

지식 베이스를 이용한 교육용 염색체 분석 시스템

Chromosome Analysis System based on Knowledge Base for CAI

박정선, 신용원*

(주)비트컴퓨터, 마산대학 보건행정학과*

Jeong Seon Park, Yong Won Shin*

BIT COMPUTER CO.,LTD., Department of Health Administration, Masan College*

ABSTRACT

The task for chromosome analysis and diagnosis by experienced cytogenetists are being concerned as repetitive, time consuming job and expensive.

For that reason, chromosome analysis system based on knowledge base for CAI has been established to be able to analyze chromosomes and obtain necessary advises from the knowledge base instead of human experts.

That is to say, knowledge base by IF THEN production rule was implemented to a knowledge domain with normal and abnormal chromosomes, and then the inference results by knowledge base could enter the inference data into the database.

Experimental data were composed of normal chromosomes of 2,736 patients'cases and abnormal chromosomes of 259 patients'cases that have been obtained from GTG-banding metaphase peripheral blood and amniotic fluid samples.

The completed system provides variously morphological information by analysis of normal or abnormal chromosomes and it also has the advantage of being able to consult with user on chromosome analysis and diagnosis.

Keyword: chromosome, knowledge base, inference, expert

1. 서론

인간의 유전정보를 지닌 염색체는 선천성 장애 질환의 진단, 악성종양의 진단과 발암물질 연구 등에 있어서 임상적으로 중요하다(J Marget, 1991).

특히, 환자 염색체의 핵형을 분류하여 진단을 내리는 경우에는 정상 염색체와 비정상 염색체의 특징과 관련 증후군, 분염기술, 그리고 ISCN(Mitelman Felix, 1995) 명명법과 같은 기본 지식과, 환자의 증상이나 전체상태, 염색체 영상, 가족력, 임신경력 등과 같은 매우 특수화된 정보가 요구된다(김한화, 1998, 김도현, 1993, 박숙자, 1992)

이러한 정보를 획득하는 세포유전학 검사실에서는 염색체의 배양, 분류, 그리고 진단을 내리는 반복작업을 수행하며, 작업과정에는 고감도의 마이크로 카메라, 각 필름에 따른 특수 현상액, 고감도 흑백 인화지, 전용암실 등을 사용한다.

또한 염색체는 다양한 형태학적인 정보를 제공하기 때문에 염색체를 분류하고 진단을 내리는 행위는 매우 복잡하고 신중함을 요구하며 염색체를 이용한 의사결정이 필요한 경우에는 세포유전학자가 학습한 체계적인 지식과 더불어 현장에서 경험하여 축적한 경험적 지식이 요구된다.

본 연구에서는 염색체의 정상과 비정상에 대하여 현재까지 국내외 관련 정보들을 유용하고 체계적인 정보로 가공, 정리하여 초보 사용자와 전문가에게 조언하고, 다양한 형태학적 정보를 제공하는 것을 목적으로 지식베이스를 이용한 교육용 염색체 분석 시스템을 구축하였다.

따라서 구축 시스템은 전문가의 축적된 지식과 체득한 경험을 기반으로 한 염색체 분석과 진단 정보를 교육하고 학습하게 함으로써 초보 사용자

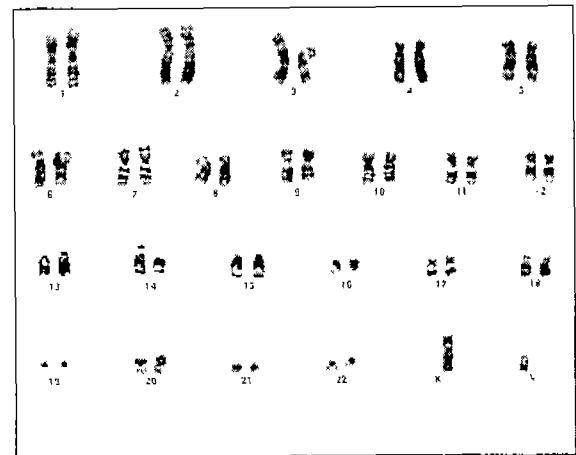
가 부족한 전문가의 풍부한 지식과 경험을 단기간에 공유하여 학습기간이 단축된다. 뿐만 아니라 염색체 분석결과를 이용한 진단 및 이에 대한 설명은 사용자와 시스템간에 협의 시스템(consultation system)의 역할을 담당한다.

2. 연구 대상

2.1 염색체 분석

인간의 중기 염색체는 그림 1과 같이 46 개로 구성된다. 이와 같은 염색체의 핵형분류(karyotyping)는 임상병리사가 수작업을 하거나 컴퓨터에 의한 자동 분류 시스템을 이용하는데, 주로 염색체의 동원체 지수, 상대길이와 같은 특징적인 값들과 분염된 밴드패턴을 인식하여 분류한다.

[Fig. 1. Human metaphase chromosomes, GTG-banding]



수작업의 경우 임상병리사는 환자로부터 샘플을 채취하여 표본의 전처리 후 최소한 유사분열 중기 염색체 20 상 이상을 현미경하에서 눈으로 관찰하고 이 가운데 5~6 상 정도를 현미경에 부착된 자동 카메라 시스템에서 35mm 고감도 마이크로 필름으로 사진을 찍고 각 마이크로 필름에 적합한

현상액으로 현상한 후 고감도 흑백 인화지로 인화한다. 사진으로 인화된 염색체는 하나씩 오리를 하고 검사실에서 전문가가 축적한 지식과 경험을 토대로 각 염색체별로 분류한 후 보고서 용지에 정확하게 배열과 붙이기를 한다. 최종 보고서에는 검체 데이터, 검사실 데이터, 환자 데이터, 핵형분류 결과, 전문가의 진단 및 이에 대한 부연설명을 기록하고, 사진과 함께 복사본을 만들어 보관한다.

3일 정도 소요되는 염색체의 채취와 배양 후에도 이러한 일련의 수작업은 많은 시간과 노력이 소요되고, 검체에 따라서 환자 염색체에 대한 최종 진단까지는 대략 7일 이상이 걸린다(J Marget, 1991).

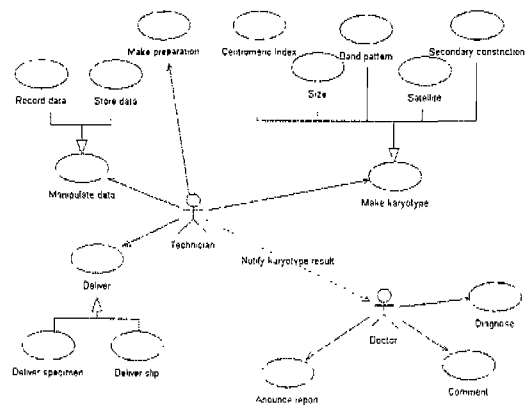
상용화된 여러 회사의 자동 염색체 분석 시스템(Applied Imaging, 1994, PSI, 1995, Meta systems)을 사용할 경우에는 수작업의 대부분이 자동화되었으나, 염색체 분석에서 요구되는 상세한 밴드분석, 분석결과가 가장 잘 나올 수 있는 중기 염색체의 적절한 선택, 관찰된 염색체의 수적 이상이나 구조적 이상에 관계된 진단학적 의사결정이 필요한 경우에는 세포유전학자나 자격을 갖춘 전문가의 개입이 요구된다.

2.2 염색체 분석 모델링

염색체를 분류하고 진단하기 위해서 환자 검체의 접수, 운반, 배양, 보고서 작성, 표본제작, 핵형분류, 그리고 통지를 수행하는 임상병리사의 업무행위와, 통지에 따라 진단, 진단에 대한 부연설명, 최종 보고서를 완성하는 의사의 업무 행위 및 각자의 요구사항을 컴퓨터속의 가상 환경으로 일대일 대응시키기 위해서는 작업의 전체흐름을 정규화하는 모델링이 필요하므로 그림 2와 같이

UML(Unified Modeling Language)(Jake S, 1999, Martin F, 1997)의 use-case 다이어그램으로 표현하였다.

[Fig. 2 A use-case diagram showing behaviors of technician and doctor in cytogenetic laboratory]



3. 연구방법

3.1 지식 베이스 시스템의 기본 개념

지식 베이스 시스템이란 특정분야에서 전문가의 지식을 이용하여 인텔리전트한 의사결정을 내릴 수 있는 컴퓨터 시스템을 의미한다. 적용영역은 주로 의료진단, 지질탐색, 유기화학, 컴퓨터 디자인, 회사 경영과 관리에 이르기까지 다양하다. 즉 지식 베이스 시스템은 빠르고 적절한 의사결정을 내릴 수 있고 확장성, 융통성, 설명력을 갖추고, 불완전한 정보를 이용한 작업 가능성 등을 장점으로 지니고 있다.

전형적인 지식 베이스 시스템은 지식 베이스, 추론엔진, 지식획득과 설명, 사용자 인터페이스로 구성되는데, 지식 베이스와 추론엔진이 지식 베이스 시스템의 핵심이라고 할 수 있다. 지식 베이스는 빠르게 변화하는 단기 정보인 사실(fact)과, 현재 알려진 것에서 새로운 사실이나 가설을 생성하

는 장기 정보인 규칙(rule)들로 구성된다. 대부분의 지식 베이스 시스템은 이러한 지식 베이스를 표현하기 위해서 IF THEN 생성 규칙을 사용한다.

< Table 1. A production rule for Trisomy Syndrome >

<p>Rule to identify disease of Trisomy 21 Classic Type</p> <p>IF KaryotypeResult = "47,XX,+21"CF 90</p> <p>OR KaryotypeResult = "47,XY,+21"CF 100</p> <p>THEN Diagnostic Disease IS Trisomy 21 Classic Type Syndrome</p> <p>Rule to identify disease of Trisomy 21 Variant Type</p> <p>IF KaryotypeResult = "47,XY,+21,inv(9)(p11;q12)"CF 70</p> <p>OR KaryotypeResult = "46,XY,-14,+t(14;21)"CF 80</p> <p>OR KaryotypeResult = "46,XY,-15,+t(15q;21q)"CF 70</p> <p>OR KaryotypeResult = "46,XY,-21,+t(21q;21q)"CF 50</p> <p>OR KaryotypeResult = "46,XY,-22,+t(21q;22q)"CF 50</p> <p>OR KaryotypeResult = "45,XX,-13,-21,+t(13q;21q)"CF 50</p> <p>THEN Diagnostic Disease IS Trisomy 21 Variant Type Syndrome</p>

생성 규칙은 IF 부의 조건절 또는 전제절, THEN 부의 귀결절로 이루어지는데 전제절이 만족되면 귀결절에서 정의된 행동(action)이 발생하여 결론이 도출된다.

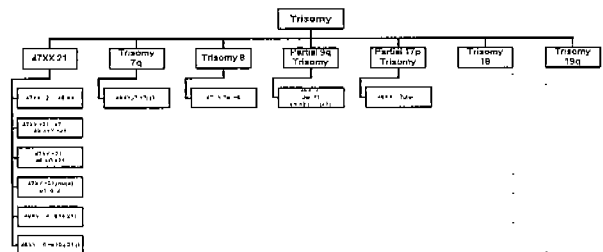
추론엔진은 주어진 정보집합이나 사실과 관련된 지식 선택을 제어하여 결론에 도달한다. 추론엔

진에는 사실에서 결론을 추론하는 전향 추론(forward chaining inference)과, 가설에서 알려진 사실로 전파되는 후향 추론(backward chaining inference) 및 이들이 결합된 혼합형 추론(hybrid chaining inference)이 있다. 사실이 규칙의 전제절에 사용될 때마다 규칙이 독립적으로 시작되기 때문에 하나 이상의 규칙을 동시에 시작하여 병렬추론을 할 수 있다(Zahedi F, 1993, Paul H et al. 1985)

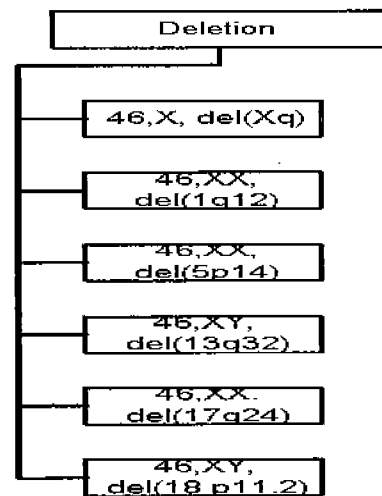
3.2 지식 베이스 시스템 모델

본 연구에 사용한 염색체들의 진단학적 소견을 기초하여 분류한 염색체들의 비정상은 수적인 이상과 구조적인 이상으로 분류 후 그림 3과 4에 예로써 나타내었다.

[Fig. 3. The classification of numerical abnormalities: Trisomy]

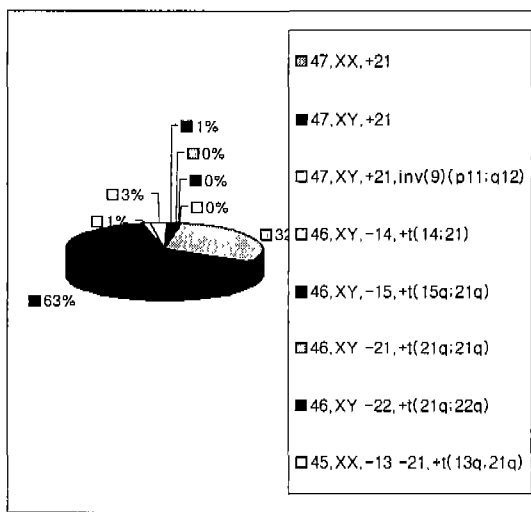


[Fig. 4. Classification of structural abnormalities: Deletion]



체계적 지식을 기반으로 축적된 경험적 지식을 사용하여 염색체를 분석하고 질병을 진단하는 전문가들은 분석된 염색체를 보고 수적인 이상 가운데 가장 발생빈도가 높은 다운 증후군 즉 Trisomy 21을 의심하는 경우가 있다. 이 Trisomy 21은 표준형(classic type)과 변이형(variant type)이 있는데, 표준형은 일반적인 일차 삼염색체성으로 남자나 여자의 정상 핵형에서 21번 염색체가 하나 증가한 47,XX 또는 XY,+21의 핵형을 보인다. 따라서 표준형은 두 가지의 경우만 고려해도 되지만 Trisomy 21 변이형을 의심하는 경우, 발생 가능한 세포유전학적 이상을 모두 고려하여야 한다.

[Fig. 5. Distribution for Trisomy 21 of the implemented system]



먼저, 21번 염색체가 세 개로 이루어져 47,XX 또는 XY,+G형으로 전체 염색체 수의 증가를 보 이면서 다른 염색체의 구조적 이상을 동반한 일차 삼염색체성의 경우가 있다.

다음으로, 21번 염색체의 전좌에 의해서 21번 염색체 장완이 다른 염색체에 부착되어 실질적으로 별개인 두 개의 21번 염색체가 되고 다른 염색 체에 전좌된 21번 염색체 장완 때문에 전체적으로

21번 염색체가 세 개의 효과를 가지는 경우가 있다. 즉 대표적인 이차 삼염색체성인 46,XX 또는 XY,-14,+t(14;21)를 보면, 염색체 개수는 46 개로 정상이나 14번 염색체가 하나 없는 대신에 구조적 인 이상으로 14번 염색체와 21번 염색체가 로버트 소니언 전좌를 일으키는 경우가 있다. 이외에도 이 차 삼염색체성에는 불균형 전좌의 경우로 46,XX 또는 XY, -15,+t(15q;21q), 46,XX 또는 XY,-21,+t(21q;21q),46,XX또는 XY,-22,+t(21q;22q) 등이 있고, 균형 전좌에는 45,XX 또는 XY,-13,-21,+t(13q;21q) 등이 있으며 모두 Trisomy 21 변이형으로 간주할 수 있다.

따라서 전문가가 Trisomy 21을 진단하고자 할 때는 발생 가능한 모든 경우 가운데 한 경우만 일 어나도 Trisomy 21 표준형 또는 변이형으로 진단 가능하므로 연산자 OR을 사용한 표 2와 같은 생 성 규칙이 구성된다. 이때 각 진단결과의 우선 순 위는 확실성의 정도를 이용하여 결정한다. 일반적 으로 확실성의 정도 표시는 주관적인 경우가 많지 만 본 연구에서는 그림 5와 같이 구축 시스템에 저장된 전체 Trisomy 21 표준형과 변이형에 대해 서 각각의 유병율(incidence)을 조사하여 유병율이 높은 순서에 따라 상대적인 비중(weight)을 두어 확신계수를 나타내었다. 이는 보다 신속하게 탐색 경로를 설정하여 추론과정을 수행할 수 있는 장점 을 갖는다.

4. 연구 결과

4.1. 실험 데이터

본 연구에 사용된 실험 데이터는 크게 두 가지 로 구분된다. 첫 번째는 염색체 분류의 벤치마크로 간주되는 CPR(Copenhagen Dataset)에서 1988년~

1990년에 핵형분류를 수행한 양수샘플 염색체를 받아서 사용하였다(Jennings AM, 1993, Errington PA et al. 1993)

두 번째는 국내병원들의 세포유전학 연구실에서 1988년~1998년까지 수집한 말초혈액과 양수샘플 염색체들을 분석하였다.

< Table 2. Distribution of the analyzed chromosomes >

	No. of Cases	%
Normal Chromosomes	2,736	91.4
Abnormal Chromosomes	259	8.6
Numerical Abnormalities	169	5.6
Structural Abnormalities	73	2.4
Combined Abnormalities	17	0.6
Total	2,995	100

4.2 구축도구

시스템 구축은 Windows NT서버와 Windows 98 클라이언트 환경에서 수행되었다. 핵형분류가 수행된 염색체 영상들은 HP ScanJet 3c/T에서 스캐닝 후, 분할처리를 거쳤다. 프로그래밍한 염색체 영상 처리 프로그램과 IMAQ Vision Builder를 사용해 추출한 염색체 특징들은 Microsoft Excel 2000, SPSS for Windows, Microsoft ACCESS 데이터베이스 2000에서 작업하여 SQL 서버로 이식하였다.

지식 베이스는 Visual Rule Studio 1.5에서 작성하였고, 전체적인 프로그램은 Microsoft Visual Studio에서 완성하였다.

4.3 교육용 염색체 분석 시스템의 구축

교육용 염색체 분석 시스템은 다음과 같이 구축되었다.

[Fig 6. Detail information tab for implemented system]

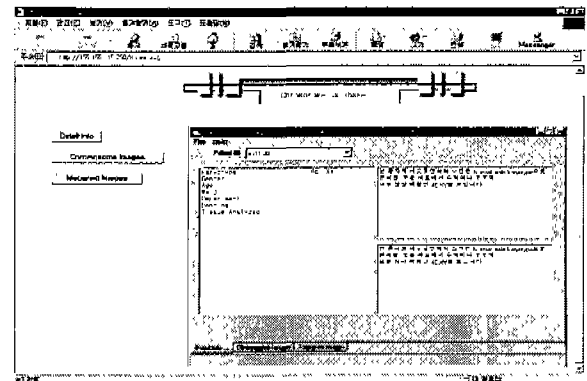


그림 6의 좌측에 있는 상세 정보 단추를 선택하고 환자 ID를 클릭하면 선택된 환자의 성별, 나이, 병실, 의뢰부서, 분염방법, 분석조직과 같은 하위항목과 그 결과가 좌측에 나타나고, 우측에는 염색체의 진단결과와 진단에 대한 부연설명이 나타난다.

[Fig. 7. Ideogram images tab for implemented system]

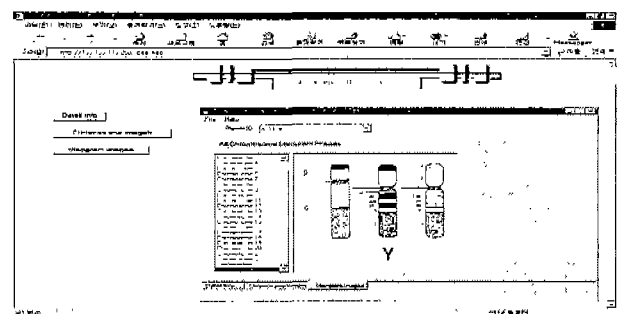


그림 7과 같이 환자의 염색체 영상이 필요할 때는 염색체 영상 탭을 클릭하는데, 사용자의 요구에 따라 각 염색체 영상의 확대된 화면을 볼 수 있고, 각 염색체를 더블 클릭하면 표준화된 이디오그램 영상들을 볼 수 있다.

[Fig 8. Chromosome images tab for implemented system]

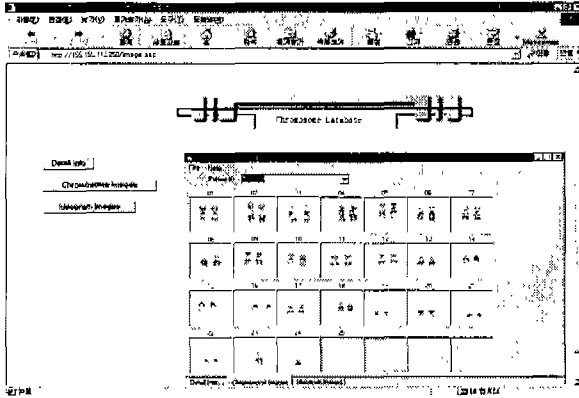


그림 8에 보이는 이디오그램 영상 탭을 클릭하면 환자의 실제 염색체와 각 염색체에 해당하는 표준화된 이디오그램 영상을 서로 비교할 수 있다.

[Fig. 9. Handbook composed of ISCN 1995 and chromosome abnormal diagnosis]

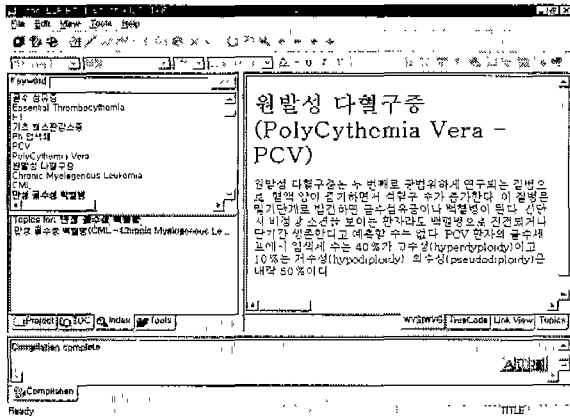


그림 9에서 보듯이 사용자가 프로그램을 사용하다가 ISCN 명명법이나 염색체의 분류와 진단에 관련된 의문사항이 생길 경우에는 제공되는 ISCN 1995와 염색체 이상 진단 핸드북의 정보내용을 직접 검색하거나, 인덱싱 키워드를 좌측상단에서 클릭하여 참조한다.

5. 결론

지식 베이스를 이용한 염색체 분석 시스템은 염색체를 분류하고 진단할 때 정보간의 충돌이 있거나 불완전한 정보인 경우에도 지식 베이스에서의 추론을 통해 결론을 추출하여 정상 염색체와 비정상 염색체의 다양한 형태학적 정보를 제공하였다. 뿐만 아니라 기존의 시스템에서는 이루어지지 않았던 분석결과를 이용한 진단 및 이에 대한 설명은 사용자와 시스템간의 상호작용을 통하여 협의 시스템(consultation system)의 역할을 담당하였다.

향후 새로운 염색체에 적용하기 위해서는 삼배체나 사배체와 같은 희귀 증례를 포함하고, 전문가의 지식이 자동으로 획득되어야 증례에 민감하고 변경과 진화가 용이한 지식 베이스 시스템이 구성될 것이다.

참고문헌

1. J Margaret, Barch, MS, CLSp(CG), *The ACT Cytogenetics Laboratory Manual* 2nd ed. Raven press, New York, 1991.
2. Mitelman Felix, *ISCN 1995 An International System for Human Cytogenetic Nomenclature(1995)*, Springer Karger, 1995.
3. 김한화, *기초의학유전학* 제2판 정문각, 1998.
4. 김도현, 박준영, "세포유전학의 임상적용", *을지 의보*, 16권 2호(1993), 3~13.

5. 박숙자, “임상영역에서의 염색체분석의 의의”, *부산 소아과학회지*, 5권 1호(1992), 87~97.
6. *Cytovision user manual*, Applied Imaging, U.K., 1994.
7. *PowerGene user manual*, PSI, U.S.A., 1995.
8. *Ikarus user manual*, Meta systems, Germany.
9. S Jake, *Professional VB UML-design and development*. WROX, 1999.
10. F Martin, Kendall S, *UML Distilled*. Addison Wesley Longman, Inc., 1997.
11. F Zahedi, *Intelligent system for business: expert systems with neural networks* 1st ed. Wadsworth, Inc., America, 1993.
12. H Paul, David K, *Expert systems. Artificial intelligence in business*. John Wiley & Sons, Inc., New York, 1985.
13. AM Jennings, Graham Jim, “A NN approach to automatic chromosome classification”, *Phys Med Bio*, Vol.38(1993), 959~970.
14. PA Errington, Graham J, “Application of artificial NN to chromosome classification”, *Cytometry*, , Vol.14(1993), 627~639.