SONY
Sony CSL

SHOE

Sequence HOmology in higher Eukaryotes

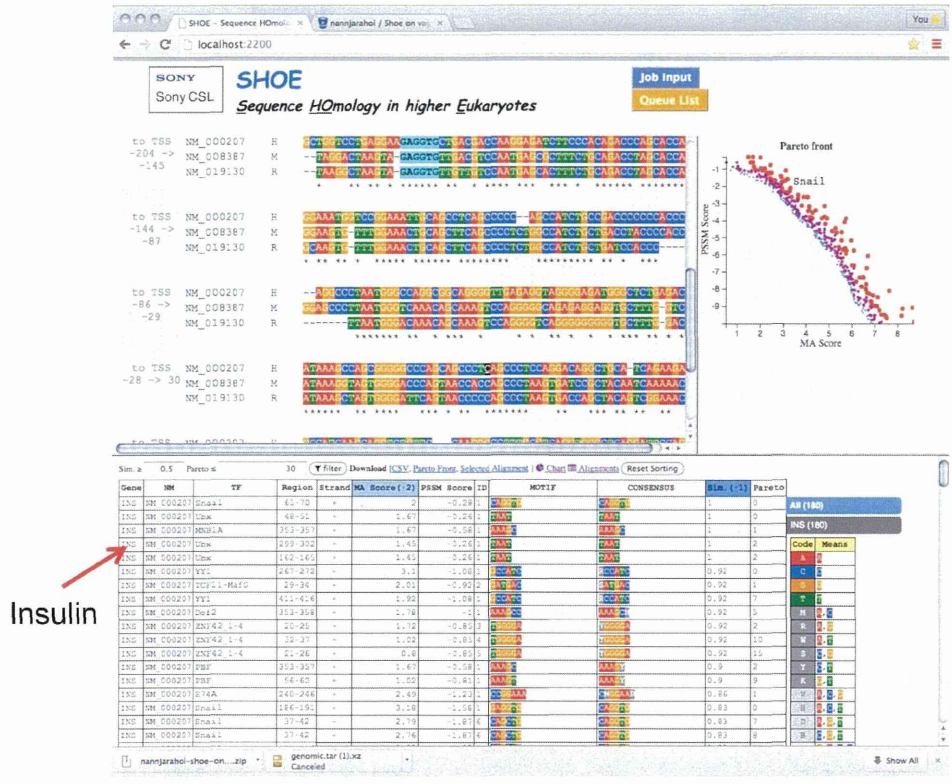
Job Input
Queue List

Compare

positive #1 mTOR (H.sapiens) control #2 mTOR (H.sapiens) compare

Results

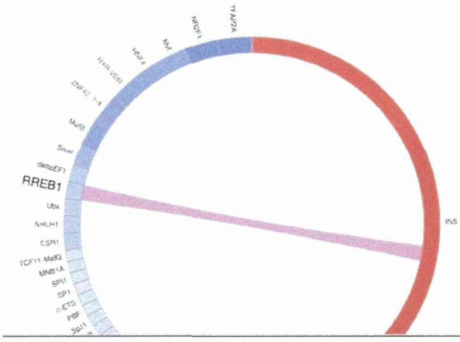
Job	Date	Gene List	Repeat Mask	UpStream	Mode	DownStream	Scoring	Pareto	Delete
#1 mTOR (H.sapiens)	2015/12/16 16:34:03 ↓ 2015/12/16 16:34:40	INS (458)	yes	2000	mode2	200	jaspar	30	delete
#2 mTOR (H.sapiens)	2015/12/16 16:36:58 ↓ 2015/12/16 17:01:38	INS (1) IGF1 (1) IRS1 (116) PIK3R2 (33) PIK3R3 (53) PTEN (3) PDPK1 (135) AKT3 (3) RRAGA (16) HIF1A (24) RPS6KB1 (27) RPS6KB2 (10) EIF4B (48) EIF4EBP1 (54) EIF4E (1) MAPK1 (5) MAPK3 (8)	yes	2000	mode2	200	transfac32	30	delete



SONY
Sony CSL

SHOE
Sequence HOmology in higher Eukaryotes

Job Input
Queue List

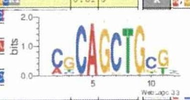


Sim. ≥ 0.5 Pareto ≤ 30 filter Download [CSV, Pareto Front, Selected Alignment] | Chart Alignments Reset Sorting

Gene	NM	TF	Region	Strand	MA Score	PSSM Score	ID	MOTIF	CONSENSUS	Sim.	Pareto
INS	NM_000207	deltaEF1	65-70	-	2.43	-0.74	1	CAGGTG	CACCTD	0.56	0
INS	NM_000207	Snail	186-191	-	3.18	-1.58	1	GAGGTG	CAGGTG	0.83	0
INS	NM_000207	Ubx	48-51	-	1.67	-0.26	1	TAAT	TAAT	1	0
INS	NM_000207	TFAP2A	106-114	-	4.06	-2.64	1	GCCCCCAGC	GCCBBRRCS	0.63	0
INS	NM_000207	Myf	266-277	-	5.96	-5.15	1	AGCCATCTGCCG	MRRCARCWGSWG	0.58	0
INS	NM_000207	SPIB	51-57	-	3.29	-2.1	1	GTTGGAA	WSMGGAA	0.71	0
INS	NM_000207	Snail	65-70	-	2	-0.28	1	CAGGTG	CAGGTG	1	0
INS	NM_000207	RELA	242-251	-	4.79	-3.15	1	GGAAATTGCA	GGGRATTTCC	0.65	0
INS	NM_000207	YY1	267-272	-	3.1	-1.08	1	GCCATC	RCCATC	0.92	0
INS	NM_000207	NHLH1	108-119	-	5.73	-3.87	1	CCCCAGCTGTGA	YMGCAGCTGCKY	0.61	0
INS	NM_000207	RREB1	210-229	-	8.72	-9.89	1	TTCCCACAGACCCAGCACCA	MCMCHAMNC		
INS	NM_000207	E74A	240-246	-	2.49	-1.23	1	CCGGAAA	CMGGAAR		
INS	NM_000207	IRF2	242-259	-	8.21	-9.13	1	GGAAATTGCAGCCTCAGC	SGAAAGYGA		
INS	NM_000207	MNB1A	353-357	-	1.67	-0.58	1	AAAGC	AAAGC		
INS	NM_000207	ESR1	131-150	-	6.76	-8.25	1	TCTGGCCACC--GGGCCCCT	CCAGGTCAC		
INS	NM_000207	TCF11-MafG	29-34	-	2.01	-0.92	2	GATGAC	SATGAC		
INS	NM_000207	REL	242-251	-	4.79	-3.79	1	GGAAATTGCA	SGGRNWTTCC	0.5	1
INS	NM_000207	Snail	225-230	-	3.17	-1.58	1	CACCAG	CAGGTG	0.5	1

All (180)
INS (180)

Code	Means
A	A
C	C
G	G
T	T
M	A, C
R	A, G
D	A, G, T
B	C, G, T

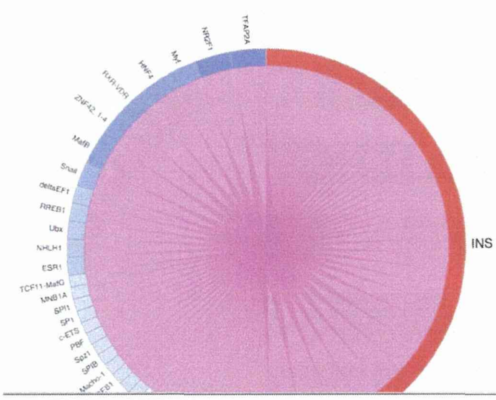


19

SONY
Sony CSL

SHOE
Sequence HOmology in higher Eukaryotes

Job Input
Queue List



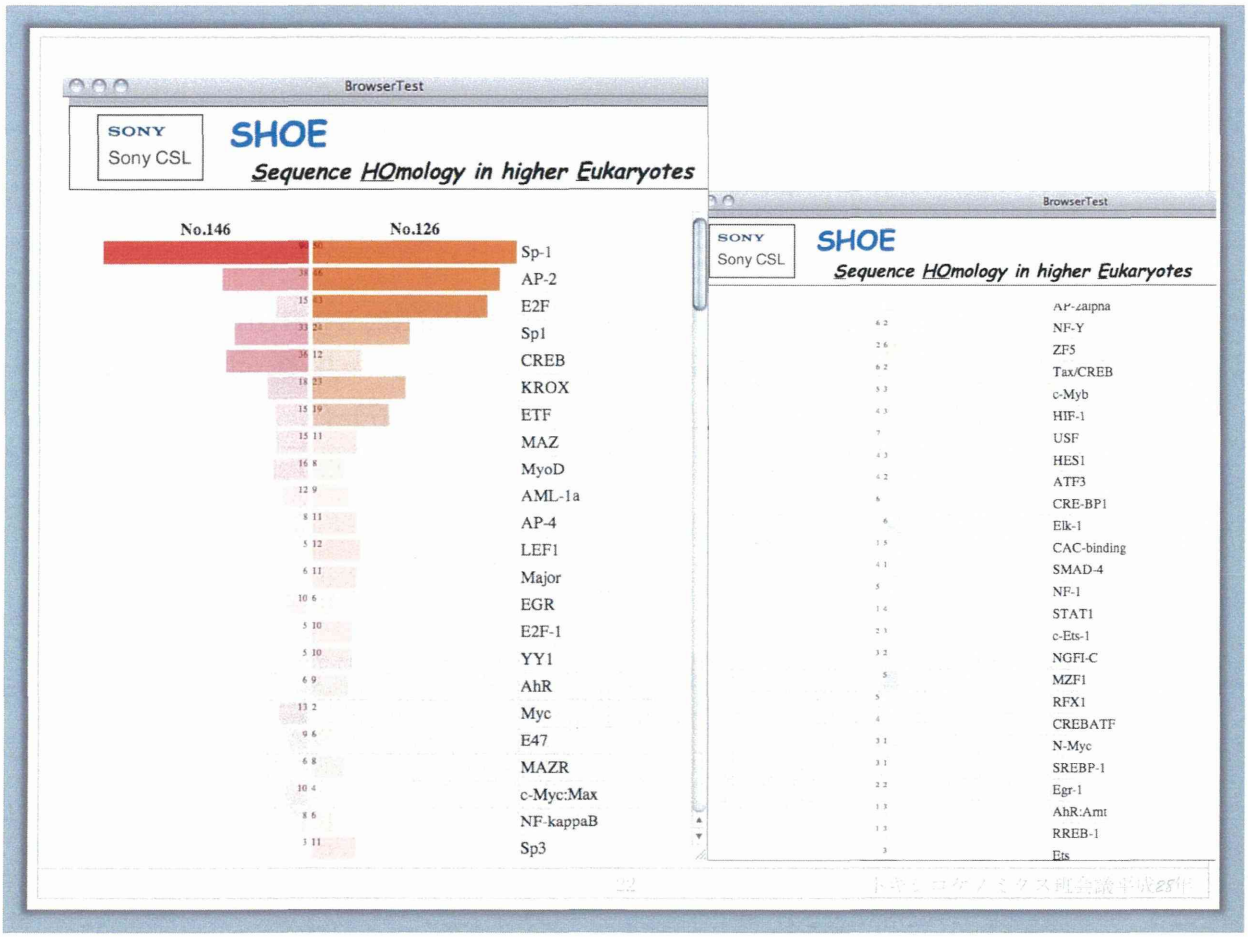
Sim. ≥ 0.5 Pareto ≤ 30 filter Download [CSV, Pareto Front, Selected Alignment] | Chart Alignments Reset Sorting

Gene	NM	TF	Region	Strand	MA Score	PSSM Score	ID	MOTIF	CONSENSUS	Sim. (·1)	Pareto
INS	NM_000207	MNB1A	353-357	+	1.67	-0.58	1	AAAGC	AAAGC	1	1
INS	NM_000207	Ubx	299-302	+	1.45	-0.26	1	TAAT	TAAT	1	2
INS	NM_000207	Snail	65-70	+	2	-0.28	1	CAGGTG	CAGGTG	1	0
INS	NM_000207	Ubx	162-165	+	1.45	-0.26	1	TAAT	TAAT	1	2
INS	NM_000207	Ubx	48-51	+	1.67	-0.26	1	TAAT	TAAT	1	0
INS	NM_000207	YY1	411-416	+	1.92	-1.08	1	GCCATC	RCCATC	0.92	7
INS	NM_000207	ZNF42_1-4	21-26	+	0.8	-0.85	5	TGGGGA	YGGGGA	0.92	15
INS	NM_000207	ZNF42_1-4	32-37	+	1.02	-0.85	4	TGGGGA	YGGGGA	0.92	10
INS	NM_000207	YY1	267-272	+	3.1	-1.08	1	GCCATC	RCCATC	0.92	0
INS	NM_000207	Dof2	353-358	+	1.78	-1	1	AAAGCC	AAAGCM	0.92	5
INS	NM_000207	TCF11-MafG	29-34	+	2.01	-0.92	2	GATGAC	SATGAC	0.92	1
INS	NM_000207	ZNF42_1-4	20-25	+	1.72	-0.85	3	TGGGGA	YGGGGA	0.92	2
INS	NM_000207	PBF	353-357	+	1.67	-0.58	1	AAAGC	AAAGY	0.9	2
INS	NM_000207	PBF	56-60	+	1.02	-0.81	1	AAAGT	AAAGY	0.9	9

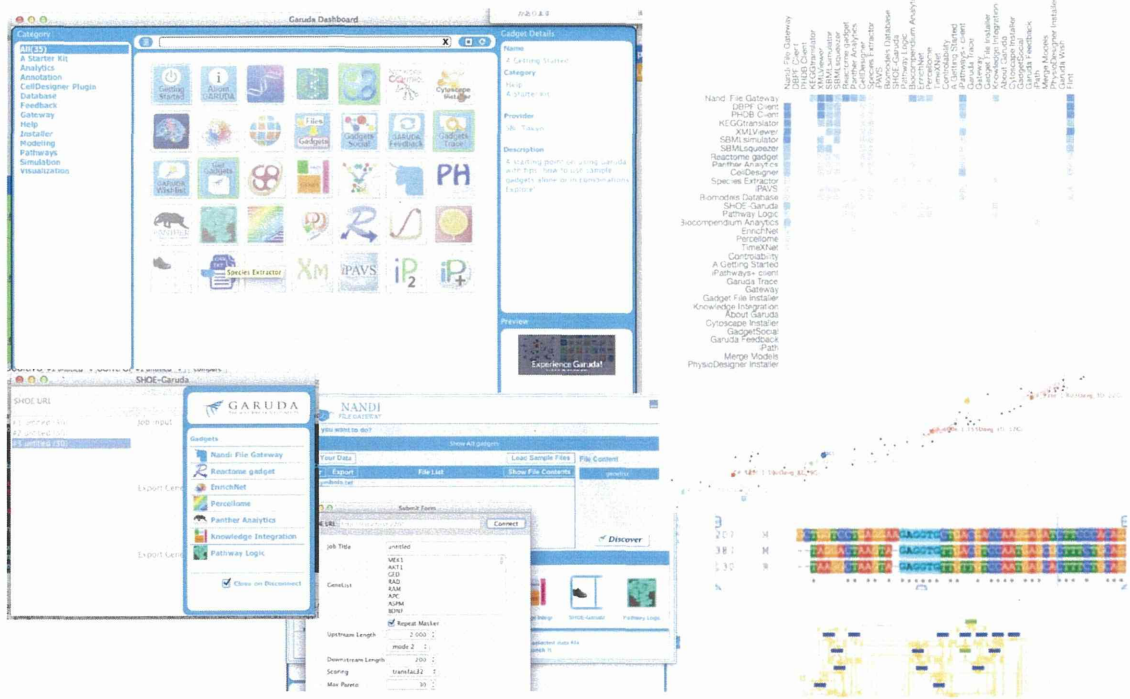
nannjarahoi-shoe-on....zip
genomic.tar (1).xz
Canceled
Show All

20


BrowserTest
SONY
Sony CSL
SHOE
Sequence HOmology in higher Eukaryotes
Job Input
Queue List
User Input 236
In SHOE 73
Multiple alignment 64
PSSM 61
Transcription Factor 113
Download


BrowserTest
SONY
Sony CSL
SHOE
Sequence HOmology in higher Eukaryotes
No.146
No.126
Sp-1
AP-2
E2F
Sp1
CREB
KROX
ETF
MAZ
MyoD
AML-1a
AP-4
LEF1
Major
EGR
E2F-1
YY1
AhR
Myc
E47
MAZR
c-Myc:Max
NF-kappaB
Sp3
BrowserTest
SONY
Sony CSL
SHOE
Sequence HOmology in higher Eukaryotes
NF-Y
ZF5
Tax/CREB
c-Myb
HIF-1
USF
HES1
ATF3
CRE-BP1
Elk-1
CAC-binding
SMAD-4
NF-1
STAT1
c-Ets-1
NGFI-C
MZF1
RFX1
CREBATF
N-Myc
SREBP-1
Egr-1
AhR:Amt
RREB-1
Ets

SHOE is connected to Garuda platform



自動で行う環境構築

- Makefile, makeに対応する形でVagrantfile, vagrant upなどのコマンドを使う。設定ファイルに必要なツールを自動ダウンロードができるようになっている。所定OSのインストールまで含めてすべて自動でやってくれる
  - OSイメージのインストールすらユーザ操作の必要がない
  - 開発環境を数行のコマンドを打つだけで終わらせる

```
$ git clone https://github.com/10up/varying-vagrant-vagrants.git
$ cd varying-vagrant-vagrants
$ vagrant up
(please wait for about twenty minuites)
$ open http://192.168.50.4/
```

Future plans

- Create Plugin to CellDesigner to make a workflow
  - Percellome ->SHOE Transcrption Regulation analysis ->CellDesigner transcription regulation map (in progress)
  - Percellome -> AGCT Cluster Analysis->SHOE Transcrption Regulation analysis -> Transcription regulation map on CellDesigner
- Establish connections with other Garuda gadgets (pathway, literature mining)

Collaborators

- Richard Nock (NICTA & the Australian National University, Canberra, Australia)
- Frank Nielsen (Sony Computer Science Laboratories Inc.)
- Kazuhiro Shibanai (Tokyo Institute of Technology)
- Kodai Takata (Tokyo Institute of Technology)
- SBI Garuda Team

Xsight: an ensemble machine learning based platform for big data analysis in toxicology studies

The Systems Biology Institute
SBX

We developed, "Xsight", a computational learning system that integrates a large number of machine-learning models for big data analysis

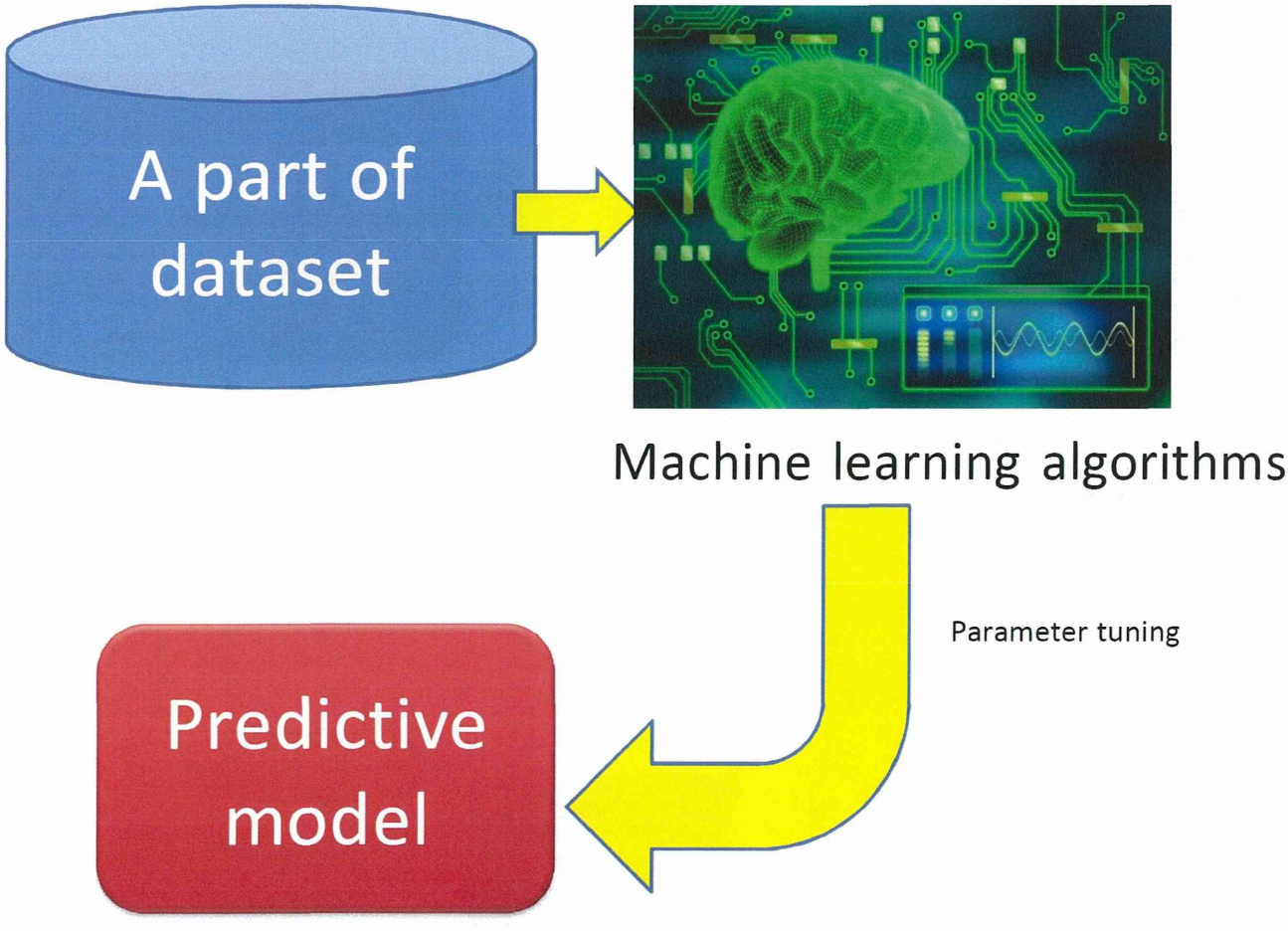
Machine learning

- Procedure of machine learning is composed of two steps.

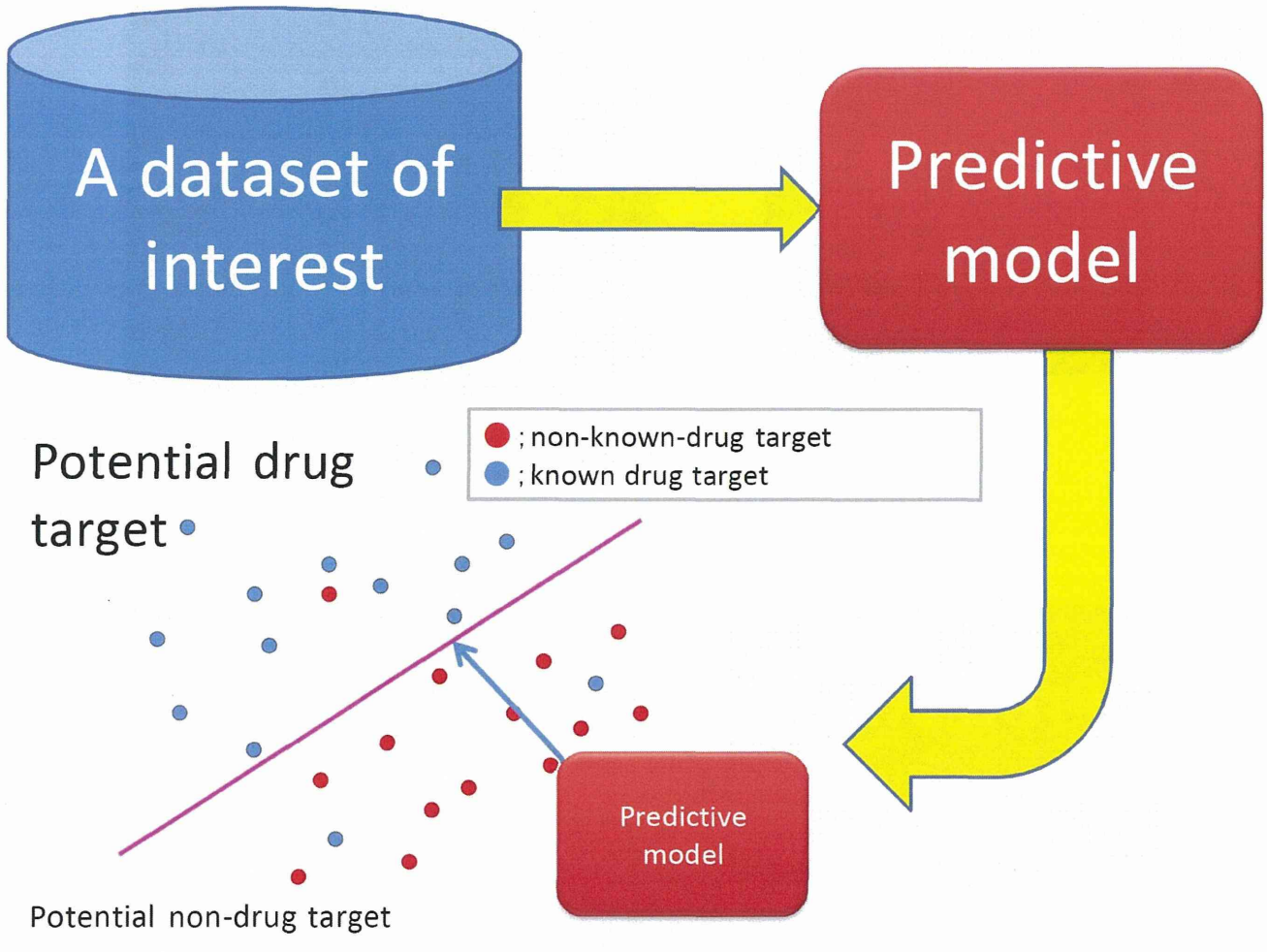
  Step 1: **Training; Use a part of data to build predictive model**

  Step 2: **Test and Prediction;** Using remaining dataset to test their predictive performance. Then, **apply predictive model to a dataset of interest to make predictions, i.e., inference of potential side effect of a compound as well as potential candidate drug targets**

Training: Model building



Prediction: prediction based on models obtained from step 1



A large number of machine learning techniques based on various statistical techniques.

One machine learning method will generate one predictive model.

Bagged_CART	Multivariate_Adaptive_Regression_Splines
Bagged_Flexible_Discriminant_Analysis	Naive_Bayes
Boosted_Generalized_Additive_Model	Nearest_Shrunken_Centroids
Boosted_Generalized_Linear_Model	Neural_Network
Boosted_Linear_Model	Neural_Networks_with_Feature_Extraction
Boosted_Logistic_Regression	Oblique_Trees
Boosted_Tree	partDSA
CART	Partial_Least_Squares
Conditional_Inference_Random_Forest	Penalized_Discriminant_Analysis
Conditional_Inference_Tree	Penalized_Linear_Discriminant_Analysis
Cost-Sensitive_CART	Penalized_Logistic_Regression
Extreme_Learning_Machine	Penalized_Multinomial_Regression
Flexible_Discriminant_Analysis	Quadratic_Discriminant_Analysis
Gaussian_Process	Quadratic_Discriminant_Analysis_with_Stepwise_Feature_Selection
Generalized_Additive_Model_using_Splines	Radial_Basis_Function_Network
Generalized_Linear_Model	Random_Forest
Generalized_Linear_Model_with_Stepwise_Feature_Selection	Regularized_Random_Forest
Generalized_Partial_Least_Squares	Robust_Regularized_Linear_Discriminant_Analysis
Glmnet	Robust_SIMCA
Greedy_Prototype_Selection	ROC-Based_Classifier
High_Dimensional_Discriminant_Analysis	Self-Organizing_Maps
k-Nearest_Neighbors	Shrinkage_Discriminant_Analysis
Learning_Vector_Quantization	SIMCA
Least_Squares_Support_Vector_Machine_with_Radial_Basis_Function_Kernel	Sparse_Partial_Least_Squares
Linear_Discriminant_Analysis	Stabilized_Linear_Discriminant_Analysis
Linear_Discriminant_Analysis_with_Stepwise_Feature_Selection	Stochastic_Gradient_Boosting
Maximum_Uncertainty_Linear_Discriminant_Analysis	Support_Vector_Machines_with_Class_Weights
Mixture_Discriminant_Analysis	Support_Vector_Machines_with_Linear_Kernel
Model_Averaged_Neural_Network	Support_Vector_Machines_with_Radial_Basis_Function_Kernel
Multi-Layer_Perceptron	Gaussian_Process_with_Radial_Basis_Function_Kernel
Multivariate_Adaptive_Regression_Spline	Sparse_Linear_Discriminant_Analysis

Ensemble learning

Integrating a larger number of predictive model from a large number of machine learning methods to make predictions