



SAS/Genetics™ Software

Performing genetics research with dedicated analysis tools

As a statistical geneticist or biostatistician, you may struggle with adequate tools to help you perform your research. Available analysis software may be useful, but you can waste time customizing it to perform specific tasks. The results may be disappointing and it may not provide adequate support when you need help. An ideal solution would handle your specialized research needs – efficiently and easily – and come with seasoned experts who could respond to your questions on a 24/7 basis.

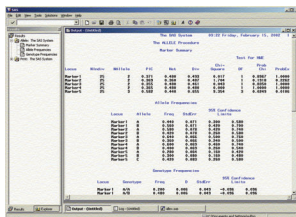
SAS/Genetics™ software is designed specifically for genetics research. It enables you to study genetic markers individually, examine the relationship between multiple markers and find associations between markers and traits using case-control or family data. And all your results can be depicted graphically – so they're easily understandable and shareable with others – with output from numerous procedures interpreted simultaneously.

Most importantly, SAS/Genetics comes from SAS. The world's leading pharmaceutical and biotechnology organizations have chosen SAS as their strategic partner – trusting us to help them use technology to know more about their data and their processes and to better anticipate results. We've been in the business of knowledge discovery for over a quarter-century and we'll be here over the life of your investment.

A powerful new product among SAS' scientific discovery solutions

SAS/Genetics™ contains five statistical procedures (or PROCs) and one macro. Based on SAS statistical methods, these PROCs are accessed through base SAS software.

PROC ALLELE



Marker	Allele	Frequency	Information	Chi-Square	P-Value
D3S1358	11	0.25	0.0000	0.0000	0.9999
	12	0.75	0.0000	0.0000	0.9999
D5S818	10	0.25	0.0000	0.0000	0.9999
	11	0.75	0.0000	0.0000	0.9999

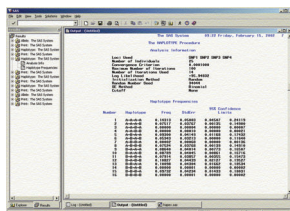
This procedure provides summary statistics to characterize individual markers. PROC ALLELE calculates measures of marker informativeness, which can help determine which markers are most valuable for association or linkage studies. You can produce tables of allele and genotype frequencies, including bootstrap confidence intervals for hypothesis testing. An asymptotic and exact test can be used for testing markers for Hardy-Weinberg equilibrium (HWE).

PROC ALLELE also explores associations between two markers or pairs of alleles at two markers. It can calculate:

- Linkage disequilibrium between pairs of alleles using haplotypes entered by the user.
- Maximum likelihood estimates of the two-locus haplotype frequencies.
- Composite linkage disequilibrium coefficients that do not require haplotype frequencies.

Additionally, you can perform an asymptotic and exact version of an overall test for no linkage disequilibrium between two markers with any number of alleles.

PROC HAPLOTYPE

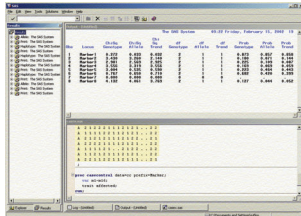


Marker	Haplotype	Frequency	Information	Chi-Square	P-Value
D3S1358	11-12	0.25	0.0000	0.0000	0.9999
	12-11	0.75	0.0000	0.0000	0.9999
D5S818	10-11	0.25	0.0000	0.0000	0.9999
	11-10	0.75	0.0000	0.0000	0.9999

Haplotypes are sets of specific alleles that occur together on the same chromosome, as opposed to genotypes which occur in pairs arising from different chromosomes. This procedure estimates haplotype frequencies for more than two loci using genotype data to reconstruct possible haplotypes. Assuming HWE, the expectation-maximization (EM) algorithm is used to calculate the maximum likelihood estimates of the haplotype frequencies. Various standard error estimation methods are available, as well as predicted probabilities of individual haplotypes. These estimated haplotype frequencies can then be used in tests for linkage disequilibrium across markers or for association with a binary variable such as disease status (case-control data).

resee

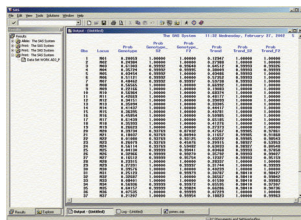
PROC CASECONTROL and PROC FAMILY



These procedures test markers for association with a dichotomous trait such as disease status. PROC CASECONTROL uses random samples of unrelated individuals, while PROC FAMILY analyzes family data consisting of either parents and an affected child, or sibling pairs made up of one affected and one unaffected sibling.

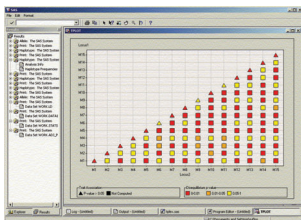
- Several case-control tests can be used: the genotype case-control test, the allele case-control test, and the linear trend test with or without a variance inflation correction.
- Family-based tests include the original transmission/disequilibrium test (TDT), the sibling TDT (S-TDT), the sibling disequilibrium test (SDT) and the reconstruction-combined TDT (RD-TDT). All tests can handle markers with any number of alleles.

PROC PSMOOTH



Marker-association tests performed by PROC CASECONTROL and PROC FAMILY examine one marker at a time, but linkage disequilibrium between markers that are close together causes correlations between these tests. PROC PSMOOTH accounts for this correlation by smoothing the p-values, where smoothing involves using a function of p-values from consecutive markers to form a new p-value. PROC PSMOOTH can be applied directly to output data sets from PROC CASECONTROL and/or PROC FAMILY, or to user-created data sets containing p-values from linkage or association analyzes. It can also adjust p-values for multiple testing.

%TPLOT macro



Included in the SAS autocall library, this macro creates a triangular plot for graphically displaying marker test results in an easy, interpretable manner.

- Output from PROC ALLELE can be combined into this single plot with output from PROC CASECONTROL, PROC FAMILY, PROC PSMOOTH, or a user-created data set containing p-values from marker tests.
- You can also represent p-values from the HWE test, tests for linkage disequilibrium tests between pairs of markers, and linkage or association tests between each marker and the trait of interest, all smoothed and adjusted using PROC PSMOOTH.

arch

From SAS, an **industry standard**

SAS is committed to helping customers extract from their data the knowledge they need to make intelligent decisions. Toward that end, we incorporate award-winning training and technical support as part of every solution. SAS-certified training professionals provide expert knowledge tailored to meet your needs, using your own data in your work environment. Additionally, advanced training is available at your site or through our regional training centers.

SAS is the world's largest privately held software company, serving more than 37,000 customer sites in 110 countries. More than 3.5 million users have seen the power of SAS products turn raw data into rare insights.

For more information about SAS/Genetics from SAS visit us on the Web at:

www.sas.com/industry/pharma/index.html



World Headquarters
and SAS Americas
SAS Campus Drive
Cary, NC 27513 USA
Tel: (919) 677 8000
Fax: (919) 677 4444
U.S. & Canada sales:
(800) 727 0025

SAS International
PO Box 10 53 40
Neuenheimer Landstr. 28-30
D-69043 Heidelberg, Germany
Tel: (49) 6221 4160
Fax: (49) 6221 474850

www.sas.com