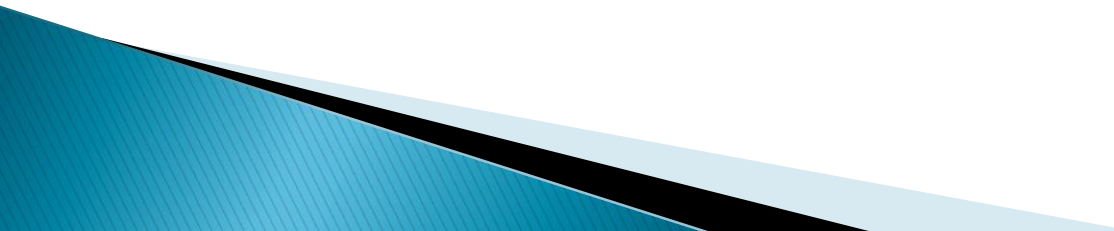


第13回バイオインフォマティクス研究会

マイクロアレイデータの2群間比較
発現量による補正法

今日の内容

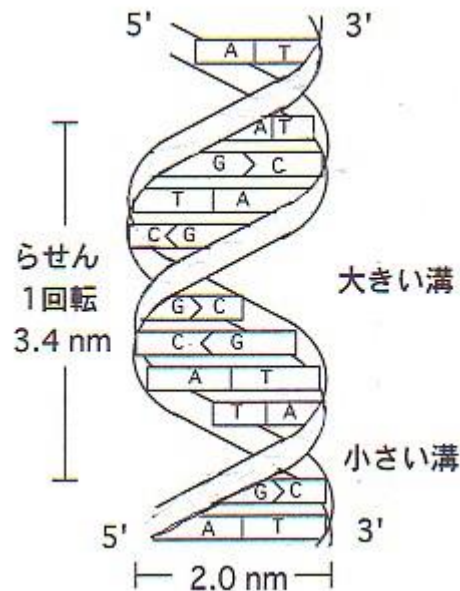
- ▶ バイオインフォマティクス
 - ▶ 二群間比較
 - ▶ ROC曲線（識別力の判定）
 - ▶ マイクロアレイデータの二群間比較（WAD法）
 - ▶ 具体例（HOXA9 shRNA vs control）
- 

バイオインフォマティクス

生物学＋情報科学

生命の持つ情報を情報科学の方法論で解析する学問分野

塩基配列 ATGTCGCAT...



出現頻度

A→0.25

T→0.25

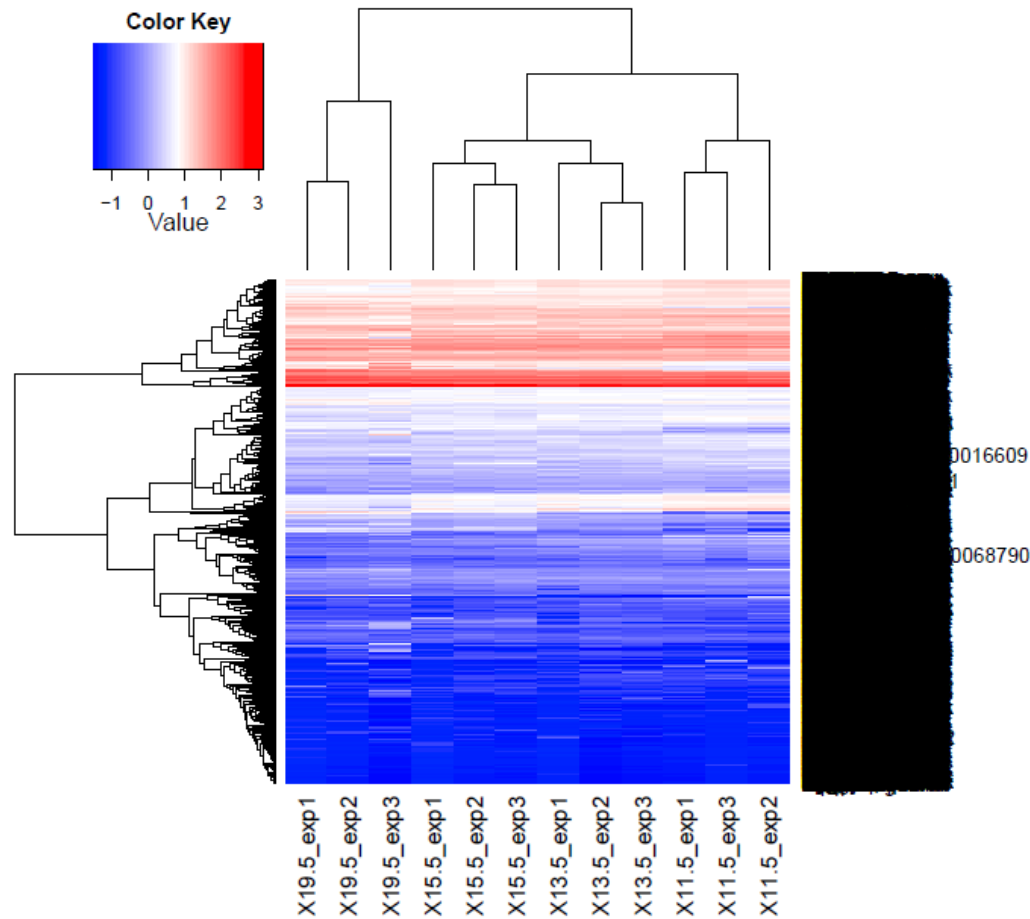
G→0.25

C→0.25

情報としての価値が低い

頻度の偏りを手掛かりに生物学的な意義や機能を探る

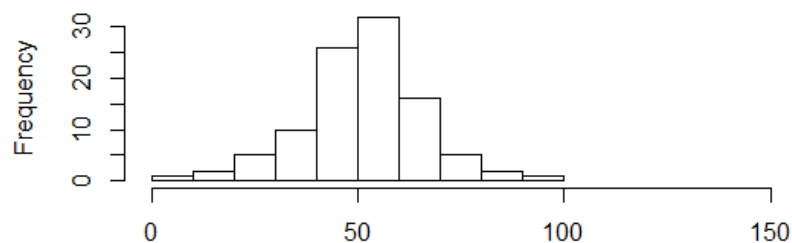
実験のハイスループット化



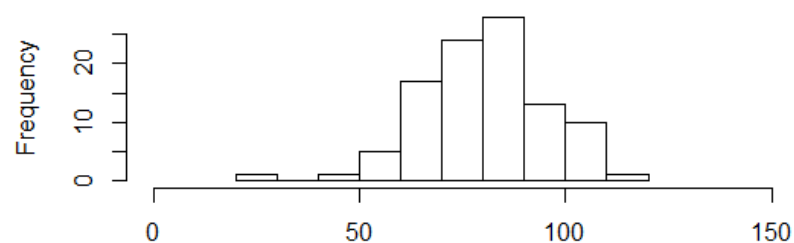
より高度な情報処理が必要

二群間比較

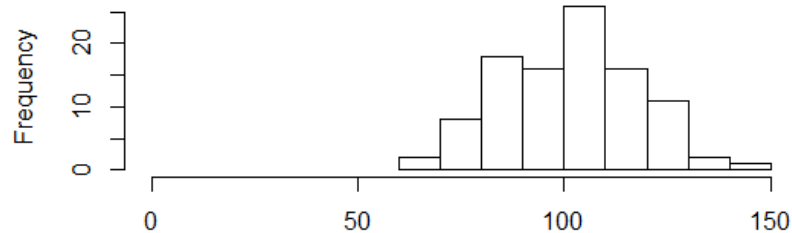
Protin X in condition A



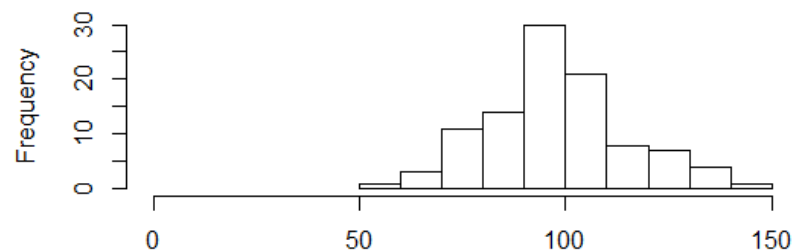
Protin Y in condition A



Protin X in condition B

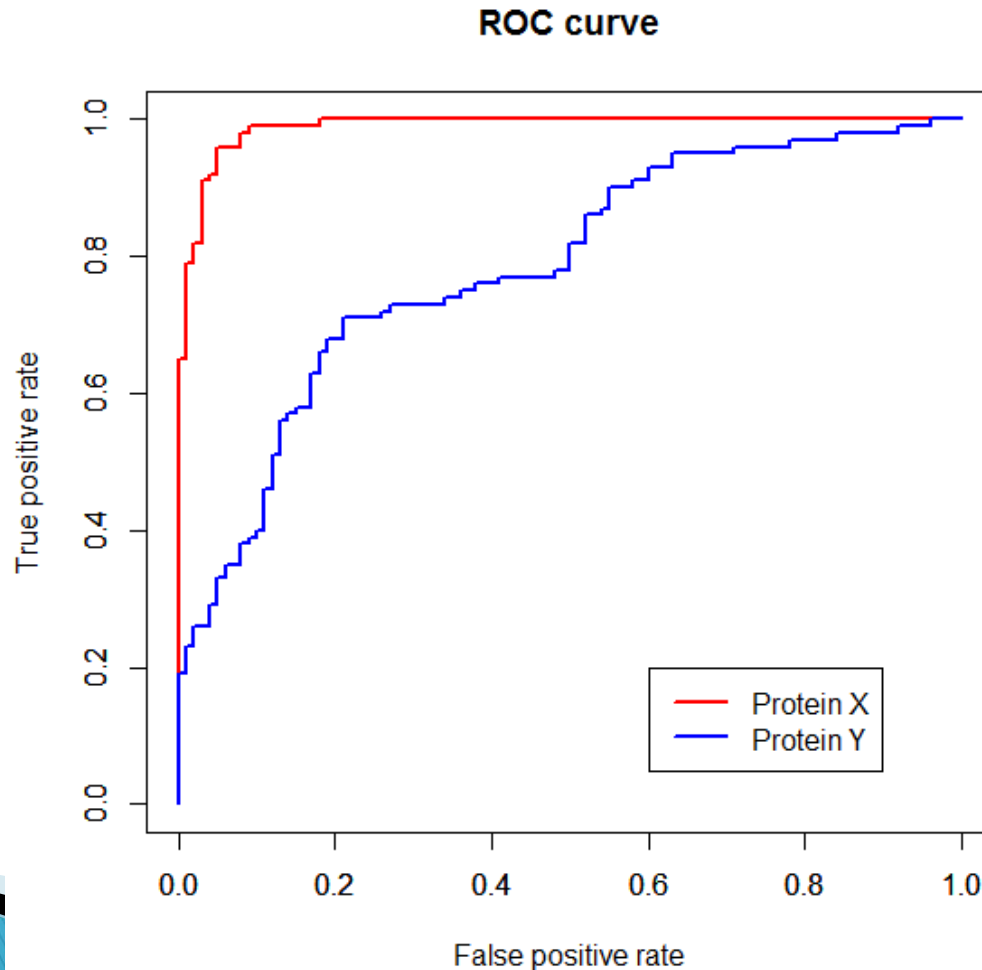


Protin Y in condition B

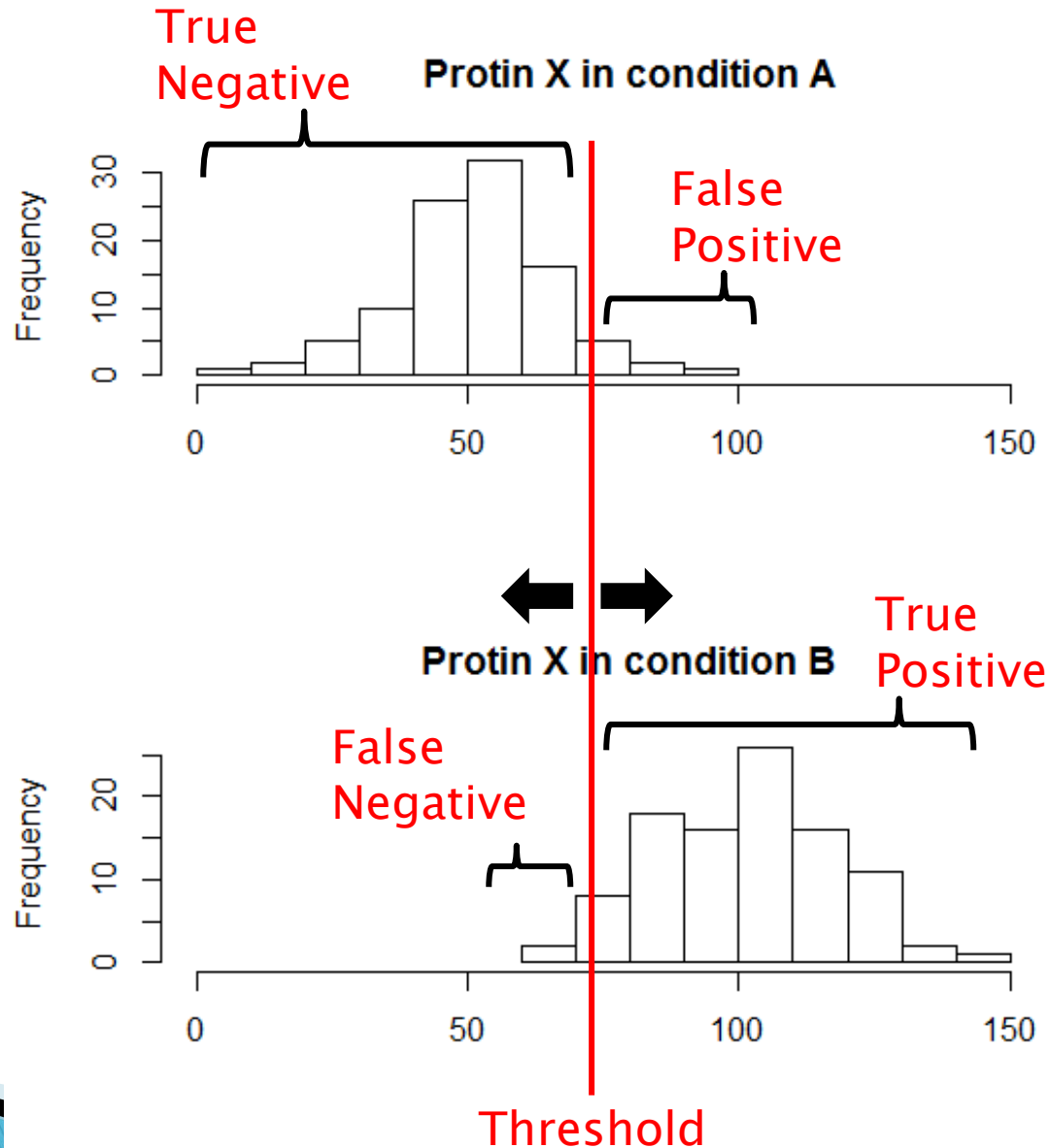


Receiver Operating Characteristic (ROC) curve

観測された信号からあるものの存在を判定する際の基準となる特性

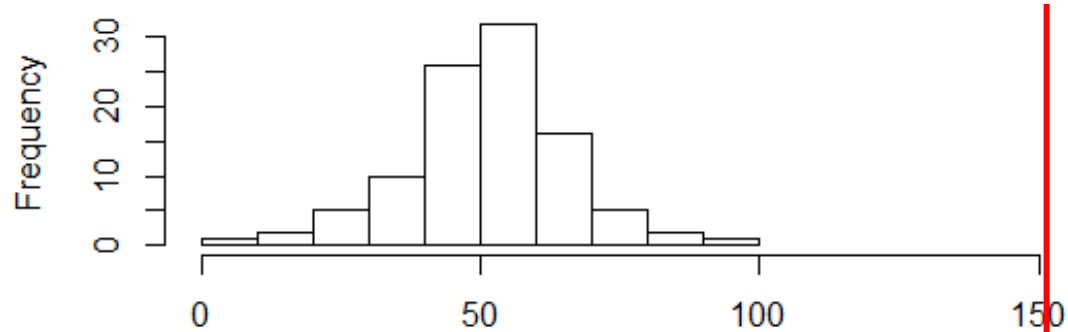


ROC curve



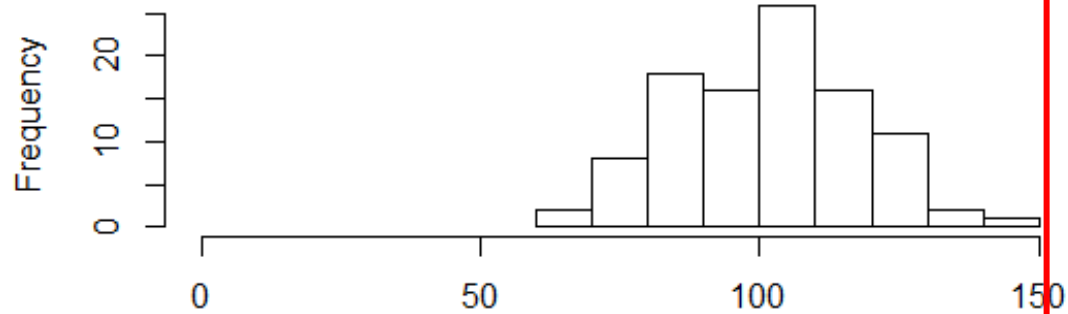
ROC curve

Protin X in condition A



$FP = 0$

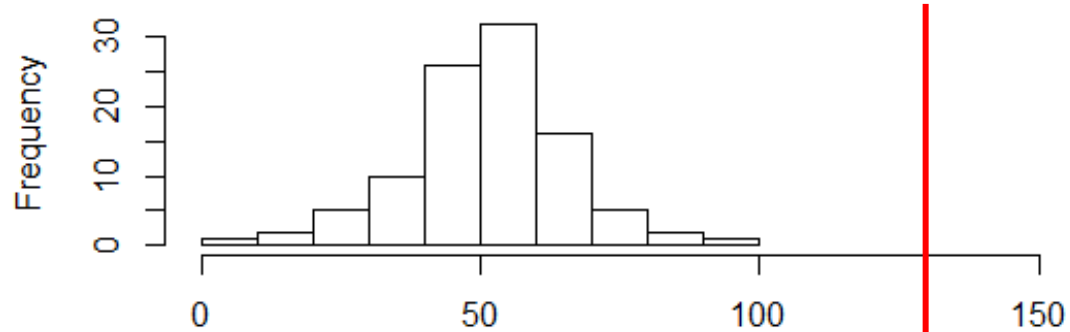
Protin X in condition B



$TP = 0$

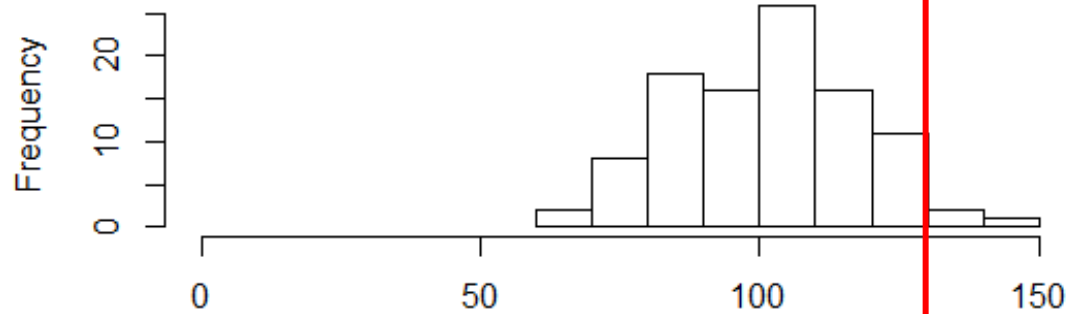
ROC curve

Protein X in condition A



$FP = 0$

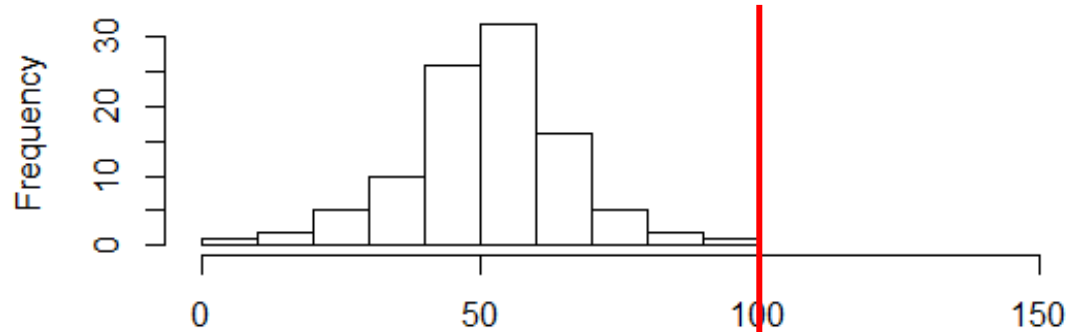
Protein X in condition B



$TP > 0$

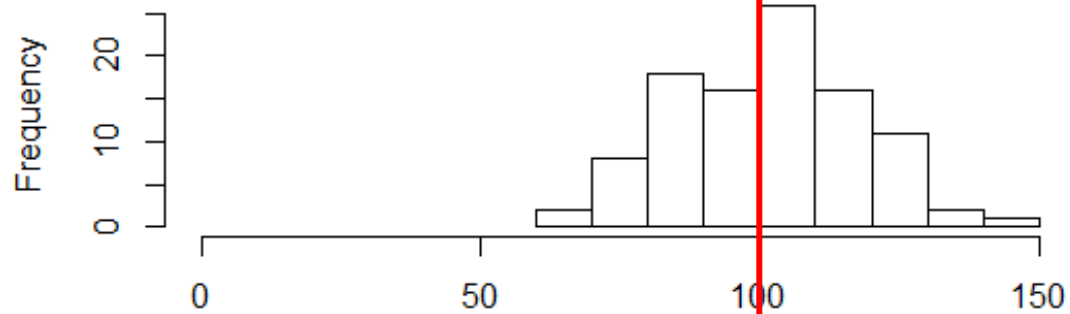
ROC curve

Protein X in condition A



$FP = 0$

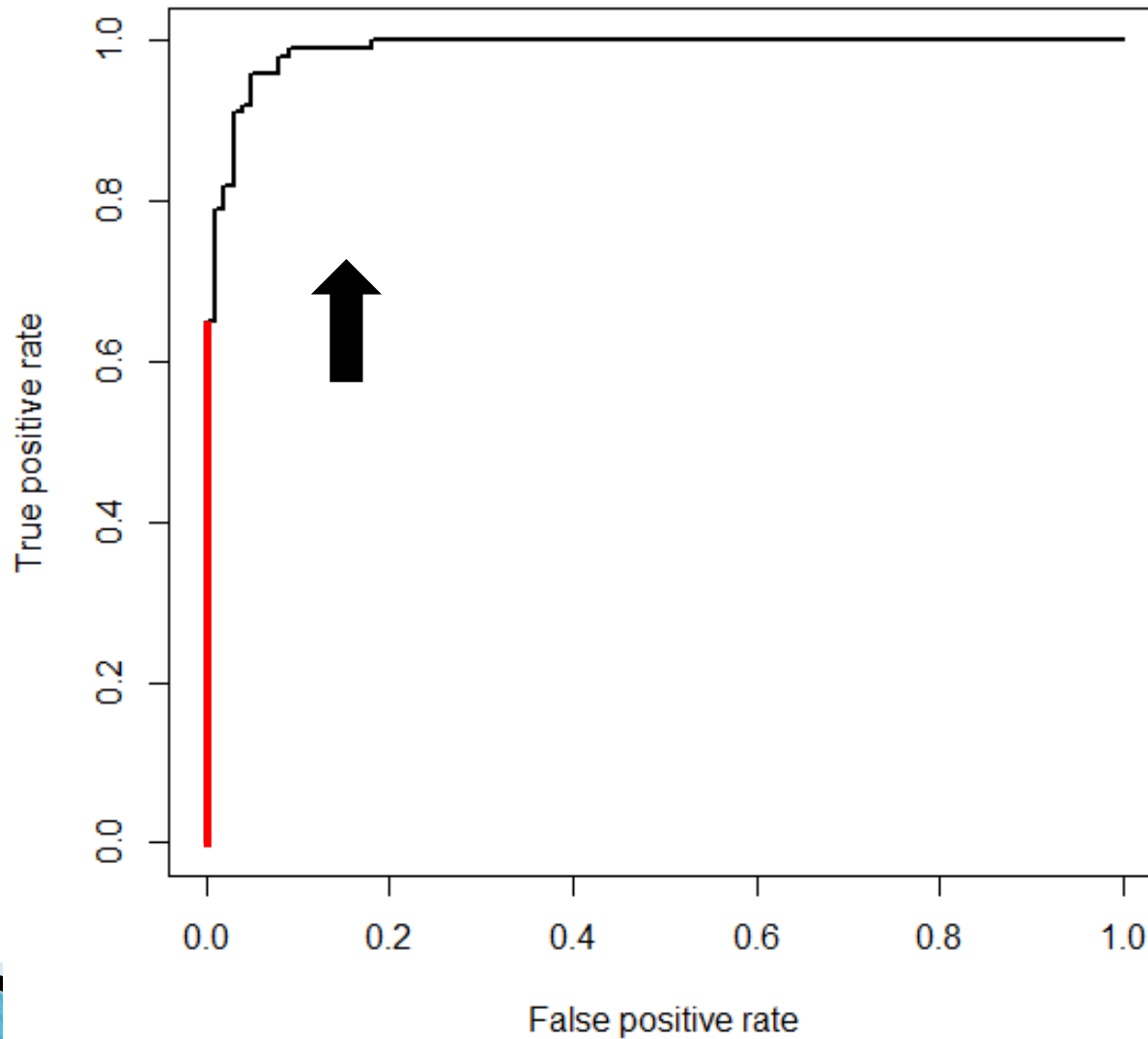
Protein X in condition B



$TP > 0$

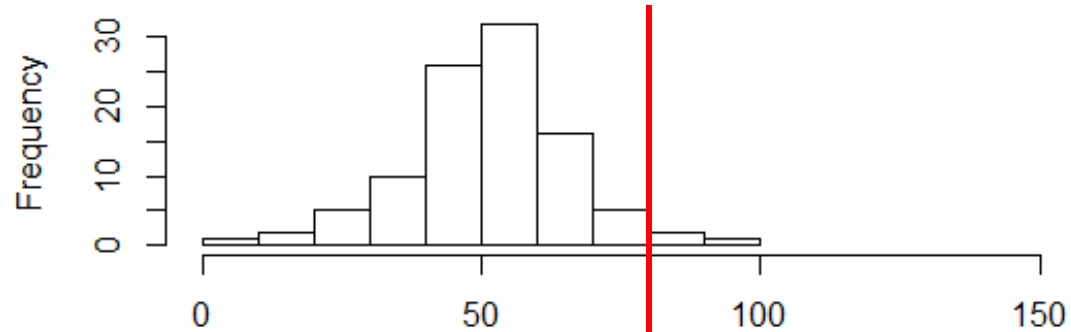
ROC curve

ROC curve



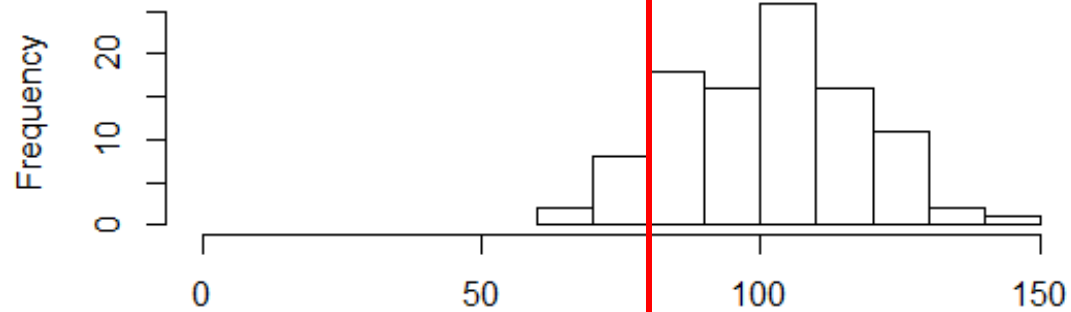
ROC curve

Protein X in condition A



FP > 0

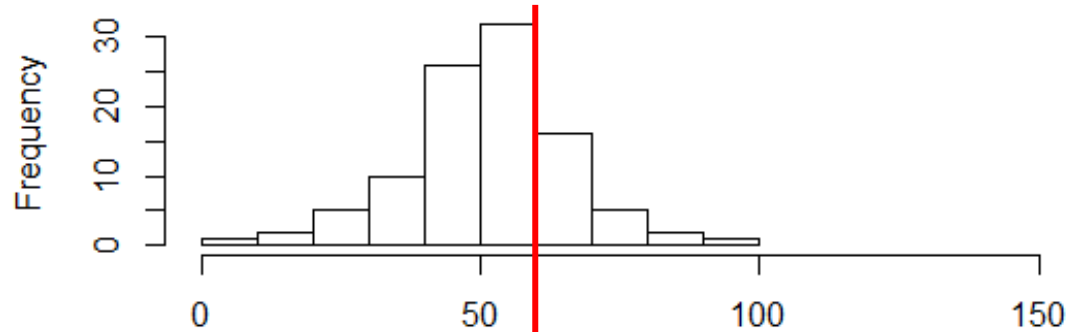
Protein X in condition B



TP > 0

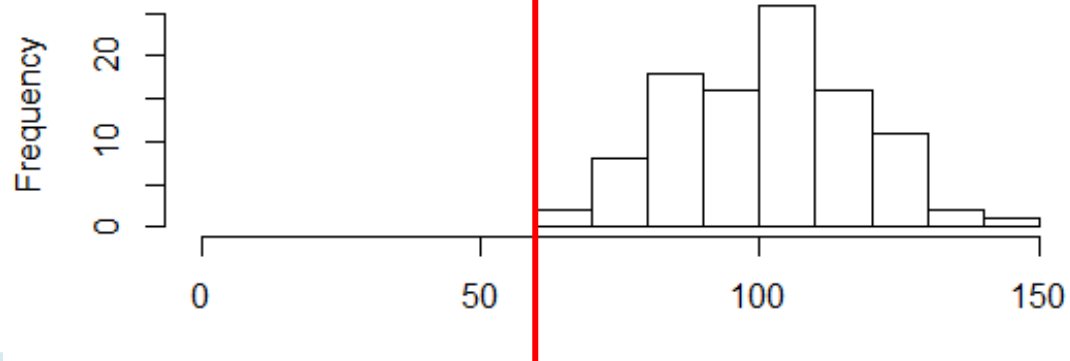
ROC curve

Protin X in condition A



$FP > 0$

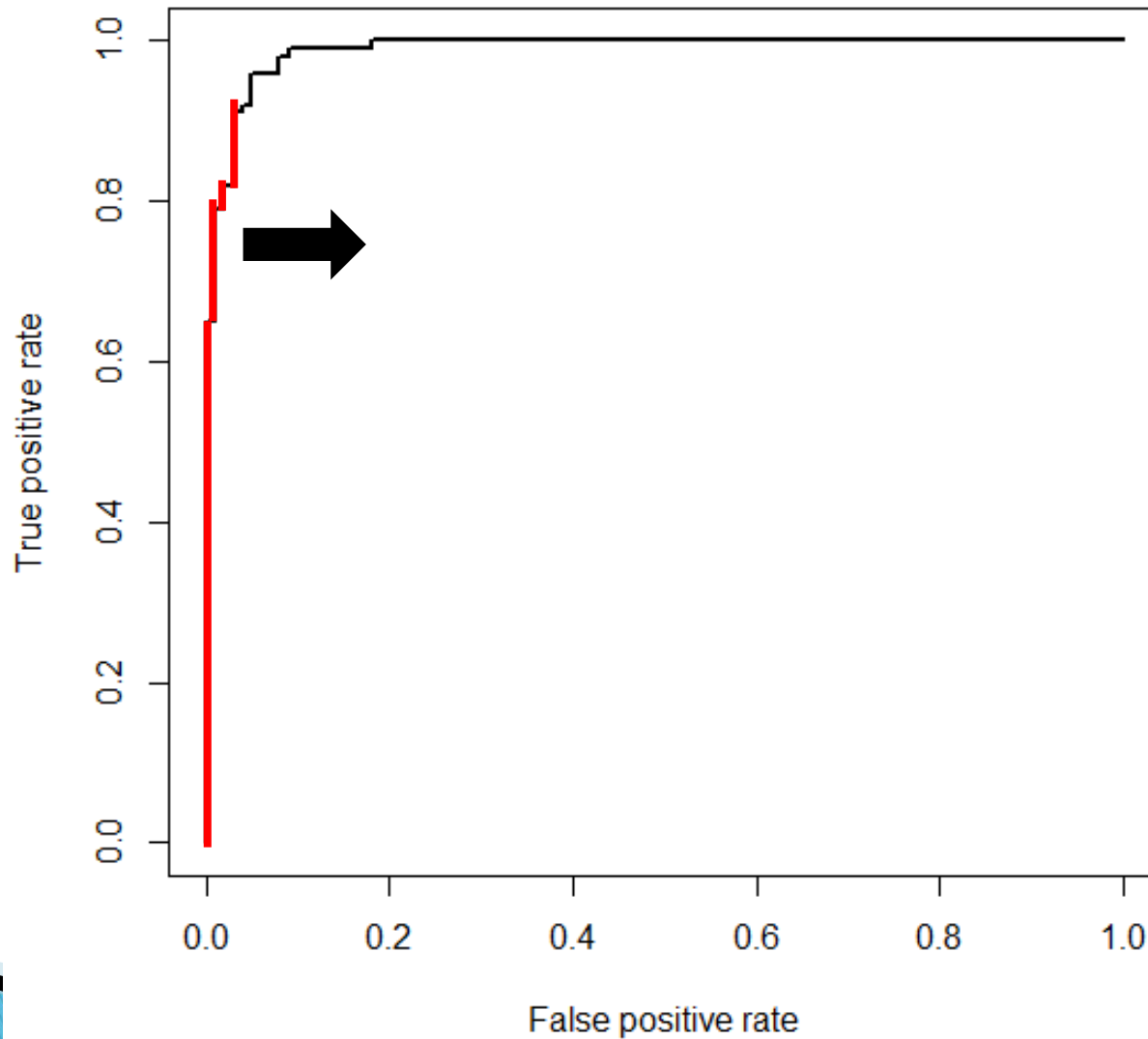
Protin X in condition B



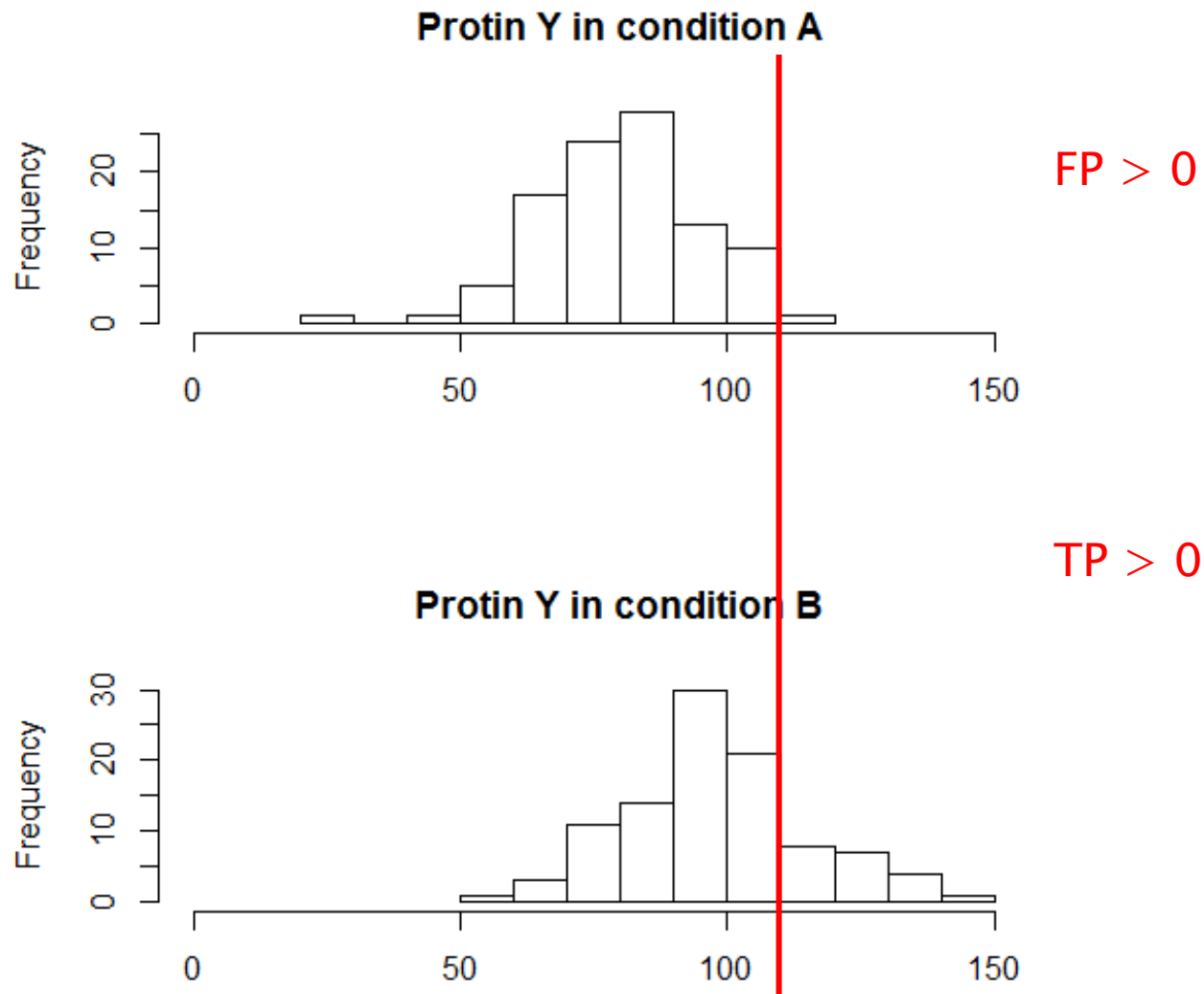
$TP > 0$

ROC curve

ROC curve

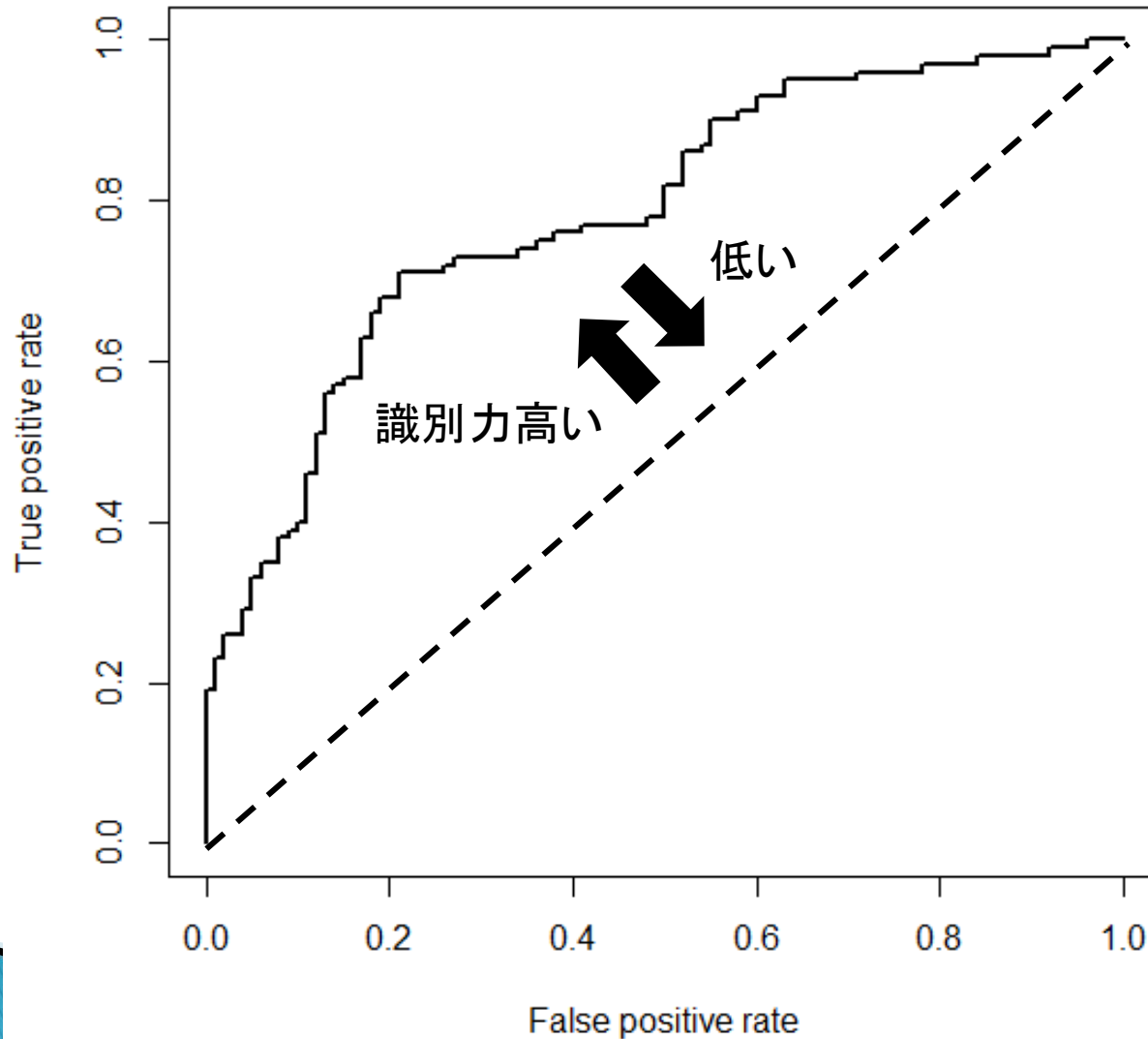


ROC curve



Area Under ROC Curve (AUC)

ROC curve



$$0.5 < \text{AUC} < 1.0$$

Weighted Average Difference (WAD法)

- ▶ 発現比によるランク付けを発現量で補正する方法

assumption : “Strong signals are better signals”

$$WAD(i) = \left(x_i^A / x_i^B \right) \times \frac{\frac{\left(x_i^A + x_i^B \right)}{2} - \min}{\max - \min}$$

WAD値

min : 0.1

max : 1000

condition A	B	
0.2	→ 2	$(2/0.2) \times \{(2+0.2)/2 - 0.1\} / (1000 - 0.1) \doteq 0.01$
20	→ 200	$(200/20) \times \{(200+2)/2 - 0.1\} / (1000 - 0.1) \doteq 1.0$

http://ncbi.nlm.nih.gov

GEO Datasets

受信トレイ - junakabayashi x MSN Japan x Home - PubMed - NCBI x (Rで)マイクロアレイデータ解析 x

www.ncbi.nlm.nih.gov/pubmed

NCBI Resources How To Sign in to NCBI

PubMed.gov
US National Library of Medicine
National Institutes of Health

GEO DataSets GSE13714 Search

Advanced Help

PubMed
PubMed comprises more than 23 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

PubMed Commons
PubMed's new commenting system
More

Using PubMed
[PubMed Quick Start Guide](#)
[Full Text Articles](#)
[PubMed FAQs](#)
[PubMed Tutorials](#)
[New and Noteworthy](#)

PubMed Tools
[PubMed Mobile](#)
[Single Citation Matcher](#)
[Batch Citation Matcher](#)
[Clinical Queries](#)
[Topic-Specific Queries](#)

More Resources
[MeSH Database](#)
[Journals in NCBI Databases](#)
[Clinical Trials](#)
[E-Utilities](#)
[LinkOut](#)

You are here: NCBI > Literature > PubMed Write to the Help Desk

GETTING STARTED NCBI Education
RESOURCES Chemicals & Bioassays
POPULAR PubMed
FEATURED Genetic Testing Registry
NCBI INFORMATION About NCBI

GSE13714_RAW.tar GPL1261-14790.txt GSE37907_RAW.tar

Show all downloads...

Affymetrix CEL file

受信トレイ - junakabayashi x MSN Japan x GEO Accession viewer x (Rで)マイクロアレイデータ解析 x

www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13714

Street address 1 Blackfan Circle
City Boston
State/province MA
ZIP/Postal code 02115
Country USA

Platforms (1) GPL571 [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array

Samples (12)
[Less...](#)

- GSM344801 MOLM14_1F3HOXA9sh_suppression_rep1
- GSM344802 MOLM14_1F3HOXA9sh_suppression_rep2
- GSM344803 MOLM14_1F3HOXA9sh_suppression_rep3
- GSM344804 MOLM14_2A5HOXA9sh_suppression_rep1
- GSM344805 MOLM14_2A5HOXA9sh_suppression_rep2
- GSM344806 MOLM14_2A5HOXA9sh_suppression_rep3
- GSM344807 MOLM14_GFPcontrolsh_control_rep1
- GSM344808 MOLM14_GFPcontrolsh_control_rep2
- GSM344809 MOLM14_GFPcontrolsh_control_rep3
- GSM344810 MOLM14_GFPcontrolsh_control_rep4
- GSM344811 MOLM14_GFPcontrolsh_control_rep5
- GSM344812 MOLM14_GFPcontrolsh_control_rep6

Relations
BioProject PRJNA110389

Analyze with GEO2R

Download family	Format
SOFT formatted family file(s)	SOFT ?
MINiML formatted family file(s)	MINiML ?
Series Matrix File(s)	TXT ?

Supplementary file	Size	Download	File type/resource
GSE13714_RAW.tar	20.1 Mb	(http)(custom)	TAR (of CEL)

Raw data provided as supplementary file
Processed data included within Sample table

GSE13714_RAW.tar x GPL1261-14790.txt x GSE37907_RAW.tar x

Show all downloads... x

CEL file

受信トレイ - junakabayashi x MSN Japan x GEO Accession viewer x (Rで)マイクロアレイデータ解析 x

www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13714

Overall design RNA was purified from t(9;11) MOLM-14 AML cells 44h after transduction in triplicates with 2 of the two most effective HOXA9shRNA constructs (3 x 1F3-HOXA9shRNA; 3 x 2A5-HOXA9shRNA) or GFP-controlshRNA (6x).

Contributor(s) [Ehler J, Krivtsov AV, Stubbs MC, Wright B, Davis TN, van den Heuvel-Eibrink](#)

GSM3448

File Home Share View

Clipboard: Copy, Paste, Cut, Copy path, Paste shortcut, Move to, Copy to, Delete, Rename

Organise

New: New item, Easy access, New folder

Open: Properties, Open, Edit, History

Select: Select all, Select none, Invert selection

This PC > Downloads > GSM3448

Search GSM3448

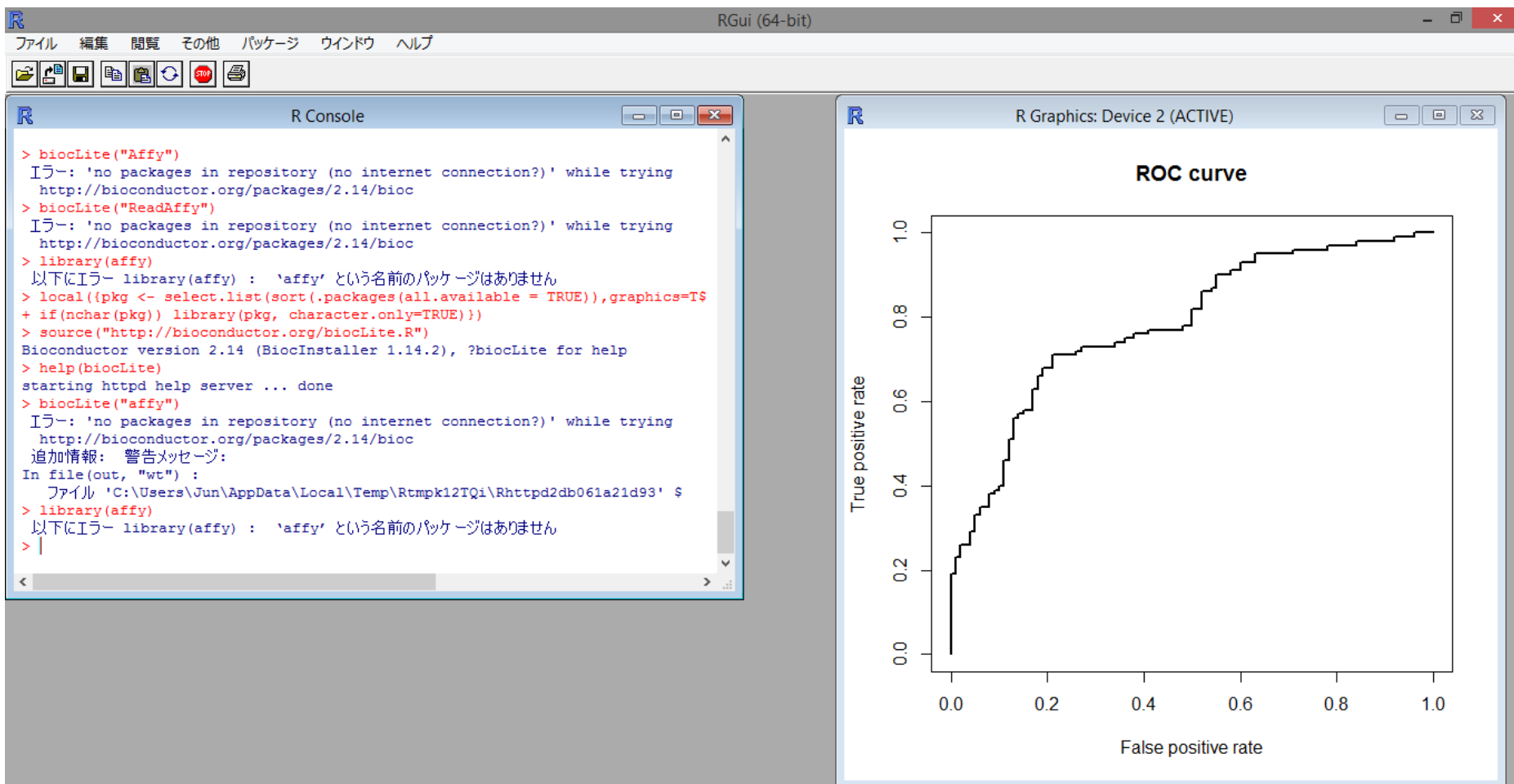
Name	Date modified	Type	Size
GSM344801.CEL	2008/11/24 22:11	CEL File	5,235 KB
GSM344802.CEL	2008/11/24 22:11	CEL File	5,235 KB
GSM344803.CEL	2008/11/24 22:11	CEL File	5,235 KB
GSM344804.CEL	2008/11/24 22:11	CEL File	5,235 KB
GSM344805.CEL	2008/11/24 22:11	CEL File	5,237 KB
GSM344806.CEL	2008/11/24 22:11	CEL File	5,237 KB
GSM344807.CEL	2008/11/24 22:11	CEL File	5,236 KB
GSM344808.CEL	2008/11/24 22:11	CEL File	5,235 KB
GSM344809.CEL	2008/11/24 22:11	CEL File	5,236 KB
GSM344810.CEL	2008/11/24 22:11	CEL File	5,236 KB
GSM344811.CEL	2008/11/24 22:11	CEL File	5,236 KB
GSM344812.CEL	2008/11/24 22:11	CEL File	5,235 KB

12 items State: Shared

GSE13714_RAW.tar GPL1261-14790.txt GSE37907_RAW.tar

Show all downloads...

統計処理ソフトR



- ▶ MOLM-14 MLL cell
- ▶ HOXA9 shRNA or control shRNA

Gene	GSM344801.CEL	GSM344802.CEL	GSM344806.CEL	GSM344807.CEL	GSM344808.CEL	GSM344809.CEL
DDR1	5.224917553	4.98522	5.186036	4.652569	4.90695	4.665124
RFC2	4.743742853	4.745553	4.701053	5.375102	5.223988	5.457174
HSPA6	6.020715827	6.285368	6.096924	7.168145	7.046261	7.125773
PAX8	7.375744313	7.265474	7.501778	6.564478	6.751406	6.600312
GUCA1A	3.054374801	3.070379	2.882453	2.62843	2.69892	2.505949
...	...	...	...	...	...	...

HOXA9 shRNA

control

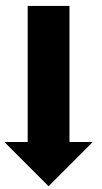
GEO accession number GSE13714

HOXA9 KD/cont

gene	KD	cont
ARL17A	1.999793	1.476522
TOM1L1	1.58211	1.05741
SLC30A4	1.957213	1.432438
UGT8	2.391064	1.866087
LHCGR	1.687855	1.162073
PTPRK	2.147615	1.62019
CYLC2	1.889134	1.358931
SHANK2	1.857913	1.319694
OASL	2.511979	1.966135
HTN3	1.708472	1.159469
PRRX1	2.080513	1.523294
CYP4A11	2.019291	1.461205
TIGD1L	1.752859	1.156133
SCD5	3.114234	2.51651
LOC100128640	2.230409	1.617674
NMBR	2.049015	1.433451
BAGE	1.73548	1.11801
CYP4A22	1.890733	1.271096
TTPA	2.025562	1.401402
SPINLW1	1.947012	1.254959
LOC100286895	2.752968	2.033344

WAD法

gene	KD	cont
KLRK1	2.495853	2.05735
APLNR	2.406628	1.932852
ATP8B1	2.795895	2.438949
SLC35E1	3.127605	2.828033
NFAT5	2.55761	2.131336
TTC18	2.600246	2.185885
LOC100128640	2.230409	1.617674
ZNF528	2.38649	1.867257
UGT8	2.391064	1.866087
IFIT3	2.721085	2.298406
FBXW12	3.015392	2.659894
ZNF721	3.152595	2.803151
OASL	2.511979	1.966135
ZNF835	2.589609	2.072417
OPHN1	3.543285	3.236864
TCTN2	3.174498	2.811978
ZNF816	2.906966	2.47826
CYP1A2	2.734135	2.226856
MEFV	3.13325	2.699734
LOC100286895	2.752968	2.033344
SCD5	3.114234	2.51651



KD > Cont

WAD法と発現比の比較

36 microarray data sets

遺伝子を二群に分類

1. マイクロアレイ以外の実験で発現増加が確認されている遺伝子
2. それ以外

WAD値と発現比についてAUCを求める

Method	average AUC
WAD	96.737 %
Expression Ratio	94.659 %

Kadota K *et al.* Algorithms Mol Biol. 2008

結論

- ▶ 遺伝子発現の二群間比較
発現比によるランク付け、
発現量による補正 WAD法
- ▶ 識別力の判定(AUCによる判定)
- ▶ WAD法は発現比のみより識別力が高い