

予測と判別

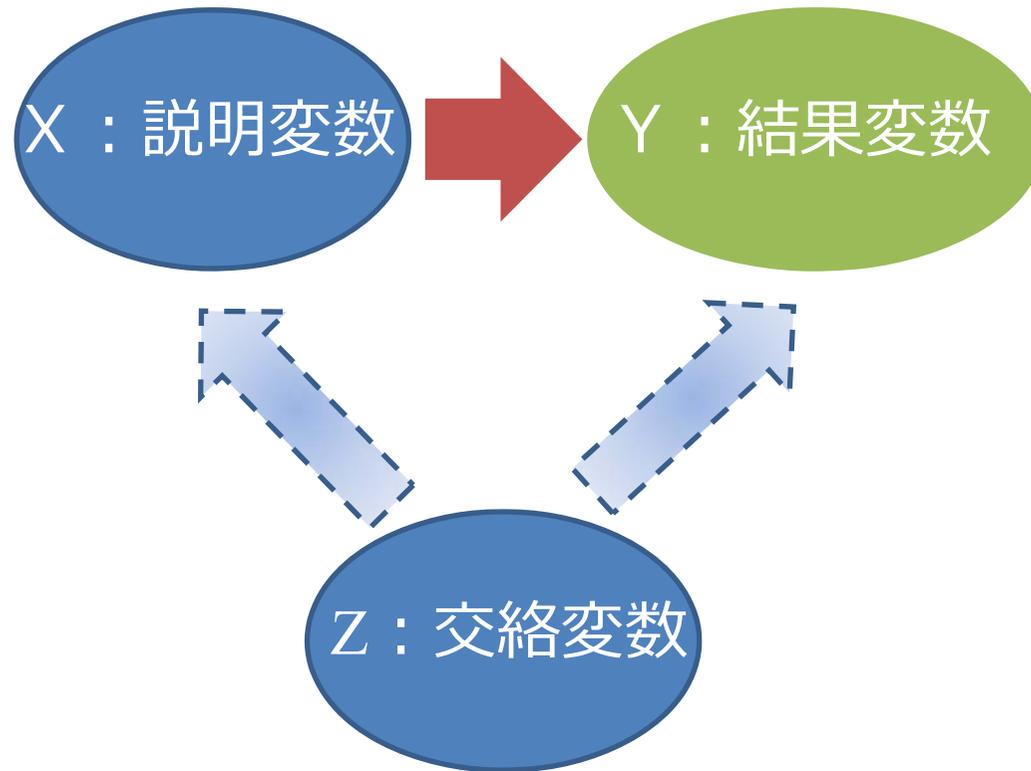
—分子マーカーデータへの応用を例に

国立国際医療研究センター

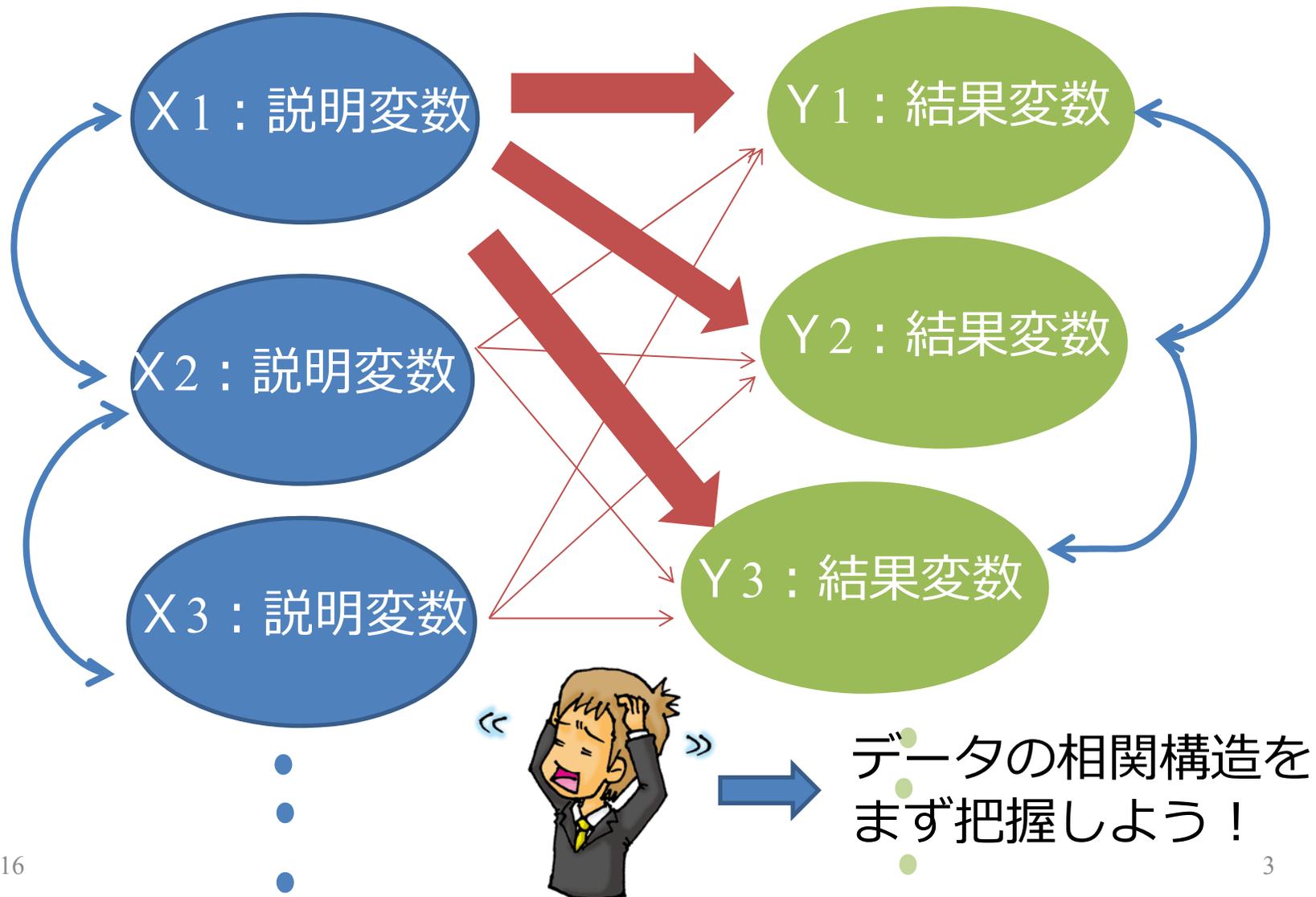
医学統計研究室

田中 紀子

データを分析する

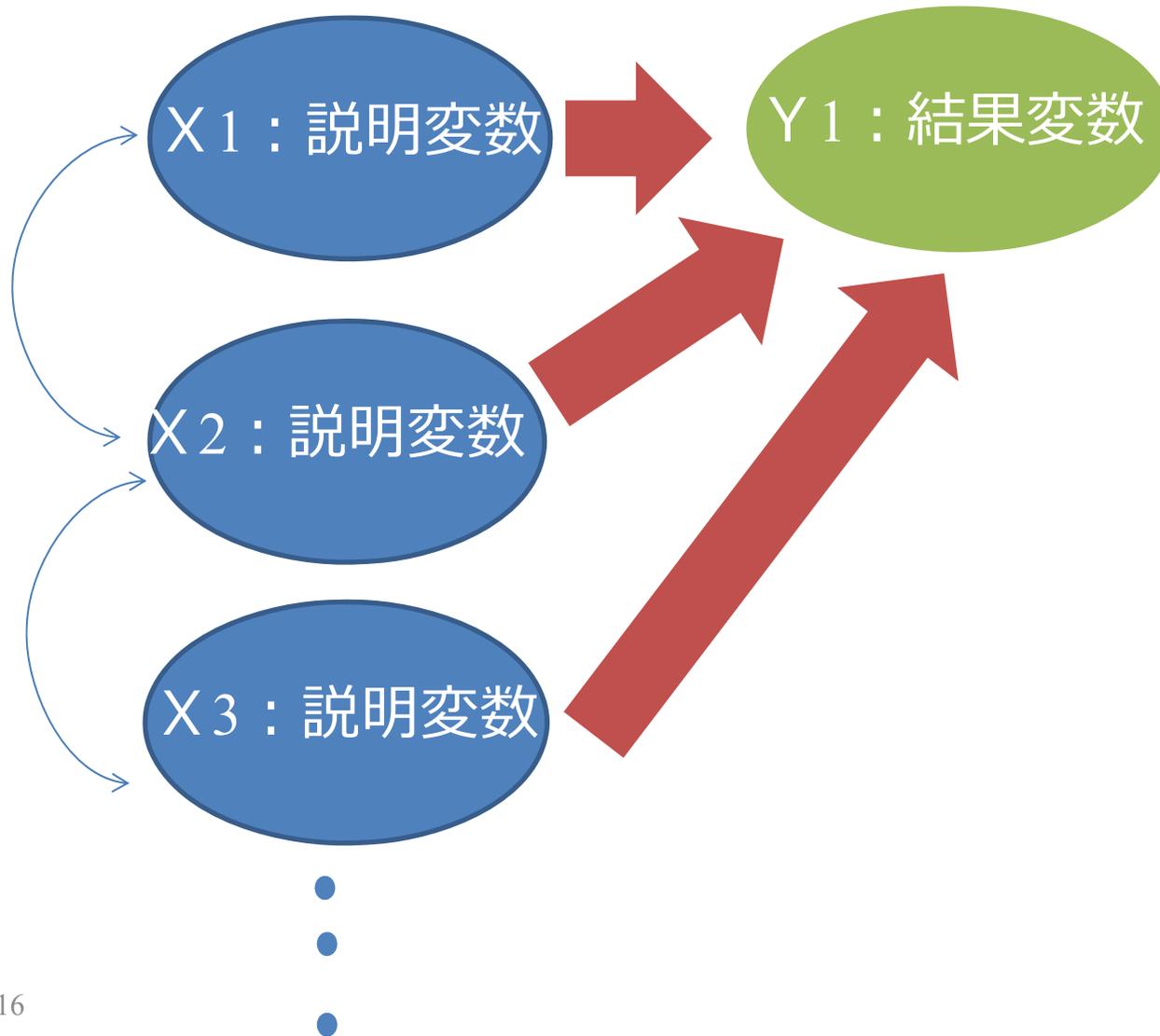


データを分析する



2015/3/16

データを分析する



2015/3/16

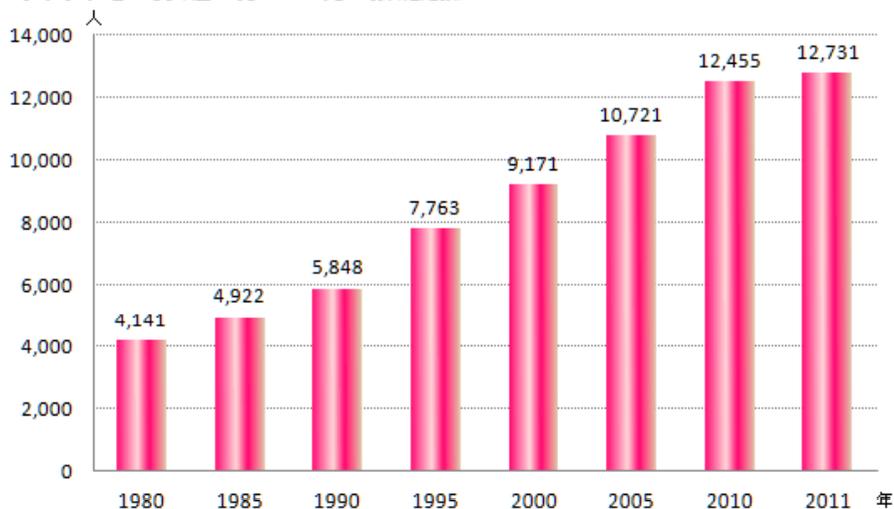
4

今日のおはなし

- 予測？判別？
- 例：NSABP-B31試験のコリラ
タイプ研究

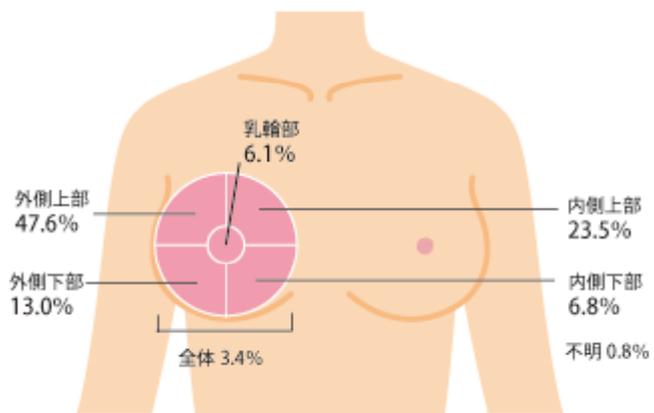
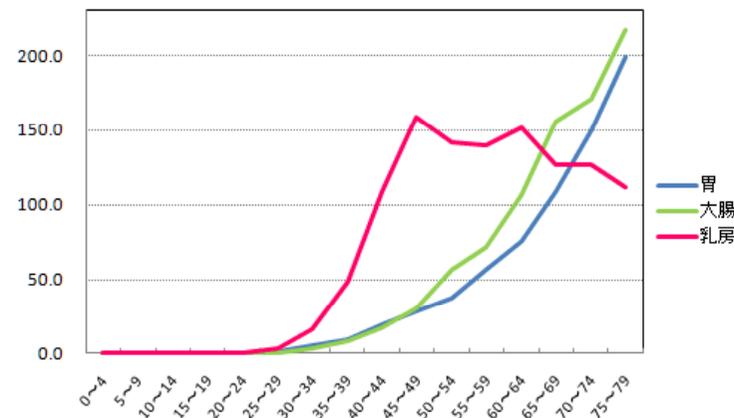
乳がん

厚生労働省人口動態統計<2012年9月6日発表>より
ますます増える女性の乳がんの死亡数(確定数)



ピンクリボンフェスティバルHPより

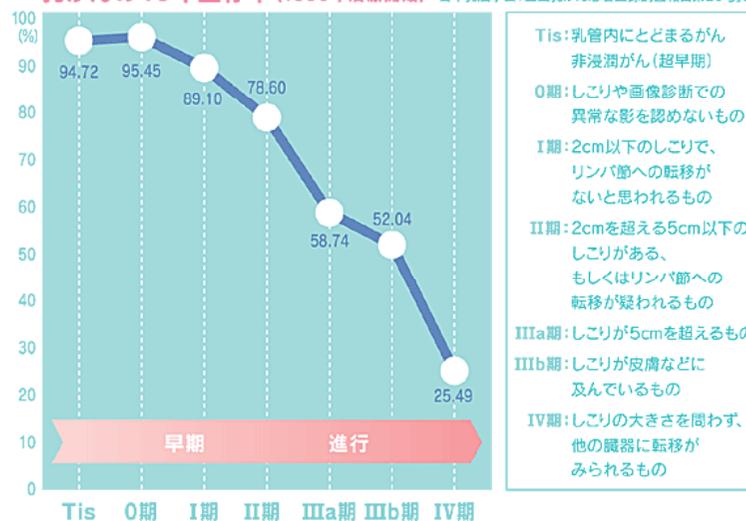
国立がん研究センターがん対策情報センター統計
部位別年齢階級別がん罹患率:女性(2005年)
(対人口10万人)



※重複症例あり

全国乳がん患者登録調査報告 第32号2000

乳がんの10年生存率(1990年治療開始) 日本乳癌学会「全国乳がん患者登録調査報告第29号」より



Tis:乳管内にとどまるがん
非浸潤がん(超早期)
0期:しこりや画像診断での
異常な影を認めないもの
I期:2cm以下のしこりで、
リンパ節への転移が
ないと思われるもの
II期:2cmを超える5cm以下の
しこりがある、
もしくはリンパ節への
転移が疑われるもの
IIIa期:しこりが5cmを超えるもの
IIIb期:しこりが皮膚などに
及んでいるもの
IV期:しこりの大きさを問わず、
他の臓器に転移が
みられるもの

2015/3/16

ノバルティスファーマHPより

乳がんの治療

ファイザーHPより

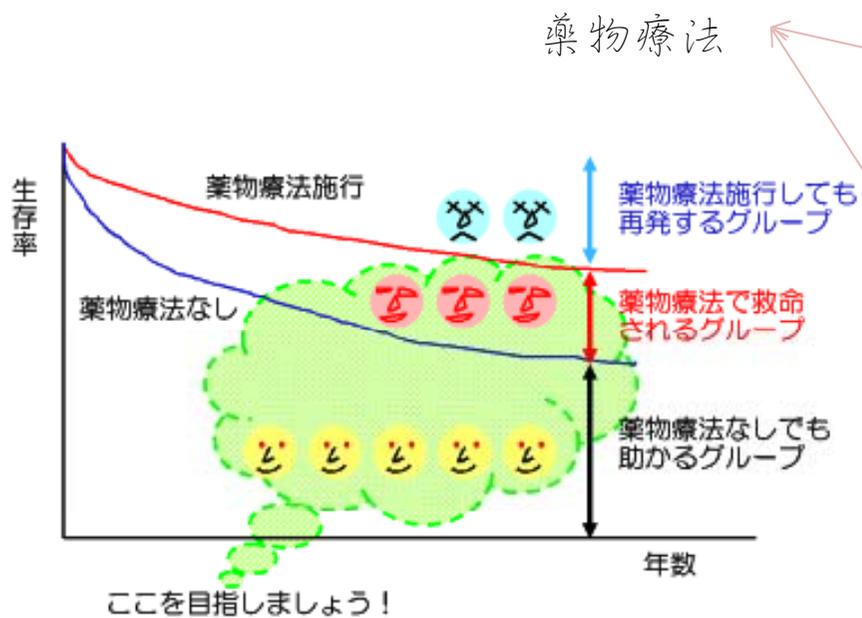
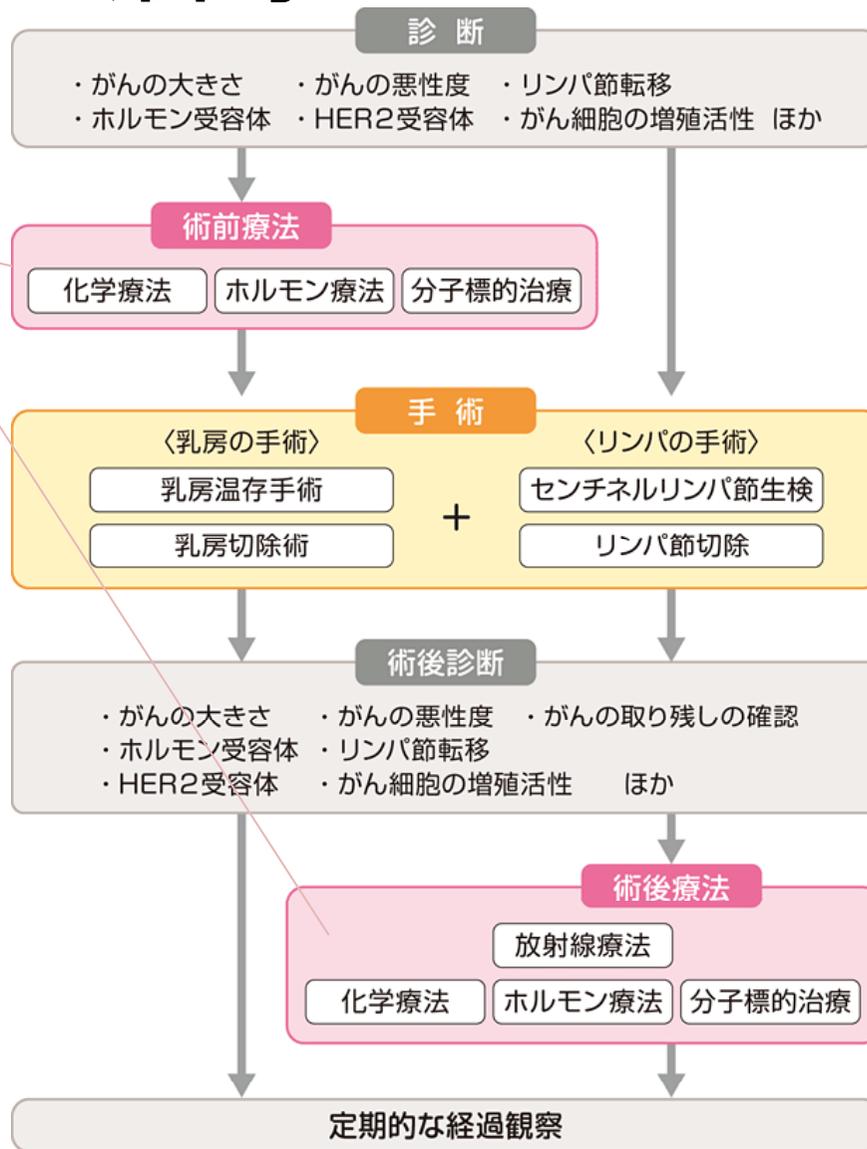


図6 乳癌における薬物療法の有用性

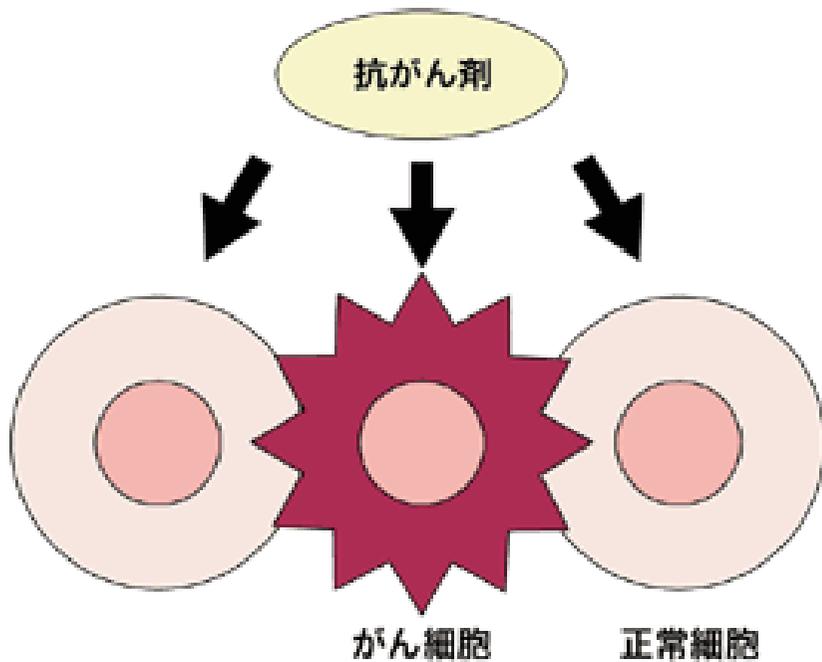
大阪医療センターHPより

2015/3/16

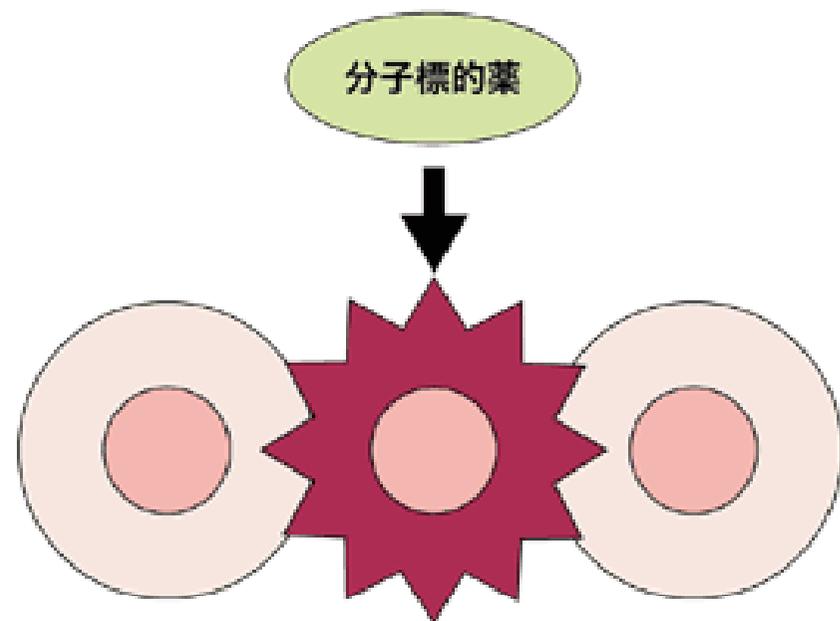


分子標的薬

がん細胞だけでなく正常細胞にも障害を与える

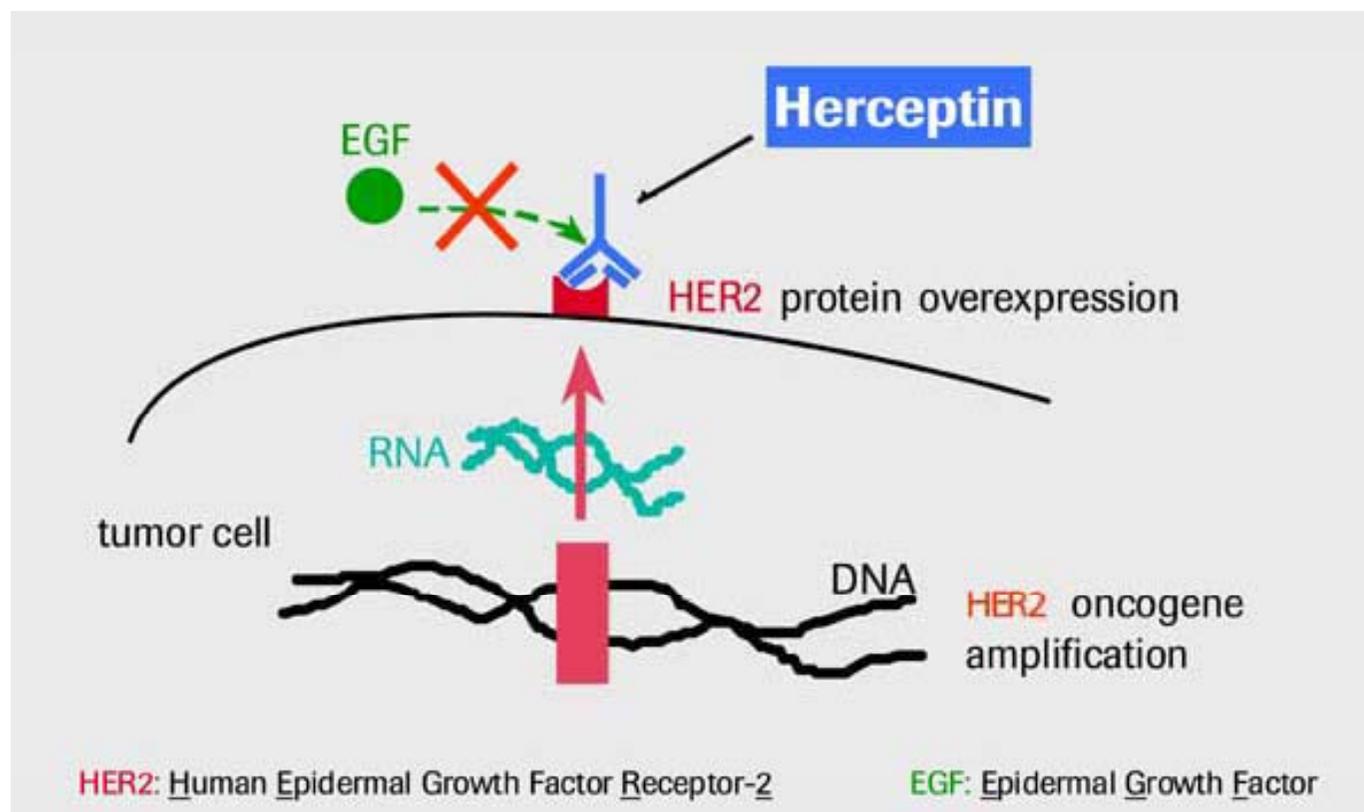


がん細胞のみを標的として狙い撃つことができる



<http://tai-gan.net/index.php?bunsi>

ハーセプチン(Trastuzumab)



<http://trialx.com/curetalk/2012/02/what-is-trastuzumab-herceptin/>

2015/3/16

9

ハーセプチンの効果

The New England Journal of Medicine

Copyright © 2001 by the Massachusetts Medical Society

VOLUME 344

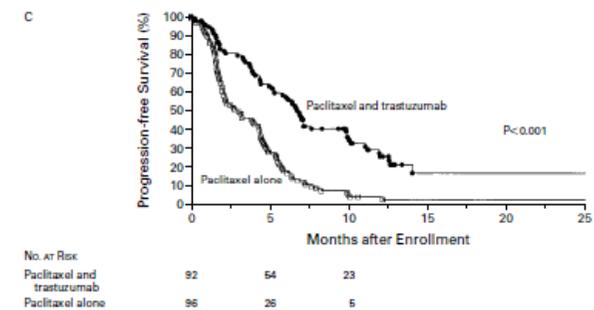
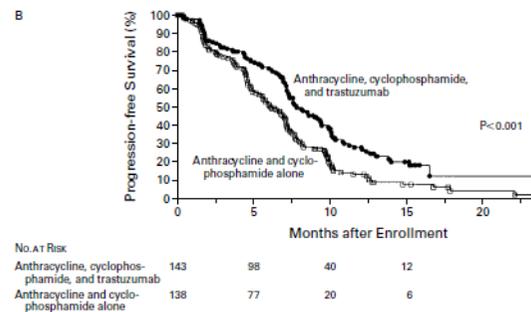
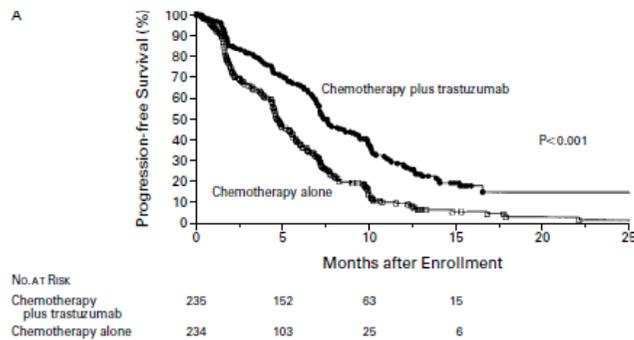
MARCH 15, 2001

NUMBER 11



USE OF CHEMOTHERAPY PLUS A MONOCLONAL ANTIBODY AGAINST HER2 FOR METASTATIC BREAST CANCER THAT OVEREXPRESSES HER2

DENNIS J. SLAMON, M.D., PH.D., BRIAN LEYLAND-JONES, M.D., STEVEN SHAK, M.D., HANK FUCHS, M.D.,
VIRGINIA PATON, PHARM.D., ALEX BAJAMONDE, PH.D., THOMAS FLEMING, PH.D., WOLFGANG EIERMANN, M.D.,
JANET WOLTER, M.D., MARK PEGRAM, M.D., JOSE BASELGA, M.D., AND LARRY NORTON, M.D.*



2015/3/16

10

ハーセプチンの安全性

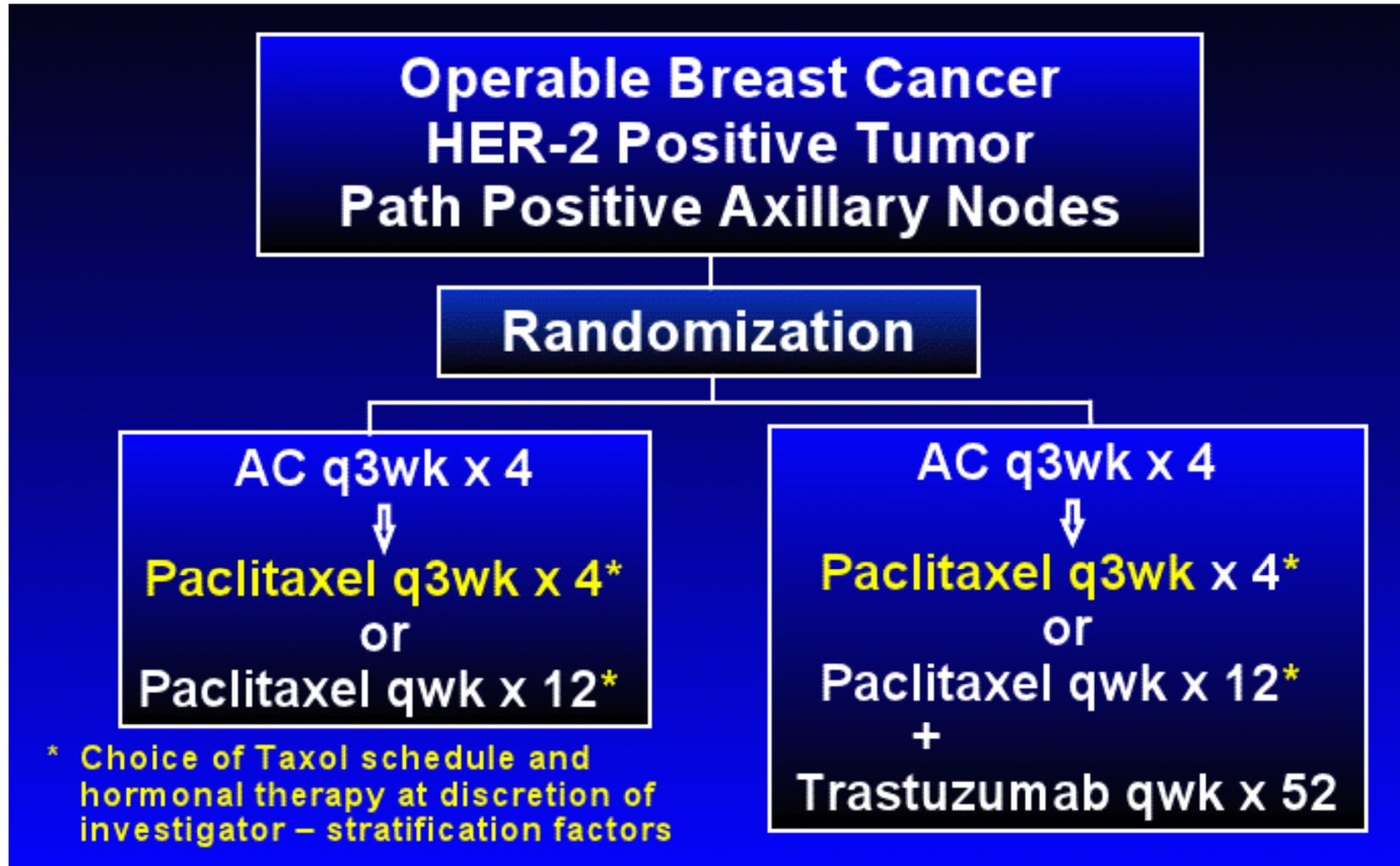
Slamon D, NEJM 2001

TABLE 4. ADVERSE EVENTS THAT OCCURRED IN MORE THAN 10 PERCENT OF PATIENTS AS A GROUP.*

TYPE OR LOCATION OF ADVERSE EVENT	CHEMOTHERAPY PLUS TRASTUZUMAB (N=234)	CHEMOTHERAPY ALONE (N=230)	AN ANTHRACYCLINE, CYCLOPHOSPHAMIDE, AND TRASTUZUMAB (N=143)	AN ANTHRACYCLINE AND CYCLOPHOSPHAMIDE ALONE (N=135)	PACLITAXEL AND TRASTUZUMAB (N=91)	PACLITAXEL ALONE (N=95)
	percentage with event (percentage with severe event)					
Any type						
Abdominal pain	27 (3)	20 (3)	23 (2)	18 (2)	34 (3)	22 (4)
Asthenia	57 (7)	56 (7)	54 (7)	55 (7)	62 (8)	57 (8)
Back pain	31 (4)	22 (4)	27 (2)	16 (2)	36 (8)	30 (5)
Chest pain	24 (3)	24 (4)	20 (3)	21 (2)	30 (3)	27 (5)
Chills	38 (<1)	8 (<1)	35 (<1)	11 (2)	42 (1)	4 (0)
Fever	53 (8)	29 (4)	56 (11)	33 (7)	47 (2)	23 (1)
Headache	41 (4)	30 (4)	44 (3)	31 (5)	36 (7)	28 (2)
Infection	47 (2)	29 (2)	47 (2)	30 (2)	46 (1)	27 (2)
Pain	57 (6)	50 (7)	57 (4)	42 (8)	60 (10)	61 (6)
Heart failure	22 (10)	5 (2)	27 (16)	8 (3)	3 (2)	1 (1)
Digestive tract						
Anorexia	28 (<1)	22 (2)	31 (0)	26 (2)	24 (1)	16 (2)
Constipation	32 (1)	28 (3)	36 (2)	28 (3)	25 (0)	27 (2)
Diarrhea	45 (1)	27 (3)	45 (1)	25 (3)	45 (1)	30 (3)
Nausea	66 (5)	66 (7)	76 (6)	79 (10)	50 (3)	48 (3)
Stomatitis	22 (<1)	21 (0)	30 (1)	31 (3)	10 (0)	7 (0)
Vomiting	47 (5)	40 (7)	53 (3)	49 (8)	37 (9)	28 (5)
Hematologic and lymphatic systems						
Anemia	27 (2)	19 (2)	35 (3)	25 (2)	14 (1)	10 (1)
Leukopenia	41 (11)	26 (9)	52 (15)	33 (11)	24 (6)	17 (5)
Musculoskeletal system						
Arthralgia	20 (4)	14 (2)	8 (<1)	10 (<1)	37 (9)	21 (4)
Myalgia	23 (3)	22 (3)	13 (<1)	13 (<1)	38 (7)	36 (6)
Nervous system						
Paresthesia	29 (<1)	23 (<1)	17 (0)	11 (0)	47 (2)	39 (1)
Respiratory tract						
Increased coughing	43 (<1)	26 (<1)	43 (<1)	28 (0)	42 (0)	22 (1)
Dyspnea not related to heart failure	36 (3)	25 (3)	42 (4)	24 (4)	28 (1)	26 (1)
Pharyngitis	27 (0)	16 (<1)	30 (0)	18 (0)	22 (0)	14 (2)
Skin						
Alopecia	57 (26)	58 (35)	58 (25)	59 (42)	56 (26)	56 (26)
Rash	31 (<1)	17 (<1)	27 (0)	17 (<1)	38 (1)	18 (1)

*The analysis of adverse events excluded five patients who were never treated.

NSABP B-31



アドリアマイシン+ シクロフォスファミド (エ
ンドキサン)

2015/3/16

12

NSABP B31 Primary Aim

デザイン：前向き非盲検ランダム化2群比較試験

対象：病理学的にリンパ節陽性で手術可能なHER2タンパク陽性乳がん患者

AC q3x4 + Taxol q3x4/ q1x12+ Herceptin q1x52(1yr)
Vs

AC q3x4 + Taxol q3x4/ q1x12

- **心イベント発症率に変化が出るかを検証すること**
- **観察期間5年を最短として、DFS（無病生存率）に差があるかどうか検証すること**

NABP B31 結果(有効性 : 4year follow up)

VOLUME 29 · NUMBER 25 · SEPTEMBER 1 2011

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Four-Year Follow-Up of Trastuzumab Plus Adjuvant Chemotherapy for Operable Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer: Joint Analysis of Data From NCCTG N9831 and NSABP B-31

Edith A. Perez, Edward H. Romond, Vera J. Suman, Jong-Hyeon Jeong, Nancy E. Davidson, Charles E. Geyer Jr, Silvana Martino, Eleftherios P. Mamounas, Peter A. Kaufman, and Norman Wolmark

NCCTG と合同で有効性に関する評価を発表
DFSに関して、ハーセプチンの効果は統計学的に有意 (P<.001)
39% の死亡率減少を達成した

DFS OS
HR 0.52(0.45-0.60) HR 0.61(0.50-0.75)

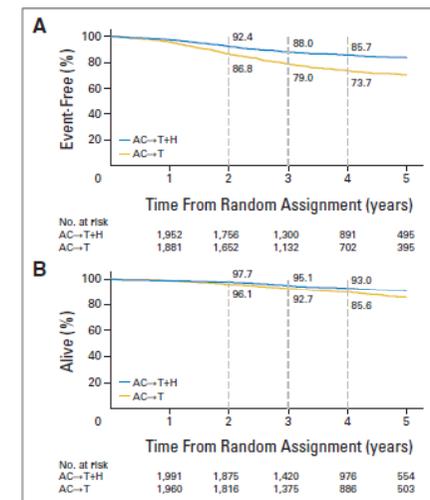


Fig 3. Kaplan-Meier estimates of (A) event-free survival and (B) overall survival. Disease events include local, regional, or distant recurrence; contralateral breast cancer; second primary cancers; or death as a result of any cause. Overall survival is measured from the time of study enrollment to last contact or death. AC, doxorubicin and cyclophosphamide; H, trastuzumab; T, paclitaxel.

2015/3/16

14

NABP B31 結果(安全性 : 7year follow up)

VOLUME 30 · NUMBER 31 · NOVEMBER 1 2012

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

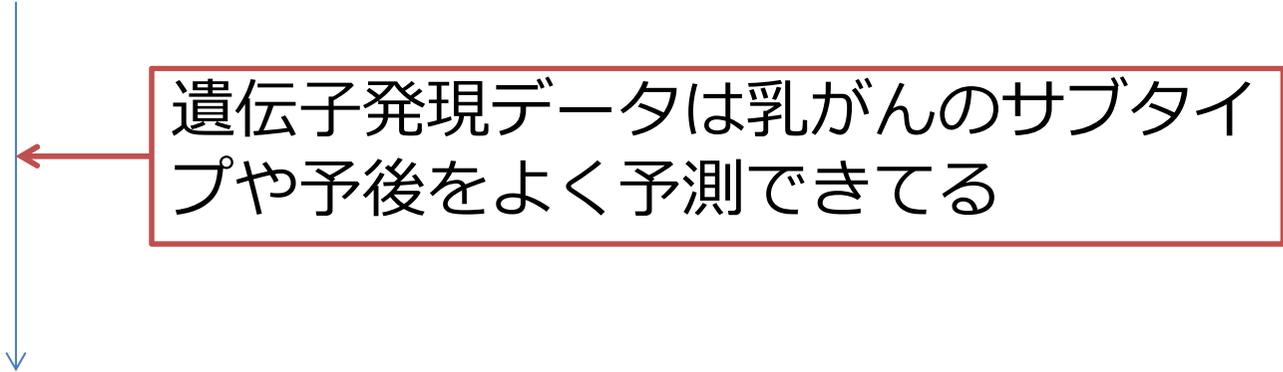
Seven-Year Follow-Up Assessment of Cardiac Function in NSABP B-31, a Randomized Trial Comparing Doxorubicin and Cyclophosphamide Followed by Paclitaxel (ACP) With ACP Plus Trastuzumab As Adjuvant Therapy for Patients With Node-Positive, Human Epidermal Growth Factor Receptor 2–Positive Breast Cancer

Edward H. Romond, Jong-Hyeon Jeong, Priya Rastogi, Sandra M. Swain, Charles E. Geyer Jr, Michael S. Ewer, Vikas Rath, Louis Fehrenbacher, Adam Brufsky, Catherine A. Azar, Patrick J. Flynn, John L. Zapas, Jonathan Polikoff, Howard M. Gross, David D. Biggs, James N. Atkins, Elizabeth Tan-Chiu, Ping Zheng, Greg Yothers, Eleftherios P. Mamounas, and Norman Wolmark

ハーセプチン治療群 (n=944)で4%、コントロール群で(n=743)で1.3%の心イベントを観測
ハーセプチン治療を中止した後、ほとんどの患者でLVEF値が正常範囲に修復

New research question

- トラスツズマブも効かない人がいる？
- HER2+は本当にトラスツズマブ投与の基準として妥当？

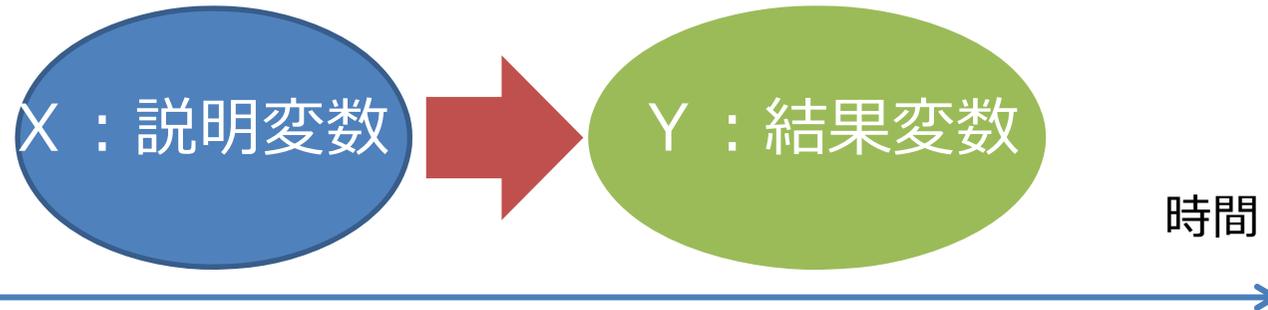


遺伝子発現データは乳がんのサブタイプや予後をよく予測できてる

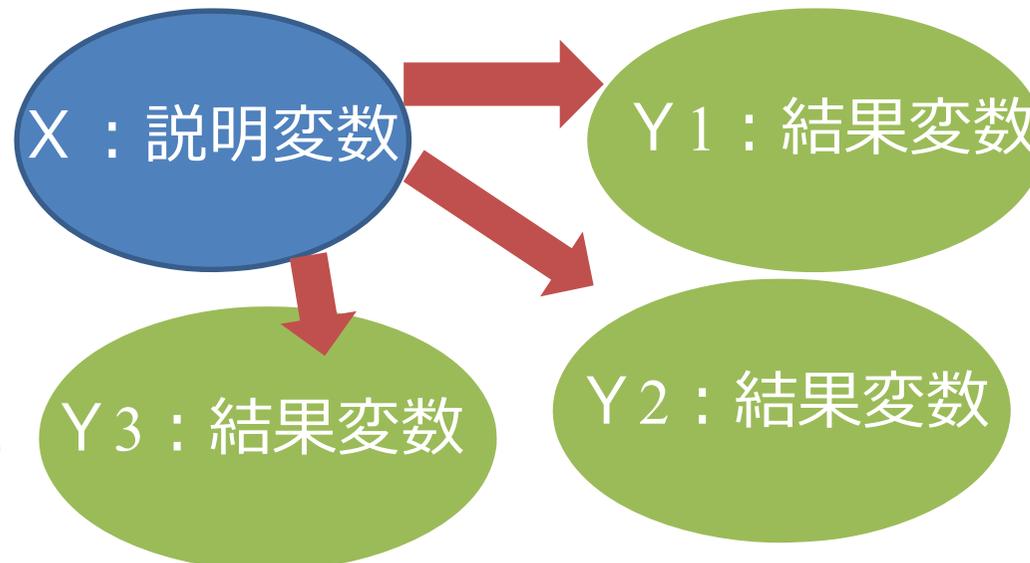
遺伝子発現データを分子マーカーとして治療効果の予測ができないのでしょうか？

予測と判別

予測 (prediction)



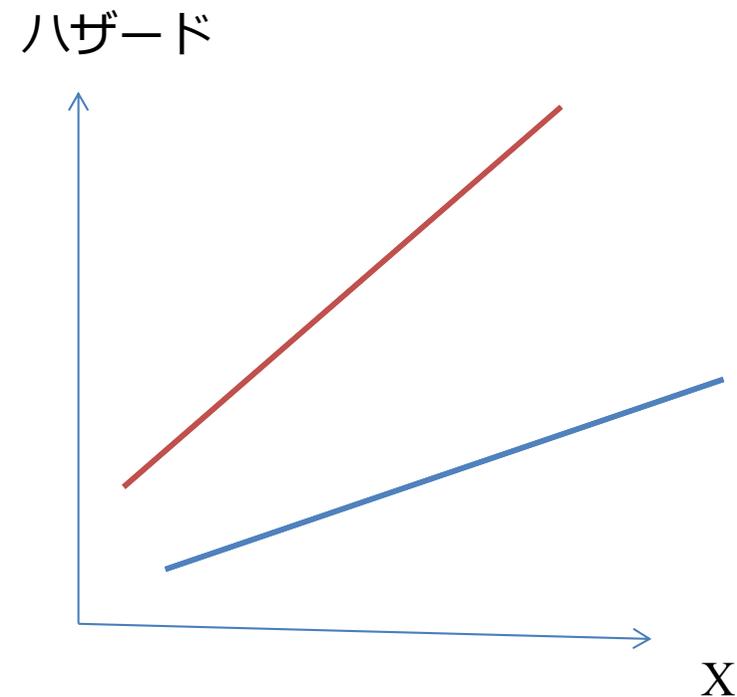
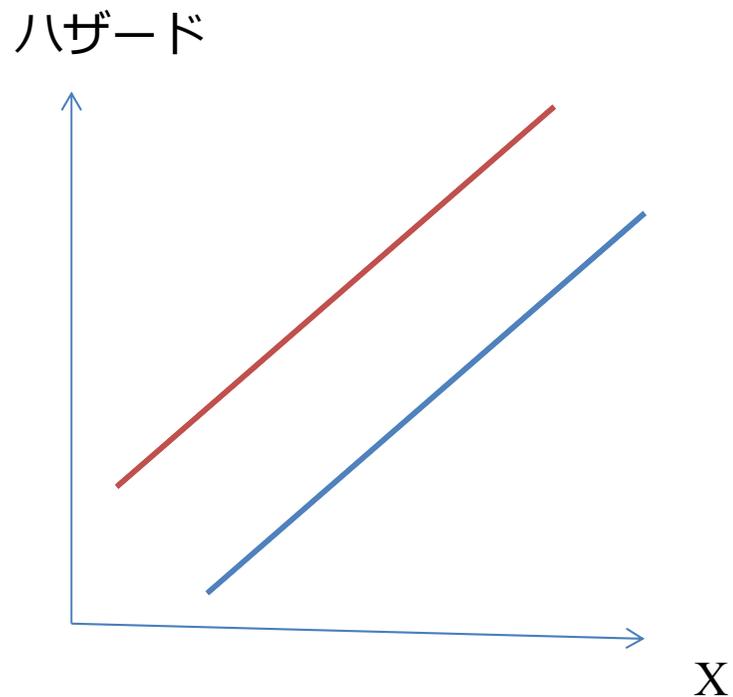
判別 (discrimination/classification)



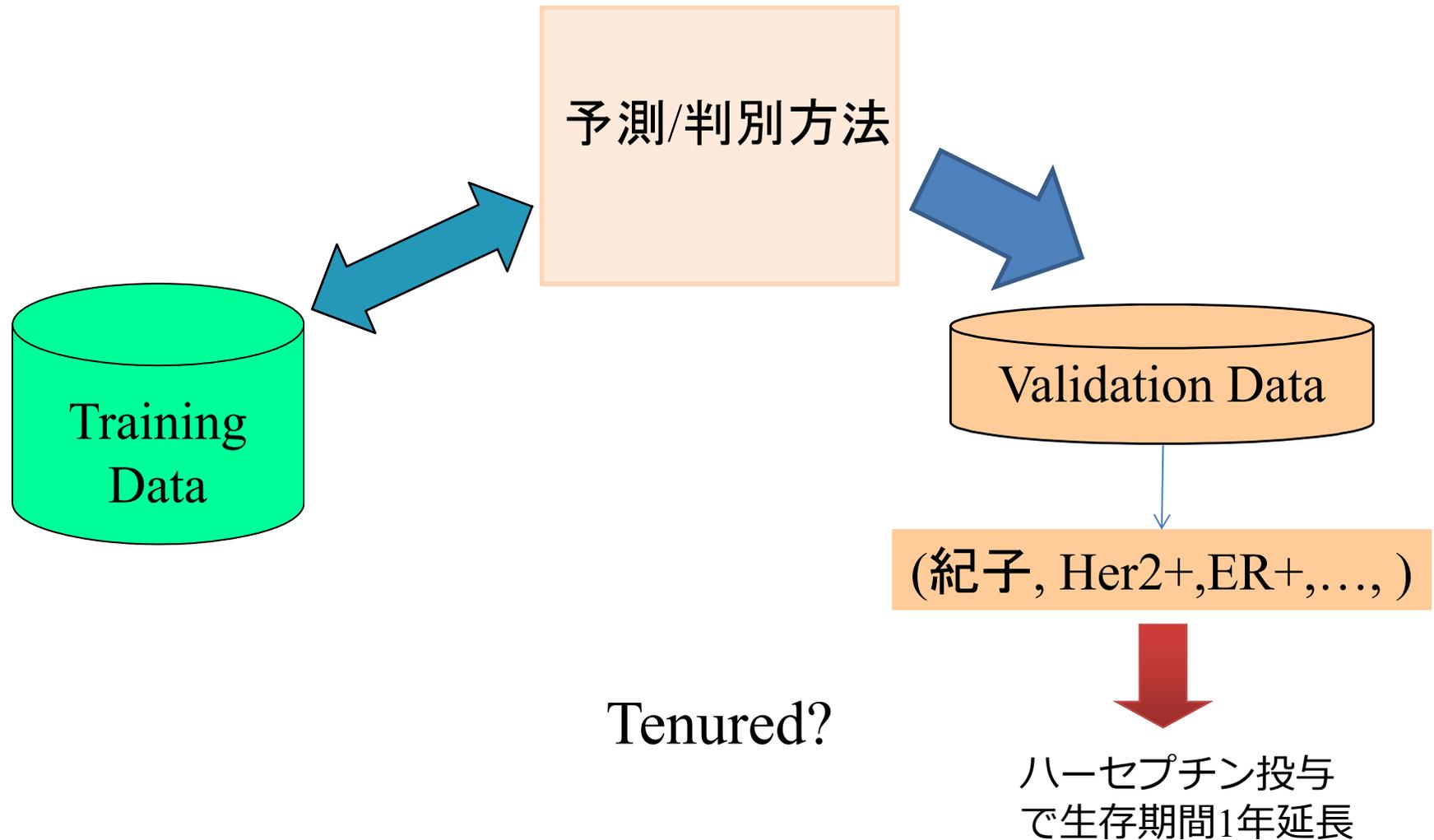
予後予測と治療効果予測

- 予後予測
 - ある時点から将来の患者の状態を予測すること
 - 予後予測因子：
 - リンパ節転移状況 組織型 HER2
OncotypeDX . . .
- 治療効果予測
 - ある治療を行って（将来）効果が得られるかどうかを予測すること
 - 治療効果予測因子：
 - ??

予後予測因子と治療効果予測因子

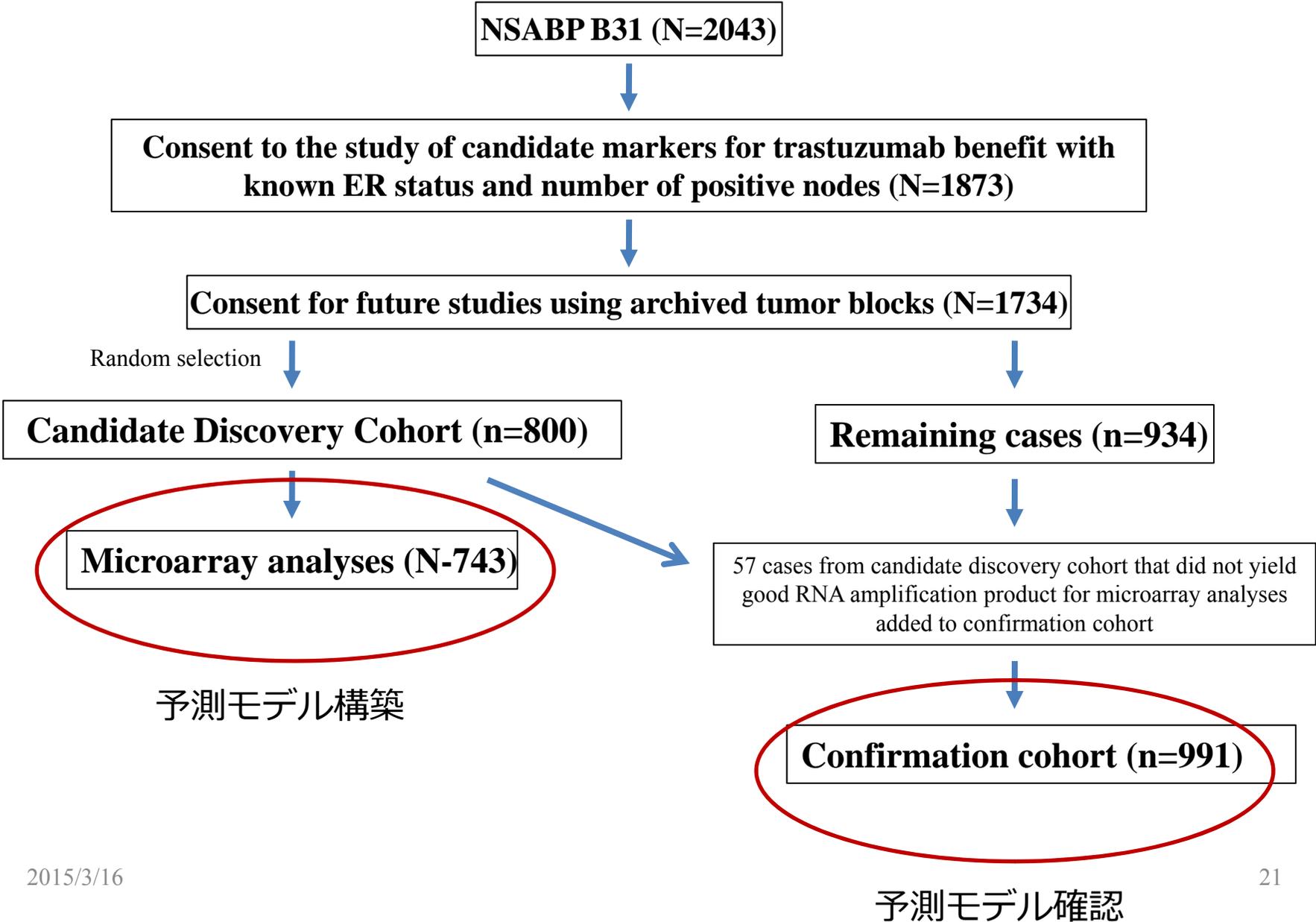


予測/判別モデル構築方法



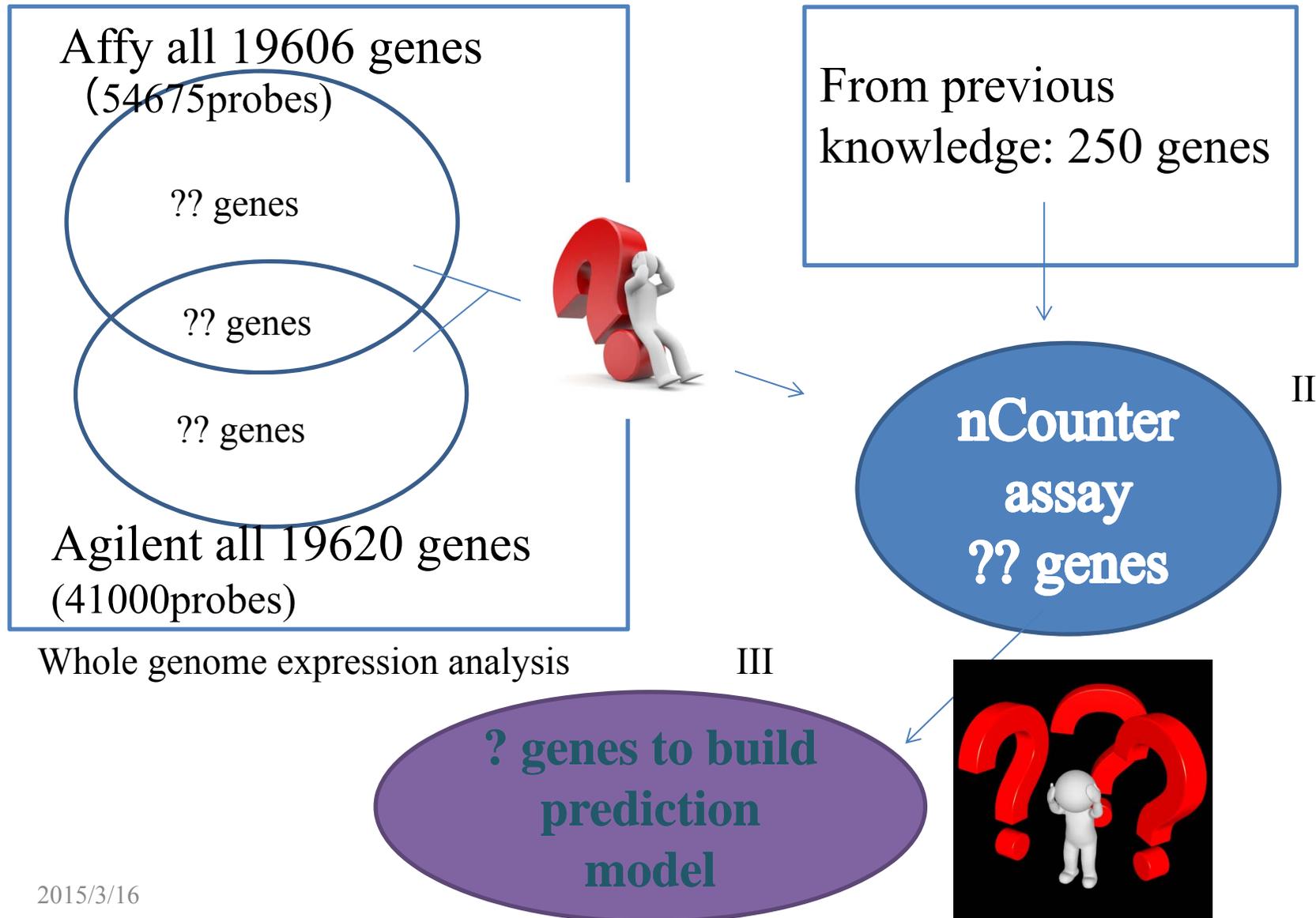
2015/3/16

20



遺伝子発現データを用いた 治療効果予測モデルの構築

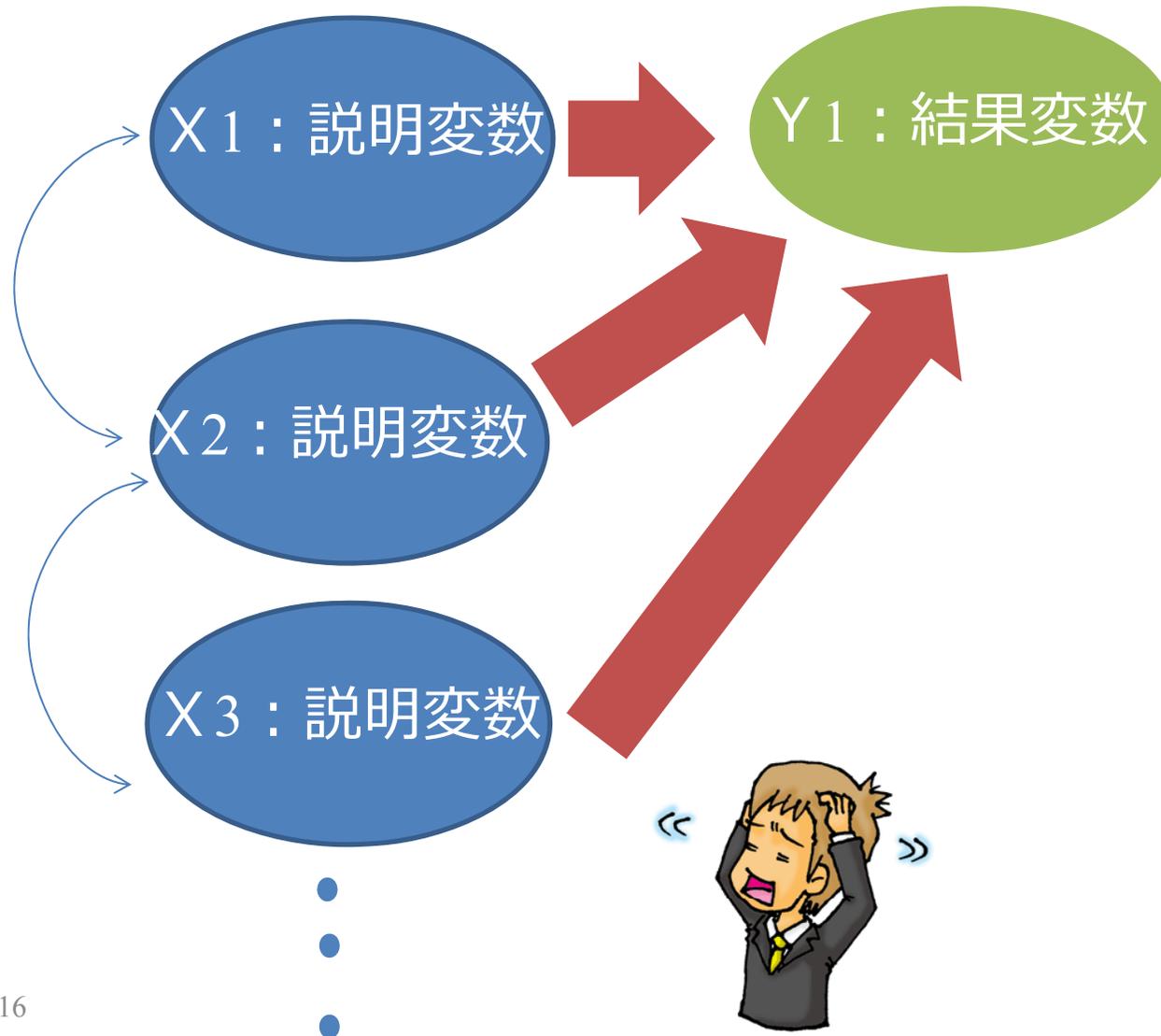
I



2015/3/16

22

I : 数十万変数からの変数選択



2015/3/16

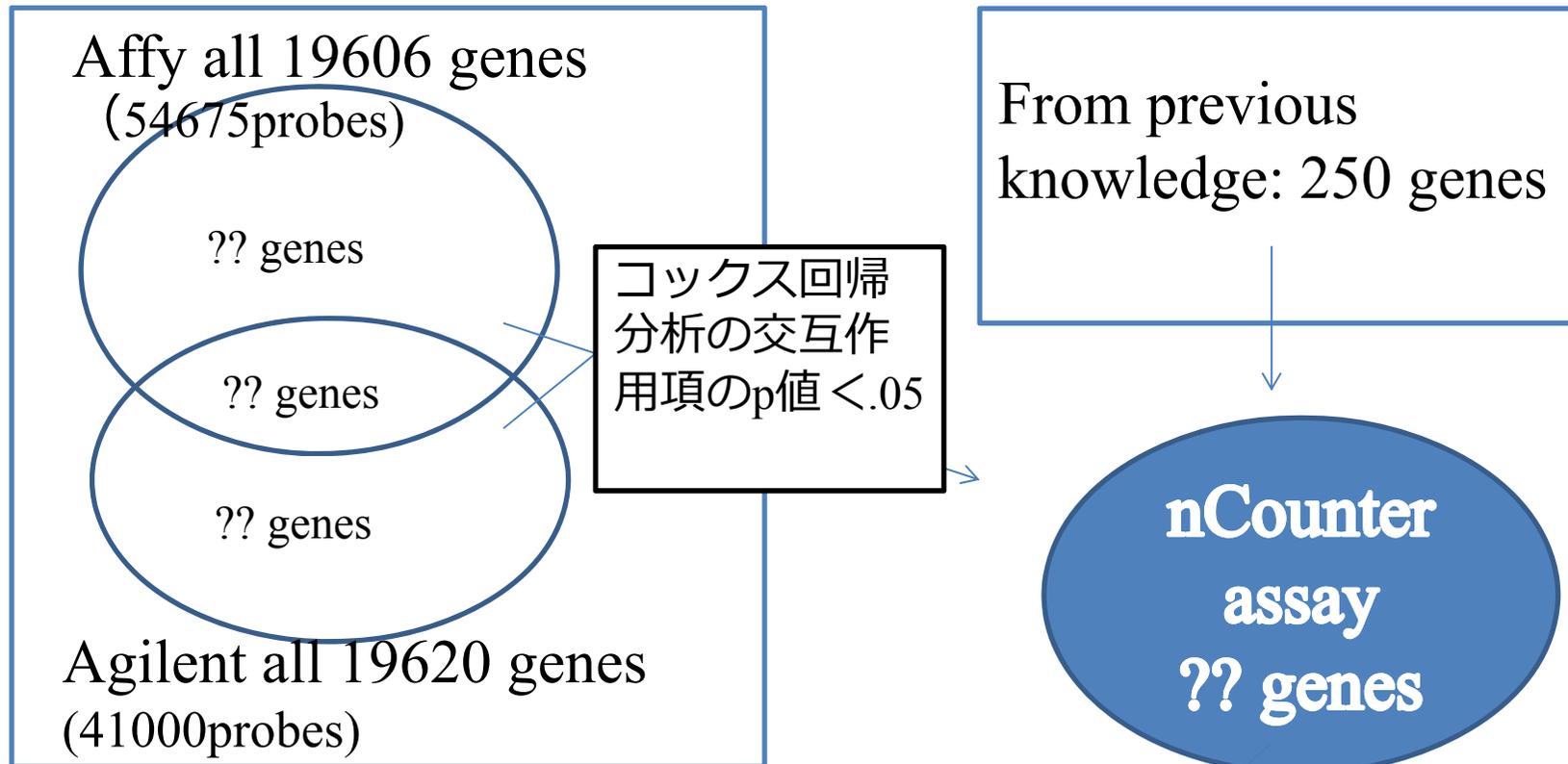
23

予測・判別に用いる方法

- 回帰分析
 - 線形/非線形回帰
 - ロジスティック回帰
- 判別分析
 - 線形/非線形判別
 - ベイズ流判別
- SVM (support vector machine)
- 決定木 (decision tree)
 - CART
 - MART
 - Random Forest...
- ニューラルネットワーク

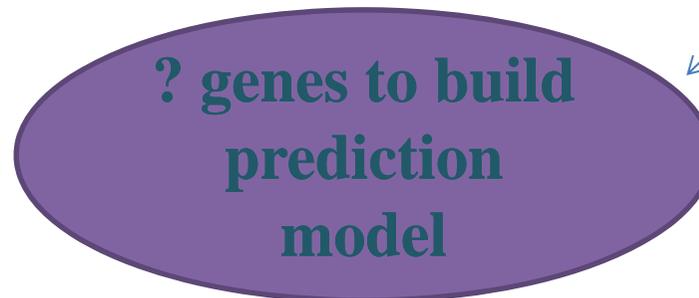
遺伝子発現データを用いた 治療効果予測モデルの構築

I



Whole genome expression analysis

III

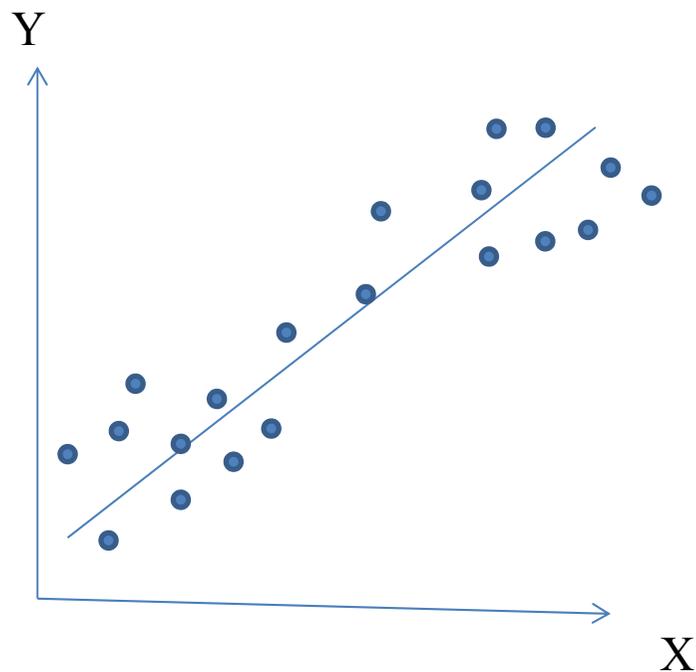


2015/3/16

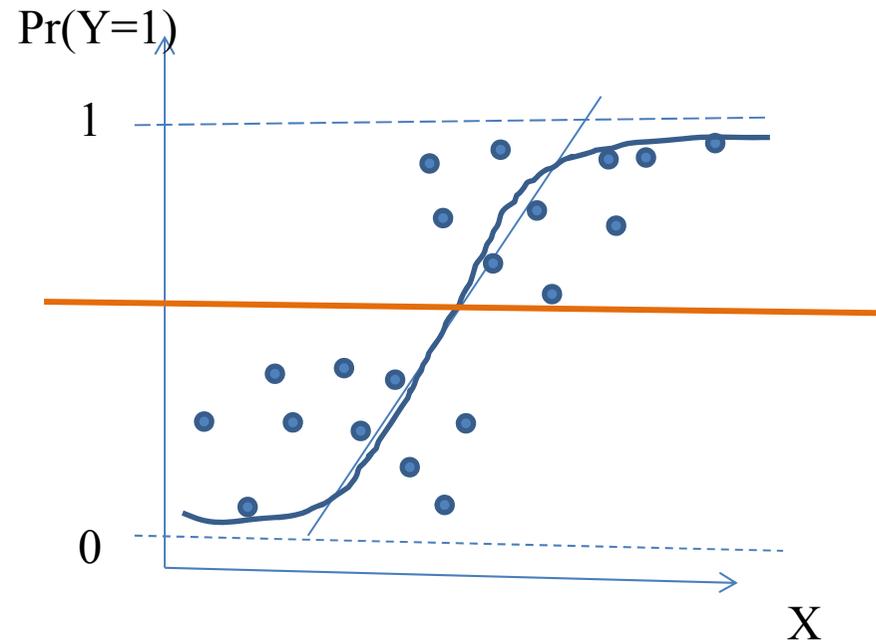
25

線形回帰による予測

- 両方連続量

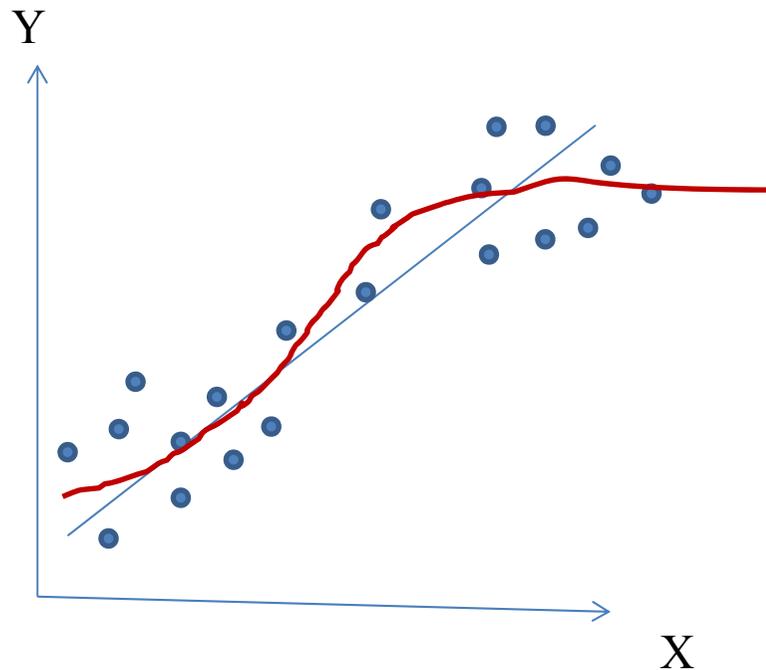


- 結果変数が2値
(ロジスティック判別)

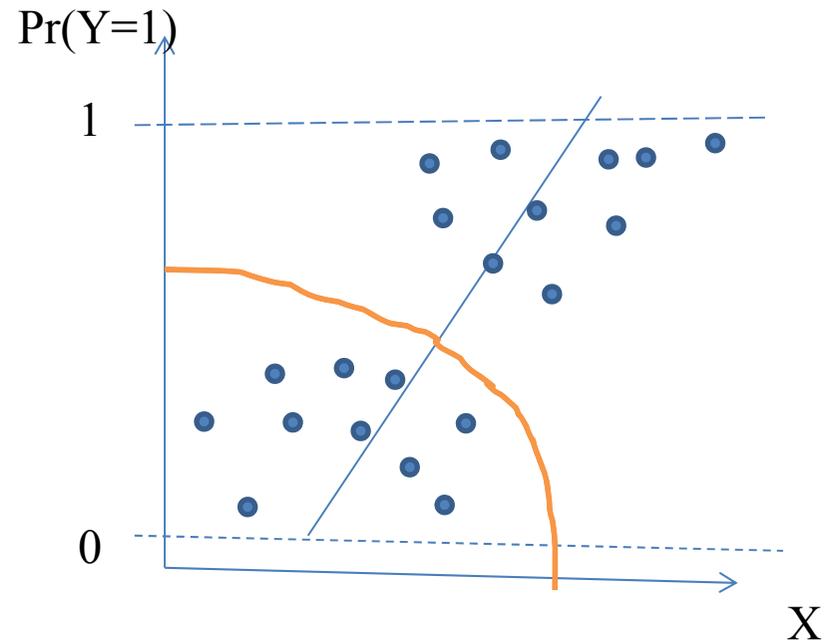


非線形回帰による予測

- 両方連続量

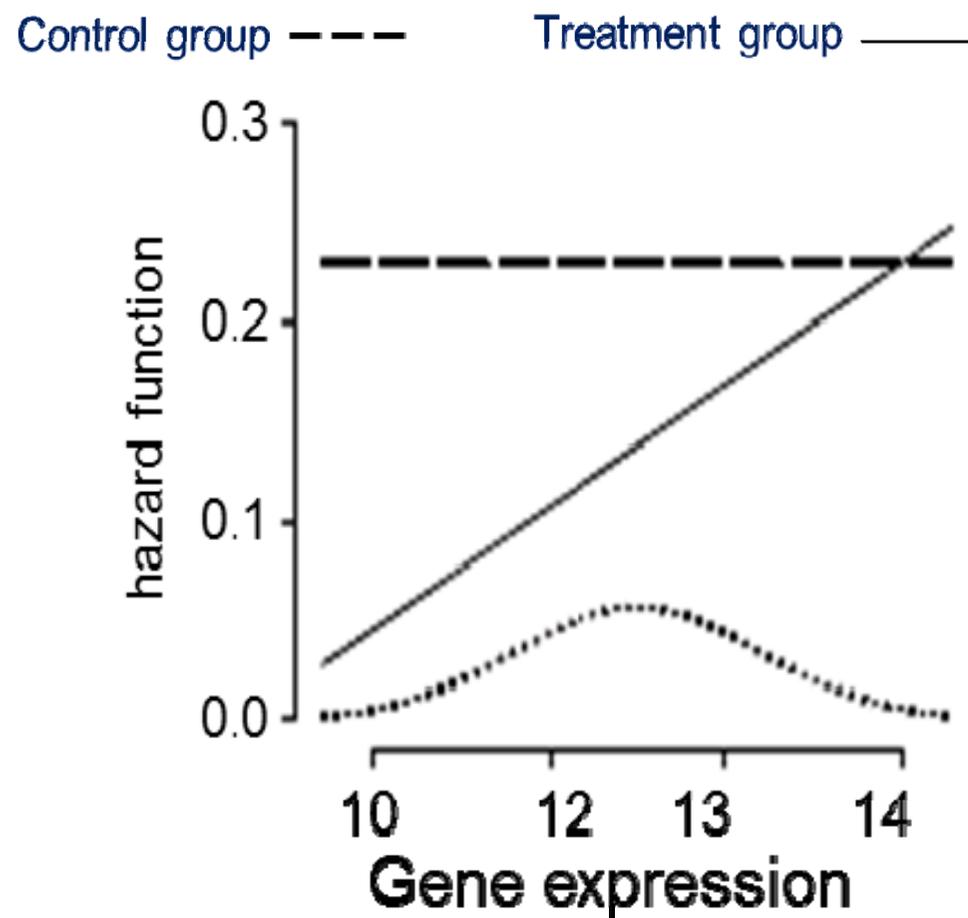


- 結果変数が2値
(非線形判別)



治療効果の予測--線形交互作用

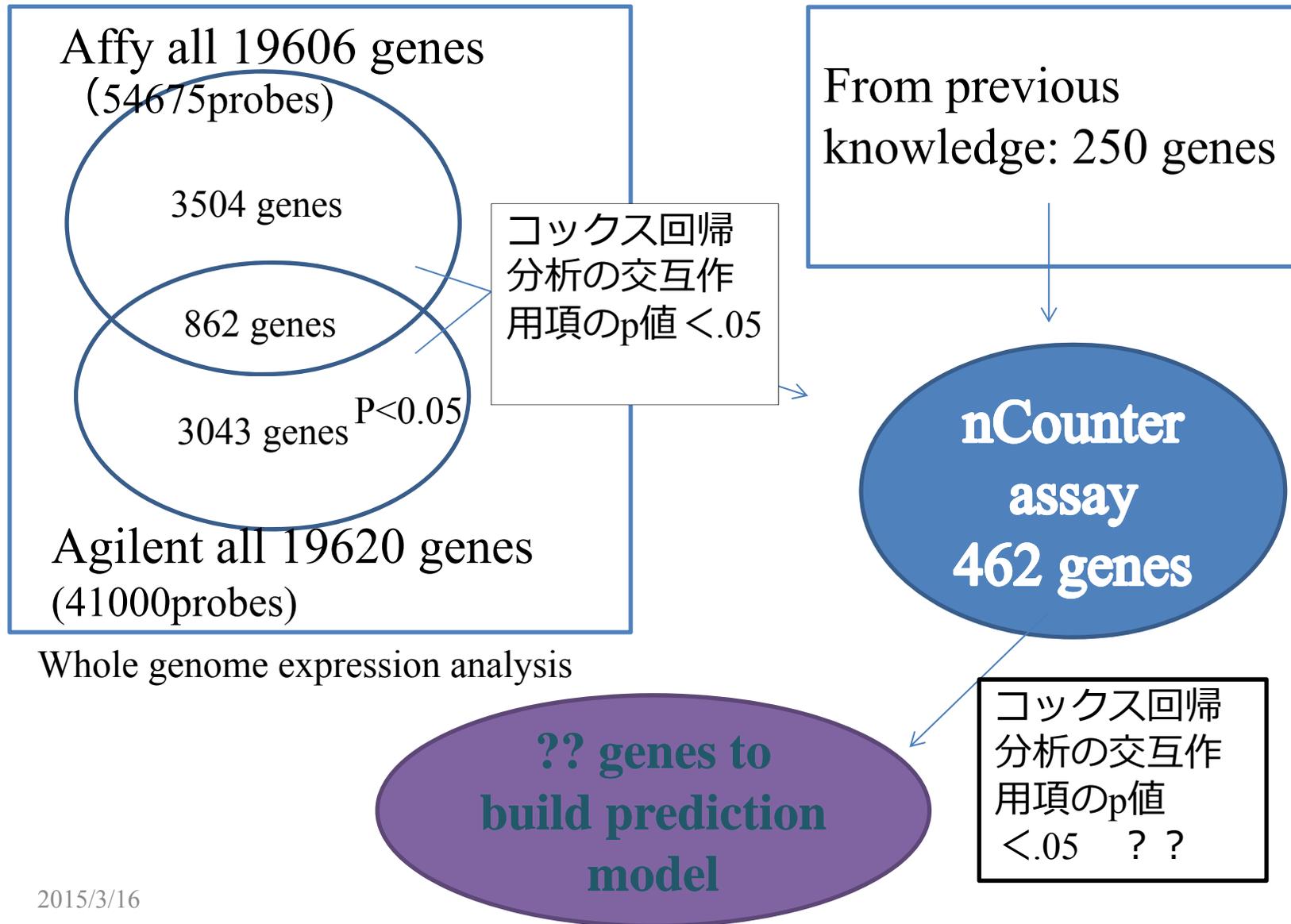
Bonetti and Gelber, 2004



I I コックス回帰分析による治療効果予測

$$\begin{aligned} \text{Log}h_i(t) = & \log\lambda_0(t) + \underbrace{\beta_1}_{\text{(治療効果)}} \times \underbrace{x_1}_{\text{(治療の種類)}} \\ & + \underbrace{\beta_2}_{\text{(j 遺伝子の効果)}} \times \underbrace{x_2}_{\text{(j 遺伝子の発現量)}} \\ & + \underbrace{\gamma}_{\text{(j 遺伝子の治療予測効果)}} \times \underbrace{x_1}_{\text{(治療の種類)}} \times \underbrace{x_2}_{\text{(j 遺伝子の発現量)}} \end{aligned}$$

Develop platform for clinical testing and selecting genes



コックス回帰分析の結果

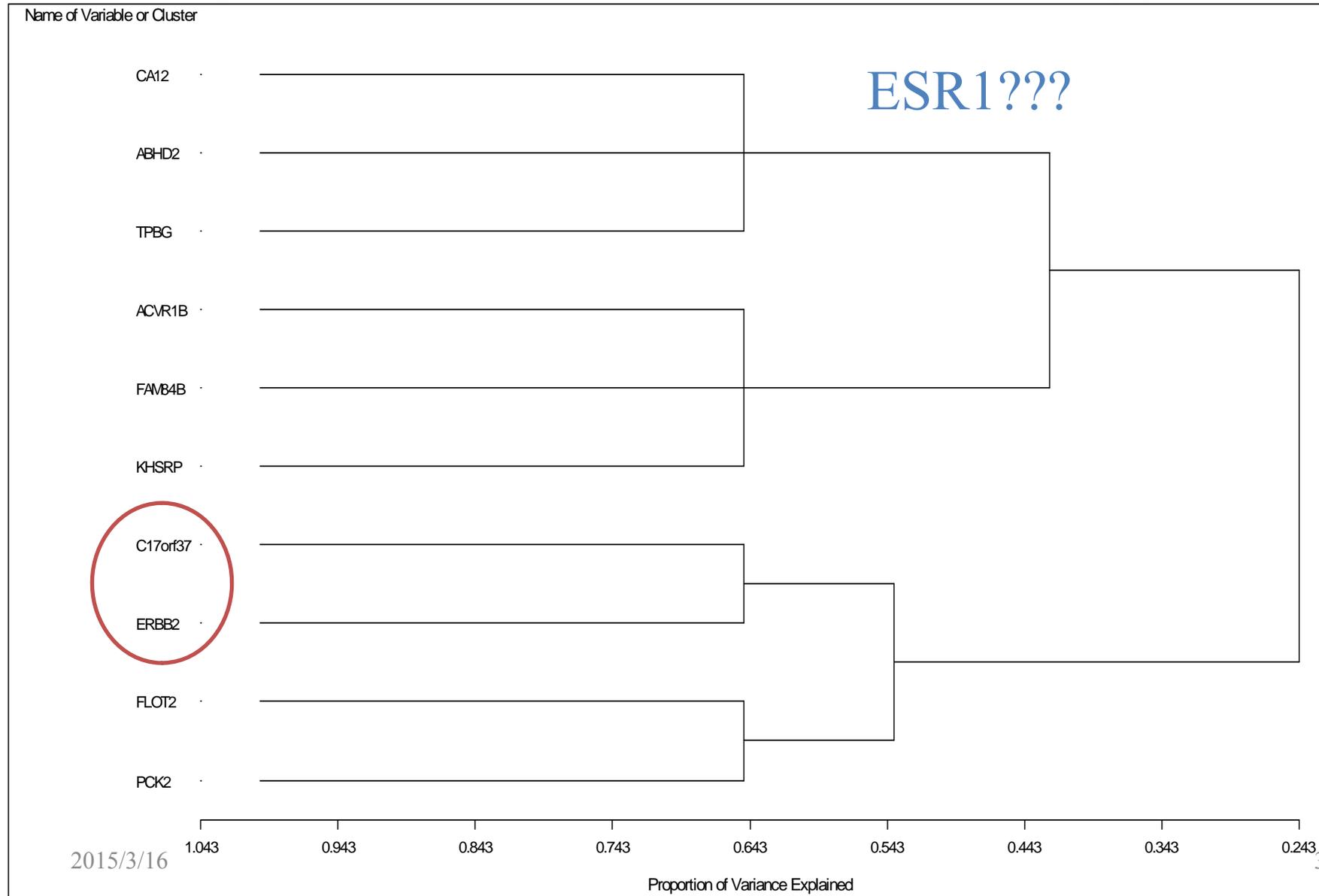
genesymbol	cv_support	mean p-value	max p-value	min p-value
FLOT2	100	0.0025	0.0054	0.0002
UNC119	100	0.0049	0.01	0.0008
TUBB2C	100	0.0051	0.0136	0.0008
XYLT1	100	0.0054	0.0131	0.0016
CA 12	100	0.0059	0.0269	0.0007
GATA3	100	0.007	0.0154	0.001
GTF3C2	90	0.0078	0.0509	0.0003
SLC39A14	100	0.0088	0.0223	0.0014
FTH1	100	0.0145	0.0347	0.0024
SUPT6H	100	0.0155	0.0385	0.0013
ACVR1B	100	0.0156	0.0349	0.0041
DKFZP434A	90	0.0166	0.0533	0.005
ILF2	90	0.0188	0.0825	0.0012
DNAJC4	90	0.0194	0.0591	0.0056
ABHD2	100	0.02	0.0477	0.002
ZACN	100	0.0214	0.0476	0.0093
TPBG	90	0.0239	0.0976	0.0041
FAM84B	100	0.0242	0.0396	0.0034
SPDEF	90	0.0243	0.0562	0.0042
DAD1	80	0.0277	0.0808	0.0074
CASC3	80	0.0297	0.1148	0.0039
MYADM	90	0.03	0.0535	0.0044
PTTG1	90	0.0316	0.1292	0.0079
UHMK1	80	0.0329	0.0827	0.0059
TMBIM6	60	0.0346	0.0666	0.0059
THOP1	80	0.0348	0.0911	0.006
ANGPTL2	90	0.0364	0.0863	0.0058
ISOC1	80	0.0366	0.139	0.005
TMSB10	90	0.0379	0.086	0.0131
PIK3CA	90	0.0388	0.2252	0.0056
SLC7A2	70	0.0401	0.107	0.0097
ORC6L	60	0.0407	0.1022	0.0088
SPP1	60	0.0408	0.0607	0.0116
CD9	60	0.0411	0.0881	0.0083
PCK2	70	0.0426	0.095	0.009
CEACAM1	70	0.0433	0.097	0.0125
RPL21	60	0.0437	0.0896	0.0159
C17orf37	70	0.0442	0.1008	0.0084
KHSRP	70	0.0458	0.1119	0.016
RASSF7	70	0.0462	0.1588	0.0111
RPL34	70	0.0477	0.1475	0.0127
ERBB2	60	0.0485	0.1114	0.0064
RPL23A	60	0.0489	0.1281	0.0116
NUF2	60	0.0497	0.1363	0.0083
EGFR	50	0.0516	0.0997	0.0122

AFFY/AGIでも有意差があった遺伝子

44 genes are still candidates

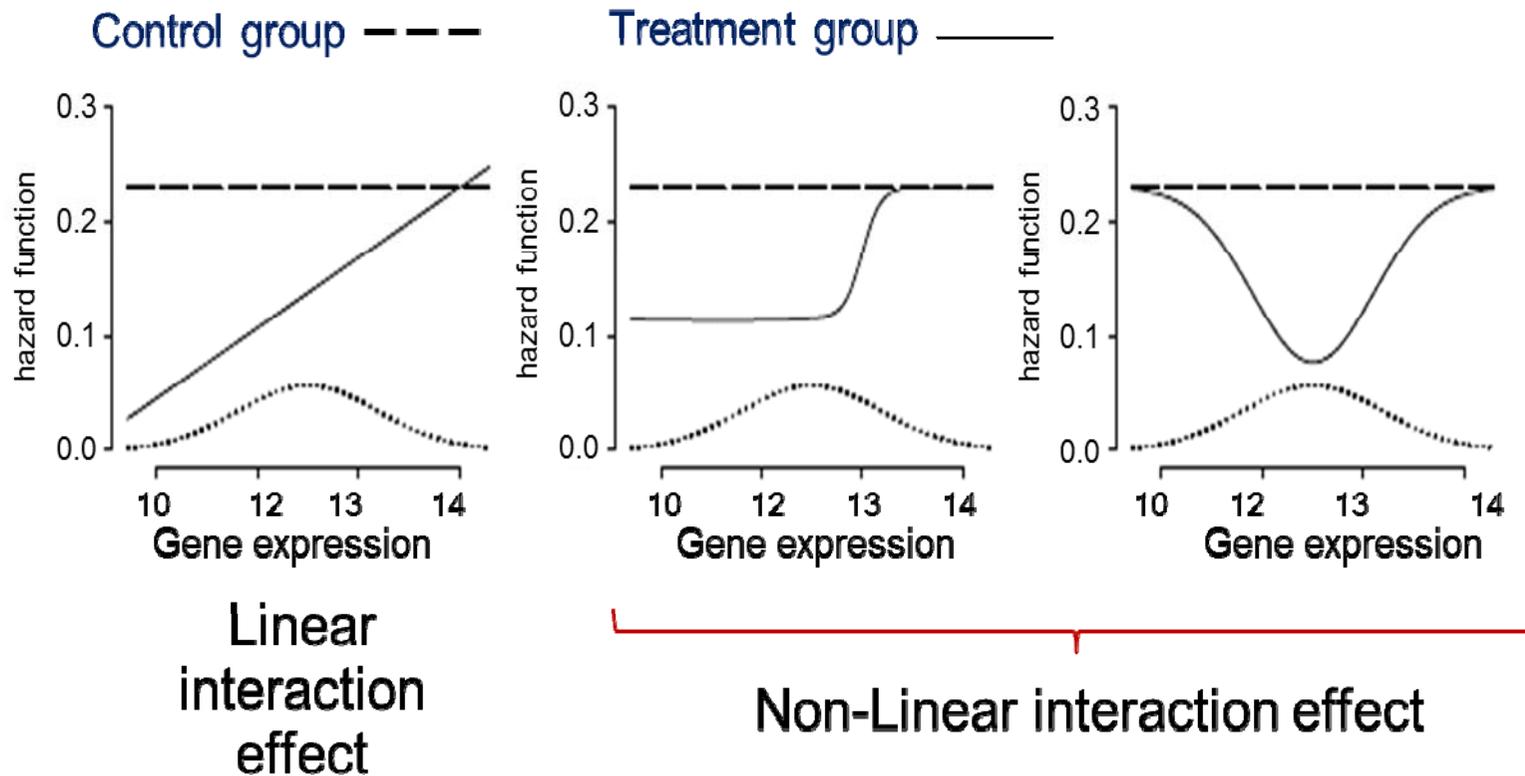
クラスター分析の結果

Significant candidates with 100% CV support



非線形交互作用

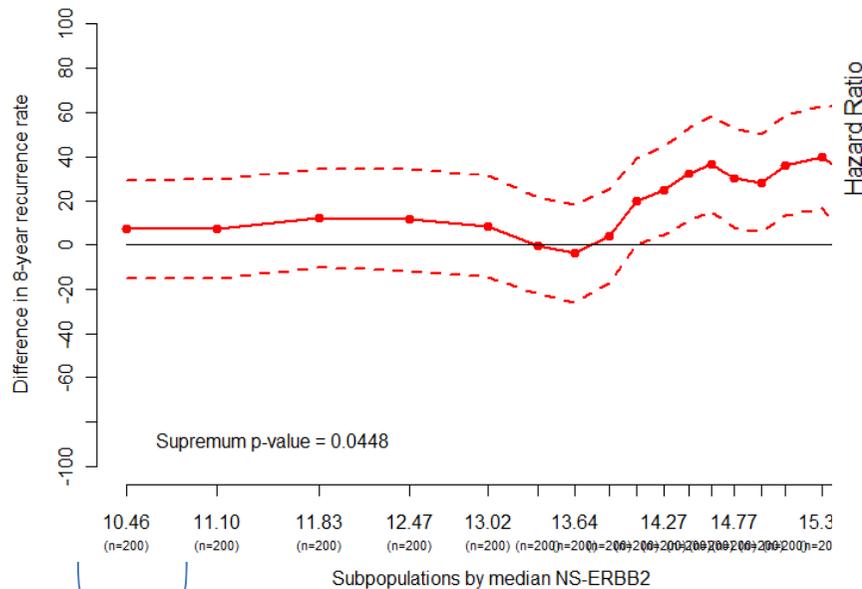
Bonetti and Gelber, 2004



Subpopulation Treatment Effect Pattern Plot (STEPP)

TANAK et al. 2012. AACR

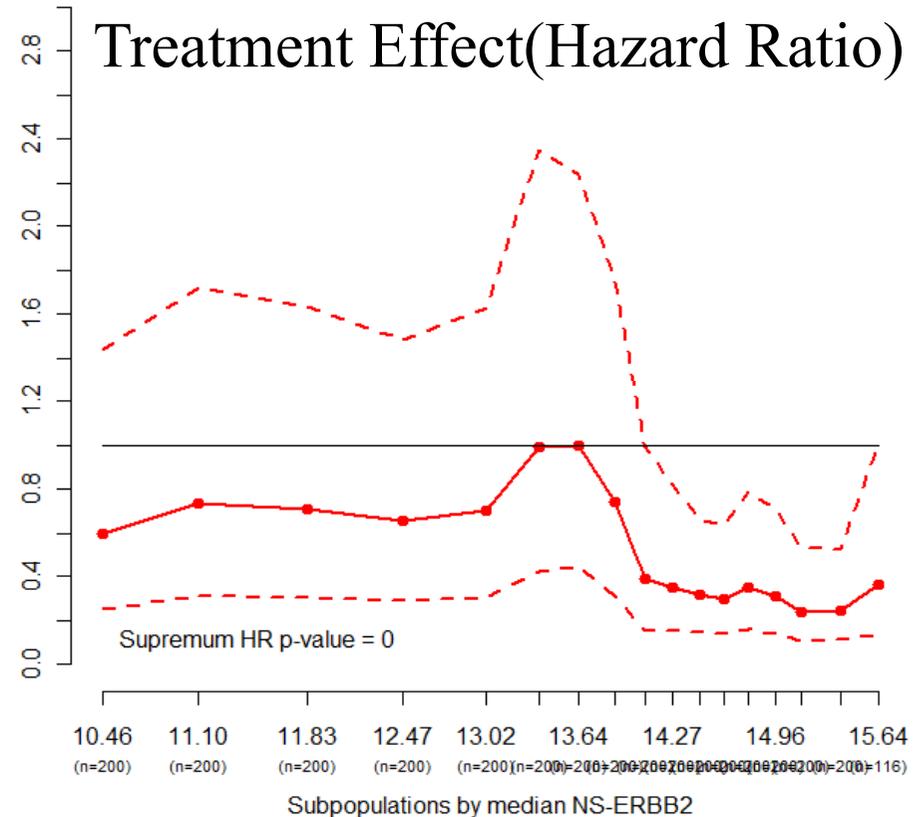
Difference of survival probability



N=200

Overlapping=50

Treatment Effect(Hazard Ratio)



2015/3/16

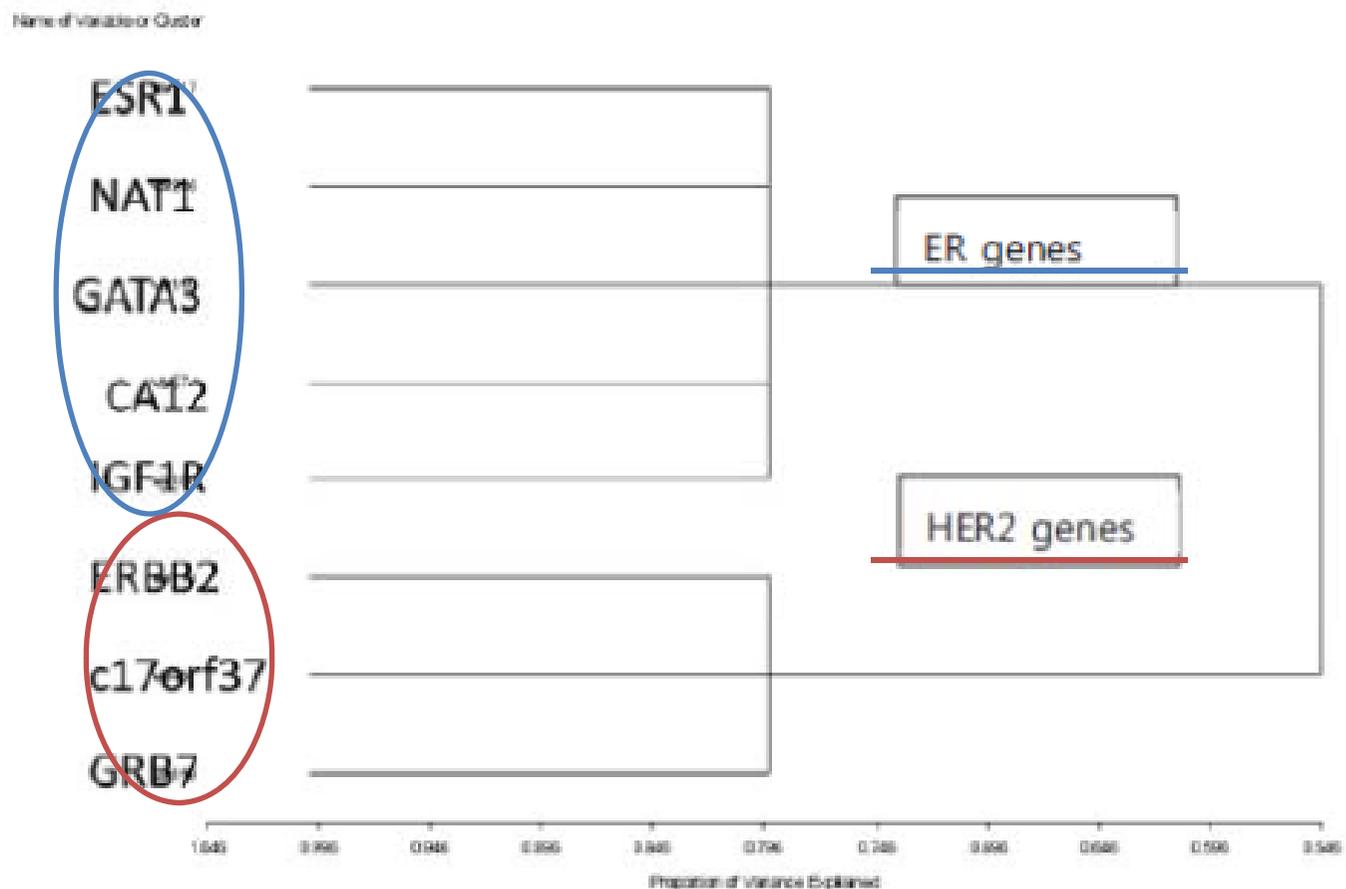
34

候補遺伝子の選び直し

Gene Symbol	Correlation with <i>ERBB2</i>	Minimum Interaction P Value
ERBB2	1	0.025
GRB7	0.912	0.06
C17orf37	0.833	0.0003
KRT7	0.498	0.047
TMEM45B	0.453	0.29
ORMDL3	0.448	0.076
C1orf93	0.427	0.1
SPDEF	0.4	0.013
VEGFA	0.395	0.24
FGFR4	0.347	0.35

Gene Symbol	Correlation with <i>ESR1</i>	Minimum Interaction P Value
ESR1	1	0.064
TBC1D9	0.757	0.49
CA12	0.733	0.0024
IGF1R	0.731	0.042
GATA3	0.727	0.0036
THSD4	0.727	0.12
NAT1	0.701	0.075
SLC39A6	0.685	0.21
SCUBE2	0.637	0.47
SIAH2	0.632	0.19

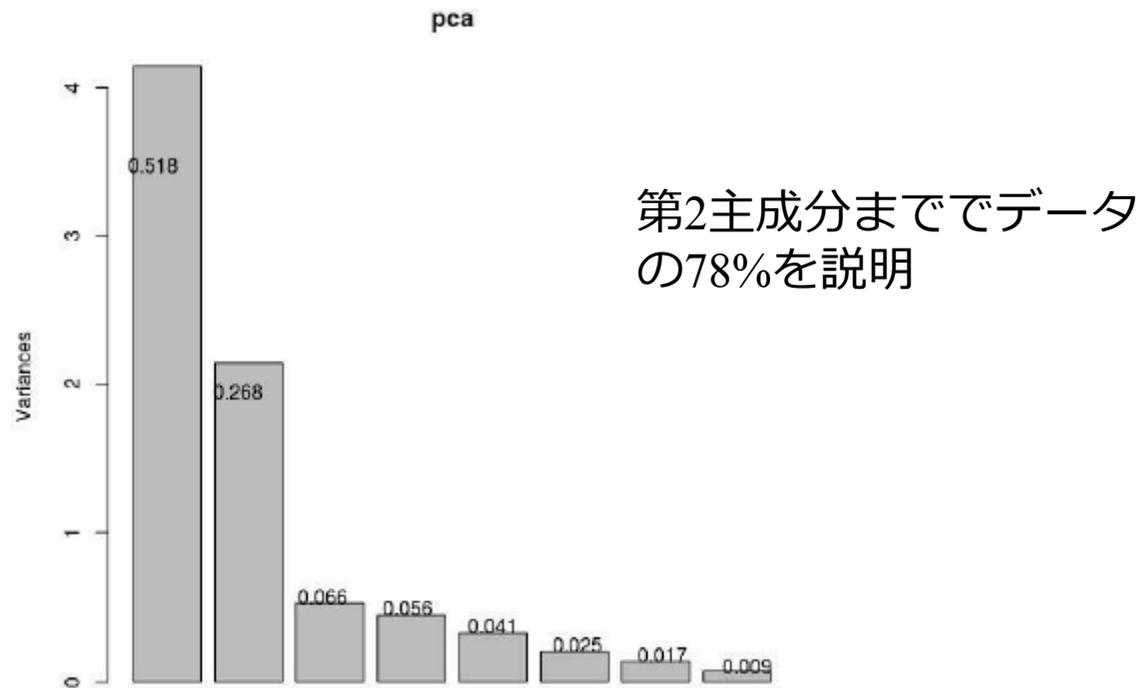
クラスター分析の結果



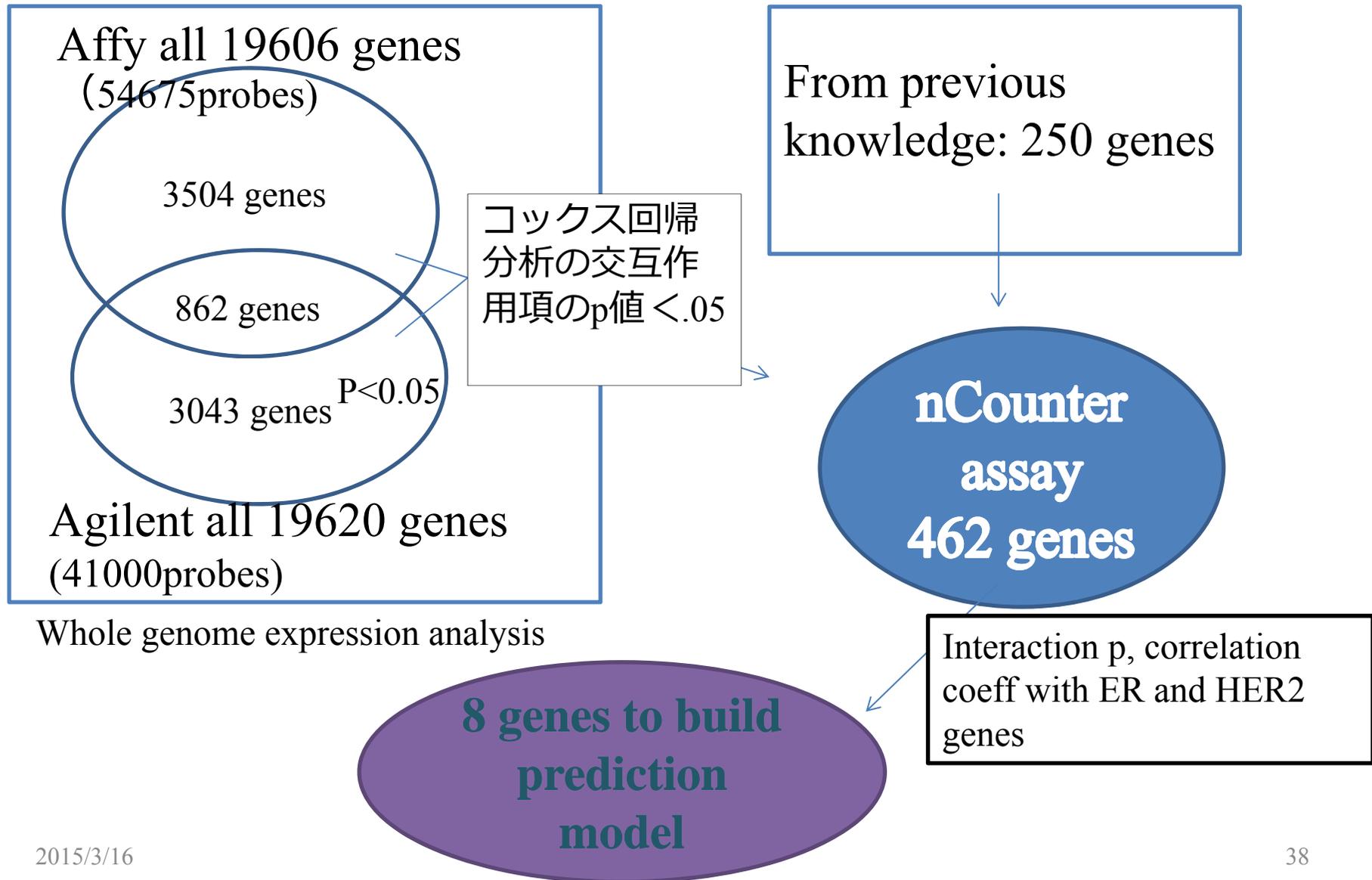
主成分分析の結果

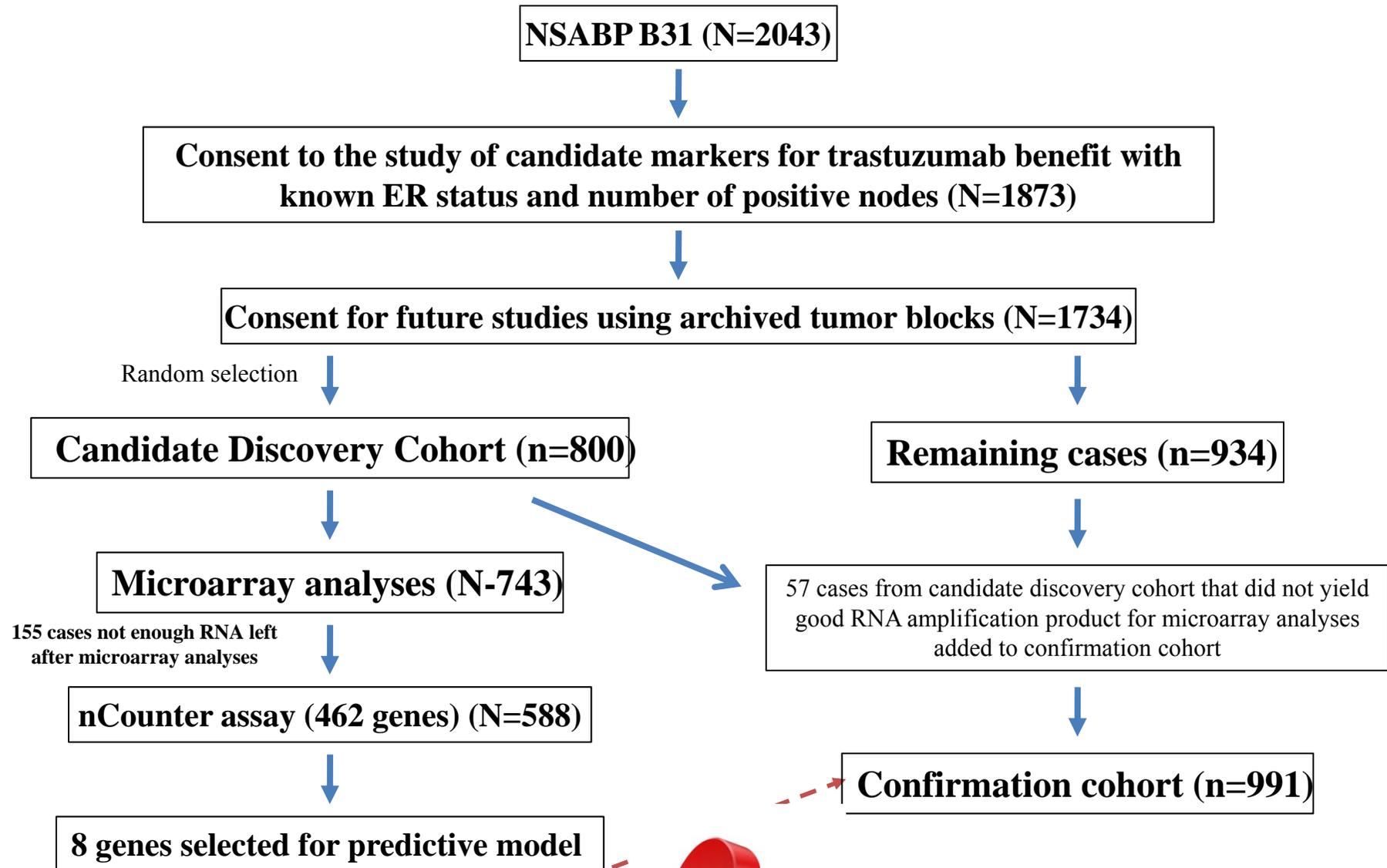
-Principal Component Analysis (PCA)-

- 主成分分析--互いに相関のある多くの変数から互いに無相関な少数個の変数（主成分）にデータを要約（=データの次元を縮小：dimensionality reduction）する手法。



Develop platform for clinical testing and selecting genes





予測・判別に用いる方法

- 回帰分析
 - 線形/非線形回帰
 - 主成分回帰（予後予測のための主成分コックス回帰：SuperPC)
 - ロジスティック回帰
- 判別分析
 - 線形/正準判別/非線形判別
 - ベイズ流判別
- SVM (support vector machine)
- 決定木 (decision tree)
 - CART
 - MART
 - Random Forest...
- ニューラルネットワーク

判別分析(Discriminant Analysis)

- 多変量を使って群間のマハラノビス距離を最大にするように標本を群に分類
- 基本的に、すべての変数に正規分布を仮定
- 分類された小集団における分散共分散構造は同じ

正準判別分析 (Canonical Discriminant Analysis)

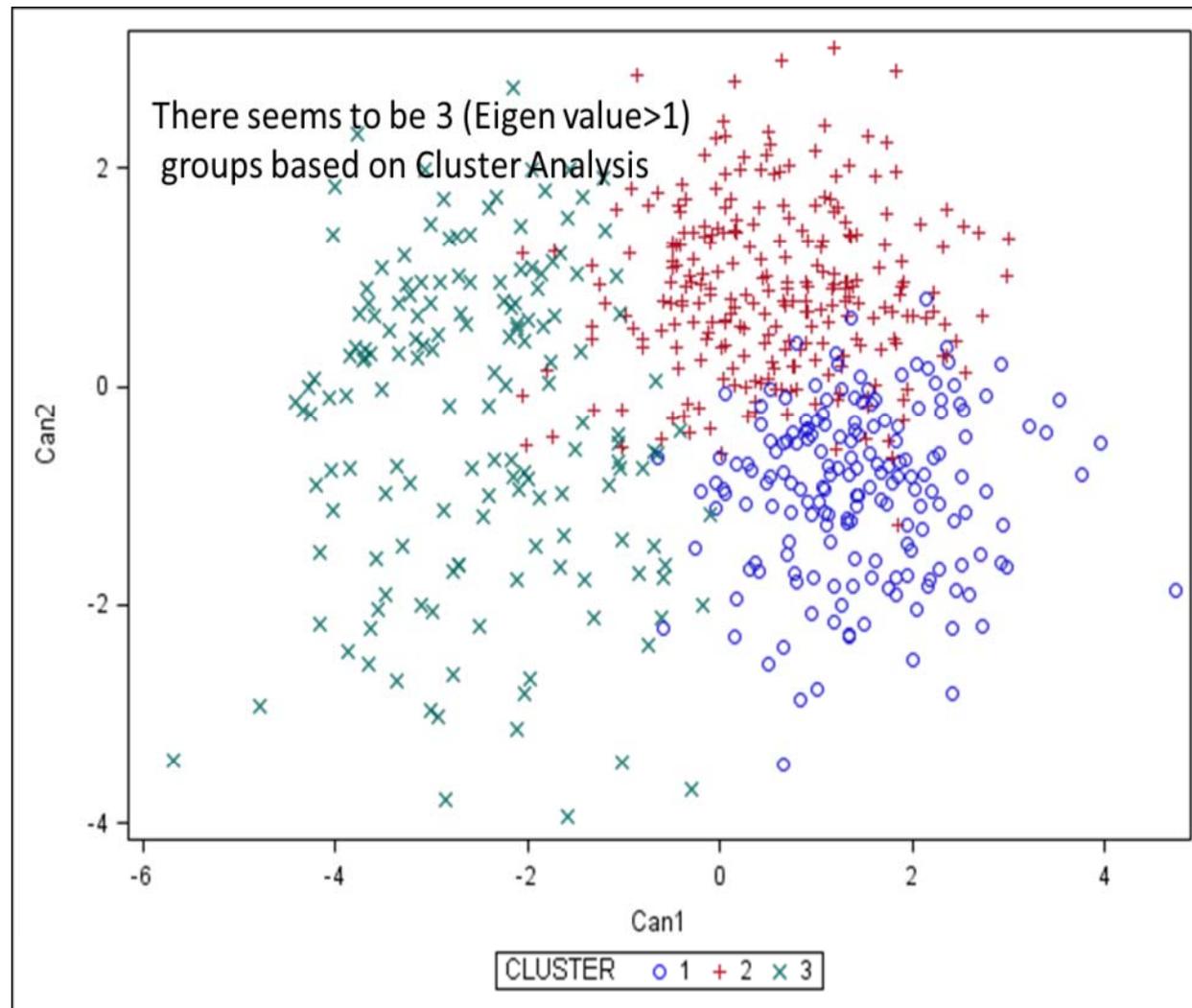
- 群内の相関をなるべく大きく群間の相関をなるべく小さくするように判別
- 正準判別関数に基づいて判別

$$F_{ki} = u_0 + u_1 X1_{ki} + u_2 X2_{ki} + \dots + u_p Xp_{ki}$$

i オブザーベーション
k 群
p 変数

- ここから求められる、次元を縮小した正準変量 (canonical variable) で群分け

正準判別分析の結果

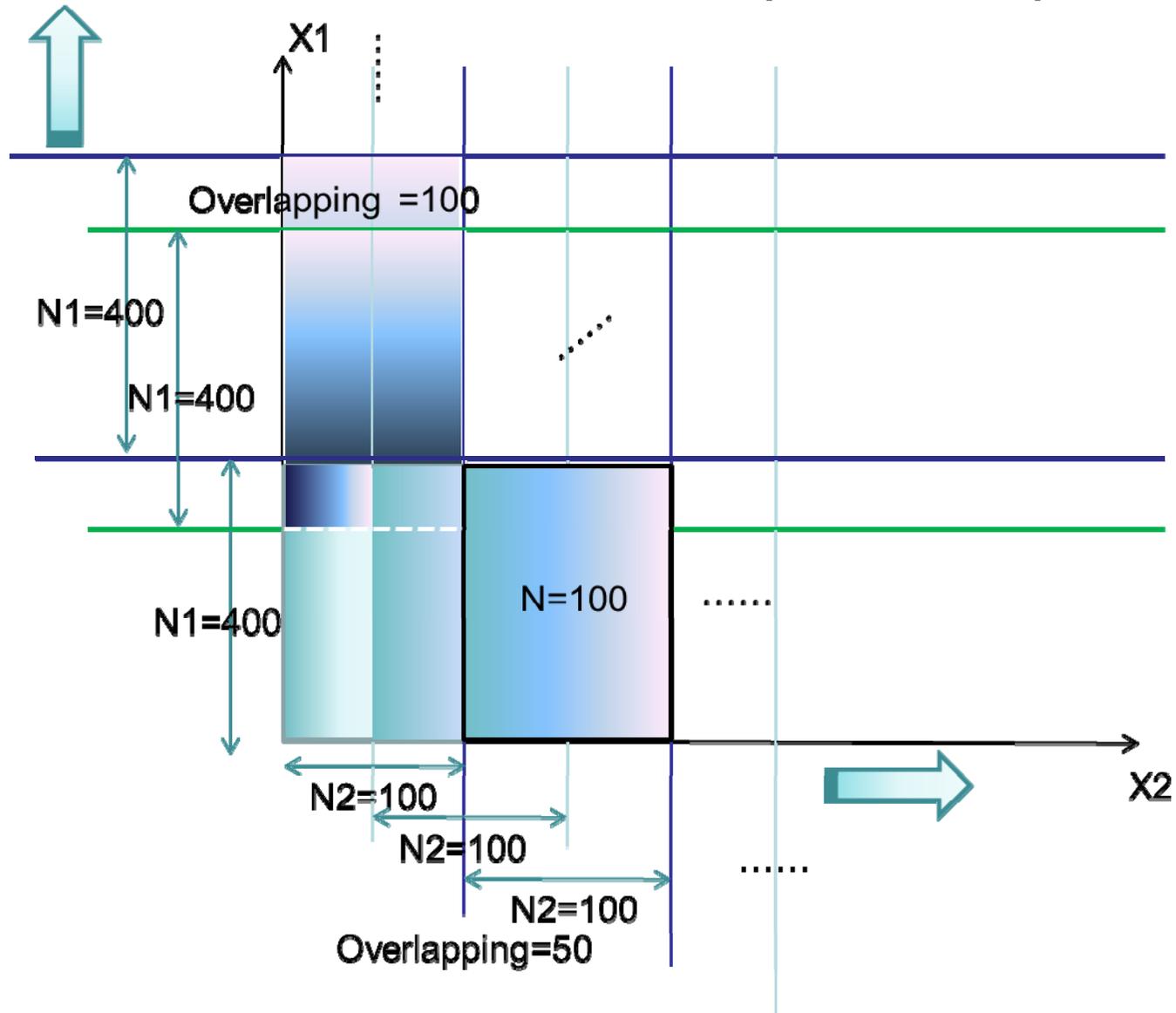


2015/3/16

Pogue-Geile et al. 2013. JNCI

43

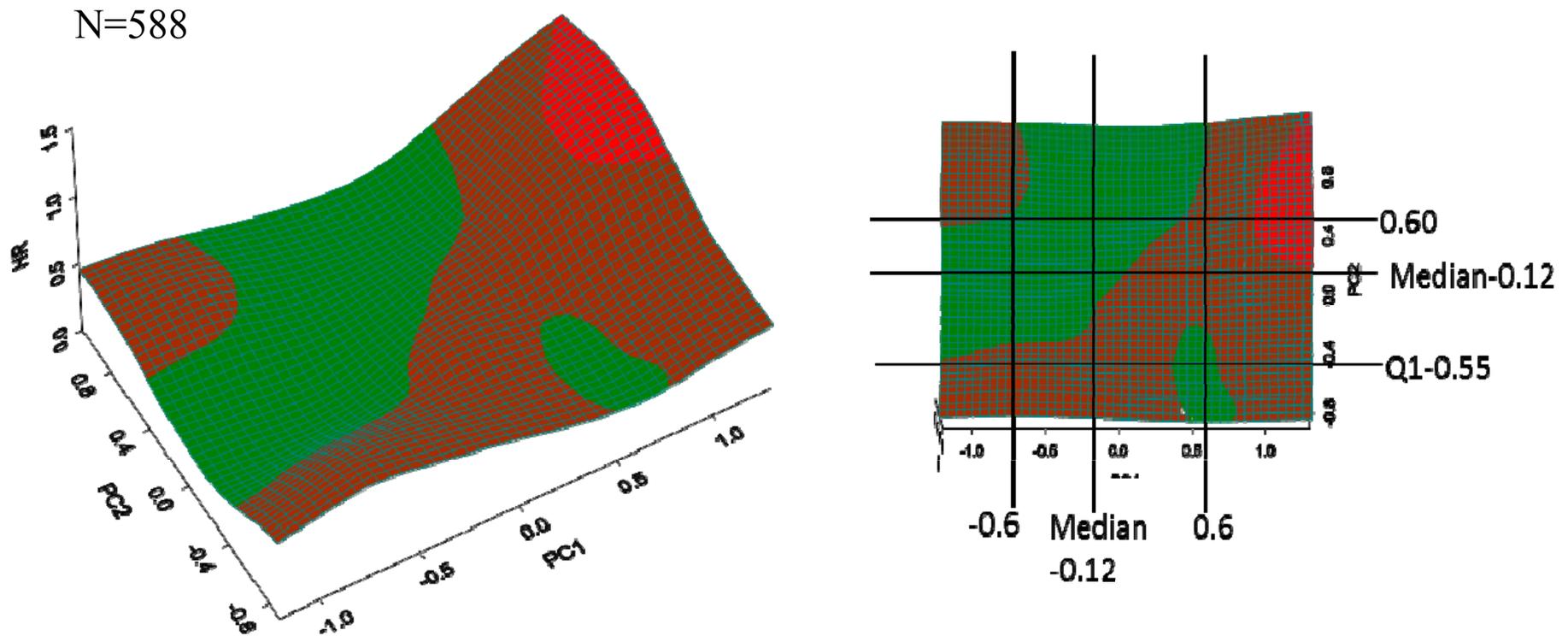
Three-Dimensional STEPP (TDSTEPP)



2015/3/16

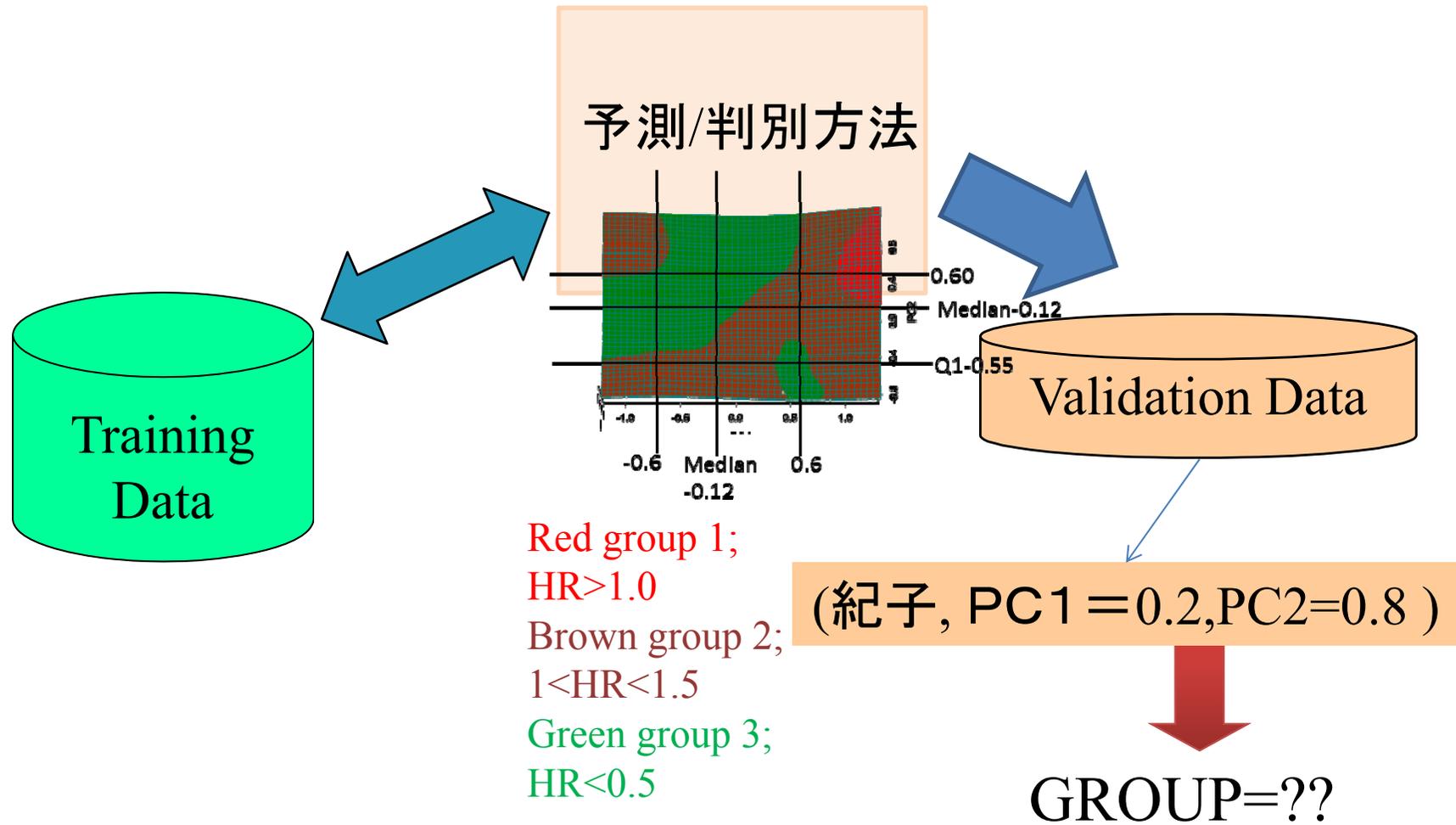
44

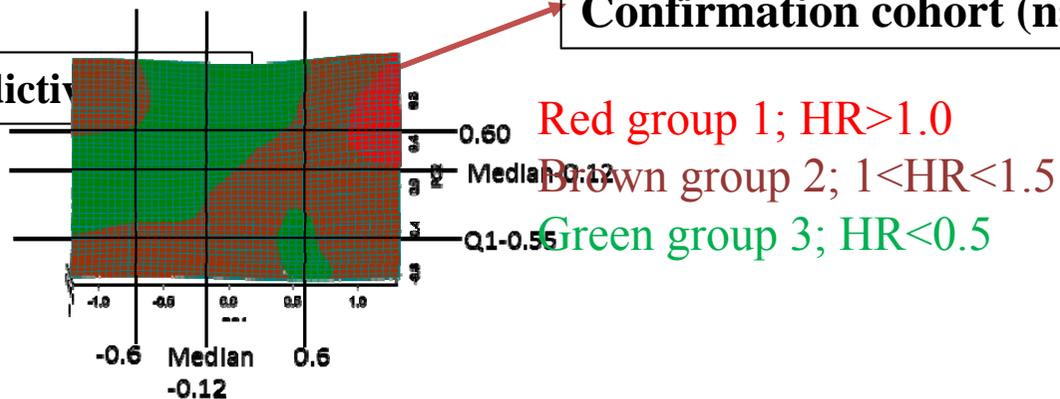
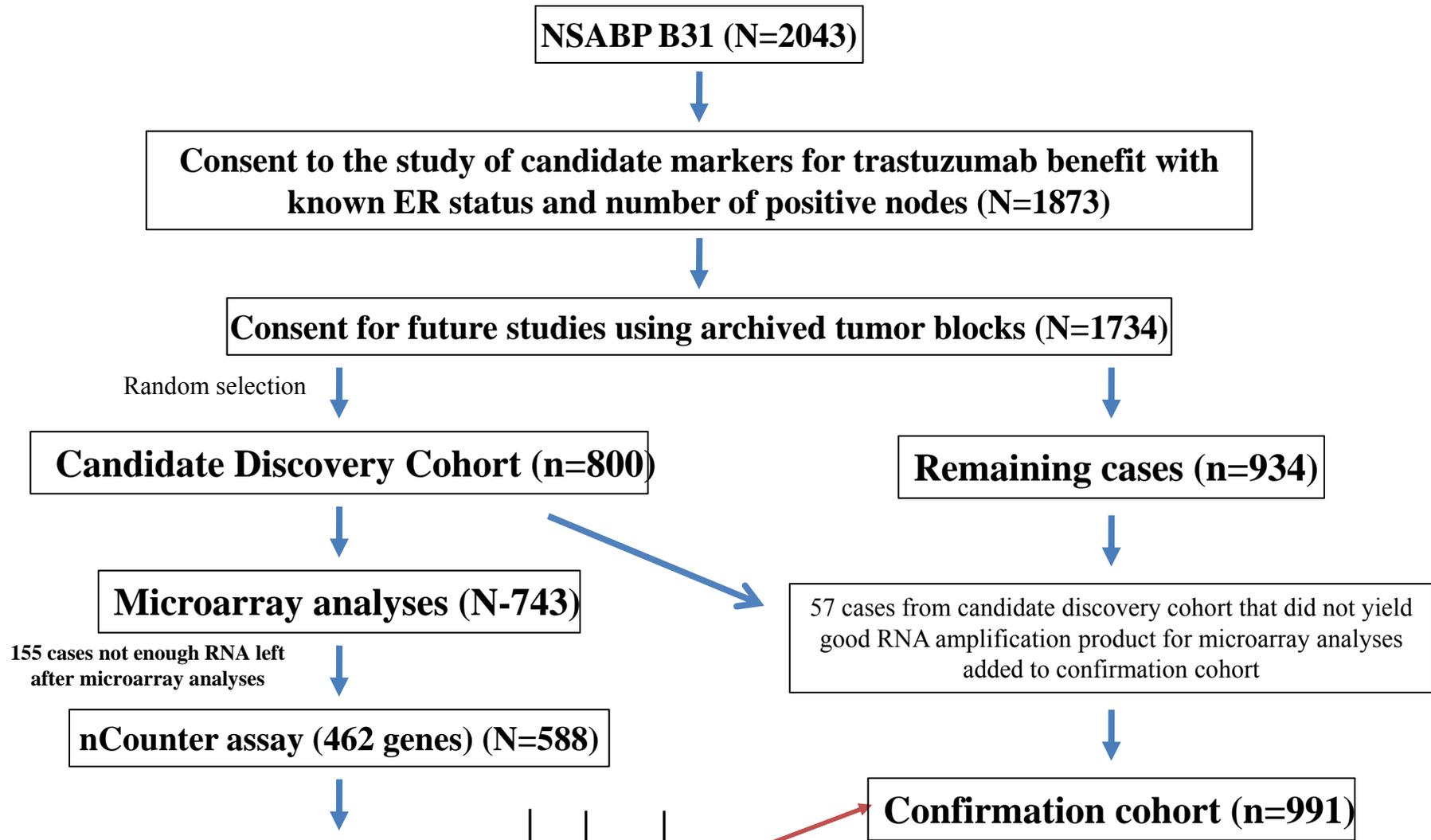
TDSTEPP (Three Dimensional STEPP) for B31 trial data (選択された遺伝子の発現量による治療効果予測結果)



Red group 1; $HR > 1.0$
Brown group 2; $1 < HR < 1.5$
Green group 3; $HR < 0.5$

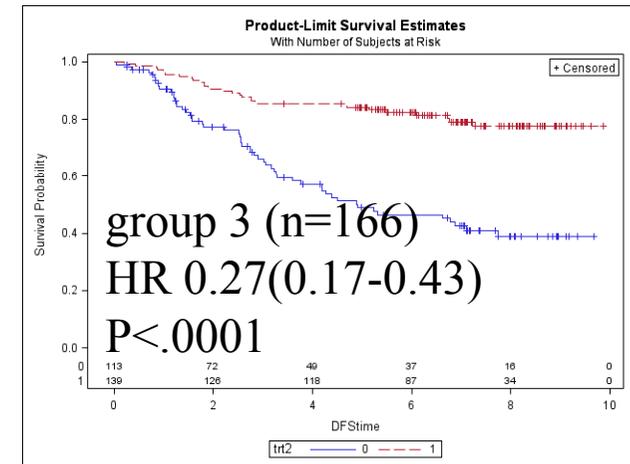
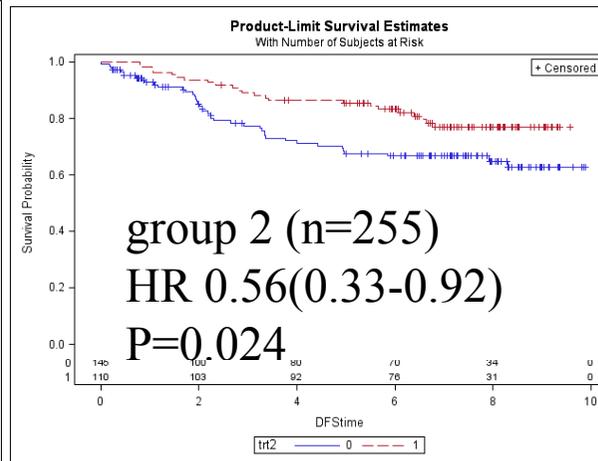
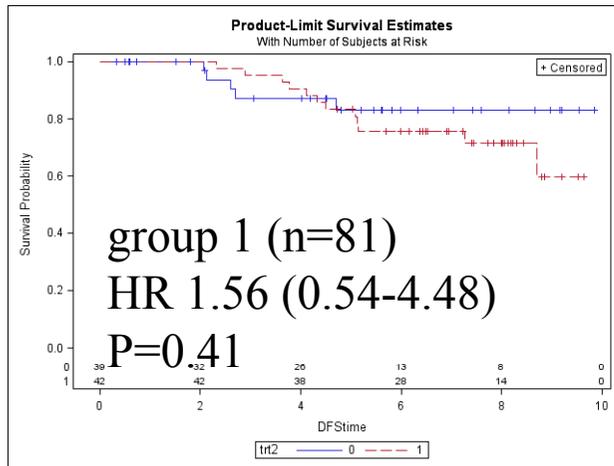
予測/判別モデル構築方法



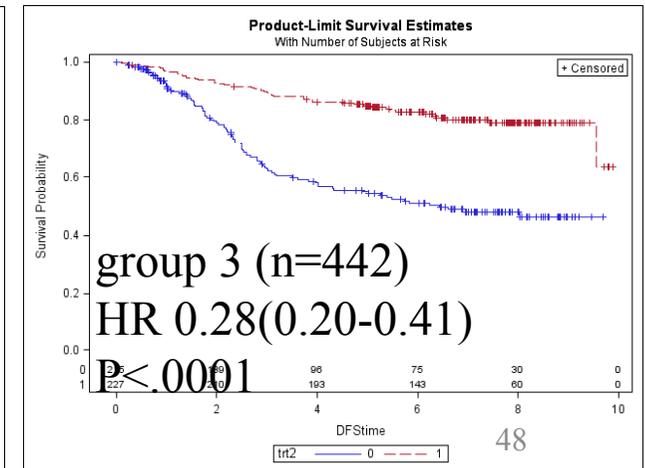
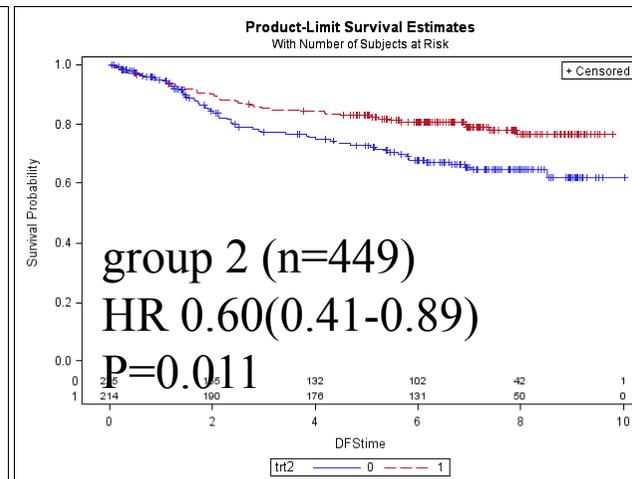
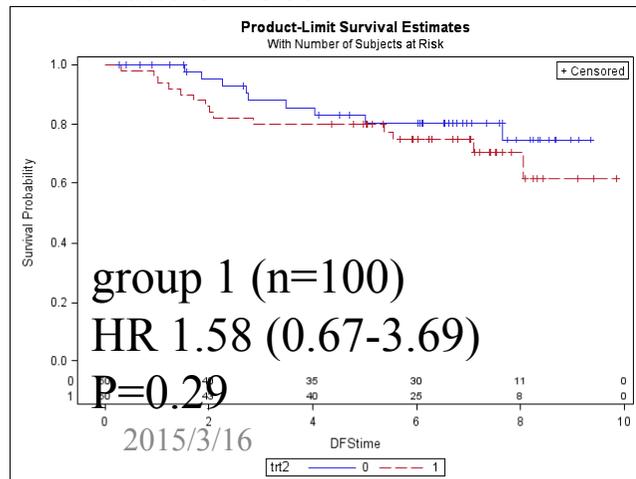


KM Plots for 3 groups defined by TDSTEPP

Discovery Set



Validation Set

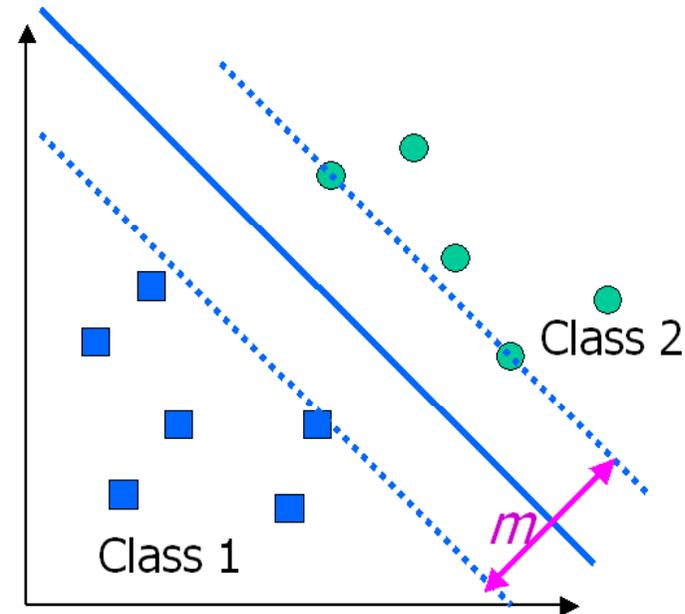
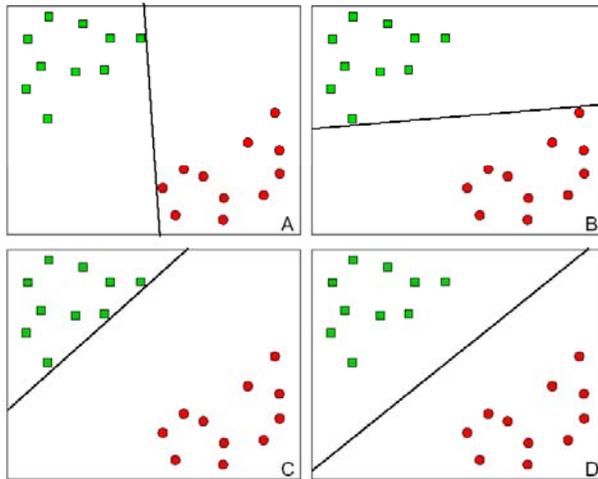


予測モデルの妥当性確認

- 鳥瞰図の利点
 - 当てはまりすぎなどを視覚的に確認可能
 - スコア計算可能な非線形の式を導出可能
- 鳥瞰図の欠点
 - 結構恣意的？
 - 他の方法での判別力を確認してみよう！

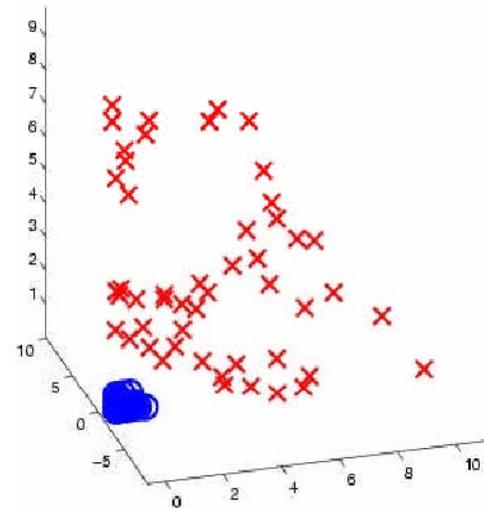
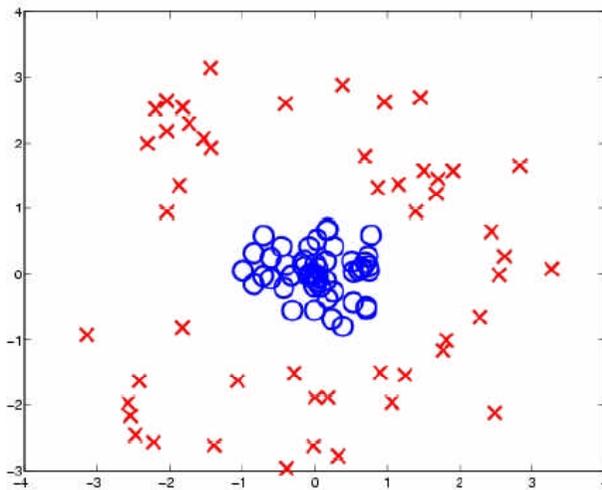
SVM (Support Vector Machine)

- SVMは2群に分離したときに、分離するための平面と最も近いデータとの距離（マージン）を最大にするような分離超平面を選んで分離する



Nonlinear SVM

- 非線形の場合でも、次元を上げたり、カーネル関数によるデータ変換を行うことで分離可能



$$K(x,y) = (xy)^2$$

- Matlab & BioConductorで実装可能

SVMによる判別力の確認

- 80%の一致率

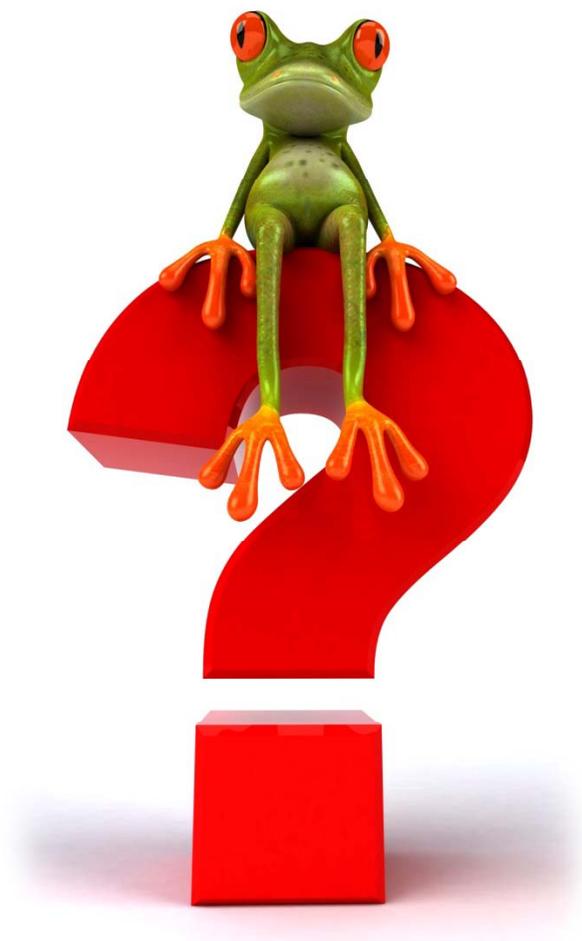
予測や判別モデルのよさを図る指標

- 精確度
 - 判別：誤判別率
 - 回帰： r^2 、Cインデックス、誤差
- 計算速度
 - トレーニング速度
 - 実用速度
- 頑健性
 - ノイズや欠測からの影響度
- 拡張性:
 - 大規模データへの適用可能性
- 解釈可能性

- その他
 - 簡便性、文化へのなじみ、大衆性

まとめ

- 予測→ある時点でのある変数からある結果を導く法則（式など）
- 判別→ある変数で既知の群分けができるかどうか、その法則（式など）を見つける
- 似たようなことをやっているが、考え方と当てはめるデータが違います
- 予測モデルや判別モデルは、変数をいっぱい入れればコンピュータで自動的に作れるものではありません



2015/3/16

55

参考文献

- Pogue-Geile K, Kim C, Jong HJ, Tanaka N, et al. Predicting Degree of Benefit from Adjuvant Trastuzumab in NSABP Trial B-31. J Natl Cancer Inst.2013. 105(23):1782-8
- Bonetti M, Gelber RD. Patterns of treatment effects in subsets of patients in clinical trials. Biostatistics 2004;5:465-81.