ゲノム研究の解析
– 臨床試験の附随研究としての –

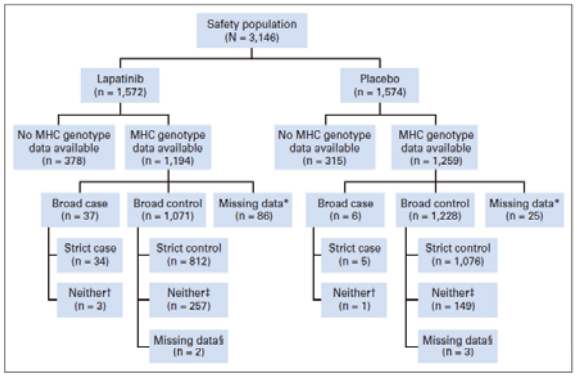
JCOGデータセンター統計部門
柴田大朗

JCOG臨床試験セミナー中級編 2015年10月10日

Outline

- 臨床試験データによるゲノム解析研究の特徴
- ゲノム研究におけるデータの取り扱い
- JCOGでの事例とプロトコール記載の概略
- 臨床情報との関連の統計解析方法の例
- omicsデータのQC-Bioinformaticianの指摘
- プロトコール作成時の留意点

Motivating example



VOLUME 32 · NUMBER 22 · AUGUST 1 2014
JOURNAL OF CLINICAL ONCOLOGY ORIGINAL REPORT

Prospective Validation of *HLA-DRB* Carriage As a Predictive Risk Factor Liver Injury

Daniel J. Schaid, Colin F. Spraggs, Shannon K. McDonnell, Laura Dianne M. Finkelstein, Erica Rappold, Joan Curran, Lon R. Cardo

Daniel J. Schaid and Shannon K. McDonnell, Mayo Clinic, Rochester, MN; Colin F. Spraggs and Charles J. Cox, GlaxoSmithKline Research and Development, Stevenage, Hertfordshire; Joan Curran, GlaxoSmithKline Research and Development, London, United Kingdom; Laura R. Parham, GlaxoSmithKline Research and Development, Research Triangle Park, NC; Bent Ejlertsen, Rigshospitalet, Copenhagen, Denmark; Dianne M. Finkelstein and Paul E. Goss, Massachusetts General Hospital, Boston, MA; and Erica Rappold and Lon R. Cardon, GlaxoSmithKline Research and Development, Philadelphia, PA.

Published online ahead of print at www.jco.org on March 31, 2014.

Processed as a Rapid Communication manuscript.

Supported by GlaxoSmithKline and by the Avon Foundation (D.M.F. and P.E.G.).

D.J.S. and C.F.S. contributed equally to this work.

Terms in blue are defined in the glossary, found at the end of this article and online at www.jco.org.

Presented as a poster discussion at the 35th Annual San Antonio Breast Cancer Symposium, San Antonio, TX, December 4-8, 2012.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Clinical trial information: NCT00374322.

Corresponding author: Colin F. Spraggs, PhD, GlaxoSmithKline Research and Development, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, United Kingdom; e-mail: colin.f.spraggs@gsk.com.

0732-183X/14/3222w-2296w/$20.00

DOI: 10.1200/JCO.2013.52.9967

ABSTRACT

Purpose
Liver injury is a serious adverse event leading to permanen patients. This study aimed to validate previously associ (MHC) variants as predictors of risk of liver injury by using trial of lapatinib in human epidermal growth factor recept (Tykerb Evaluation After Chemotherapy [TEACH]: Lapat Early-Stage Breast Cancer).

Patients and Methods
The frequency of ALT elevation cases was compared amc randomly assigned to lapatinib. Cumulative ALT elevation t compared between carriers and noncarriers of specified

Results
In lapatinib-treated patients, there was a significant differer carriers and noncarriers. The highly correlated alleles *HL* (study frequency, 22.4%) were associated with ALT elevat 37) and controls (n = 1,071). These associations strengthe and in Hy's Law cases. In lapatinib-treated patients, Institute-Common Terminology Criteria for Adverse Even limit of normal) was 2.1%; HLA allele carriers had an incr was reduced to 0.5%, comparable to ALT elevation for all in ALT case incidence in the lapatinib arm showed no lapatinib treatment.

Conclusion
These results validate *HLA-DRB1*07:01* allele carriage as risk of lapatinib-induced liver injury and implicate an immune tion could support clinical management of patients experie lapatinib treatment.

J Clin Oncol 32:2296-2303. © 2014 by American Society of Clin

INTRODUCTION

Lapatinib (Tykerb/Tyverb, GlaxoSmithKline, London, United Kingdom), an inhibitor of human epidermal growth factor receptor 2 (HER2) and epidermal growth factor receptor, is approved in combination with capecitabine or letrozole for the treatment of HER2-overexpressing metastatic breast cancer (MBC).[1,2] Extensive clinical experience with lapatinib has demonstrated an acceptable safety profile. However, lapatinib treatment is associated with liver injury, with a low incidence for National Cancer Institute-Common Terminology Criteria for Adver isolated ALT ele threatening concu bilirubinemia (H hepatotoxicity req treatment; thus, prove the safety r premature disco Drug-induced liv and can affect pa dict risk of liver i Previous pharma confirmed associ

Table 2. Broadly Defined ALT > 3× ULN Cases and Controls and MHC Genotype Associations Among Lapatinib-Treated Patients

		Cases		Controls				
MHC Marker	Risk Genotype Category	Count	%	Count	%	OR†	95% CI	*P*‡
TNXB rs12153855	CC/CT genotype	18	9.1	180	90.9	4.68	2.41 to 9.09	1.3×10^{-5}
	Noncarrier	19	2.1	889	97.9	1.00 (reference)		
	Total	37		1,069				
*HLA-DRB1*07:01*	Carrier (heterozygous or homozygous)	29	11.7	218	88.3	14.12	6.36 to 31.32	2.4×10^{-13}
	Noncarrier	8	0.9	849	99.1	1.00 (reference)		
	Total	37		1,067				
*HLA-DQA1*02:01*	Carrier (heterozygous or homozygous)	29	11.7	218	88.3	14.08	6.35 to 31.24	2.6×10^{-13}
	Noncarrier	8	0.9	847	99.1	1.00 (reference)		
	Total	37		1,065				
*HLA-DQB1*02:02*	Carrier (heterozygous or homozygous)	23	11.9	170	88.1	8.64	4.36 to 17.13	7.9×10^{-10}
	Noncarrier	14	1.5	894	98.5	1.00 (reference)		
	Total	37		1,064				

Abbreviations: MHC, major histocompatibility complex; OR, odds ratio; ULN, upper limit of normal.
†OR and 95% CIs are from logistic regression.
‡Fisher's exact test *P* value. Genes are ordered according to their position on chromosome 6: *TNXB* is approximately 469 kb telomeric from *HLA-DRB1*, and the HLA genes *DRB1*07:01*, *DQA1*02:01*, and *DQB1*02:02* span approximately 88 kb.

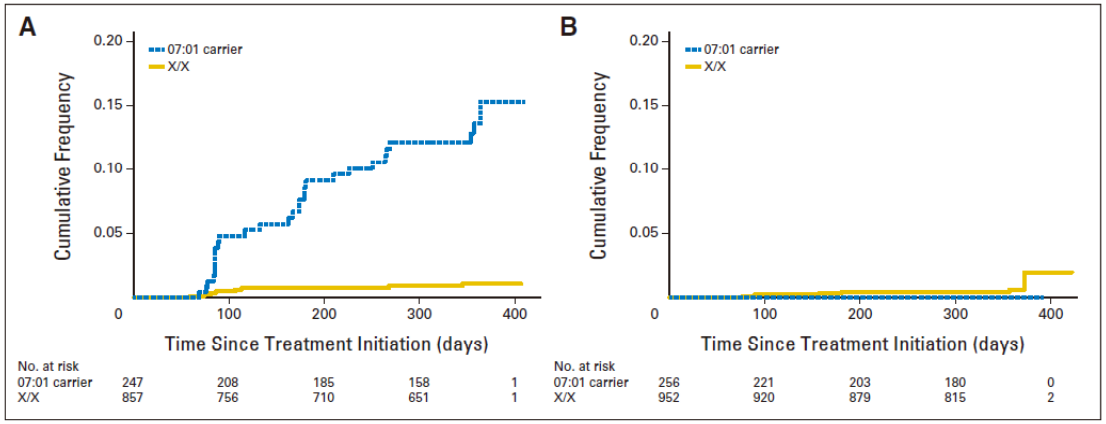


Fig 2. Cumulative incidence of ALT elevations to more than 3× upper limit of normal in *HLA-DRB1*07:01* allele carrier and noncarrier (X/X) subsets for (A) lapatinib treatment compared with (B) placebo treatment. These cumulative frequencies account for censoring as a result of incomplete follow-up, estimated by Kaplan-Meier survival curves subtracted from one. The numbers along the bottom of the *x*-axis are numbers of patients observed over time, carriers above noncarriers.

Schaid et al. JCO, 2014.

もし臨床試験でなくカルテ調べだったら？

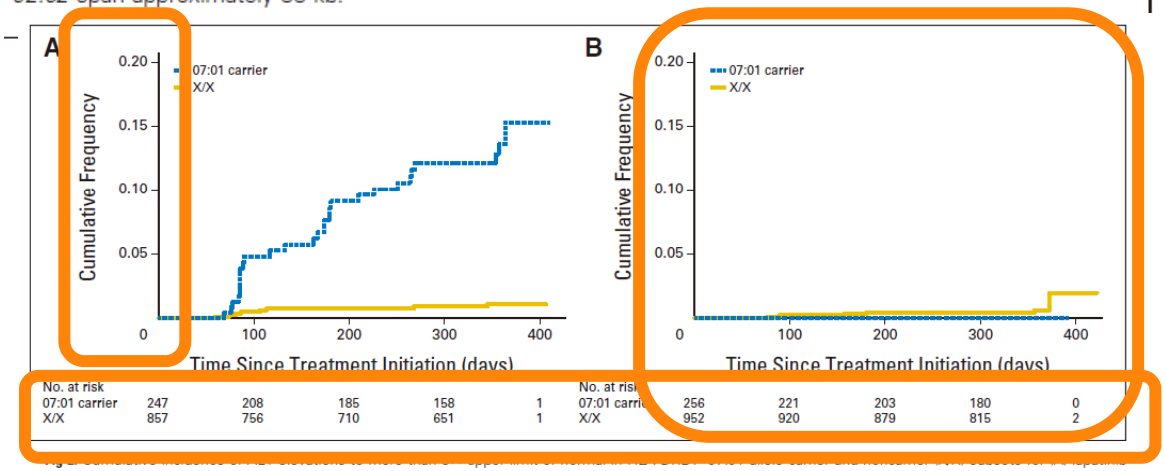
- AE発生割合・リスク比・AE累積発生割合・at risk数の算出は困難
- AE発現例と非発現例を比較可能な形で収集することは困難
- AE発現・AE記録前に死亡するなど、競合リスクの問題があったとしても対処不能
- 非投与例のデータ収集はほぼ不可能

Table 2. Broadly Defined ALT > 3× ULN Cases and Controls and MHC Genotype Associations Among Lapatinib-Treated Patients

	Cases		Controls				
egory	Count	%	Count	%	OR†	95% CI	P‡
	18	9.1	180	90.9	4.68	2.41 to 9.09	1.3×10^{-5}
	19	2.1	889	97.9	1.00 (reference)		
	37		1,069				
omozygous)	29	11.7	218	88.3	14.12	6.36 to 31.32	2.4×10^{-13}
	8	0.9	849	99.1	1.00 (reference)		
	37		1,067				
omozygous)	29	11.7	218	88.3	14.08	6.35 to 31.24	2.6×10^{-13}
	8	0.9	847	99.1	1.00 (reference)		
	37		1,065				
omozygous)	23	11.9	170	88.1	8.64	4.36 to 17.13	7.9×10^{-10}
	14	1.5	894	98.5	1.00 (reference)		
	37		1,064				

nplex; OR, odds ratio; ULN, upper limit of normal.

according to their position on chromosome 6: *TNXB* is approximately 469 kb telomeric from *HLA-DRB1*, and the *02:02 span approximately 88 kb.



treatment compared with (B) placebo treatment. These cumulative frequencies account for censoring as a result of incomplete follow-up, estimated by Kaplan-Meier survival curves subtracted from one. The numbers along the bottom of the *x*-axis are numbers of patients observed over time, carriers above noncarriers.

Schaid et al. JCO, 2014.

臨床試験の附随研究としてのゲノム研究の特徴

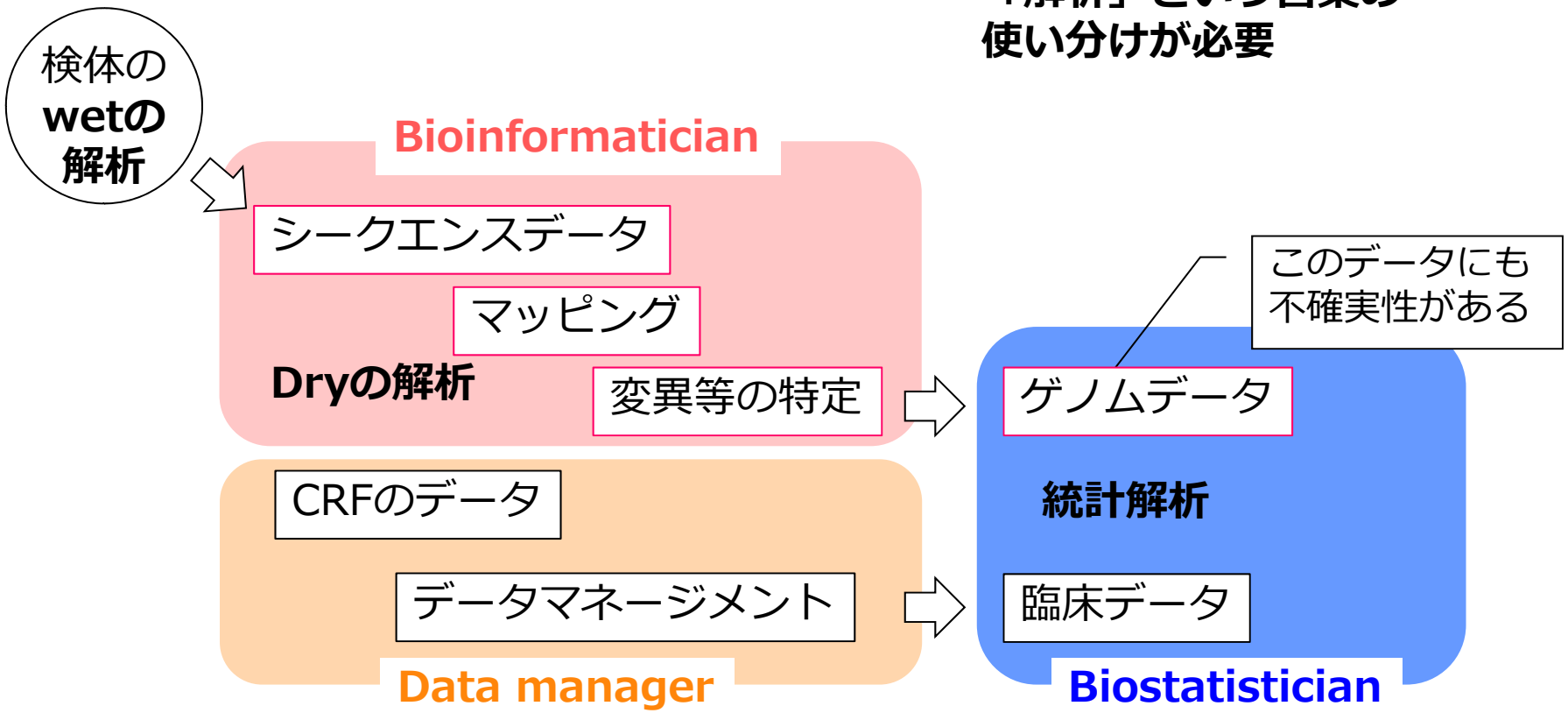
- メリット
  - 適格規準が定められている
    - 診断・対象選択規準がはっきりしている
  - エンドポイントの評価法・評価規準が統一されている
  - 治療法が統一されている
    - 疾患の理解を深める分析に加え、治療選択・治療継続の意思決定に寄与する分析が可能
  - 観察開始の原点が明確にできる
  - 原因-結果の流れに沿った前向きのデータが得られる
    - 検討できる事項や解析方法の選択肢が広がる
  - ランダム化試験の場合、治療間で厳密な比較対照が取れる
    - 検討できる事項が広がる、結果が信頼できる
- デメリット
  - 対象者数が限られる
  - 予後の解析が可能になるまでに時間を要する
  - ゲノム研究のためのサンプルサイズ設計が（通常）困難

Outline

- 臨床試験データによるゲノム解析研究の特徴
- **ゲノム研究におけるデータの取り扱い**
- JCOGでの事例とプロトコール記載の概略
- 臨床情報との関連の統計解析方法の例
- omicsデータのQC-Bioinformaticianの指摘
- プロトコール作成時の留意点

データの流れ

「解析」という言葉の使い分けが必要



- 赤枠部分にもデータの不確実性がある
- 赤枠部分の「変異等の特定」にも、比較対照の設定が重要（非がん部との比較など）

Outline

- 臨床試験データによるゲノム解析研究の特徴
- ゲノム研究におけるデータの取り扱い
- **JCOGでの事例とプロトコール記載の概略**
- 臨床情報との関連の統計解析方法の例
- omicsデータのQC-Bioinformaticianの指摘
- プロトコール作成時の留意点

JCOGでの附随研究事例

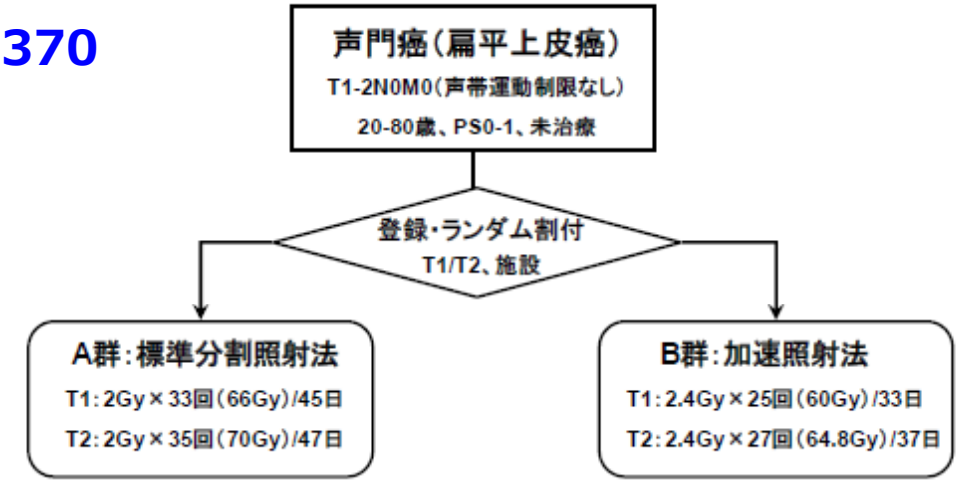
- 放射線治療グループ
  JCOG0701A(2010.5.26)
- 脳腫瘍グループ
  JCOG0911A(2010.10.25)
- 骨軟部腫瘍グループ
  JCOG0905A(2011.8.31)
- 食道がんグループ
  JCOG0502A(2014.12.25)

https://secure.jcog.jp/dc/doc/member/protocol/

JCOG0701A

- **放射線感受性遺伝子候補190遺伝子1,300 SNPsのうち約1,000のSNPs**を解析し、
  声門癌放射線治療による
  放射線治療による急性期および晩期有害事象の発現と程度に関与している可能性が高いSNPsを同定する
  探索的研究

n=370



解析方法(測定方法)の記載とその意図

- **SNPsタイピング方法**
- 放射線感受性遺伝子候補190遺伝子1,300SNPsを対象として、そのうち約1,000のSNPsを中心に放射線医学総合研究所 重粒子医学センターにてヌクレオチドの質量差を検出する
  MassARRAYシステムによりタイピングを行う

- タイピング結果により**以下の条件に合致するSNPマーカーは統計解析から除外**
  1）アレル頻度< 5%
  2）Hardy-Weinberg平衡が成立していない場合 (χ2検定、p < 0.001)

Dryの解析部分（統計解析以前）に行われている「統計学的手法」を用いたデータの処理も可能な限り記す

解析方法(臨床情報との関連の統計解析)の記載とその意図

- 1. 解析対象となるデータの要約、一般化を否定する状況に無いことの確認
- 2. 遺伝子多型間の比較可能性の確認

→ 解析の下準備

- 3. 遺伝子多型と有害事象の関係の分析
- 4. 結果の解釈を補足するための検討

→ 研究の主たる目的を達成するための統計解析方法を明記

- 5. 有意水準および解析対象者数に占める有害事象発現例数の見込み

→ 研究の設定に関する記述

- 6. 結果の解釈

→ “だから何？”に応える記述

5. 有意水準および解析対象者数に占める有害事象発現例数の見込み

- 本附随研究は、**先行研究で有害事象との関連が報告されているSNPsに関する解析**（**タイピングを行うマーカーは実験データ、論文データを基にした候補遺伝子アプローチ**）と、**新たなSNPsを探索する解析**とからなる。検証的ならびに探索的な解析ともに**FDRを5%**として多重性の調整を行う。
- 急性期有害事象のエンドポイントである**放射線治療による急性粘膜炎のうちGrade 2以上の発現頻度**は、過去の報告から約30%と予想され、本附随研究の参加見込み症例を250例とした場合に**75例**となる。
- 晩期有害事象のエンドポイントである**音声機能の変化の**場合、Grade 2の付帯条件によりGrade 2以上の発現頻度は20-30%と予想され、**50−75例**と推定される。

推論の精度に関する記述

検出力の記述に代えて、イベント件数の見込みを記述

※この対象者数・イベント数で、どの位の差がどの位の検出力になるのかを記述する手もある

6. 結果の解釈

- **これまで報告のあったSNPsに関しては本研究により検証的なvalidationを行うという位置づけ**
- これまでに報告がなく、本研究で**新たに放射線治療の有害事象との関連が示唆されたSNPsについては探索的な位置づけ**
  - この研究に引き続き予定している全乳房短期照射の臨床試験の遺伝子多型解析において、**validation**を行う。

- 本研究で検証的な位置づけであるSNPsは以下の10 SNPs

  SOD2:rs4880　　TGFB1:rs1800469

  ・・・（略）・・・

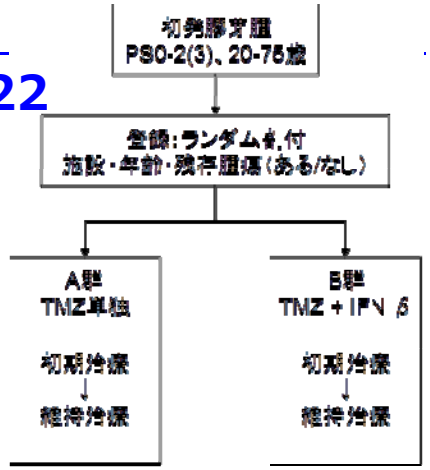
  RAD9A:rs2286620　　LIG3:rs3744355

「本研究は探索的なものである」という決まり文句で意味のない言い逃れをせず、検証と言える範囲とそうでない範囲の線引きをする

「探索的」なものについても、どうケリをつけるつもりなのか（当然未来のことは不確定であるが、その時点での見込み）を記している

JCOG0911A

n=122



- 目的1: 予後予測に最適なMGMT測定方法の検討
- 目的2: IT療法の治療効果とMGMT発現の関連の検討
- 目的3: 膠芽腫の既知・未知バイオマーカー探索
  - 腫瘍細胞の第10染色体長腕(10q)欠失(略) EGFR過剰増幅やPTEN異常等の既知の予後因子を個々に評価
  - mRNA, microRNA発現の網羅的解析、SNP-Microarrayによる全染色体の網羅的解析を行うことで、膠芽腫に対する予後因子、効果予測因子となりうる既知・未知バイオマーカーを探索

どの解析を誰が担当するかを事前に切り分ける
（プロジェクトが動くよう、具体的に段取りをつける）

- **統計解析施設**
  - 目的1、目的2の統計解析：JCOGデータセンター
  - 目的3の統計解析：試料解析研究事務局
  - 目的3の統計解析のうち、試料解析結果と診療データを統合したデータパッケージを対象として行われる解析部分のダブルチェック： JCOGデータセンター

JCOG0502A

- **調べること**
  ①化学放射線療法群における、
  がん組織の遺伝子ステータスと**CR割合**の関連
  ②化学放射線療法群における、
  生殖細胞系列遺伝子変異と**有害事象**の関連
  ③手術群における、**がん組織の遺伝子ステータス**と
  全生存期間、無増悪生存期間の関連
  ④化学放射線療法群における、**がん組織の遺伝子ステータス**と
  全生存期間、無増悪生存期間の関連
- **解析の順序**
  - まず本附随研究計画段階で判明しているデータで①を調べる
  - ②は、データが固定し、有害事象のデータの公表が終わった段階で解析を行う
  - ③と④は、主たる解析が終了し、主たる結果の公表が終わった段階で解析を行う

統計解析方法に関して
プロトコールに記載している項目

- 目的
- エンドポイント
- データの形式
  - 二値、順序カテゴリー、連続値、生存期間、etc.
  - カットオフの定義 and/or 決め方
- 対象となる患者・検体の選択方法
- 解析方法
  - 対象集団、解析方法、解析の条件、etc.
- 感度解析
- 解析結果の解釈

Outline

- 臨床試験データによるゲノム解析研究の特徴
- ゲノム研究におけるデータの取り扱い
- JCOGでの事例とプロトコール記載の概略
- **臨床情報との関連の統計解析方法の例**
- omicsデータのQC-Bioinformaticianの指摘
- プロトコール作成時の留意点

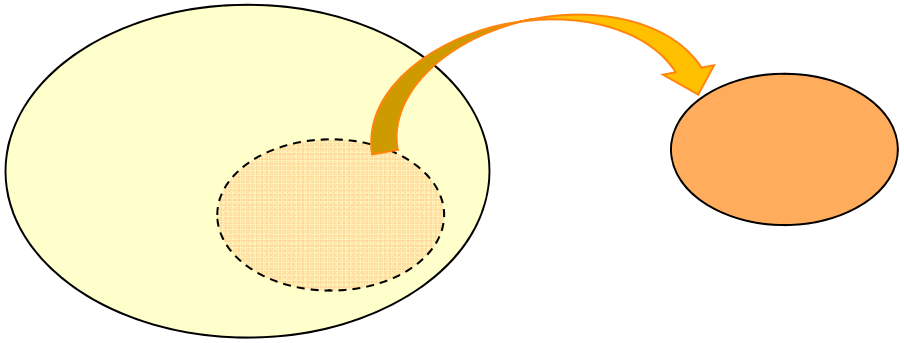
JCOG0701A
解析方法(臨床情報との関連の統計解析)

- **1. 解析対象となるデータの要約、一般化を否定する状況に無いことの確認**
- 「全登録例と本附随研究の対象となる全症例」、「本附随研究の対象とならなかった症例と本附随研究の対象となる全症例」との間で、背景因子、予後、有害事象発現状況の比較

- **2. 遺伝子多型間の比較可能性の確認**
- **本附随研究は臨床試験対象症例を扱うので治療法の統一、有害事象の判定ポイントは厳密に設定されている**が、まず、年齢、性別、施設、治療法など診療情報と急性期有害事象（Grade 0-1 vs. Grade 2以上）、晩期有害事象（Grade 0-1 vs. Grade 2以上）との関連をそれぞれ解析し、**有害事象発症に遺伝的要因以外の因子が関わっている可能性について検討**

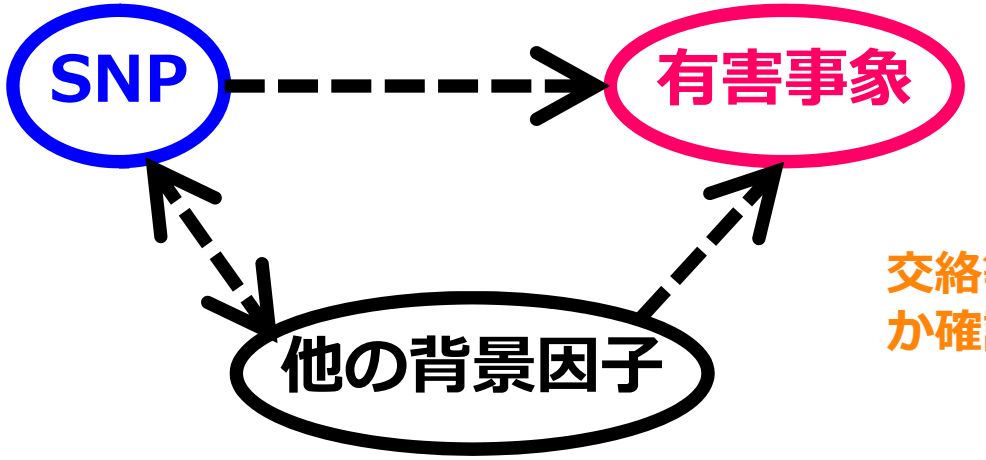
解析方法(臨床情報との関連の統計解析)

- 1. 解析対象となるデータの要約、一般化を否定する状況に無いことの確認



偏っていないか確認

- 2. 遺伝子多型間の比較可能性の確認



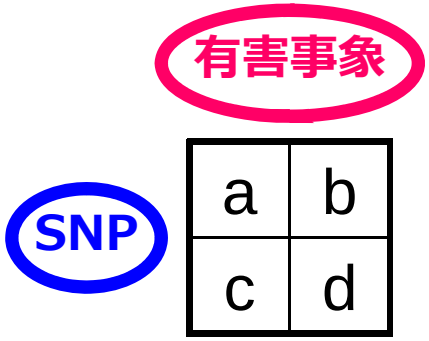
交絡等が起きていないか確認

解析方法(臨床情報との関連の統計解析)

- **3. 遺伝子多型と有害事象の関係の分析**
- 特定の診療情報と有害事象発症との間に関連が認められない場合は、遺伝子多型で構成した2群間で、急性期有害事象あるいは晩期有害事象のそれぞれについて、Grade0-1とGrade2以上の有害事象発症に差があるか否か、**Fisher正確検定**により関連解析を行う。この場合allele, genotype (優性モデル、劣性モデル)それぞれについてどのモデルが最適であるか検討する。
- また、検出力を上げるためSNPsの組合せ、特に同一遺伝子座でのSNPsの組合せによるハプロタイプ解析を行い、**リスクハプロタイプを予測**する。
- また、診療情報を説明変数に加えた**対数線型モデル**により、感度分析及び診療情報を調整した効果の推定を行う。

解析方法(臨床情報との関連の統計解析)

- 3. 遺伝子多型と有害事象の関係の分析



ここでは簡単のため、
AA, Aa, aaの
3カテゴリー
ではなく、
優性モデル or
劣性モデルの
2カテゴリーでの
解析を例示

Fisher's exact test

ロジスティック回帰：
オッズ比 a/b ÷ c/dをモデル化
→リスク比が計算出来ない場合や、
有害事象発現割合が小さい場合に使う

対数線形モデル：
リスク比 a/(a+b)÷c/(c+d) をモデル化
→本研究は前向きにデータを集めているの
でリスク比を算出可能
SNPの影響を調べたいのでオッズ比より
リスク比を出すほうが自然

ケース・コントロール研究（有害事象あり・なしの情報に基づき対象を選択するデザイン）ならばロジスティック回帰＆オッズ比を用いる
本研究では臨床試験のデータで前向きに（有害事象発現前の情報で）集団が特定出来ているのでロジスティック回帰＆オッズ比ではなく、対数線型モデル＆リスク比を用いる

解析方法(臨床情報との関連の統計解析)

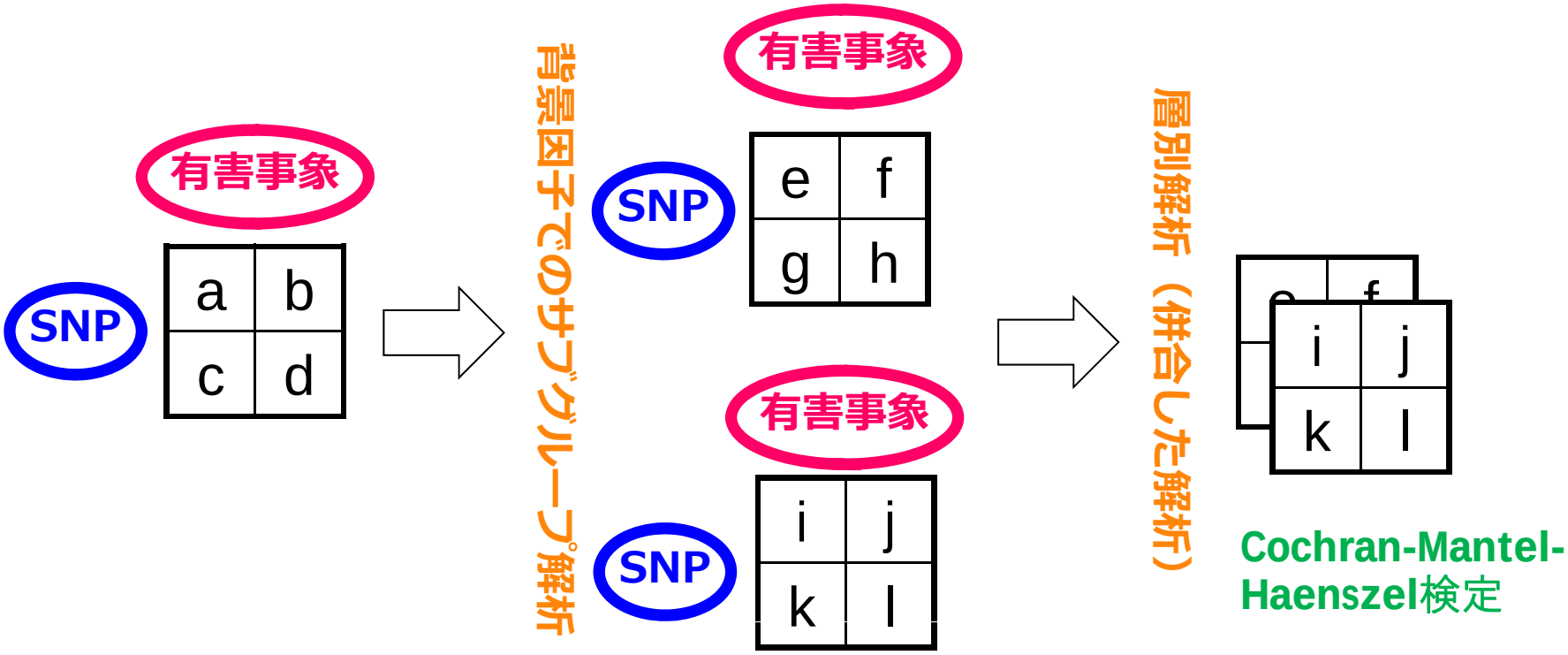
- **3. 遺伝子多型と有害事象の関係の分析　（つづき）**
- 特定の診療情報と有害事象発症との間に関連が認められる場合、当該因子による**サブグループ別の解析**を行い各々で効果を推定
- 交互作用が見られない場合、これらの因子を層とした**Cochran-Mantel-Haenszel流のχ2検定**並びに**層を併合した効果を推定**

SNPs以外の要因（診療情報）によって生じる交絡の影響を排除するため、サブグループ解析と層別解析を実施

- 有害事象発症との間に関連がある因子を説明変数に加えた**対数線型モデル**により、診療情報を調整した効果を推定
- 有害事象発症との間に関連が認められなかった因子も説明変数に加えた**対数線型モデル**で感度分析

解析方法(臨床情報との関連の統計解析)

- 3. 遺伝子多型と有害事象の関係の分析　（つづき）



SNPs以外の要因（診療情報）によって生じる交絡の影響を排除するため、サブグループ解析と層別解析を実施

対比 contrast

- 用量反応性を評価する際や、その応用としてカットオフ値を探索する際に用いる解析方法
- 例：検査の結果が3カテゴリーで(-, ±, +)と表現される場合に、outcomeの用量反応パターン

それぞれのパターンを、以下のような数字の組み合わせで表現する（これを対比contrastと呼ぶ）

(-1,0,1)　(-2,1,1)　(-1, -1, 2)

これらから統計的に当てはまりの良いものを選択

FDR: false discovery rate

- 多数の仮説に対して検定を行った場合に、検定で有意になった結果の確からしさを示す指標
- 検定で有意になり棄却された仮説の中には以下の2つのケースが含まれる
  - 帰無仮説（差なし）・・・含まれて欲しくない
  - 対立仮説（差あり）
- 帰無仮説がたくさん含まれているような論文はあてにならない！
- 「有意になった仮説が○個ありました」という主張の確からしさを示す情報を添えたい→FDR等を使う
- 「検定で有意になった仮説の中に“帰無仮説”(＝本来、棄却すべきでない仮説)が含まれる割合の期待値」

αエラー 5%とFDR 5%の違い

αエラー5%の意味

	帰無仮説（差なし）	対立仮説（差あり）
有意差あり	a	b
有意差なし	c	d
	a+c	

a/(a+c)=5%

ある帰無仮説を**多数回［（a+c)回］の試験**で検証した際、間違って有意と結論づける試験が生じる確率が5%

※ a, b, c, dは「繰り返す試験の数」

FDR 5%の意味

	帰無仮説（差なし）	対立仮説（差あり）	
有意差あり	a	b	a+b
有意差なし	c	d	

a/(a+b)=5%

多数個［（a+b)個］の有意になった仮説の中に、間違って帰無仮説が含まれている確率が5%

※ a, b, c, dは「調べる仮説の数」

クラスタリング

- クラスタリング：データを数理的な規準によりいくつかの部分集合に分ける方法
- クラスタリング手法の分類の切り口
  - 部分集合に分ける際に真の分け方に関する外的情報*を利用するか否か（「奏効例、非奏効例」などの情報）
    - Supervised（教師あり）
    - Non-supervised（教師なし）
  - 集団を分ける際に階層構造を入れるか否か
  - 部分集合間の遠い・近いを判断するための基準として何を用いるか、どのようなアルゴリズムを用いるか
  - 患者を分けるか、変数を分けるか

より詳しく学びたい方はICRwebの以下の講義を視聴して下さい
9. 相関構造の分析ー主成分分析とクラスター分析《講師：田中紀子先生》
http://www.icrweb.jp/course/view.php?id=258

Outline

- 臨床試験データによるゲノム解析研究の特徴
- ゲノム研究におけるデータの取り扱い
- JCOGでの事例とプロトコール記載の概略
- 臨床情報との関連の統計解析方法の例
- **omicsデータのQC-Bioinformaticianの指摘**
- プロトコール作成時の留意点

データは確かか？

①赤い△は誤入力
（本当は男性→女性と入力）
②青い○は誤入力
（本当は女性→男性と入力）
③青い○はKlinefelter症候群
が疑われる

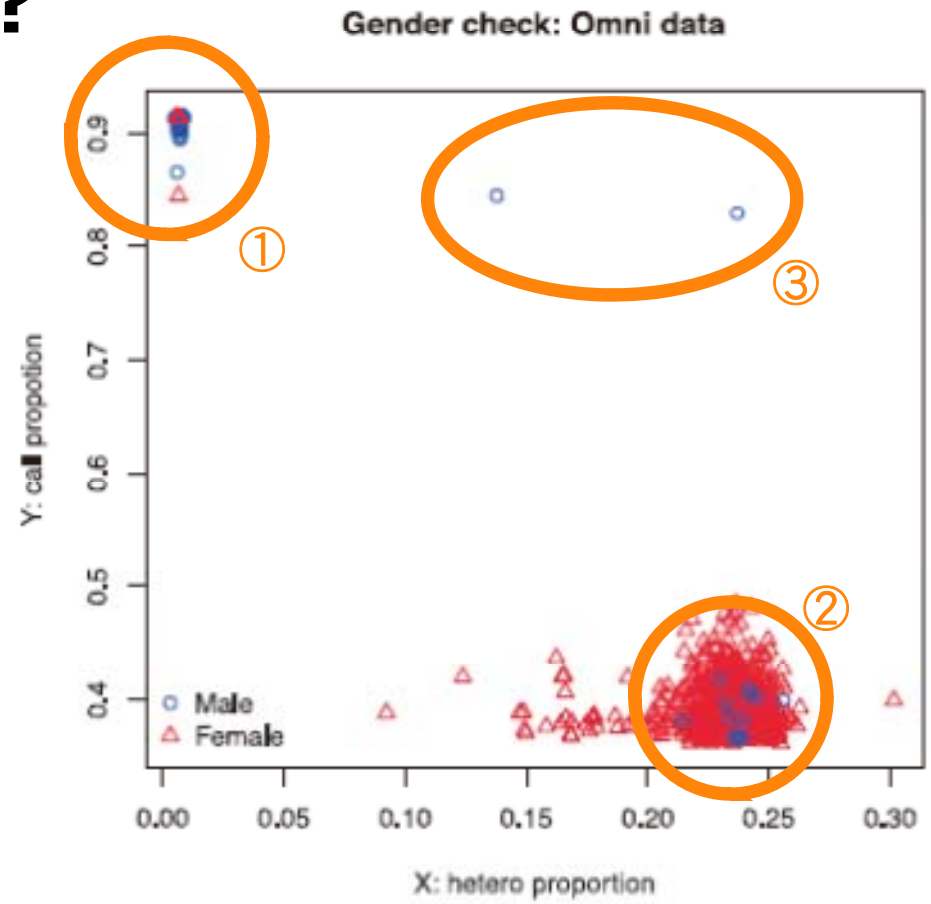


図 6 横軸に X 染色体のヘテロ割合，縦軸に Y 染色体の SNP call 割合とした散布図. 青丸:臨床情報が男性，赤三角:臨床情報が女性.

知久季倫 et al.,核酸omicsデータのquality control解析,
みずほ情報総研 技報 6(1):14-29.

データの品質管理が重要

- Bioinformaticianの指摘①

“臨床情報はMicrosoft excel形式で届くことが多く、これまでの経験から情報解析者として

- 1. 全角英数字、全角スペース
- 2. 数字やフラグ箇所への自然言語での書き込み
- 3. 色で区別
- 4. 表記の揺れ
- 5. 臨床情報は時系列データであり、論文までに最低3回は全計算をやり直す

らを意識して注意している”

知久季倫 et al.,核酸omicsデータのquality control解析，みずほ情報総研 技報 6(1): 14-29.

JCOG研究の場合には、データマネージャーがこのような問題が生じないよう、対応している

データの品質管理が重要（つづき）

- Bioinformaticianの指摘②

“臨床情報と臨床検体の突き合わせも重要であり、

1. 経験的に200人に1人くらいは性別が一致しない
2. 同じサンプルが異なるIDで複数回提供される
3. 地域的に異なる施設間でも血縁者が含まれる
4. 健常人として提供された血液サンプルが白血病だった
5. 日本人ではなかった
6. がん部、非がん部のペアが別人
7. がん部と非がん部の標識ラベルの付け間違え
8. 他者のDNAの混入

らをQCとして確認している。上記の制度は検体提供施設によって大きな差がある。”

知久季倫 et al.,核酸omicsデータのquality control解析，みずほ情報総研 技報 6(1): 14-29.

6～8は深刻な問題であり、適切な検体収集手順の規定が必要
繰り返し検体を測定する場合などは時期のラベル付与方法も問題になる

Outline

- 臨床試験データによるゲノム解析研究の特徴
- ゲノム研究におけるデータの取り扱い
- JCOGでの事例とプロトコール記載の概略
- 臨床情報との関連の統計解析方法の例
- omicsデータのQC-Bioinformaticianの指摘
- **プロトコール作成時の留意点**

プロトコール作成時の留意点

- 「解析」の書き分けが必要
  - Wetの解析、dryの解析、臨床情報との関連の統計解析
- 「ゲノム解析は探索的なものであり、統計解析方法を事前に記載することは出来ない」は不適切
  - 事前に想定していた解析以外の解析を行うのは事実であり、bioinformatics部分の解析方法の変化は早いため事前の想定通りにはいかないことも事実である
  - しかし、研究開始時点で少なくともひとつの一連の解析の流れは書き下せる
  - 「探索的」はいい加減な研究を正当化するための言葉ではない
- 誰が何をするか、いつするか、を明らかにする
- 検体の収集方法を（取り違え等を避けるよう）具体的に定める
- Bioinformatics、データ管理、生物統計学の専門知識が必要なところは、専門家とコラボする