(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2009-0076113
(43) 공개일자 2009년07월13일

(51) Int. Cl.
C12Q 1/68 (2006.01)
(21) 출원번호 10-2008-0001873
(22) 출원일자 2008년01월07일
심사청구일자 없음

(71) 출원인
조선대학교산학협력단
광주광역시 동구 서석동 375, 조선대학교 산학기획팀
(72) 발명자
최규정
광주시 동구 서석2동 375번지
김혁주
전북 익산시 신용동 344-2
(뒷면에 계속)
(74) 대리인
황주명, 성정현, 위병갑

전체 청구항 수 : 총 5 항

(54) 마이크로어레이 데이터 분석 방법

(57) 요 약

본 발명은 방대한 양의 마이크로어레이 데이터를 신속하고 정확하게 분석함으로써 보다 정확한 유전자 진단을 제공할 수 있는 마이크로어레이 데이터 분석 방법을 제공하는 것을 목적으로 한다. 본 발명에 따른 마이크로어레이 데이터 분석 방법은 샘플로부터 분리된 DNA에 형광을 표지하는 단계; 기판 상에 복수의 프로브가 고정되어 있는 마이크로어레이에 상기 형광 표지된 DNA를 제공하는 단계; 상기 프로브와 결합하지 않은 DNA를 제거하는 단계; 상기 마이크로어레이의 형광 세기를 검출하는 단계; 및 교락 설계법(confounding designs) 또는 요인 설계법(factrorial designs)을 이용하여 형광 세기 데이터를 분석하는 단계를 포함한다.

(72) 발명자

이윤동

서울 광진구 화양동 1번지

수드히르, 굽타

미합중국 일리노이주 60115 데칼브 노던 일리노이 주 유니버시티디비젼 오브 스테틱스

특허청구의 범위

청구항 1

샘플로부터 분리된 DNA에 형광을 표지하는 단계;

기판 상에 복수의 프로브가 고정되어 있는 마이크로어레이에 상기 형광 표지된 DNA를 제공하는 단계;

상기 프로브와 결합하지 않은 DNA를 제거하는 단계;

상기 마이크로어레이의 형광 세기를 검출하는 단계; 및

교락 설계법(confounding designs) 또는 요인 설계법(factrorial designs)을 이용하여 형광 세기 데이터를 분석하는 단계를 포함하는 마이크로어레이 데이터 분석 방법.

청구항 2

제 1항에 있어서,

상기 교락 설계법은 교락된 행열 설계법(confounded row-column designs)인 것을 특징으로 하는 마이크로어레이 데이터 분석 방법.

청구항 3

제 1항에 있어서,

상기 요인 설계법은 균형 요인 설계법(balanced factorial desings)인 것을 특징으로 하는 마이크로어레이 데이터 분석 방법.

청구항 4

제 1항에 있어서,

상기 교락 설계법 또는 요인 설계법은 복수의 인자들 및 상기 각 인자의 복수의 레벨들을 선정하는 단계를 포함하는 것을 특징으로 하는 마이크로어레이 데이터 분석 방법.

청구항 5

제 4항에 있어서,

상기 인자는 마이크로어레이 슬라이드, 샘플의 종류, 혼성화 시간, 실험자의 성별, 실험 시각 및 실험시의 날씨로부터 선택되는 하나 이상인 것을 특징으로 하는 마이크로어레이 데이터 분석 방법.

명 세 서

발명의 상세한 설명

기 술 분 야

<1> 본 발명은 마이크로어레이 데이터 분석 방법에 관한 것으로, 보다 상세하게 본 발명은 다수의 인자를 고려하여 마이크로어레이를 효과적으로 분석할 수 있는 마이크로어레이 데이터 분석 방법에 관한 것이다.

배 경 기 술

<2> 최근 cDNA 마이크로어레이는 생물학적 샘플들에서 수천개의 유전자들의 상대적 발현 레벨을 동시에 연구하기 위하여 매우 널리 이용되고 있다. 마이크로어레이 슬라이드는 수천개의 알려진 유전자 또는 특정 유전자의 단편들을 스팟 형태로 그의 표면에 구비하고 있다. 상기 스팟들은 프로브라고도 불리는데, 그들은 DNA, cDNA 또는 올리고뉴클레오티드일 수 있다.

<3> 연구 중인 2개의 생물학적 샘플, 즉 타겟으로부터 유도된 cDNA는 형광 염색, 예컨대, cy3(녹색) 및 cy5(적색)로 표지되고 마이크로어레이에 인가된다. 일반적으로, 질병을 갖는 조직과 정상 조직이 2개의 생물학적 샘플이다. 상기 표지된 타겟 cDNA 분자는 타겟의 세포 내에서 발현된 유전자에 대응하는 어레이 상의 프로

브와 결합한다. 상기 결합 과정은 혼성화라고 한다. 상기 프로브 및 타겟 분자들은 단일 스트랜드의 DNA이다. 녹색(적색) 타겟 스트랜드 및 프로브 스트랜드는 서로 상보적인 경우 결합하거나 혼성화할 것이다.

<4> 각 프로브에 혼성화된 타겟의 양은 특수 스캐너를 이용하여 측정된다. 상기에서 측정된 수치들은 녹색 및 적색 빛의 세기의 형태이다. 일반적으로 상기 빛 세기는 로그 변형된다. 마이크로어레이 실험의 주요 목적은 상기 빛 세기 데이터를 이용하여 2개의 샘플 조직에서 상이하게 발현되는 유전자를 확인하는 것이다.

<5> 상기 빛 세기 데이터를 분석하기 위하여 여러 방법이 적용되었다. 하지만, 종래의 분석 방법들은 방대한 양의 데이터를 신속하고 정확하게 분석함에 있어서 여전히 한계를 드러내고 있는 실정이다.

발명의 내용

해결 하고자하는 과제

<6> 본 발명은 상기 종래기술의 문제점을 해결하기 위하여 안출된 것으로, 본 발명의 목적은 방대한 양의 마이크로어레이 데이터를 신속하고 정확하게 분석함으로써 보다 정확한 유전자 진단을 제공할 수 있는 마이크로어레이 데이터 분석 방법을 제공하는 것이다.

과제 해결수단

<7> 상기 본 발명이 이루고자 하는 기술적 과제를 달성하기 위하여, 본 발명은 샘플로부터 분리된 DNA에 형광을 표지하는 단계; 기판 상에 복수의 프로브가 고정되어 있는 마이크로어레이에 상기 형광 표지된 DNA를 제공하는 단계; 상기 프로브와 결합하지 않은 DNA를 제거하는 단계; 상기 마이크로어레이의 형광 세기를 검출하는 단계; 및 교락 설계법(confounding designs) 또는 요인 설계법(factrorial designs)을 이용하여 형광 세기 데이터를 분석하는 단계를 포함하는 마이크로어레이 데이터 분석 방법을 제공한다.

<8> 상기 교락 설계법은 교락된 행열 설계법(confounded row-column designs)일 수 있다.

<9> 상기 요인 설계법은 균형 요인 설계법(balanced factorial desings)일 수 있다.

<10> 상기 교락 설계법 또는 요인 설계법은 복수의 인자들 및 상기 각 인자의 복수의 레벨들을 선정하는 단계를 포함할 수 있다.

<11> 상기 인자는 마이크로어레이 슬라이드, 샘플의 종류, 혼성화 시간, 실험자의 성별, 실험 시각 및 실험시의 날씨로부터 선택되는 하나 이상일 수 있다.

효 과

<12> 본 발명의 마이크로어레이 데이터 분석 방법에 따르면 방대한 양의 마이크로어레이 데이터를 신속하고 정확하게 분석함으로써 보다 정확한 유전자 진단을 제공할 수 있는 장점을 갖는다.

발명의 실시를 위한 구체적인 내용

<13> 본 발명에 따른 마이크로어레이 데이터 분석 방법은 다양한 실험계획법을 이용한다. 이하, 다양한 실험계획법을 이용한 마이크로어레이 데이터 분석 방법에 대하여 보다 상세하게 설명한다.

<14> 2^2 cDNA 실험

<15> 먼저, 특정 질병에 관련된 유전자의 생물학적 영향을 연구하기 위한 2^2 cDNA 실험에 대해 설명한다. 2개의 인자를 사용하고, 첫 번째 인자는 2개의 레벨(2종류의 돌연변이주)을 갖는 세포주, 및 2개의 레벨(0시간 및 24시간)을 갖는 시간이다. 상기 2개의 인자를 각각 F_1 및 F_2라 하고, 각 인자의 레벨을 0 및 1로 코딩한다.

<16> 본 실험의 목표들 중 하나는 상호작용 효과를 연구하는 것이다. 상호작용의 존재는 24시간과 0시간에서 2개의 샘플에서 상이하게 발현되는 유전자를 확인하는데 도움을 준다. 실험 중인 4개의 처리 조합에 대응하는 4개의 생물학적 샘플이 존재한다. 하지만, 하나의 마이크로어레이 슬라이드에는 단지 2개의 샘플만이 혼성화될 수 있다.

<17> 슬라이드 대 슬라이드의 변이성을 설명하기 위하여, 슬라이드는 블락킹 인자로 취급된다. 따라서, 2개

이상의 샘플이 연구 중인 경우 마이크로어레이 슬라이드는 크기 2의 불완전 블락을 형성한다. 2개의 주요 효과 및 상호작용을 추정하게 하는 연결 설계는 최소 4개의 슬라이드를 포함할 것이다.

<18> 예컨대, F_1 주요 효과는 2개의 블락 [00, 01] 및 [10, 11] 사이에서 교락된다(confounded). 여기서, 블락 [00, 01]은 처리 조합 00 및 01 등에 대응하는 2개의 조직 샘플과 혼성화한 마이크로어레이 슬라이드를 나타낸다. 유사하게, 2개의 블락 [00, 10] 및 [01, 11]은 F_2 주요 효과를 교락한다. 상기 4개의 블락에 대응하는 4개의 슬라이드는 정보의 손실 없이 F_1F_2 상호작용을 추정하게 하고, 2개의 주요 효과 각각은 50% 효율로 추정된다.

<19> 한편, 슬라이드에 추가하여 2개의 염색약이 두 번째 블락킹 인자를 형성할 수 있다.

<20> 블락 크기 2의 교락 설계(Confounded Designs with Block Size Two)

<21> 각 2개의 레벨을 갖는 m개의 인자들, $F_1,F_2,\cdots,F_m$을 포함하는 인자 실험을 고려한다. 처리 조합(treatment combinations)은 $a_1a_2\cdots a_m$으로 표시하고, 상기에서 a_i=0 또는 1이고, 이는 i번째 인자의 코딩된 레벨을 나타낸다. D는 교락법을 이용하여 얻어지는 m-인자 실험을 위한 크기 2의 2^{m-1} 블락을 갖는 단일 반복 설계(single replicate design)를 의미한다. 또한, D의 블락 사이의 교락된 인자 효과(주요 효과 및/또는 상호작용)는 $B_i(i=1,2,\cdots,2^{m-1}-1)$를 나타내고, $A_i(i=1,2,\cdots,2^{m-1})$는 교락되지 않은 인자 효과를 나타내고, 따라서 D에서 추정가능하다.

<22> (t_{i1},t_{i2})는 D의 i번째 블락에서 2개의 처리 조합을 나타내고, (y_{i1},y_{i2})는 그들에 대응하는 관측을 나타낸다$(i=1,2,\cdots,2^{m-1})$. 블락 총합 $y_{i1}+y_{i2}(i=1,2,\cdots,2^{m-1})$의 대조는 블락 효과를 추정하고, 인자 효과 $B_i(i=1,2,\cdots,2^{m-1}-1)$는 교락된다. 반면, 교락되지 않은 인자 효과 $A_i(i=1,2,\cdots,2^{m-1})$는 블락 내의 비교 $y_{i1}-y_{i2}(i=1,2,\cdots,2^{m-1})$를 이용하여 추정된다.

<23> $$y_d=[(y_{11}-y_{12}),(y_{21}-y_{22}),\cdots,(y_{b1}-y_{b2})]'$$

<24> 상기 식에서, $b=2^{m-1}$은 블락의 수를 나타낸다. $h_j=(h_{j1},h_{j2},\cdots,h_{jb})'$를 $E(h'_jy_d)=A_j(j=1,2,\cdots,2^{m-1})$로 둔다. 그렇다면, 명백하게 $h_j(j=1,2,\cdots,2^{m-1})$는 크기 2^{m-1}의 상호 직교 벡터의 완벽한 세트를 형성한다.

<25> 또한, $Var(h'_jy_d)=h'_jh_j\sigma^2$이고, 상기에서, σ^2는 차이 $y_{i1}-y_{i2}$의 일정 변이를 나

타낸다. 2^{m-1} 어레이를 포함하는 마이크로어레이 실험의 경우에 있어서, 블락내 비교 $y_{i1}-y_{i2}$는 어레이 상에 혼성화하는 2개의 처리 조합에 대한 $\log_2$ 발현 비율을 나타낸다.

<26> 모든 2^{m-1} 인자 효과가 추정가능한 부분적 교락 설계는 상이한 세트의 인자 효과들을 교락하는 추가적인 반복을 추가함으로써 얻어질 수 있다. 이하에서 설명할 실시예에서는 2^m 실험을 이용한 교락법을 예시하지만, 본 발명은 그에 한정되지 않으며 일반적으로 프라임 넘버의 레벨 수를 갖는 인자들에도 적용될 수 있다.

<27> 인자 마이크로어레이에 적용(Application to Factorial Microarrays)

<28> m 인자 마이크로어레이 실험을 위하여, 상기 유전자 발현 수치용 선형 모델은 다음과 같이 주어진다.

<29>
$$y_{ijk}=\mu_j+\alpha_{ij}+\delta_{jk}+\tau_{ijk}+\varepsilon_{ijk},$$

<30>
$$i=1,2,\ldots,b;j=1,2,\ldots,g;k=1,2$$

<31> 상기 식에서, i번째 유전자에 대하여, μ_j는 전체 평균이고, α_{ij}는 I번째 어레이 효과이고, δ_{j1} 및 δ_{j2} 은 각각 염색약 인자의 레벨 1 및 2의 효과를 나타낸다. $F_{g,}\tau_{ijk}$는 처리 조합의 효과를 나타내고, i_k 및 ε_{ijk}은 평균 0 및 편차 σ_j^2로 독립적으로 분포된 랜덤 에러를 나타낸다. 처리 대조는 블락내 비교 $y_{ij}-y_{i_2j}$로부터 추정되기 때문에, 상기 차이에 대한 모델은 $(y_{i1}-y_{i2})=(\delta_1-\delta_2)+(\tau_{i1}-\tau_{i2})+(\varepsilon_{i1}-\varepsilon_{i2})$로 주어진다.

<32> < 2^2실험>

<33> 1. 2^2에 대한 인자 실험의 주요 설계

<34> (1) 블락 크기 2를 갖는 블락의 수 : = $\frac{2^2}{2}=2^{2-1}$

<35>

<36> [표 1] 2^2실험

<37>

D_{11} : [00,01],[10,11]
D_{12} : [00,10],[01,11]
D_{13} : [00,11],[01,10]

<38> (2) 교락된 효과

<39> 교락된 인자들의 주요 설계에 대한 방법은 다음과 같이 다루고 상기 표 1로 주어진다.

<40> ⓧ : 각 인자에 대한 기호 산물, 0 및 1은 각각 - 및 +로 표시된다.

<41> 1) 주요 효과 : 블락 기호들을 곱한 후 "+"를 나타내는 것들

<42> 2) 상호작용 효과: 블락 기호의 곱셈으로부터 적어도 2개의 인자에 대한 기호들을 곱한 후 "+"를 나타내는 것들

<43> 3) 교락된 인자들에 대한 주요 설계의 예

<44> $$D_{11} : 00ⓧ01 = (--ⓧ-+) = (+,-) \Rightarrow F_1$$

<45> $$D_{12} : 00ⓧ10 = (--ⓧ+-) = (-,+) \Rightarrow F_2$$

<46> $$D_{13} : 00ⓧ11 = (--ⓧ++) = (-,-) \Rightarrow F_1F_2$$

<47> (3) 마이크로어레이 설계

<48> 1) 통상적인 루프 설계(CL)

<49> 주요 설계는 반복 없이 D_{11} 및 D_{12}로 이루어진다 (4 슬라이드).

<50> $$D_1=\begin{pmatrix} 00 & 11 & 10 & 01 \\ 01 & 10 & 00 & 11 \end{pmatrix}$$

$$D_{11} \qquad D_{12}$$

<51> 2) 크로스 스왑 설계(CS)

<52> 주요 설계는 D_{13}의 2개의 반복으로 이루어진다 (4 슬라이드).

<53> $$D_2=\begin{pmatrix} 00 & 01 & 11 & 10 \\ 11 & 10 & 00 & 01 \end{pmatrix}$$

<54> 상기 설계 형상은 주요 설계를 조정 및 재배열한 후에 얻어졌고, 그에 의해 2개의 행은 동일한 처리 조합 수를 갖고 상기 행들에 대한 인자 합은 동일하다.

<55> $D^2,(00 \oplus 01 \oplus 11 \oplus 10)(\mathrm{mod}2)=00:row1$ 에서,

<56> $$(11 \oplus 10 \oplus 00 \oplus 01)(\mathrm{mod}2)=00:row2$$

<57> 상기에서, 기호 $\oplus$는 인자 합으로 정의된다.

<58> 3) 16 슬라이드를 이용한 CL의 4 반복

<59> $$D_3=\begin{pmatrix} 00\ 11\ 10\ 01 : 01\ 10\ 00\ 11 : 00\ 11\ 10\ 01 : 01\ 10\ 00\ 11 \\ 01\ 10\ 00\ 11 : 00\ 11\ 10\ 01 : 01\ 10\ 00\ 11 : 00\ 11\ 10\ 01 \end{pmatrix}$$

<60> 4) 16 슬라이드를 이용한 CL 및 CS의 반복

$$D_4=\begin{pmatrix} 00\ 11\ 10\ 01 : 01\ 10\ 00\ 11 : 00\ 01\ 11\ 10 : 11\ 10\ 00\ 01 \\ 01\ 10\ 00\ 11 : 00\ 11\ 10\ 01 : 11\ 10\ 00\ 01 : 00\ 01\ 11\ 10 \end{pmatrix}$$

<61>

<62> 5) 16 슬라이드를 이용한 D_{11}의 3 반복, D_{12}의 2 반복 및 D_{13}의 3 반복

$$D_5=\begin{pmatrix} 00\ 10\ 01\ 11\ 00\ 10 : 10\ 01\ 00\ 11 : 01\ 11\ 00\ 01\ 11\ 10 \\ 01\ 11\ 00\ 10\ 01\ 11 : 00\ 11\ 10\ 01 : 10\ 00\ 11\ 10\ 00\ 01 \end{pmatrix}$$

$$D_{11}\quad D_{11}\quad D_{11}\quad D_{12}\quad D_{12}\quad D_{13}\quad D_{13}\quad D_{13}$$

<63>

<64> < 2^3실험>

<65> 1. 2^3에 대한 인자 실험의 주요 설계

$$\frac{2^3}{2}=2^{3-1}=4$$

<66> (1) 크기 2의 블락 수 : =

<67> (2) D_{21}의 블락 구성

<68> 1) 주요 블락 : 000, 및 나머지 7 처리 조합 중 하나 선택

<69> 2) 다른 블락 : 나머지 6개로부터 쌍을 만들어 모드 (2)에 의해 주요 블락에 대한 인자 합과 등가이도록 한다.

<70> [표 2] 2^3실험

<71>

D_{21}	(000, 100)(001, 101)(010, 110)(011, 111)	F_2,F_3,F_2F_3
D_{22}	(000, 010)(001, 011)(100, 110)(101, 111)	F_1,F_3,F_1F_3
D_{23}	(000, 001)(010, 011)(100, 101)(110, 111)	F_1,F_2,F_1F_2
D_{24}	(000, 111)(001, 110)(010, 101)(011, 100)	F_1F_2,F_1F_3,F_2F_3
D_{25}	(000, 011)(001, 010)(101, 110)(100, 111)	$F_1,F_2F_3,F_1F_2F_3$
D_{26}	(000, 101)(001, 100)(011, 110)(010, 111)	$F_2,F_1F_3,F_1F_2F_3$
D_{27}	(000, 110)(010, 100)(011, 101)(001, 111)	$F_3,F_1F_2,F_1F_2F_3$

<72> (3) 교락된 효과

<73> 교락된 인자들의 주요 설계에 대한 방법은 다음과 같이 다루고 상기 표 2로 주어진다.

<74> 1) D_{21} : 000ⓧ100 = (---ⓧ+--) = (- + +) [F_2,F_3,F_2F_3]

<75> 2) D_{24} : 000ⓧ111 = (---ⓧ+++) = (- - -) [F_1F_2,F_1F_3,F_2F_3]

<76> 3) D_{27} : (000ⓧ110) = (---ⓧ++-) = (- - +) [$F_3,F_1F_2,F_1F_2F_3$]

<77> (4) 16 슬라이드를 이용한 마이크로어레이 설계

<78> 1) 표 2로부터 4개의 주요 설계를 무작위로 선택하고, 각 블락을 조정하고 재배열하여 행 1 및 행 2가 동일한 각 처리 조합의 수를 갖도록 한다.

<79> 예 : $D_{24}D_{25}D_{26}D_{27}$ 사용

$$D_1=\begin{pmatrix} 111\ 110\ 010\ 011 : 000\ 001\ 101\ 100 : 000\ 001\ 011\ 010 : 110\ 100\ 101\ 111 \\ 000\ 001\ 101\ 100 : 011\ 010\ 110\ 111 : 101\ 100\ 110\ 111 : 000\ 010\ 011\ 001 \end{pmatrix}$$

D_{24} D_{25} D_{26} D_{27}

<80>

<81> 2) 표 2에서 주요 설계를 적절히 조합하여 구성

<82> i) 2개의 임의 주요 설계 및 다른 주요 설계의 2 반복을 조합하여 구성

<83> ii) 2개의 임의 주요 설계의 2 반복을 조합하여 구성

<84> < 2^4실험>

<85> 1. 에 대한 인자 실험의 주요 설계

<86> (1) 크기 2의 블락 수: = $\frac{2^4}{2} = 2^{4-1} = 8$

<87> (2) D_{31}의 블락 구성

<88> 1) 주요 블락 : 0000, 및 나머지 15개의 처리 조합으로부터 1 선택

<89> 2) 다른 블락 : 나머지 14개로부터 쌍을 만들어 모드 (2)에 의해 주요 블락에 대한 인자 합과 등가이도록 한다

<90> 3) 모든 가능한 주요 슬라이드 :

<91> [0000, 0001] [0000, 0010] [0000, 0011] [0000, 0100]

<92> [0000, 0101] [0000, 0110] [0000, 0111] [0000, 1000]

<93> [0000, 1001] [0000, 1010] [0000, 1011] [0000, 1100]

<94> [0000, 1101] [0000, 1110] [0000, 1111]

<95> 4) 주요 설계의 일반적인 구성 방법:

<96> 하기 표 3과 같이 주요 슬라이드 $[a_1, b_1]$로부터 선택하여 A1 $a_1 \oplus b_1 (\mathrm{mod}2)$

<97> [표 3] 2^4 실험

<98>

D_{31}	(0000, 0001) (0010, 0011) (0100, 0101) (0110, 0111) (0100, 1100) (0101, 1101) (0110, 1110) (0111, 1111)	$F_1.F_2, F_3, F_1F_2,$ $F_2F_4, F_3F_4, F_2F_3F_4$
D_{32}	(0000, 0010) (0001, 0011) (0100, 0110) (0101, 0111) (1000, 1001) (1010, 1011) (1100, 1101) (1110, 1111)	$F_1.F_2, F_4, F_1F_2,$ $F_1F_4, F_2F_4, F_1F_2F_4$
D_{33}	(0000, 0011) (0001, 0010) (0100, 0111) (0101, 0110) (1000, 1011) (1001, 1010) (1100, 1111) (1101, 1110)	F_1, F_2, F_1F_2, F_3F_4 $F_1F_3F_4, F_2F_3F_4, F_1F_2F_3F_4$
D_{34}	(0000, 0100) (0001, 0101) (0010, 0110) (0011, 0111) (1000, 1100) (1001, 1101) (1010, 1110) (1011, 1111)	$F_1.F_3, F_4, F_1F_3$ $F_1F_4, F_3F_4, F_1F_3F_4$
D_{35}	(0000, 0101) (0001, 0100) (0010, 0111) (0011, 0110) (1000, 1101) (1001, 1100) (1010, 1111) (1011, 1110)	$F_1.F_3, F_1F_3, F_2F_4$ $F_1F_2F_4, F_2F_3F_4, F_1F_2F_3F_4$

D_{36}	(0000, 0111) (0001, 0110) (0010, 0101) (0011, 0100) (1000, 1110) (1001, 1111) (1010, 1100) (1011, 1101)	$F_1.F4, F_1F_4, F_2F3$ $F_1F_2F_3, F_2F_3F_4, F_1F_2F_3F_4$
D_{37}	(0000, 0111) (0001, 0110) (0010, 0101) (0011, 0100) (1000, 1111) (1001, 1110) (1010, 1101) (1011, 1100)	$F_1.F_2F_3, F_2F_4, F_3F_4$ $F_1F_2F_3, F_1F_2F_4, F_1F_3F_4$
D_{38}	(0000, 1000) (0001, 1001) (0010, 1010) (0011, 1011) (0100, 1100) (0101, 1101) (0110, 1110) (0111, 1111)	$F_2.F_3, F_4, F_2F_3$ $F_2F_4, F_3F_4, F_2F_3F_4$
D_{39}	(0000, 1001) (0001, 1000) (0010, 1011) (0011, 1010) (0100, 1101) (0101, 1101) (0110, 1111) (0111, 1110)	$F_2.F_3, F_1F_4, F_2F_3$ $F_1F_2F_4, F_1F_3F_4, F_1F_2F_3F_4$
D_{310}	(0000, 1010) (0001, 1011) (0010, 1000) (0011, 1001) (0100, 1110) (0101, 1111) (0110, 1100) (0111, 1101)	$F_2.F_4, F1F_3, F_2F_4$ $F_1F_2F_3, F_1F_3F_4, F_1F_2F_3F_4$
D_{311}	(0000, 1011) (0001, 1010) (0010, 1001) (0011, 1000) (0100, 1111) (0101, 1110) (0110, 1101) (0111, 1100)	$F_2.F_1F_3, F1F_4, F_3F_4$ $F_1F_2F_3, F_1F_2F_4, F_2F_3F_4$
D_{312}	(0000, 1100) (0001, 1101) (0010, 1110) (0011, 1111) (0100, 1000) (0101, 1001) (0110, 1010) (0111, 1011)	$F_3.F_4, F_1F_2, F_3F_4$ $F_1F_2F_3, F_1F_2F_4, F_1F_2F_3F_4$
D_{313}	(0000, 1101) (0001, 1100) (0010, 1111) (0011, 1110) (0100, 1001) (0101, 100) (0110, 1011) (0111, 1010)	$F_3.F_1F_2, F_1F_4, F_2F_4$ $F_1F_2F_3, F_1F_3F_4, F_2F_3F_4$
D_{314}	(0000, 1110) (0001, 1111) (0010, 1100) (0011, 1101) (0100, 1010) (0101, 1011) (0110, 1000) (0111, 1001)	$F4.F_1F_2, F_1F_3, F_2F_3$ $F_1F_2F_4, F_1F_3F_4, F_2F_3F_4$
D_{315}	(0000, 1111) (0001, 1110) (0010, 1101) (0011, 1100) (0100, 1011) (0101, 1010) (0110, 1001) (0111, 1000)	$F_1F_2.F_1F_3, F_1F_4, F_2F_3$ $F_2F_4, F_3F_4, F_1F_2F_3F_4$

<99> <3× 2 실험>

<100> 1. 3× 2에 대한 인자 설계의 주요 설계

<101> [표 4] 3× 2 실험

<102>

D_{41}	(00, 01) (10, 11) (20,21)
D_{42}	(00, 11) (10, 21) (20, 01) (01, 10) (11, 20) (21, 00)
D_{43}	(00, 10) (10, 20) (20, 00) (01, 11) (11, 21) (21, 01)

<103> (1) 구성 방법 : 일반화된 사이클 방법 사용

<104> ⅰ) 10씩 추가하고 첫 번째 순환까지 계속

<105> ⅱ) 10, 10, 11씩 추가하고 순환까지 계속

<106> 상기에서, 20 및 21이 각각 10 및 11 대신에 사용될 수 있다.

<107> (예)

<108> 1) D_{41} (00, 01) $\oplus$ 10(mod(3, 2)) = (10, 11) $\oplus$ 10 mod(3, 2)

<109> = (20, 21) $\oplus$ 10 mod(3, 2) = (00, 01) : main block

<110> D'_{41} (00, 01) $\oplus$ 20 mod(3, 2) = (20, 21) $\oplus$ 20 mod(3, 2)

<111> = (10, 11) $\oplus$ 20 mod(3, 2) = (00, 01) circulation

<112> $\therefore D_{41} \cong D'_{41}$

<113> 2) D_{42} (00, 11) $\oplus$ 10(mod(3, 2)) = (10, 21) $\oplus$ 10 mod(3, 2)

<114> = (20, 01) $\oplus$ 11 mod(3, 2) = (01, 10) $\oplus$ 10 mod(3, 2)

<115> = (11, 20) $\oplus$ 10 mod(3, 2) = (21, 00) $\oplus$ 11 mod(3, 2)

<116> = (00, 11) : 순환 주요 블락

<117> (2) 다른 주요 설계

<118> i) D_{44}(00, 20) (10,00) (20,10)

<119> [처음으로부터, $\oplus$ 10 mod(3, 2) , $\oplus$10 mod(3, 2), $\oplus$ 11 mod(3, 2)]

<120> (01, 21) (11,01) (21,11) (00,20) [주요 블락의 순환]

<121> [처음으로부터, $\oplus$ 10 mod(3, 2), $\oplus$ 10 mod(3, 2), $\oplus$ 11 mod(3, 2)]

<122> ii) D_{45} (00, 21) (10, 01) (20, 11) (01, 20) (11, 00) (21, 10) (00,21) [주요 블락의 순환]

<123> [처음으로부터, $\oplus$10, $\oplus$10, $\oplus$11, $\oplus$10, $\oplus$10, $\oplus$11 순환]

<124> 여기서 , $D_{44} \cong D_{43}, D_{45} \cong D_{42}$

<125> 따라서, 3 블락을 이용한 BS 설계의 수는 5가 아니라 3이고, 6 블락을 이용한 XL 및 AL 설계의 수는 5가 아니라 3임을 알 수 있다.

<126> <마이크로어레이 설계>

<127> (1) BS : 6 슬라이드를 이용한 D_{41} 의 2-반복

<128> $$D_1\begin{pmatrix} 001020:011121 \\ 011121:001020 \end{pmatrix}$$

<129> (2) 12 슬라이드를 이용한 D_{41} 및 D_{42}의 2 반복

<130> $$D_2\begin{pmatrix} 001020011121:001020011121 \\ 011121001020:112101102000 \end{pmatrix}$$

<131> 교락된 행열 설계(confounded row-column designs)

<132> <모델 및 예비사항>

<133> v 처리, p 행 및 q 열을 갖는 행열 설계 D를 고려한다. r은 처리 반복의 상수를 나타낸다. 행열 설계 모델은 다음과 같이 주어진다.

<134> $Y_{ijk} = u + \tau_i + \rho_j + v_k + \varepsilon_{ijk}$,

<135> 상기에서, Y_{ijk}는 i번째 처리가 인가된 j번째 행 및 k번째 열로부터 관측을 나타내고, τ_i, ρ_j, 및 v_k은 각각 i번째 처리, j번째 행, 및 k번째 칼럼의 효과를 나타내고, ε '은 평균 0 및 편차 σ^2을 갖는 비연관 랜덤 에러를 나타낸다.

<136> N1 및 N2는 처리-행 및 처리-열 발생 매트릭스를 나타낸다. N1 및 N2에 대응하는 성분 설계는 D1 및 D2로 표시된다. 처리 파라미터 τ의 벡터를 추정하기 위한 인트라-블락내 감소된 표준 식은 C_τ=Q로 주어지고,

$$C = rI - \frac{1}{q}N_1N_1' - \frac{1}{p}N_2N_2' + \frac{r^2}{pq}11',$$

<137>

<138> 여기서, I는 항등 매트릭스를 나타내고 1은 1's의 열 벡터를 나타내고 각 크기는 υ이다.

$$C = C_1 + C_2 - C_b,$$

<139>

<140> 여기서,

$$C_1 = rI - \frac{1}{q}N_1N_1',$$

$$C_2 = rI - \frac{1}{p}N_2N_2',$$

$$C_b = r\left(I - \frac{1}{v}11'\right)$$

<141>

<142> 는 각각 성분 설계 D_1, D_2, 및 D_b의 블락내 정보 매트릭스이다. 여기서 D_b는 행 및 열 분류를 무시함으로서 얻어지는 하나의 블락을 갖는 설계를 나타내고, 인시던스 매트릭스 N_b=r1이다.

<143> <구성 방법>

<144> 각각 s 레벨을 갖는 m개의 인자들 F_1, F_2, ... , F_m을 포함하는 S^m 인자 실험에 대해 살펴본다. 처리 조합은 m 개(a_1, a_2 · · · a_m)으로 표시되고, 여기서 어떤 인자의 s 레벨은 0, 1, ..., s-1로 표시된다. S^m 인자 실험에 대한 행열 설계 D를 고려한다. 행의 수(p) 및 열의 수(q)가 다음의 표시를 허용하는 것으로 가정한다: $p=S^{m1}$, $q=S^{m2}$, $pq=S^{m1+m2}=rS^m$. 따라서,

<145> $r=S^{m1+m2-m}$

<146> r≥1이기 때문에,

<147> m1 + m2 ≥ m.

<148> A_1, A_2, ..., A_{m-m2}는 D의 S^{m1} 행들 사이의 교락된 독립 상호작용을 나타낸다. S^{m-m2} - 1 처리 자유도는 D의 $p=S^{m1}=rS^{m-m2}$ 행들 사이에서 교락된다. 행들 사이에 교락된 효과들(각 효과는 s-1의 자유도를 가짐)의 총 수는 $(S^{m-m2}-1)/(s-1)$로 주어진다. 따라서, D의 행들 사이에 교락된 일반화된 상호작용의 수는 다음과 같이 주어진다.

$$g_1 = \frac{s^{m-m_2} - 1}{s - 1} - (m - m_2).$$

<149>

<150> 상기 일반화된 상호작용은 A_{m-m2+1}, A_{m-m2+2}, ..., $A_{m-m2+g1}$와 같이 표시된다.

<151> B_1, B_2, ..., B_{m-m1}는 D의 S^{m2} 열들 사이의 교락된 독립 상호작용을 나타낸다. 유사하게 D의 S^{m2} 열들 사이에 교락된 일반화된 상호작용의 수는 다음과 같이 주어진다.

$$g_2 = \frac{s^{m-m_1} - 1}{s - 1} - (m - m_1).$$

<152>

<153> 상기 g2 일반화된 상호작용은 B_{m-m1+1}, B_{m-m1+2}, ..., $B_{m-m1+g2}$와 같이 표시된다.

<154> <적용예>

<155> p=q=2^2인 행열 설계르르 이용하여 2^4 실험을 구성한다. s=2, m=4, mm1=m2=2, m-m1=m-m2=2로 둔다. $A_1=F_1F_2$, $A_2=F_3F_4$, $B_1=F_1F_2F_3$, $B_2=F_2F_3F_4$로 둔다. $A_3=F_1F_2F_3F_4$, $B_3=F_1F_4$. 행 및 열 키 블락들 및 생성된 행열 설계는 다음과 같이 주어진다:

Column	Row key block			
key block	0000	1100	0011	1111
0000	0000	1100	0011	1111
0110	0110	1010	0101	1001
1101	1101	0001	1110	0010
1011	1011	0111	1000	0100

<156>

<157> 상기 단일 반복 설계는 상호작용 F_1F_2, F_3F_4, $F_1F_2F_3F_4$, $F_1F_2F_3$, $F_2F_3F_4$ 및 F_1F_4를 완전히 교락한다. 모든 주요 효과, 및 모든 다른 상호작용은 완전 효율로 추정된다.

<158> 본 발명에 따른 마이크로어레이 분석 방법의 전부 또는 일부는 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CDROM, 자기 테이프, 플로피디스크 및 광데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드로 저장되고 실행될 수 있다.

<159> 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.