약 력

1. 인적사항

성 명	양 재 일
소속기관	(주)마크로젠
직 위	과장
전자메일	jiyang@macrogen.com

2. 학력/경력

연 도	학교 / 기관	전공 / 직위	학위 / 비고
1991. 3 - 1998. 2	전북대학교	통계학	학사
2000. 3 - 2002. 8	전북대학교	전산통계학	석사
1997.10 - 1999.11	농림부 농업통계사무소	직원	
2001.12 - 2002. 5	(주)케이테크 멀티미디어 DB연구소	연구원	
2002. 8 - 2005. 5	(주)선도소프트	과장	통계팀
2005.05 - 현재	(주)마크로젠	과장	분석팀장

3. 주요연구실적(개조식, 간단하게)

• 2005.05 - 2005.12 : 한국인 유전체형 분석사업 2005-1(질환후보유전자의 한국인 SNP 발굴)

4. 발표시 사용 기자재

* LCD projector의 사용을 원칙으로 합니다.
* LCD 사용을 위해 CD나, 저장 매체에 담아 오시는 것을 권장하며, Zip드라이브는 학회에서 준비하지 않습니다.

Leading the Biochip Revolution!

Macrogen, Inc
Biochip Division

1

Contents

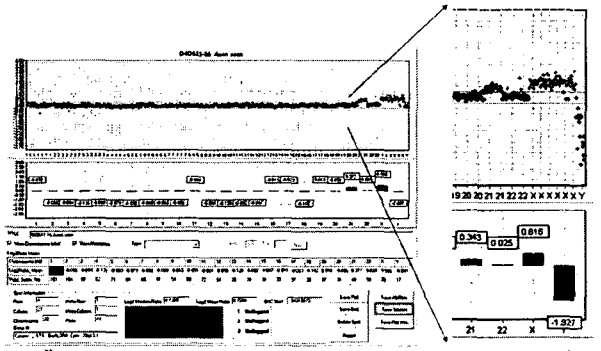
- Introduction to chip analysis services
- What is SNP?
- Macrogen's SNP discovery
- SNP association study
  - What is R?
  - SNP Analysis Method
- Conclusion

MACROGEN Leading the Biochip Revolution! 2

Chip Analysis Services

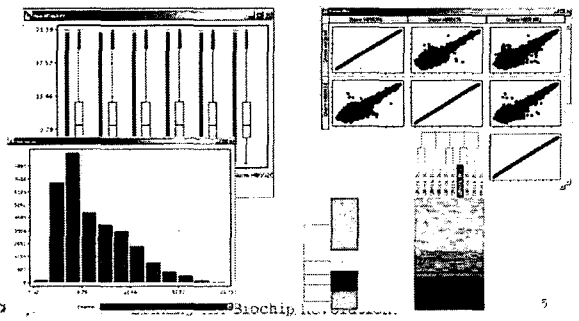
MACROGEN Leading the Biochip Revolution! 3

Do you know Macrogen Karyo chip?



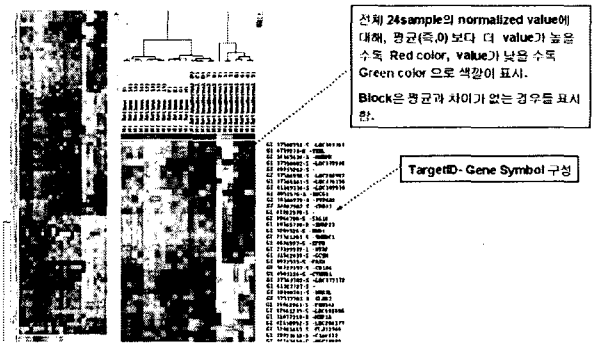
MACROGEN Leading the Biochip Revolution! 4

Analyze ABI1700 export in Avadis



MACROGEN Biochip ... 5

Illumina Expression Chip



MACROGEN Leading the Biochip Revolution! 6

What is SNP?

Single Nucleotide Polymorphism

- Single Nucleotide Polymorphism (SNP) arises from mutation.
- Mutation nucleotides become SNPs when observed frequency > 1% in a population.
  - SNP: DNA single base variations found >1%
  - Mutation: DNA single base variations found <1%

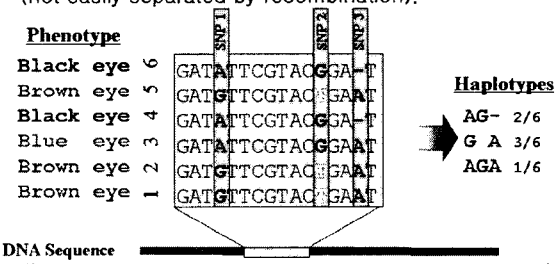
General Population	ACTTAGCTT ←	94%
SNP	ACTTAGCTC ←	6%
General Population	ACTTAGCTT ←	99.9%
Mutation	ACTTAGCTC ←	0.1%

Single Nucleotide Polymorphism

- All humans share 99.9% the same DNA sequence
  - SNPs occur about every 600 base pairs.
  - 90% of human genome variation comes SNPs.
  - The human genome contains about 3 million SNPs.
- Because of the A-T/C-G complement, a SNP can have only two variants: (AT) or (CG).
- A SNP is a variable with two states:
  - Major allele: Allele (i.e., (AT) or (CG)) > 50%.
  - Minor allele: Allele < 50%.

Haplotype

- A set of closely linked SNPs located on one chromosome, which tend to be inherited together (not easily separated by recombination).

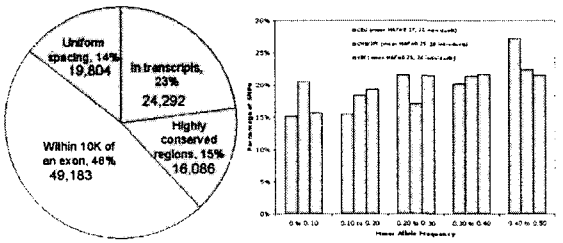


Illumina's Genotyping Assays

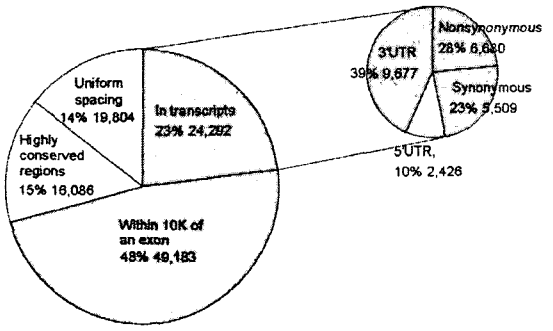
- GoldenGate™ Assay
  - For custom SNP projects
  - For standard panel projects up to 12,000 loci
    - Linkage projects
    - MHC
    - Mouse Linkage coming soon
  - Multiplexed from 384 to 1536 (and multiples thereof)
    - 96plex in development
  - Industry standard for completeness & quality
- Infinium™ Whole Genome Genotyping Assay
  - For standard SNP projects
  - Multiplexed from 10,000 to hundreds of thousands

Human-1 Content Strategy

- >109K SNPs total
- >73K SNPs in transcripts or within 10kb of an exon
- ~26 kb intermarker spacing



Breakdown of Transcript SNPs



Macrogen's SNP Discovery

한국인 유전체형 분석사업 2005-1

- 질환 후보유전자의 한국인 SNP 발굴
  - 단일염기서열(SNP)발굴
  - Haplotype 정보 구축
  - 한국인의 유전정보 국가자원화

  - Korean SNP DB 확보 및 구축
  - 맞춤의학 기본토대 마련
  - 신약개발 추진

분석대상 유전자 및 유전자정보

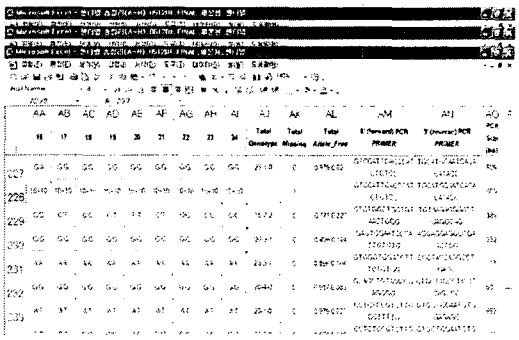
- 7개 질환군별 유전체센터 당 30개 유전자
  - 폐 및 호흡기, 면역질환, 골격계, 당뇨내분비
  - 선천성기형, 조혈계, 뇌신경계
- 국립보건연구원 90개 유전자

- UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway)
- NCBI GenBank (http://www.ncbi.nlm.nih.gov)
- 프로모터 부분 1.5 이상 분석
- 모든 엑손 과 좌우 인접 인트론 지역의 최소 50bp 이상 분석
- *3' Franking region 500bp*

SNP Discovery Result

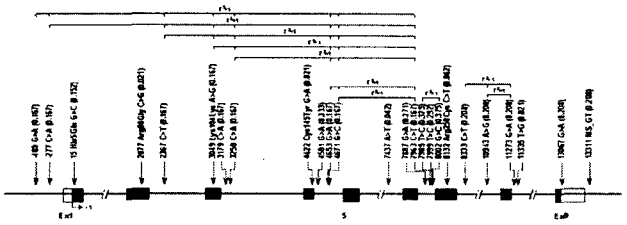
	Promoter	CDS Syn	CDS Non-Syn	3'UTR	5'UTR	Intron	3'Flank	Total	Novel	dbSNP
호흡기질환	82	38	42	45	13	110	20	350	152	198
면역질환	114	51	48	54	12	145	16	440	178	262
골격계질환	110	47	25	51	11	155	12	421	172	249
당뇨내분비질환	106	34	29	92	11	128	5	405	167	238
선천성질환	119	24	32	51	10	71	4	311	124	187
조혈계질환	103	27	30	43	10	130	9	362	166	196
뇌신경계질환	102	48	29	67	13	148	18	425	191	234
국립보건연구원	394	177	120	253	44	530	21	1,539	584	954
총계	1,130	448	355	666	124	1,427	105	4,243	1,734	2,508

SNP Discovery Result



Gene Map

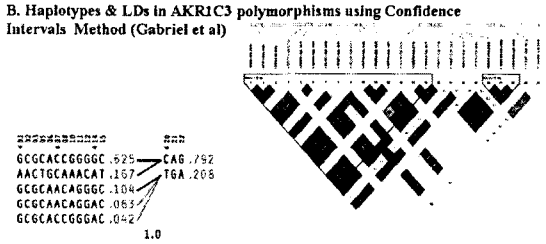
A. Map of AKR1C3



Haplotype Analysis

- Haplotyping (Haploview)
- – Marker Block using Confidence Intervals (Gabriel et al, Science, 2002)
- strong LD: D'≥ 95% confidence

B. Haplotypes & LDs in AKR1C3 polymorphisms using Confidence Intervals Method (Gabriel et al)

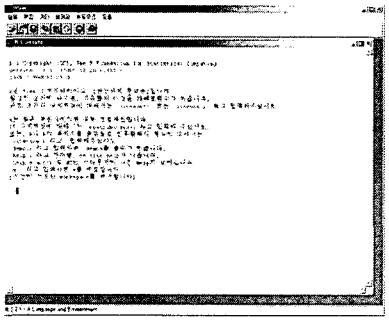


SNP Association Study

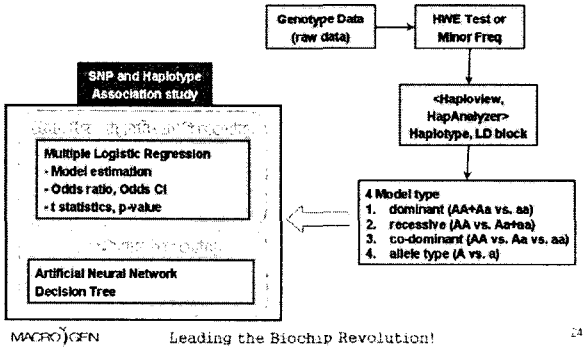
Advantage using R

- Steady increase in popularity
- Open source for customization
- Used by the world's top biostatisticians
- Reproducibility of the analysis
- High quality and flexible graphics output
- Many add-on packages
- Cross-platform
  - Windows, linux, MaxOS X

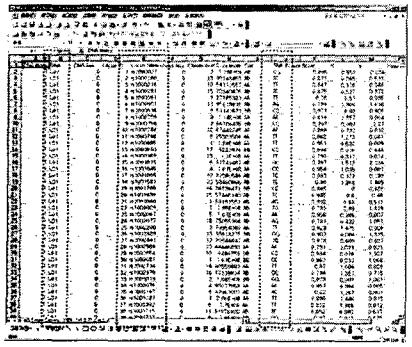
Start R Console Windows



SNP Analysis Method

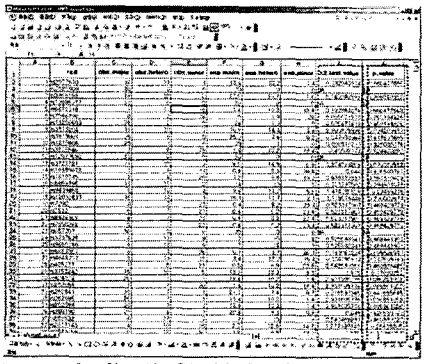


Genotype data



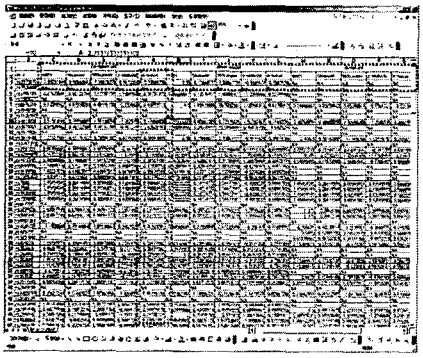
MACROGEN Leading the Biochip Revolution! 25

HWE test



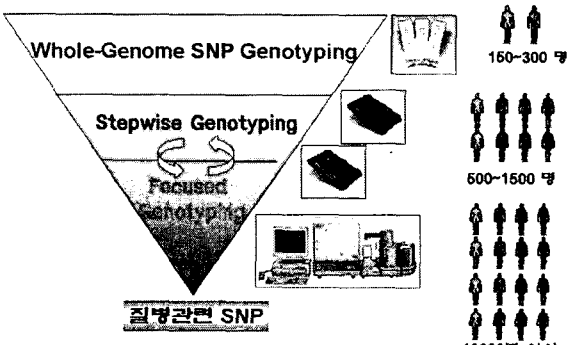
MACROGEN Leading the Biochip Revolution! 26

Model Report



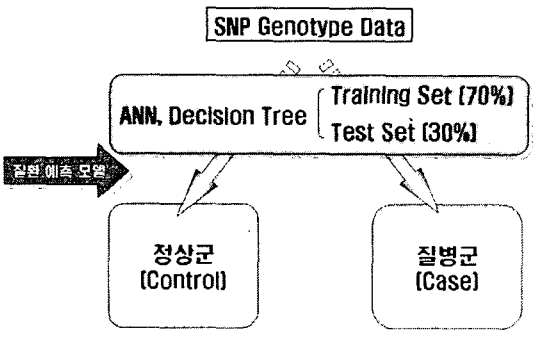
MACROGEN Leading the Biochip Revolution! 27

Multi-stage Genotyping



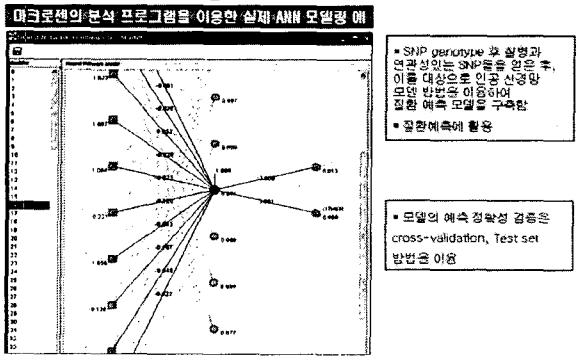
MACROGEN Leading the Biochip Revolution! 28

Classification Model



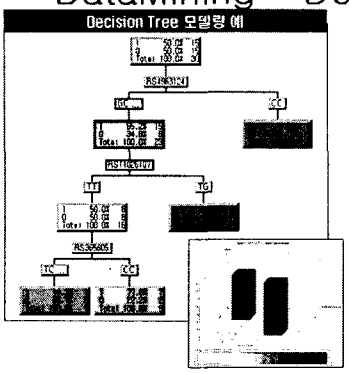
MACROGEN Leading the Biochip Revolution! 29

DataMining – Neural Network



MACROGEN Leading the Biochip Revolution! 30

DataMining – Decision Tree



- Decision Tree 활용한 예제
- 결론
  - rs4963124 : CC - case.
  - rs4963124 : GC, GG
    rs11026107 : TG - control.
  - rs4963124 : GC, GG
    rs11026107 : TT
    rs365605 : CC - control.
  - rs4963124 : GC, GG
    rs11026107 : TT
    rs365605 : TC, TT - case.

# of Sample : 30,
# of Polymorphism :255

RS4963124, RS11026107, RS365605
질환군과 대조군 분류.

Conclusion

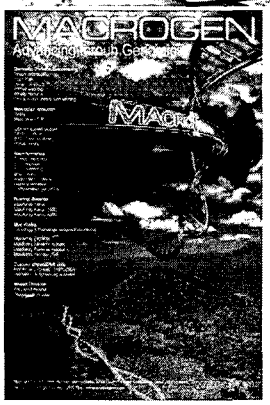
Model & Expectation

Illumina SNP Chip(GoldenGate, Infinium)

SNP 데이터 분석:
HWE, Haplotype, LD, Tagging SNP data 분석 ...

질환과의 연관성 분석:
Chi-Square, Logistic Regression, ANN, Decision Tree...

모델 생성 및 질환 예측



Thank You!