SNP 분석 software

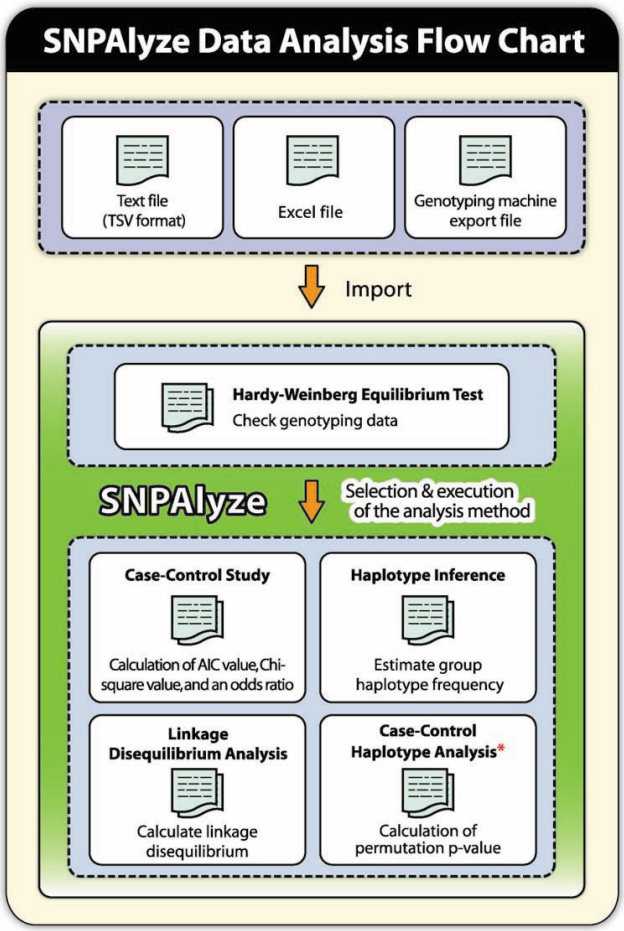
SNPAlyze® Ver.4.1

유전자 다형성 연구의 데이터 마이닝 툴!!

TaKaRa Code DN001~003

* 본 제품은 Dynacom사의 제품입니다.

SNPAlyze는 방대한 유전자 다형성 데이터에서 보다 유용한 정보를 추출하는 데이터 마이닝용 소프트웨어이다.
SNP 및 Microsatellite의 Genotyping data에 통계 처리를 추가하고, 유전자성 질환 마커가 되는 유전자 다형성 검출 및 질환 관련 유전자를 분류하기 위하여 개발되었다.
Case-Control Study, Linkage Disequilibrium Analysis, Haplotype Inference, Case-Control Haplotype Analysis(Permutation test), Hardy-Weinberg Equilibrium Test의 5가지 방법을 이용할 수 있다.



특징

- Excel과 동등한 간편한 조작성
- 데이터의 Import가 간단
- 세계 최고속 Haplotype 추정*
- 많은 유전자좌의 Haplotype 추정**
- Chi-square value, p값 외 AIC로 대응
- Linkage Disequilibrium의 graphical display
- 일반적인 연쇄불평형계수(D, D′, r2)로 AIC(LD) 대응
- Diplotype distribution의 추정
- Permutation test의 이용
- Microsatellite에 대응

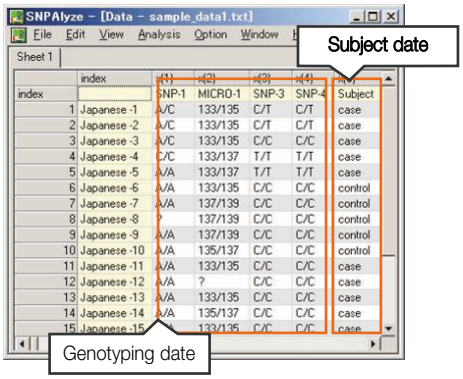
*2004년 10월 자사 조사 결과
**Pro version에만 존재하는 기능

주요 기능

(1)Genotyping data Import

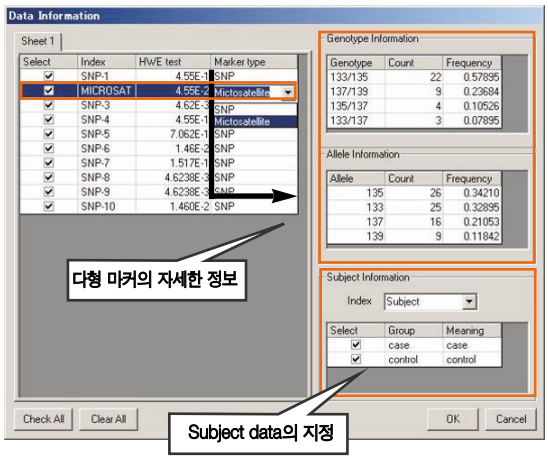
SNPAlyze는 아래와 같은 형식의 파일을 해석 할 수 있다.
- Microsoft excel file
- TSV(Tab Separated Text) file
- Biotage사 PSQ export file
- ABI사 PRISM7900 export file

Genotyping data는 2 allele의 조합으로 입력 가능하다(A/C, 133/135등). 각 시료가 속하는 집단을 표현하는 Subject data를 입력하여 그룹별 해석이 가능하게 된다.



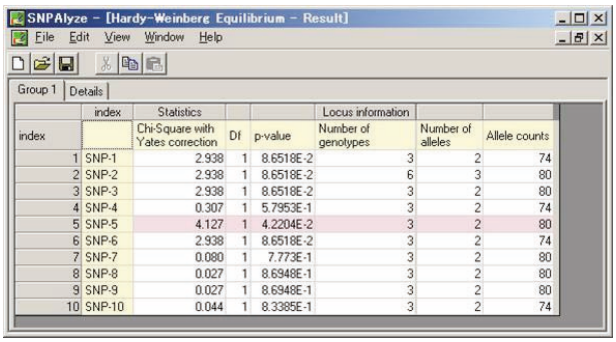
[다형성 정보의 참조]

Genotyping data를 집계하고, 각 Genotype 및 Allele의 수와 그 빈도를 표시한다.
Hardy-Weinberg Equilibrium Test의 결과나 다형성의 종류(SNP or Microsatellite)에 대한 정보도 표시 된다.
이러한 정보를 참고하여 해석 대상으로 하는 다형성 마커를 선택한다.
또, 병의 유무에 따라 시료를 집단 별로 해석하는 경우는 각 시료가 속한 집단의 정보가 입력되어 있는 Subject data를 지정한다.



(2)Hardy-Weinberg Equilibrium Test

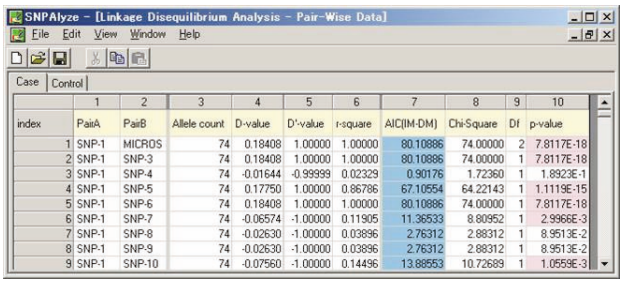
Hardy-Weinberg Equilibrium(HWE)는 도태나 돌연변이 등의 요인이 없고, 개체수가 충분히 큰 집단에서 성립된다.
SNPAlyze에서는 Chi-square goodness of fit test로 HWE가 성립하는지 검정 한다. HWE의 성립은 SNPAlyze로 이용 가능한 통계 수법의 전제 조건이다. HWE가 성립되지 않는 다형성 마커를 해석에 이용하는 것은 바람직하지 못하다.



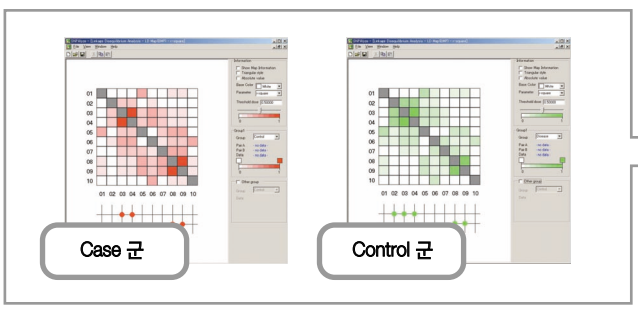
index		Chi-Square with Yates correction	Df	p-value	Number of genotypes	Number of alleles	Allele counts
1	SNP-1	2.938	1	8.6518E-2	3	2	74
2	SNP-2	2.938	1	8.6518E-2	6	3	80
3	SNP-3	2.938	1	8.6518E-2	3	2	80
4	SNP-4	0.307	1	5.7953E-1	3	2	74
5	SNP-5	4.127	1	4.2204E-2	3	2	80
6	SNP-6	2.938	1	8.6518E-2	3	2	74
7	SNP-7	0.080	1	7.773E-1	3	2	80
8	SNP-8	0.027	1	8.6948E-1	3	2	80
9	SNP-9	0.027	1	8.6948E-1	3	2	80
10	SNP-10	0.044	1	8.3385E-1	3	2	74

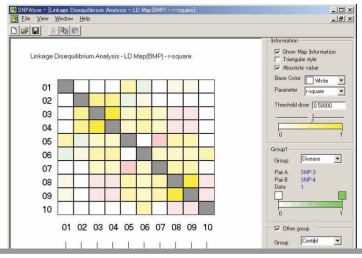
(3)Linkage Disequilibrium Analysis

어떤 집단에 있어서 다형성이 존재하는 두 유전자좌의 Allele 분배가 독립적이지 않은 경우를 연쇄불평형이라고 하며, 어떤 질환의 원인변이와 SNP/Microsatellite와의 사이에 존재하는 연쇄불평형을 이용하여 원인변이의 영역을 추정하는 것이 연쇄불평형해석이다. SNPAlyze에서는 연쇄불평형의 정도를 나타내는데 일반적으로 이용하고 있는 D, D′, r2로 AIC(LD)를 출력하고, 연쇄불평형에 대한 graphical view를 제공한다. 다른 집단의 결과를 반복하여 특정 집단에 특이적인 연쇄불평형을 검출할 수 있다.



index	PairA	PairB	Allele count	D-value	D'-value	r-square	AIC(IM-DM)	Chi-Square	Df	p-value
1	SNP-1	MICROS	74	0.18408	1.00000	1.00000	80.10886	74.00000	2	7.8117E-18
2	SNP-1	SNP-3	74	0.18408	1.00000	1.00000	80.10886	74.00000	1	7.8117E-18
3	SNP-1	SNP-4	74	-0.01644	-0.99999	0.02329	0.90176	1.72360	1	1.8923E-1
4	SNP-1	SNP-5	74	0.17750	1.00000	0.86786	67.10554	64.22143	1	1.1119E-15
5	SNP-1	SNP-6	74	0.18408	1.00000	1.00000	80.10886	74.00000	1	7.8117E-18
6	SNP-1	SNP-7	74	-0.06574	-1.00000	0.11905	11.36533	8.80952	1	2.9966E-3
7	SNP-1	SNP-8	74	-0.02630	-1.00000	0.03896	2.76312	2.88312	1	8.9513E-2
8	SNP-1	SNP-9	74	-0.02630	-1.00000	0.03896	2.76312	2.88312	1	8.9513E-2
9	SNP-1	SNP-10	74	-0.07560	-1.00000	0.14496	13.88553	10.72689	1	1.0559E-3





Case군, Control군의 양쪽에서 연쇄불평형이 강한 경우는 황색(녹+적)으로 표시되어진다.
어느 쪽이든 한 쪽에서 연쇄불평형이 강한 경우 녹색 또는 적색으로 표시되어진다.

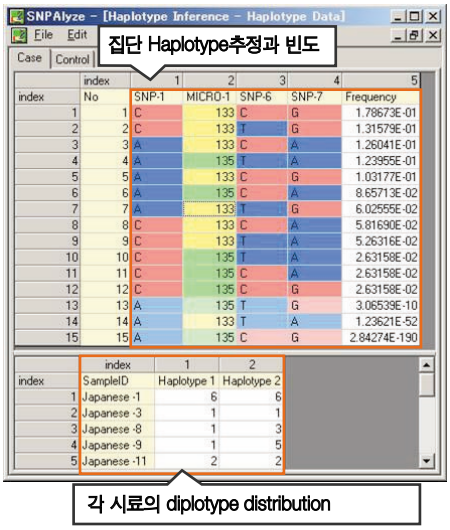
(4)Haplotype Inference

Haplotype은 동일 염색체상에 존재하는 복수의 유전자좌의 Allele이 정렬되는 방법이다. 질환이나 개인의 체질과 깊게 연관되어 있다고 알려져 있어 Haplotype를 조사하는 것은 매우 중요하다.

-〉 Case-Control Haplotype Analysis

SNPAlyze에서는 Expectation-Maximization(EM) Algorithm을 이용하여 집단 Haplotype 후보로 그 빈도를 추정한다.

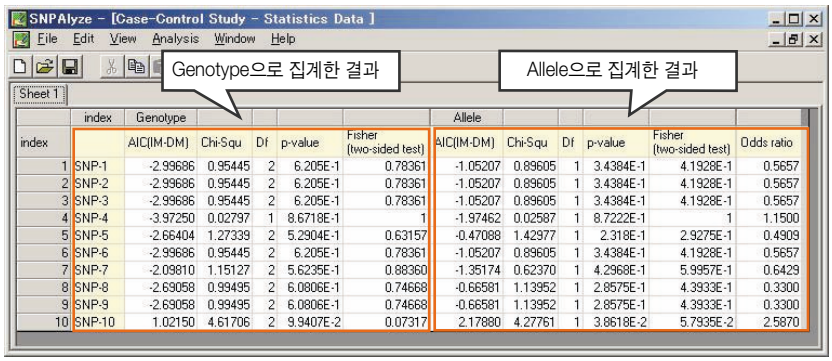
각 시료가 가질 수 있는 가능성이 가장 높은 Haplotype을 조합하여 diplotype distribution를 추정할 수 있다(Ver. 4.1 Pro에서는 최대 40 Locus의 Haplotype 추정 가능).



(5)Case-Control Study

질환의 유무에 따라 Case군과 Control군을 지정하고 유전자 데이터등과의 관련성을 해석하는 것이 Case-Control Study이다.

각각의 유전자좌와 질환과의 관련해석을 하기 때문에 단독으로 질환과 관련 있는 SNP/Microsatellite를 검출할 수 있다.

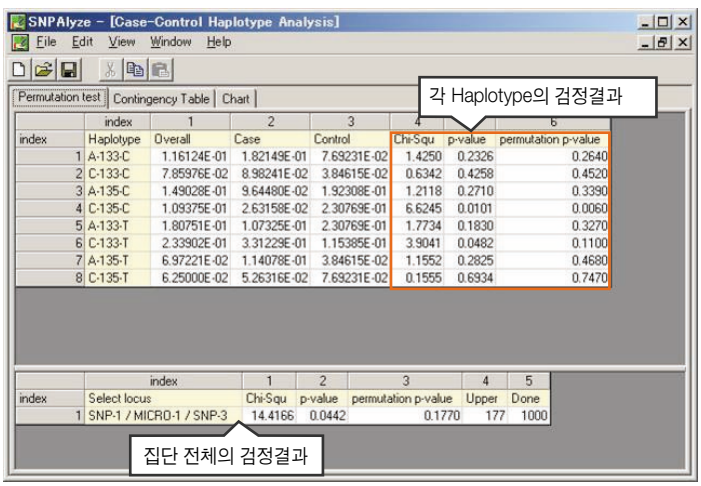


각 집단의 SNP/Microsatellite 데이터를 Genotype 및 Allele으로 집계하고 검정한다. 질환과 관련이 있는 SNP/Microsatellite를 검출한다.

(6)Case-Control Haplotype Analysis

질환관련변이와 연쇄불평형인 유전자 다형성으로 구성되어진 Haplotype은 질환관련변이를 가진 Case군에 많게 나올 것으로 예측한다. 이 상태를 통계적으로 해석하는 방법이 Case-Control Haplotype Analysis이다.

SNPAlyze에서는 집단 각각에 대해서, EM algorithm보다 Haplotype 빈도를 추정하고 Permutation test에 따라 검정한다. 집단전체의 검정결과로 각 Haplotype의 검정결과도 출력 가능하다.



타사 소프트웨어와의 비교

		SNPAlyze(Ver.4)	Arlequin(V.2001)	Cocaphase(V2.40)	SNPHAP(V1.0)
User interface		GUI	GUI	CUI	CUI
Data import	Text file	○	○	○	○
	Excel file	○	×	×	×
	Others	PSQ96 PRISM7900	Genepop, phylip, Biosys, Mega…	Genehunter	—
Data Output/save	Text file	○	○	○	○
	Excel file	○	×	×	×
	Others	BMP	html	—	—
OS	Windows	○	○	○	—
	Others	—	Mac, UNIX	UNIX	UNIX

SNPAlyze® Ver.4.1 종류별 비교

	Lite	Standard	Pro
최대 시료수	250	1,000	10,000
Hardy-Weinberg-Equilibrium Test	○	○	○
Case-Control Study	○	○	○
Linkage Disequilibrium Analysis	○	○	○
Haplotype Inference	○ (최대15 loci)	○ (최대 30 loci)	○ (최대 40 loci)
Case-Control Haplotype Analysis		○	○
Bootstrap		○	○
TaKaRa Code	DN001	DN002	DN003

필요 시스템

	Lite / Standard	Pro
System	Windows 98/Me NT4.0/2000/XP*	Windows 2000/XP
RAM	128MB이상	512MB이상
Disk space	32MB이상	1GB이상
Others	CD-ROM drive Interface** Windows 98/Me/2000/XP : USB port Windows NT4.0 : Parallel port (DB25)	

*대량의 데이터를 해석하는 경우는 Windows NT4.0/2000/XP을 사용하시기 바랍니다.
**본 제품은 protect key로 사용자 관리를 합니다. 따라서 소프트웨어 실행 시에 protect key를 접속하는 USB port 1 개 또는 Parallel port가 필요합니다.

참고문헌

1) Shimo-onoda K, Tanaka T, Furushima K, Nakajima T, Toh S, Harata S, Yone K, Komiya S, Adachi H, Nakamura E, Fujimiya H, Inoue I. Akaike's information criterion for a measure of linkage disequilibrium. *J Hum Genet*. 2002; **47**(12):649-55.

2) Ohmori H, Makita Y, Funamizu M, Chiba S, Ohtani K, Suzuki Y, Wakamiya N, Hata A. Haplotype analysis of the human collectin placenta 1 (hCL-P1) gene. *J Hum Genet*. 2003; **48**(2):82-5.

3) Tanaka T, Ikari K, Furushima K, Okada A, Tanaka H, Furukawa K, Yoshida K, Ikeda T, Ikegawa S, Hunt SC, Takeda J, Toh S, Harata S, Nakajima T, Inoue I. Genomewide linkage and linkage disequilibrium analyses identify COL6A1, on chromosome 21, as the locus for ossification of the posterior longitudinal ligament of the spine. *Am J Hum Genet*. 2003 Oct; **73**(4):812-22.