/
- SNAP: a tool to plot regional association plots or regional LD plots:
  http://www.broadinstitute.org/mpg/snap/

- The R software package!