

平成 22 年度厚生労働省委託事業

平成 22 年度次世代医療機器評価指標作成事業

テーラーメイド医療用診断機器

(DNA チップ等を用いる遺伝子発現解析装置)

審査 WG 報告書

平成 23 年 3 月

座長 神田 忠仁

理化学研究所 新興・再興感染症研究ネットワーク推進センター
チームリーダー

目次

1.	RNA プロファイルにもとづく診断装置の評価指標策定について	2
2.	テーラーメイド医療診断機器（DNA チップ等を用いる遺伝子発現解析装置）委員等名簿	4
3.	平成 22 年度会議議事概要	5
4.	各タスクフォースからの提言	
	I. DNA チップ等を用いた RNA プロファイル解析装置の評価指標案	14
	II. DNA チップ等の遺伝子発現解析装置の製造販売承認申請における臨床性能評価に関する考慮すべき事項	19
	III. 診断装置と診断方法を普及させる仕組みに関する事項（提言）	27
5.	参考資料	31
	（関連論文翻訳）	
1.	予後バイオマーカーおよび予測バイオマーカーの評価に関する保存標本の使用（J Natl Cancer Inst. 2009 101:1446-52.）	31
2.	妥当性確認試験の要点（Nat Rev Clin Oncol. 2010 7:327-34.）	42
3.	バイオマーカーと代替エンドポイント—統計学的妥当性確認の課題（Nat Rev Clin Oncol. 2010 7:309-17.）	48
	（審査 WG 発表資料）	
1.	国内外の規制と米国での認可状況（評価基準の作成に向けて）	59
		鈴木孝昌
2.	Gene Signature の検討 臨床の立場から	戸井雅和 74
3.	遺伝子検査標準化の現状	登 勉 82

RNA プロファイルにもとづく診断装置の評価指標策定について

理化学研究所・感染症研究ネットワーク推進センター
神田忠仁

試料に含まれる RNA (mRNA 及び microRNA) を高感度で検出・定量する技術や、数万種の RNA のレベルを同時に測定・比較する技術が開発され、信頼性も向上している。この技術を使って、疾患と関連する遺伝子群の探索や、予後や治療薬への感受性・副作用と関連する遺伝子群の探索とそれらの遺伝子の転写レベルの検討がすすみ、蓄積されたデータを疾病の検出、病態の分類、患者に適した治療薬の選定などへ応用する研究が急速に進展している。例えば、腫瘍における特定の遺伝子群由来 RNA レベルの相対量比 (プロファイル) を調べ、化学療法に対する感受性や予後を推定する方法が提案されている。一部の腫瘍では、プロファイル解析の臨床有用性が示され、診断補助装置として市場導入することが期待されている。このような状況を踏まえ、我々は、RNA のプロファイルをもとに疾病の検出や患者の予後、治療薬への感受性、副作用の程度などを予測する診断装置の信頼性、有用性を保証するための指標について議論を始めた。

患者検体から抽出した RNA 試料に含まれる特定の RNA 群の相対レベルを測定し、そのプロファイルを複雑なアルゴリズムで解析して得た医療情報の有用性・信頼性を評価するには、どのような情報が必要か？我々は、1) 装置が複数の RNA 分子の相対レベルを確実に測定し、再現性良くプロファイルが得られることと、2) RNA プロファイルをもとに装置が導き出す判定結果の臨床的意義を分けて整理することとした。さらに、既存の評価系では、臨床的意義の評価が難しい場合が想定されると考え、3) 新たな枠組みの構築を含め、臨床性能の評価系についても議論をすすめた。

検体中の特定の RNA 分子のレベルを測定するには、RNA ないし逆転写反応で得た cDNA を適切なプローブと結合させて生じた複合体を定量する方法か、リアルタイム PCR による方法が使われる。前者による測定では、装置の信頼性、安定性などは平成 20 年 4 月に公開された「DNA チップを用いた遺伝子型判定用診断薬に関する評価指標」と重複する部分がある。後者は多種の RNA 分子の定量には向かないものの、測定技術は確立している。DNA に比べて不安定な RNA の取り扱い、適切な標準物質の設定、ダイナミックレンジ内でのカットオフ値の設定などに留意しながら、評価指標案の策定を進めた。

臨床性能を示すには、疾患や予後の診断、治療薬の有効性や副作用などの診断基準を統一させた臨床データと RNA プロファイルとの関連を示す必要がある。症例が少ない疾患の場合は、海外の臨床データや過去に保存された臨床検体の利用も想定しながら、収集可能な臨床データの質と量、装置による予測と臨床データが一致する程度などを議論し、さらに装置の有用性や発展性を議論した。

臨床検体における特定の RNA プロファイルから導いた臨床情報の有用性の評価は、医薬品医療機器総合機構によるこれまでの診断薬の評価とは異なる部分がある。迅速に信頼できる評価を行うには、どのような枠組みが必要か、海外の例を参考にしながら考察した。

以上のように議論をすすめ、論点を整理して研究報告書にまとめた。RNA のプロファイルを診断や治療戦略の策定に応用する技術は、個々の患者に最適な治療を選択するいわゆるテーラーメイド医療の基盤であり、一方では先端医療産業の側面もある。技術の進歩を的確に医療現場に還元し、また医療情報を技術の発展に役立てる仕組み

としても機能する評価指標を策定することが求められており、この報告書がその一歩となると考えている。

テーラーメイド医療用診断機器（DNAチップ等を用いる遺伝子発現解析装置）

審査WG委員等名簿（敬称略）

座長

神田 忠仁 理化学研究所 新興・再興感染症研究ネットワーク推進センター
チームリーダー

副座長

戸井 雅和 京都大学医学部附属病院 乳腺外科 教授

委員(五十音順)

佐々木 博己 国立がんセンター研究所 多層オミックス・バイオインフォーマ
ティクス分野 ユニット長
高橋 隆 名古屋大学大学院医学系研究科 神経疾患・腫瘍分子医学研究セン
ター 腫瘍病態統御部門・分子腫瘍学分野 教授
寺前 紀夫 東北大学理学部化学専攻分野 分析化学研究室 教授
登 勉 三重大学臨床検査医学分野 教授
古川 洋一 東京大学医科学研究所 先端医療研究センター 臨床ゲノム腫瘍学
分野 教授

アドバイザー

油谷 浩幸 東京大学先端科学技術研究センター ゲノムサイエンス分野 教
授

会計

柏 喜代美 JBCRG (Japan Breast Cancer Research Group) 事務局

厚生労働省

関野 秀人 医薬食品局 審査管理課 医療機器審査管理室長
東 健太郎 医薬食品局 審査管理課 医療機器審査管理室 新医療材料専門
官
吉成 知也 医薬食品局 審査管理課 医療機器審査管理室 技官

独立行政法人 医薬品医療機器総合機構 (PMDA)

田村 敦史 医療機器審査第二部長
宮本 大誠 医療機器審査第二部 審査役代理
磯部 総一郎 審査マネジメント部長
水上 良明 審査マネジメント部 主任専門員

国立医薬品食品衛生研究所(事務局)

松岡 厚子 医療機器部 部長
宮島 敦子 医療機器部 室長
鈴木 孝昌 遺伝子細胞医薬部 室長

平成 22 年度 次世代医療機器評価指標作成事業
テーラーメイド医療用診断機器 (DNA チップ等を用いる遺伝子発現解析装置)
審査 WG (第 1 回) 議事概要

1. 開催日時 平成 22 年 11 月 2 日(火) 17 時～19 時
2. 開催場所 東京八重洲ホール 302 会議室
3. 出席者 (敬称略・座長以下五十音順)
審査 WG 委員：神田 忠仁、戸井 雅和、佐々木 博己、高橋 隆、寺前 紀夫、
登 勉、古川 洋一
アドバイザー (開発 WG 委員)：油谷 浩幸
厚生労働省：関野 秀人、吉成 知也
医薬品医療機器総合機構：宮本 大誠、水上 良明
審査 WG 事務局：松岡 厚子、鈴木 孝昌、宮島 敦子
オブザーバー：中江 裕樹 (バイオチップコンソーシアム)、木山 亮一 (産業
技術総合研究所)
4. 配布資料
資料 1-1 平成 18 年度次世代医療機器評価指標策定事業 次世代医療機器評価指標
の公表について (DNA チップ)
資料 1-2 In Vitro Diagnostic Multivariate Index Assays (IVDMIA)
資料 1-3 Class II Special Controls Guidance Document: Instrumentation for
Clinical Multiplex Test Systems (Instrumentation)
資料 1-4 Class II Special Controls Guidance Document: Gene Expression
Profiling Test System for Breast Cancer Prognosis (MammaPrint)
資料 1-5 Class II Special Controls Guidance Document: Cardiac Allograft
Gene Expression Profiling Test Systems (Allomap)
資料 1-6 510K 承認時の decision summary (MammaPrint)
資料 1-7 510K 承認時の decision summary (Pathwork)
資料 1-8 510K 承認時の decision summary (Allomap)
資料 1-9 Oncotype DX 関連資料 (IVDMIA のガイダンスに対する、製造業者の
Genomic Health 社のコメント)
資料 2-1 利益相反関連の書類について
資料 2-2 <話題提供資料> 中江 裕樹 氏
資料 2-3 <話題提供資料> 鈴木 孝昌 氏
資料 2-4 <話題提供資料> 戸井 雅和 委員
資料 2-5 平成 21 年度 開発ガイドラインワーキンググループ委員会の検討結果

参考資料 (委員のみ)
「平成 18 年度次世代医療機器評価指標策定事業 テーラーメイド医療用診断機器
(DNA チップ) 審査 WG 報告書」
5. 議事概要
(1) 開会の挨拶 (事務局)

- (2) 次世代事業の説明（厚生労働省）
- (3) 座長挨拶、委員紹介
- (4) 話題提供
 - ・中江 裕樹 氏（バイオチップコンソーシアム）
「バイオチップの開発動向」
 - ・鈴木 孝昌 氏（国立医薬品食品衛生研究所）
「国内外の規制と米国での認可状況（評価基準の作成に向けて）」
 - ・戸井 雅和 委員（京都大学医学部附属病院）
「Oncotype DXに関する検討 臨床の立場から」
- (5) 討議

中江氏、鈴木氏、戸井委員の情報提供に対して質疑応答がなされた後、発表にもとづいて、討議を進めた。

 - ・米国 FDA に認可された MammaPrint, Pathwork, AlloMap と、認可は受けていないが米国において保険適用になっている Oncotype DX について、討議が行われた。
 - ・米国の CLIA 認証制度と LDT、および保険適応の問題、日本国内での薬事承認をめぐる状況の違いに関して、質疑がなされた。
 - ・作成しようとするガイドラインの想定される適応範囲について議論された。
 - ・日本における現状について理解を深めるため、第 2 回の審査 WG において、登委員、佐々木委員、宮本氏(PMDA)より各分野に関する情報提供をお願いすることにした。
- (6) 次回の予定（第 2 回）

平成 22 年 12 月 2 日(木) 10 時～12 時 東京八重洲ホール 411 会議室

平成 22 年度 次世代医療機器評価指標作成事業
テーラーメイド医療用診断機器（DNA チップ等を用いる遺伝子発現解析装置）
審査 WG（第 2 回） 議事概要

1. 開催日時 平成 22 年 12 月 2 日(木) 10 時～12 時
2. 開催場所 東京八重洲ホール 411 会議室
3. 出席者（敬称略・座長以下五十音順）
審査 WG 委員：神田 忠仁、戸井 雅和、佐々木 博己、高橋 隆、寺前 紀夫、
登 勉、古川 洋一
厚生労働省：関野 秀人、吉成 知也
医薬品医療機器総合機構：宮本 大誠、水上 良明
審査 WG 事務局：鈴木 孝昌、宮島 敦子
オブザーバー：木山 亮一、安佛 尚志、橋本 亮一（産業技術総合研究所）

4. 配布資料

- 資料 1 平成 22 年度 次世代医療機器評価指標作成事業 テーラーメイド医療用診断
機器（DNA チップ等を用いる遺伝子発現解析装置） 審査 WG（第 1 回）議事概
要
- 資料 2 平成 18 年度次世代医療機器評価指標策定事業 次世代医療機器評価指標の公
表について（DNA チップ）（再配布）
- 資料 3 「平成 18 年度次世代医療機器評価指標策定事業 テーラーメイド医療用診断
機器（DNA チップ）審査 WG 報告書」（抜粋 P40-46）
- 資料 4 医療機器開発ガイドライン策定事業
経済産業省委託事業 実施機関：独立行政法人産業技術総合研究所
- 資料 5 <話題提供資料> 登 勉 委員
- 資料 6 <話題提供資料> 佐々木 博己 委員
- 資料 7 <話題提供資料> 宮本 大誠 氏

- 参考資料 1 テーラーメイド医療用診断機器（DNA チップ）開発ガイドライン 2007
-遺伝子型（ジェノタイピング）検定用 DNA チップに関して-
- 参考資料 2 テーラーメイド医療用診断機器（DNA チップ）開発ガイドライン-遺伝子
発現解析用 DNA チップに関して-（平成 22 年度第 2 回開発ガイドライン
WG 委員会検討内容）

5. 議事概要

- (1) 開会の挨拶（事務局）
- (2) 座長挨拶
- (3) 開発 WG の状況について
木山 亮一 氏（産業技術総合研究所）
「医療機器開発ガイドライン策定事業」
- (4) 話題提供
 - ・登 勉 委員（三重大学大学院医学系研究科）
「遺伝子検査標準化の現状」
 - ・佐々木 博己 委員（国立がんセンター研究所）

「日本の企業による診断用DNAチップ開発の現状から見た評価項目」

- ・宮本 大誠 氏 (医薬品医療機器総合機構)

「体外診断用医薬品を審査する立場から」

(5) 討議

- ・木山氏、登委員、佐々木委員、宮本氏の情報提供に対して質疑応答がなされた後、発表にもとづいて、ガイドライン作成にあたり論点とすべき内容について討議を進めた。その結果、以下の内容に関して、意見交換がなされた。
 - 臨床性能試験の重要性
 - 臨床性能試験としての要求事項
 - 臨床性能試験として認められる内容（前向き試験と後ろ向き試験）
 - 前向き臨床試験としての予後予測データ取得の難しさ
 - 体外診断薬としての承認と保険適応の関係（home-brew テストとしての利用）
 - 条件付（仮）承認の可能性など、新しい診断法、装置を普及させる仕組みの必要性
 - 開発側ガイドラインと審査ガイドラインの相互関係
 - 先進医療の利用と開発支援
 - 使用目的、効能効果の限定
-
- ・第1回および第2回の審査WGにおいて、海外および日本における状況についての情報がそろい、現状認識と重要課題に関する理解が深まってきたことから、2回のWGの結果をふまえて、座長と事務局で論点の整理を行い、第3回審査WGにおいてガイドライン作成に向けた討議を進めることにした。

(6) 次回の予定（第3回）

平成23年1月13日(木) 10時～12時 東京八重洲ホール 800会議室

平成 22 年度 次世代医療機器評価指標作成事業
テーラーメイド医療用診断機器 (DNA チップ等を用いる遺伝子発現解析装置)
審査 WG (第 3 回) 議事概要

1. 開催日時 平成 23 年 1 月 13 日(木) 10 時～12 時
2. 開催場所 東京八重洲ホール 800 会議室
3. 出席者 (敬称略・座長以下五十音順)
審査 WG 委員：神田 忠仁、戸井 雅和、佐々木 博己、高橋 隆、登 勉、古川 洋一、
アドバイザー (開発 WG 委員)：油谷 浩幸
厚生労働省：関野 秀人、東 健太郎
医薬品医療機器総合機構：宮本 大誠、水上 良明
審査 WG 事務局：松岡 厚子、鈴木 孝昌、宮島 敦子
オブザーバー：木山 亮一、本間 一弘 (産業技術総合研究所)

4. 配布資料

- 資料 1 平成 22 年度 次世代医療機器評価指標作成事業 テーラーメイド医療用診断機器 (DNA チップ等を用いる遺伝子発現解析装置) 審査 WG (第 2 回) 議事概要
- 資料 2 平成 21 年度 開発ガイドラインワーキンググループ委員会の検討結果 (再配布)
- 資料 3 テーラーメイド医療用診断機器 (DNA チップ) 開発ガイドライン-遺伝子発現解析用 DNA チップに関して- (平成 22 年度第 2 回開発ガイドライン WG 委員会検討内容) (再配布)
- 資料 4 「平成 18 年度次世代医療機器評価指標策定事業 テーラーメイド医療用診断機器 (DNA チップ) 審査 WG 報告書」 (抜粋 P40-46) (再配布)
- 資料 5 <佐々木委員 情報提供> DNA チップ解析・診断の工程/ELISA や CLIA を基盤にした対外診断薬の例
- 資料 6 <寺前委員 コメント>

<戸井委員 情報提供>

- 参考論文-1 What should physicians look for in evaluating prognostic gene-expression signatures? Subramanian J, Simon R. Nat Rev Clin Oncol. 2010 7:327-34.
- 参考論文-2 Gene-expression-based prognostic assays for breast cancer. Kim C, Paik S. Nat Rev Clin Oncol. 2010 7:340-7.
- 参考論文-3 Biomarkers and surrogate end points--the challenge of statistical validation. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Nat Rev Clin Oncol. 2010 7:309-17.
- 参考論文-4 American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S,

Somerfield MR, Hayes DF, Bast RC Jr; American Society of Clinical Oncology. J Clin Oncol. 2007 25:5287-312.

参考論文-5 Use of archived specimens in evaluation of prognostic and predictive biomarkers. Simon RM, Paik S, Hayes DF. J Natl Cancer Inst. 2009 101:1446-52.

- 参考資料-1 審査 WG (第 1 回) <話題提供資料> 鈴木氏
- 参考資料-2 審査 WG (第 1 回) <話題提供資料> 戸井委員
- 参考資料-3 審査 WG (第 2 回) <話題提供資料> 登委員
- 参考資料-4 審査 WG (第 2 回) <話題提供資料> 佐々木委員
- 参考資料-5 審査 WG (第 2 回) <話題提供資料> 宮本氏

5. 議事概要

- (1) 座長挨拶
- (2) 討議

「論点整理」(別紙 1) について確認し、「会議の進め方」(別紙 2) に従って、評価指標作成に向けて、以下の内容に関して、意見交換がなされた。

- I. 遺伝子発現測定装置としての物理化学的承認基準に関する事項
 - 開発ガイドラインと審査ガイドライン(評価指標)の整合性
 - アルゴリズム
 - 標準品と信頼性の確保
- II. 臨床性能の評価に関する事項
 - レトロスペクティブ/プロスペクティブ試験
 - 装置によって得られるデータの再現性と臨床的有用性の再現性
 - Oncotype DX を例として、承認申請における臨床性能の評価について
- III. 診断装置と診断方法を普及させる仕組みに関する事項
 - 保険収載と医療経済学的効果
 - レトロスペクティブ/プロスペクティブデータの扱い
 - 仮承認、条件付きシステム
 - 先進医療の利用と開発支援

次世代審査 WG (第 3 回) における討議の結果、報告書を、「会議の進め方」の分類に従って、下記の方向でまとめることにした。

I. 遺伝子発現測定装置としての物理化学的承認基準に関する事項
事務局、寺前委員、神田委員が中心となって、評価指標のたたき台となる原案を作成する。

II. 臨床性能の評価に関する事項
戸井委員、佐々木委員、古川委員が中心となって、当該機器から得られる情報の使用目的、目的が達せられることを示すのに必要な臨床データの量と質など、現状についてまとめる。

III. 診断装置と診断方法を普及させる仕組みに関する事項

登委員、高橋委員、宮本氏が中心となって、当該機器の評価と医療現場へ導入・普及させるための枠組みについて提言をする。

(3) 今後の予定

各項目について、分担委員が中心となって、たたき台を2月中旬を目処に作成する。

たたき台をメールで関係者全員に送付し、意見交換を行う。

最終のまとめを、3月18日（金）までに作成する。

(別紙 1)

平成 22 年度 次世代医療機器評価指標作成事業 テーラーメイド医療用診断機器
(DNA チップ等を用いる遺伝子発現解析装置) 審査 WG (第 3 回)

論点整理

I. 臨床性能の評価に関する事項

1. 臨床性能試験の要求事項

装置としての精度保証、一体型の専用機器と汎用機器の区別、
標準品の選定、アルゴリズムの評価など

臨床的意義を示すために必要なデータ

症例数

統計手法の妥当性

探索的に行った研究結果の利用

レトロスペクティブ/プロスペクティブな試験

先進医療の枠組みでの臨床性能試験の利用

ホームブリューテスト、LDT の位置付け、GLP 適合の利用等

海外データの利用

2. 臨床性能を第三者機関が評価する可能性

3. 国際整合性

II. 承認申請に関する事項 (診断装置と診断方法を普及させる仕組み)

1. 承認申請を受けるメリット

保険収載

仮承認的な承認の利用

保険収載に頼らない方法

医療経済学的考察

2. ガイドラインに関する事項

実用的なガイドラインの扱う範囲

開発ガイドラインとの整合性

(別紙 2)

会議の進め方

- I. 遺伝子発現測定装置としての物理化学的承認基準に関する事項
 - DNA チップと測定装置を一体化して扱う場合と汎用測定装置を使う場合
 - 検体、RNA 等に関する事項
 - 検体の前処理、保存
 - 測定系に関する事項
 - 測定原理、アルゴリズム
 - 性能の信頼性確保に関する事項
 - 感度、再現性、品質管理、対照品
 - 評価方法
 - 既存の方法との比較、標準品
 - 開発ガイドラインとの整合性

- II. 臨床性能の評価に関する事項
 - 装置によって得られるデータの臨床応用
 - 診断情報、予後の推定、治療薬への反応、モニタリング
 - 臨床応用の信頼性の確保
 - 症例数、統計手法の妥当性、探索的に行った研究結果の利用
 - レトロスペクティブ/プロスペクティブな試験
 - 先進医療の枠組みでの臨床性能試験の利用
 - Home Brew Test、Laboratory Developed Test (LDT)の位置付け
 - GLP 認証の利用等
 - 海外データの利用
 - 臨床性能を第三者機関が評価する可能性
 - 国際整合性

- III. 診断装置と診断方法を普及させる仕組みに関する事項
 - 承認を受けるメリット
 - 保険収載
 - 保険収載に頼らない方法
 - 医療経済学的考察

I. DNA チップ等を用いた RNA プロファイリング装置の評価指標案

理化学研究所 新興・再興感染症研究ネットワーク推進センター 神田忠仁
東北大学理学部化学専攻分野 分析化学研究室 寺前紀夫
国立医薬品食品衛生研究所 医療機器部 松岡厚子
国立医薬品食品衛生研究所 医療機器部 宮島敦子
国立医薬品食品衛生研究所 遺伝子細胞医薬部 鈴木孝昌

1. はじめに

DNA チップ等を使い、様々な遺伝子の転写レベル (mRNA 及び microRNA) を同時に測定する技術が急速に進歩している。疾患の発症に直接かかわる遺伝子群や治療薬への応答に関わる遺伝子群が同定され、その転写状況と病態の特性や治療薬の効果・副作用の関連も明らかにされつつある。これらの技術と蓄積されたデータは、個々の患者の予後の推定や最適な治療を選択する基盤となる。特定の遺伝子群の転写量の相対比 (RNA プロファイル) の解析装置を、その有効性と信頼性を確認して臨床現場に導入するために、診断補助装置として評価する指標が必要となっている。本指標案では、DNA チップを用いた遺伝子型判定装置の評価指標 (薬食機発第 0404002 号室長通知「DNA チップを用いた遺伝子型判定用診断薬に関する評価指標」、平成 20 年 4 月 4 日) の考え方を踏襲しつつ、RNA プロファイル解析に固有の事項を加えた。

2. 本評価指標の対象

血液や病変部等の RNA プロファイルから診断や治療に役立つ医療情報を得るために、DNA チップ等による RNA 量の測定系と RNA 量比を基に医療情報を導き出す専用の解析系から構成される装置を評価の対象とする。装置から得られる情報の臨床的意義については、個別の事例ごとに臨床データをもとに検証することになる。

3. 評価指標の位置づけ

本評価指標は、現時点で重要と考えられる事項を示したものである。今後の技術革新や知見の集積等を踏まえて改定されるものであり、申請内容に対して拘束力を持つものではない。本評価指標が対象とする製品の評価に当たっては、個別の製品の特性を十分理解した上で、科学的な合理性を背景にして、柔軟に対応する必要がある。

4. 評価に当たって留意すべき事項

(1) 品目の概要に関する事項

判定に用いる RNA が由来する遺伝子の名称及びコードするタンパク質や RNA の機能について明らかとなっている情報を記載すること。

1) 臨床的意義

プロファイル解析から得られる医療情報の有効性が、既に論文等により科学的に実証されている場合は、文献等を引用して説明すること。解析対象とする個々の RNA 分子についてもプロファイル解析における寄与度や個別の臨床的意義などについて説明すること。また海外での臨床試験を参考とする場合には、日本人集団での結果の同一性に関して、可能な範囲で考察すること。

2) 対象とする患者の範囲と解析内容の既定

対象とする患者を添付文書において明確に規定すること。プロファイル解析によって得られる医療情報の意義、治療への利用法について添付文書の中で具体的に記載・説明すること。

3) 測定原理

RNA 分子を定量する手法の原理を詳細に示すこと。RNA を増幅する場合は、増幅手法を説明し、増幅によってプロファイルが歪む可能性について記載すること。測定装置における測定原理を示すとともに測定原理が記載された論文、特許等の文献があれば引用し、必要な実側データと仕様に関する資料を添付すること。既存の測定機器を用いる場合には、使用可能な機種を限定するとともに、その資料を添付すること。

4) 遺伝子の選択

測定する RNA 分子種を選択した根拠を示すこと。

5) プライマー、プローブ等の塩基配列

逆転写反応に用いるプライマー配列及び各 RNA 定量のための特異的プライマー及びプローブ等の塩基配列を示すこと。偽遺伝子や類似配列を持つ他の遺伝子由来 RNA の存在を含めて、それらの配列を選択した妥当性を説明すること。ミスマッチプローブ等を判定に利用する場合には、その配列を設定した根拠を説明すること。

6) DNA チップ構成

DNA チップを用いる場合には、プローブの空間的配列と固定方法の特徴を詳細に説明すること。マイクロプレート等でリアルタイム PCR 法による検出を行う場合には、測定する遺伝子の位置情報を明記すること。

7) DNA チップに搭載される対照物質

陰性対照は、シグナルのバックグラウンド算定の根拠となるので、陰性対照と陽性対照の設定と妥当性を説明すること。例えば、陰性対照として検出する RNA 配列と類似の配列を複数搭載することが望ましいが、バックグラウンドシグナル値、陽性対照物質による内部標準等を用いて測定データの補正を行う場合には、その原理、手法について実測データを用いて詳細に説明すること。

8) アッセイ条件

ハイブリダイゼーション、洗浄、乾燥等の反応条件（温度、時間、緩衝液の組成等）の概略（アッセイのプロトコル及び標準手順を含む）を記載し、非特異反応が生じる可能性等も説明すること。

9) ソフトウェア

蛍光や電流値などのシグナル強度から解析機器に組み込まれたソフトウェア等によってプロファイルを得る場合は、測定アルゴリズムの妥当性に関して説明すること。ソフトウェアの動作に関するバリデーションの方法を示すこと。

10) 判定アルゴリズム

プロファイルから医療情報を導くためのアルゴリズムとその構築方法の詳細について、用いたデータセットを含めて説明すること。いったんアルゴリズムを確定した後は、その後の評価の過程において、その内容を変更するべきではなく、変更の必要が生じた場合には改めてバリデーションを行うこと。この場合、科学的な妥当性が示せれば、はじめのアルゴリズムの確立に用いたデータセットをバリデーションに用いることも可能である。アルゴリズム構築の際にも、トレーニングセットとは独立したバリデーション用データセットを用いて検証することが、その後の臨床性能試験の評価に役立つ。

(2) 仕様及び安定性に関する事項

1) 品質管理の方法

DNA チップを用いる場合は、固定されたプローブの塩基配列とデザインした塩基配列の同一性を、実測データを用いて説明すること。

特異的核酸増幅反応（PCR）を用いる場合は、使用するプライマーの純度、配列を確認すること。

用いる手法が、対象 RNA のレベルを測定する感度、正確性及び同時再現性を保証する標準試験を設定し、標準試験の成績からこれらの項目を検証する具体的な方法を、実測データを用いて説明すること。

2) 感度、特異性、測定範囲

一定の RNA（または cDNA）のコピー数を含む試料を希釈して測定し、定量的検出限界を示すこと。また、段階希釈試料を用いて、直線性を認めるデータの範囲を示し、測定濃度範囲を規定するとともに、補正が必要な場合にはその方法と根拠を説明すること。遺伝子工学技術によって作製した核酸や組織や培養細胞等により得られた RNA を標準試料として使う場合は、臨床検体由来試料の濃度や純度等に留意すること。

非特異的反応やバックグラウンドシグナルの安定性や均一性を検討し、誤判定の可能性を説明すること。試料中の RNA を測定できる最小検体量（RNA の最小必要量）を示すこと。必要に応じ、許容される最大検体量について検討すること。プロファイルの直線性の評価は、RNA レベルの異なる二つの試料を一定の割合にて混合した試料を用いることにより、検証することができる。また、データの精度を保障するための、データ受け入れ基準について明確にすること。

3) データの標準化

試料間で同一のレベルを示す RNA 分子種を複数用いてデータを標準化することが望まれるが、それらの選択根拠と標準化の手法を説明すること。標準化のための RNA 分子種を選択が難しい場合は、全 RNA 量の総和等により標準化（グローバルノーマライゼーション）を行うことになるが、その方法の詳細と妥当性に関して説明すること。標準化の際のバックグラウンド値の測定方法とシグナル強度の補正方法を説明すること。

4) 測定装置の較正

一定のシグナルを安定して発生する較正用 DNA チップもしくはそれに順ずる標準試料を装置の較正に用い、動作バリデーションを定期的に行うことができる場合は、較正用チップ、標準試料の妥当性を説明すること。これらが利用できない場合には、陽性及び陰性較正用試料を用いた測定値の評価等によって動作確認をとる方法を示すこと。較正用試料の妥当性を説明すること。測定・解析装置を一体として評価する際に、これらの情報が必要になることに留意すること。

5) 安定性に関する資料

DNA チップや測定用キットの保存条件、有効期限を設定し、その妥当性を説明すること。使用者が調製する試薬類がある場合は、調製方法や品質管理の方法を示すこと。

(3) 性能に関する事項

1) プロファイル取得の精度

RNA の定量精度は、個々の RNA について信頼性の確立されたリアルタイム PCR 法などや、精度がリアルタイム PCR 法によって確実に担保されている DNA チップを用いて検証する。同等の既存品がある場合には、それとの比較によって検証する。

この際、RNAの絶対量が一致する必要はなく、標準試料または異なる試料間での量比が一定となることを示すことで、精度は担保される。

2) 検体と共に測定する対照試料

陽性対照試料、陰性対照試料を選定し、選定した理由を説明すること。公的機関より適切な標準試料が供給されている場合には、これを用いること。そして、それらを用いた精度管理の方法を、実測データを用いて説明すること。

3) 再現性、頑健性

標準試料を用いた繰り返し測定によるシグナル検出及びRNAの相対量比の再現性に関する検討を行うこと。同一施設内で場所や作業者を替えた測定や、複数施設における測定によって、再現性を確認すること。必要に応じて頑健性に関する情報、外部精度管理の方法に関する情報を提供すること。

4) コンタミネーション対策、データ取り違え対策

検体の前処理や測定においてPCR等による核酸の増幅過程が含まれる場合、コンタミネーションによる誤判定の可能性とそれらを排除するための方策を、必要に応じて実測データを用いて説明すること。また、キャリーオーバーを否定する試験を実施して、コンタミネーション対策の妥当性を示すこと。バーコード等を使ったデータ管理システムにより、検体情報および解析結果の対応に誤りが起こらないような方策が求められる。検体や試薬類の入れ間違いなど人為的ミスを防ぐ方策と、それを確認できる方法に関して検討を行うこと。

5) 検体の調製

使用する試料が不安定なRNAであり、検体の質がプロファイルに大きな影響を与えるため、検体の取り扱いには細心の注意が必要である。高品質なRNA試料を得るために、採取する検体の種類に応じて、採取、保管、運搬等に関する適切な取扱い方法を設定し、その妥当性を説明すること。特にRNAの分解を防ぐ方策を講じることが望ましい。

検体の種類（血液、組織等）に留意しながら、検体からRNAを抽出する方法と得られた試料の品質を評価するための方法または参考値（量、純度、分解度等）を示すなど、試料の評価基準を明確にすること。また、調製した試料の安定性について説明すること。必要に応じて反応を妨害する物質（血清中のトリグリセリド、ヘモグロビン、ビリルビン、脂質などや投薬された薬物、検体採取に用いた抗凝固剤等）について予め評価しておくこと。

6) 測定装置

基本的に、DNAチップ等と測定装置は一体として評価される。専用の測定装置を使用する場合には、その測定装置が一般的な医療機器として満たすべき性能を持ち、医療機器として認可される必要がある。汎用性のある測定機器を用いる場合で、その装置が医療機器としての認可を得ていない場合には、使用できる機器の限定と性能を担保する方法に関して、説明をすること。

(4) 臨床性能に関する事項

当該装置を使用して得られる情報の臨床的有効性は、適切に計画された臨床性能試験の成績を持ってこれを担保する。臨床性能試験に要求される事項については、別途詳細に記載する。

(5) リスク分析に関する事項

操作過程において、人為的及び機械的ミス、非特異反応等が発生する要因を分析し、必要に応じて添付文書にて注意喚起を行うなどの対策を講じること。誤った医

療情報が得られた場合に起こりうる、診断、治療上のリスクについて、文献等を使って評価すること。判定結果を別の手法を用いて個別に確認するための方法について、積極的に提示すること。診断結果は直接ゲノム情報を含むものではないが、それに準ずる情報を含むと考えられるため、個人情報としての取り扱いに留意し、倫理面で配慮すること。

(6) データの保存と医療情報の表示方法に関する事項

得られた医療情報に加え、測定した各 RNA の補正、標準化前後のシグナル値を含めて保存し、検証を可能とすることが望ましい。医療情報の開示方法については、あらかじめ添付文書等でその形式と臨床的意義に関する説明を明記することが望ましい。また、全工程が良好に進んだかどうかの判定項目を示すとともに、医療情報の根拠となるデータ（指標とした RNA の発現レベル等）を可能な限り開示する。

II. DNA チップ等の遺伝子発現解析装置の製造販売承認申請における 臨床性能評価に関する考慮すべき事項

京都大学医学部附属病院乳腺外科 戸井雅和
国立がん研究センター研究所多層オミックス・バイオインフォマティクス分野
佐々木博己
東京大学医科学研究所先端医療研究センター臨床ゲノム腫瘍学分野 古川洋一

1. 背景

DNA チップは、DNA、蛋白質、脂質などの高分子化合物および低分子化合物や病理組織片など様々な種類の並列解析要素を多数、基板上に載せたものの総称である。そのうち DNA チップ（別名、マイクロアレイ）とは、DNA をスライドガラス等の基板にスポットもしくは直接合成して、整列させたものの総称である。蛍光標識した検体の DNA または RNA をチップ上のプローブとハイブリダイゼーションすることにより、検体に含まれる DNA または RNA を定量することができる。1996 年ごろから普及しはじめた。

癌を例にすると RNA を対象とする発現解析の用途は 2 つに大別できる。1 つは探索的研究 (Gene finding) である。これは早期発見や再発予測に役立つマーカーの分離、癌の発生や進展に関与する遺伝子の同定、癌細胞と間質の相互作用の把握、病理学的分類と遺伝子発現に及ぼす信号伝達経路の対応、予後不良症例における創薬ターゲットの同定など新しい治療法の開発に役立つ。もう 1 つは分類や予測の研究 (Class discovery と Prediction) である。これは疾患そのものの同定と重症度の分類、同一部位の腫瘍におけるサブタイプの同定、治療感受性の予測などによる個別化医療の実現に役立つことが期待されている。

2006 年に米国のマイクロアレイ品質管理コンソーシアム (MAQC-I) は、種々のマイクロアレイプラットフォーム (メーカーや研究機関ごとのアレイ解析システム) の性能を人工調製した標準試料や薬剤処理したラットの生体試料で評価したところ、概ね再現性もよく、臨床診断機器として認可も近いと結論した。しかし、現在、世界で承認されているアレイ解析装置は、Affymetrix 社 (米) の Gene profiling reagent and profiling array-GCS3000Dx-AMDS のみである。これは GMP 対応 U133plus2.0 アレイ、cRNA の標識キット、工程管理ソフト AMDS 搭載の装置 GCS3000Dx を含んでいる。この Affymetrix 社のプラットフォームを利用した具体的な診断用チップ (コンテンツ) としては、AML Profiler (蘭、AML を 7 種に分類)、MapQuant Dx Genomic Grade (仏、乳癌の病期 I, II, III の分類)、Pathwork Tissue of Origin (米、原発不明癌の起源同定) がある。前 2 者は FDA の承認 (PMA 又は 510K) は受けていないが、EU で Reference Lab Test として提供されている (Kim C, Palk S. Nat Rev Clin Oncol 7: 340-347, 2010)。一方、Pathwork Tissue of Origin は承認 (510K) されている。しかし、その運用は LDT (Laboratory Development Test) のように、特定の施設で行われている。Almac Diagnostics 社 (英) などは、この装置で診断用アレイ開発を行っている。今後、シスメックス (株) などわが国での販売権を持つ企業からもコンテンツの開発が期待される。Agilent 社 (米) の装置は医療機器としてはまだ未承認であるが、FDA から 2007 年 2 月に承認された DNA チップがある。早期乳癌 (T1, T2/N0) の術後 5 年以内の再発リスクを予測する方法である。現在、MammaPrint という商品名で Agendia 社 (蘭) がサービスを行っている。最近、同社は乳癌を 3 タイプに分ける BluePrint なども開発した。その他、Febit 社 (独) も積極的に装置、コンテンツの開

発を行っている。また、DNA チップ以外にも RT-PCR を基礎としたホルモン受容体陽性乳癌の化学療法の効果を予測する製品 Oncotype DX（米）が実用化されている。さらには、2008 年に NanoString 社（米）は蛍光分子イメージングによるデジタル発現プロファイリング法を開発した。この装置の名称は nCounter といい、定量 RT-PCR よりも精度が良いと報告されている。一度に約 2000 遺伝子の発現量を測定することが可能である。

このような世界的な動向もあり、わが国でも、2007 年 10 月に米国の MAQC に相当する JMAC（日本マイクロアレイコンソーシアム）が 67 社によって立ち上げられた。

このような背景から発現解析型 DNA チップ等の製造販売承認申請に係わる審査評価指標策定への要望が高まっている。ここでは、製造販売業者による申請の万全の準備と医薬品医療機器総合機構（PMDA）による審査に役立てることを目的とし、臨床性能の評価に関する考慮すべき点を整理したので報告する。なお整理にあたっては、2007 年に 5 月 9 日付けで発表された乳癌の予後予測に関する遺伝子発現プロファイル検査に関する FDA のガイダンス「Class II Special Controls Guidance Document: Gene Expression Profiling Test System for Breast Cancer Prognosis」を参考にした。しかし、このガイドラインは乳癌の予後に特化されたもののため、本稿では、より多様な臨床性能に関する考え方や評価に拡大する必要があったことを明記する。また、医療イノベーション推進のためにはグローバルな展開が必要であることから、わが国の制度に合致しつつ、国際整合性が考慮された評価が行われるための指標となることを目標にした。この報告が国民の利益につながる安全性が担保された次世代体外診断薬の遅延なき導入に向けた申請、評価の助けとなることを望む。

2. 製品（DNA チップ等遺伝子発現解析装置とそのプロトコール）に関して（Device Description）

DNA チップ等の遺伝子発現解析診断システムの性能を適切に示す為に、以下の項目の情報を記載しなければならない。また、PMDA の審査を助けるような製品の概要を説明する図や写真、既存の診断法との比較表などを添付するべきである。また、製品の性能や臨床的意義を示した査読のある論文やその他参考になる総説、説明書・パンフレットがあれば必ず提示する。

(1) 使用目的（Intended Use）

どのような試料からどの遺伝子（群）の発現をどのように調べ、如何なる臨床性能を発揮するのか明記する。また、検査の対象とする被験者の臨床病理学的情報（疾患の有無、疾患名、人種、年齢、性別、病態、病期、重症度など）に関しても記載する。発揮する臨床性能が定性的か定量的であるか、についても明記する。

(2) 検査方法（Test Methodology）

- 検査のプラットフォームは何か（例、DNA チップか、RT-PCR ベースか、その他か）、
- 検査が限られた解析施設（検査会社等を含む）で行われることを予定する場合は、施設に関する情報も記載する。
- DNA チップの場合：その工程は、1) DNA チップの製造、2) 試料の採取・保存・輸送、3) RNA の調製と質の評価、4) 標識反応、5) ハイブリダイゼーション、6) 洗浄、7) 発現強度の測定、8) 測定アルゴリズムによる標準化等の操作、9) 判定アルゴリズム

による診断、10) 診断結果報告書の作成、に分けることができる。ステップ数が多く、装置や器具の組み合わせも多いのが特徴である。また、すべての工程が最終的には臨床性能に反映され、逆に臨床性能をもって、すべての工程が評価できることもこの分子診断の特徴とも云える。したがって、各項目について詳細な検査方法の説明が必要である。その際、使われる試薬、キットについては研究用試薬 (research-use-only, RUO) は推奨しない。

1) DNA チップの説明

基板上に整列化したプローブの種類 (オリゴヌクレオチドまたは cDNA かなど) ・配列・構成、どのような基板にどのように固定または合成したのかについて、記述する。

2) 試料の採取・保存・輸送

DNA チップの遺伝子発現解析装置に供する試料は、生検や外科的切除による生組織やホルマリン等による固定組織の他、血液・骨髄液・腹腔洗浄液などの体液、汗・尿・便などの排泄物などと多様である。そのため、解析対象とする試料の範囲、その採取・保存・輸送法に制限や推奨される方法がある場合はその詳細を申請 (試料) 毎に記述する。

3) RNA の調製法と質の評価

組織、血液、骨髄液のように数十マイクログラムの RNA が得られるものから、排泄物のように定量できないほど微量なヒト由来 RNA が得られる試料がある。そのため、RNA の調製法に関しては、試料毎に試薬、器具等を含めた詳細なプロトコルを記述する。RNA の質や量の評価を必要とする場合は、その方法と解析に使用できる限界 (または推奨範囲) を明示する。

4) 標識反応

DNA チップと反応する核酸は蛍光物質等で標識される。蛍光物質や酵素などを含め標識に用いる試薬や、推奨する器具や装置、標識方法に関する詳細なプロトコルを記述する。

5, 6) ハイブリダイゼーションと洗浄

標識された核酸は DNA チップとのハイブリダイゼーションとそれに続く洗浄操作で全反応工程を終える。ハイブリダイゼーションや洗浄のための装置は、従来からフィルターで行われていたサザンロットやノーザンロットハイブリダイゼーションで使われていた装置を改変させたもので、専用の装置とは必ずしも云えない。また、各ステップの時間にも幅がある。標識された核酸と DNA チップとのハイブリダイゼーションやそれに続く洗浄操作で使用する推奨 (あるいは指定) 装置や器具、それぞれの反応液の組成や量、反応温度と反応時間、攪拌条件などを含んだ詳細なプロトコルを記述する。

7) 発現強度の測定

ハイブリダイゼーション・洗浄を終えるとレーザー光や電氣的・化学的検出方法によって、DNA チップ上の全てのプローブから信号を検出する工程に移る。シグナルを検出する為の方法 (電氣的・化学的検出など)、出力データの種類、検出に用いる推奨 (あるいは指定) 装置、データ検出の為の出力条件やパラメーター (測定範囲) など、詳細な測定方法を記述する。

8) 測定アルゴリズムによる標準化等の操作

検出された生データから最終的な判定を行うまでの、データの変換や評価の為の方法を明示する。DNA チップ解析では、1 枚の DNA チップから得られるシグナル値 (蛍光強度) は、内部コントロールプローブ (ハウスキーピング遺伝子)、外部コントロールプローブ (ラムダファージなど)、スパイク配列プローブ (植物遺伝子など試料

中の RNA とハイブリダイズしない配列) および mismatches 配列プローブ (人工的な非ハイブリ配列) の値からバックグラウンド値などの基準値を設定し、第一次標準化がなされる。さらにチップ上の解析対象となる全遺伝子プローブのシグナル値を一定の値にするなどの方法で第二次標準化を行い、DNA チップ間のデータを比較解析できるようにするのが一般的である。多検体比較解析では全てのデータで統一的な標準化を行う場合もある。これらの標準化はゲノムワイド DNA チップでもカスタム DNA チップでも同様に行われ、判定アルゴリズムの構築や臨床性能に大きな影響を及ぼす。したがって、この標準化を行うための測定アルゴリズムを説明することが必要である。特定の内部コントロールがある場合には、システムにおけるその役割とその使用法を明記する。また外部コントロールを用いる場合は、推奨 (あるいは指定) 物質やその用法についても記載する。複雑な計算を要するアルゴリズムや権利化されていないブラックボックス部分についても、可能な限り概要の説明を加えるべきである。さらに、標準化後のシグナル値の対数変換等の補正が判定アルゴリズムの構築に必要な場合は、それも明記する。

9, 10) 判定アルゴリズムによる診断と診断結果報告書の作成の項目は、次の 3、「臨床的意義を示すために考慮すべき点」で詳述する。

- RT-PCR などをベースとしたその他のプラットフォームの場合：上記のうち 2) 試料の採取・保存・輸送と 3) RNA の調製と質の評価、に関しては DNA チップと基本的要件は同じである。1) DNA チップの製造に相当する部分として、プライマーやプローブ等に関する記載が必要となる。4) 標識反応、5) ハイブリダイゼーション、6) 洗浄、7) 発現強度の測定、8) 測定アルゴリズムによる標準化等の操作に関しては、サイクル数や蛍光強度の測定など、それぞれのプラットフォームに応じた説明が必要となる。9) 判定アルゴリズムによる診断と 10) 診断結果報告書の作成に関しては、DNA チップの場合と共通する項目として後述する。

(3) DNA チップ等自体の品質保証 (Quality Control)

DNA チップの性能を担保するための品質に関して、以下の項目を記載する。

- DNA チップ上に配置されたプローブを正しく識別しているかどうかについての品質評価
- 複数のプローブと結合する場合、各プローブの特異性・非特異性結合の検討事項
- 複数のプローブを配置する際の誤配置や交差汚染の防止や、作成における均一性・不均一性に関する評価事項

3. 臨床的意義を示すために考慮すべき点

遺伝子型診断用 DNA チップ解析と比較し、発現解析型 DNA チップ解析のもつ臨床的意義は広範に及ぶ。ここでは発現解析型 DNA チップの用途を分類し、臨床的意義を示すために必要なデータの提示について国際整合性を含めてまとめた。

(1) 発現解析型 DNA チップの診断法の開発での用途

血液や洗浄液などの体液や排泄物を試料として解析した場合、癌細胞などの病変細胞の存在診断のためのマーカーや病態に相関する代理 (surrogate: サロゲート) マーカーの検出に使うことができる。それによって、疾病の発症リスク診断、早期発見やスクリーニング、重症度診断、医薬品の副作用・効果の判定およびモニタリングに役立つ。

つ可能性がある。病変部の生組織や固定組織を使った場合も、癌を含む特定の疾患のサブタイピングや重症度診断、治療効果や副作用の予測などに役立つ可能性がある。したがって、申請者はこれらの多様な用途を把握し、それぞれの特徴に合わせた診断アルゴリズムの構築と臨床的意義の提示を行う必要がある。

(2) 臨床的意義と判定アルゴリズムの多様性

臨床的意義は大きく2つに分けることができる。1つは、医師による確定診断や日常的な監視と併せた診断を補完するものである。例えば、患者の早期発見やスクリーニング、癌のような生死につながる疾患以外の重症度診断などが該当する。この場合、画像診断、内視鏡的所見、病理診断などの既存の診断の補完である。特に早期発見やスクリーニングは医療経済効果も高く、市場も大きい。他に方法がない場合や既存の方法に問題がある場合、非常に有用である。もう1つは、治療方針の決定に結びつく診断で、患者の予後やQOLを左右するものである。例えば、癌のサブタイピングや病期分類、治療効果や副作用の予測などが該当する。申請にも評価にも慎重さが求められるが、他に効果的な既存方法がないことが多く、患者にとっては早急な導入が必要とされる診断法である。

判定アルゴリズムまたは判定基準の設定を分類すると以下ようになる：

- 1) 癌細胞などの病変細胞の存在診断では、陽性か陰性かを定性的に判定するためのシグナル強度（遺伝子発現量）を基にした判定閾値（Cut-off 値）の設定が求められる。
- 2) サブタイピングや重症度診断のような病態分類では、遺伝子発現プロファイルを基にした分類器（Classifier）の構築が求められる。
- 3) 治療効果や予後判定では、比較的多くの遺伝子の発現量を基にした2群（高リスクと低リスク）の予測器（Predictor）や3群以上（高中低など）の階層化予測器の構築が求められる。

このように、臨床的意義と判定アルゴリズムには多様性があるので申請側も審査側も十分な理解の下に適切に対応する必要がある。

(3) 判定アルゴリズムや判定基準の設定等について

DNAチップ等の遺伝子発現解析装置を用いた診断では、複数の遺伝子の発現パターンから上述の判定アルゴリズムを構築する必要がある。

そのアルゴリズムの妥当性を評価するために、次の事項が必要とされる。

- アルゴリズムの構成と判定（あるいは分類）対象事項。
- アルゴリズムに用いるプローブ名（遺伝子名）、それらのシグナル強度（遺伝子発現量）を用いたアルゴリズムの構成。
- アルゴリズムによる判定（あるいは分類）のための閾値、およびその設定根拠。
- 既存の機械学習プログラムや分類プログラムでの場合、その名称や参考文献。
- 新規プログラムまたは修正版の場合、アルゴリズムの作製や評価に用いたデータセット、その為に用いたサンプルの選択基準、サンプルサイズに関する統計学的妥当性。
- 結果報告の見本、およびそれぞれの判定（あるいは分類）結果の解釈に関する説明書の見本を添付する。説明書には、その結果示す感度、特異性、正診率などの情報が、具体的かつ分かり易く記されている必要がある。

加えて判定アルゴリズムの検証のために、次の臨床性能試験を示すことが求められる。

(4) 国際整合性を考慮した臨床性能試験について

FDA は、背景で紹介した Affymetrix 社の DNA チップで開発された Pathwork Tissue of Origin や Agilent 社の DNA チップを基に開発された MammaPrint をレトロスペクティブに集めた試料（過去の検体）を用いた臨床性能試験結果から承認（510K）している。また、レトロスペクティブに集めた検体で示された臨床性能が、治療法の変更もなく今後も有用なものであれば良いとの見解を表している（Guidance for Industry and FDA Staff: Class II Special Controls Guidance Document: Gene Expression Profiling Test System for Breast Cancer Prognosis 13-16 行、ページ 14、2010 年 LDT に関するワークショップにおける FDA の講演）。したがって、過去の検体を用いた臨床性能評価であっても、その DNA チップを用いた診断アルゴリズムが生み出す結果が現在または将来ともに適用できるものであれば、過去の検体を使った臨床性能試験を評価資料として考慮することができる。例えば遺伝子発現による病理組織学的な診断などがこれに当たる（例、Pathwork Tissue of Origin）。また検体の得られた時期と現在とで、当該疾患の生物学的特性が変わらない検体や、治療法やその臨床経過に変化がないものに関しても、これに準じて評価が可能である（例、MammaPrint）。ただしこれら過去の検体を用いた性能評価を添付する場合には、過去の検体を用いた評価が、現行の医療における評価との間に大差がないことを示す根拠を添付しなければならない。

後向き試験にしても前向き試験にしても、臨床性能試験の中で以下の項目を提示する必要がある。

- 1) 臨床試験プロトコール：疾患、患者数、人種、治療法、施設、エンドポイント（例えば、再発・転移までの期間や率、全または無病 3 年または 5 年生存率など）を明記する。
- 2) 客観性を保つための検体と臨床病理情報の整理：バイアスのない患者集団であるかの検討や臨床情報のブラインド化を明記する。例えば、一般的な患者集団の臨床病理情報との比較、半数の検体で判定アルゴリズムを構築し、半数の検体で検証するなど。
- 3) 海外データ：海外発の製品である前述の MammaPrint や Oncotype DX などは既に膨大な海外での臨床性能試験のデータが集積されている。海外でのデータがある場合には、参考値としてそのデータ（もしくは文献）の提出が求められる。ただし、海外で有用性が認められた検査を評価する際には、国内メーカーが開発する場合と同様に、日本法人などの製造販売拠点があり、製品の安全性を担保できる場合に限る。また評価の際には、日本人の臨床検体に対する試験データを加えて評価することが望ましい、各国の国民が医薬品に求めるリスクベネフィットバランスは異なっており、今後、わが国での海外のデータを活用するための考え方を整理する必要がある。
- 4) 診断結果：癌細胞などの病変細胞の存在診断では、疾患のスクリーニングの場合は日常検査になるため、何%の確率である疾患の疑いがあるなど。病院での検査では、例えば体液中に癌細胞がいる疑いが強く、陽性患者の何%は 2 年以内に再発し、5 年生存率が何%と低いなど。組織を用いた治療効果や予後判定では、単に有為差をもって治療効果が良い群に予測されたという報告ではなく、何%の患者が 5 年生存できる治療感受性群と診断されたなどの具体的な診断結果を提示する。同時にその統計学的根拠も説明する。
- 5) 再現性の確認または検証：体外診断用医薬品の臨床性能試験成績に関する資料として、2 施設以上、150 検体（症例）が通常求められており、これに準拠した試

験による確認や検証が望まれる。しかし、判定に用いた遺伝子セットの発現パターンに生物学的な根拠がある場合や、統計学的に解析が可能な場合は、この限りではない。生物学的根拠に関しては、実験的な証明に基づくことが重要であり、文献の提示が必要である。また学会の推薦など専門家の判断を示すことも重要である。統計学的解析に関しては、閾値の設定、分類器、予測器の構築に用いる検体（トレーニングセット）とその検証セットを交換した際の食い違いがないかなどの検討結果を明記する必要がある。例えば、5年生存率やハザード比が10%以内にあることなどを明記し、判定アルゴリズムが調べた検体の中で「収束した結果を示す」ことが必要である。

しかし治療効果や予後の予測の確認・検証では、短期間に少数の施設で十分な数の試料を集め試験を行うことが困難なことが少なくない。一方、患者の利益や医療経済を考えれば、社会として「早期導入」が望まれるものも存在する。わが国には米国で行われているLDTのような仕組みはないが、上述のような「収束された判定アルゴリズム」の申請に関しては、例えば、市販後に積極的にコンソーシアムを組んで前向き試験を行い、データを蓄積して、その結果を（学術発表等で報告することを含め）当局への報告することを条件に、試験的検査機器として条件付き承認することを提案したい。臨床試験で行われる自己規制条件によるストップルールと同様に、前向き試験において条件付で承認された製品の性能が発揮できなかった場合は、サービス・販売を中止することになる。一方、前向き試験体の後向き使用によって診断アルゴリズムの改良が可能であった場合は、再申請を可能とする。この場合も、条件付承認となる。これは、LDTでの使用後の本承認（PMA 又は 510K）に膨大な試験を要求するFDAの審査システムと矛盾しない。

一方、DNAチップのような次世代診断機器は高価で、作業工程も多く、データの提出も単純ではない。したがって、検査部（科）が充実していて、専任の技師や実験補助員を継続的に雇用できる医療施設でのみ、使うことができると見るのが現実的である。そのため、普及には検査センターで解析する仕組みが必要であり、上述の市販後の前向き試験は、2項先進医療の枠組みで行い、複数の拠点病院から検査センターへ検体を送ることで実施されるものとする。この普及の仕組み作りの提言に関しては、他稿（III. 診断装置と診断方法を普及させる仕組みに関する事項）を参考にさせていただきたい。

4. 性能の特徴 (Performance characteristics) に関する記述

- (1) プローブやプライマー等の特異性に関して、記述する。例えば、ゲノムワイド解析型DNAチップなら、既に評価されているチップとの比較、限られた数のプローブでできているミニDNAチップなら、定量的RT-PCR等との比較が可能である。
- (2) 推奨する試料収集・保存・輸送法について、例えば、生検または外科的切除試料か、生か、凍結か、固定するかなどの区別、保存や輸送の方法、温度や時間に関する評価事項を明記する。その他、試料に要求される条件として、適切な大きさ、癌細胞の含有量、壊死や炎症部分の比率、コンタミネーションなどについて記述する。
- (3) RNA抽出に関して、推奨する試薬とその有効期限・安定性、安全性などを明記する。
- (4) RNAの品質管理に関しては、例えば、キャピラリー電気泳動等の手法で、臨床性能に及ぼす質を28S/18SリボソームRNAの比率やRNA Integrity Number (RIN)等でDNAチップ解析前に評価する必要がある場合はそれを明記する。一方、DNA

チップ解析後の内部コントロールプローブ等のデータから発現強度の測定に及ぼすRNAの質を評価するシステムを採用している場合は、それを明記する。

- (5) 工程の管理：全反応工程または多くの工程の管理（再現性、試薬や装置の安定性など）を較正用試料（RNA）によって行える場合はその方法を明記する。標識反応では、その効率（標識または増幅）が発現強度に及ぼす影響の評価指標の1つとして採用する場合は、その閾値等を明記する。一方、DNAチップ解析後の外部コントロールやスパイクプローブ等のデータから発現強度の測定や臨床性能に及ぼす影響を評価するシステムを採用している場合は、それを明記する。ハイブリダイゼーションと洗浄では攪拌の条件、時間、温度の管理に関する人為的ミスの予防に必要な注意点がある場合は、それも明記する。スキャナーは、専用装置としてクラスI登録（製造販売届出）が要求され得る。しかし、先行して承認されている酵素免疫測定法（ELISA）や化学発光免疫測定法（CLIA）によるウイルスや病原菌の検出では、酵素反応から検出まで全自動で行う一体型の専用装置を使うもの、クラスI届出された複数の検出装置を推奨機器として指定しているもの、推奨機器を指定せずに承認されているものが混在している。一方、DNAチップの装置については、ハイブリダイゼーション・洗浄装置、スキャナーおよび工程管理ソフトなどがシステム化されている機器はAffymetrix社のGCS3000Dx-AMDSのみである。上述の免疫測定法の承認例を考慮すると、推奨する各スキャナーでのレーザー光の出力条件、時間の他、各スキャナー間でのデータの較正法も明記し、臨床性能の互換性を示すことができれば、必ずしもクラスI届出した専用装置を必要としない。しかし、医療機器として届出された専用装置の場合に当然実施する再現性に影響を与える性能の定期的な確認は義務である。また、機器間差が想定される場合には補正方法を組み込むことも必要である（想定されない場合はその根拠を明記することが必要）。さらにモデルチェンジに伴う補正や再現性の確認も機器メーカーと相談しながら実行する体制についても明記することも必要である。

5. データの保存と提出方法での考慮すべき点

最終的な診断結果のみならず、測定した各遺伝子の補正、標準化前後のシグナル値を含めて保存し、各データ処理の段階毎に動作確認を可能とすることが望ましい。個人情報保護法に基く、匿名化管理も必要である。

診断結果の提出方法に関しては、全工程が良好に進んだかどうかの判定項目を示すことが必要である。その結果、診断可能、不能の区別を示すことになる。さらに個別の診断結果の報告に関しては上述のように、かなり具体的な報告となることが想定される。すなわち、患者からみた価値の増加を分かり易く説明すると、受診者によっては非常にショックを受ける診断結果である場合もあるので、臨床家と申請者のみならず、当局も含めて調整する必要がある。

追記：わが国で開発された製品のグローバルな事業展開や海外製品の承認、普及を考えた場合、主要国の審査ガイドラインの日本語版も必要だが、わが国の評価指標の英語版も必要と考える。これらは、PMDAのホームページからダウンロードできることが望ましい。

III. 診断装置と診断方法を普及させる仕組みに関する事項（提言）

三重大学大学院医学系研究科検査医学分野 登 勉
名古屋大学大学院医学系研究科附属神経疾患・腫瘍分子医学研究センター
腫瘍病態統御部門・分子腫瘍学分野 高橋 隆

1. 高度医療評価制度

医学医療の高度化やこれらの医療技術を臨床に応用するため、薬事法の承認が得られていない医薬品・医療機器の使用を伴う先進的な医療技術を「高度医療」として認め、保険診療と併用できることとし、薬事法上の承認申請に繋がる科学的評価が可能なデータの収集を迅速化することを目的とする制度である。その対象となる医療技術は（1）薬事法上の承認又は認証を受けていない医薬品・医療機器の使用を伴う医療技術と（2）薬事法上の承認又は認証を受けている医薬品・医療機器の承認内容に含まれない目的の使用（いわゆる適応外使用）を伴う医療技術の2つである。高度医療評価制度に係る申請等の取扱いや実施上の留意事項については、「高度医療に係る申請等の取扱い及び実施上の留意事項について」（平成20年3月31日付け医政発第0331022号厚生労働省医政局長通知）に示されている。

高度医療評価制度による薬事承認までのロードマップを図1と2に例示した。臨床研究の結果に基づいて高度医療が申請され、薬事承認申請に至るまでには膨大な時間と経費が予想される。しかしながら、保険診療と併用できることは、国民の選択肢を拡げ、利便性を向上するという観点から歓迎すべきことと考える。

2. 先進医療

先進医療は、健康保険法等の一部を改正する法律（平成18年法律第83号）において、「厚生労働大臣が定める高度の医療技術を用いた療養その他の療養であって、保険給付の対象とすべきものであるか否かについて、適正な医療の効率的な提供を図る観点から評価を行うことが必要な療養」として、「評価療養」の1つとされた。薬事法上の承認申請に繋がる科学的評価が可能なデータの収集を迅速化することを目的とする高度医療評価制度に対して、将来的な保険導入のための評価を行うことを目的としている。先進医療は、「第2項先進医療」と「第3項先進医療（高度医療）」に分類され、前者は、薬事法上の未承認又は適応外使用である医薬品又は医療機器の使用を伴わず、未だ保険診療の対象に至らない先進的な医療技術と定義されている。一方、後者は、薬事法上の未承認又は適応外使用である医薬品又は医療機器の使用を伴い、薬事法による申請等に繋がる科学的評価可能なデータ収集の迅速化を図ることを目的とした先進的な医療技術と定義され、前項の高度医療と同義である。

第2項先進医療は、平成23年3月1日現在で90種類（第3項先進医療技術31種類を除く）が実施されている。しかしながら、これら90種類の先進医療のうちで遺伝子検査を医療技術として用いる多くの場合、薬事法上の承認又は認証を受けた診断薬や診断キットでなく自家調製であるHome-brew assayである。定義上は、これらの先進医療は第3項に分類されるべきものであるが、第2項として処理され、その中から既に保険導入されたものもある。

3. 保険収載された遺伝子検査における課題

平成22年4月の診療報酬改定で改定された遺伝子検査関連項目を表1に示した。表中赤で示した項目は、既に承認された診断薬や診断キット（以下、キット）が存在するものであり、それ以外は、保険収載されているが薬事承認されたキットのないもの

である。PCR 法、DNA シーケンス法、FISH 法またはサザンブロット法などの方法が記載されているが、自家調製試薬類あるいは薬事承認されていないキットを検査に用いて診療報酬を得ることは可能である。精度管理された試薬・方法での遺伝子検査が理想であるが、現実には異なっている場合があり、非常に大きな問題である。

また、遺伝子検査の大半は、診療報酬点数が 2,000 点であり、開発コストや検査実施コストを回収できない点数になっていることも、今後の遺伝子検査の臨床応用を考える場合の課題である。診療報酬点数については、上記のコストと医療経済効果も考慮した評価が望まれる。

米国疾病管理予防センターCDC は、遺伝子検査について Analytic Validity、Clinical Validity、Clinical Utility、Ethical, Legal and Social Issues (ELSI) の 4 項目を評価する ACCE モデルを公表している。本邦においても ACCE モデルに準じた評価が実施されることが望まれるが、医療機関、検査施設、民間企業等で独自に検査法の開発・実施が行われ、いわゆる Home-brew assay による検査が中心である現状では、統一した評価の実現までには至っていない。

さらに、遺伝子検査項目は同じであっても、使用方法・機器の違いが非常に大きく、測定方法についても開発者・企業によって独自性があり、測定者の技術に格差があるといった点が検査法の標準化や精度管理を困難にしている。現状では、分析的妥当性、臨床的妥当性、そして臨床的有用性の確立した遺伝子検査の提供は難しいということになる。

4. 診断方法を普及させる仕組み

図 3 に遺伝子検査の開発から臨床応用までの新ルート案を示した。日本では厳密な意味の Laboratory Developed Tests (LDT) は存在しないし、Home-brew assay と同じではないが、図 3 では便宜上同じものとして表記した。Home-brew assay が先進医療として利用される場合には、精度保証や分析的妥当性が重要になる。したがって、PMDA がこういう働きをして、非常にオンタイムにやっていただければいいのですが、なかなかそうもいかないということで、もし可能であれば、NPO 法人、例えば Japan Molecular Diagnostic Standards (仮称)、が開発者・企業から提供された Home-brew assay の分析的妥当性を検査・認証し、その結果をもって先進医療専門家会議が先進医療技術として、Home-brew assay の開発者が直接に、或いは、技術移転を行ってものに対して承認を与えて、第 2 項先進医療として臨床応用される仕組みを提案する。それによって、現状分析の項で指摘した如くの問題点を持つ、似て非なる精度が保証されていない検査の実施を防止可能とする。そして、最終的には保険収載へと繋がるのが、本邦における遺伝子検査の普及にとって現実的ではないかと考える。なお、このような検査法の国内における研究開発と普及を促進するためには、上述したように、診療報酬点数の算定において、開発及び実施コスト並びに医療経済効果を考慮した評価を行って、研究開発へのインセンティブとする必要がある。

図1. 薬事承認までのロードマップ(治験)

試験薬または試験機器: ○○○(製品名: × ×)
 高度医療での適応疾患: △△△

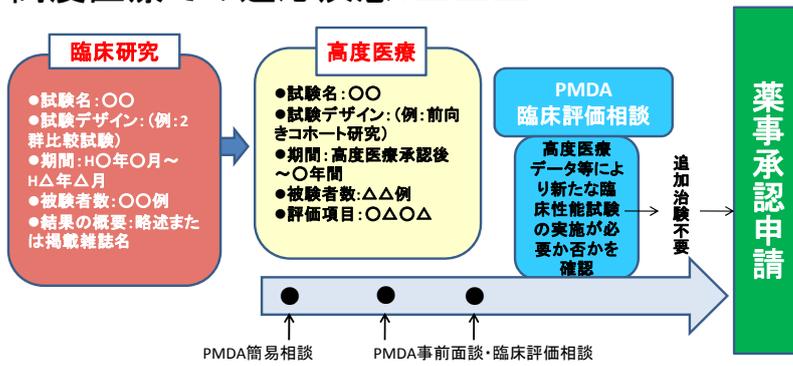


図2. 高度医療承認後のロードマップ

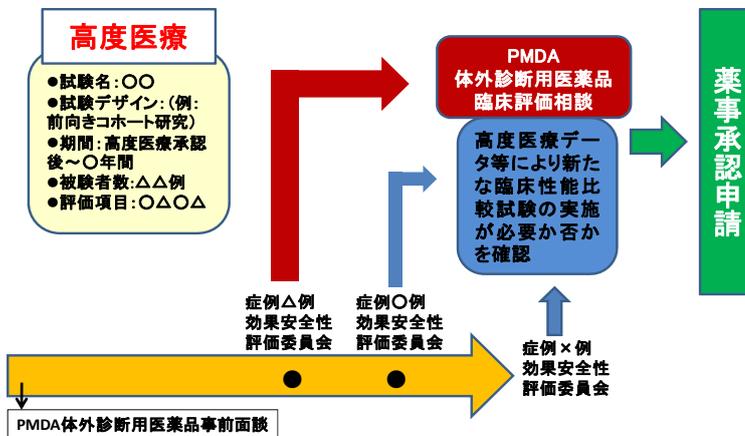
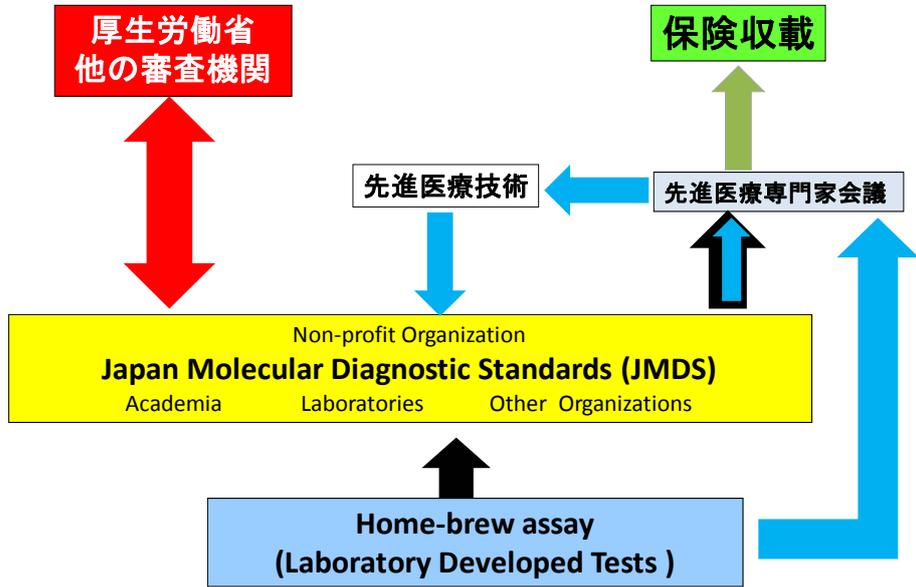


表1. 平成22年4月 診療報酬改定(遺伝子検査関連項目)
(薬事承認済診断薬は赤字で示した.)

D006-2	血液細胞核酸増幅同定検査(造血器腫瘍核酸増幅同定検査) 2,000点 PCR法、LCR法またはサザンプロット法による。(Major bcr-abl、PML-RARA など)
D006-3	Major bcr-abl mRNA核酸増幅検査 1,200点 DNAプローブ「FR」Amp-CML(TMA法)
D006-4	遺伝学的検査 4,000点 PCR法、DNAシーケンス法、FISH法またはサザンプロット法による。
D006-6	免疫関連遺伝子再構成 2,400点 PCR法、LCR法またはサザンプロット法による
D006-7	WT1 mRNA核酸増幅検査、サイトケラチン(CK)19mRNA、UDPグルクロン酸転移酵素遺伝子多型 2,000点 リアルタイムRT-PCR法(WT1 mRNA測定キット「オーツカ」)、OSNA法(LAMP法によるCK19 mRNAの増幅・定量)、インペーダー法(インペーダーUGT1A1アッセイ)
D004-2	悪性腫瘍組織検査 1 悪性腫瘍遺伝子検査 2,000点 固形腫瘍の腫瘍細胞を検体とし、PCR法、SSCP法、RFLP法等を用いる。肺癌及び大腸癌におけるEGFR遺伝子検査又はK-ras遺伝子検査、悪性骨軟部組織腫瘍におけるEWS-Fli1、TLS-CHOP、又はSYT-SSX遺伝子検査、消化管間葉系腫瘍におけるc-kit遺伝子検査、家族性非ポリポーシス大腸癌におけるマイクロサテライト不安定性検査、悪性黒色腫センチネルリンパ節生検に係る遺伝子検査

薬事承認された
診断薬のない遺
伝子検査が保険
収載されている

図3. 遺伝子検査の開発から臨床応用への新しいルート案



参考資料

(関連論文翻訳)

1. 予後バイオマーカーおよび予測バイオマーカーの評価に関する保存標本の使用

(原著)

Use of archived specimens in evaluation of prognostic and predictive biomarkers.

Simon RM, Paik S, Hayes DF. J Natl Cancer Inst. 2009 101:1446-52.

臨床的に使用可能な腫瘍バイオマーカーの開発は複雑である。われわれは、保存標本を使用したものをはじめとする腫瘍マーカー試験について、バイオマーカー試験のデザイン、実施、分析および評価の洗練されたシステムを提言する。これには保存標本を用いた試験をはじめとする腫瘍マーカー試験のための根拠レベルの階層的尺度が組み込まれている。予後バイオマーカーや予測バイオマーカーの医学的有用性を評価するには完全に前向きは無作為化臨床試験が標準法であるが、このような試験は費用がかかるため、間接的だが、より効率的な保存標本を用いた「前向き後ろ向き」臨床試験デザインについて検討する。われわれは特に、次の点を条件とする新たな指針を提案する。1) 十分な症例数の保存組織が十分量入手でき、十分な統計学的検出力があり、評価対象患者が元の試験の対象患者を明らかに代表している、前向き試験(予測因子については一般に無作為化試験)を対象にする、2) 検証されるテストは保存組織を用いた分析の妥当性が確認されていなければならない、3) バイオマーカーの評価計画は、保存組織を用いた分析の実施前に確定(記載)されており、目的、手法の確立した単一の classifier 指標の評価に焦点を絞ったものでなければならない、4) 保存標本を用いてえられた結果は、類似の別の試験の標本を用いてその妥当性を確認する必要がある、の4点である。

全身治療は癌患者の多くにおいては無効である。有効性が高いといわれる化学療法でも、無病生存率や全生存率を5~10パーセント改善するにすぎない。また、転移性癌に対する化学療法でも、治療した患者のごく一部にしか持続的な有効性が認められないことが多い。便益を享受する患者よりも多くの癌患者が高額の治療費、薬剤の毒性に暴露されている。このような過剰治療は生命を脅かす疾患に立ち向かっていることを考えれば理解できるが、現在よりも個別化された治療方針を導入することができれば、患者にも医療費にも大いに有益であろう。バイオテクノロジーと分子遺伝学が発達したが、癌患者の治療法の決定に情報を与える新たなマーカーの受け入れはあまり進んでいない。有効なバイオマーカーの導入が限定的であるのは、ひとつには腫瘍マーカー検査に対する費用還付の方が医療保険会社による治療よりも大幅に低いことによるが、ほかにマーカーの有用性に関する前向き試験が不足しており、予後マーカーおよび予測マーカーに関して報告された後ろ向き試験の多くが再現性および信頼性を欠くものであったからでもある(1, 2)。

いくつかの委員会や執筆者が任意のマーカーの評価や報告に用いることができる指針を提案してきた。例えば1996年、米国臨床腫瘍学会腫瘍マーカー指針委員会は腫瘍マーカーの臨床的有用性を明らかにするために用いることができる5段階の根拠レベル Level Of Evidence (LOEs)を提案した(3)。このLOE尺度は臨床で腫瘍マーカーの使用を推奨できるかどうかの決定や、腫瘍マーカーに関する臨床試験デザインおよび

実施の尺度として広く引用され用いられている(4, 5)。マーカー試験結果の報告基準はREMARK 基準と呼ばれ、いくつかの雑誌に掲載され、少なくとも数誌が投稿要項にREMARK を組み入れた(6, 7)。

本稿では、予後マーカー(治療とは無関係に予後を予測する)、治療効果の予測マーカー(最善の治療選択に関する予測)いずれでも腫瘍マーカー試験に関与する方法論的な問題の本質を検討する。また、予後バイオマーカーおよび予測バイオマーカーの臨床的妥当性や医学的有用性評価の信頼性を損なわず、保存標本を使用することができる条件を提案する。

腫瘍マーカーの有用性に関する前向き無作為化試験

新規医療介入マーカーの臨床的有用性を確認するための標準的試験は、前向き無作為化臨床試験である。予後または予測マーカーの評価について、同試験が提言されている(8-13)。候補となる予測バイオマーカーの医学的有用性は最近の状況では、新規薬剤を含む治療法を対照治療と比較する無作為化臨床試験において、事前に確定した分析計画を用い十分なサイズの患者サブグループを対象にし、マーカーの状態(陽性か陰性か)による新規治療法の有益性を評価することによって明らかにすることができる。

検査自体が研究的介入である前向き臨床試験を、予後または予測マーカーの妥当性確認試験と考えることが可能である。これらは、Simon, Wang によってマーカー戦略デザインと呼ばれたように、治療に当該検査を用いる群と用いない群に無作為に患者を割り付け、治療法が決定されるよう試験がデザインされる場合がある(14)。このような試験においては、無作為化によって対照群に割り付けられた患者に標準的予後因子と治療指針を用いた治療、被験群に割り付けられた患者には、標準的予後因子を併用する場合もあるが、その検査、またはマーカーを用いて治療を行う。検査は無作為化により被験群に割り付けられた患者に限って実施され、全転帰を無作為化群2群について比較評価する。転帰は、新規検査が対照群には用いられていないため、全体で比較しなければならない。多くの場合、これはデザインから探り出すことができる情報を大幅に限定することになる。標準治療が医師によって異なる場合は特に、結果が交絡し希釈されることがある。

マーカー戦略デザインは、無作為化に要する患者数に関して、極めて非効率的である。無作為化試験におけるサンプルサイズ要件は、特定の統計学的検出力で検出しようとする治療効果のサイズの二乗の逆数に比例することが多い。マーカー戦略デザインの場合に、無作為化により割り付けられた群に関わらず患者の多くが同じ治療を受けることになる。そのため、無作為化された2群の間の全治療効果が評価できるに過ぎず、効果量は一般にきわめて小さい。分析が検査陰性患者に対して標準治療をしないことが劣っていないことを示すものであればサンプルサイズの問題が大きくなり、サンプルサイズを巨大にしても結果が説得力のあるものになる可能性は低い。

これに代わるアプローチの場合には、全患者に前もってマーカーを検査する必要がある。標準治療に基づく治療割り付けとは異なり、検査に基づく治療割り付けを実施した患者のサブセットに焦点を絞って評価することができる。たとえば第II期の患者に対しては化学療法を実施するが第I期患者には実施しないことが標準治療であり、検査が病期に無関係に化学療法が有効な可能性が高い患者を特定することを目的とすると、検査陽性患者に化学療法を実施し、検査陰性患者には実施しないことになる。この場合、無作為化により割り付けられた患者は検査陽性の第I期患者及び検査陰性の第II期患者に限られる。このデザインでは、患者のこのようなサブセットを個別に対象として化学療法の有効性を評価することができる。しかしこのデザインでは標準的な予後変数の機能として標準治療が決まっていることを前提とする。

全患者をあらかじめ検査する方法は、ヨーロッパで実施されている The Microarray in Node-Negative Disease may Avoid Chemotherapy (MINDACT) 試験および北アメリカで実施されている the Trial Assigning Individualized Options for Treatment (Rx) (TAILORx) 試験という 2 件の最近の臨床試験で用いられている (15, 16)。両試験のデザインは複雑でいくらか異なるが、いずれもリンパ節転移がなくエストロゲン受容体陽性の女性乳癌患者で、所定の遺伝子発現に基づくリスクスコアによって再発のリスクが低いと予測された患者を対象群にして、標準治療である化学療法を実施しないことの医学的有用性を検討している。MINDACT 試験は 70 遺伝子シグナチャーを評価し、TAILORx 試験では 21 遺伝子シグナチャーを評価している。このデザインは無作為化マーカー戦略試験デザインより有効であるが、両試験とも数千例の患者参加が必要で、最初の結果が得られるまでに約 10 年を要する。TAILORx 試験および MINDACT 試験は実施に数百万ドルまたは数百万ユーロかかり、最新の技術革新速度をもってしても試験終了時には時代遅れになっている可能性がある。

決定的試験により特定の薬剤や薬剤クラス、化学療法など治療法一般における有益性が明らかにされた後に新規マーカーが明らかになることがある。われわれは、多くの場合、以前に実施された適切な臨床試験の保存検体を用い、厳密にデザインされた前向き試験の目的、第 1 種の過誤を制御し、統計学的検出力の保持することが可能である、と考える。実際、新たなマーカーが特定の薬剤の有益性を予測できるという予備的証拠が豊富ならば、結腸直腸癌で KRAS について実施されたように、薬剤を評価するために行われた無作為化臨床試験の保存検体を用いてマーカーを分析することができる (17, 18)。

適切な保存組織が入手でき、信頼性を損なわずに使うことができれば、有効な癌診断法の開発が容易になり、推進されて、患者に少なくない利益がもたらされる可能性がある。しかし、保存組織を用い信頼性のない検討をすると、患者にリスクを生じる。われわれはここで、この資源を信頼できる方法で使用するに関する要件を明らかにすべく努めた。保存標本を用いた腫瘍マーカー研究の質を厳密に分析できるように、以前発表された LOE 尺度を改訂することを提言する。

前向き試験対後ろ向き試験：解釈の問題

生物医学者や生物統計学者は「前向き」試験の方が「後ろ向き」試験よりも好ましいと教えられているが、前向きと後ろ向きの別は「実験的」と「観察的」の区別に混同されることが多い。われわれは腫瘍学の分野の予後バイオマーカーおよび予測バイオマーカーの試験に関しては、後ろ向きという用語は誤解を招く場合があると考えられる。

腫瘍疫学分野では、後ろ向き症例対照試験と前向きコホート試験、いずれも実験的より観察的な試験である。両試験とも無作為化によって要因暴露に割り付けるのではないため、要因暴露と疾患の間に観察された相関が、因果関係の根拠としては実験的試験で得られるほど強いものではない。疫学的試験の重要なことはその性質が非実験的であることにあり、前向きか後ろ向きかではない。

治療学では、後ろ向き分析の多くが無作為化ではなく患者因などに基づいて治療を選択しており、非実験的である。書面のプロトコールがなく、焦点を絞っていないことが多く、導き出された結論が偽陽性である可能性に対して対照を置かず数多くの患者サブセットや評価項目が比較されている。これに対し前向き無作為化臨床試験では、治療割り付けの内部対照を置き、転帰やエンドポイントなどのデータ収集を注意深く行い、データ分析前に焦点を絞った分析計画を作成する。

バイオマーカー試験の多くが、たまたま入手できてマーカーの検査ができたにすぎない便宜的な標本を用いており、対象例の適格性、検出力算出、マーカーカットポイント設定や分析計画作成は事前には実施されていない。このような試験では大きく偏った結論に至る可能性が高く、まさに非難をこめて「後ろ向き」と称されるに値する。しかし「後ろ向き」試験がかつて実施された前向き試験の保存標本を用いるようにデザインされた場合やマーカー試験の実施前に一定の条件が書面のプロトコールに前もって明記されている場合には、「前向き後ろ向き」試験と考えられるであろう。このような試験には、標本および分析がたまたま利用できた単純な便宜的試験より、マーカーの臨床的有用性の解明に対して重みがあるに違いない。しかし、同一の前向き試験から得た保存組織を用い、さまざまな候補バイオマーカーについて多数の試験を実施すると、特定のバイオマーカーに関する単一の完全に前向きな試験よりも偽陽性の結論に至る可能性が高くなる。そのため、多数の前向き後ろ向き試験で得られた特定のバイオマーカーについての結果を独立に確認することが重要である(下記参照)。

マーカーの医学的有用性を明らかにするための保存組織の使用

予後バイオマーカーおよび予測バイオマーカーの評価において保存標本を用いる場合は、Henry と Hayes が提案した腫瘍マーカーの臨床的重要性に関する3つの条件、を考慮することが有用である(2)。1)マーカーの特異的な背景や用途が明確である 2)マーカー「陽性」患者と「陰性」患者の間で、転帰または治療効果の強さにおいて、臨床医や患者が2人の別の患者に対する治療法が異なることを受け入れられるほど十分に差がなければならぬ 3)その強さの推定が信頼できるものでなければならぬ、という3点である。

これらの基準はマーカーの臨床的有意有用性を明らかにすることにおいて重要である。診断検査に応用されるので「妥当性確認」という用語が有用である。Hunter らは、遺伝子検査に関して妥当性確認を3つのタイプに区別した(19)。「検査の分析的妥当性の問題、当該遺伝子型を正確性と信頼性をもって計測する能力の確認、臨床的妥当性または検査の能力に関して、関連疾患を発見したり予測したりする能力、臨床応用された場合、検査の臨床的有用性、リスクと便益の均衡における能力であり、その検査を用いることの臨床的意義や医学的適応が明らかで、検査結果が示す転帰や治療効果が治療決定に影響を及ぼすほど十分に大きいというような、実用性確認が必要である。予後マーカーの後ろ向き試験の多くにみられる深刻な欠陥が、患者がマーカー使用の医学的適応を明らかにするために選択されたのではないことである。このような試験では、マーカーの臨床転帰との相関が明らかになる可能性はあるが、その医学的適応は明らかにできない。

マーカーの効果の強さを信頼性をもって見極めようとする、さらに次の3つの条件に分けることができるであろう。1)マーカー検査の技術的、分析的特徴は正確で強固であり、再現性がある、2)臨床試験のデザインおよび分析は、正確に目的とする臨床応用の有用性を検討するにふさわしく、十分である、3)結果は、2件以上の試験で互いに類似するが別々の患者集団を対象に、ほぼ同じ強さが推定されることを確認または検証しなければならないという3点である。ここに挙げた条件はいずれも、保存標本を用いた後ろ向き試験の多く、特に便宜的試験では大きな偏りの影響を受ける可能性があるものである。前向き後ろ向き試験であるとしても、この問題ひとつひとつに慎重に注意を払うことによって便宜的試験で得られる偏りや矛盾した結果を減少させることができ、ひいては有用な腫瘍マーカーの臨床応用を加速できると考える。

分析に関する問題

「分析的検証」とは一般に、検査または分析値の再現性および確実性を指す。これには通常、組織の収集、処理、保存、調製などの前分析因子、試薬選択やインキュベーションの時間や条件、読み取り法(カットポイントの設定など)などの分析的因子、双方について、変動をできるだけ小さくすることがある(20, 21)。

保存組織を用いた臨床的バイオマーカーの評価が解釈可能なものであるためには、保存検体から得た検査結果が実際の臨床の場に発生することを反映していなければならない。真の臨床標本と保存組織にどれほどの差がありうるかを示す例を次に挙げる。

- 1) 前分析的問題 過去に採取された検体が、特に手元に保存している場合には、現在とは異なる方法で扱われている可能性がある。異なる例として、採取前診断生検(遺伝子の発現や組織処理に影響する可能性がある)、検体を採取してから処理までの時間(固定、凍結など)、固定や凍結の方法、標本の保存(温度、室内空気への暴露、ブロック保存かスライド保存かなど)および何回凍結と解凍を繰り返したかなどが挙げられる。
- 2) 分析的問題 腫瘍マーカー試験が実地臨床を変えるためには、検査自体が臨床的に使用できなければならない。実地臨床での治療を変えようとする試験において、担当者は、正確で再現性があることが明らかにされている試薬、条件およびカットポイントの使用、を注意深く前もって計画しなければならない。このような検討事項には、試薬供給源、濃度、インキュベーション時間などさまざまな要素がある。試験担当者は、その検査について、保存標本と臨床検体との間の検査結果の一致性を統計学的に示す必要がある。検体が検査のために組織マイクロアレイに調製されたか切片全体として調製されたか、抗体除去の実施の有無や実施方法などが例に挙げられる。

不完全な分析妥当性確認または前分析妥当性確認の結果惹起される偏りに関する注意点として、保存標本を用いたマーカー試験では治療および患者転帰などの臨床データを全部、伏せた状況で分析する必要がある、ことがあげられる。

臨床試験デザイン

必要条件に記載したように、試験担当者は検査の用途について明確な考えを持っている必要がある。一般に、予後因子は治療を加える必要があるか否かを決定する際に有用で、予測因子はある特定のタイプの治療が有効であるか否かを予測判断する際に有用である。予後マーカーの医学的有用性を明らかにするには無作為化試験が必須であるとは限らない。たとえば、再発スコアが低い集団を対象に極めて良好な予後の転帰を確認すべくデザインされた TAILORx 臨床試験では、再発のリスクが低い患者には化学療法を実施しないという前向き単群試験の結果が用いられた。分析前に因子がよく調整され、現状に則した医療で、典型的な手法で臨床データが収集されれば、十分大きな無治療患者集団から得た保存組織を用いて腫瘍マーカーサブグループごとの再発を予測し、マーカーの臨床的有用性を判断することは許容されるであろう。

ある治療法の単群第2相臨床試験から得られた抗腫瘍効果に関するデータは、治療効果予測に関するバイオマーカーの臨床的妥当性を明らかにするために用いることができる。しかし、予測マーカーの医学的有用性を明らかにするには通常、生存または無進行生存を主要評価項目とする大規模無作為試験が必要とされる。

LOEs の改訂の提案

本来の米国臨床腫瘍学界 LOE 尺度では、「後ろ向き試験」というものは LOE II 以下とされていた(3)。われわれはここに、バイオマーカーの臨床的有用性を分析するため

に用いることができる試験のタイプにより正確な定義を提供し、保存標本を用いた後ろ向き試験によってレベル I の根拠が得られるように LOE 尺度を改訂することを提案する。バイオマーカーの医学的有用性に関する LOE は、患者や標本、分析評価および統計学的分析計画などの重要な因子に関連している(表 1, 2)。

特定の状況でのバイオマーカーの臨床的有用性は、科学的には前向き無作為化臨床試験によって検討するのが最も好ましい(表 1、カテゴリー A)。あらかじめ作成されたプロトコールに従って前向きに患者を試験に登録し、治療し、追跡する。その試験は腫瘍マーカーの問題に対して特異的に前向きに十分な力を持ち、標本はリアルタイムでマーカーのために採取し、処理し、分析する。しかし上記のような「マーカー戦略デザイン」には重大な限界があるため、無作為化試験は通常それを用いることはない。カテゴリー A の前向き試験から得られた結果をまた別の試験でさらに確認することはいつでも歓迎されるが、このような試験からの説得力ある結果は決定的と考えられ、他に妥当性評価試験が必要とされることはない。この方法は、最初の全米臨床腫瘍学会の LOE 尺度に LOE I として含まれており、やはり標準法である。

改訂 LOE 尺度では、レベル 1 のデータを得るための第 2 の戦略が可能である。治療上の問題、あるいは他のマーカーに関する、前向き試験から入手した保存標本を用いて、腫瘍マーカー試験を実施し、現在のマーカーの問題に対応することである(表 1、カテゴリー B)。予後良好であるために追加治療を必要としない患者を特定するための予後マーカーを評価する場合、臨床試験は無作為化を必要としないことがある。たとえば TAILORx 試験では、再発スコアの低い群には内分泌治療に限って実施し、再発のリスクが 21 遺伝子再発スコアによる予測と同じくらい低いか検討している。効果予測マーカーの評価では、前向き試験は、通常、その治療をしかるべき対照治療群と比較した無作為化試験でなければならない。試験デザイン A では患者を前向きに登録し、治療、追跡、標本を前向きに採取し、処理し、一般的標準法を用いて保存する。腫瘍マーカーの問題は試験実施中または終了後に明らかにされるであろうが、腫瘍マーカー仮説の特定は完全に試験外の結果に基づくものである。試験から保存された組織は、完全に特定された統計学的解析計画を用い、新しいマーカーの評価に焦点を絞った新たなプロトコールが作成されるまで、解析に用いてはならない。試験開始前に、検査は、保存組織を用いることに関して、分析的にも前分析的にも妥当性が確認されている必要があり、臨床データを伏せた状況で、解析されなければならない。このような臨床試験は治療上の課題についてデザインされているため、マーカーの関係で治療法の統計学的有意性を証明するには検出力が不足することが多い(22)。しかしそれは、予測バイオマーカーに関して予期されるように、「検査陽性」患者に対する大きな治療効果を信頼性をもって明らかにするには十分なサイズであろう。しかし十分に留意をしても、このような試験の結果は完全な前向き方法によって得られる結果よりも偶然の産物である可能性の方が高い。

試験対象患者集団が一般患者集団を代表するという保証はないが、使用できる保存標本が試験対象患者全体を代表することが望ましい。普遍的に適用できる最低必要条件はない。われわれは相関性試験には全対象患者の少なくとも 2 / 3 が含まれるか、対象患者は選択の偏りを防ぐ方法で抽出する必要があると考える。たとえば研究者が資源利用をなるべく少なくしようとした場合や複数の試験内標本セットを検査および妥当性確認に使用しようとした場合、その集団全体の既知の重要な予後因子または予測因子を反映する試験の標本を抽出するために数学的無作為化を用いるだろう(5)。

カテゴリーBの試験を治療変更に必要なものとするためには、その結果が同一ではないとしてもほぼ同じ方法でデザインされ、実施され、分析されたもうひとつ別の試験の保存標本に基づく第二のカテゴリーB試験から得られた標本を用いて確認されなければならない。この2つの試験の結果は臨床治療の変更と同じように説得力がなければならない。さらに、この妥当性確認試験は同一のマーカを明確に特定する同じ検査法またはほぼ同じ検査法を用いて実施する必要がある。たとえば、遺伝子異常に対する直接シーケンシングやタンパク発現を明らかにする免疫組織学的検査、機能分析など、p53について検査する方法が研究者によって異なっている。このような分析評価はp53の大幅に異なる指標であるため、利用できるデータは解釈がきわめて難しい(5)。妥当性確認試験というものはまた、同一の主要評価項目に対するものでなければならない。この評価項目は医学的有用性を反映するものである必要がある。

Hayesらは、前向き無作為化臨床試験で採取した保存標本1500近くを用いて、リンパ節転移がありエストロゲン受容体陽性ヒト上皮成長因子受容体2陰性の患者にはドキシソルピシン、シクロホスファミド併用療法を4サイクル実施した後、補助パクリタキセル両方を追加しても上乗せ効果が認められなかったことを報告した(23)。この結果は興味深いものであったが、別の類似したデザインの前向き無作為化臨床試験ではこの結果が確認できず(24)、補助パクリタキセル療法に関する患者選択の問題は未だ解決されていない(25)。このように、この問題はなお、表2のLOE IIであると考えられる。一方で、最近KRAS変異の存在とセツキシマブやマニツムマブなど上皮成長因子受容体に対するモノクローナル抗体が無効であることとの間に相関が認められたが(17, 18)、これは医学的有用性を明らかにするためカテゴリーBの保存検体を用いた成功例である。進行結腸直腸癌患者の治療において、複数の前向き無作為化試験により、この抗体単独または化学療法との併用投与がわずかにしる統計学的に有意に有効であることが示された(26)。これに先立って、LOE IIやIIIの試験により、セツキシマブおよびパニツムマブが癌に野生型KRASが認められる患者に限って活性を示すことが示唆された(27)。現在、このデータは、大規模前向き無作為化臨床試験で得られた保存検体を用いて実施した後向き試験によって妥当性が確認されたと考えられ、われわれの改訂尺度ではLOE Iに達するであろう(表1, 2)(28)。

カテゴリーCバイオマーカー試験では、前向き患者登録を利用、標準治療に従って治療し追跡する。標本は一般的な標準実施手順を用いて前向きに採取し、処理し、保存するが、患者登録終了後に分析を実施する(表1)。このような標本を用いた腫瘍マーカー試験には、前向きに十分に力を持たないことが多い。治療の割り付けや標本採取およびデータ収集が制御されていないため、このような方法は、一般に患者、標本および転帰をはじめとする臨床データの選択の偏りの影響を受けやすい。厳密に制御された登録の中にはこのような懸念が当てはまらないものがあるであろう。カテゴリーC試験の方がカテゴリーA、Bより認識されない偏りによる交絡が起きやすく、結果が偶然によるものである可能性が高い。その後2つ以上の試験でほぼ同じ結果が得られた場合(表2)、カテゴリーC試験がLOE IIと認められることがある。しかし、よほど説得力がある状況を除いて、カテゴリーC試験が治療を変更するに十分と考えられる可能性は低い。

カテゴリーD試験というのがこれまで報告された腫瘍マーカー分析で最も多いタイプである(表1)。不特定の目的で採取、処理し、さまざまな方法で保存してあり、たまたま解析に利用できた標本を用いた便宜的試験である。このタイプの試験で得られた結果は不安定であり、偶然の産物に過ぎない可能性が高い。

要旨

新たな医学的介入を臨床医療に取り入れるには、いかなるものでもレベル I の根拠を背景とする場合に限るのが理想的であり、その根拠が前向き無作為化臨床試験によって得られたものであることが理想的である。しかし、このような試験は必ずしも実際的ではない。腫瘍マーカーの場合、治療ガイドラインや他の診断法の利用可能性により新たな臨床試験を実施することがきわめて困難になることがある。試験の一部として、治療の標準であると考えられている治療を実施しない可能性があるからである。倫理的であると思われても、このような試験は通常実施に何年もの期間を要し、莫大な費用がかかる。新薬の開発には多くの場合、結腸直腸癌に対する EGFR 阻害薬の予測バイオマーカーとしての KRAS 変異のように、分析的に妥当性が確認されたコンパニオン診断検査がないか、しかるべき生物学的尺度が当該薬剤の中心的試験開始時に明らかにされていないことがある (17, 18, 28)。

このため、予後マーカーや効果予測マーカーの医学的有用性を明らかにするためには、質の高いデータセットから保存組織標本を用いて解析することが極めて重要なことがある。大規模前向き臨床試験から得た保存組織標本の使用が適していると考えられる。しかし、その評価が単に仮説を立てるよりも有用であるには、幾つかの条件を満たす必要がある。

- 1) 保存組織が利用できなければならない。バイオマーカー評価の対象患者が pivotal 試験の対象患者を十分に代表することは、適切な検出力のある解析を可能にする上において、極めて重要である。普遍的に適用できる最小必要条件を明示することはできないものの、対象となる試験の全患者の 3 分の 2 から得た標本が解析可能であることを提案する。
- 2) 検査の分析的妥当性について多くのデータがあり、保存標本から得られた結果がリアルタイムで採取した標本の解析から得られた結果と極めて類似していることを確認する必要がある。臨床データを伏せて分析を実施する必要がある。
- 3) バイオマーカー評価の分析計画は、バイオマーカー分析開始前に完成していなければならない。試験開始時から、バイオマーカー試験の分析計画および分析のための標本を抽出した試験のデザインがコンパニオン診断の評価にふさわしいものである必要がある。分析は、単一で完全に特定された診断分類に的を絞らなければならない。多遺伝子分類には、個々の構成因子、重量およびカットポイントを組み合わせた数学的フォームはあらかじめ特定しておく必要がある。分析は一般に、探索的であってはならず、偽陽性の結論を導くような分析はさけるべきである。
- 4) 結果は、少なくとも 1 件以上の同一検査法を用いたほぼ同じデザインの試験で妥当性を確認しなければならない。

臨床医は個々の患者に治療を選択するために改良された手段が必要とする。同じ原発部位の癌でも、分子的病態発生、臨床経過および治療の奏効性は多様である。治療の開発、評価および使用に対する現在のアプローチの結果、多くの患者が有効とは言えない薬剤で治療されている。癌のゲノミクスおよびバイオテクノロジーの発達によって、有効性の高い治療法の開発や治療に関する情報のための予後バイオマーカーや予測バイオマーカーの開発の機会が広がった。このような機会には、患者および治療費に莫大な便益をもたらす可能性がある。しかし、癌生物学が複雑で、薬剤のバイオマーカー開発が複雑になってきているため、腫瘍学がこれまでより予測的なものへと移行するのは大変困難である。場合によっては、まったくの前向き方式で予後バイオマーカーや予測バイオマーカーの医学的有用性を評価することが、倫理的または現実的に不可能なことがある。

癌患者は徹底的かつ信頼できる方法で評価された後、速やかに有益な予後判定検査や予測試験を確実に提供されることが不可欠である。われわれは本稿で、この分野の予後バイオマーカーおよび予測バイオマーカーの妥当性確認についての不確実性を明らかにしようとし、腫瘍学でバイオマーカーの医学的有用性を評価するために広く用いられている LOE 方式の最新化を提案した。われわれは、この改訂は妥当性確認試験を改善するために重要であり、場合によっては重要な診断法の推進に重要であると考ええる。

研究資金

一部 Fashion Footwear Charitable Foundation of New York/QVC Presents Shoes on Sale (D.F.H.) の援助を受けて実施した。

注

著者らはデータの分析および解釈、原稿執筆および出版社への投稿の決定について完全に責任を持つ。

著者らは、同僚および米国臨床腫瘍学会腫瘍マーカー指針委員会の委員が本研究に大いに貢献し、その議論がきわめて役立ったことに感謝申し上げたい。D.F. Hayes が研究を報告し、アストラゼネカ、グラクソスミスクライン、ファイザー、ノバルティス、Ayerst-Wyeth、ベリデックス(ジョンソン アンド ジョンソン)、サノフィアベンティス、Predictive Biocsciences、Curalogie SAB、Incyte Corporation、DNA repair および Compendia Bioscience の協力を得た。

本稿は 2008 年 12 月 5 日受領、2009 年 8 月 4 日に修正され、2009 年 8 月 27 日に受理された。

表 1. 根拠のレベルの決定を構成する腫瘍マーカー試験の要素

カテゴリー	A. 前向き	B. 保存検体を用いた前向き	C. 前向き・観察的	D. 後ろ向き・観察的
臨床試験	腫瘍マーカーを目的とする PCT デザイン	腫瘍マーカーを目的にデザインされていない前向き試験であるが腫瘍マーカーの有用性に適応できるデザイン 予測マーカーへの適応には PRCT が必要	前向き観察的登録、治療および追跡は制御されていない	試験に前向きの要素がない
患者および患者データ	PCT で前向きに登録、治療、追跡を実施	前向きに登録し治療し、臨床試験で追跡。特に予測に関する有用性を考慮した	前向きに登録するが治療と追跡は標準治療	治療も追跡も前向きではない。後ろ向きカルテ調査によって患者データを収集。

		場合は、当該治療に関するPRCT		
標本採取、処理および保存	リアルタイムで特定のマーカについて標本を採取、処理および保存する	標本は一般的SOPsを用いて前向きに収集、処理および保存する。試験完了後に解析	標本は一般的なSOPsを用いて前向きに収集、処理および保存する。試験完了後に解析	前向きSOPsなしに標本を採取、処理し保存する。
統計学的デザインおよび解析	腫瘍マーカに関する問題に対して十分に力を持つ試験	治療上の問題について十分に力を持つ試験で、腫瘍マーカの問題には不十分な力を持つ解析前に作成されたマーカに関する問題の分析計画に焦点を絞っている	前向きに十分な力がまったくない。後ろ向き試験デザインで試験標本の抽出による交絡の影響を受ける。解析前に作成されたマーカに関する分析計画に焦点を絞っている	前向きに十分な力がまったくない。後ろ向き試験デザインで試験標本の抽出による交絡の影響を受ける。解析前に作成されたマーカに関する焦点を絞った解析計画なし
妥当性確認	結果が偶然の産物である可能性は低い実施が好ましいが、妥当性確認は必要ではない。	結果が偶然である可能性はAより高いがCより低い。1件以上の妥当性確認試験を要する	結果が偶然である可能性がきわめて高い。続いて妥当性試験を実施する必要がある。	結果が偶然である可能性がきわめて高い。続いて妥当性試験を実施する必要がある。

*PCT=前向き対照試験、PRCT=前向き無作為化対照試験、SOPs=標準的手順

表2. 腫瘍マーカ試験の要素を用いた改訂版根拠レベル決定因子

根拠のレベル	表1のカテゴリー	利用可能な妥当性試験
I	A	不要
I	B	結果が一致した場合1件以上
II	B	不要または矛盾した結果
II	C	結果が一致した場合2件以上
III	C	不要または結果が一致した場合1件または結果不一致
IV-V	D	NA †

*Hayes らによって提案された原案から改訂した根拠レベル(LOEs)

† NA=LOE IV および V 試験は医学的有用性の確認に対して十分であることはないので適用できない

2. 妥当性確認試験の要点

(原著)

What should physicians look for in evaluating prognostic gene-expression signatures? Subramanian J, Simon R. *Nat Rev Clin Oncol*. 2010 7:327-34.

ASCO 腫瘍マーカーガイドライン委員会はマーカーの評価に用いることができる5段階の Level of Evidence (LOE) を提言した²⁵。レベルは I から V にわたり、レベル I (LOE I) は決定的な証拠としている。新規シグネチャーの臨床的有用性を評価するためのレベル I の根拠は、そのシグネチャーを検討するためにデザインされた前向き臨床試験の説得力のある結果から得ることができる。本稿では、予後予測のための試験デザインを考察する。予測シグネチャーに関する試験デザインについては、レビューが二つ発表されているので参照されたい^{26,27}。分析は妥当性検証試験開始前に、標準化され分析的妥当性が確認されている必要がある。また、シグネチャーは完全に特定されていなければならない、探索的解析で評価すべき多数のシグネチャーのひとつであってはならない。

シグネチャーの妥当性確認のための前向き試験

マーカー戦略デザインは、予後シグネチャーの医学的有用性に関し、妥当性を前向きに直接確認することを目的とするが、効率は悪いことが多い²⁸。これらの試験デザインでは、シグネチャーを使用して検査する群としない群、それぞれに患者を無作為に割りつける。検査しない患者には標準的な予後因子、治療指針を用いて治療を決定する。検査する群に割りつけられた患者には標準的予後因子にシグネチャーを併用して治療方針を決定する。試験は無作為化した2群の全般的転帰を比較することによって評価する。

しかし、マーカー戦略デザインは、有効性が低いことが多い。患者の多くが無作為化によって割りつけられた群に無関係に同じ治療を受ける可能性があるため、シグネチャーを用いることで治療法が変わる患者のサブセットについては、転帰の差を検出するに十分な統計学的な検出力をもたせるには、膨大な症例数が必要になることがある²⁶。マーカー戦略デザインでは標準治療群のシグネチャーを測定しないため、このようなサブセットの転帰は直接的に評価することができない。また、シグネチャーによって低リスクと予測された患者に対して化学療法を実施しないことが化学療法を追加することに劣らないという分析結果であると、この有効性の低さは増幅され、膨大な症例数を対象にしても結果は説得力のあるものにならない可能性が高い。マーカー戦略デザインは、対照群にはシグネチャーを測定しないためにサブセット分析が不可能であることから、マーカーの治療上の意味が複雑な場合に情報量が乏しいことがある。

マーカー戦略デザインの欠点を補うために、修正マーカー戦略デザインを用いることができる。この修正デザインでは、シグネチャーは全例に測定し、シグネチャーに基づいた治療法が標準と異なる場合に限り無作為化する。化学療法を実施しないことが IA 期患者の標準治療である NSCLC を例にとり解説すると、IA 期 NSCLC に対する予後シグネチャーの妥当性を確認するには、次のように試験をデザインすることができる。シグネチャーは IA 期 NSCLC 患者全例について測定、シグネチャーによって低リスクと予測された患者を試験から除外する。高リスクと予測された患者を無作為化によって化学療法を実施する群、実施しない群とに割りつけ、転帰を比較する。このデザインは、標準治療が標準予後変数として規定されていることが前提となる。あらかじめ

め全例に検査を実施するこの方法は、乳癌患者を対象に *Oncotype DX*[®]再発スコアを評価するために実施されている Trial Assigning Individualized Options for Treatment (Rx) (TAILORx) 試験で用いられている²⁹。

TAILORx 試験の主目的は、確立された治療指針では補助化学療法を実施することになる女性患者のうち *Oncotype DX*[®]再発スコアが 11-25 と中等度の患者を対象に、ホルモン療法単独治療が補助化学療法とホルモン療法の併用に劣らないかを明らかにすることである。副次的目的には、低再発率群(再発スコア<11)を対象に長期予後およびホルモン療法単独の有効性を明らかにすることがある。*Oncotype DX*[®]の初期妥当性確認試験に用いられた再発スコアの範囲は、低リスク(<18)、中リスク(18-30)および高リスク(≥31)であった。中リスク群と高リスク群に対する治療が不十分になる可能性をできるだけ低くするため、TAILORx では各群を決定するスコア幅は低くした(低リスク[<11]、中リスク[11-25]、高リスク[≥25])²。新たなカットオフは試験手順にあらかじめ設定し、試験中に探索的に変更することはない。この試験のデザインは修正マーカー戦略デザインに基づいており、対象となる患者全例に前もってシグネチャーの検査を実施する。低リスクと予測された患者はホルモン療法単独治療に、高リスクと予測された患者は化学療法とホルモン療法、併用療法に割りつける。中リスクと予測される患者は、ホルモン療法単独または化学療法とホルモン療法の併用に無作為に割りつけられる。再発スコアの予測が正確であれば、低リスク群の再発率は低く、化学療法の追加によってえられる利益は極めて小さいはずである。中リスク群については、5年無病生存率が化学療法群 90%、ホルモン療法単独群は 87.0%以下と 3%の絶対的減少が認められれば、受け入れがたいと考えられる²⁹。

修正マーカー戦略デザインを用いた第2の臨床試験が MINDACT (Microarray in Node-Negative Disease May Avoid Chemotherapy) 試験で、リンパ節転移陰性乳癌患者に対する補助治療の指針となる 70 遺伝子予後シグネチャー MammaPrint[®] (アジェンディア、アムステルダム、オランダ)を前向きに評価するためにデザインされたものである³⁰。

保存標本を使用したシグネチャーの妥当性確認

新規シグネチャーの臨床的有用性を検証する上で確かな判断基準は、前向き臨床試験で得られる説得力のある結果に基づくものであるが、このような試験では標準的治療を実施しない場合があるため、必ずしも実施可能ではない。例えば、標準治療である化学療法を実施しなくてもよい低リスクの II 期乳癌患者を特定することを目的とすると、このような状況が持ち上がる。また前向き臨床試験では患者数千例が必要とすることがあり、実施に何年もかかり、きわめて高コストであることがある。TAILORx 試験は 2006 年に開始し、2013 年にしか報告されないと予測されている²⁹。試験終了時には検査が時代遅れになっている可能性がある。

大規模でしかるべくデザインされた数多くの臨床試験の患者から一様に採取した標本が使用できる場合、前向き(分析前)に特定した単一のシグネチャーについて、分析的に妥当性が確認された分析に関して、探索的ではなく焦点を絞ってこのような試験を分析すると、保存標本を用いても高レベルの臨床的有用性の根拠が得られることができ、シグネチャーを臨床の現場に導入する速度を速めることができる可能性がある。最近、LOE 尺度の改訂が提案され、保存標本を用いた妥当性確認試験のデザイン、必要条件などが設定された²⁴。偏りを防ぐため、これらの試験は Box2 にまとめた厳密な条件の下に実施する必要がある。

保存標本を用いた妥当性確認試験の結果は、最初の試験とほぼ同じか、概ね同じようにデザインされ、実施され、分析された別の試験からの標本を用いて確認する必要がある。妥当性確認試験は同一または類似した分析を用いて実施する必要がある、同じ評価項目を検討し、評価項目が医学的有用性を反映していなければならない。臨床治療の変更に足るレベルの証拠を得るには、これらの妥当性試験の結果は一致している必要がある、同等に説得力のある結果でなければならない。結果の一致が無い場合は、シグネチャーについて最高でもレベル II の証拠と考えるにとどまる。

最近、*Oncotype DX*[®]再発スコアをリンパ節転移のある ER 陽性乳癌患者のうち化学療法がほとんど有益でない患者を特定するために用いることができるかを明らかにするために、SWOG-8814 試験で得た保存標本を用いた妥当性試験が実施された³¹。この試験の結果は、リンパ節転移の数とは無関係に、高再発スコア女性乳癌患者においてはタモキシフェンにアントラサイクリンを主とする化学療法を追加すると多いに有益であるのに対し、低再発スコアの女性患者には無益か、わずかしかな有益ではないことを示していた。リンパ節転移のある乳癌患者の一部に標準的化学療法を実施しないことになるので、この場合、妥当性確認試験のために前向き無作為化臨床試験を実施することは難しかったと考えられる。

NSCLC の予後シグネチャー

NSCLC の現在の治療決定指針は TNM 病期分類および他のいくつかの臨床組織病理学的パラメーターに基づいている^{14,32}。これまで概説してきたとおり、次のような場合に NSCLC に遺伝子発現に基づく新規予後シグネチャーが臨床的に有用であると考えられるであろう。第一に、高リスクで完全切除術後の IA 期患者のうち補助化学療法が有益な患者、次に IB 期または II 期の患者のうち化学療法を実施しなくても再発リスクが低い患者を標準的危険因子よりも有効に洗い出すことができる場合である。

近年われわれは、NSCLC の予後遺伝子発現シグネチャーの開発や妥当性に関する試験を評価する調査を実施した¹⁶。この調査の目的は、当該シグネチャーの臨床的有用性が標準的治療指針によるものに勝るといふ証拠を得ることである。この試験は、これまで考察したさまざまな点を反映する基準に基づいて評価された。評価に用いた 3 つの主要な基準は、プロトコルの適切性、予後予測モデルの統計学的妥当性、シグネチャーの医学的有用性の確認である。

調査した研究はいずれも新規予後シグネチャーの開発試験であった。調査により、試験の多くが患者選択にも症例数の設定にも十分重要留意していないことが明らかになった。試験の 50%超が予後シグネチャーの開発に用いたのと同じデータを予測精度の評価に用い、偏った再代入結果を示していた。さらに試験の大部分が、シグネチャーの独立した妥当性確認試験を実施できるような特定したモデルを報告していなかった。また、試験の大半が予後シグネチャー開発に使用していないデータに基いた妥当性確認結果を報告していたが、このような妥当性確認では IA、IB、II 期の NSCLC に対する治療決定の改善に関する新規シグネチャーの有用性を十分示すことができず、既知をしのぐ遺伝子発現シグネチャーの有用性を検証できなかった。このことは大変重要である。

これまで、NSCLC の予後シグネチャーはいずれも標準治療指針に組み込むことができるほどの医学的有用性を示していない。最近、I 期 NSCLC 患者の治療について補助化学療法と経過観察を比較する前向き臨床試験 CALGB-30506 が開始された。この試験で、遺伝子シグネチャーは I 期患者の中から補助化学療法が有効な患者を選択する際に有用か、否かを評価するために、層別化因子として用いられている^{33,34}。

Oncotype DX[®]から得られる教訓

Oncotype DX[®]は、21 遺伝子を分析する診断検査で、ホルマリン固定パラフィン包埋の乳癌組織を用いる。この検査の結果は再発スコアとして表される。開発後、この検査は再現性およびパラフィン保存組織に対する頑健性について分析的妥当性が確認され、NSABP B-14 試験のタモキシフェン群から得られた独立した ER 陽性でリンパ節転移のない患者群でにおいて臨床的妥当性が確認された^{35,36}。

Oncotype DX[®]は現在、TAILORx 試験で前向きに評価されている。この試験の評価で、留意すべき点は以下の通りである。Oncotype DX[®]は検査が分析的妥当性を確認されている、臨床病理研究室で通常通り処理されたホルマリン固定パラフィン包埋組織標本にかぎる、臨床的に重要な患者群であるホルモン受容体陽性でリンパ節転移がなく、タモキシフェン治療を受けている患者に対して開発され、妥当性確認が行われていることの3点である。初期妥当性確認試験の結果により、低再発スコアの患者の再発リスクは十分に低く、治療決定に使用できることが示されたものである^{2,29}。

乳癌では数多くの予後シグネチャーが開発され、4つのシグネチャーが市販されている {MammaPrint[®], Oncotype DX[®], Theros Breast Cancer IndexSM (Bio-Theranostics), MapQuant DxTM (Ipsogen)}³⁷。Fanら³⁸は乳癌の遺伝子シグネチャーの比較試験において、シグネチャーは遺伝子同一性に関して重複はしていなかったが、個々の患者の転帰予測においては高率に一致していたことを明らかにした。Wirapatiら³⁹はいくつかの公表された試験結果を用いて乳癌2,833例を対象にメタアナリシスを実施した。この試験の結果は、比較した9つの予後シグネチャーいずれも、この大規模なデータセットについてほぼ同じ予後予測力を示し、予後予測能力が大部分分化関連遺伝子によるものであることを示した。

乳癌に対して市販されているシグネチャーは予後と関連していることが示されており、解析研究室はCLIA(Clinical Laboratory Improvement Amendments)認定を受けている³⁵。MammaPrint[®]は、米国食品医薬品局(US FDA)により、クラス2、51万ドル製品として承認された⁴⁰。しかし、ASCOは、Oncotype DX[®]以外のシグネチャーは明確に特定された用途について臨床的有用性を確認するためにさらに検討を重ねる必要があるとしており、リンパ節転移がない、エストロゲン受容体陽性でタモキシフェン投与を受ける乳癌患者の再発リスクの予測にはOncotype DX[®]に限って推奨している⁴¹。MammaPrint[®]が急速凍結組織を必要とするのに対し、Oncotype DX[®]検査はホルマリン固定パラフィン包埋組織標本のみを必要とし、標準的な臨床の場で実用的である。Oncotype DX[®]の開発の成功から学ぶべき点はいかなる遺伝子発現シグネチャーの試験にも重要であり、Box3に詳細に記載する。

結語

遺伝子発現シグネチャーがあれば、個々の癌患者に対する治療決定を確実に最適なものにすることができる。しかし、癌生物学の複雑さや高次元データ解析の複雑さおよび予後シグネチャーの開発および妥当性確認の焦点が絞られていないことが予測腫瘍学への移行にあたっての厄介な問題である。予後シグネチャーは初めから、治療上重要な用途に焦点を絞って開発し、妥当性を確認する必要がある。開発試験で遺伝子シグネチャーの妥当性確認に成功しても、臨床応用には十分とは言えない。前向き無作為化臨床試験からの証拠がある場合に限り、新規予後シグネチャーを臨床応用するのが理想である。現在、乳癌を対象にこのような臨床試験が2件実施されている。Oncotype DX[®]を評価するTAILORxとMammaPrint[®]を評価するMINDACTである^{29,30}。しかし、このような前向き試験は常に可能とは限らず、(多くの場合)重要な予後的決定手段を臨床で使用できるようにするために最も有効な方法ではないことがある。

厳密な条件下に実施すれば、適切な臨床試験から得られた保存標本を用いた妥当性確認試験によって临床上の有用性についてレベルの高い証拠を得ることができる²⁴。事実、TAILORx、MINDACTの主要な目的のひとつは大規模組織バンクを作ることである^{29,30}。このような組織バンクは、将来の乳癌の予後研究にとって、究極的には将来の臨床治療にとって高品質なデータ源となるであろう。われわれはこの概説に記載した指針が、遺伝子発現シグネチャーの報告を評価する際の一助となり、研究者が遺伝子発現シグネチャーについて新たな開発試験や妥当性確認試験を計画する際の一助となることを願ってやまない。

Box2 保存標本を用いた試験が満たさなければならない条件

- ・妥当性試験の基礎となる保存標本を採取した前向き臨床試験は、もし予後シグネチャーを評価するために前向きにデザインするのであれば使用したであろうものと同じ登録基準で同じ構造であることが不可欠である。しかし、試験対象患者数は多くの場合少ない。
- ・妥当性試験は特定された単一シグネチャーを評価しなければならない。シグネチャーのパラメーターがいずれかひとつでも妥当性確認中に最適化されたら、最適化の補正のため特別な分析法を使用しない限りは、その試験はもはや妥当性確認試験にふさわしくない。
- ・解析は保存組織について使用するにあたり分析的に妥当性が確認されている必要がある。組織採取や処理、保存および調製などの前分析因子および試薬選択、インキュベーション時間および条件、読み出し方法などの分析因子の双方による変動をできるだけ小さくする必要がある。
- ・評価に用いた各臨床試験の対象患者数が多く、その大部分からの標本入手が必要である。
- ・保存組織は、新規シグネチャーの評価に焦点を絞って新たに手順書と完全に特定された統計学的分析計画が作成されるまで分析してはならない。
- ・分析は臨床データを伏せて実施しなければならない。
- ・ここ記載した様式で保存標本の評価をするためには、それぞれに十分な臨床試験が2件以上使用できなければならない。

Box 3 *Oncotype* DX[®]検査開発の要点

重要な臨床的用途に焦点を絞る

ER陽性でリンパ節転移がなく、ホルモン療法を実施している乳癌患者のうち、化学療法を実施する必要がないほど再発リスクが低い患者を選定すること。

検査の分析的妥当性の確認

Oncotype DX[®]検査の分析特性は検証され、*Oncotype* DX検査で定義される検査性能諸元は100単位尺度の内2再発単位の標準偏差以内で個々の患者の定量的再発スコア値を報告できると結論された。さらに、試験デザインの分析によって、機器や検査者、試薬および日々の基礎変動による分析の不正確さが小さいことが明らかにされた⁴⁰。

シグネチャーの開発に用いられた患者群ではなく、大規模で治療上ふさわしい個別の患者群を対象とした妥当性確認試験

21 遺伝子シグネチャーは、NSABP B-14試験のタモキシフェン治療群から得たER陽性でリンパ節転移がない患者668例について妥当性が確認された。この妥当性確認試験

の結果、カプランマイヤー法で10年間無遠隔再発患者の割合が低リスク群で93.2%、高リスク群で69.5%であったことを示し($P < 0.001$)、シグネチャーの予後予測力を示した⁴¹。低リスク群(93.2%)の再発率は十分低く、シグネチャーが治療上適切であると考えられる。

便宜性

この検査にはホルマリン固定パラフィン包埋組織標本だけが必要であり、標準的臨床医療に実用的である。

調査基準

PubMedデータベースを用いて2010年2月28日以前に英語で発表された論文を検索して情報を得た。検索語には「開発試験」「妥当性確認試験」に「予後遺伝子発現シグネチャー」、「Oncotype DX」、「MammaPrint」、「TAILORx」、「MINDACT」および「乳癌予後シグネチャー」があった。さらに関連する情報を得るために、必要に応じて主要な論文の参考文献を参照した。NSCLCシグネチャーを確認するための調査基準は参考文献16に概説されている。NSCLCシグネチャーに関してこれ以上の検索は実施しなかった。

3. バイオマーカーと代替エンドポイント—統計学的妥当性確認の課題

(原著)

Biomarkers and surrogate end points—the challenge of statistical validation. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Nat Rev Clin Oncol. 2010 7:309–17.

要旨

バイオマーカーおよび代替エンドポイントは、臨床腫瘍学の分野において利用可能性はあるが、統計学的妥当性確認が大きな課題であり、しっかりと確認されたバイオマーカーは少ない。治療と無関係に予後を予測する予後バイオマーカーに関するデータは、小規模な後ろ向き試験によって得ることが可能であるが、多施設で妥当性を確認することはより難しいことが明らかにされている。特定の治療に対する反応性を予測するバイオマーカーの方が妥当性確認に膨大なデータを必要とし、大規模無作為化試験やメタアナリシス等が必要である。代替エンドポイントは妥当性確認が一層難しく、代替エンドポイントが治療と無関係に真のエンドポイントを予測すること、代替エンドポイントに関する治療効果が真のエンドポイントに関する効果を確実に予測することの双方のデータを示す必要である。われわれは、予後バイオマーカー、予後バイオマーカーおよび代替エンドポイントの特性を考察し、妥当性確認に必要な統計学的方法およびデザインについてレビューを行った。妥当性確認のための統計学的必要条件を満たすことができない場合は、エンドポイントの生物学的妥当性あるいは代替指標が裏付けとなることがある。確かな統計学的な妥当性が確認されない場合のバイオマーカーや代替エンドポイントの実際的な評価、採用の過程、標準等については未だ見解は一致していない。

緒言

癌の臨床研究において、バイオマーカーや代替エンドポイントの重要性は増している。バイオマーカーが予後評価や、特定の治療に対する個々の患者の反応性の予測に使用することができるのに対し、代替エンドポイントは生存率などの古典的なエンドポイントより迅速に、ときにより正確に、新たな治療の有効性を評価することを可能にするものである。しかし、バイオマーカーや代替エンドポイントの採用については克服しなければならない課題は数多くあり¹、発見、検定、分析の適格性確認から統計学的妥当性確認、臨床試験における成功、さらに最終的には臨床での日常的使用に至るまで幅広い。われわれはここで、この過程のうち最も難しい段階の一つである統計学的妥当性の問題に焦点を当てる。

定義

この概説で用いたバイオマーカーおよびエンドポイントの定義は表1に示す。バイオマーカー定義ワーキンググループによって採用された定義によれば²、バイオマーカーとは“a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” 「正常の生物学的過程、病学的過程または治療的介入に対する薬理学的反応の指標として客観的に測定でき評価できる特徴」を言う。われわれはこの広いカテゴリーのなかで、創薬初期に用いられる薬物力学的バイオマーカーや薬物動態的バイオマーカーではなく、将来の状態

病態を予測するバイオマーカー、いわゆる予後バイオマーカーおよび予測バイオマーカーに焦点を絞って考察する。予後バイオマーカーとは疾患が進むと思われる経過を、治療とは無関係に予測するものである。たとえば固形癌治療では、リンパ節転移の有無に関わらず治療によって患者の生存期間を延ばすことができるが、リンパ節転移があると、転帰が不良であることが予測される。これに対し、予測バイオマーカーというものは治療に対する反応性を予測する。たとえばホルモン受容体の有無によって乳癌の内分泌治療に対する反応性を予測することができる。また、バイオマーカーのなかには乳癌のホルモン受容体のように予後的でも予測的でもあるものがある。

バイオマーカーは患者の感覚や機能、生存に関する情報を把握する臨床エンドポイントと対比されることがある³。代替エンドポイントは、バイオマーカーに基づくことがあるが、実験的治療の効果について、臨床的エンドポイントを迅速で感受性の高いエンドポイントに置換することを目的としている。われわれはこの概説で、予後バイオマーカーおよび予測バイオマーカー、ひいては代替エンドポイントの統計学的妥当性確認を達成する上での課題について考察する。

「妥当性」という用語は混乱を引き起こし、論文でさまざまな意味で使用されてきた。バイオマーカーの発見および採用の全段階について命名をさらに標準化する必要がある⁴。執筆者や規制機関のなかには、バイオマーカーや代替エンドポイントの信頼性を確認する過程に「適格性確認」という用語を用いることがある⁵⁻⁷。このほかに「妥当性確認」と「適格性確認」という用語はいずれもバイオマーカー分析の有効性を確認する過程に用いられてきた^{1,6}。しかしわれわれは、本稿の目的のために分析妥当性のこのような重要点を無視して、妥当性が確認されたバイオマーカーを、頑健な統計学的方法によって、ある臨床的エンドポイントと関連していること（予後バイオマーカー）、臨床エンドポイントについて治療効果を予測すること（予測バイオマーカー）または治療効果を評価するための臨床的なエンドポイントに代えることができること（代替エンドポイント）に注意を向ける。

予後バイオマーカーおよび予測バイオマーカー

バイオマーカーには画像に基づく指標や生理学的指標があるが、標的療法の時代では、細胞性バイオマーカー、分子バイオマーカーおよび遺伝子バイオマーカーがますます重要になってきている。腫瘍学の分野では、乳癌の研究が細胞性バイオマーカー、分子バイオマーカー、遺伝子バイオマーカーの探索の先駆けである。特に Early Breast Cancer Trialists' Collaborative Group (EBCTCG)は、ホルモン受容体陽性腫瘍の再発および死亡ハザード比の経時的データを、ホルモン受容体陰性腫瘍と比較して、長期間集積した⁸。EBCTCGが実施した分析により、ホルモン受容体の有無が乳癌の予後マーカーであることが確認された。タモキシフェンやアロマターゼ阻害薬などの内分泌療法がホルモン受容体を発現した腫瘍に限って有益であることも確認した。すなわち、ホルモン受容体の状態は、治療に対する反応性の予測マーカーでもある⁸。現在の乳癌研究では、癌の分子的不均一性をきわめて詳しく探索しており^{9,10}、このような研究が、ひいてはバイオマーカーの発見や既知のバイオマーカーの使用の改善につながる可能性がある。乳癌以外で、妥当性が確認された予後バイオマーカーおよび予測バイオマーカーの数は依然としてわずかであるが、この分野に集積されてきている知見は、分子腫瘍学の進歩が最終的に患者選択と癌治療に大きな変化を起こす可能性を示唆している。

生物学的考察が予後バイオマーカーおよび予測バイオマーカーの初期特定に重要な役割を演じており、その評価と臨床医療への採用を通して、臨床試験の統計学的分析とともに重要である。例えば HER2/*neu* の過剰発現は乳癌のいくつかの型の生物学に不

不可欠であり、HER2/*neu*の状態はHER2経路を標的とした薬剤の有効性を予測すると見込まれる¹¹⁻¹⁵。乳癌におけるHER2/*neu*の過剰発現や消化管間質腫瘍のKITの変異、慢性骨髄性白血病にみられる融合遺伝子BCR-ABL（フィラデルフィア染色体）は、全て、腫瘍生物学的に予後のバイオマーカーであり、治療効果に関する予測バイオマーカーでもある例である。さらに、ここに挙げた遺伝子はいずれも癌の浸潤的特性を惹起し、治療標的となる可能性がある^{16,17}。

これに対して、疾患の病態発生に不可欠ではないバイオマーカー候補の方が予後予測力がある可能性が低い、場合によっては（有益性が高いことよりむしろ）有益性がないことを予測できる。たとえば結腸直腸癌のKRAS変異は疾患に必須ではないが、抗EGFRモノクローナル抗体の有益性がないことを予測し、そのことによって患者選択について予測バイオマーカーとしてきわめて有用である^{18,19}。

生物学的考察がバイオマーカーの採用を後押しすることがあるが、注意深く解釈せねばならないことが重要であり、場合によって誤解を招くことがある。たとえば最近1件以上の試験で、HER2/*neu*の過剰発現がなくても抗HER2治療が臨床的に有益な可能性があることが示唆され^{20,21}、これは現在わかっているHER2の生物学に基づいて予期することができない。確認されれば、この観察結果によって抗HER2治療の主要な分子標的との直接相互作用とは別の作用機序が明らかになるであろう。結局、生物学は臨床試験や統計学的分析によるバイオマーカーの妥当性確認試験の代わりにはならない。

予後バイオマーカーの妥当性確認

バイオマーカーおよび代替エンドポイントの妥当性確認は、統計学的な視点から言えば、最初にマーカーと目的とする転帰の間に相関が存在することを明らかにすることに始まり、続いてその関係について独立した統計学的妥当性確認を実施するのが典型的である。バイオマーカーを予後バイオマーカーとして妥当性を確認するには、試験開始時のマーカーの有無またはバイオマーカーの経時的変化と、治療に無関係な臨床的エンドポイントとの間の相関を明らかにしなければならない（表2）²²。これは統計学的には比較的簡単な要件であり、特別な試験デザインは一切必要ない。実際、小規模後ろ向き試験が十分なデータ源となりうることが多い。真の予後マーカーを特定するために必要な事象の数は、転帰のハザード比が極端になり（1から大きく離れる）、マーカーがある患者の割合が50%に近づくにつれて減少する。

初期の相関の確認から強固な妥当性確認への移行の問題点は、MammaPrint[®]（オランダ癌研究所およびアジェンディア BV、アムステルダム、オランダ）遺伝子シグネチャーに見られるように極めて大きい²³⁻²⁵。シグネチャーはマイクロアレイに基づく70遺伝子分析で、乳癌患者の予後を予測することを目標に開発された（図1）²³。比較的小規模の78例の標本を用いた後ろ向き試験で、そのシグネチャーは切除後5年以内の転移発生率の強力な予後マーカーであることが明らかにされた。MammaPrint[®]シグネチャーによる予後不良患者は、予後良好患者に対して5年以内遠隔転移のオッズ比が15.0（95% CI 3.3-56, P<0.001）であった²⁴。

70遺伝子シグネチャーの予後不良を予測する感度は高いが（転移のある患者の91%に予後不良シグネチャーが認められた）、特異度は中程度であること（予後良好シグネチャーが認められた患者のうち転移しなかったのは59%にとどまった）²⁴は注目に値する。シグネチャーの予測精度に目を向けると、シグネチャーの陽性予測値は0.63（予後不良シグネチャーが認められた患者の約3分の2に5年以内に転移がみられると予期される）で、陰性予測値が0.9（予後良好シグネチャーが認められた患者の10人に1人に限って5年以内に転移すると予期される）であった。この結果は、MammaPrint[®]シグネ

チャーには予後良好患者に対し侵襲的な化学療法を実施することを防ぐのに役立つ可能性があるが、患者が転移を来すか否かについて治療決定の唯一の基盤となるほど正確な予測因子ではないことを示唆する。

MammaPrint[®]試験の結果は、臨床的に広く使用することができる予後マーカーを特定するには、きわめて有意なP値とオッズ比が必ずしも十分条件ではないことを示し、教訓的である^{26,27}。事実、マーカー陽性およびマーカー陰性の Kaplan-Meier 曲線の明確で目に見える分離およびきわめて有意なログランク検定が、予後マーカーの価値の誤解を招くような評価である場合がある。古典的臨床的因子について補正した上で、(層別化もしくは多変量モデルによって)ある予後マーカー候補に目的とする特定の臨床的転帰に対する大きな効果が認められても、マーカーの予測精度が臨床応用を正当化するに足ることを意味するものではない。予測精度および説明変動の測定が事実、一般に必要とされているが、報告されることはまれである^{27,28}。

予後バイオマーカーの初期特定に続いて、多施設妥当性確認を実施するか、データセットをひとつだけしか利用できない場合に再抽出法を用いて、交差妥当性確認を実施する必要がある。MammaPrint[®]遺伝子シグネチャーの場合、ビルジュイフ、パリ、オックスフォード、ロンドンおよびストックホルムにある施設の独立した標本を対象に、ミラノで集中的に病理学的調査を実施し、ブリュッセルで統計学的分析をし、ローザンヌでマイクロアレイ分析をした大規模な妥当性確認試験を実施した^{29,30}。この解析では施設間でシグネチャーの予後についての有用性に有意な変動は認められなかったが、ハザード比が元のアムステルダム試験よりも弱いことが明らかにされた。たとえば元の試験では遠隔転移の経時的ハザード比が6.07であったのに対し、妥当性確認試験では2.13であった(生存ハザード比はそれぞれ17.46、2.63)²⁸。ここで認められた乖離の考えられるさまざまな原因のうち最も興味深いものは追跡期間であり、元の試験が6.7年であったのに対し、妥当性確認試験では13.6年と2倍の長さであった。この結果は遺伝子シグネチャーの予後効果は時間に強く依存することを示唆し、シグネチャーは疾患が遅れて進行するリスクが高い患者よりも早期に進行するリスクが高い患者を特定するのにきわめて優れているといえる^{29,30}。この妥当性確認試験は、MammaPrint[®]遺伝子シグネチャーが確かに、年齢や腫瘍の大きさ、病期、エストロゲン受容体の有無およびリンパ節転移の有無など、既知の臨床的因子や病理学的因子から引き出される情報に付加的な予後情報をもたらすと結論したが、全般的臨床的有用性については未だ確認されていない。

バイオマーカーの多施設統計学的確認ができて、その臨床的有用性の最終確認にはなお、臨床試験での無作為化された前向きな根拠が必要である。特に、最適な治療が古典的パラメーターでは明らかにできない患者に対するバイオマーカーの有用性を明らかにするには、前向き試験が必要である。たとえば再発のリスクが低い患者には標準治療に限って必要であろうが、再発のリスクが高い患者には研究的治療を必要とする。しかし、バイオマーカーや臨床病理学的因子によるリスクが中等度の患者には、治療決定に関して不明確性であろう。バイオマーカーの治療決定に演じる役割を明らかにするには、このような患者を標準治療群か研究治療群に無作為に割り付ける必要がある。早期乳癌に対するこのようなバイオマーカーに基づく治療の試験には、現在実施されている MINDACT 試験³¹ および TAILORx 試験³² がある。

予測バイオマーカーの妥当性確認

MammaPrint[®]シグネチャーは、有望な予後バイオマーカーの統計学的妥当性確認の困難さを物語っている。しかし予測マーカーの方が、マーカーと転帰の相関の初期検証とそれに続く頑健な妥当性確認のいずれも難しい(表2)。

バイオマーカーの基礎値や経時的変化が治療の有効性や毒性を示す場合、規定された臨床的エンドポイントによって評価された通り、バイオマーカーを予測的と考えることができる¹⁹。予測バイオマーカーの統計学的特定には、バイオマーカー値が高い患者と低い患者の双方を含む患者を対象とした無作為化試験によるデータを必要とする。最も高いレベルの根拠は「相互作用」デザインで、患者をバイオマーカー値によって層別化し、2つの治療のうち1つに無作為に割りつけた試験に由来するものである。この妥当性確認試験の方法は現在実施されている Marker Validation of Erlotinib in Lung Cancer (MARVEL)試験で用いられている。EGFRの有無について患者を検査し、非小細胞肺癌の二次治療としてエルロチニブ(Tarceva[®], F. Hoffmann-La Roche Ltd, バーゼル、スイス)とペメトレキセド(Alimta, Eli Lilly and Company, インディアナポリス、インディアナ)に無作為に割りつけたものである³³。この試験では、マーカー陽性患者とマーカー陰性患者で治療効果が異なることを明らかにすることを目的とし、相互作用試験を用いて2群を個別に分析した。一般に、信頼できる相互作用の検出には数多くの事象が必要とされ、このため大規模な患者集団が必要とされる³⁴。このことから現実的には、予測マーカーの妥当性を確認できる「相互作用」試験が少数である可能性が高い。

このことから、二者択一の治療決定に役立つ新規バイオマーカーの大部分の妥当性確認は相互作用試験に代わる方法を用いることになる。「選択」デザインが採用されることが多く、この場合マーカー陽性患者に限って妥当性確認試験に登録される^{35, 36}。このような試験には、治療が有効な集団を特定するためのマーカーの有用性を確認する能力があるが、マーカー陰性患者に対して有益ではないことについての情報が得られないため、マーカーが真に予測的であることを意味するものではない。このことの重要な例が乳癌の再発遅延または再発予防に対するトラスツズマブ(ハーセプチン[®]、ジェネンテック社、カリフォルニア州サンフランシスコ)の効果である。腫瘍にHER2/*neu*の増幅がみられる患者には、トラスツズマブ治療の有益性がいくつかの大規模無作為化試験で確認されている¹¹⁻¹⁵。しかしこの治療がHER2/*neu*の増幅が見られない腫瘍患者にもほぼ同じ効果を示すことが示唆されている^{20, 21}。具体的には、NSABP B-31²⁰試験およびNCCTG9831²¹試験に登録された患者の約10%に地域の研究拠点でHER2/*neu*が増幅していることが確認されたが、中央の研究拠点で再試験をしたところ認められなかった。患者のこのサブセットの解析によって、このような患者に中央研究拠点でHER2/*neu*の増幅が認められた患者と同等の効果が認められたことが示唆された。この観察結果に対して、次のようないくつかの説明ができる。遺伝子の増幅がなくHER2が過剰発現している、トラスツズマブが既知のHER2変異を介することなく作用する、HER2/*neu*増幅の評価に研究的アーチファクトがある、比較的小さなサブセットに基づく結果が単純に偶然によるものである、などである。この特定の事象の理由がいずれであれ、マーカー陽性の患者に限って治療が有効であることが前提であっても、適度に小規模のマーカー陰性群を対象とする相互作用試験が適応となる場合があることが示唆される。

「選択」アプローチは数多くの現在の試験で使用されており、特にN0147³⁷試験、PETACC-8³⁸試験およびC80405³⁹試験をはじめKRAS野生型結腸直腸癌患者を対象に、進行癌の治療または補助療法としてEGFR阻害剤を検討するために始められた試験で用いられている。ほかにステージ2の結腸癌患者を対象としたECOG E5202⁴⁰試験がその例である。この試験では、マイクロサテライト不安定性(フルオロピリミジン抵抗性の予測バイオマーカー候補)や正常の18q染色体(予後バイオマーカー)のある腫瘍患者には補助療法を実施しないが、患者にマイクロサテライト安定性および18q染色体異常が認められた場合、5フルオロウラシル、ロイコボリンおよびオキサリプラチンを

用いた標準的補助療法にベバシズマブを加える群と加えない群にランダムに割りつけた。Sargent ら^{35,36,41}は予測バイオマーカーの前向き妥当性を確認するこの試験デザインの限界について詳しく考察している。簡潔に述べると、選択デザインは予測効果を確認することができないが、予測デザインではその統計学的検出力が十分ではないことがあるというものである。

予測バイオマーカーの妥当性を確認するための無作為化試験の問題点を考えると、終了した無作為化試験の後ろ向き解析がやはり最も重要な根拠源であろう。しかしこのような後ろ向き解析で説得力のある根拠を得るためには、分析的詳細(カットポイント、統計学的方法など)を記載した前向きの手順と解釈上の指針(統計学的有意性のレベル、効果の強さなど)を作成しなければならない。何回か実施することができるのであれば、結果は全体を通じて一致しなければならない。この「後ろ向き前向き」アプローチは最近、結腸直腸癌に対して用いられ、KRAS 変異の有無がセツキシマブ(アービタックス[®]、Merck KgaA、ダルムシュタット、ドイツ)およびパニツムマブ(ベクチビックス[®]、アムジェン、サウザンドオークス、カリフォルニア)という2つの抗EGFRモノクローナル抗体が全く無効であることを予測することが多数の試験で一致して示された^{18,19,42-44}。

代替エンドポイントの妥当性確認

現時点で臨床腫瘍学には、承認された代替エンドポイントがわずかであり、腫瘍の反応性や固形癌の分子マーカーおよび遺伝子マーカーに基づくものはない⁴⁵。これに対し非固形癌では、血液学的完全緩解が長く時間対疾患進行および全生存率の代替と考えられてきた^{46,47}。代替エンドポイントの統計学的妥当性確認というのは、代替エンドポイントと臨床的エンドポイントの間の明白な相関の初期実証とそれに続く頑健性検証のいずれもきわめて困難な過程である(表2)。代替エンドポイントに関する一般的な論文調査をWeirとWally⁴⁸およびLassere⁴⁹が報告している。

Prenticeはまず、代替的關係を確認するには代替が臨床的エンドポイントを予測する必要があり、代替エンドポイント候補と臨床的エンドポイントの双方に治療が有意に有効でなければならず、代替エンドポイントに対する治療効果が臨床エンドポイントに対する治療効果を十分に捕えていなければならないと提言した⁵⁰。この必要条件の后者は非現実的であると考えられ、その結果代わりの前提が開発された。妥当性基準というのは未だ統計学的研究が力を注いでいる領域であるが、現在の結論は、妥当性確認が「相関アプローチ」に基づくことができるということである。これには、無作為化試験やメタアナリシスによって、一方で代替が疾患転帰に対して予後的であり、その一方で代替に対する介入の効果は真のエンドポイントに対する効果と十分相関していることを検証することが含まれる^{48,39,51-57}。

このように妥当性を確認するには、候補代替マーカーはまず予後マーカーと同じように、特定の介入に無関係に転帰を予測することを示す必要がある。代替性のこのような面は一般に個体レベルの代替(転帰代替ともいう)と言われ、個々の患者についてマーカーまたは代替転帰が生存などの最終的に目的とする転帰が十分相関していなければならないという意味である。次にさらに重要なのが、候補代替マーカーに対する治療効果が真の臨床エンドポイントと密接に相関していなければならないということである。この面は通常、臨床試験で患者群に認められるので、試験レベルの代替(効果代替ともいう)と言われる。

統計学的に個体レベル代替および試験レベル代替は、通常配分されたエンドポイントについて互いに独立であり、このためそれぞれ個別に妥当性を確認する必要がある⁵⁸。標準相関係数が統計学的相関を定量化するのに最も頻繁に用いられる方法であるが、

ほかに情報理論に由来する代替尺度を利用することができ、統一のパラダイムとして有望である⁵⁹。個体レベル代替は単一の試験で確認することができるが、試験レベル代替の欠点として、典型的にはいくつかの無作為試験のメタアナリシスに基づくものでなければならないことが挙げられる⁶⁰⁻⁶²。しかし、メタアナリシスを実施に十分な数の試験が使用できない場合、大規模試験の結果を、個々の国や試験施設など小規模な分析単位に分けることができる。この方法は、前立腺特異的抗原が進行前立腺癌患者の生存に対する代替とは考えられないことを示すために用いられた^{63,64}。

試験レベル代替については最近、代替閾値効果(STE)という概念が取り入れられた。STEは真のエンドポイントに対する効果を予測するのに必要な代替に対する最小の治療効果と定義される⁶⁵。代替にSTEより大きな効果をもたらすことができる治療が、臨床転帰にも比較的大きな効果をもたらすことができると考えられることから、このアプローチでは自然に臨床的視点から代替を解釈することができる。進行結腸直腸癌では、フルオロピリミジンを主とした治療について無進行生存期間(PFS)および全生存期間が強く相関しており(個体レベルの代替)PFSに対する治療効果と全生存期間に対する治療効果が相関しており(試験レベルの代替)、PFSが全生存期間の代替として受容できることが明らかにされている(図2および3)⁶⁶。時間的に一連のフルオロピリミジンを主とする治療を評価した無作為化試験10件を基に、代替閾値効果は0.86または影響力の強い試験を除外して0.77であり、新規治療が腫瘍進行のハザードを23%以上減少させれば、生存に有益である可能性がきわめて高い(図3)⁶⁵。

しかし、代替エンドポイントの妥当性評価に関する主要な問題点は、特定の治療または特定の組み合わせの治療に関する妥当性であるということに由来する。新規作用機序の新たな治療について、これまでの治療で確認されている代替関係を適用可能であるか明らかではない。たとえば進行結腸直腸癌の第一次試験では、ベバシズマブ(アバスタチン[®]、ジェネンテック、サンフランシスコ、カリフォルニア)、パニツムマブ、セツキシマブなどの新規標的治療に関してPFSは未だ、全生存期間の代替として用いることができると確認されていなかった。これまでの治療について確認された代替を、新たに出現した各治療の臨床的評価の代替とするのが正当であると考えることが合理的であるのかという問題が浮上する。治療法が時間とともに発達すると言う事実からさらに問題が生まれる。たとえばPFSは、現在多様な救済療法を可能にした新たな細胞傷害性薬剤や標的治療が導入される前に、結腸直腸癌に対して古典的な5フルオロウラシルを主とする化学療法での代替として妥当性が確認されている。このような治療が結腸直腸癌治療の進化上、早期に使用可能になっていたら、5フルオロウラシルを主とする化学療法についてPFSが全生存期間の代替となりうるということが確認された可能性は低い⁶⁶。

事実臨床腫瘍学の分野で標準治療が発展すると、実際の治療法の数が増加して生存期間が長くなるに従って、短期エンドポイントおよび全生存率の代替の確認が困難になるのが避けられない。進行乳癌の領域の最近の試験で、腫瘍反応性、PFS、疾患制御および時間対疾患進行がいずれも全生存率と個体レベルで相関していたが、妥当性のある代替エンドポイントとして適格とするに足る試験レベルでの強い相関が認められたものは一切なかった⁶⁷。この状況の例が、進行乳癌を対照としたベバシズマブの大規模試験で、ベバシズマブ治療がきわめて有意なPFS有益性をもたらしたが、全生存率には有益性が認められなかったことである⁶⁸。

このような観察結果は、現在いくつかの治療法がある固形癌の全生存に対して、PFSを代替として正式に認めることは難しいかも知れないことを示唆しているが、これはPFSがそれ自体でエンドポイントとして有用性がないということの意味するものではない。PFSは実に、多数の積極的治療法が利用可能である(いずれも全生存率を改善す

る可能性がある)ことを考えると薬剤評価の唯一の感受性のある(かつ現実的な)エンドポイントであろう⁶⁹。

厳密な妥当性確認か実用的な妥当性確認か

現在、腫瘍学の分野で妥当性が確認された予測バイオマーカーおよび代替バイオマーカーが不足していることは、本稿で考察した統計学的な問題のみならず、発見と評価の過程の全段階の問題を示している⁷⁰。米国国立癌研究所早期発見研究ネットワークは、早期癌発見のバイオマーカー発見のために5つの明確な相を提言している^{70,71}。表3にこの5つの相をいずれのバイオマーカーの開発にも当てはめ、例としてMammaPrint[®]の現在の状態を各相について概説する。妥当性確認の最大の問題のひとつが、政府に資金提供を受けた機関による試験であれ、企業による試験であれ、主要な試験全部から高品質の生物学的標本および反応の標準化された尺度の双方を利用することができないことである。欧州医薬品審査庁(EMA)およびFDAなどの規制機関は企業や学界のパートナーにこの問題を緩和する条項を作成させることを考える必要がある。たとえば多試験組織バンクやデータベースを設立すれば、バイオマーカーの研究を後押しし、後ろ向き解析の資源を提供することができる。米国のNIHバイオマーカーコンソーシアムの設立は、この方向に向けた歓迎されるが小さな一歩である⁷²。

問題はあつたが、次の数年はさまざまな程度の統計学的妥当性のバイオマーカーの候補がますます数多く集積する可能性が高い。報告の標準を順守することは妥当性確認試験の質の向上につながるであろう⁷³。このほか個々の候補の予測力は低いが、転帰や反応を予測するために、個々のバイオマーカーに加えてバイオマーカーのグループが開発され、集中的にもちいられる(たとえば多変量解析または他の予測モデル)可能性が高い。生物学的考察によって妥当性確認が不完全であっても予測バイオマーカーの妥当性を高めることができ、ひいては実際的な受容可能性を向上させることができる。たとえば現在、バイオマーカーと治療薬の間の相互作用に関して正式な検証はないが、HER2/*neu*の状態がトラツズマブの有効性の予測バイオマーカーとして受け入れられている⁷⁴。これはHER2/*neu*陰性腫瘍患者に対する有益性に関する説得力のある証拠は未だにないが、トラツズマブの有益性が大きく、HER2/*neu*陽性腫瘍患者について確認されていることが主な理由である。たとえば、バイオマーカーが存在する場合にある特定の治療を必要とする転帰に対する感受性が高いバイオマーカー(このマーカーのある患者のほとんどに治療を実施する必要がある)に関するデータを、バイオマーカーが存在しない場合に治療に対する高い特異性が低下するバイオマーカーのデータ(このマーカーのない患者のほとんどには治療を実施すべきでない)と考え合わせるといように、いくつかのバイオマーカーの情報を組み合わせて用いることはまた、臨床的方針決定に実際的に有用性がある可能性が高い。

代替エンドポイントに関しては、PFSや疾患制御に関する類似する尺度の妥当性評価に認められる問題は、治療法の幅が広がれば、疾患制御と全生存率の双方を大きく改善することができる治療に関しても増える可能性が高い。逆説的に、臨床試験の速度を速めるには早期に観察できるエンドポイントが必要であるが、新規治療が最も有効な状況ではそれが全生存率に対する代替となりうる可能性がきわめて低い。この逆説は、このような新規治療が、一度一次試験治療が終了した患者の全部ではないとしても大半に実施され、そのために全生存に対する新規治療の有益性を統計学的に確認することが困難になることに直面する。いくつかの治療法をはじめとする治療戦略を比較する試験によって、たとえばその薬剤を第一次治療の一部として投与すべきか後の治療時に投与すべきかを決定するなど、新規薬剤の真の効果の評価が改善されてき

た^{75,76}。しかし、きわめて有効な新規薬剤をある治療が失敗した後に投与することの交絡作用は、依然として全生存の解析についての問題点である。

ここに挙げた問題点のために、候補代替が統計学的妥当性に基づく評価だけでなく調和に関する国際会議で提案されたように、臨床試験で確認された生物学的妥当性や有用性に関して評価することができる実際的な方法が必要である⁷⁷。このことから、代替エンドポイントをどのように系統的に頑健に、幅広い実用的な枠組みのなかで共通の標準について評価することができるかという問題が浮かび上がる。Lassere は生物学的、疫学的、統計学的、臨床試験の重みづけした評価およびリスク便益の根拠に基づいて、代替バイオマーカーとエンドポイントの間の関係の強さを数値的に評価するための秩序だった図式を提案した⁴⁹。このほか Lathia ら⁷⁸は候補代替の多面的評価を考察し、代替の採用のための柔軟な方法を述べている。さらに明確化し標準化することが必要である。その一方で、生物学的考察に裏付けられた個体レベルデータおよび試験レベルデータが、未だに候補代替の妥当性を評価する基本的な基準である。

結論

予後バイオマーカー、予測バイオマーカーおよび代替エンドポイントはいずれも腫瘍学に必要なが、わずかし確認されておらず、試験や臨床でのその有効性の証拠は依然として限られている。初期特定に必要なのが比較的小規模な後ろ向きデータである予後バイオマーカーでさえ、多くの候補が独立した妥当性確認試験の後に不備であると明らかにされる。予測マーカーの妥当性確認には大規模多施設無作為化試験が必要であるのに対して、全バイオマーカーのうち代替エンドポイントが最も特定が困難であり、妥当性確認に無作為化試験のメタアナリシスが必要である。困難であるが、数多くの試験が腫瘍学の分野のバイオマーカーや代替エンドポイントを特定し妥当性確認をすべく実施されており、有効な標的治療の探索が継続されるにつれて他に多くの候補が生まれる可能性が高い。現実的にはバイオマーカーや代替エンドポイントの採用はどのような状況でも徹底的な統計学的妥当性確認によることができず、統計学的、臨床的および生物学的考察の組み合わせを用いた証拠に基づくものでなければならない。これによって、次にバイオマーカーおよびエンドポイントの採用の幅広い過程をいかに構造化し評価すべきかという問題が生まれる。

要点

- ・予後バイオマーカー候補は比較的簡単に特定できるが、多施設の妥当性確認試験が実施されることはまれであった。
- ・予測バイオマーカーの妥当性確認には、大規模無作為化臨床試験およびメタアナリシスに基づく膨大なデータが必要である。
- ・代替エンドポイントは、予後について真のエンドポイントについて代替的であることと、治療効果に関しても真のエンドポイントと相関していること、双方が確認されたデータが必要である。
- ・バイオマーカーまたは代替エンドポイントの生物学的信頼性は、完全な統計学的妥当性確認がない場合でも、その採用を裏付けることがある。
- ・強固な統計学的妥当性確認がない場合のバイオマーカーの実用性評価、採用に関する、最適な方法については、意見の一致に至っていない。

表 1. バイオマーカーおよび代替エンドポイントの定義
用語 定義

バイオマーカー 正常の生物学的過程、病理学的過程または治療に対する薬理学的反応の指標として客観的に測定し、評価することができる特性²
 予後バイオマーカー 治療とは無関係に疾患の経過を予測するバイオマーカー
 予測バイオマーカー ある特定の治療に対する反応性を予測するバイオマーカー
 臨床エンドポイント 患者の感覚、機能、生存についての情報をもたらす尺度³
 代替エンドポイント 臨床エンドポイントと治療効果を早期に正確に予測する尺度
 妥当性確認 強固な統計学的方法を用いて予後マーカー、予測マーカーや代替エンドポイントが臨床応用に必要かつ十分な条件を満たすことを確認すること

表2 予後および予測バイオマーカーおよび代替エンドポイントの例

バイオマーカーの種類	治療および臨床試験での使用	特定	妥当性確認	例
予後バイオマーカー	治療選択、患者選択および層別化	簡単だが欠陥や偏りが認められることが多い	頻繁であるが方法への回帰や初期特定試験の欠陥により不十分なことが多い	一般状態不良、肝酵素上昇、結腸直腸癌の多発転移
予測バイオマーカー	治療選択、患者選択および層別化	困難、無作為化試験を要する	一般的ではない、大規模無作為化試験を要する	結腸癌のセツキシマブおよびパニツムマブの活性欠如を予測する KRAS 変異 ^{18,19} 。乳癌のタモキシフェンおよびアロマターゼ阻害薬の効果予測するホルモン受容体状態 ⁷⁹ 。乳癌のトラツズマブおよびラパチニブの効果予測する HER2/ <i>neu</i> 増幅 ¹¹⁻¹⁵
代替エンドポイント	治療選択、治療評価	きわめて困難、メタアナリシスまたは大規模無作為化試験を要する	まれ、メタアナリシスまたは大規模無作為化試験を要する	結腸癌のフルオロピリジンを主とした治療法について全生存の代替としての無進

				行生存 ⁶⁶ 。白血病患者の疾患進行時間に対する血液学的完全緩解 ^{46,47}
--	--	--	--	--

図1 予後マーカー：乳癌のアムステルダム70遺伝子シグネチャーに関する初期の所見

図2 進行結腸直腸癌の無進行生存(PFS)および全生存(OS)

図3 進行結腸直腸癌の代替エンドポイント：試験レベル(効果)代替および代替閾値効果

表3 バイオマーカー開発および妥当性確認の相⁷¹

相	題名	目的	MammaPrint [®] の例
1	臨床前探索的試験	有望なバイオマーカーの特定	該当なし
2	臨床検査開発	バイオマーカーの測定に用いる臨床検査の開発および妥当性確認	78例を対象とした初期試験 ²³
3	後ろ向き妥当性確認試験	利用可能な患者群、腫瘍バンクおよび他の保存検体を対象としたバイオマーカーの効果の定量化	同一施設の234例 ²⁴ および他の5施設の326例 ²⁹ を対象とした後ろ向き試験
4	前向き妥当性確認試験	前向き試験によるバイオマーカーの効果の確認	現在実施中の前向き試験 ⁸¹
5	臨床的有用性	前向き試験によるバイオマーカーの臨床的有意性の確認	現在実施中の前向き試験 ⁸¹

調査基準

2004年1月1日から2009年4月1日までに公表された論文をPubMedデータベースを用いて検索した。使用した検索用語は論文題名および要旨に含まれる「代替エンドポイント」「バイオマーカー」「妥当性確認試験」および「適格性確認」とした。結果は関連する論文およびPubMedが指定した関連論文についてスクリーニングした。著者らは個人的知見および経験に基づいて、調査結果にさらに論文を提供した。

国内外の規制と米国での認可状況 (評価基準の作成に向けて)

国立医薬品食品衛生研究所
遺伝子細胞医薬部
(審査WG事務局)

鈴木 孝昌

体外診断薬のリスクに基づくクラス分類表

国際分類	リスクによる医療機器の分類	分類	リスク	製造販売規制
クラス I	一般的な較正用標準物質が存在し、製品の製造管理及び品質管理の一貫で行う較正が比較的容易(一般用検査薬以外) (例)血液 ヘモグロビン、GOT、GTP、インスリン	クラス I	小	承認・認証不要 (届出/自己認証)
クラス II	クラス III 以外の一般用検査薬 クラス I、III 以外クラス II (例) ビタミンB ₁₂ 、妊娠検査薬、尿試験紙、抗核抗体	クラス II	中	登録認証機関による認証 (認証基準に適合するものに限る) 大臣による承認 (総合機構による審査)
クラス III	診断情報リスクが比較的大きく、情報の正確さが生命維持に与える影響が大い	クラス III	大	
クラス IV				(例) HIV 抗体、グルコース自己検査、PSA、HLA、風疹ウイルス抗体、HER-2 遺伝子

医療機器のリスクに基づくクラス分類表

国際分類	リスクによる医療機器の分類	分類	リスク	製造販売規制
クラスⅠ	不具合が生じた場合でも、人体へのリスクが極めて低いと考えられるもの (例) 体外診断用機器 鋼製小物、X線フィルム、歯科技工用用品	一般医療機器	極めて低い	承認・認証不要 (届出/自己認証)
クラスⅡ	不具合が生じた場合でも、人体へのリスクが比較的低いと考えられるもの (例) MRI、電子式血圧計、電子内視鏡、消化器用カテーテル、超音波診断装置、歯科用合金	管理医療機器	低い	登録認証機関による認証 (認証基準に適合するものに限る)
クラスⅢ	不具合が生じた場合、人体へのリスクが比較的高いと考えられるもの (例) 透析器、人工骨・関節、人工呼吸器、バルーンカテーテル	高度医療機器	中・高	大臣による承認 (総合機構による審査)
クラスⅣ	患者への侵襲性が高く、不具合が生じた場合、生命の危険に直結するおそれがあるもの (例) ペースメーカー、人工心臓弁、ステント			

医療機器のクラス分類ルール

(平成16年7月20日付け医薬発第0720022号医薬局長通知別紙)

2. 分析機器のクラス分類のルール

クラスⅢ

誤った診断結果が得られた場合に、人の生命及び健康に重大な影響を与えるおそれがある検査項目を測定する分析機器。

①自己検査用測定機器

②主たる反応系を内蔵する専用分析機器のうち、体外診断用医薬品で承認を必要とする検査項目を測定するもの

クラスⅡ

主たる反応系を内蔵する専用分析機器のうち、標準品の無いもの(クラスⅢ品目を除く。)

クラスⅠ

その他の分析装置(注)

注) 主たる反応系を内蔵する専用分析機器について、標準品があるもの

(備考)

新検査項目、新測定原理、新たに自己検査用に移行するもの、新たな検査項目を測定する主たる反応系を内蔵する専用分析機器は大臣承認とし、承認の際にクラス分類を決定するものとする。

**体外診断薬としてはクラスⅢ、医療機器としてはクラスⅠ
ただし、分析系すべてが機器に組み込まれれば、医療機器としてクラスⅢ**

米国における体外診断薬規制の特徴

- 医療機器として審査される
- リスクに応じてクラス分類（クラスI, II, III）
- PMA (クラスIII)、510(K) (クラスII)
- CLIA認証とLDT (Laboratory Developed Test) の問題
- (保険制度の違い)

米国における2つの診断薬市場化パスウェイ

LDT

(Laboratory Developed Test)

単一ラボにて、開発、評価、
実施される検査

CMS

(Centers for Medicare & Medicaid Services)

CLIA

(Clinical Laboratory Improvement Amendment)

試験機関を評価するもので
検査自体の評価ではない！
(臨床評価が必要ない！)

自由裁量

(Enforcement Discretion)

研究用試薬

ローカルな場合が多い

内容

担当機関

規制

評価対象

FDAの立場

ラベリング

需要

IVD

(In Vitro Diagnostics)

一般に普及させるため
の検査キット:

FDA

510K PMA

検査自体を評価するもので
試験機関の評価ではない！
(臨床評価が重要)

承認、認証

診断用試薬

広く普及

FDAによるLDTの規制動向

LDTのクラシカルな例としては、
染色体検査、免疫組織染色

LDTは一般に規制対象外

Analyte Specific Reagent
(抗体など)

ASR Rule (1997) 検査に使う試薬を規制する

ゲノム解析技術の進展により
LDTがより複雑になっていった

↓
キット全体をASR とするなど正しく認識されず、期待された効果がなかった。

ASRの認可をとらず、Research Use Onlyとラベルして、医療機関に販売した

IVDMIA Draft Guidance (2006)

ASR Q & A Guidance (2007)

Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS)による指摘

↓
LDTとして開発されてもFDAの認証を取らないといけないことにする。

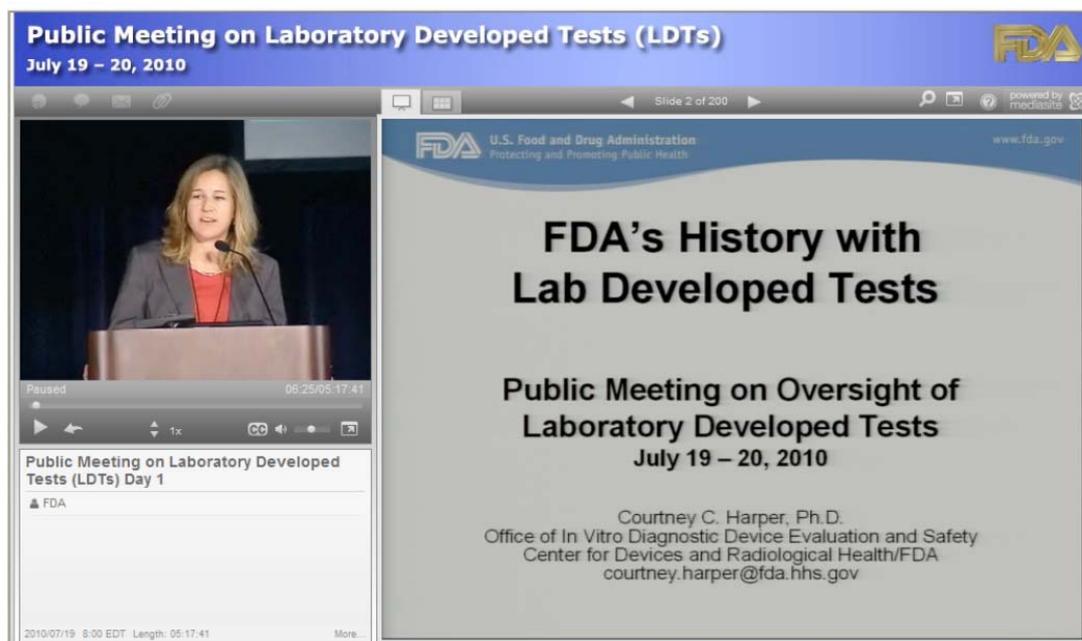
DTC (Directly Offered to Consumers) Genetic Testの信頼性問題

FDA/CDRH Public Meeting:
Oversight of Laboratory Developed Tests (July 19-20, 2010)

FDAによるLDTの規制強化に関する公開ミーティング

The screenshot shows the FDA website page for the public meeting. The header includes the U.S. Department of Health & Human Services logo and the FDA logo. The main content area features a navigation menu with categories like Home, Food, Drugs, Medical Devices, etc. The page title is "FDA/CDRH Public Meeting: Oversight of Laboratory Developed Tests (LDTs), Date July 19-20, 2010". A notice states: "NOTICE: MEETING LOCATION HAS BEEN CHANGED. The public meeting has been changed to an alternative site in the Washington, DC metro area to accommodate the significant public response received and requests to attend the public meeting. Details on the new meeting site can be found below." A list of links is provided: Date, Time and Location; Webcast; Background; Session Descriptions; Federal Register Notice; Agenda; Transcripts; and Contacts. The "Date, Time and Location" section specifies the meeting was held on July 19 (8:00 a.m. to 5:00 p.m.) and July 20 (8:00 a.m. to 5:30 p.m.) at The Marriott Inn and Conference Center, UMUC, 3501 University Blvd E, Hyattsville, MD 20783. The "Webcast" section states the meeting was webcast at no charge and provides links for Monday, July 19; Tuesday, July 20; and a Closed Caption Webcast for both days.

Webcastとして公開中



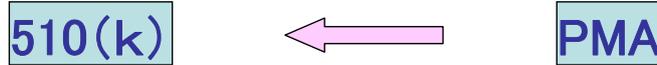
FDAに認可されたIVDMIA

- **MammaPrint[®]** (2007年2月)
 - Agilent microarray (70遺伝子)
 - 乳癌の遠隔転移(予後)予測
- **Pathwork[®] Tissue of Origin** (2008年7月)
 - Affymetrix PathChip (1550遺伝子)
 - 15種類の典型的な癌のタイプを分類
- **AlloMap[®] Molecular Expression Testing** (2008年8月)
 - リアルタイムPCR (11+9 control遺伝子)
 - 心臓移植の拒絶の予測



510(k)とPMAの比較

Special Control Guidance



申請	Premarket Notification Submission 市販前届け	Premarket Approval Application 市販前申請
クラス分類	クラスII	クラスIII
審査内容	実質的同等性	臨床試験
審査期間	3-4ヶ月	10-12ヶ月
審査結果	Clearance	Approval

IVDに関する規制情報の検索

U.S. Department of Health & Human Services | www.hhs.gov

FDA U.S. Food and Drug Administration | A-Z Index | Search

Home | Food | Drugs | Medical Devices | Vaccines, Blood & Biologics | Animal & Veterinary | Cosmetics | Radiation-Emitting Products | Tobacco Products

FDA Home > Medical Devices > Databases

IVD - In Vitro Diagnostics

510(k) | Registration & Listing | Adverse Events | Recalls | PMA | Classification | Standards
CFR Title 21 | Radiation-Emitting Products | X-Ray Assembler | Medsun Reports | CLIA

IVD - In Vitro Diagnostics [about IVD](#) | [help](#)

Enter a search term in the space below.

Optional:

Approval Date From: To: 10/31/2010

search clear 10 Records per Report Page

sort by: Approval Date Device Name

IVD - In Vitro Diagnostics searches the following databases.

510(k) Premarket Notification	PIA - Premarket Approval	CLIA - Clinical Laboratory Improvement Amendments
-------------------------------	--------------------------	---

We welcome your [comments and feedback](#) about IVD - In Vitro Diagnostics.

Page Last Updated: 10/19/2010

Home | About FDA | Contact Us | A to Z Subject Index | Web Site Policies | FOIA | Accessibility | No FEAR Act

Combination Products | Advisory Committees | Science & Research | Regulatory Information | Safety | Emergency Preparedness | International Programs
News & Events | Training and Continuing Education | Inspections/Compliance | State & Local Officials | Consumers | Industry | Health Professionals

FDAにおける510(k)関連資料

- MammaPrint®** *Class II Special Controls Guidance Document: Gene Expression Profiling Test System for Breast Cancer Prognosis (May, 2007)*
 - K062694 Original (凍結組織を使用) (Jan, 2007)
 - K070675 Modification (新鮮組織を使用) (Jun, 2007)
 - K080252 Modification (アレイHD化とスキャナー追加) (Jul, 2008)
 - K081092 Modification (適応年齢を60歳から87歳へ) (Dec, 2009)
- Pathwork® Tissue of Origin**
 - K080896 Original (凍結組織を使用) [vs BioPlex 2200] (Jul, 2008)
 - K092967 Modification (FFPE用Kitを使用) (Jun, 2010)
- AlloMap®** *Class II Special Controls Guidance Document: Cardiac Allograft Gene Expression Profiling Test Systems (October, 2009)*
 - K073482 Original (血液を使用) (Aug, 2008)

510(K)記載内容の比較(1)

	MammaPrint	Pathwork	Allomap
B. 測定対照	70遺伝子(乳癌)	15種の癌の遺伝子発現	20遺伝子(乳癌)
D. タイプ	マイクロアレイ(Agilent)	マイクロアレイ(Affymetrix)	定量的RT-PCR
H. 適応	凍結癌組織を使って、その発現情報から患者の遠隔転移のリスクを評価する	15種類の診断のついた典型的な癌の発現情報をデータベース化し、類似性を評価する。	末梢単核球の遺伝子発現から、心臓移植時の拒絶反応の予測を補助
適応外	61歳未満のステージI, IIの乳癌患者 (腫瘍サイズ?5cm LN-) 臨床病理学的に判断のつかない種類の癌 予後や治療効果を予測するものではない 原発か転移かの区別 15種以外の癌の判別	bladder, breast, colorectal, gastric, hepatocellular, kidney, non-small cell lung, ovarian, pancreatic, prostate, thyroid carcinomas, melanoma, testicular germ cell tumor, non-Hodgkins lymphoma, soft tissue sarcoma 診断、治療の感受性の予測、治療の選択	15歳以上 移植後55日以上
使用機器	Agilent 2100 Bioanalyzer (2台を特定) Agilent DNA microarray scanner (1台)	Affymetrix GeneChip System (K080995)	Biomek FXp (3台特定) ABI 7900HT (8台特定)

510(K)記載内容の比較(2)

	MammaPrint	Pathwork	Allomap
I. デバイスの説明	Agilent製カスタムアレイを使用 1900-feature subarrays 232遺伝子triplicate (予測用70遺伝子含む) + 915 標準化遺伝子、289品質評価スポット 予後良好な44癌サンプルをリファレンスとして 使用 Low Risk, High Risk, Low Risk Borderline, High Risk Borderlineに分類	Affymetrix製PathChipを使用 500,000 oligonucleotide probe 14,500 遺伝子に対応 machine learning methodにて選択した1550遺 伝子の情報を診断に利用 121プローブセットを標準化と品質評価に利用 2039サンプルでの解析結果をデータベース 化して照合 癌種ごとにSimilarity Score(0-100)を算出。 30以上で陽性。	単一ラボ(XDx)にて実施 血液をキットで処理して、凍結状態で送付 CPT™ Tube+Reagents+Frozen Shipper Pack
J. 同等性	なし	BioPlex™ 2200 Medical Decision Support Software (k043341)	なし
K. 参照ガイダンス	CDRH Guidance: Guidance for the content of premarket submissions for software contained in medical devices	ClassII SCGD: Gene Expression Profiling Test System for Breast Cancer Prognosis ClassII SCGD: Drug Metabolizing Enzyme Genotyping System	Draft Guidance: In Vitro Diagnostic Multivariate Index Assays ClassII SCGD: Gene Expression Profiling Test System for Breast Cancer Prognosis Interference Testing in Clinical Chemistry; Approved Guideline (EP7-A2)
L. テスト原理	リファレンスRNA(予後良好な44サンプル)と Cy3/Cy52色比較 Agilent Future Extraction Softwareで数値化 し、独自のアルゴリズムでMammaPrint Index(-1~+1)を算出 +0.4より大;Low Risk, +0.4以下;High Risk ボーダーライン(0.365-0.435)は再試験、平均 が0.3775-0.4225ならボーダーラインと結論	組織100mg。癌>60%。壊死<20% RNA 1 μg チップ、試薬等をキット化し、スキャンまでを 各機関で行い、イメージデータをPathwork社 に送付 専用のソフトウェアで解析 15のSimilarity Scoreを算出し、レポートをpdf ファイルで返送	採血(8mL)からPBMCの分離と溶解までを キットを使って各機関で Lysateを凍結してXDxへ送付 qRT-PCR、専用ソフトによる解析 スコア化して、結果を医師に返送

510(K)

M. Performance Characteristics の記載項目

1. Analytical performance

- Precision/Reproducibility
- Linearity/assay reportable range
- Traceability, Stability, Expected values (controls, calibrators, or methods)
- Detection limit
- Analytical specificity
- Assay cut-off

2. Comparison studies

- Method comparison with predicate device
- Matrix comparison

3. Clinical studies

- Clinical Sensitivity
- Clinical specificity
- Other clinical supportive data (when a. and b. are not applicable)

4. Clinical cut-off

5. Expected values/Reference range

TISSUE OF ORIGIN TEST PERFORMANCE

Available Diagnosis	Positive Percent Agreement (% ratio, 95% CI)	Negative Percent Agreement (% ratio, 95% CI)	Non-Agreement (% ratio, 95% CI)	Indeterminate (% ratio, 95% CI)	Area Under ROC Curve
Bladder	78.6% (22/28) [59.0, 91.7]	100.0% (517/517) [99.3, 100.0]	14.3% (4/28) [4.0, 32.7]	7.1% (2/28) [0.9, 23.5]	0.996
Breast	94.1% (64/68) [85.6, 98.4]	98.5% (470/477) [97.0, 99.4]	5.9% (4/68) [1.6, 14.4]	< 0.1% (0/68) [0.0, 4.3]	0.979
Colorectal	94.6% (53/56) [85.1, 98.9]	99.2% (485/489) [97.9, 99.8]	5.4% (3/56) [1.1, 14.9]	< 0.1% (0/56) [0.0, 5.2]	0.980
Gastric	76.0% (19/25) [54.9, 90.6]	99.8% (519/520) [98.9, 100.0]	16.0% (4/25) [4.5, 36.1]	8.0% (2/25) [1.0, 26.0]	0.970
Hepatocellular	92.0% (23/25) [74.0, 99.0]	99.8% (519/520) [98.9, 100.0]	< 0.1% (0/25) [0.0, 11.3]	8.0% (2/25) [1.0, 26.0]	0.999
Kidney	94.9% (37/39) [82.7, 99.4]	99.8% (505/506) [98.9, 100.0]	2.6% (1/39) [0.1, 13.5]	2.6% (1/39) [0.1, 13.5]	0.975
Melanoma	84.6% (22/26) [65.1, 95.6]	99.8% (518/519) [98.9, 100.0]	7.7% (2/26) [0.9, 25.1]	7.7% (2/26) [0.9, 25.1]	0.999
Non-Hodgkin's Lymphoma	93.9% (31/33) [79.8, 99.3]	99.4% (509/512) [98.3, 99.9]	3.0% (1/33) [0.1, 15.8]	3.0% (1/33) [0.1, 15.8]	0.999
Non-small Cell Lung	90.3% (28/31) [74.2, 98.0]	98.6% (507/514) [97.2, 99.5]	6.5% (2/31) [0.8, 21.4]	3.2% (1/31) [0.1, 16.7]	0.991
Ovarian	94.2% (65/69) [85.8, 98.4]	99.4% (473/476) [98.2, 99.9]	2.9% (2/69) [0.4, 10.1]	2.9% (2/69) [0.4, 10.1]	0.996
Pancreas	76.0% (19/25) [54.9, 90.6]	99.8% (519/520) [98.9, 100.0]	16.0% (4/25) [4.5, 36.1]	8.0% (2/25) [1.0, 26.0]	0.953
Prostate	88.5% (23/26) [69.8, 97.6]	100.0% (519/519) [99.3, 100.0]	3.8% (1/26) [0.1, 19.6]	7.7% (2/26) [0.9, 25.1]	0.998
Soft-tissue Sarcoma	83.9% (26/31) [66.3, 94.5]	99.4% (511/514) [98.3, 99.9]	9.7% (3/31) [2.0, 25.8]	6.5% (2/31) [0.8, 21.4]	0.972
Testicular Germ Cell	82.1% (23/28) [63.1, 93.9]	100.0% (517/517) [99.3, 100.0]	3.6% (1/28) [0.1, 18.3]	14.3% (4/28) [4.0, 32.7]	0.999
Thyroid	91.4% (32/35) [76.9, 98.2]	99.6% (508/510) [98.6, 100.0]	5.7% (2/35) [0.7, 19.2]	2.9% (1/35) [0.1, 14.9]	0.976
Overall	89.4% (487/545) [86.5, 91.8]	99.6% (507/509) [98.6, 100.0]	6.2% (34/545) [4.4, 8.6]	4.4% (24/545) [2.8, 6.5]	

*As described in the Guide to Report Interpretation, if two or three SS are greater than or equal to 30, then one of those results indicates the likely tissue of origin. In the clinical validation study, this occurred in 11 of 545 specimens or 2%.

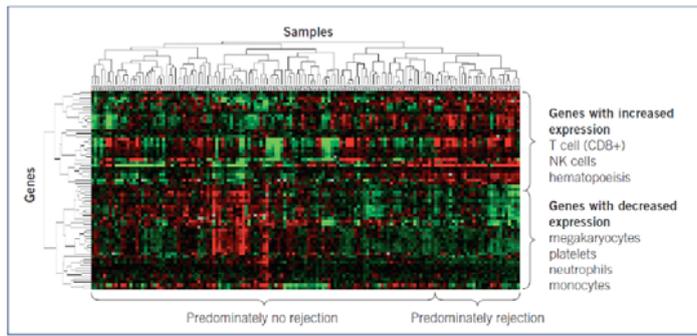
Results of Off-panel Tumor Types

Specimen	TOO Result	Total Specimens (n = 143)	TOO Reports with 2 SS ≥ 30	Total Indeterminates		Total False Positives	
		n		n	%I	n	%
Endometrium		49	2	1	2.0	48	98.0
Esophagus		28	1	6	21.4	22	78.6
Small Cell Lung		4	0	0	0.0	4	100.0
Ovarian Germ Cell		2	0	1	50.0	1	50.0
Sq. Head & Neck		18	0	7	38.9	11	61.1
Uterine cervix		42	2	6	14.3	36	85.7

15種以外の癌は間違って分類される

Specimen n	Total False Positives	Distribution of Similarity Scores ≥ 30 across the 15 tissues on the TISSUE OF ORIGIN panel															
		N	BL	BR	CO	GA	GC	LI	KI	LY	LU	ME	OV	PA	PR	SC	TH
Endometrium	48		1	2							2		43				2
Esophagus	22		1	1	4	14					2			1			
Small Cell Lung	4										4						
Ovarian Germ Cell	1						1										
Sq. Head & Neck	11		2	2							6						1
Uterine cervix	36		6	9	10				1	6		5				1	

Figure 4: Gene Identification Analysis



Heat Map of Rejection Genes: qRT-PCR measurements of 66 genes whose differential expression can distinguish between the presence and absence of acute cellular rejection. Red indicates increased expression of the gene. Green indicates decreased expression of the gene.

Figure 6: Rejection Pathways

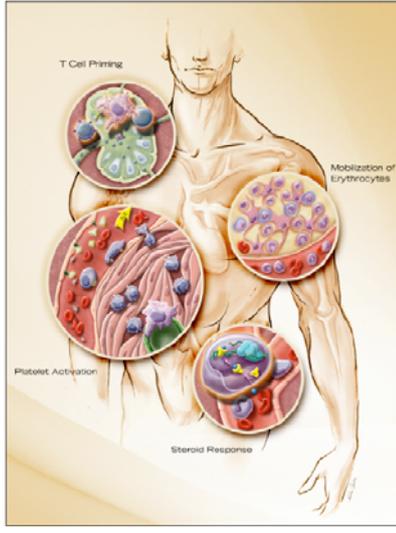


Figure 5: Targeting Specific Genes for the AlloMap Test

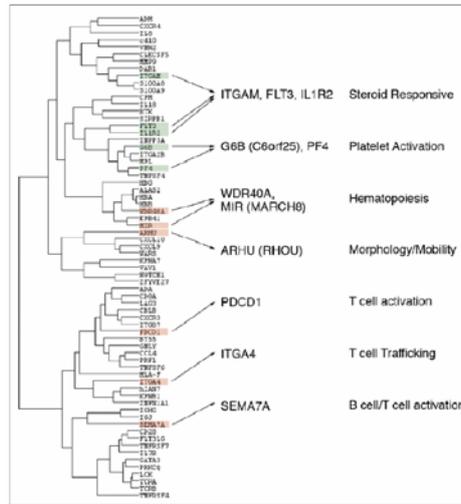


Figure 7: Differential Expression of AlloMap Genes in Rejection Samples

Pathway and Gene	Gene Expression Level
T cell priming <i>ITGA4</i>	↑
Integrin alpha-4 α subunit of VLA-4; involved in T cell trafficking and adhesion	↑
Programmed cell death T cell costimulatory molecular (inhibitory); CD28 family	↑
Proliferation and mobilization of erythrocytes <i>MARCH8</i>	↑
Cellular mediator of immune response (MIR) E3 ubiquitin ligase	↑
WDR40A WD repeat domain 40A Uncharacterized protein of the WD-repeat protein family	↑
Platelet activation <i>PF4</i>	↓
Platelet factor 4 Chemokine-like molecule expressed in platelets	↓
C6orf25 G6b inhibitory receptor Putative inhibitory receptor of the Ig superfamily expressed in platelets	↓
Steroid response <i>IL1R2</i>	↓
Interleukin-1 receptor type II IL-1 decoy receptor inhibits cytokine signaling; steroid-dependent expression	↓
ITGAM Integrin alpha-M α subunit of MAC-1; involved in cell trafficking	↓
FLT3 FMS-like tyrosine kinase Signaling molecule expressed in monocytes	↓
Unknown role <i>SEMA7A</i>	↑
Semaphorin 7A Expressed by T cells, B cells, and immature granulocytes	↑
RHOU Ras homolog gene family, member U Member of the Rho GTPase family involved in the modulation of cytoskeleton organization	↑

AlloMap®

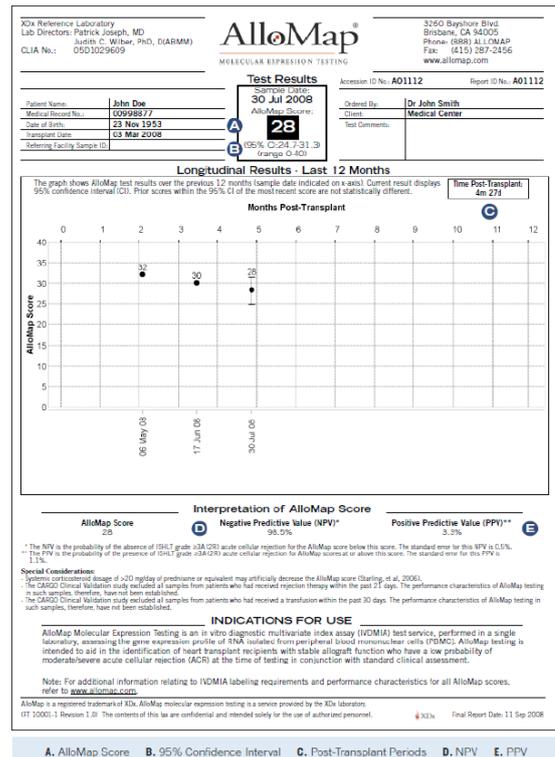
診断結果の表示

Table 2: AlloMap Testing Clinical Performance Characteristics**

Post-Transplant Period >2 - 6 months (n=166 samples)		AlloMap Score**	Post-Transplant Period >6 months (n=134 samples)	
NPV <3A(2R) ± SE	% Pts Below PPV ≥3A(2R) ± SE		NPV <3A(2R) ± SE	% Pts Below PPV ≥3A(2R) ± SE
97.9% ± 0.0%	100.0%	39	98.3% ± 0.0%	97.7%
97.9% ± 0.0%	100.0%	38	98.2% ± 0.0%	96.5%
98.1% ± 0.2%	97.8%	37	98.4% ± 0.2%	91.7%
98.1% ± 0.2%	97.3%	36	98.7% ± 0.3%	5.4% ± 3.2%
98.1% ± 0.2%	94.5%	35	98.7% ± 0.4%	4.0% ± 2.2%
98.2% ± 0.3%	91.7%	34	98.9% ± 0.4%	4.1% ± 1.7%
98.1% ± 0.3%	89.4%	33	99.1% ± 0.4%	3.8% ± 1.3%
98.0% ± 0.3%	85.6%	32	99.0% ± 0.5%	2.9% ± 0.9%
98.2% ± 0.1%	81.0%	31	99.8% ± 0.6%	54.1%
98.6% ± 0.4%	77.2%	30	98.7% ± 0.6%	50.6%
98.6% ± 0.4%	73.7%	29	99.0% ± 0.7%	40.8%
98.5% ± 0.5%	68.3%	28	98.9% ± 0.7%	39.1%
98.7% ± 0.5%	63.6%	27	98.7% ± 0.9%	31.6%
99.0% ± 0.4%	61.4%	26	100.0% ± 0.0%	26.8%
99.3% ± 0.5%	56.0%	25	100.0% ± 0.0%	22.1%
99.1% ± 0.6%	47.5%	24	100.0% ± 0.0%	18.4%
99.0% ± 0.6%	41.8%	23	100.0% ± 0.0%	2.0% ± 0.1%
98.9% ± 0.7%	38.8%	22	100.0% ± 0.0%	1.9% ± 0.1%
98.8% ± 0.8%	33.6%	21	100.0% ± 0.0%	9.8%
100.0% ± 0.0%	24.3%	20	100.0% ± 0.0%	8.1%
100.0% ± 0.0%	<22.4%	19	100.0% ± 0.0%	55.4%

* (KDx Laboratory Services Guide, 2008)

Figure 2: AlloMap Test Report



ORIGINAL ARTICLE

Gene-Expression Profiling for Rejection Surveillance after Cardiac Transplantation

Michael X. Pham, M.D., M.P.H., Jeffrey J. Teuteberg, M.D., Abdallah G. Kfoury, M.D., Randall C. Starling, M.D., M.P.H., Mario C. Deng, M.D., Thomas P. Cappola, M.D., Andrew Kao, M.D., Allen S. Anderson, M.D., William G. Cotts, M.D., Gregory A. Ewald, M.D., David A. Baran, M.D., Roberta C. Bogaev, M.D., Barbara Elashoff, M.S., Helen Baron, M.D., James Yee, M.D., Ph.D., and Hannah A. Valentine, M.D., for the **IMAGE Study Group***

ABSTRACT

BACKGROUND

Endomyocardial biopsy is the standard method of monitoring for rejection in recipients of a cardiac transplant. However, this procedure is uncomfortable, and there are risks associated with it. Gene-expression profiling of peripheral-blood specimens has been shown to correlate with the results of an endomyocardial biopsy.

METHODS

We randomly assigned 602 patients who had undergone cardiac transplantation 6 months to 5 years previously to be monitored for rejection with the use of gene-expression profiling or with the use of routine endomyocardial biopsies, in addition to clinical and echocardiographic assessment of graft function. We performed a noninferiority comparison of the two approaches with respect to the composite primary outcome of rejection with hemodynamic compromise, graft dysfunction due to other causes, death, or retransplantation.

From Stanford University Medical Center, Stanford (M.X.P., H.A.V.), and VA Palo Alto Health Care System, Palo Alto (M.X.P.) — both in California; the University of Pittsburgh Medical Center, Pittsburgh (J.J.T.); Intermountain Medical Center and Intermountain Healthcare, Salt Lake City (A.G.K.); Cleveland Clinic, Cleveland (R.C.S.); Columbia University Medical Center, New York (M.C.D.); Hospital of the University of Pennsylvania, Philadelphia (T.P.C.); Mid America Heart Institute, Saint Luke's Hospital, Kansas City, MO (A.K.); University of Chicago Medical Center (A.S.A.) and Northwestern University (W.G.C.) — both in Chicago; Washington University School of Medicine, St. Louis (G.A.E.); Newark Beth Israel Medical Center, Newark, NJ (D.A.B.); Texas Heart Institute, Houston (R.C.B.); and XDX, Brisbane, CA (B.E., H.B., J.Y.). Address reprint requests to Dr. Valentine at the Division of Cardiovascular Medicine, Stanford University School of Medicine, Falk CVRB, 300 Pasteur Dr., Stanford, CA 94305, or at hvalentine@stanford.edu.

RESULTS

During a median follow-up period of 19 months, patients who were monitored with gene-expression profiling and those who underwent routine biopsies had similar 2-year cumulative rates of the composite primary outcome (14.5% and 15.3%, respectively; hazard ratio with gene-expression profiling, 1.04; 95% confidence interval, 0.67 to 1.68). The 2-year rates of death from any cause were also similar in the two groups (6.3% and 5.5%, respectively; $P=0.82$). Patients who were monitored with the use of gene-expression profiling underwent fewer biopsies per person-year of follow-up than did patients who were monitored with the use of endomyocardial biopsies (0.5 vs. 3.0, $P<0.001$).

CONCLUSIONS

Among selected patients who had received a cardiac transplant more than 6 months previously and who were at a low risk for rejection, a strategy of monitoring for rejection that involved gene-expression profiling, as compared with routine biopsies, was not associated with an increased risk of serious adverse outcomes and resulted in the performance of significantly fewer biopsies. (ClinicalTrials.gov number, NCT00351559.)

- 12機関が参加した共同研究
- AllomapとBiopsyの予測性を評価
- 602人の患者を前向きに評価
- Acute Rejectionの予測率は同等

AlloMap[®]

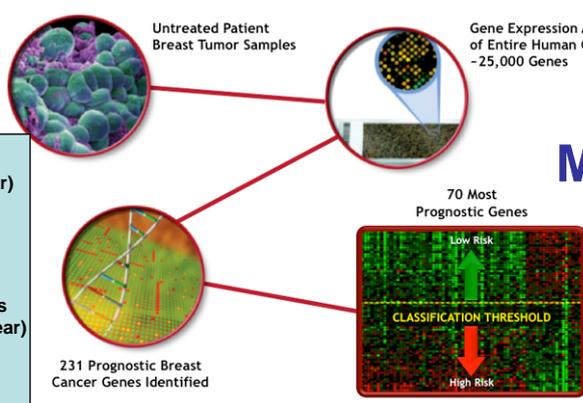
最新のバリデーション結果

Delivering better science through unbiased gene selection

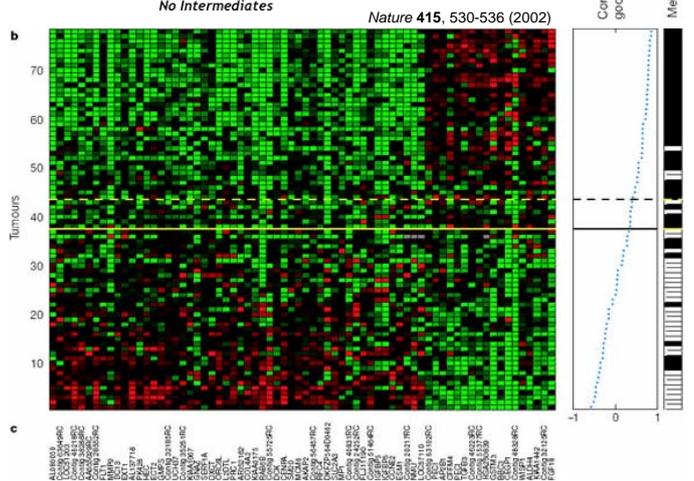
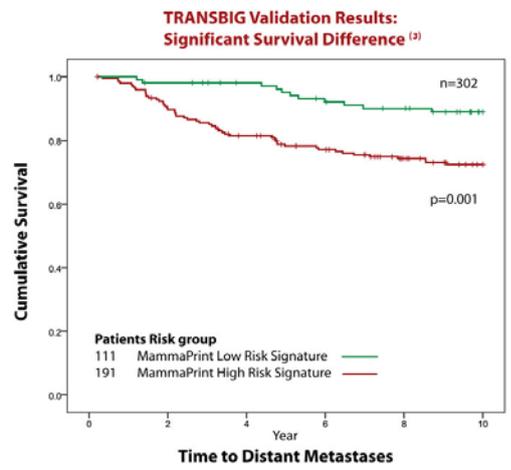


5-year time to distant metastases (no metastatic disease within 5 year)
 Low risk group: 0.95
 High risk group: 0.78

10-year time to distant metastases (no metastatic disease within 10 year)
 Low risk group: 0.90
 High risk group: 0.71



MammaPrint[®]



MammaPrintのご依頼について

国内での受託



検査予定が決定しましたら、
サンプリングキットを
ご用意ください。

- サンプリングキット内容物
1. 一体型チューブ
(チューブ大およびチューブ小)
 2. バイオプシーパンチ
 3. 匿名化IDラベル
 4. セーフティバッグ
 5. 発送用伝票
 6. 発送用封筒
 7. ユーザーズガイド
- 試料: 乳癌組織 3mm角以上5mm角以内



1 サンプリングキットを用いて
乳癌組織を採取します。



2 匿名化IDのラベルを貼付します。



3 弊社へご送付ください。



4 検査結果レポートを
試料受領後約3週間で
担当医師の先生へお届けします。

※検査解析は、オランダ・アムステルダムの
Agendia 社内のラボにて実施します。

※提出していただいた乳癌組織から十分な
RNA量が確保できない場合は、本検査サービ
スをご利用頂けない場合もございますので、
予めご了承ください。

• MammaPrintはオランダ Agendia 社の商標登録です。
• RNAチェックはDNAチップ研究所の商標登録(申請中)です。

株式会社DNAチップ研究所

About ColoPrint® for Stage II & III Colon Cancer Prognosis and Prediction

ColoPrint is a microarray-based gene expression profile for predicting the risk of distant recurrence of stage II and III colon cancer patients. Currently in the final stage of development, the combination of this profile and selected clinical factors could prove even more powerful and accurate in identifying high risk patients for more personalized clinical management.



Blueprint™: Development and Validation of the Breast Cancer Subtyping Profile

Classification of breast cancer into molecular subtypes may be important for accurate selection of therapy for patients as tumors with a seemingly similar biology but different molecular characteristics can have different outcomes. Blueprint is a multi-gene profile which has been developed for the classification of breast cancer into molecular subtypes. The profile separates tumors into Basal-type, Luminal-type and ERBB2-type subgroups. The addition of this molecular subtyping profile will further complement MammaPrint® by providing further improvement of outcome prediction.

さらに様々なアッセイが開発されている

DiscoverPrint® - A tool for the development of companion diagnostics in clinical trials of oncology therapeutics for the biotechnology and pharmaceutical industry

The importance of conducting focused, time- and cost-efficient clinical trials is crucial if potential new drugs are to be given the best possible chance of reaching the market. One of the biggest challenges in drug development, particularly in targeted oncology therapeutics, is to understand the differences in patient response rates. There have been many instances where novel therapeutics elicited a favorable response in some patients, while others showed no response.

About the Research Gene Panel

Agendia's TargetPrint Research Gene Panel is a microarray-based gene expression panel of 56 genes that have been identified as potential targets for prognosis and therapeutic response to a variety of therapies. Although these genes are still in a research phase, in the future they may hold the key to a greater level of personalized prognosis and therapy for breast cancer patients. Today, the Research Gene Panel is offered as a research tool and for Research Use Only (R.U.O.).

TargetPrint® Provides Quantification of ER, PR and HER2 Status

TargetPrint is a microarray-based gene expression test which offers a quantitative assessment of the patient's level of estrogen receptor (ER), progesterone receptor (PR) and HER2/neu overexpression within her breast cancer. TargetPrint is offered in conjunction with MammaPrint to provide the physician an even more complete basis for treatment decisions.

評価指標作成にあたり参考となる資料

- 平成18年度テーラーメイド医療用診断機器(DNAチップ)審査WG報告書
(特に、「遺伝子発現解析用DNAチップ等の製造販売承認申請において考慮すべき事項について」の項)
- DNA チップを用いた遺伝子型判定用診断薬に関する評価指標
(薬食機発第0404002 号)
- FDA ガイダンス
 - In Vitro Diagnostic Multivariate Index Assays (IVDMIA)
 - Class II Special Controls Guidance Documents (MammaPrint, AlloMap, Instrumentation for Clinical Multiplex Test Systems)
- 510(k) decision summary
 - MammaPrint, Pathwork, & AlloMap

想定される論点

- 目的 (必要性)
- ガイドラインのカバーする範囲
 - DNAチップ、RT-PCR、シーケンサー、(タンパク発現)・・・
→ (遺伝子)発現解析という観点で広く包括できるのが望ましい?
- 機器 or 診断薬
- 診断アルゴリズムの評価
- 開発ガイドラインと審査ガイドラインの関係
- ガイドラインの要求項目
- 臨床性能評価(有用性)
- LDTの取り扱い

Oncotype DXはなぜFDAで認可されていない？

- 承認申請していない。
- CLIAの認証を受け、LDTとして販売。
- 有効性は認められ、**保険適応**も受けている。



FDAで認可をとる必要(メリット)はない。



日本では？

Challenges of Translating Gene Expression Microarray Data into Clinically Useful Tests

Lisa M. McShane, Ph.D.

Biometric Research Branch
Division of Cancer Treatment and Diagnosis
National Cancer Institute

April 27, 2009

臨床性能評価の重要性

Combined Meeting of the Pediatric and Oncologic Drugs Advisory Committees

April 27, 2009

There is no substitute for a well-designed, COMPLETELY INDEPENDENT validation study.

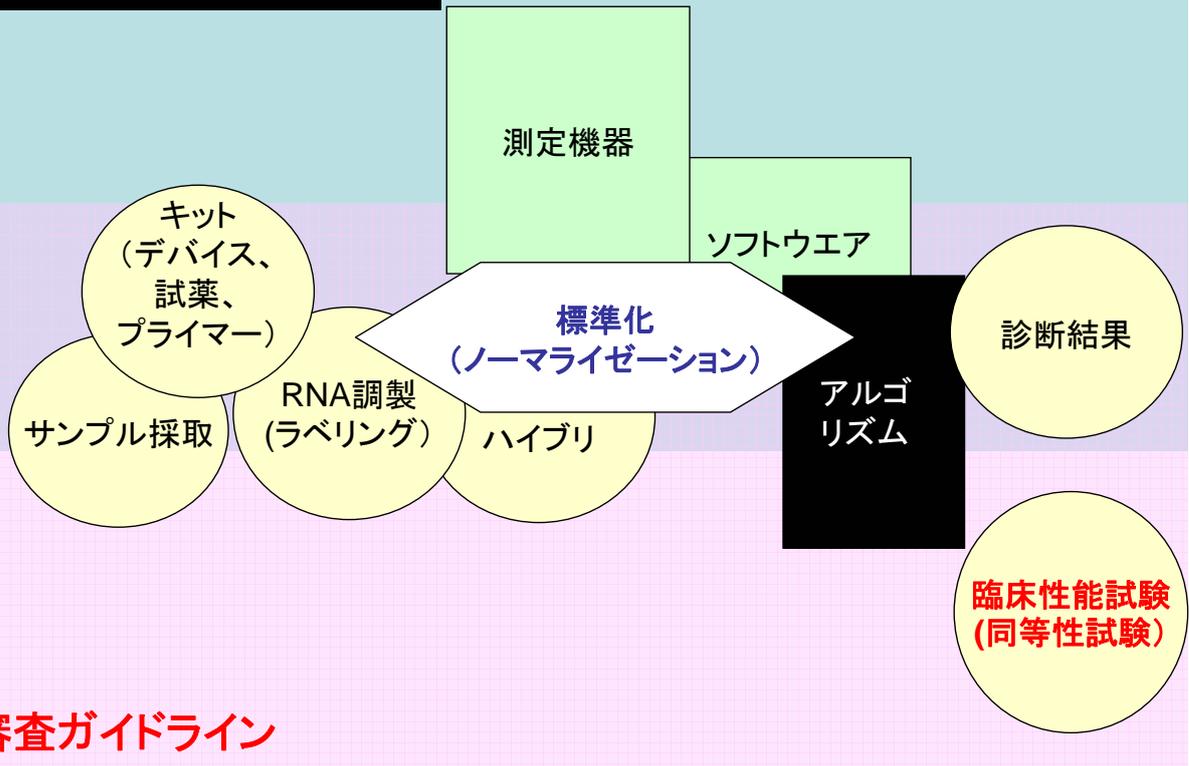
開発・審査ガイドラインの
相互関係と評価項目

開発ガイドライン

医療機器

体外診断薬

審査ガイドライン





Gene signature の検討 臨床の立場から

京都大学大学院医学研究科
外科学講座乳腺外科学
戸井雅和

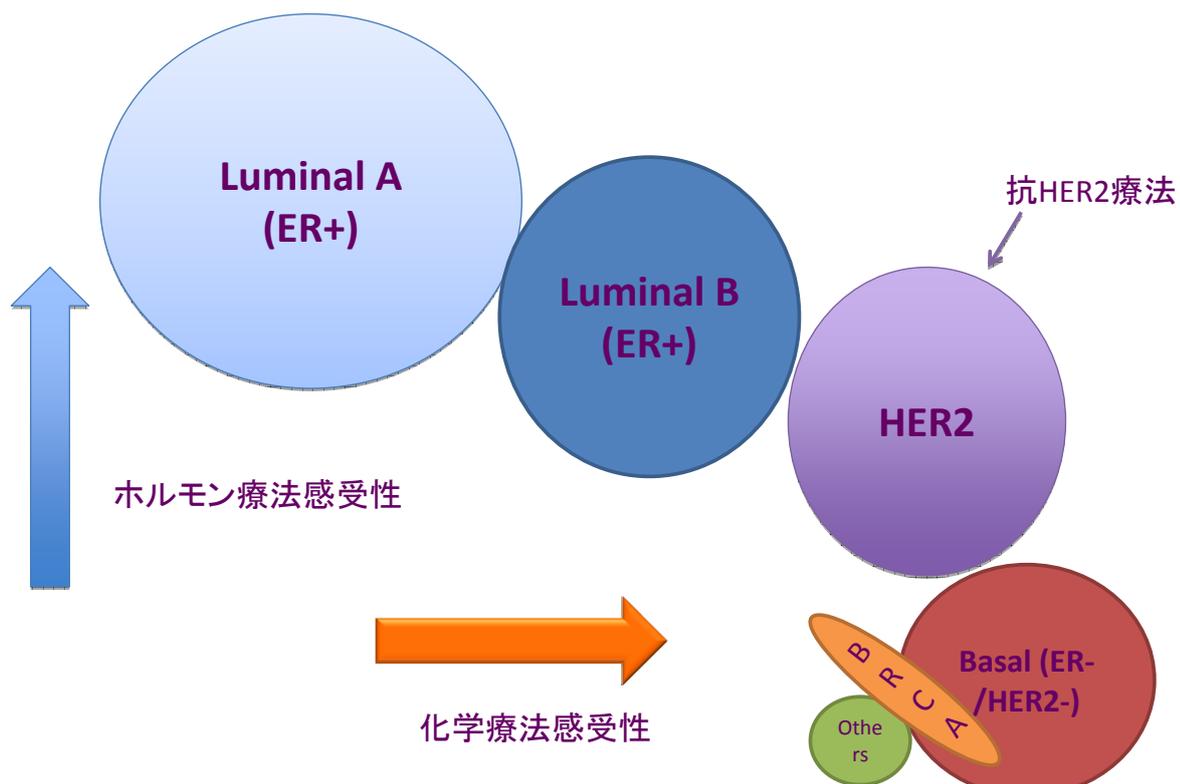
Oncotype DX[®]: 21 gene signature

- 日本における検討
- Heterogeneityに関して
- 医療経済学的評価について

Biomarker- Molecular profiling

分子プロファイル	遺伝子数
Supervised classification	
MammaPrint™	70
Rotterdam 76-gene signature	76
Wound-response signature	512
Genomic grade	97
Invasiveness gene signature	186
Oncotype DX®	21
Unsupervised classification	
Intrinsic subtype	~500

Intrinsic Subtypeと治療



21 gene signature

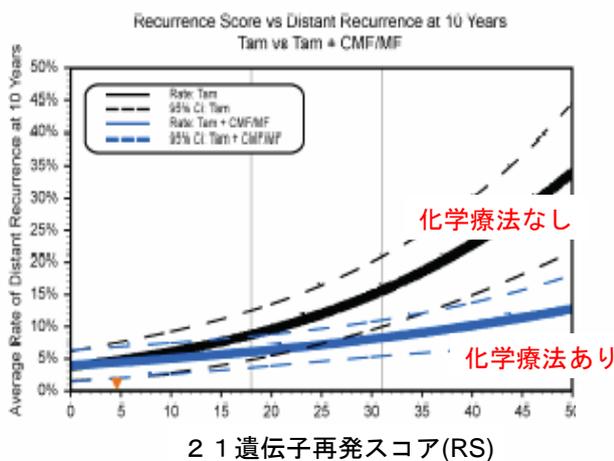
16 INFORMATIVE CANCER GENES AND 5 REFERENCE GENES

Estrogen	Proliferation	HER2	Invasion	Others	Reference
ER PR Bcl2 SCUBE2	Ki-67 STK15 Survivin Cyclin B1 MYBL2	GRB7 HER2	Stromelysin 3 Cathepsin L2	CD68 GSTM1 BAG1	Beta-actin GAPDH RPLPO GUS TFRC

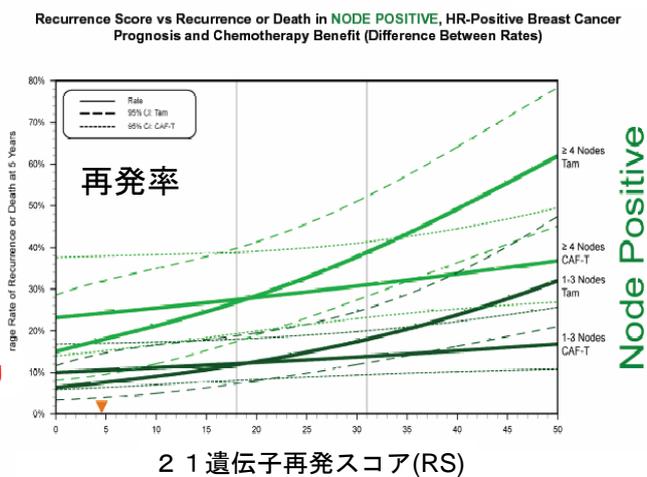
Risk category	Recurrence Score® value (0-100)
Low	< 18
Intermediate	18-30
High	≥ 31

Paik S, et al. *N Engl J Med.* 2004;351:2817.

ER+乳癌におけるRSと化学療法感受性 (21-gene signature)



リンパ節転移陰性 (Node Negative)



リンパ節転移陽性 (Node Positive)

JBCRG TR03: 適格／除外基準

- **Inclusion criteria**

- ER 陽性乳がん患者 (1993-1998)
- 全身補助療法はホルモン療法単独
- パラフィン包埋病理組織使用可能
- 十分な臨床情報

- **Exclusion criteria**

- 不適切な病理標本
- RNA (<275 ng) for RT-PCR

患者背景

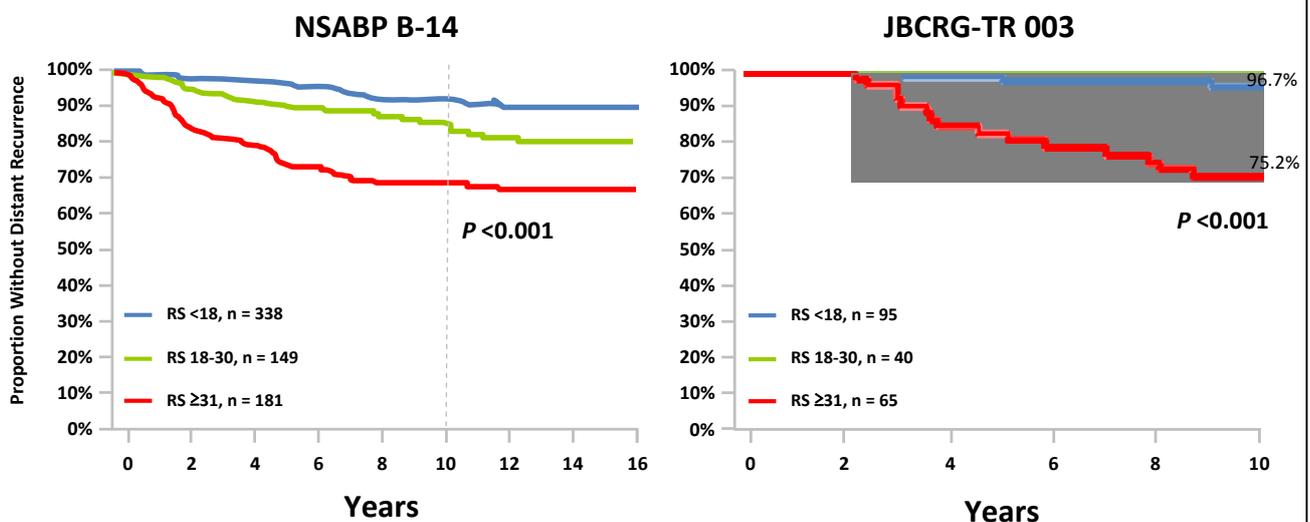
Characteristic	Node-Neg (n=200)	All (n=280)
年齢 <50 years	68 (34%)	96 (34%)
≥50 years	132 (66%)	184 (66%)
腫瘍径 ≤2 cm	92 (46%)	127 (45%)
>2 cm	108 (54%)	153 (55%)
核グレード		
1	30 (15%)	40 (14%)
2	80 (40%)	99 (35%)
3	36 (18%)	40 (14%)
Unavailable	54 (27%)	101 (36%)

Recurrence Score の分布

Recurrence Score	Node-Neg (n=200)	All (n=280)
Low (<18)	95 (48%)	131 (47%)
Intermediate (18-30)	40 (20%)	58 (21%)
High (≥31)	65 (33%)	91 (33%)



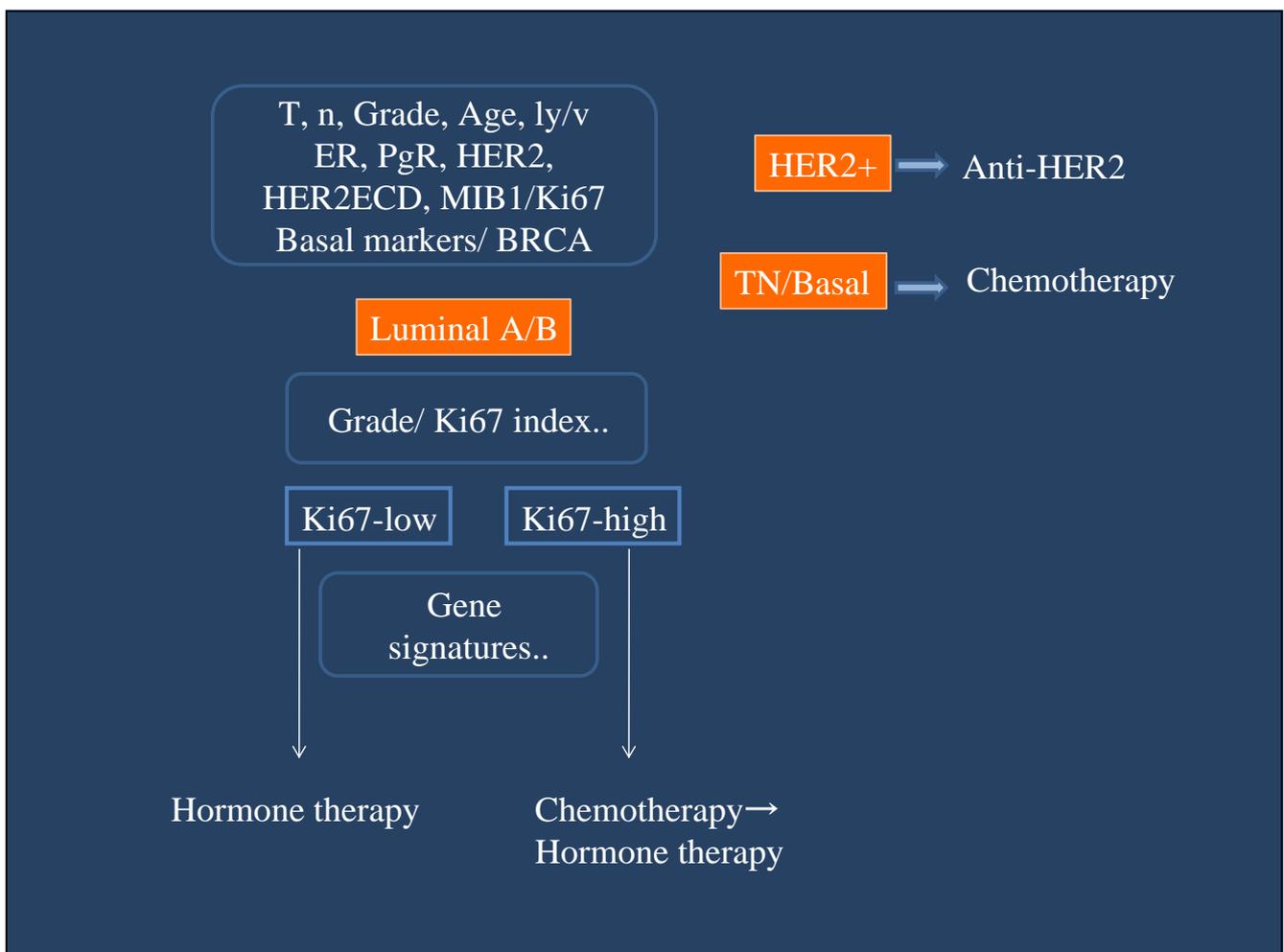
NSABP B-14 and JBCRG-TR 003



NSABP B14 & JBCRG TR03

		Low (<18)	Inter	High (≥31)
JBCRG TR03	No. of pts (%) N=200	48%	20%	33%
	10-year Distant recurrence (95%CI)	3.3 (1.1-10.0)	0 (N/A)	24.8 (15.7-37.8)
NSABP B-14	No. of pts (%) N=668	51%	22%	27%
	10-year Distant recurrence (95%CI)	6.8 (4.0-9.6)	14.3 (8.3-20.3)	30.5 (23.6-37.4)

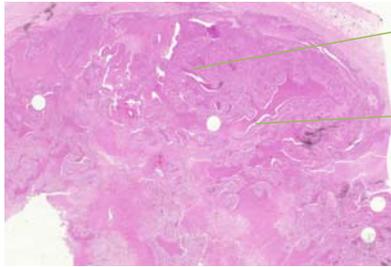
Courtesy S. NSABP to Paik



21 gene signature within block

Tumor blocks: 8 BC patients

1) 4枚の 600- μ m tissue microarray cores



Core 0

Core 1

Core 2

Core 3

RT-PCR Analysis

Day 1

Subsequent Day

2) 2枚の全セクション

▶ 全セクション 1

▶ 全セクション 2

Subsequent Day

Drury S, et al. *J Clin Pathol.* 2010;63(6):513-517.

21 gene signature within block

- Recurrence Score, ER, PR, HER2 :
大部分の症例で、Coreも全セクションもよく符合。
- Core間の違いの方が全セクション間の違いよりも大きい傾向あり。

Drury S, et al. *J Clin Pathol.* 2010;63(6):513-517.

医療経済学的評価(JBCRG-TR03)

ER+/ node-negative

Incremental cost-effectiveness ratio: ¥ 384,828 (US\$3,848)
per QALY

Social willingness-to-pay for one QALY gain from an
innovative medical intervention in Japan: ¥5,000,000/QALY
(US\$50,000/QALY)

Kondo M, et al. Breast Cancer Res Treat 2007

MammaPrint®に関する検討

- Ishitobi M et al. JJCO 2010
- 102 症例
- Node-negative
- ER+ and ER-
- Low 20%
- High 80% (63% in ER+)
- 生存成績における傾向を指摘

遺伝子検査標準化の現状

三重大学大学院医学系研究科
登 勉

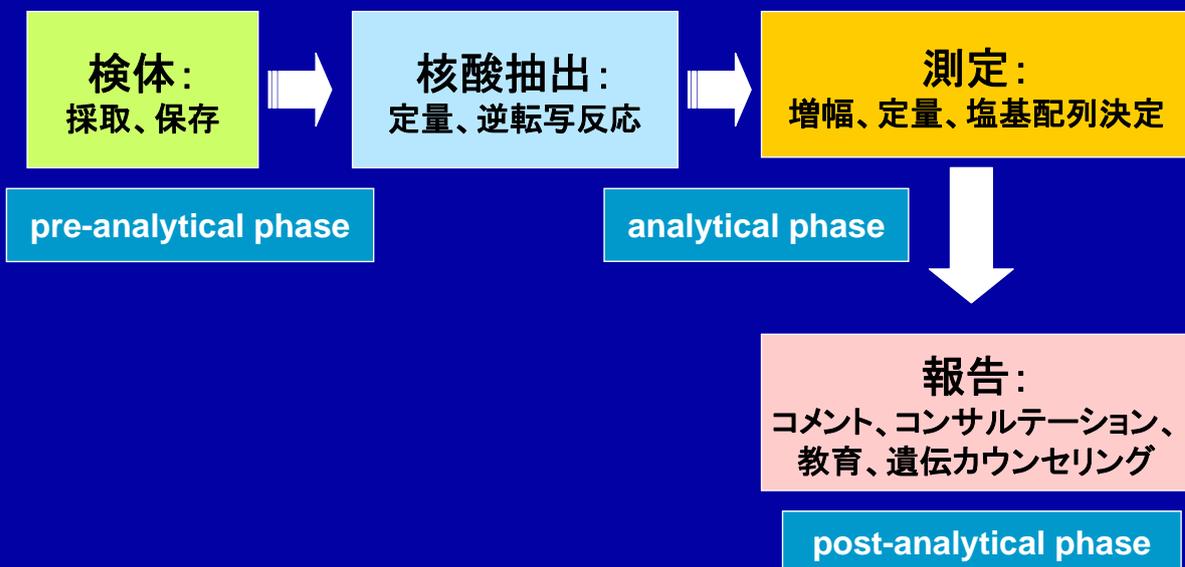
遺伝子関連検査

JCCLS日本臨床検査標準協議会
遺伝子関連検査標準化専門委員会

1. 病原体遺伝子検査
(病原体核酸検査)
2. ヒト体細胞遺伝子検査
3. ヒト遺伝学的検査
(生殖細胞系列遺伝子検査)

遺伝子検査のフローチャート

遺伝子検査は検体の採取・保存、核酸抽出、測定、結果報告の各ステップからなり、それぞれのステップに関する標準化が必要である。



Global Efforts

- Organization for Economic Cooperation and Development (OECD)
 - 2001 – Genetic testing: policy issues for the new millennium
 - 2007 – Guidelines for Quality Assurance in Molecular Genetic Testing
- ISO TC 212
 - 2005 – Japanese National Committee proposal for new standard
 - 2008 – Decision to incorporate unique features into ISO 15189:2012

OECD「分子遺伝学的検査における精度保証に関するガイドライン」

(Organization for Economic Cooperation and Development 経済協力開発機構)

趣旨：分子遺伝学的検査の質的保証を確保するため、政府、専門家機関及び臨床検査機関等における必要な取組事項を記載し、各国に政策提言を行う。

適用範囲：生殖細胞系列の遺伝学的検査

(個人識別や組織適合検査を目的、ガン細胞や病原体のDNA分析に関連したもの含まない)

OECDガイドラインの内容

項目	主な内容
検査の質保証	<ul style="list-style-type: none">・検査機関に対する認定システムの導入・認定を受けるインセンティブを付与・認定システムの国際的な協調、協力
技能試験	<ul style="list-style-type: none">・技能試験を、認定システムの中に含む・技能試験を受けるインセンティブを付与・検査機関に対するモニタリング
検査結果の報告	<ul style="list-style-type: none">・検査結果の取扱い・必要となる報告の内容
検査機関の職員に対する教育	<ul style="list-style-type: none">・検査に従事する者の水準や資格

Efforts in Europe

- Ongoing – Academic societies in laboratory medicine
 - 1994 – Nuffield Council on Bioethics, UK
 - 2000 – Genetics and health from Nuffield Council, UK
 - 2000 – Laboratory services from National Health Services, UK
 - 2000 – EU report on Human Genetics and New Technologies
 - 2003 – Gene direct from Human Genetic Commission, UK
 - 2003 – Towards QA and Harmonization of genetic testing services in the EU by Institute for Prospective Technological Studies
 - 2004 – Ethical, legal, and social aspects of genetic testing EC Expert Group
 - 2007 – OECD
 - 2009 – New German law “Human Genetic Examination Act”



EuroGentest



- *EuroGentest* is a network of excellence
- Funded by the *European Commission* in 2005.
- Aims to harmonize and improve the overall quality of genetic services in Europe.
- <http://www.eurogentest.org/>

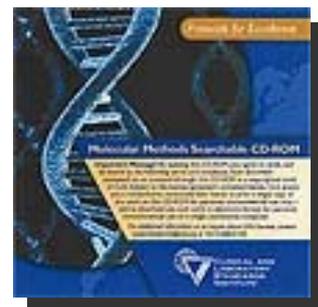


Efforts in USA

- Ongoing – Academic societies in laboratory medicine
 - 1997 – US Task Force on Genetic Testing
 - 2000 – Secretary’s Advisory Committee on Genetic Testing (SACGT)
 - 2004 – Johns Hopkins Genetics and Public Policy Center
 - 2007 – OECD
 - 2008 – Secretary’s Advisory Committee on Genetics, Health, and Society (SACGHS)
 - 2008 – The Genetic Information Nondiscrimination Act (GINA)
 - 2009 – SACGHS “Integration of Genetic Technologies Into Health Care and Public Health”
 - 2009 – CDC MMWR “Good Laboratory Practices for Molecular Genetic Testing for Heritable Disease and Conditions



CLSI Molecular Methods Guidelines



- *Molecular Diagnostic Methods for Genetic Diseases (MM1)*
- *Fluorescence In Situ Hybridization (FISH) Methods for Medical Genetics (MM7)*
- *Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine (MM9)*
- *Diagnostic Nucleic Acid Microarrays (MM12)*
- *Collection, Transport, Preparation, and Storage of Specimens for Molecular Methods (MM13)*
- *Use of External RNA Controls in Gene Expression Assays (MM16)*
- *Verification and Validation of Multiplex Nucleic Acid Assays (MM17)*

CLSI has **16 published guidelines** related to molecular methods and genetic testing and additional projects in development.

“DRAFT DOCUMENT. This draft CLSI document is not to be reproduced or circulated for any purpose other than review and comment. It is not to be considered either *final* or *published* and may not be quoted or referenced. **1 November 2010.”**

MM20-P
Vol. 0 No. 0

Quality Management for Molecular Genetic Testing; Proposed Guideline

Put Tagline Here *The tag line is one or two sentences describing the important features of the document. It should stimulate user interest.*

A guideline for global application developed through the Clinical and Laboratory Standards Institute consensus process.

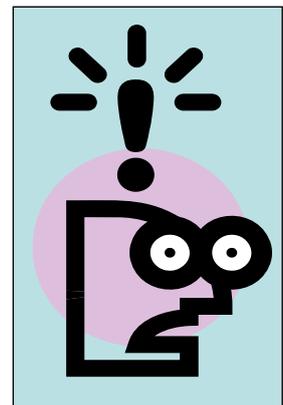


QC Materials for Genetic Testing

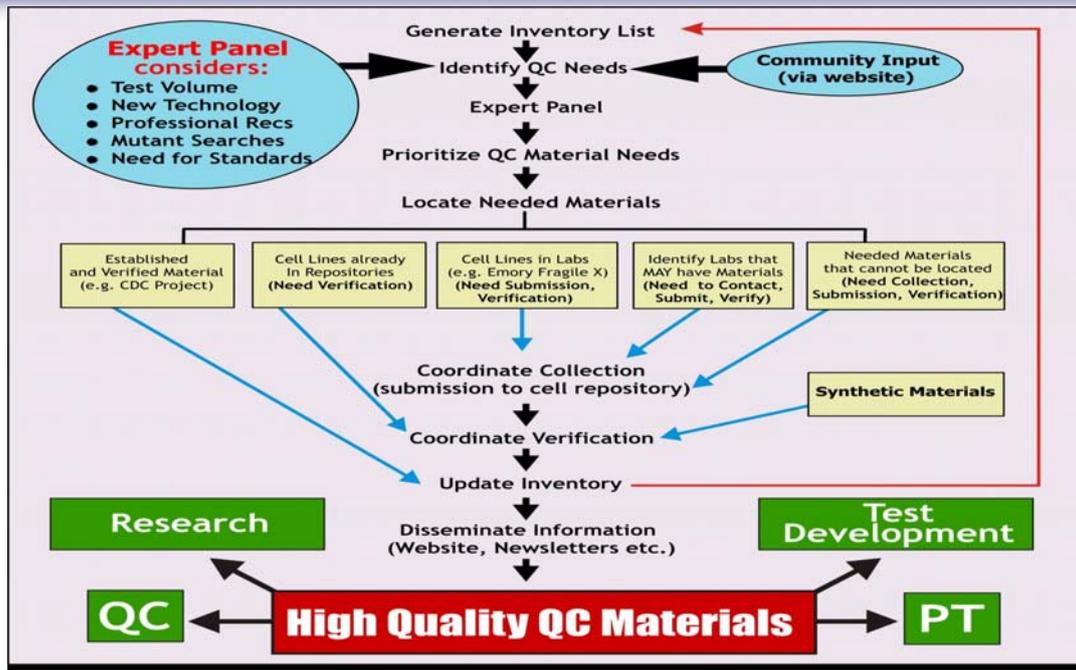
- Genetic testing volume is increasing
- Many new tests are being offered

However.....

- Positive controls are not available for most genetic tests!



GeT-RM Flowchart



www.cdc.gov/dls/genetics/rmmaterials/default.aspx



ACCE Project: Model Process for Collection, Evaluation, Interpretation, and Reporting

Disorder/Setting

- Analytic Validity
- Clinical Validity
- Clinical Utility
- ELSI



<http://www.cdc.gov/genomics/gtesting/ACCE/fbr.htm>



Japanese Committee of Clinical Laboratory Standards:JCCLS

日本臨床検査標準協議会（NPO法人）

1985年：設立モデル→米国臨床検査標準委員会
NCCLS(National Committee for Clinical Laboratory Standards:
名称変更→CLSI (Clinical and Laboratory Standards Institute)

特別会員9（官公庁）

正会員31（関連医学学会、協会、団体）

賛助会員46（企業）

特別助成団体（日本医師会）

個人会員

精度保証の必要性

- 医療機関、検査施設、民間企業等独自に検査法開発・実施
- 方法・機器の違い、測定者の技術格差→施設間差
- 分析的妥当性↓→臨床的妥当性↓→臨床的有用性の評価↓
- 有用性確立した遺伝学的検査の早期実用化↓
- 保険適応↓検査実施体制の維持・継続↓

遺伝子関連検査 現状マップ (国内版)

- 検査前フェーズ
- 検査フェーズ
- 様々な工程と役割
- 検査後フェーズ



検査前フェーズ	検査フェーズ			検査後フェーズ
	核酸検査 (感染症遺伝子)	遺伝子検査 (ヒト体細胞遺伝子)	遺伝学的検査 (生殖細胞系列遺伝子)	
検査依頼者	環境・食品等	白血病	単一遺伝子病	一般消費者
検査ネットワーク	民間企業	主治医	臨床遺伝専門医 遺伝カウンセラー	民間企業
個人情報保護	民間企業	医療・介護関係者個人情報(遺伝情報)保護ガイドライン	いでんネット/自動化学会	民間企業
検査適正化	民間企業	EBGT/エビデンスに基づく...	10医学会ガイドライン 日衛協指針	民間企業
監督指導	民間企業	検査適正化、精度管理委員会	指導監督医 精度管理責任者	民間企業
検査機関	民間企業	検査専門医/検査管理医 病院検査室(保険医療機関)	衛生検査所	民間企業
測定者	民間企業	ISO, CAP認定/認証	研究機関	民間企業
測定方法	民間企業	臨床検査技師等	研究者等	民間企業
測定・機器試薬	民間企業	臨床細胞遺伝学認定士 認定染色体技師	遺伝子分析科学認定士	民間企業
標準物質	民間企業	日本自動化学会 マニュアル	染色体検査 ガイドライン	民間企業
外部精度管理	民間企業	自動化システム 二次的(メーカー) JBA	個人遺伝情報取扱 協議会の自主基準	民間企業
報告(書)	民間企業	日臨床サーベイ 自動化学会 サーベイ	CAPサーベイ?	民間企業
検査利用	民間企業	臨床遺伝専門医 遺伝カウンセラー	個人遺伝情報取扱 協議会の自主基準	民間企業
	食品産業等	一般診療(保険診療)	遺伝診療	健康(産業)



遺伝子検査の開発と臨床応用における課題

- **サイエンス/技術面：**
開発ターゲットの探索と選定が容易ではない
- **知的財産権：**
開発コストが膨大となる
- **ヒトゲノム検査の実施頻度：**
市場性、利益性が乏しい
- **臨床的有用性の評価：**
製品化までに時間がかかる
- **診療報酬制度上の課題：**
検査実施施設の経済的メリットが明確でない

平成22年4月 診療報酬改定(遺伝子検査関連項目) (薬事承認済診断薬は赤字で示した。)

D006-2 血液細胞核酸増幅同定検査(造血器腫瘍核酸増幅同定検査) 2,000点

PCR法、LCR法またはサザンプロット法による。(Major bcr-abl、PML-RARA など)

D006-3 Major bcr-abl mRNA核酸増幅検査 1,200点

DNAプローブ「FR」Amp-CML(TMA法)

D006-4 遺伝学的検査 4,000点

PCR法、DNAシーケンス法、FISH法またはサザンプロット法による。

D006-6 免疫関連遺伝子再構成 2,400点

PCR法、LCR法またはサザンプロット法による

D006-7 WT1 mRNA核酸増幅検査、サイトケラチン(CK)19mRNA、UDPグルクロン酸転移酵素遺伝子多型 2,000点

リアルタイムRT-PCR法(WT1 mRNA測定キット「オーツカ」)、OSNA法(LAMP法によるCK19 mRNAの増幅・定量)、インペーダー法(インペーダー-UGT1A1アッセイ)

D004 -2 悪性腫瘍組織検査

1 悪性腫瘍遺伝子検査 2,000点

固形腫瘍の腫瘍細胞を検体とし、PCR法、SSCP法、RFLP法等を用いる。**肺癌及び大腸癌におけるEGFR遺伝子検査又はK-ras遺伝子検査、悪性骨軟部組織腫瘍におけるEWS-Fli1、TLS-CHOP、又はSYT-SSX遺伝子検査、消化管間葉系腫瘍におけるc-kit遺伝子検査、家族性非ポリポージス大腸癌におけるマイクロサテライト不安定性検査、悪性黒色腫センチネルリンパ節生検に係る遺伝子検査**

薬事承認された
診断薬のない遺
伝子検査が保険
収載されている

遺伝子検査の開発から臨床応用への新しいルート案

