

# 바이오 빅데이터를 활용한 원예 육종 및 식품 산업 연계

(주)인실리코젠 강병철 기술이사 \_ 2018. 10.

INSILICOGEN

[www.insilicogen.com](http://www.insilicogen.com)

주인실리코젠

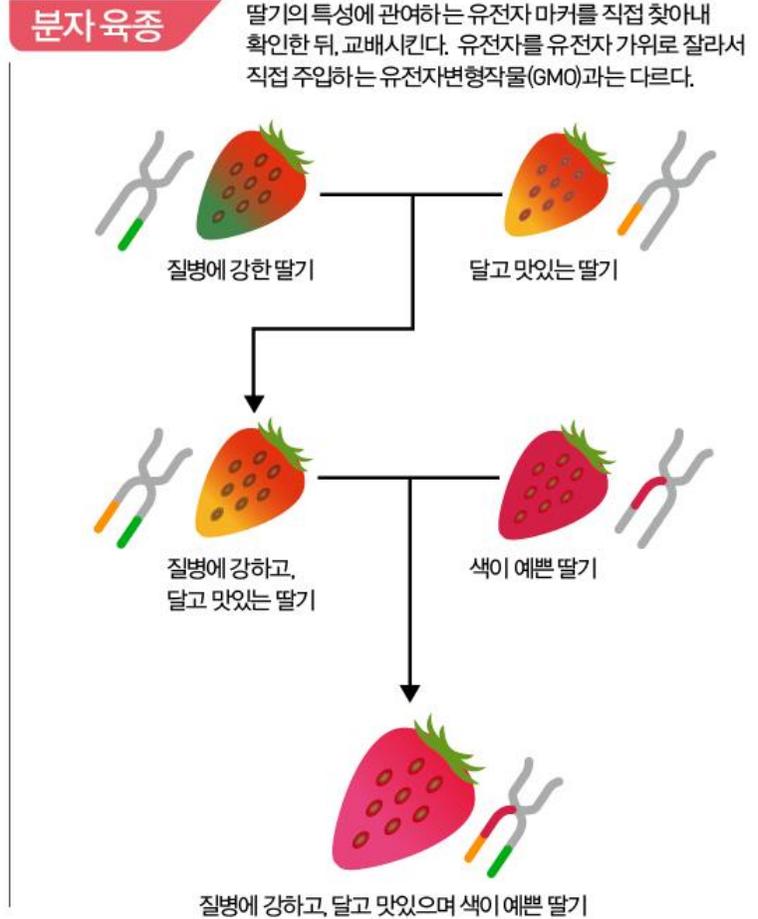
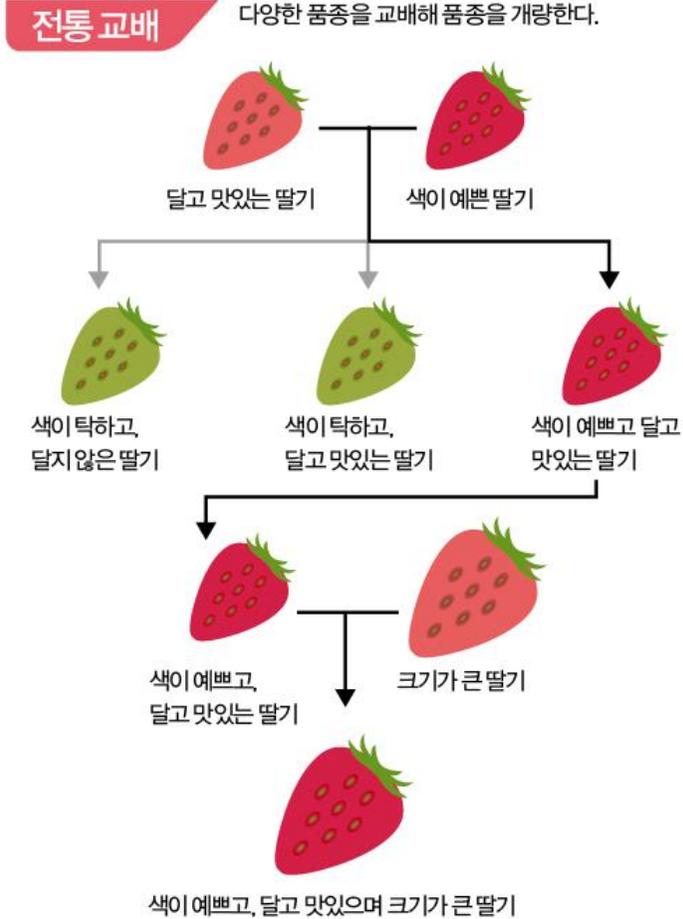
본 문서의 모든 콘텐츠는 저작권법의 보호를 받는 저작물로 별도의 저작권 표시 또는 다른 출처를 명시한 경우를 제외하고는 (주)인실리코젠에 저작권이 있습니다.  
저작권 표시 또는 기타 소유권 표시를 삭제해서도 안되며, 당사와의 협의 또는 허락없이 무단 복제, 변경, 배포를 금지합니다.  
저작권 관련 문의사항이 있으시면 [bc@insilicogen.com](mailto:bc@insilicogen.com)으로 연락바랍니다.  
© 2018 **INSILICOGEN, INC.** ALL RIGHTS RESERVED.

# Index

- 1 전통 원예 육종 및 분자 육종
- 2 기계 학습 및 빅데이터를 활용한 원예 정밀 육종
- 3 바이오 빅데이터를 활용한 식품 산업 연계

# 1. 전통 원예 육종 및 분자 육종

# 전통 육종 vs 분자 육종

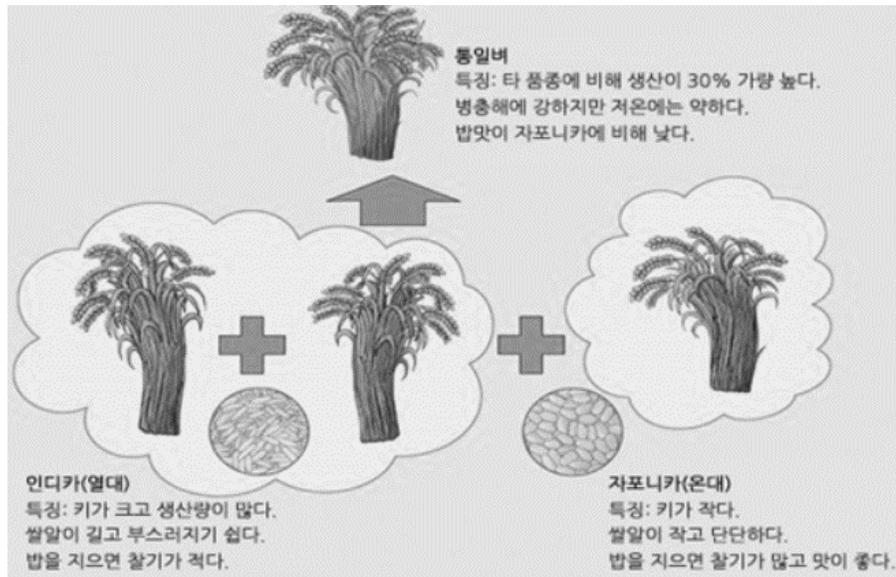


Ref. 과학동아 2018년 06호

# 전통 육종 vs 분자 육종

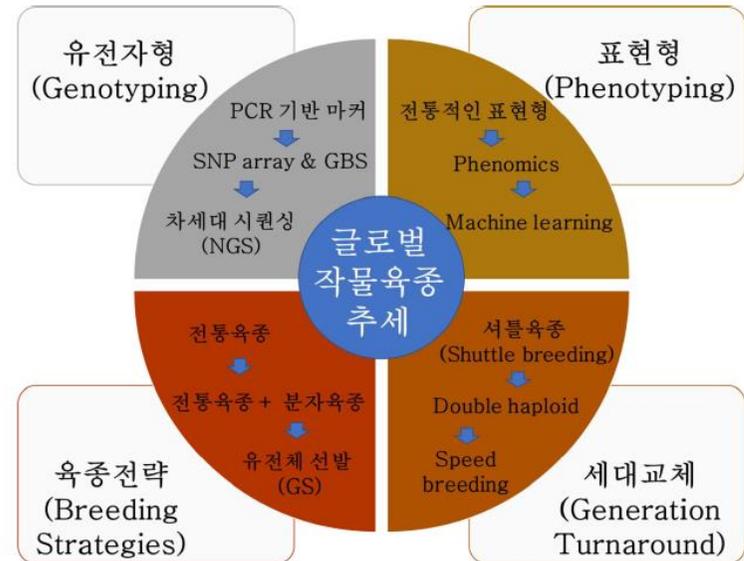
## 전통 육종

- 같은 종 내에서 다른 형질을 갖는 품종을 찾아 서로 유전자를 교환하도록 교배



## 분자 육종

- 유전체학의 발전과 기술의 발달로 바이오 마커, 유전자형 표현형을 기반으로 하는 분자 육종 기술 현황



Ref. bric view 2018, 이성희

# 동일 품종(고추)의 영양성분 맛의 차이점



## 고추

학명 : *Capsicum annuum*

주요 용도 : 매운맛을 활용한 향신료

킬로줄	167 kJ	칼로리	40 kcal
탄수화물	8.81 g	콜레스테롤	0 mg
설탕당	5.3 g	식이섬유	1.5 g
단백질	1.87 g	나트륨	9 mg
총 지방 함량	0.44 g	<b>칼륨</b>	<b>322 m</b>
포화지방	0.042 g	<b>스코빌 지수 4K - 12K SHU</b>	
다불포화지방	0.239 g	(청양고추 기준)	
불포화 지방	0.024 g		



## 파프리카

학명 : *Capsicum annuum*

주요 용도 : 단맛을 활용한 채소류 요리

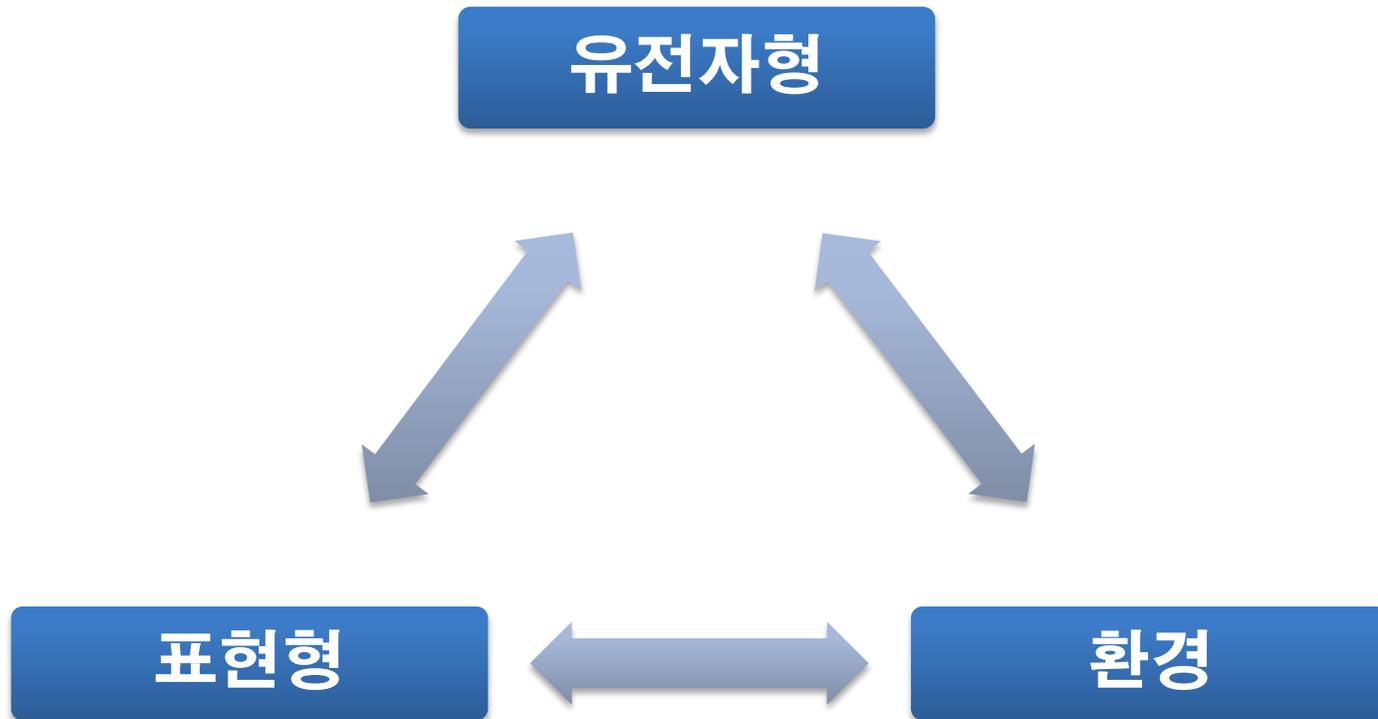
킬로줄	63 kJ	칼로리	15 kcal
탄수화물	3.43 g	콜레스테롤	0 mg
설탕당	1.78 g	식이섬유	1.3 g
단백질	0.64 g	나트륨	2 mg
총 지방 함량	0.13 g	<b>칼륨</b>	<b>130 mg</b>
포화지방	0.043 g	<b>스코빌 지수</b>	<b>0 SHU</b>
다불포화지방	0.046 g		
불포화 지방	0.006 g		

- 파프리카는 고추에 품종 이지만 주요 영양성분과 매운 맛 척도가 다름
- 스코빌 지수 : 매운 맛을 나타내는 척도

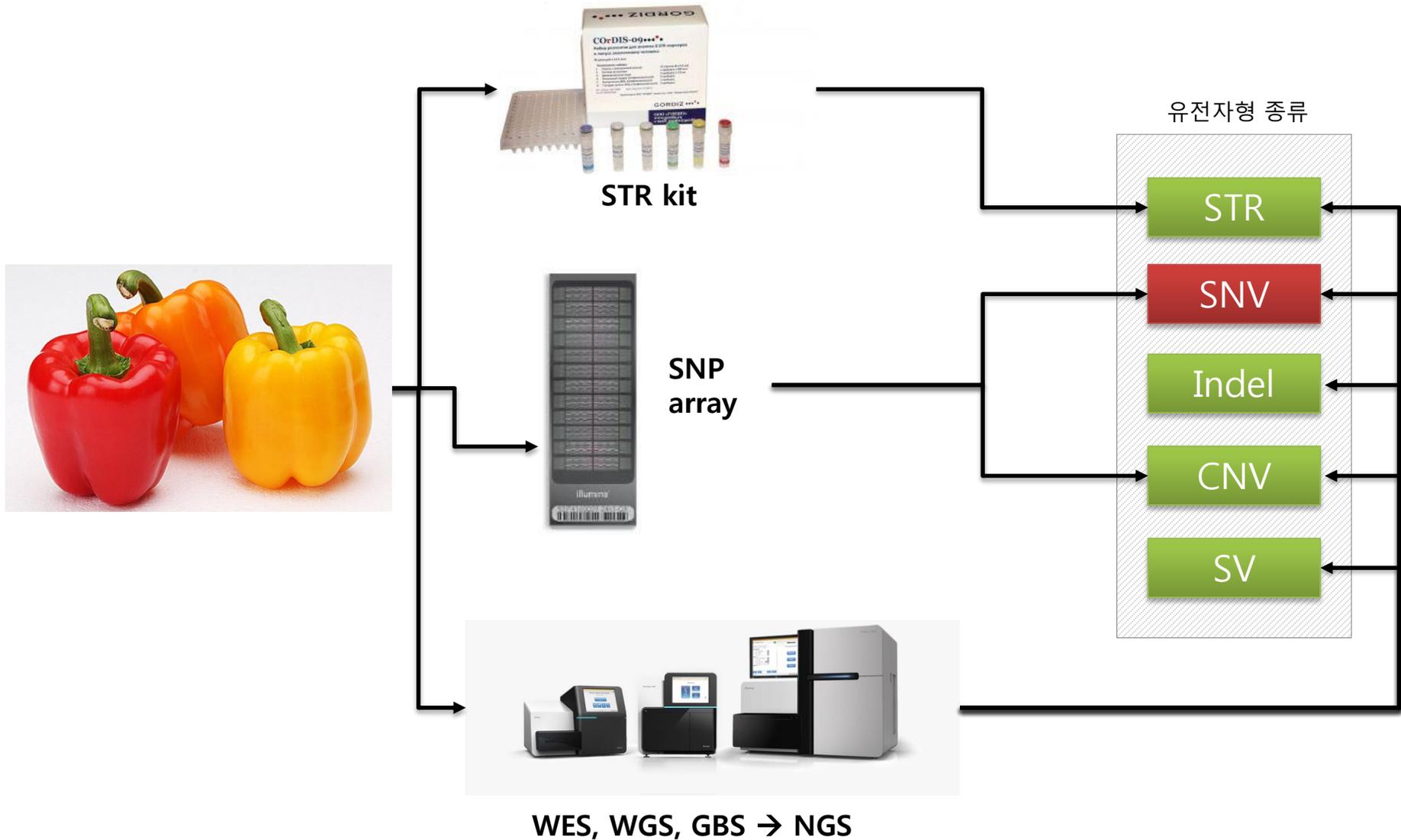
# 유전자형 - 표현형 - 환경

$$P = G + E$$

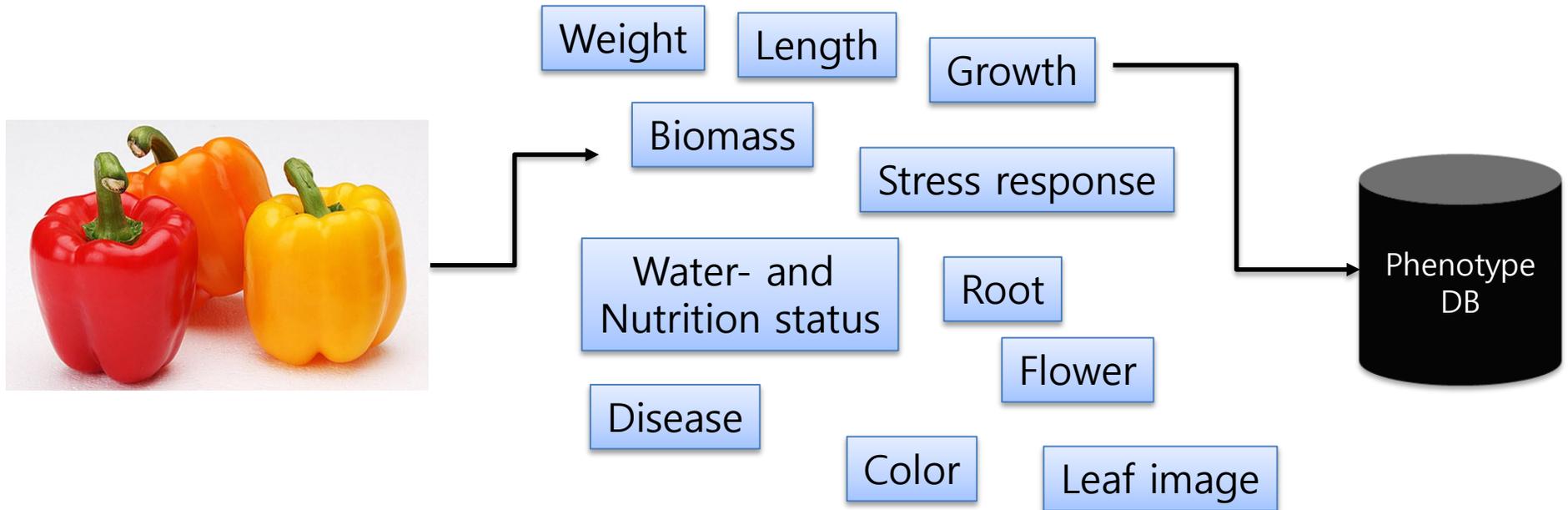
phenotypic value      genotypic value      environmental deviation



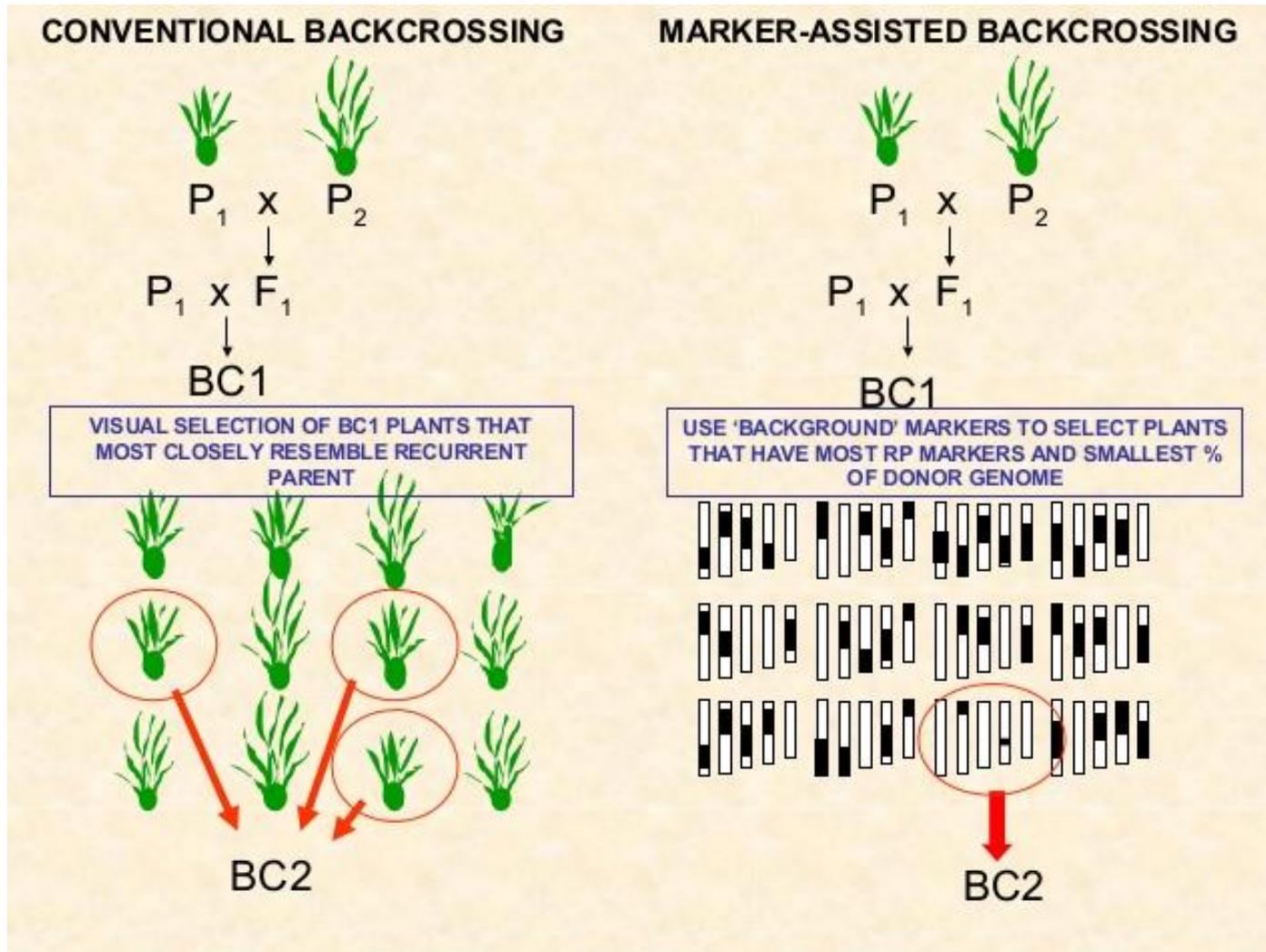
# 유전자형 정보



# 표현형 정보



# MAS (Marker assisted selection)



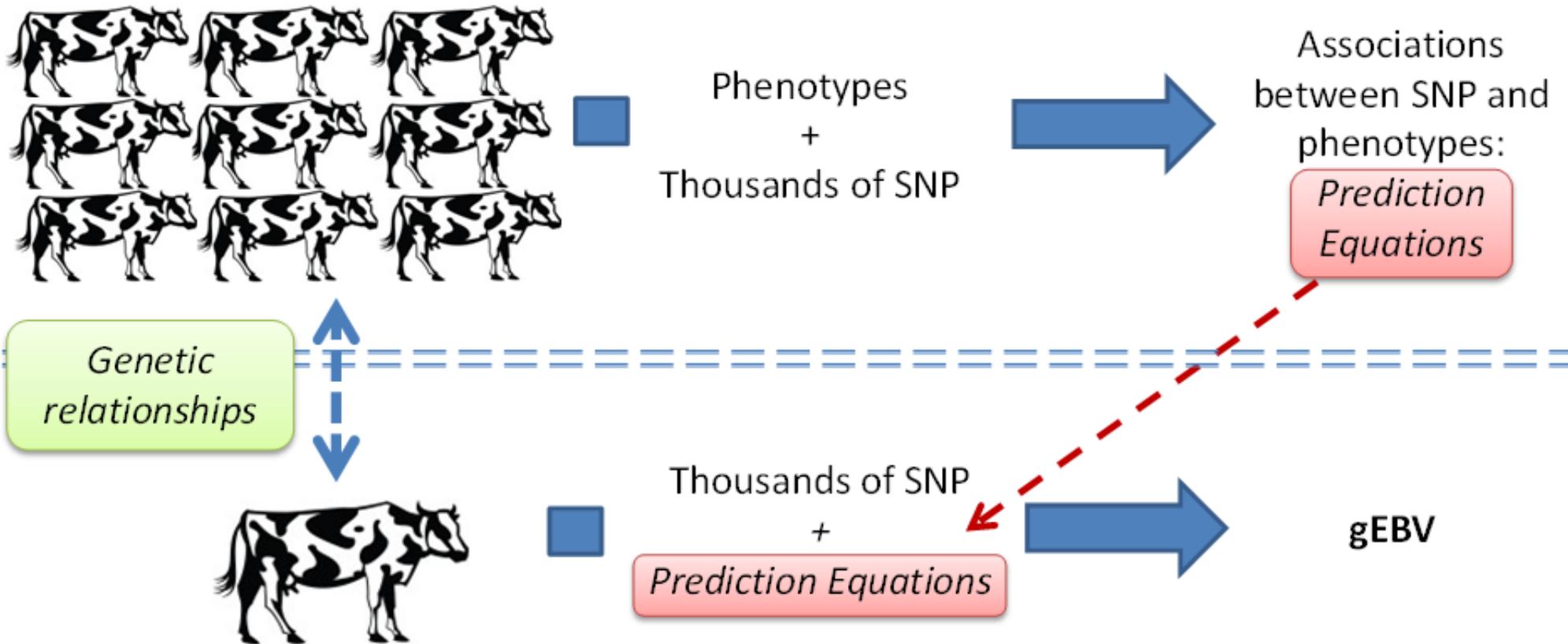
MAS uses a few loci.

But, usual phenotype is polygenic.

GS (Genomic selection) estimates all marker effects.

# GS (Genomic selection)

Reference population: Development of prediction equations



Main population: Application of prediction equations

# Best linear unbiased prediction

## BLUP

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A^{-1} \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix}$$

A-matrix (pedigree-based)

1	0.5	0.5	0	0.5
0.5	1	0.25	0	0.25
0.5	0.25	1	0	0.25
0	0	0	1	0
0.5	0.25	0.25	0	1

## gBLUP

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda G^{-1} \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix}$$

G-matrix (SNP-based)

1	0.5	0.5	0.02	0.5
0.5	1	0.2	0.015	0.2
0.5	0.2	1	0.025	0.3
0.02	0.015	0.025	1	0.025
0.5	0.2	0.3	0.025	1

Other methods:

snpBLUP, Lasso, ridge regression

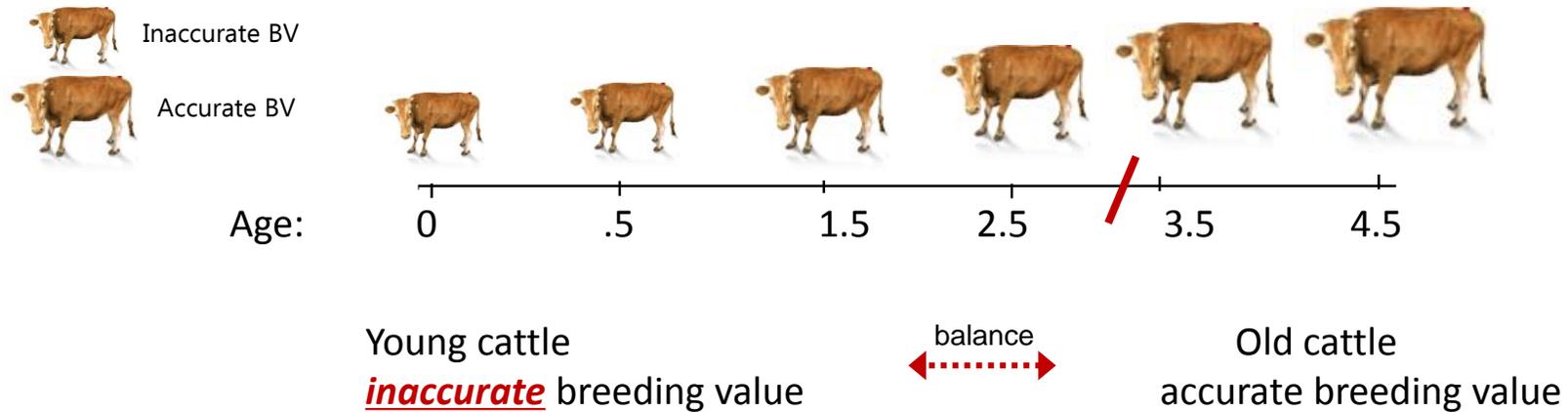
Bayesian – Bayes A, B, C, Cpi, R

support vector machines, semi-parametric kernel regressions

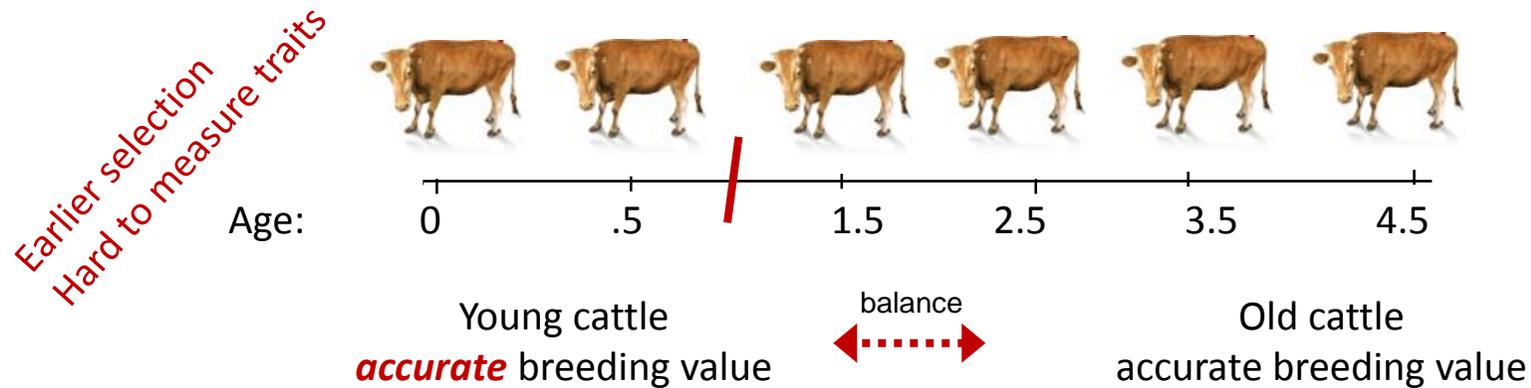
reproducing kernel Hilbert space regression

Single step

# Progeny test and genomic selection



## Genomic Selection

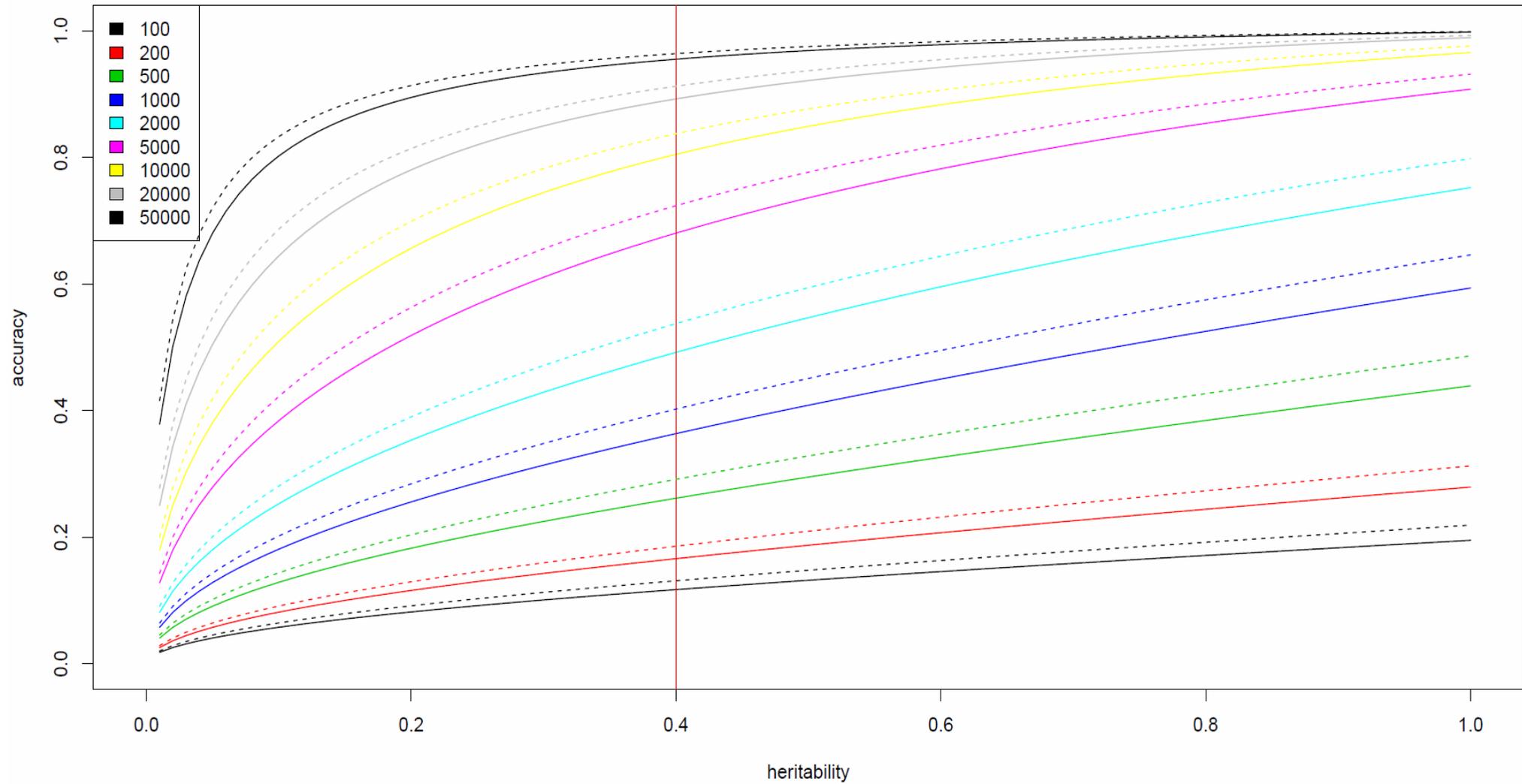


$$G = \frac{\text{Intensity} * \text{accuracy} * h^2}{\text{generation interval}}$$

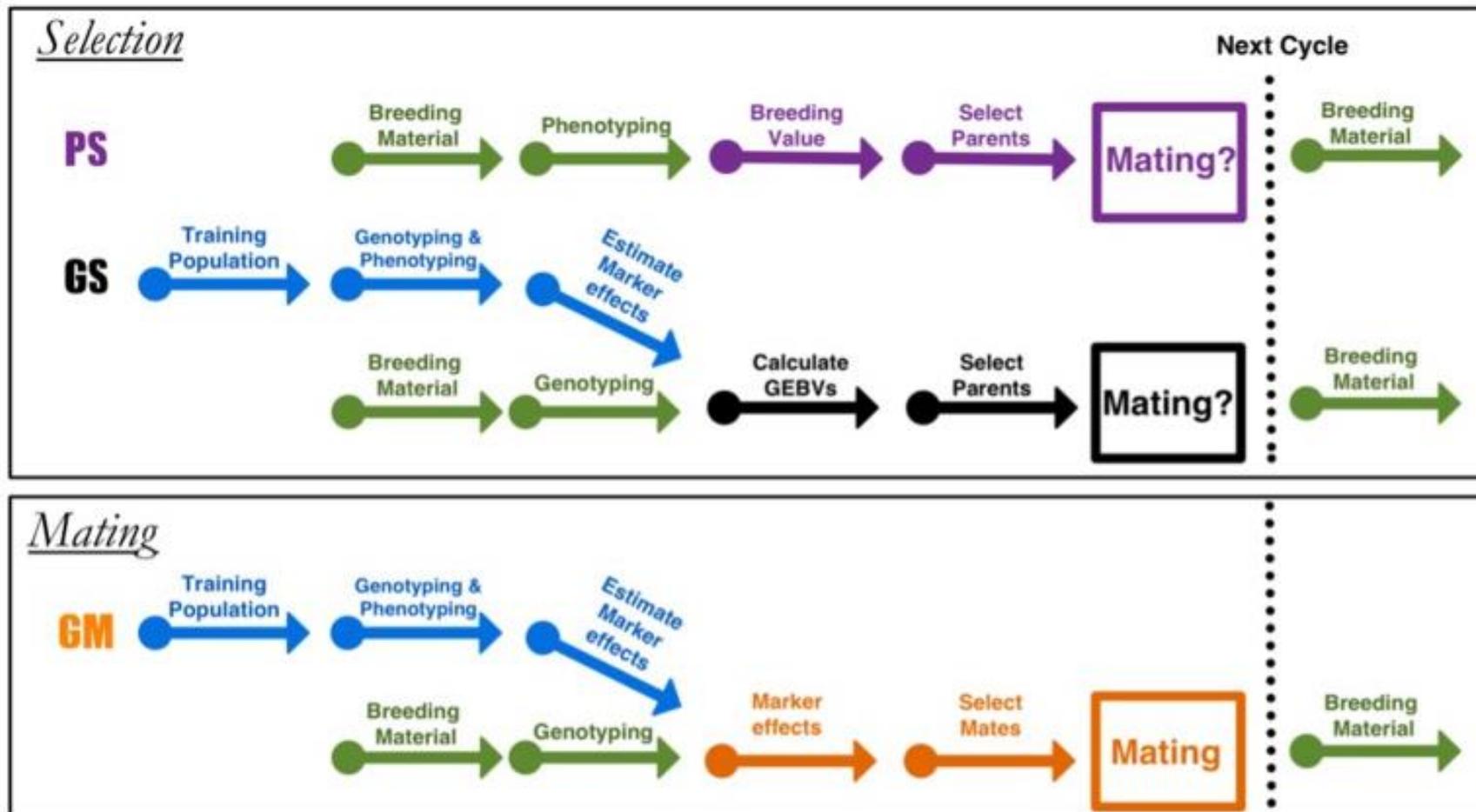
# Accuracy of genomic selection

accuracy X heritability for different number of individuals genotyped

solid line - 50K  
dashed line - 700k



# GM (Genomic mating)



**Diagram for the different breeding approaches.** Phenotypic selection (PS) and genomic selection (GS) are truncation selection methods, and genomic mating (GM) is the mating approach. Arrows indicate the different stages in a breeding cycle. In PS, starting with a set of parents as breeding material, selection is performed based on phenotypes.

**01 WEB FOLDER**

Store all kinds of your own data into our **WEB FOLDER**

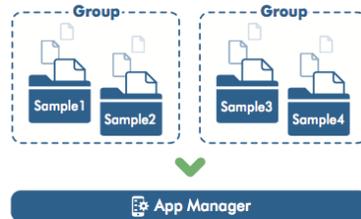
You can upload a variety of data file such as genotypes (.ped, .xlsx, vcf), phenotypes, environments, and parentage to your web folder. It will allow you to create directories, upload, rename, and delete files.



**02 DATASETS**

Manage your data by creating suitable **DATASETS**

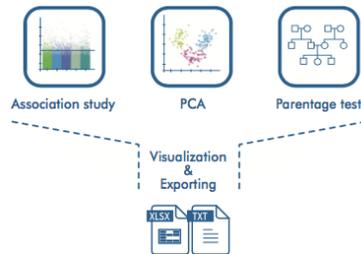
You can create your genotype, phenotype, environment, and parentage data into datasets. You can also make subset groups like case and control, and various phenotypes, so you can use them to analyze in later step.



**03 APPS**

Analyze and interpret the meaning of your data from our various **APPS**

IncoGWAS provides a variety of analyzing modules in forms of app. Each apps performs specific analysis for selected datasets and visualizes the results on the web interface. Researchers can also share the results of apps in public.



**Data Accumulation**

Are you planning to perform multiple experiments for your project?

You can merge multiple dataset for one big analysis so that you can truly find the meaning of the big data. You can also import and merge public data.



Manage your experimental results in "Dataset" units

**Convenient Web Folder**

Don't you have enough space for **GWAS analysis?**

Once, you store your large data online, you don't have to worry about disc space and where to analyze because IncoGWAS is available anywhere as long as internet is available.

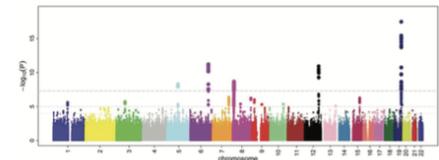


Store any file and see your data anywhere

**GWAS Analysis with Web Interface**

Are you not familiar with programming?

Our IncoGWAS comes with user-friendly web interface, which will help you to find the meaning of your data without worrying about learning complex programming languages like R, PLINK, and so on.

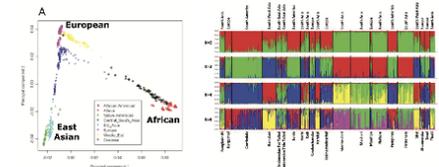


General visualization of GWAS - Manhattan plot

**Various Statistics**

Do you want to apply various statistics to your data?

Our IncoGWAS comes with various apps for many kinds of statistical analysis from allele frequency calculation to PCA, and more importantly, they will visualize your data for your better understanding.

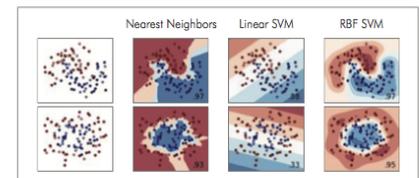


Examples of population analysis by PCA and structure

**Breeding Value Estimation**

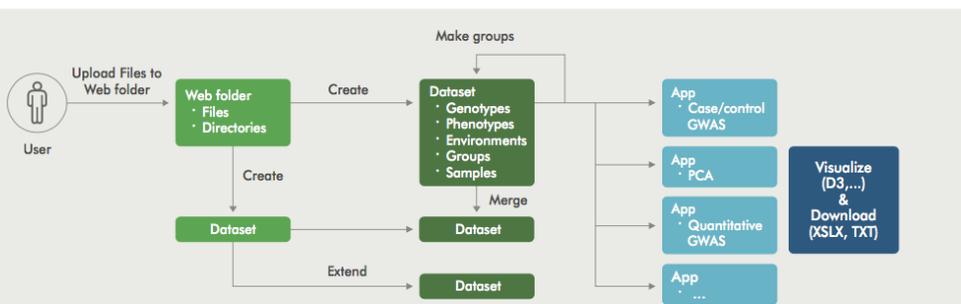
Are you looking for a platform for selective breeding?

With both genotype and phenotype data, IncoGWAS can predict phenotype from cutting edge machine-learning algorithm. Its breeding value estimation via genomic best linear unbiased prediction will help you to find suitable candidates for you.



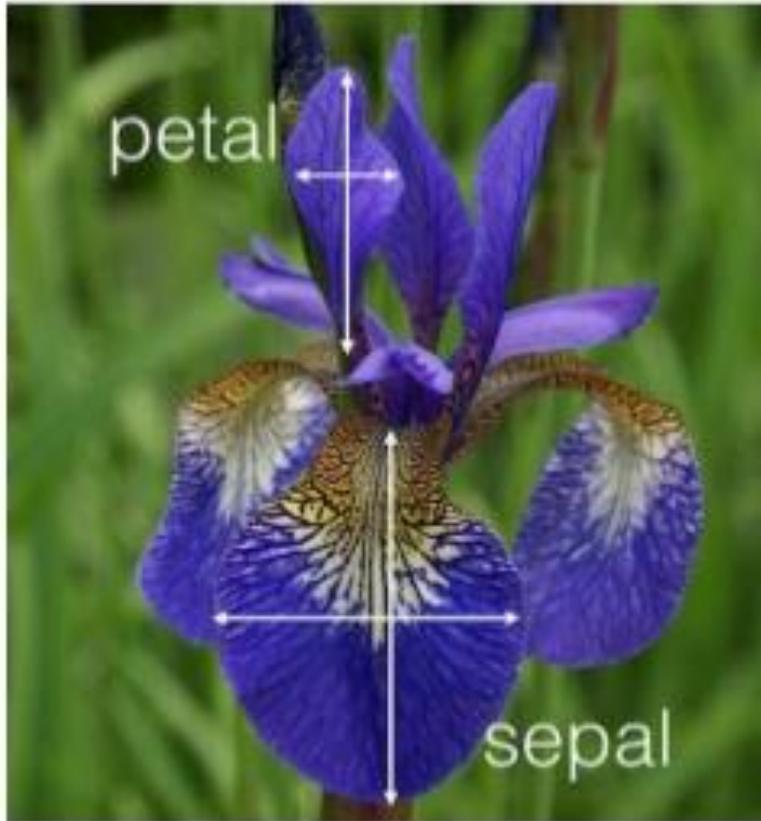
Example of classification using various machine learning algorithms

Process of IncoGWAS from data upload to analysis



## 2. 기계 학습 및 빅데이터를 활용한 원예 정밀 육종

# Supervised learning *classification* problem (using the [Iris flower data set](#))

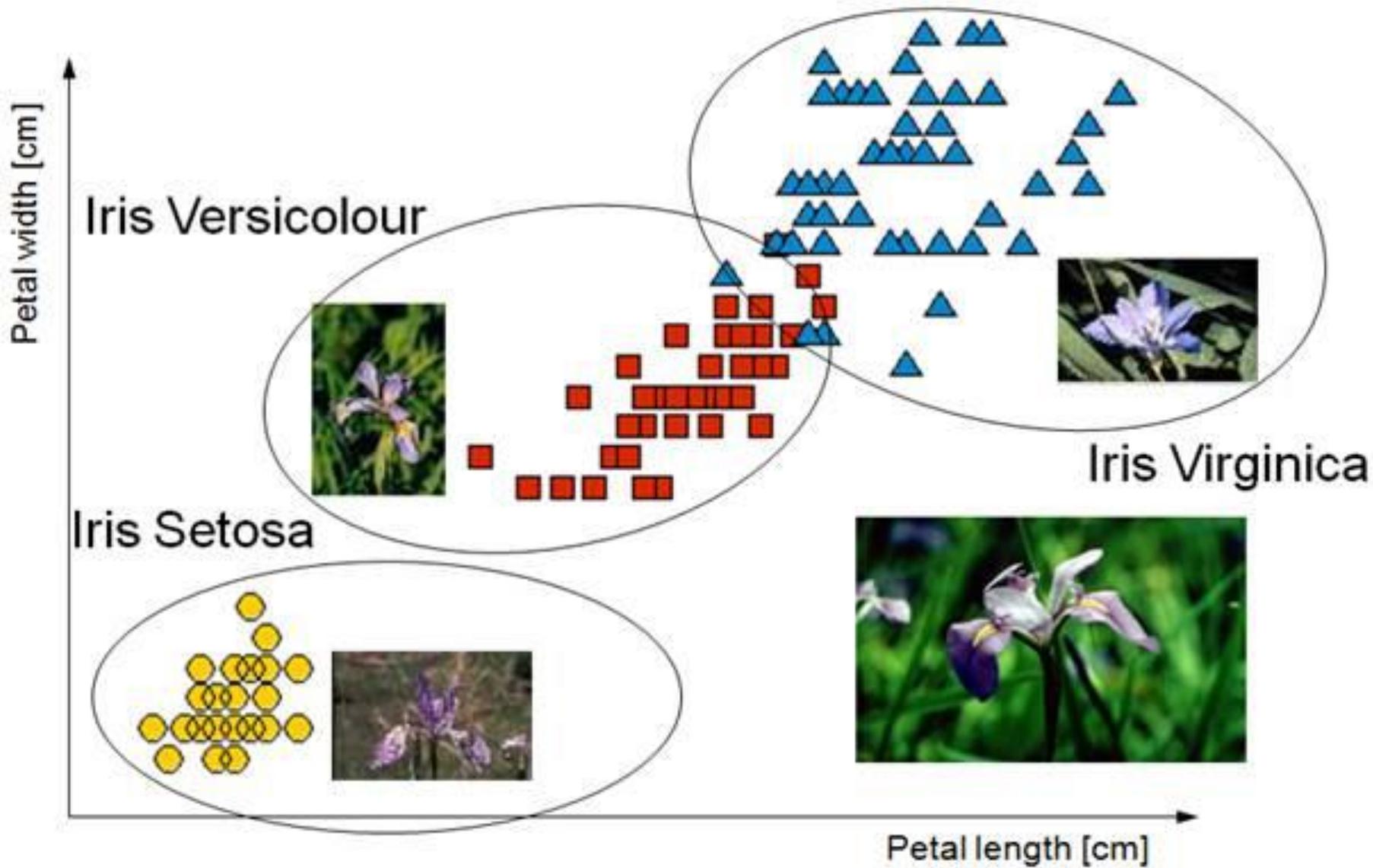


## Training / test data

### Features

### Labels

Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	Iris setosa
4.9	3.0	1.4	0.2	Iris setosa
7.0	3.2	4.7	1.4	Iris versicolor
6.4	3.2	4.5	1.5	Iris versicolor
6.3	3.3	6.0	2.5	Iris virginica
5.8	3.3	6.0	2.5	Iris virginica



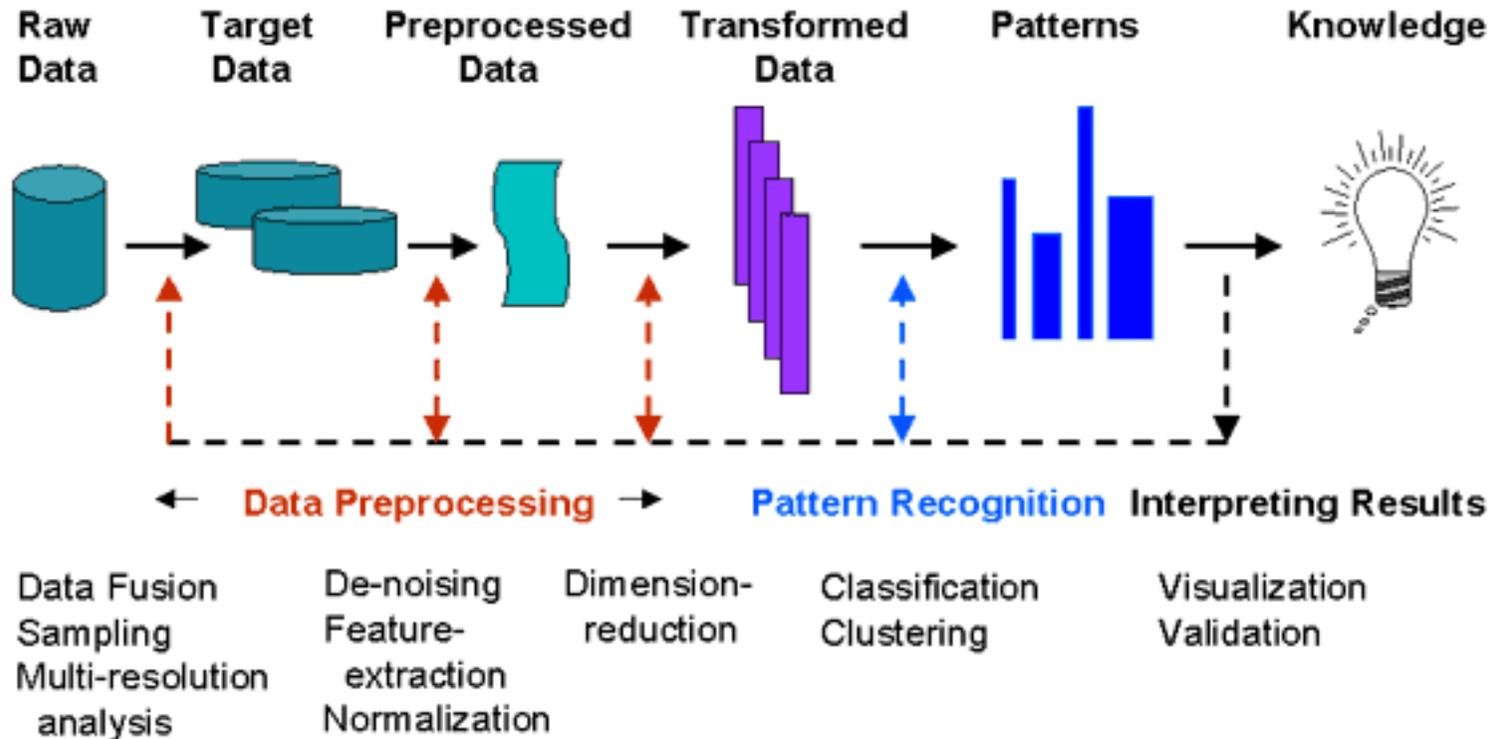
# 유전자형 데이터 행렬

		# of controls					# of cases				
# of SNPs	{	0	1	2	...	0	0	0	0	...	0
		2	0	0	...	1	0	0	2	...	0
		0	0	0	...	0	0	2	0	...	2
		0	0	0	...	0	0	0	0	...	1
		⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮	⋱	⋮
		1	0	2	...	0	2	0	0	...	0

SNP array: 580,961  
 NGS (VCF): 30,000,000

0: major allele homo  
 1: hetero  
 2: minor allele homo

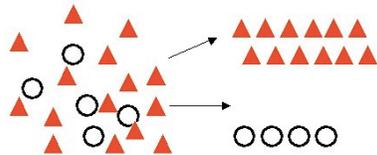
# Data Mining Process



**An iterative and interactive process**

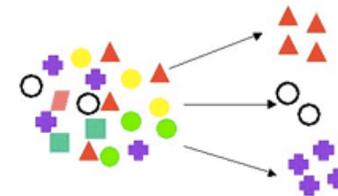
# Methods

## Classification



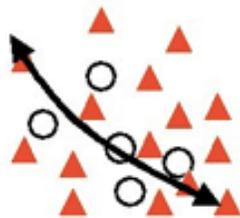
Generalizing known structure to apply to new data.

## Clustering



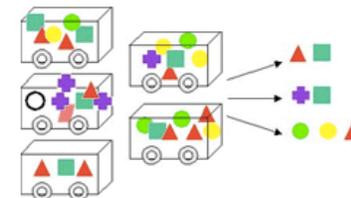
Discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.

## Regression



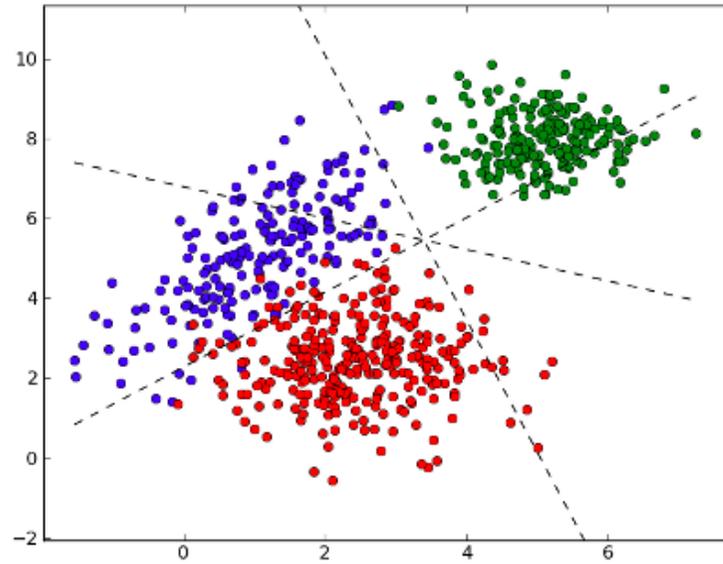
Attempts to find a function which models the data with the least error.

## Association Rules

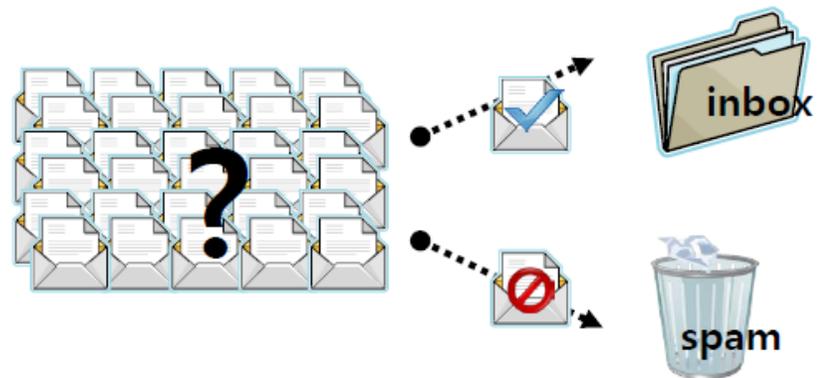


Generalizing known structure to apply to new data.

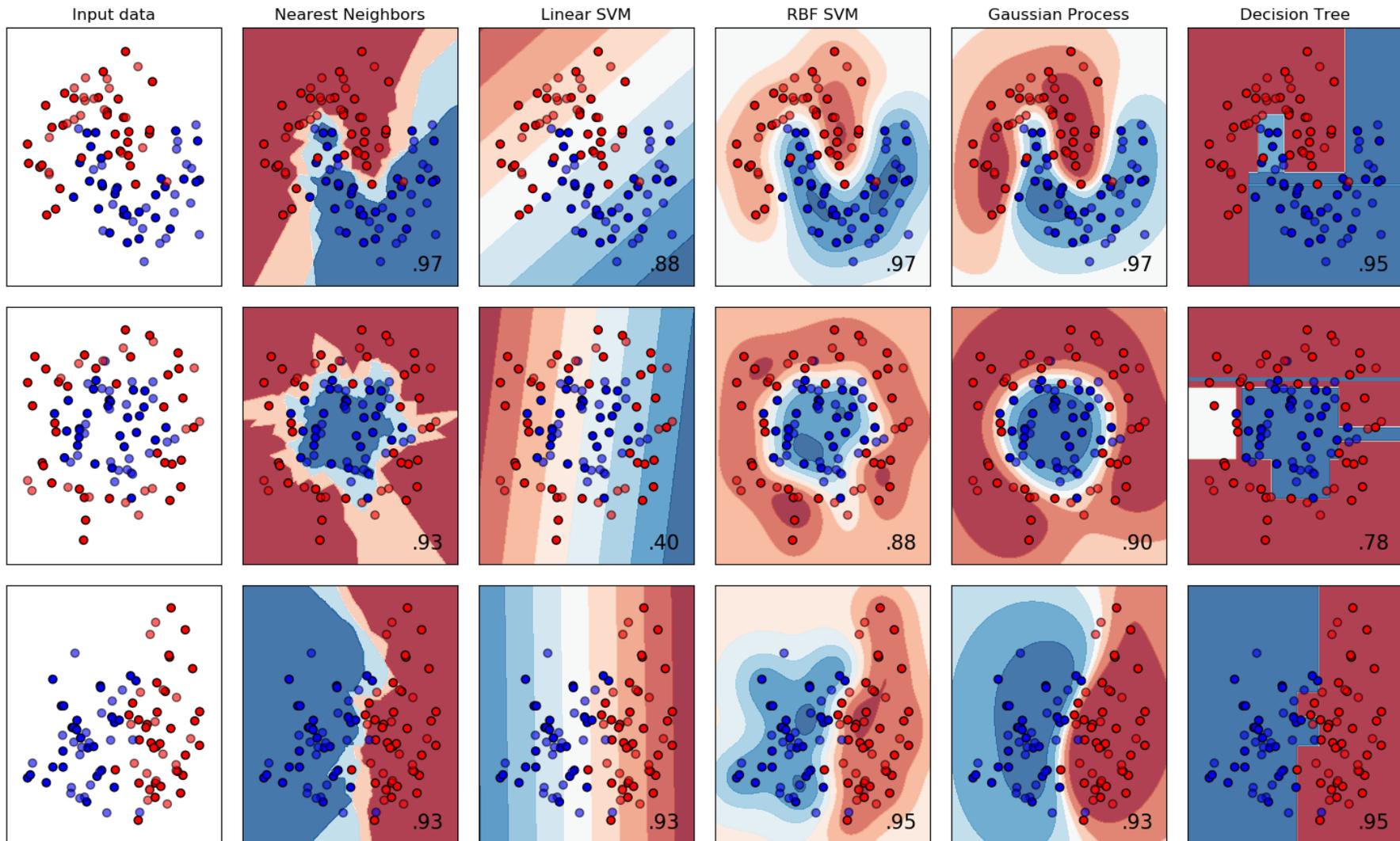
# Classification



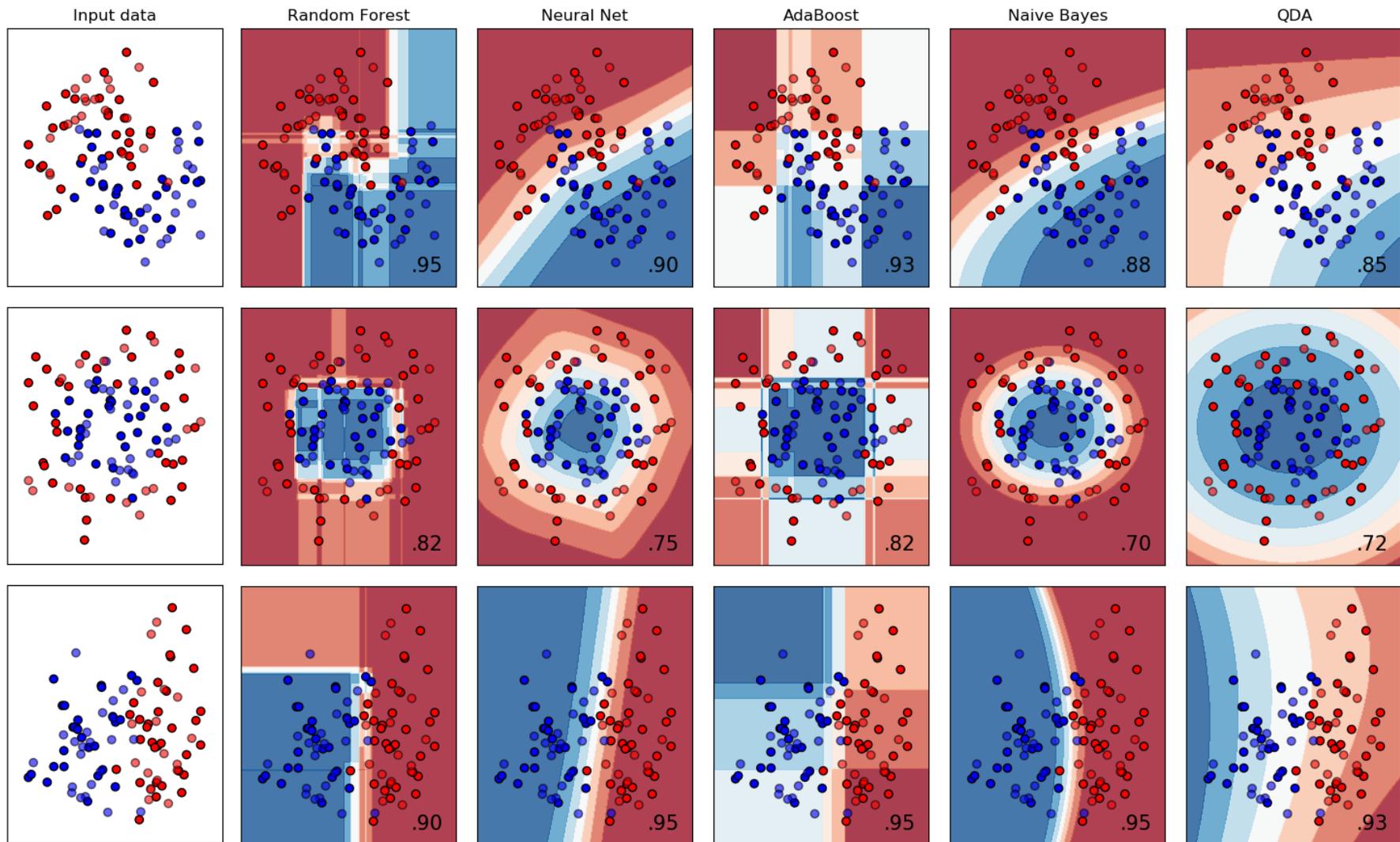
E-mail spam detection



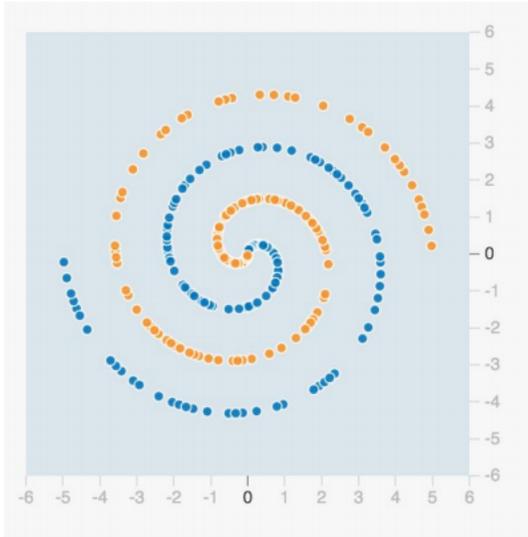
# Classifier comparison



# Classifier comparison

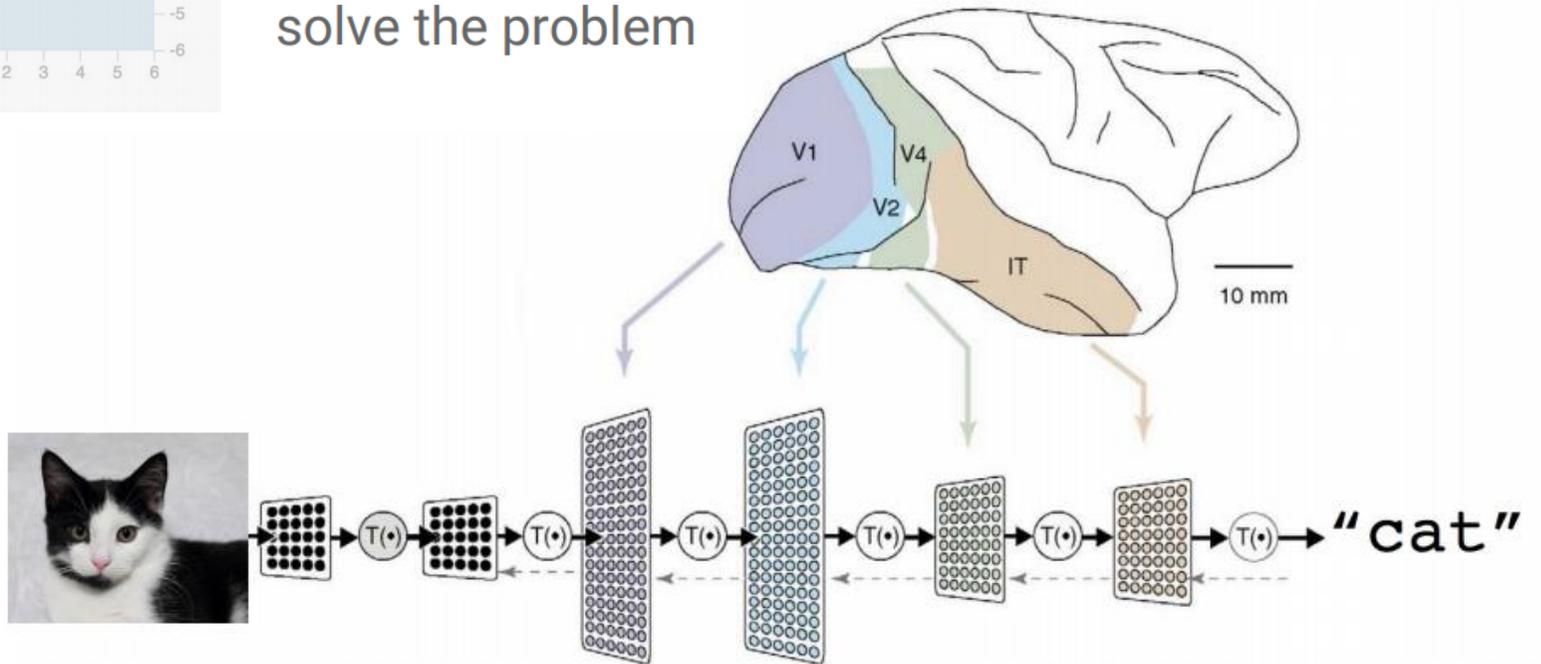


# Deep learning



How do you classify these data points?

Neural Network can **find a way** to solve the problem



# Regression

- Linear regression, k-nearest neighbors(k-NN), artificial neural networks (ANN), ...

- Polynomial curve fitting

- The basic form

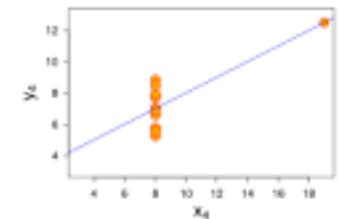
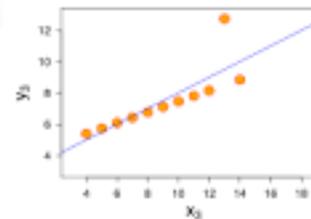
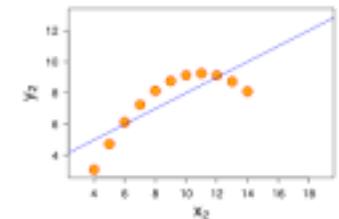
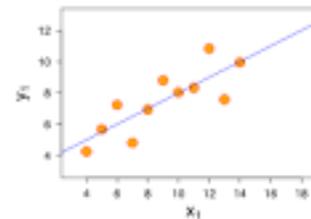
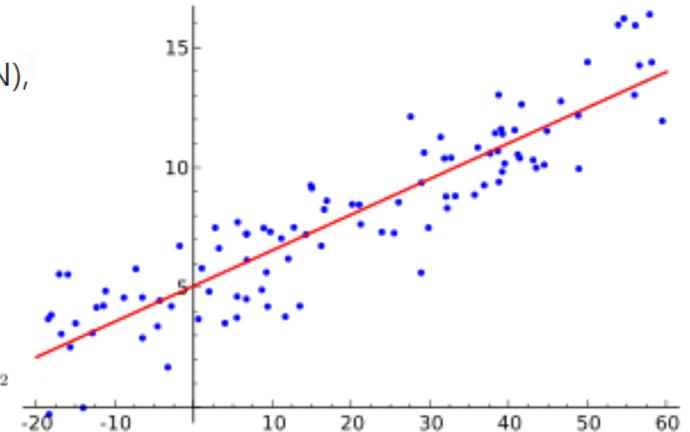
$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j \longrightarrow \min E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

- The advanced form

$$y(x, \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j\phi_j(x) \longrightarrow \min \tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

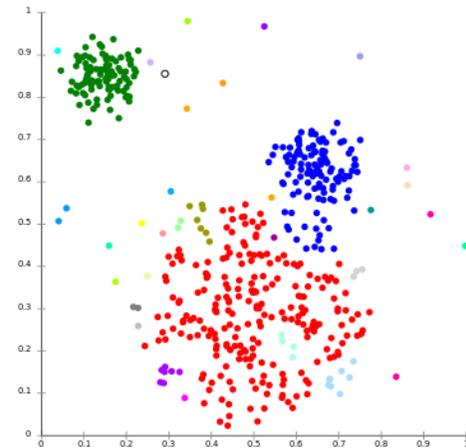
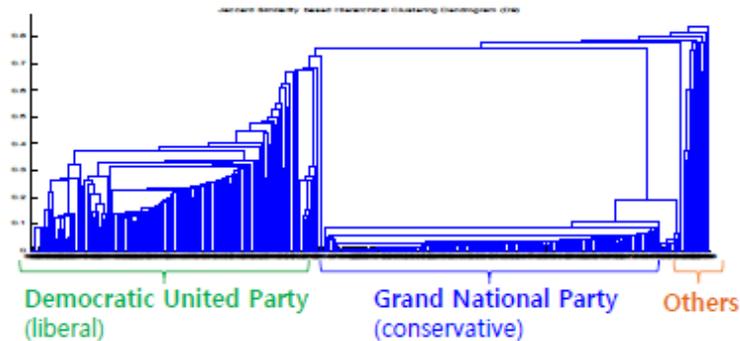
- Example:

- Tomorrow's stock price =  $\mathbf{f}$  (recent prices, economic indicators, ...)

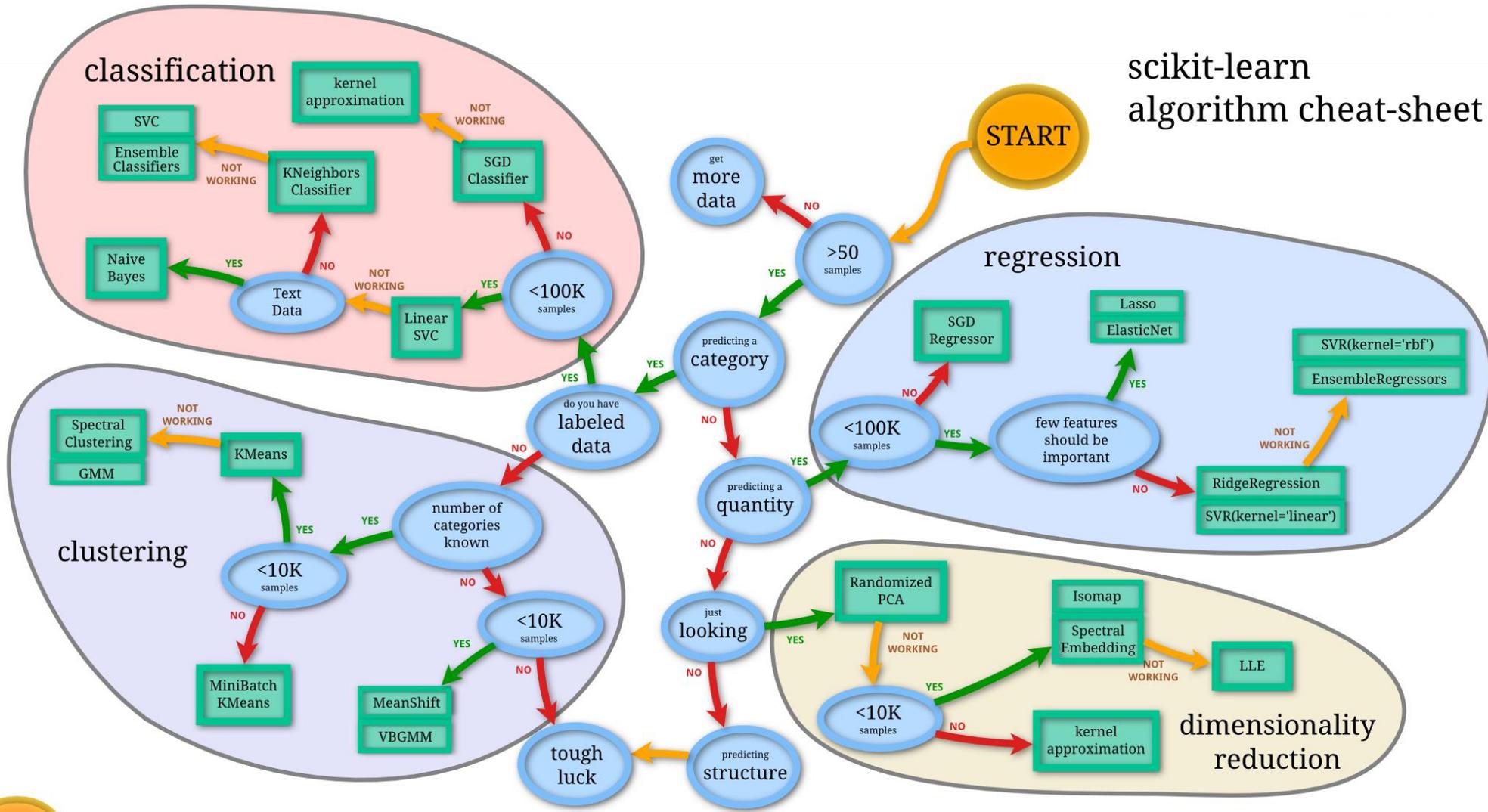


# Clustering

- Grouping of similar objects
- Unsupervised, Exploratory Knowledge Discovery
- k-means, hierarchical clustering, SOM, ...
- Ex: Politician segmentation



# scikit-learn algorithm cheat-sheet



# Feature selection and extraction

Feature selection (특징 선택) 은 모델에서 중요한 특징을 선택하는 것이고, Feature extraction (특징 추출) 은 임의의 특징을 만들어 내는 것.

모델을 만들기 전에 Feature selection을 수행하면 좋은 점

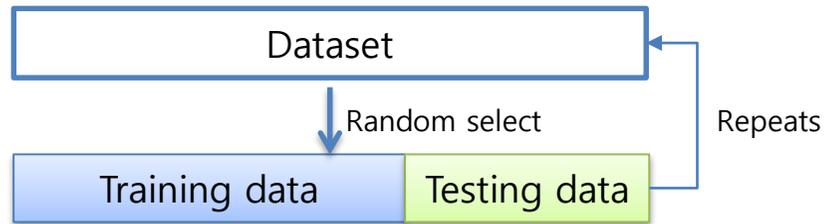
- **Reduce overfitting:** 중복 데이터를 줄인다.
- **Improves accuracy:** 정확도 향상
- **Reduces training time:** 학습시간을 줄인다.

알려진 방법들

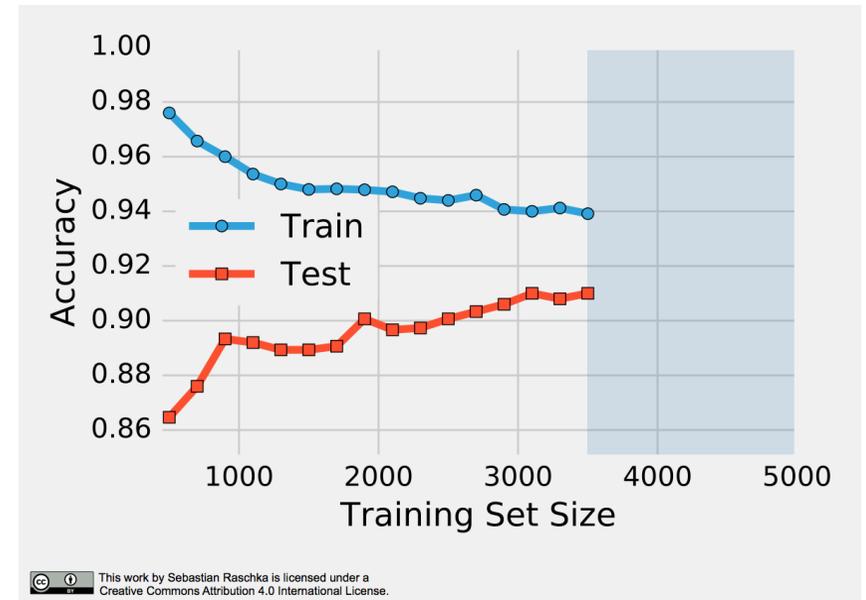
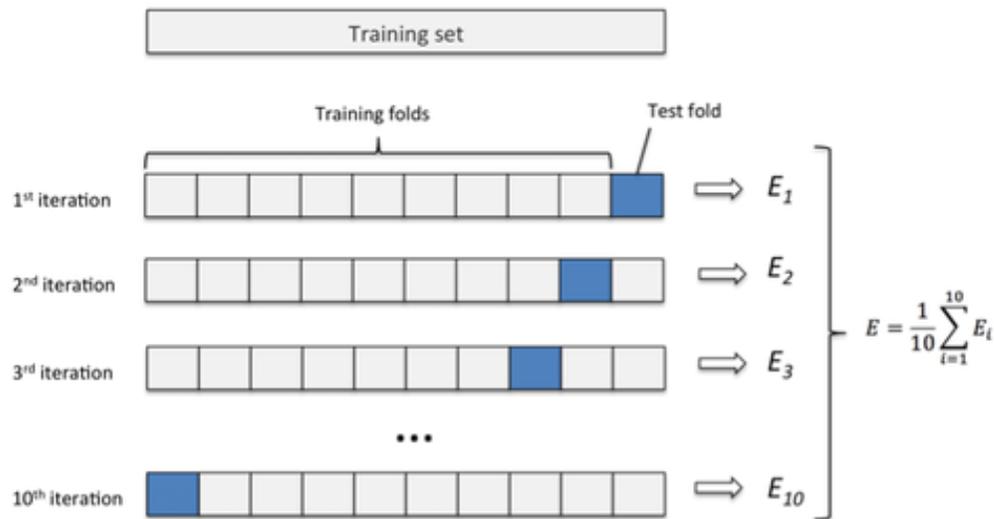
- **Univariate selection:** 각 특징별 빈도로 Chi-squared test 하여, 유의한 특징 선별
- **Recursive feature elimination:** 재귀적으로 특징 제거
- **Meta-transformer** for selecting features based on importance weights
  - L1-based feature selection
  - Tree-based feature selection
- **PCA weights**
- 일부 기계학습 알고리즘에서 important features 정보 제공 (Random forest, Extra trees classifiers,...)

# Training set and testing set

## <Random selection>

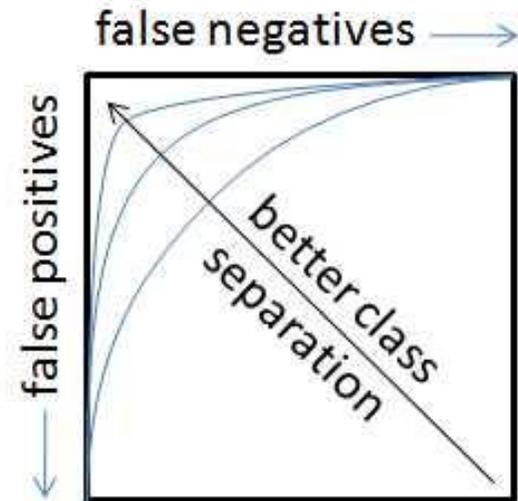


## <K-fold method>

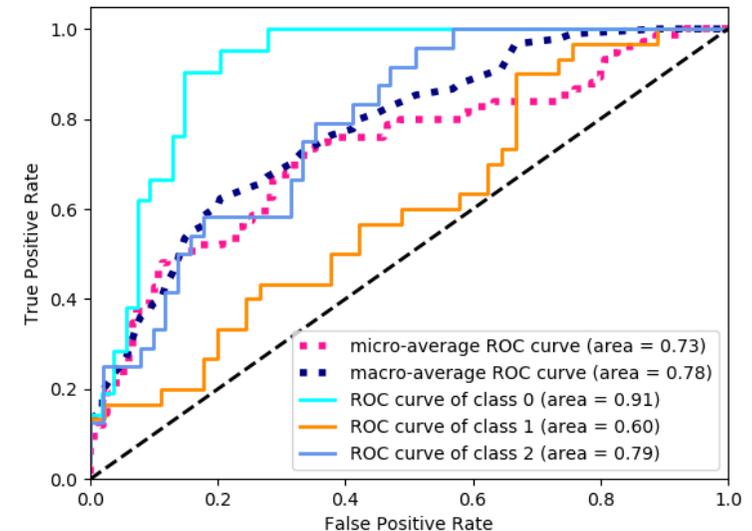


# Estimate accuracy of model for binary classification

		Patients with <b>bowel cancer</b> (as confirmed on <b>endoscopy</b> )		
		Condition positive	Condition negative	
Fecal occult blood screen test outcome	Test outcome positive	<b>True positive</b> (TP) = 20	<b>False positive</b> (FP) = 180	<b>Positive predictive value</b> = $TP / (TP + FP)$ = $20 / (20 + 180)$ = <b>10%</b>
	Test outcome negative	<b>False negative</b> (FN) = 10	<b>True negative</b> (TN) = 1820	<b>Negative predictive value</b> = $TN / (FN + TN)$ = $1820 / (10 + 1820)$ ≈ <b>99.5%</b>
		<b>Sensitivity</b> = $TP / (TP + FN)$ = $20 / (20 + 10)$ ≈ <b>67%</b>	<b>Specificity</b> = $TN / (FP + TN)$ = $1820 / (180 + 1820)$ = <b>91%</b>	

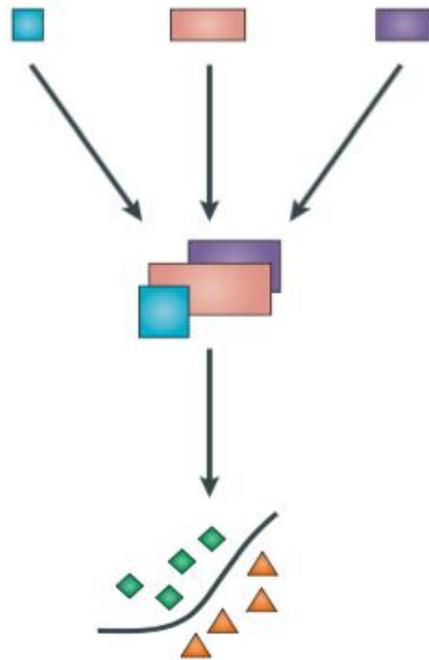


Some extension of Receiver operating characteristic to multi-class

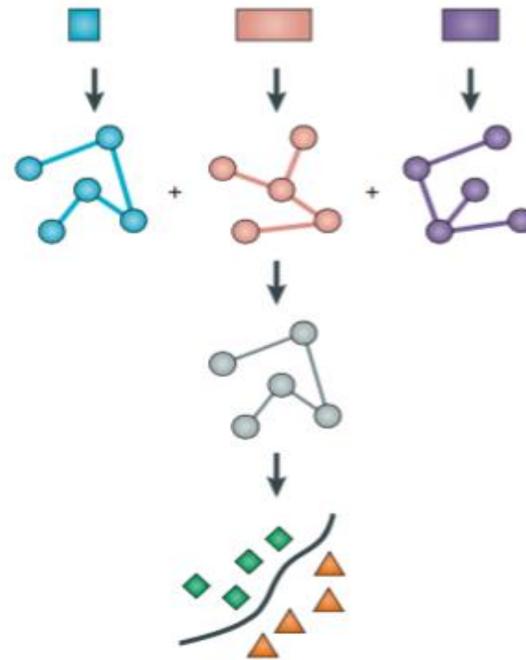


# Meta-dimensional analysis

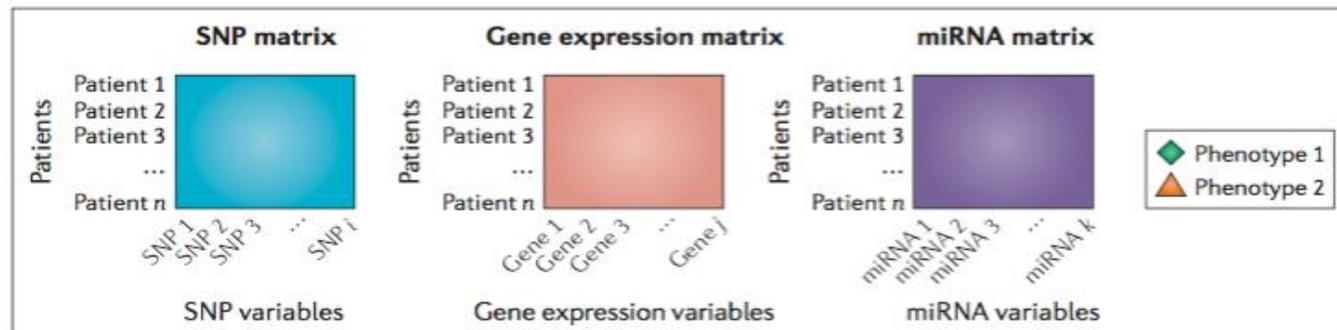
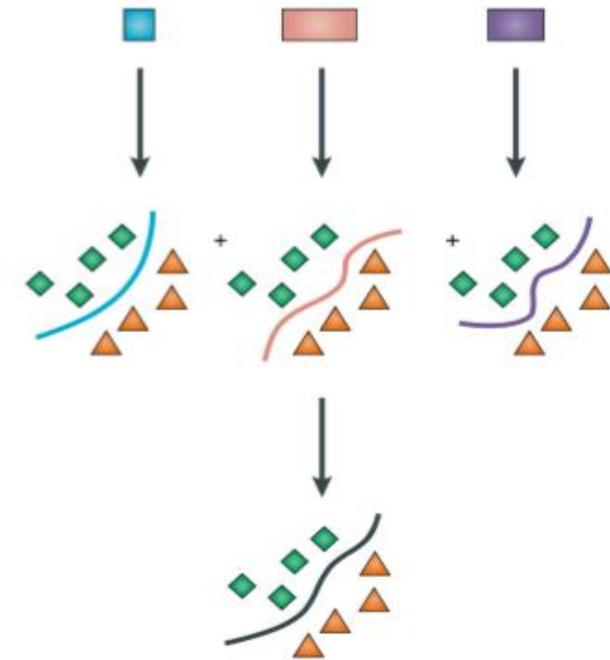
**a Concatenation-based integration**



**b Transformation-based integration**

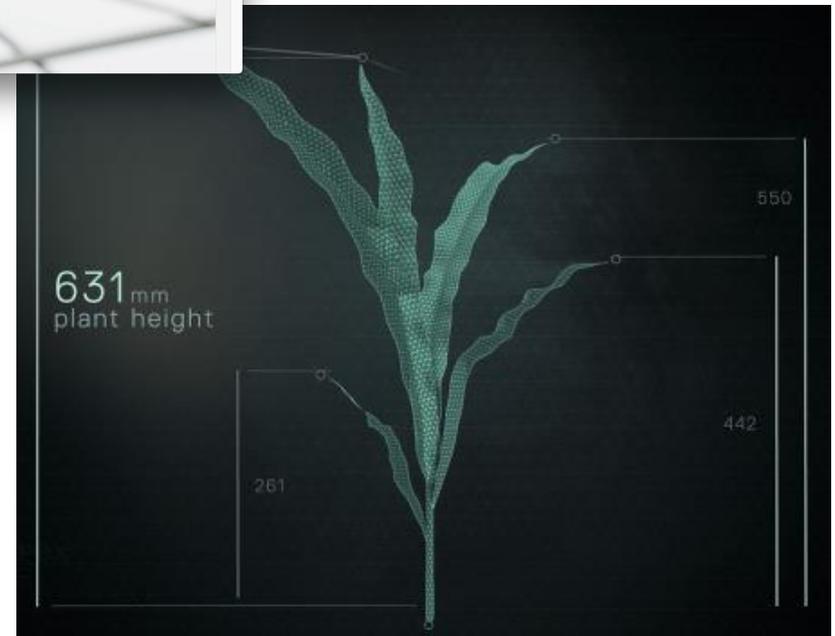
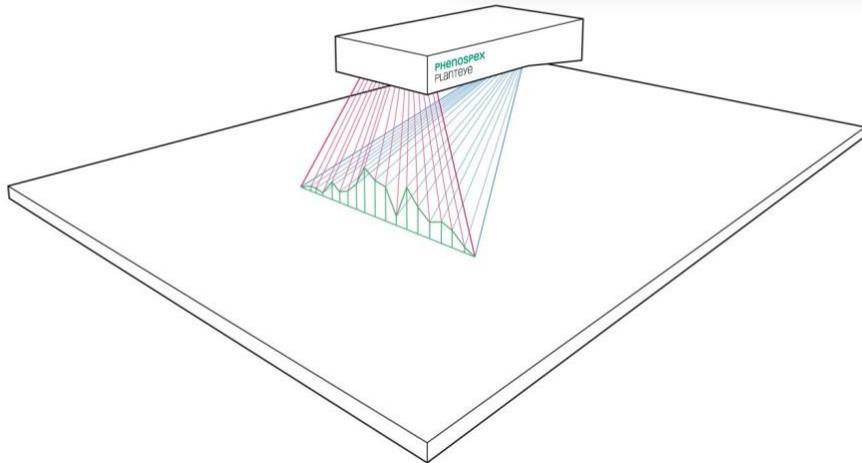
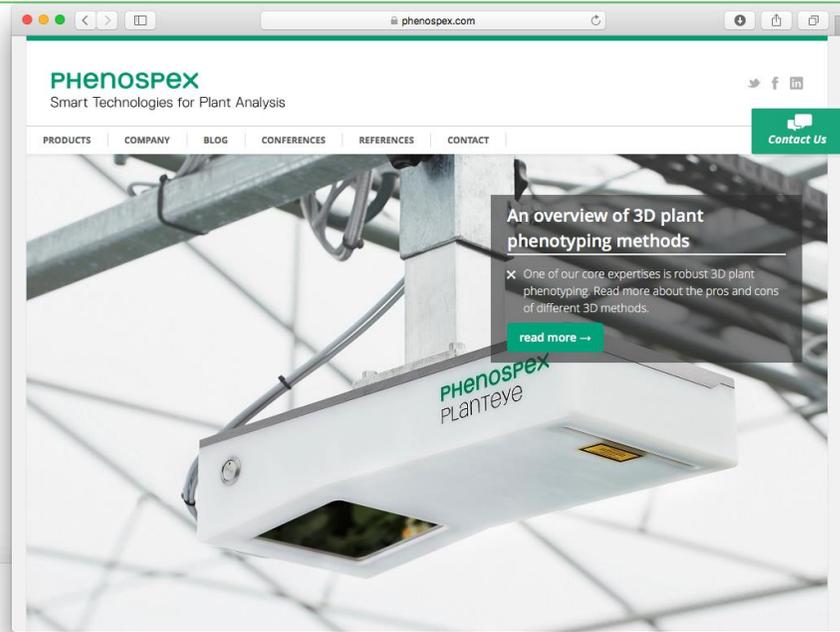


**c Model-based integration**

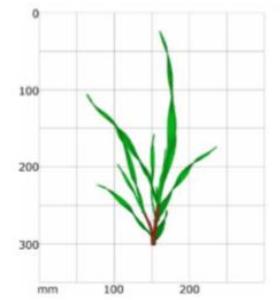
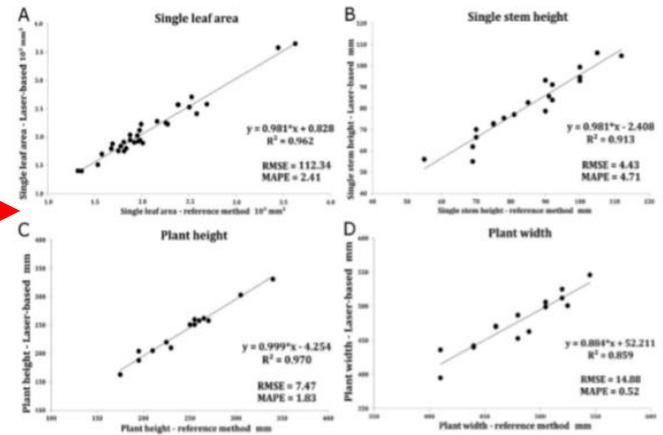
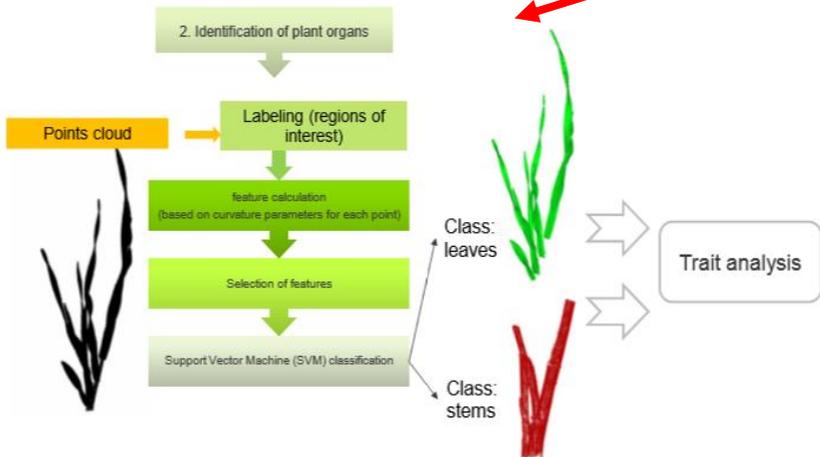


(Nature Review Genetics 2015)

# Massive phenotype data are coming...



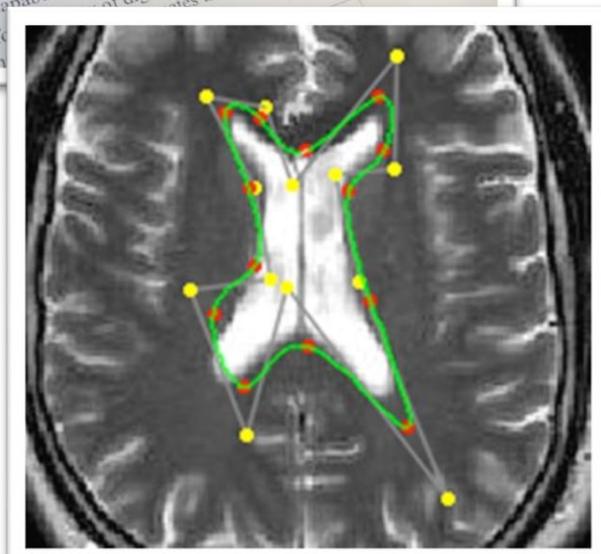
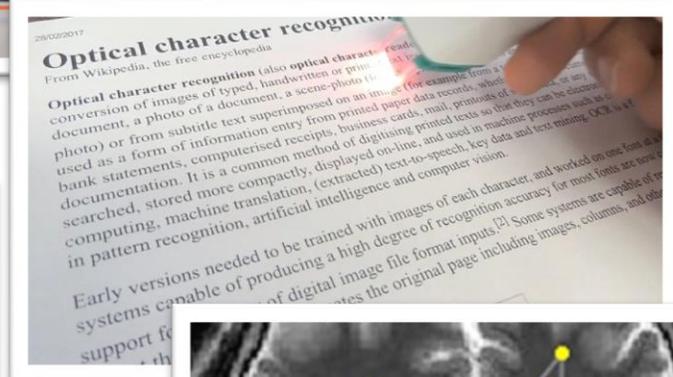
# 영상정보 표현형 측정



Examples of correlation analysis of single leaf area, stem height, plant height and plant width obtained from laser-scanner data integration with manually taken measurements.

# Image processing

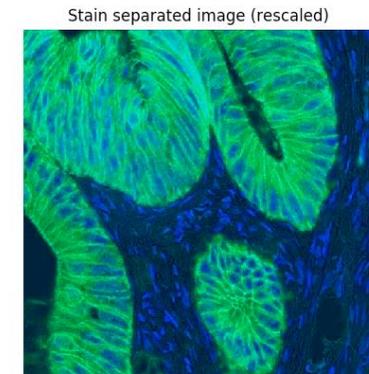
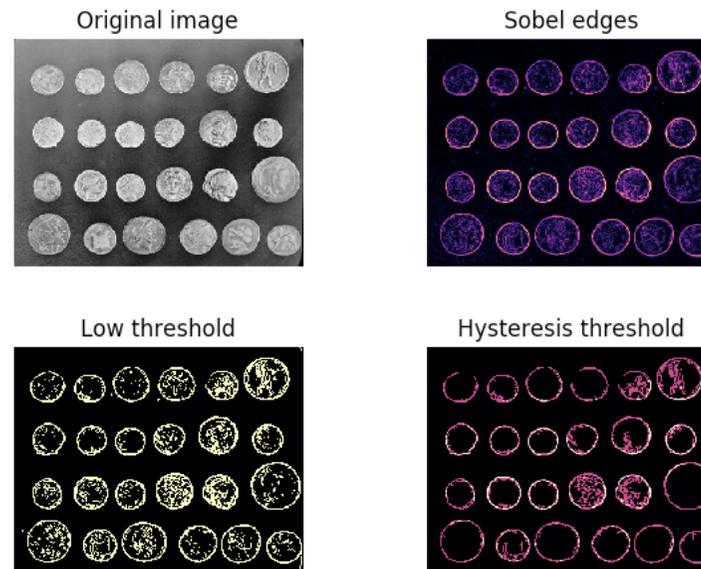
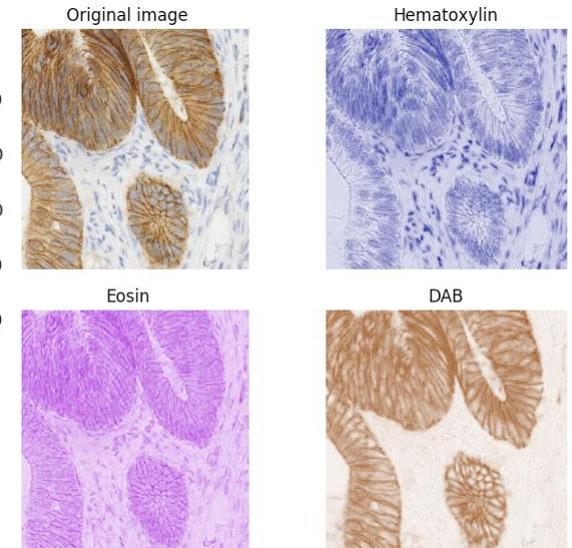
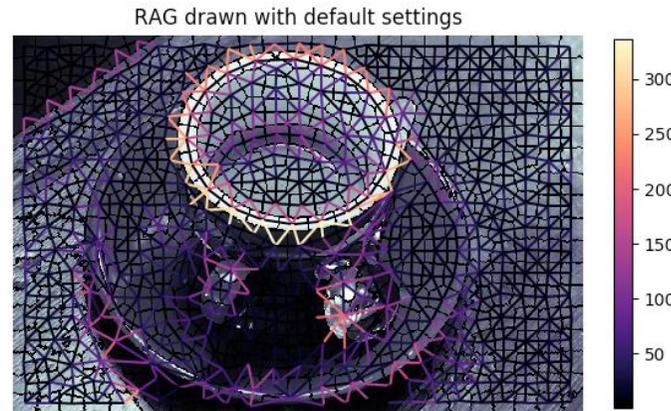
- Machine learning
- Image recognition
- OCR (optical character recognition)
- Satellite imagery
- Medical imaging
- Videophone



# Algorithms for image processing

## Algorithms

- Segmentation
- Geometric transformatio
- Color space manipulatio
- Analysis
- Filtering
- Morphology
- Feature detection



# 영상데이터 딥러닝 기술 적용 사례

## ■ 농업 및 식품 생산에 적용되는 빅데이터 및 딥러닝 기술

Computers and Electronics in Agriculture 147 (2018) 70–90



### 합성곱 신경망 알고리즘 (CNN:Convolution Neural Network) 기반 딥러닝

A. Kamilaris, F.X. Prenafeta-Boldú

Computers and Electronics in Agriculture 147 (2018) 70–90

Review

#### Deep learning in agriculture: A survey

Andreas Kamilaris\*, Francesc X. Prenafeta-Boldú

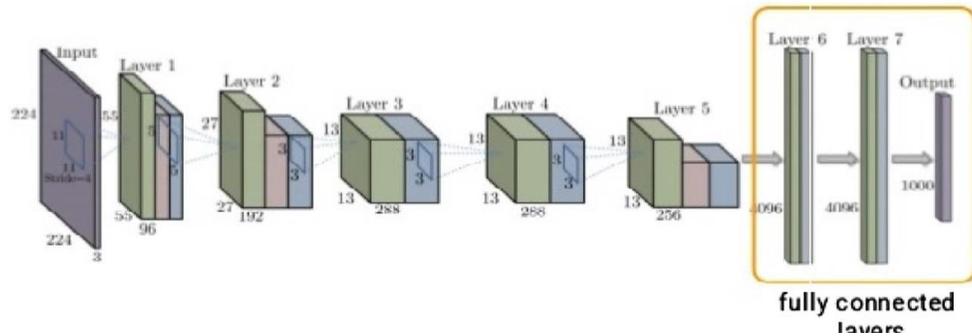
Institute for Food and Agricultural Research and Technology (IRTA), Spain

#### ARTICLE INFO

**Keywords:**  
 Deep learning  
 Agriculture  
 Survey  
 Convolutional Neural Networks  
 Recurrent Neural Networks  
 Smart farming  
 Food systems

#### ABSTRACT

Deep learning constitutes a recent, modern technique for image processing results and large potential. As deep learning has been successfully applied also the domain of agriculture. In this paper, we perform a survey of deep learning techniques, applied to various agricultural and food particular agricultural problems under study, the specific models and their pre-processing of data used, and the overall performance achieved work under study. Moreover, we study comparisons of deep learning with respect to differences in classification or regression performance. Our study provides high accuracy, outperforming existing commonly used image



A. Kamilaris, F.X. Prenafeta-Boldú

Computers and Electronics in Agriculture 147 (2018) 70–90

- 식물의 질병을 식별하기 위하여 식물 잎의 이미지를 딥러닝 기술로 Convolutional Neural Networks (CNN) 활용

- 1) 데이터의 정확한 분류를 위해 질병을 보이는 잎의 이미지 학습 데이터 전처리 (**Convolution**)
- 2) 컨볼루션을 통해 질병이 있는 이미지의 특징이 정돈된 상태에서 가장 핵심적인 부분 추출 (**Pooling**)
- 3) 전처리를 한 데이터를 가지고 실제 다층의 퍼셉트론에 데이터를 집어 넣어서 질병을 보이는 식물 잎 데이터 분류

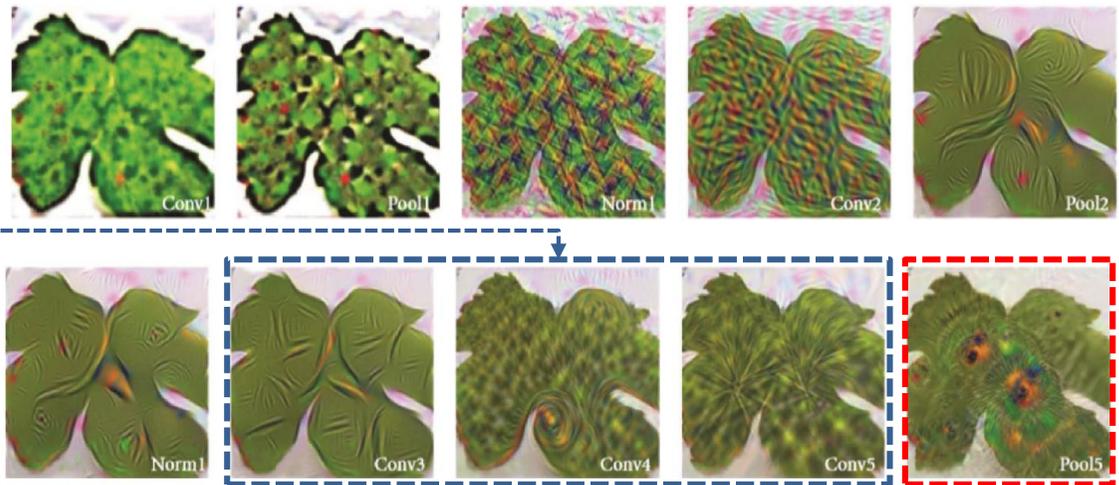
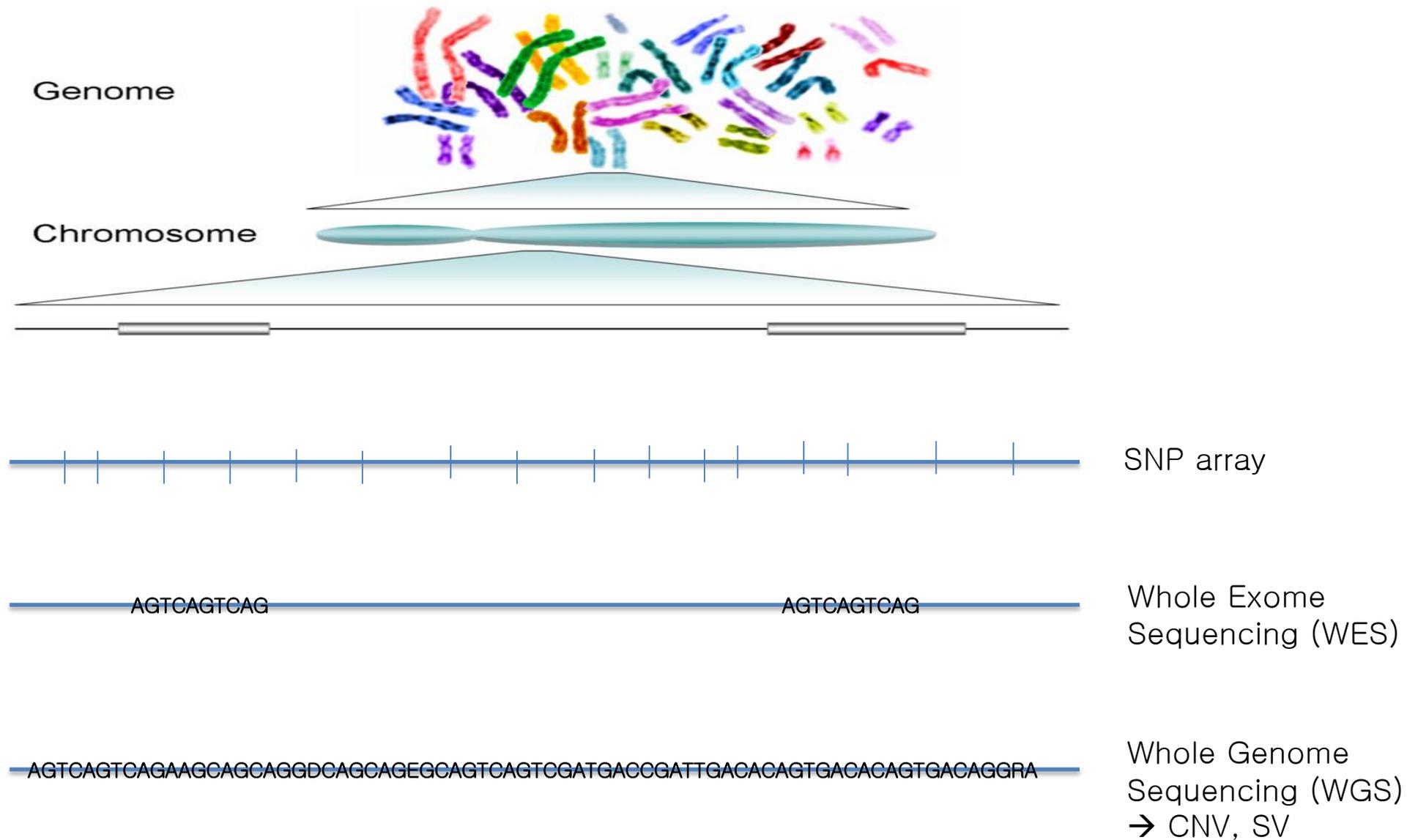


Fig. 2. Visualization of the output layers images after each processing step of the CaffeNet CNN (i.e. convolution, pooling, normalization) at a plant disease identification problem based on leaf images.  
 Source: Sladojevic et al. (2016).

# Genotyping → Genome scanning (From SNP array to NGS)



### 3. 바이오 빅데이터를 활용한 식품 산업 연계



# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ 개인의 다양성에 따른 맞춤형 식품의 필요성 (Personalized Nutrition)

**Cell**

Explore Online Now Current Issue Archive Journal Information For Authors

< Previous Article Volume 163, Issue 5, p1079–1094, 19 November 2015

ARTICLE

**Personalized Nutrition by Prediction of Glycemic Responses**

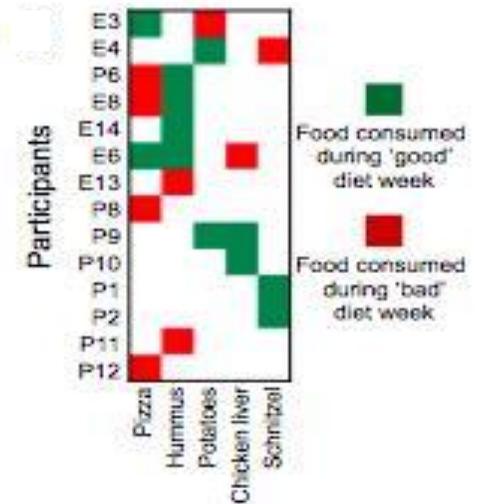
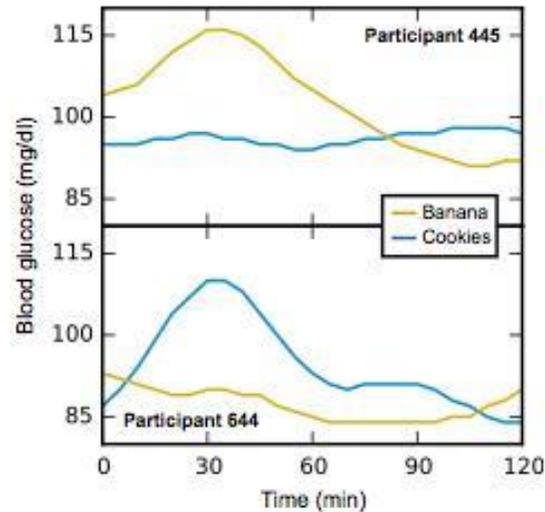
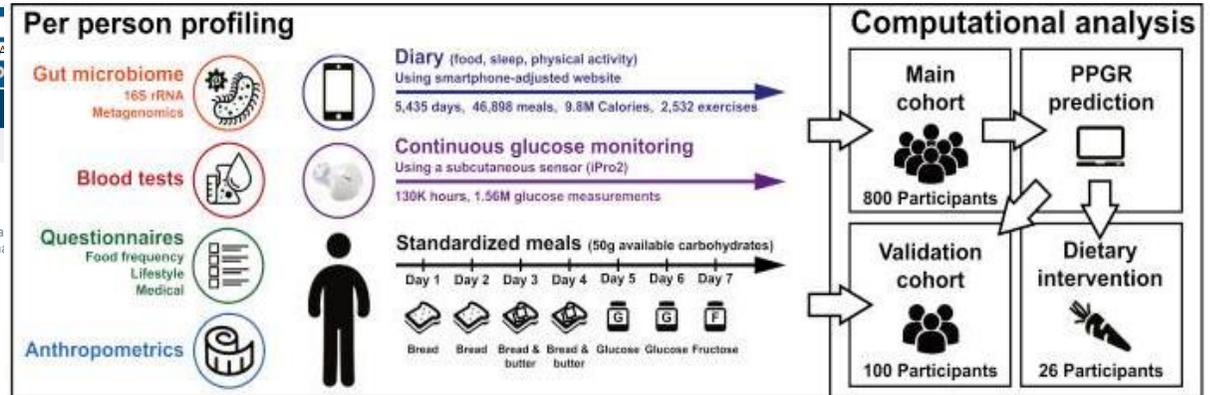
David Zeevi<sup>1</sup>, Tal Korem<sup>2</sup>, Niv Zmora<sup>3</sup>, David Israeli<sup>4</sup>, Daphna Rothschild, Adina Weinberger, Orly Ben-Yacov, Dar Lador, Ta Maya Lotan-Pompan, Jotham Suez, Jemal Ali Mahdi, Eliad Matot, Gal Malka, Noa Kosower, Michal Rein, Gili Zilberman-Schi Meirav Pevsner-Fischer, Rony Bikovner, Zamir Halpern, Eran Elinav<sup>5</sup>, Eran Segal<sup>6</sup>

Co-first author  
Co-senior author

Open Archive PlumX Metrics

DOI: <https://doi.org/10.1016/j.cell.2015.11.001> | CrossMark

Article Info



# 바이오 빅데이터를 활용한 식품 산업 연계

## 빅데이터 기반의 맞춤형 정밀 식품 추천의 필요성

### Current World Population (2018, June 20)

This page features the world population data by country with live calculation based on the birth/death ratio. Whilst every effort is made to ensure the population information is correct and up to date, there may be changes which we are not aware of.

7,477,677,162	Current World Population
202,428	Births Today
63,302,221	Births This Year
85,348	Deaths Today
26,689,585	Deaths This Year

### Top 30 Largest Countries by Population (Live)

China 1,386,269,132	India 1,296,504,958	United States 326,741,985
Indonesia 263,408,383	Brazil 205,519,352	Pakistan 208,327,005
Bangladesh 160,405,114	Russia 141,709,959	Japan 126,252,538
Mexico 126,371,784	Philippines 106,245,540	Vietnam 97,101,397
Egypt 94,666,993	Iran 84,744,089	Germany 80,229,478
Turkey 81,872,928	United Kingdom 64,773,370	Italy 61,811,958
South Korea 51,185,185	Spain 48,592,196	Colombia 47,125,906
Ukraine 43,869,836	Argentina 43,806,288	Algeria 40,263,711
Poland 38,470,100	Iraq 40,183,930	Sudan 38,250,047
Canada 35,488,389	Morocco 33,655,786	Afghanistan 34,948,475

**Past, Present & Future of Personalized Nutrition**

Current estimated market size: **\$6-18bn**

**Key Drivers**

- Increased interest in health & wellness
- Aging population
- Rapid developments in technology
- Quantified self-monitoring
- Saves time

**What Consumers Want Personalized**

- Ingredients
- Nutrients
- Health goals
- Portion size
- Taste
- Supplements
- Cooking preference
- Packaging

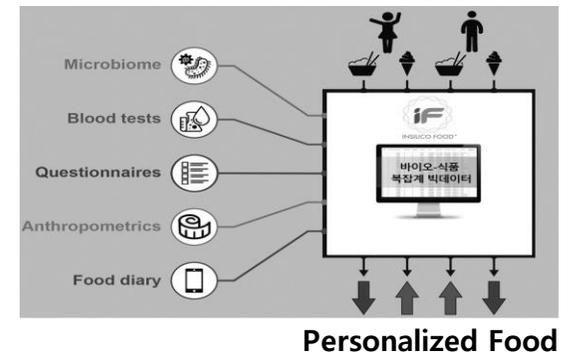
**Who consumers want to hear personalized nutrition advice from**

- 31% Family Doctor
- 28% Dietitian/Nutritionist

Ref. <https://marietteabrahams.com>

## Remain 41% Consumer + α

### Advice from Bio Big Data



# The battlefield of Big Data in Farming & Food

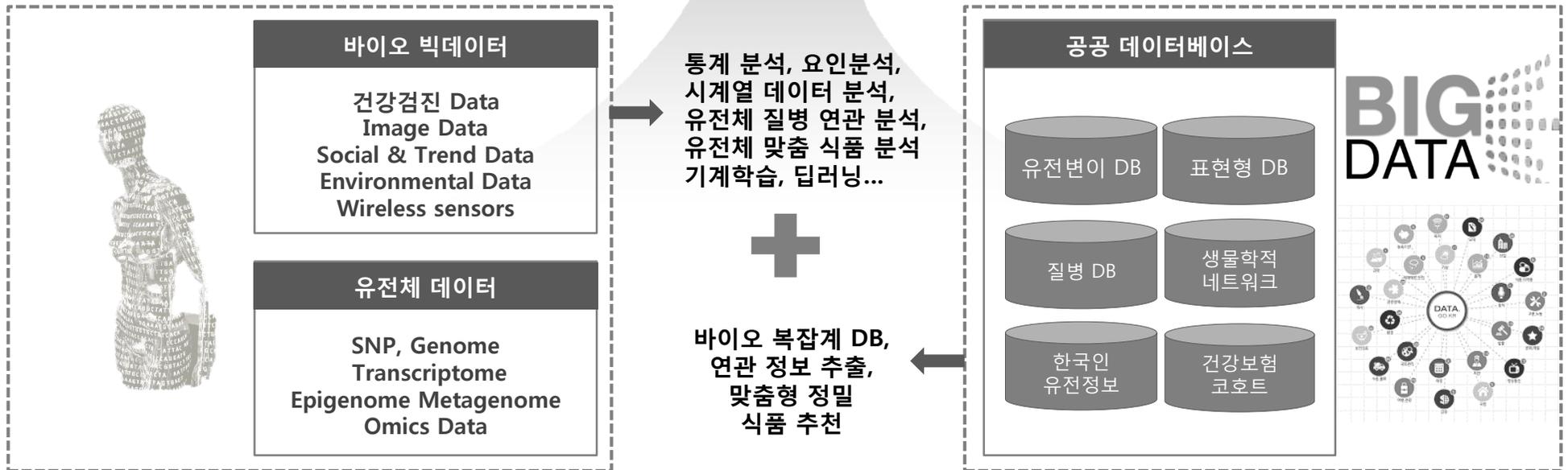
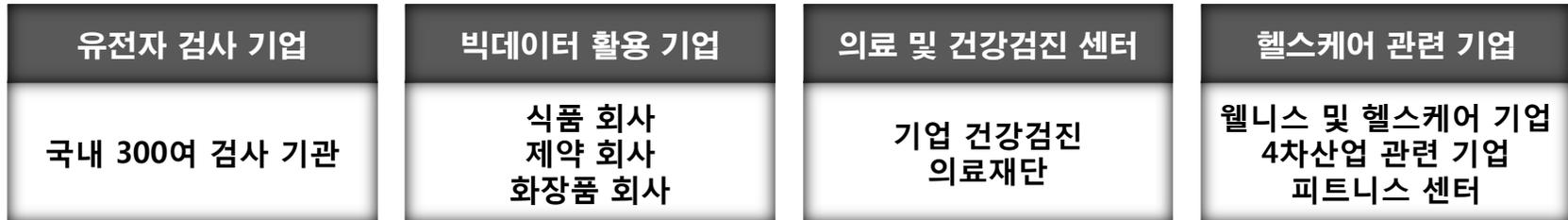


See: Wolfert et al., *Agricultural Systems* 153 (2017) 69–80

# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ 바이오 빅데이터와 공공 빅데이터와의 융합

### 바이오 식품 빅데이터를 활용한 비즈니스 분야



# 바이오 빅데이터를 활용한 식품 산업 연계

- 유전자 검사 기업 : 유전자 및 건강기능 식품 데이터 융합을 통한 맞춤형 건기식 비즈니스

바로, 유전적 차이와 식생활 습관의 차이 때문입니다.



개인의 타고난 유전자형과 식생활습관을 파악하여 환경요인을 개선하고, 자신의 건강상태에 맞는 건강관리를 함으로써 개인별 맞춤 건강관리가 가능합니다.



당신의 유전자에 따른

# 맞춤형 솔루션

도리픽 헤벌라이프 멤버에게 문의하세요!



# 바이오 빅데이터를 활용한 식품 산업 연계

- 빅데이터 활용 기업 : 비만 및 식품 빅데이터와 사물인터넷(IoT), 인공지능(AI) 융복합 비즈니스



# 바이오 빅데이터를 활용한 식품 산업 연계

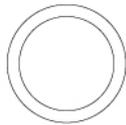
- 건강검진 : 유전자 데이터, 건강검진 데이터 및 ICT 기술 융합을 통한 헬스케어 비즈니스



# 바이오 빅데이터를 활용한 식품 산업 연계

- 헬스케어 : 의료 기술, 레시피 데이터 및 식품 빅데이터를 융합한 푸드테라피 및 음식업 비즈니스

clinic PLATE



의사의 음식 처방, 푸드테라피  
You are what you eat



# 바이오 빅데이터를 활용한 식품 산업 연계

## 4차 산업 : 유전자 데이터, 헬스 데이터 및 블록체인 기술 융합을 통한 헬스케어 비즈니스



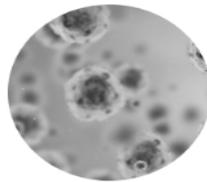
**유전자  
분석 DATA**

- 타고난 신체 특성 유전자 분석
- 어떤 유전자가 어떤 병에 결정적 영향을 미치는지도 밝힐 수 있고, 유전성 질환을 일으키는 유전자를 교체하거나 작동할 수 없도록 미리 예방 가능



**기능 의학  
분석 DATA**

- 일반적인 검사에 이상이 없음에도 불구하고 만성피로, 우울, 무기력, 어지러움증 등의 증상이 있는 현대인의 필수적인 검사
- 기능의학검사 결과를 근거로 과학적인 치료와 근본적 치료 가능



**면역력  
분석 DATA**

- 장내세균은 몸 속의 대사과정에 영향을 미치는 매우 중요한 환경적 요인
- 과학적으로 유익균과 유해균의 양을 분석하여 장내세균의 구성을 확인 하고 식생활 개선 등을 통한 장내 환경의 변화를 모니터링하여 맞춤 서비스 가능



**개인 생체  
Life Log Data**

- 개인의 일상과 행동 패턴을 데이터로 저장하고 활용할 수 있는 라이프로그(Life Log) 데이터
- 라이프로그를 통해 상황에 맞는 맞춤형 서비스를 구현하는 등 새로운 서비스나 상품 개발이 가능



**my23**

유전자 / 미네랄 영양 균형 검사 기반 헬스케어 맞춤 솔루션



**SELINA**

IoT Device 셀리나 | 성분 측정기 다이어트 케어 서비스

# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ 건강기록 데이터, 영양 성분 데이터, 의료정보 및 ICT 기술을 융합한 다이어트 비즈니스

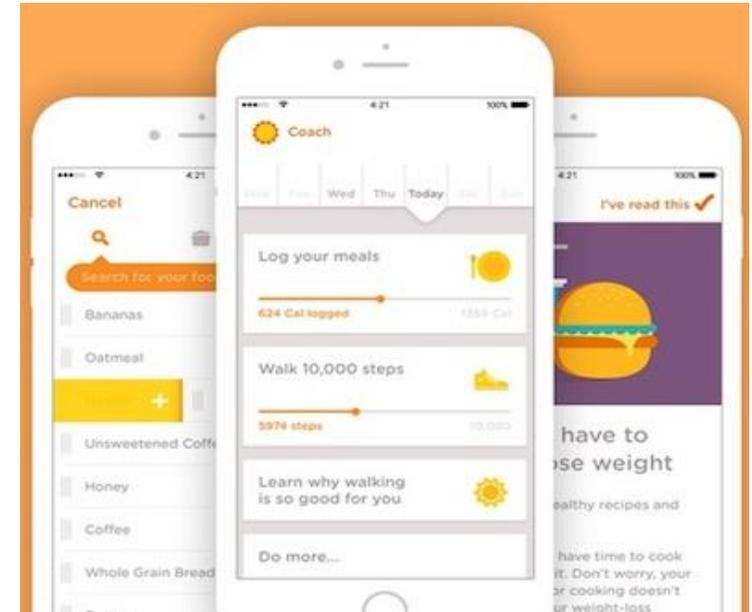
- 활동량 데이터 다음으로 대두되고 있는 분야
- 스마트폰을 활용한 손쉬운 식사 기록
- 식생활 변화를 통한 체중 감량 및 만성질환 예방에 활용



- 장소 : 건강보험공단 건강증진센터 18개 지사
- 기간 : '16년 6월 - '16년 11월
- 대상 : 만성질환 위험인자 소유자 500명



noom



- 장소 : 수지구 보건소
- 기간 : '16년 6월 - '16년 12월(본사업)
- 대상 : 3050세대 비만/전당뇨/고지혈증 소견자 총 200명 이상



- 장소 : 경기도 광주시 보건소
- 기간 : '16년 9월 - '16년 11월
- 대상 : 지역주민 및 직장인 대상 비만/전당뇨/고지혈증 소견자 총 150명



- 장소 : 포항시 남구 보건소
- 기간 : '16년 9월 - '16년 11월
- 대상 : 지역주민, 지역 사업장 근로자 및 대학생 비만/전당뇨/고지혈증 소견자 총 200여명



400만명 돌파  
한국 내 사용자  
(2017.6 기준)



4천 7백만명  
늘 서비스 누적 이용자



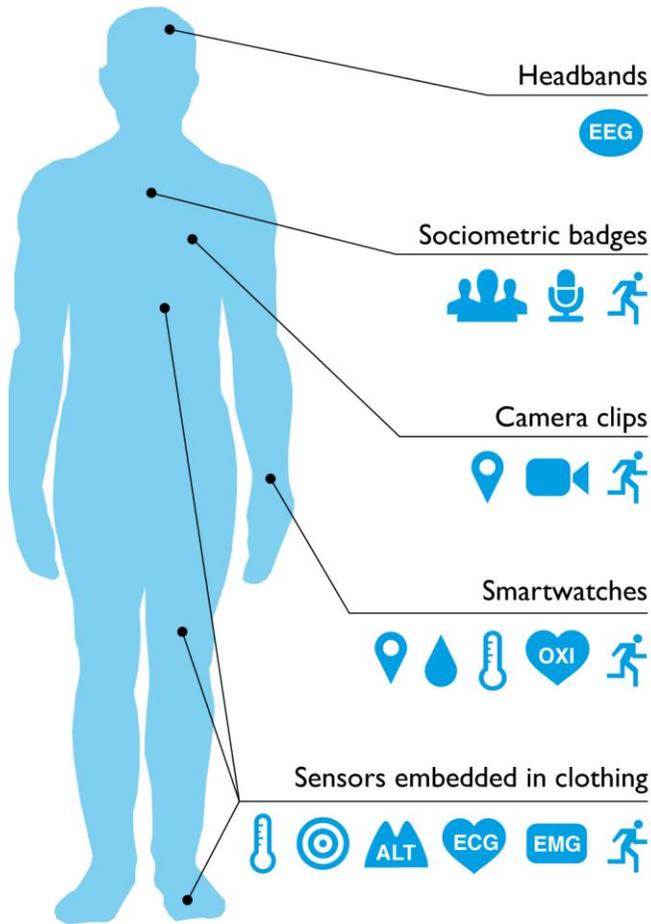
모바일 당뇨병 예방 프로그램

미국 질병예방본부(CDC) 세계 최초 공식 인증

건강, 식품, 당뇨, 고객, 비만 관련 빅데이터

# 바이오 빅데이터를 활용한 식품 산업 연계

- Wearable 라이프로그 데이터, 표현형 데이터, 유전자 데이터를 융합한 맞춤형 식단 비즈니스



- Accelerometer
- Altimeter
- Digital camera
- Electrocardiogram
- Electromyograph
- Electroencephalogram
- Electrodermograph
- Location GPS
- Microphone
- Oximeter
- Bluetooth proximity
- Pressure
- Thermometer



## PERSONALIZED NUTRITION

Building the foundation for a future food ecosystem

The screenshot shows the 'habit' website with the tagline 'FOOD, PERSONALIZED'. The navigation menu includes 'HOW IT WORKS', 'SCIENCE', 'SUCCESS STORIES', and 'SHOP NOW'. Below the navigation, there are three main sections:

- TAKE THE HABIT NUTRITION TEST:** The convenient at-home Habit Nutrition Test Kit is a comprehensive nutrition test that looks at your DNA and your blood to determine how your body handles carbs, fat and protein. We also learn about your lifestyle and health goals. [LEARN MORE](#)
- REVIEW YOUR TEST RESULTS:** Each area of your results has easy-to-understand explanations, from how you handle macros to insights into 60+ biomarkers, like those relating to heart health. You'll know right away where you're doing well and where you have opportunities to optimize your health with nutrition.
- GET TO KNOW YOUR NUTRITION PLAN:** Our science-based recommendations take the guesswork out of eating so you can make confident food choices. We'll break down what your ideal meal looks like, highlight your hero foods and create personalized daily nutrient recommendations based on your biology, lifestyle and goals.

# 바이오 빅데이터를 활용한 식품 산업 연계

- Wearable Device, 운동량 데이터, 체성분 데이터 및 식품 데이터 융합 피트니스 비즈니스



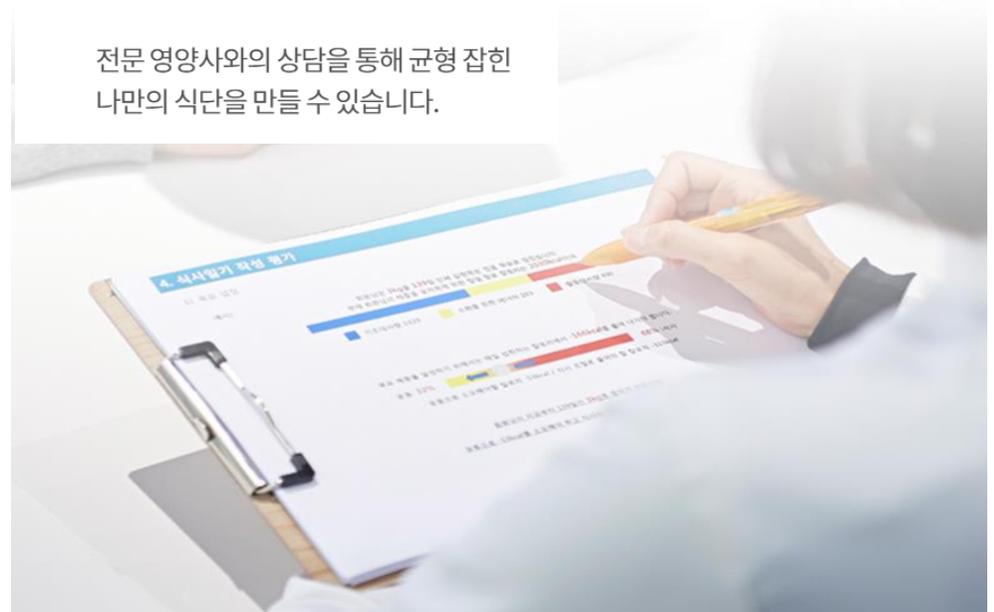
17.5 CIRCUIT TRAINING

# MILON SYSTEM

인바디밴드와 내 몸에 맞게 자동 세팅된 밀론 기구에서  
최적의 운동이 가능하며 모든 운동 기록은 데이터화됩니다.

## 임상영양사

전문 영양사와의 상담을 통해 균형 잡힌  
나만의 식단을 만들 수 있습니다.



## FOOD & BEVERAGE

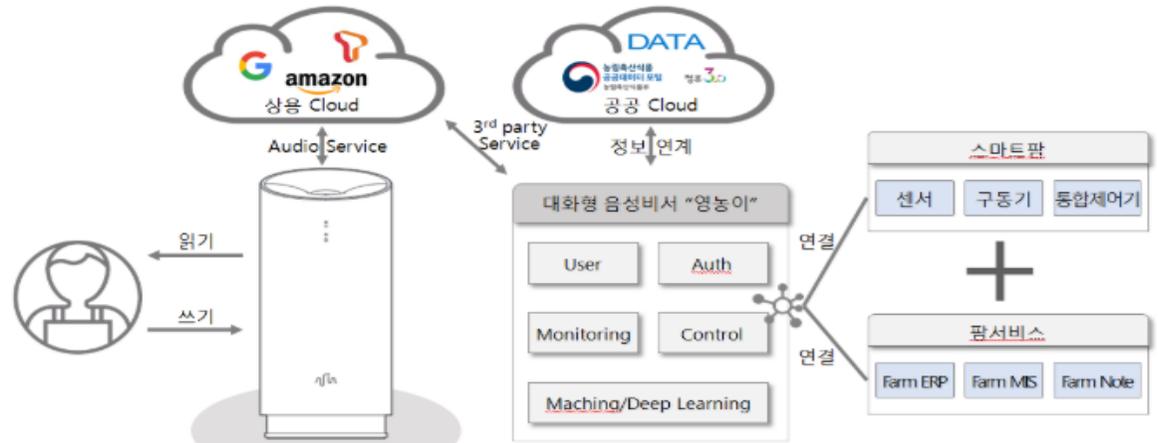
유기농 콜드 프레스 주스와 함께  
건강한 식생활을 관리할 수 있습니다.

# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ 국내 식품 빅데이터 활용



출처: 빅데이터 활용 경진대회 (2017)



출처: 공공데이터 활용 경진대회 (농림부, 2016)



출처: 제노플랜 & Stay Jungle DNA 맞춤 도시락 (2018)

# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ 국내 식품 빅데이터 플랫폼 활용

- ▶ “음료 하나를 골라도 자신의 건강 상태를 고려해 유용 성분이 함유된 제품을 선택하는 이들이 점차 늘어나면서 앞으로 이처럼 유용성이 알려진 식품을 활용한 음료 출시가 더욱 늘어날 것으로 예상
- ▶ KGC인삼공사 ‘천녹삼’: 홍삼 구매층인 40~50대 남성이 피로 해소를 중시하는 점을 감안해 녹용 성분을 넣자 출시 두 달 만에 13억5000만원 판매.
- ▶ 빅데이터를 활용하여 자신의 건강 뿐만 아니라, 생활패턴/습관 등 최적으로 개개인의 맞춤형 형태의 정보 및 제품개발이 계속해서 증가할 것으로 예상

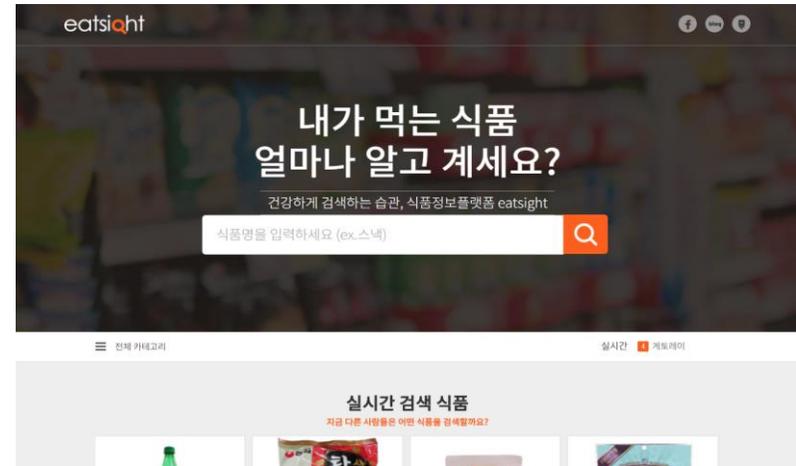
### 비타민인 <https://vitamiin.co.kr>

- 비타민인 서비스는 사용자의 건강상태에 맞는 최적의 건강식품정보를 큐레이션 해주는 서비스
- 제품의 원재료, 영양소, 건강기능식품의 복용방법 등의 정보를 제공함.



### Eatsight <https://www.eatsight.com>

- 시판되는 식품들의 영양성분 정보를 제공하고 있으며, 알레르기 반응을 일으킬 수 있는 식재료에 대한 정보를 제공하고 있음.
- 건강상 주의가 필요한 식품의 영양정보를 제공함.

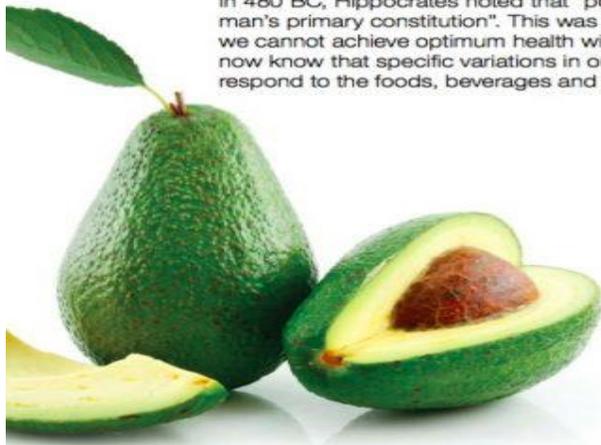


# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ Nutrigenomics를 활용한 비즈니스 및 주요 성과

### The answer may be in your genes

In 480 BC, Hippocrates noted that "positive health requires knowledge of man's primary constitution". This was just an ancient way of saying that we cannot achieve optimum health without knowing about our genes. We now know that specific variations in our genes can explain how we will respond to the foods, beverages and supplements we consume.

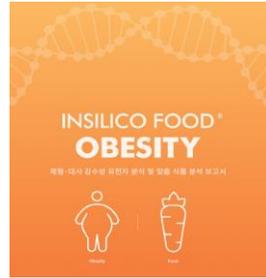


Learn how your genes can affect:



# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ 식품 빅데이터와 바이오 빅데이터를 융합한 식품분석 서비스



건강검진 맞춤 식품 분석

(5종 및 14종 일반질환)

체형 · 대사

Geno Body Care

질환

iF Disease

피부

iF Skin

샘플수집 키트



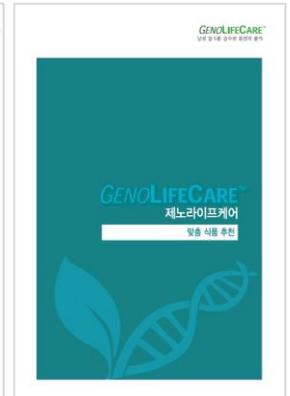
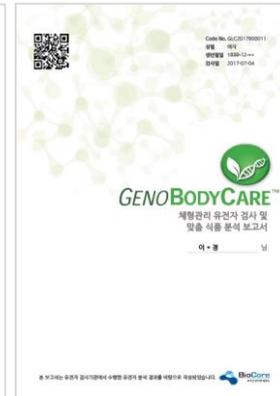
데이터베이스 시스템



iF crm



Report Service Platform



iF DB 활용 제안서



건강검진 식품분석 서비스



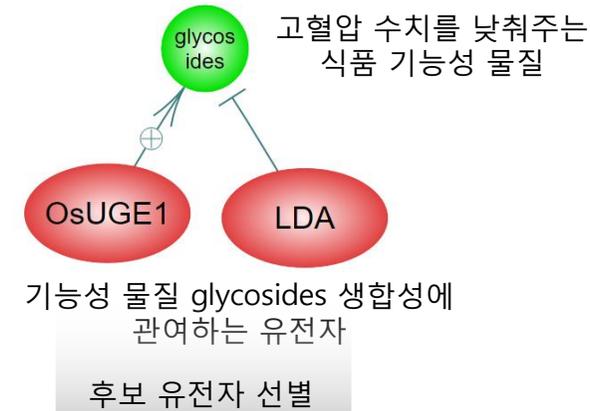
유전자검사 및 개인정보 동의서



# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ 농업 빅데이터를 활용한 기능성 소재 개발

식품 기능성 정보



### 문헌정보

*Biochim Biophys Acta*, 1999 May 18;1431(2):538-46.

#### Isolation and characterization of the gene encoding the starch debranching enzyme limit dextrinase from germinating barley.

Kristensen M<sup>1</sup>, Lok F, Planchot V, Svendsen I, Leah R, Svensson B.

#### Author information

*J Plant Physiol*, 2007 Oct;164(10):1384-90. Epub 2007 Jun 12.

#### Over-expression of OsUGE-1 altered raffinose level and tolerance to abiotic stress but not morphology in Arabidopsis.

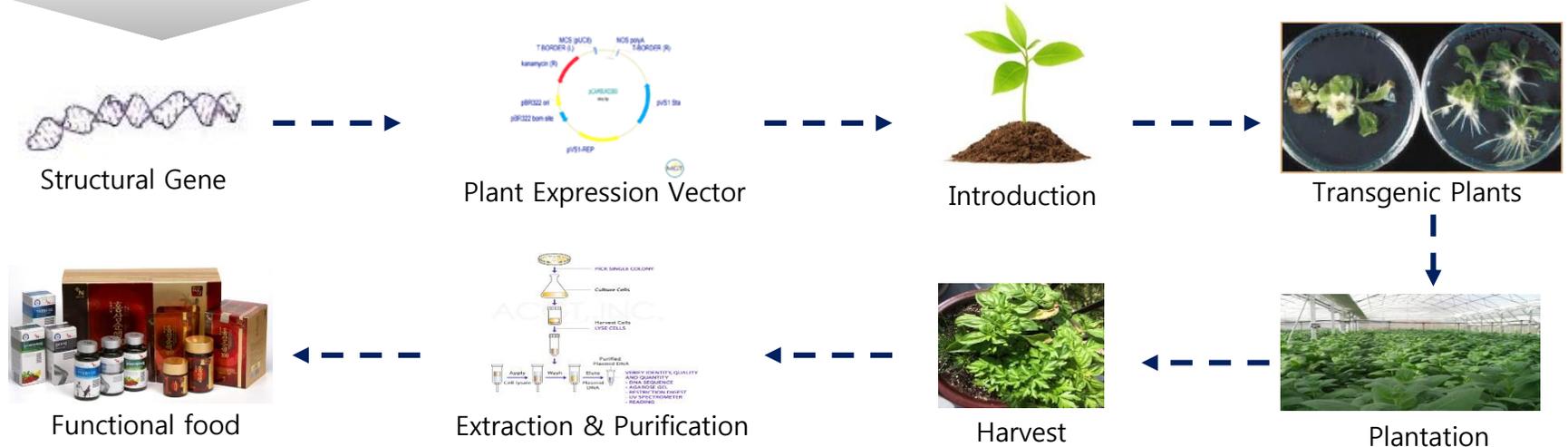
Liu HL<sup>1</sup>, Dai XY, Xu YY, Chong K.

#### Author information

#### Abstract

OsUGE-1 is known to be induced by various abiotic stresses, but its exact function in plants is unclear. In the present study, OsUGE-1 was over-expressed in Arabidopsis, transgenic plants conferred tolerance to salt, drought and freezing stress without altering plant morphology. In addition, transgenic plants showed a higher level of the soluble sugar raffinose than did wild-type plants. Our results suggest that elevated level of raffinose with over-expressed OsUGE-1 resulted in enhanced tolerance to abiotic stress. Thus, the gene may be applied to improve tolerance to abiotic stress in crops.

기능성 소재 생산



## ■ 항진균제 신소재 개발을 위한 비타민 생합성 관련 Fungal Genomics 빅데이터 연구

Review

### Vitamin Biosynthesis as an Antifungal Target

Zohar Meir and Nir Osherov \*

Department of Clinical Microbiology and Immunology, Sackler School of Medicine, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 6978, Israel; lightenzm@gmail.com

\* Correspondence: nosherov@post.tau.ac.il; Tel.: +972-3-640-9599; Fax: +972-3-640-9160

Received: 29 May 2018; Accepted: 13 June 2018; Published: 17 June 2018



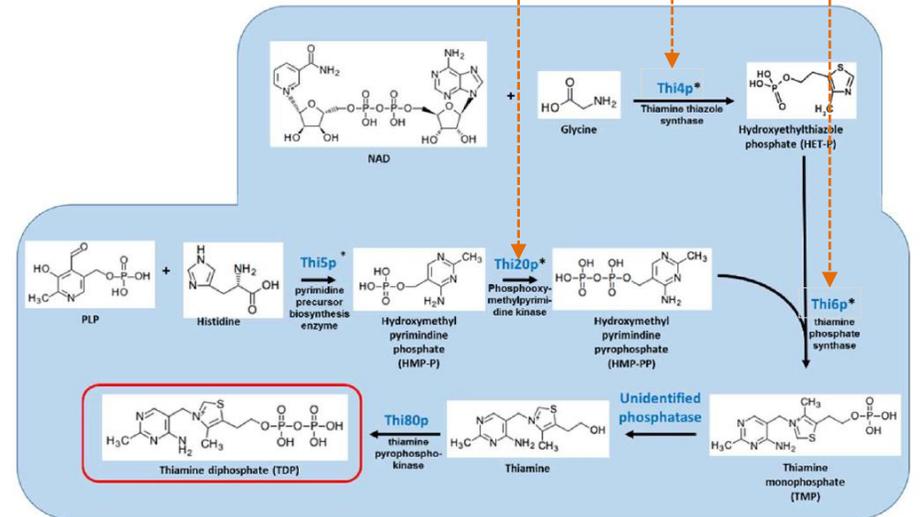
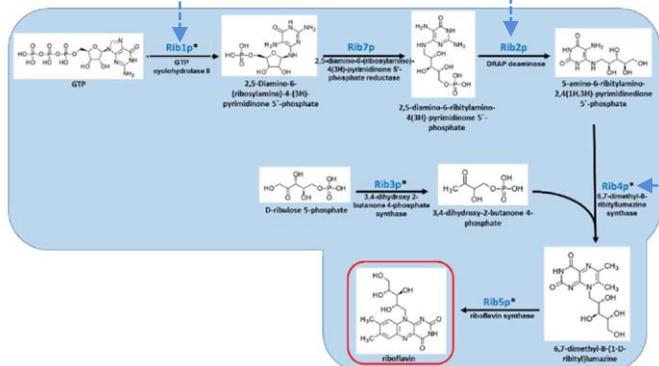
**Abstract:** The large increase in the population of immunosuppressed patients, coupled with the limited efficacy of existing antifungals and rising resistance toward them, have dramatically highlighted the need to develop novel drugs for the treatment of invasive fungal infections. An attractive possibility is the identification of possible drug targets within essential fungal metabolic pathways not shared with humans. Here, we review the vitamin biosynthetic pathways (vitamins A–E, K) as candidates for the development of antifungals. We present a set of ranking criteria that identify the vitamin B2 (riboflavin), B5 (pantothenic acid), and B9 (folate) biosynthesis pathways as being particularly rich in new antifungal targets. We propose that recent scientific advances in the fields of drug design and fungal genomics have developed sufficiently to merit a renewed look at these pathways as promising sources for the development of novel classes of antifungals.

**Keywords:** antifungals; fungal vitamin metabolism; drug target; essential genes

Table 1. Evaluation of the suitability of fungal vitamin B biosynthetic genes for antifungal development.

Pathway	Gene *	Essential for Fungal Virulence	FUNGAL CRYSTAL STRUCTURE	Fungal Inhibitors Developed
Vitamin B1 thiamine	THI4	No [33]	Yes [34]	No
	THI5	ND <sup>†</sup>	Yes [35]	No
	THI6	No [27]	Yes [36]	No
	THI20	ND	Yes [37]	No
Vitamin B2 riboflavin	RIB1	Yes [27,33]	No	No
	RIB2	Yes [9,38]	No	No
	RIB3	ND	Yes [39]	No
	RIB4	ND	Yes [40]	Yes [42]
	RIB5	ND	Yes [41]	No
Vitamin B5 pantothenate				
Vitamin B6 pyridoxine				
Vitamin B7 biotin				
Vitamin B9 folate				

\* *Saccharomyces cerevisiae*



# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ 농식품 빅데이터 플랫폼을 활용한 농업 기술과 ICT와의 융합

- 농식품의 생산정보, 식품정보 바이오정보는 계속해서 증가하고 있지만, 이들 정보간의 연계 부재
- 농업의 첨단화를 위해 정밀 농업 빅데이터와 ICT 간의 정보 연계 필요성 급증

### 빅데이터 & ICT & 농업의 만남

농산물 직황 점검, 제초제, 살충제 및 비료살포, 과실 성숙도 체크 등  
농업용 드론

트랙터-스마트폰-위성을 연결하는 GPS로 농작물 관련 정보를 실시간으로 수집 및 분석자료 제공  
GPS와 IoT 기술

토양 수분, 영양분, 염의 농도를 실시간으로 수집 및 확인 가능  
넓은 토양도 적은 비용으로 측정 가능  
RFID센서와 극소형 칩



### 영주 사과 작황 정보 서비스

HOME | LOGIN | JOIN | SITEMAP | 영주폭도 | 영주시

영주시농업 빅데이터기반 영주사과작황정보

사과작황정보 | 병해충정보 | 영주생활정보 | 커뮤니티 | 서비스안내

농업 빅데이터를 수집하여 병해충 징후 및 사과 생산 추이, 가격동향 서비스를 제공합니다

Web 515 u-IT

u-IT 정보

온도	습도	도랑수분
25.3	13.0	28.8

태풍발생 0

기상청 정보

24.5°C

비량 누적 2.1m/h  
습도 15.0%  
강수량 0.0mm

시정뉴스

영주시 관련 뉴스들을 통합적으로 소개합니다.

시정뉴스

영주시에 상여있는 볼거리를 생생하게 알려드립니다.

시정뉴스

과수원정보

국내 최대 사과생산지 영주시의 과수원을 소개합니다.

시정뉴스

도농상생일자리

영주시 일자리 관련 정보가 한가득 구민구청을 도모드립니다.

시정뉴스

사과 공공데이터 | Apple Public Data

가격 동향

생산량 추이

병해충 징후

추진 20kg 상자 상여값

가격 3년 평균 1년전

가격(원)

0329 0406 0412 0419

날짜

주관 가격 동향 분석

사과 가격이 안정세를 보이면서 가격이 하락할 것으로 보입니다.

영주시과

필과 SNS에서 수집한 영주시과 관련 데이터를 분석한 결과 영주시과와 깊은 연관을 가지는 단어는 나주배, 경상북도, 맛, 알몸 등으로 나타났습니다.

사과 민간데이터 | 사과 관련 뉴스도

영주시과

가격

생산

병해충

직판행사이전 상품 선정 APC

가격 경상북도 수상

대박 판매맛 명품 좋은 과일

사랑 서울 사과 흥동백서

특산물 행사 나주배 홍보 축제

내용 아이러브영 주사과 유령

가득 와인 산들네 역기스

발전 판매스 농장 연결 취지

영주시과

필과 SNS에서 수집한 영주시과 관련 데이터를 분석한 결과 영주시과와 깊은 연관을 가지는 단어는 나주배, 경상북도, 맛, 알몸 등으로 나타났습니다.

입력: 2015년 10월 20일

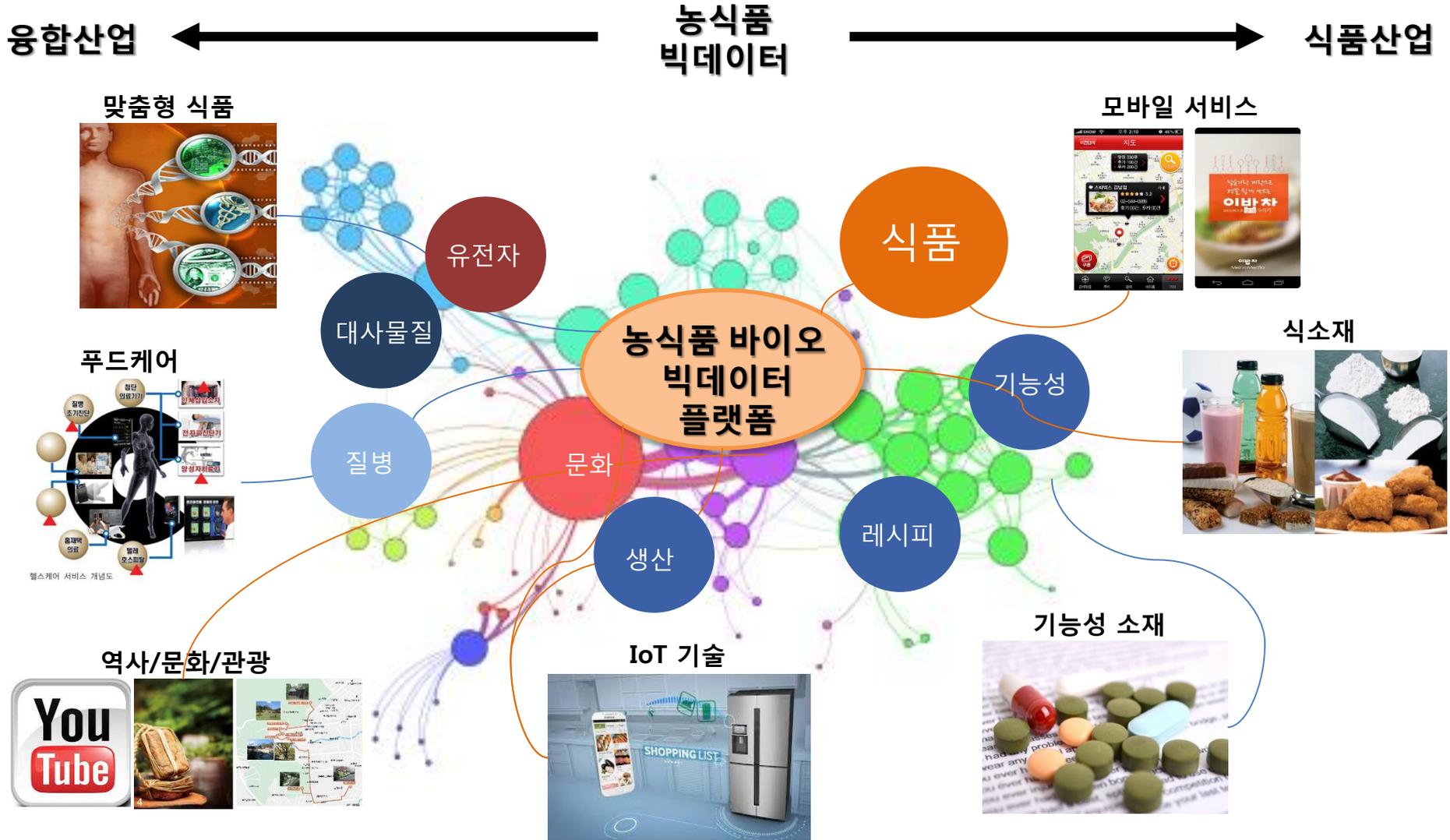
대행: +

병해충 시기

대행: +

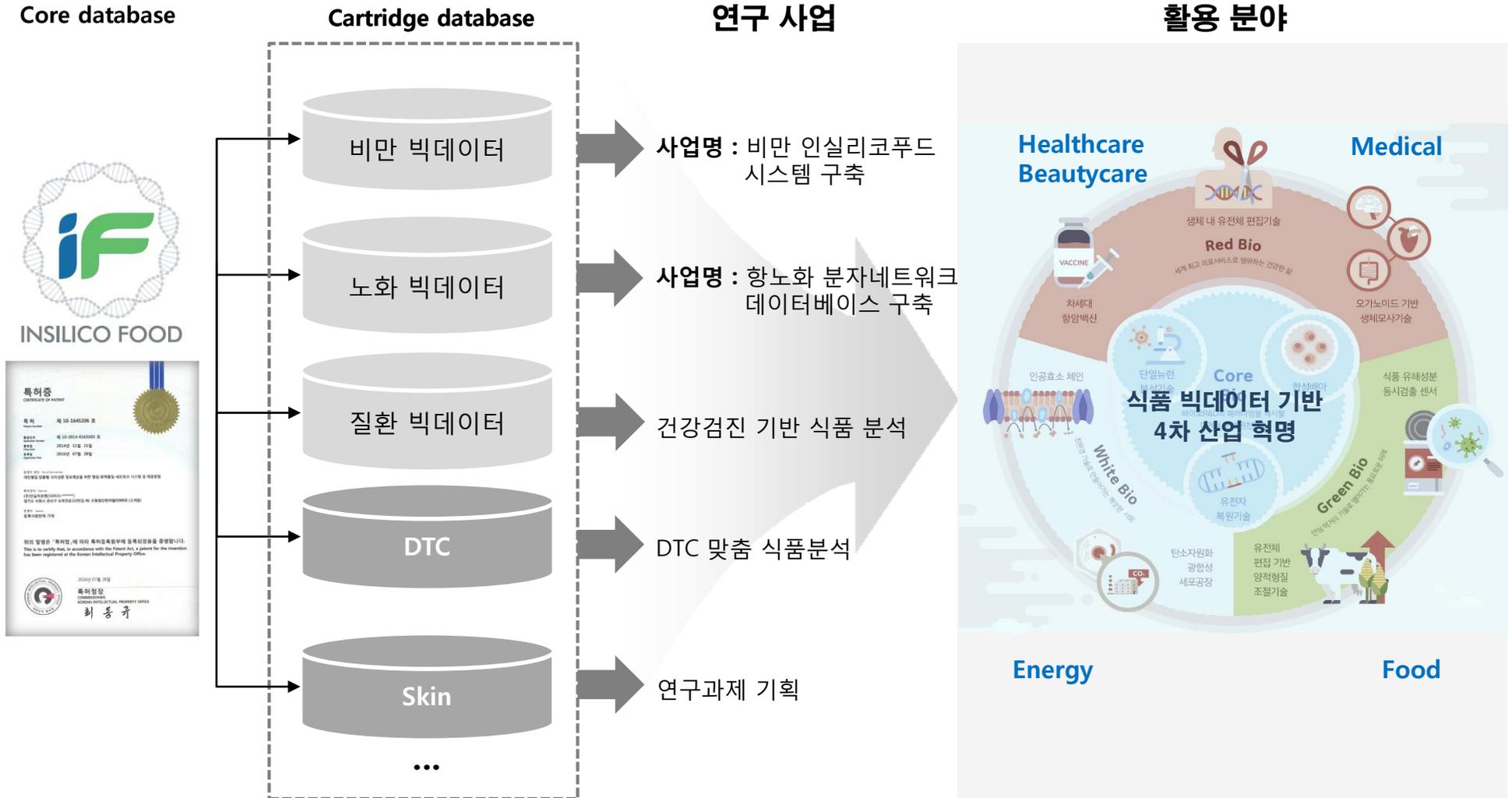
출처: 2015년 전북도청

# 바이오 빅데이터를 활용한 식품 산업 연계



# 바이오 빅데이터를 활용한 식품 산업 연계

## ■ 연구 개발 분야에서의 식품 빅데이터 활용



thank  
you

*Bioinformatics*

DEEP IN BIG

[www.insilicogen.com](http://www.insilicogen.com) E-mail [info@insilicogen.com](mailto:info@insilicogen.com) Tel 031-278-0061 Fax 031-278-0062