삼성 SDS의 Bioinformatics: 사업 및 연구/개발

정태수 박사

삼성SDS 정보기술 연구소 Bioinformatics Lab. 연구원

E-mail: epiai@samsung.co.kr

Abstract

- Overview of Bioinformatics and vision of Samsung SDS on it
- Overview of Bio Chip and its market
- Product roadmap with "Expert system for DNA chip data"
- "UniBIO" as an integrated package of DNA chip data analysis
- Demo of UniBIO

CV

2000년-현재: 삼성SDS 정보기술 연구소 Bioinformatics Lab. 연구원

2000년-2000년: 고려대학교 수학과 BK21(금융수학)

1998년-2000년: 서강대학교 강사(수학과, 전산과, 경영학과)

1997년-1998년: 일본 나고야 메이조대학교 Post-Doc(확률론)

1990년-1997년: 서강대학교 이학박사(확률.통계학)

1990년-1993년: 육군사관학교 수학과 전임강사

1988년-1990년: 서강대학교 이학석사(수학과)

1984년-1988년: 서강대학교 이학사(수학과)

삼성SDS의 Bioinformatics: 사업 및 연구/개발

삼성SDS
정보기술연구소
Bioinformatics Lab.
정 태 수
epiai@samsung.co.kr

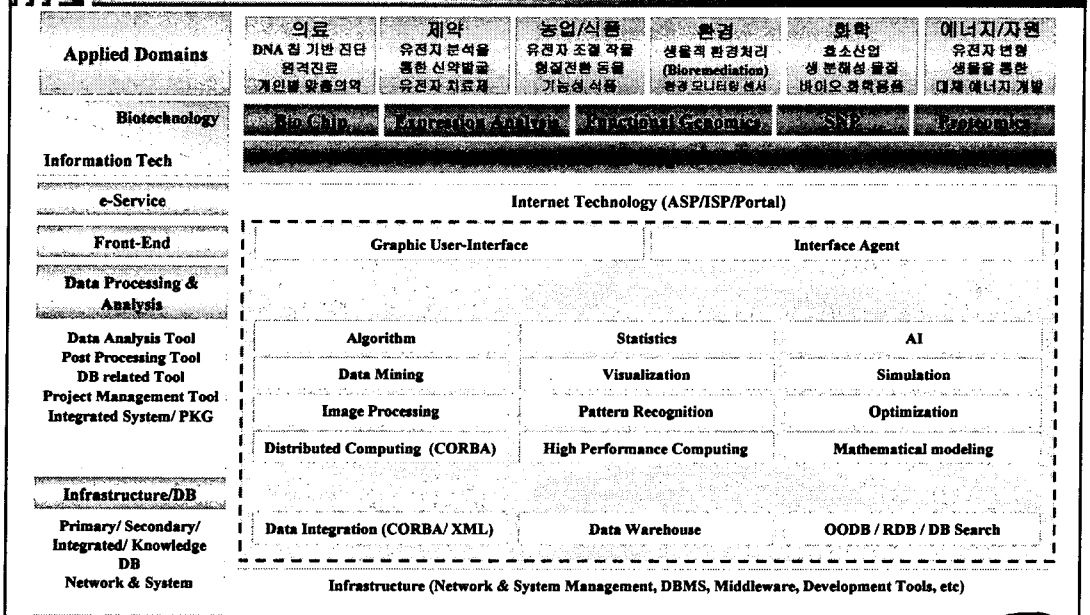
발표 순서

- Overview of Bioinformatics
- Bio Chip
- Expert system for DNA chip
- UniBIO
- Demo

Overview of Bioinformatics

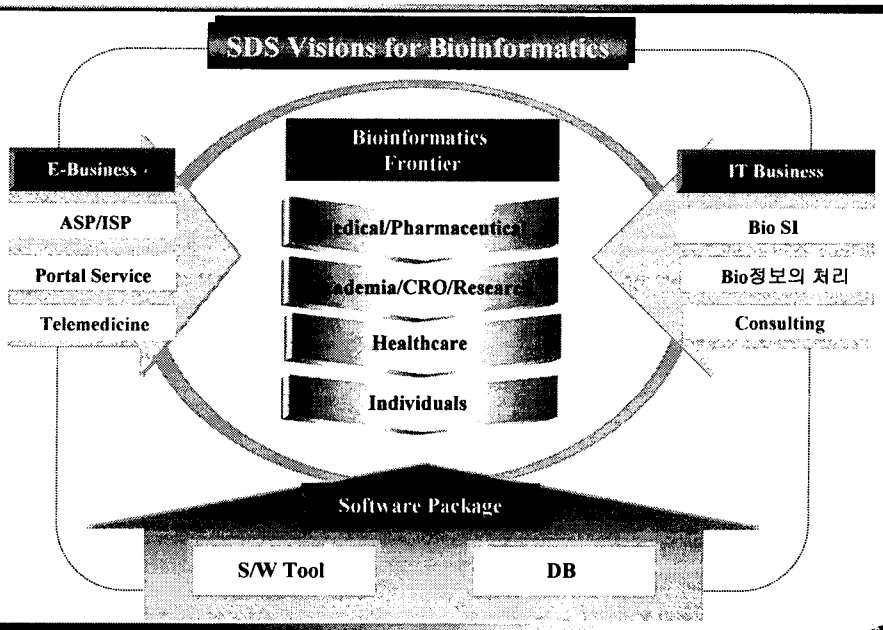
SAMSUNG

Overview of Bioinformatics



SAMSUNG

Vision of Samsung SDS on Bioinformatics



Vision of Major IT companies on Bioinformatics

Company	Viewpoint of BT	Action
IBM	We believe this will be a **very hot opportunity for IT** over the next decades	▪ **Customer: First Genetic Trust** ▪ **Life Science Div 설립** ▪ **Blue Gene 개발** ▪ **Bio consulting 사업 시작**
SUN	BT growth will more than replace any loss of dot-com revenue, especially, **the growth of computational resources and storage.** Needs will be very significant	▪ **Strategy: Deep discounts to Univ** ▪ **영국에 대규모 data center 설립** ▪ **Port DeCypher(TimeLogic)to Solaris OS** ▪ **Promote open platform on life-science based on *Java and XML***
Silicon Graphics	science research portion of SGI is 1/5 of all and **protein analysis** will produce much more data than we have today	▪ **3D visualization & data application tools**
Compaq	Major driving forces: ❑ IT investments on R&D by big pharmaceuticals and biotech companies ❑ Brand-new industry to supply information to pharmaceuticals and biotech companies	▪ **Celera Genomics/Sandia National Lab과 R&D 협력** ▪ **Invest $100million on genomics & bioinformatics company via venture capital** ▪ **Invest $10million in Geneva Proteomics**

Bio Chip

SAMSUNG

Bio Chip Market

Segmentation	1999	2001	2004	Major Trends
Microarray	$232mill	$874mill	$2.6bill	Patent issue/Consolidation/ Acquisition/Higher density array/ Lab with their own microarray
Microfluidics	$17mill	$77mill	$305mill	Alliance with liquid handling (Aclara & Packard/ Gamera & Tecam)
Other Bio chip	$23mill	$81mill	$311mill	Protein chip(Ciphergen) NanoChip(Nanogen) SNP Detection(Sequonome) Probes(Clinical Micro Sensors)
Bio Chip	$272mill	$1.03bill	$3.3bill	**65% AGR (Annual Growth rate)**

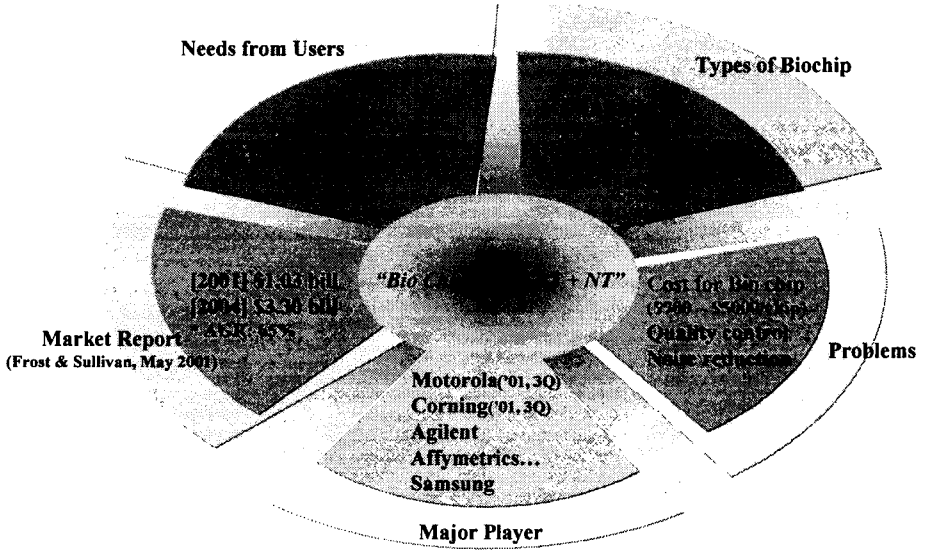
Frost & Sullivan, May, 2001

SAMSUNG

Major Bio Chip Player (Motorola)

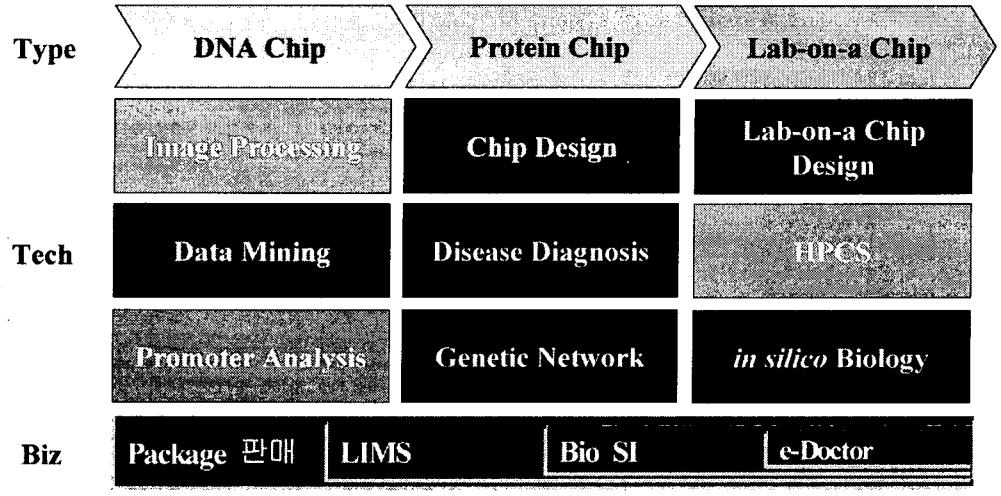
Motorola's Life Science	
Product Goal	Type: *Universal platform* for DNA-based testing "full value-chain from research to discovery to trials to diagnostics with different products targeted to different types of needs"
Target customers	Researchers/Diagnostics lab/*Large* doctor's office
Selling strategy	Mass production with *high costs*. but emphasis on *quality of chip*
Products	- CodeLink array for researchers with 10k human gene microarray (August)-mass production - Later: 10k *rat array(Oligo chip)*
Collaborator	**Compugen**: chip design service
Recent PJT	[DARPA] Microfluidics multi-chip-module genetic analysis sample preparation system

Overview of Bio Chip

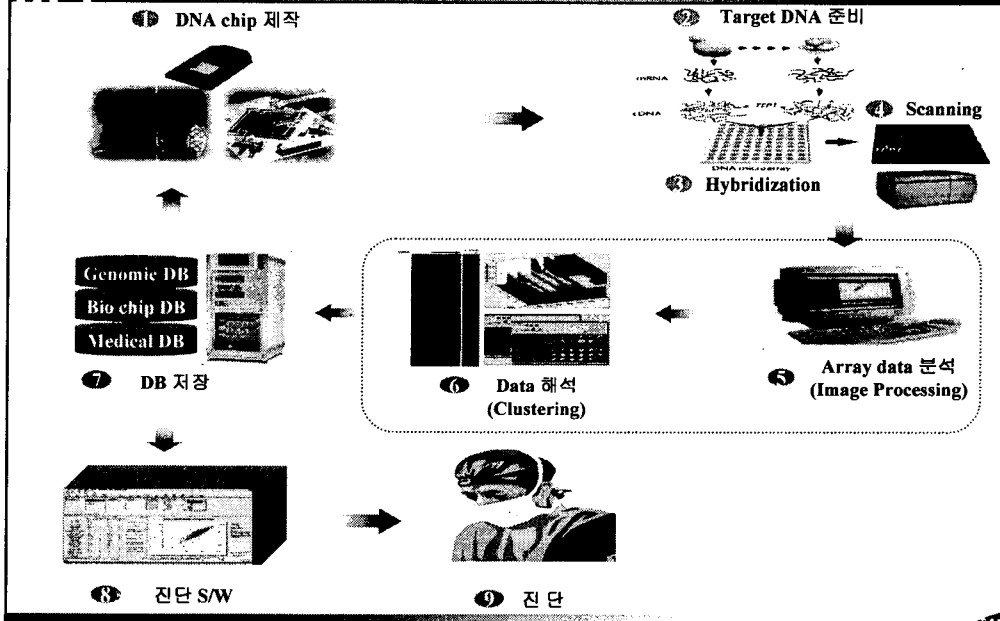


Expert System for Bio Chip

Tech. & Product Roadmap



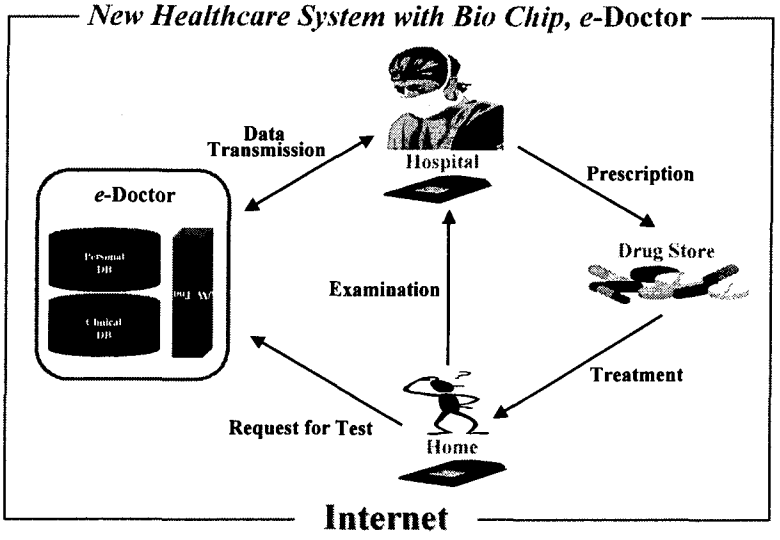
An application of DNA chip Analysis system



Microarray system

Chip Design	☐ **Problem: Which genes to put on the chip and Which probesused for each gene** ☐ **Process:** **- (cDNA) selecting a set of clones** **- (Oligo) design oligos for the regions of each gene you want to cover** ☐ **Informatics:** *AIMS/Heavy-duty sequence-analysis problem*
Chip Construction	☐ **(Input) chip layout (Output) A list of chip identifiers***(barcode)* ☐ **Process: clone or oligo picking → spotting on plates** ☐ **Informatics:** *LIMS + inventory and order management* **(Bookkeeping of many clones and oligos bought from supplying company)**
Data Acquisition	☐ **(Input) sample and chip** ☐ **Process: RNA extraction, labeling, hybridization, image capture** ☐ **Additional Information:** **- chip data from chip construction** **- sample data from sample acquisition LIMS***(manually)* ☐ *Wet phase*
Data Analysis	☐ **Image Processing** ☐ **Data Mining** ☐ **Informatics:** *AIMS*

e-Doctor: Healthcare system with Bio Chip



UniBIO

Integrated package for DNA chip data analysis

UniBIO의 특징

- Full integrated system for DNA chip data analysis
- Project management
- Automatic procedure with minimal user input
- Various file import/export
- Web interface (GenBank)
- User-friendly interface
- Batch process/Wizard
- JAVA and XML

UniBIO의 기능 (1)

Image Processing	Spot detection	❖ Manual segmentation ❖ Automatic segmentation
	Extract expression	Quantification Quality control
	Normalization	❖ Using control genes ❖ Inter-experiment normalization
	Chip quality validation	

UniBIO의 기능 (2)

Statistical Analysis	Quality control	❖ Filtering outliers ❖ Normalization procedures
	Test of differential expression	❖ Significance of test ❖ False-positive error control
	Cost/Benefit analysis	❖ Estimates number of replicates needed ❖ False-negative error control

UniBIO의 기능 (3)

Data Mining	Visualization	❖ Scatter plot ❖ Pattern graph ❖ Dendrogram
	PCA	Reduce number of dimension
	Data clustering	❖ K-Means ❖ SOM ❖ Hierarchical algorithm ❖ Bi-clustering
	Classification	❖ Classification ❖ Estimation ❖ Prediction ❖ Regression

UniBIO의 기능 (4)

Promoter Analysis	Promoter sequence extractor	❖ GenBank explorer ❖ Promoter extractor
	Promoter sequence analyzer	TF site search Motif edit TF site discovery TF site validation Promoter module search Promoter module discovery Promoter prediction

DEMO