

平成 22 年度補正研究加速枠戦略的基盤技術高度化支援事業
「血液診断バイオマーカーのための定量比較 LC-MS ロボットに
おける組込みソフトウェアの開発」

研究開発成果等報告書

平成 23 年 9 月

委託者 関東経済産業局

委託先 株式会社 MCBI

第1章 研究開発の概要

1-1 研究開発の背景・研究目的及び目標

1-1-1 研究開発の背景

本研究開発は、特定ものづくり基盤技術において達成しようとする高度化の目標、特定ものづくり基盤技術高度化指針のうち、以下の項目に対応している。

- ・組込みソフトウェアに係る技術に関する事項
- ・組込みソフトウェアに係る技術において達成すべき高度化目標

1) 川下製造業者の抱える課題及び要請

これについて、対応する事項ごとに研究開発の背景を記述する。

(1) 産業機械及び産業用ロボットに関する事項

- ・新たな適合分野への対応

近年医療分野においてもいわゆる「医療用ロボット」が実用化されつつある。大手の病院や検査センターでは、血液から血清・血漿を分離し、通常の生化学検査装置でほぼオートメーションで処理と計測が行うことができるようになってきているが、これは医療用ロボットの一つである。さらに最近では、外科手術をサポートする手術支援ロボットが実用化され、これも産業用ロボットの新たな適合領域となっている。一方、これらの装置を製造する川下製造業者（OEM 供給業者を含む）では、医学研究の高度化・高速化に伴って次々に生み出される新技術への対応も必要とされるようになってきている。その中でも、生体試料の成分を分離する液体クロマトグラフィー（LC）と田中耕一氏（島津製作所）がノーベル賞を受賞した質量分析（MS）技術の組合せ（LC-MS）は血液中のタンパク質・ペプチドの検出に多大の成果を挙げてきた。最近ではこの計測技術が診断分野に応用され、従来にない高精度の診断を可能にした。しかしながら、現状では研究室レベルの応用に留まり、臨床診断の現場で広く活用されるに至っていない。その大きな理由の一つがこの技術の広範囲な活用を可能にする組込みソフトウェアの不足である。具体的には、本研究開発で取り組む課題として、計測装置である高速 LC 装置と MS 装置が正しく稼働しているかをテスト・検証する技術、質量分析データから計測対象物を定量する技術、ヒトの健常と病気の状態における検査データを比較・評価し、医療機関との情報交換を行う技術があげられる。これら技術を充足するソフトが組込まれた LC-MS 装置が求められる LC-MS ロボットである。

(2) 川下分野横断的な共通の課題・ニーズに関する事項

- ・品質の向上

当該新規適応分野である医療用ロボットでは特に高い品質・信頼性が求められる。そのため、前述のように血液検査ロボットが正しく稼働しているかを随時テスト・検証することが可能でなければならない。プロセスを可視化し、管理者がそれを利用して直ちに対応できるような組込みソフトウェアが求められている。

- ・国際規格への対応

国際規格に基づく品質保証が必要である。研究分野ではタンパク質・ペプチドに関する質量分析データについての国際規格を国際学会（Human Proteome Organisation, HUPO）が定めている。産業

レベルではデータ処理についての国際規格がないため、研究分野に準じた国際規格とこれに基づいたデータのみを扱うようなソフトウェアが求められている。

2) 研究開発の背景

すでに述べたように、生体試料の成分を分離する LC と MS 技術の組合せ (LC-MS) は血液中のタンパク質・ペプチドの検出に多大の成果を挙げ、従来にない高精度の診断を可能にしたが、現状では研究室レベルの応用に留まっている。一方、申請者は LC-MS データを解析するソフトウェア DeView™ をすでに開発し、販売するに至っている。これは現時点で市販される唯一の複数群サンプルのデータを処理できる LC-MS 解析ソフトウェアであって、LC-MS データを一括して解析し、群間の差異を検出することができるものである。しかしながら、本ソフトウェアはいわばスタンドアローンのソフトウェアであって、以下の課題を解決する必要性が具体化してきていた。

(ア) LC-MS ロボットの信頼性の向上

(イ) LC-MS ロボットの高性能化および機能の向上

(ウ) LC-MS ロボットの情報を利用する医療機関の情報システムとの連携

このうち、(ア) は「平成 21 年度戦略的基盤技術高度化支援事業 (補正予算)」に採択され、検討が終了して解決済 (最終報告済) である。(イ) については自力での検討が途中まで終了し、本研究開発において (ウ) とともに本委託事業として継続実施した。

申請者が特許権者かつ実施権者であり、本計画で活用する特許は、特許第 4597741 号「タンパク質、部分タンパク質および/もしくは部分ペプチド、それらのプロファイルに基づく体外診断システム」である。本特許には国内外の他者による類似特許はなく、かつ他特許に抵触する可能性はない。また、PCT 出願 PCT/JP2009/004657 (WO2010/032458 : 100325 公開)「新規非アルコール性脂肪性肝疾患バイオマーカーおよび該バイオマーカーを用いた非アルコール性脂肪性肝疾患の検出方法」は、本年 4 月までに米国および欧州への国内移行出願を行った。非アルコール性脂肪性肝疾患は、従来確実な診断法がなく、本特許出願の血清を用いる方法は LC-MS ロボットによる分析に適しており、近年罹患率が増加している本疾患の早期診断に有用である。

1-1-2 研究目的及び目標

以下に上記の研究開発の背景から明らかになった研究目的及び目標を記載する。

1. LC-MS ロボットの信頼性の向上

1-1. LC におけるデータの再現性の確保

LC におけるデータの再現性の確保のために、実サンプルについての検査前後に標準サンプルのテストランを行ったとき、申請者の組込みソフトウェアは「良・不良」の判断とその根拠を提示することができる。

1-2. MS における質量分析データの正確な取扱いの保証

本研究開発の成果物である組込みソフトウェアは、質量分析データが国際学会 (Human Proteome Organisation, HUPO) が定めている国際規格に準じているかどうかを検証できる。

1-3. 標準サンプルを用いた LC-MS ロボットのパフォーマンスの検証

標準サンプルを用いて、LC-MS ロボットのパフォーマンスの総合的かつシームレスな検証が可能となる。

2. LC-MS ロボットの高性能化および機能の向上

2-1. 組込みソフトウェアによるデータ処理時間の短縮

LC によって得られた各分画の質量分析データの組込みソフトウェアによる処理の目標値は一日あたり 500-1,000 サンプルである。

2-2. サンプルから得られる情報の可視化とこれによる比較解析機能の構築

組込みソフトウェアによる可視化画像により、容易にサンプルから得られる情報を使用者に理解できるようにする。複数群の比較解析結果についても可視化して使用者の便宜を図る。

2-3. 複数のサンプルに関する統計解析機能の付与

複数群の臨床的診断の妥当性について最も有用な統計解析法は ROC 曲線 (receiver operating characteristic curve) を用いるものである。本研究開発の組込みソフトウェアには、この機能を付与する。

3. LC-MS ロボットの情報を利用する機関の情報システムとの連携をサポートする技術の構築

3-1. 医療機関への情報提示方法の構築

LC-MS ロボット内のデータから疾患カテゴリー内の各診断名に対する可能性(推定される確率)の情報を提示できるようにする。

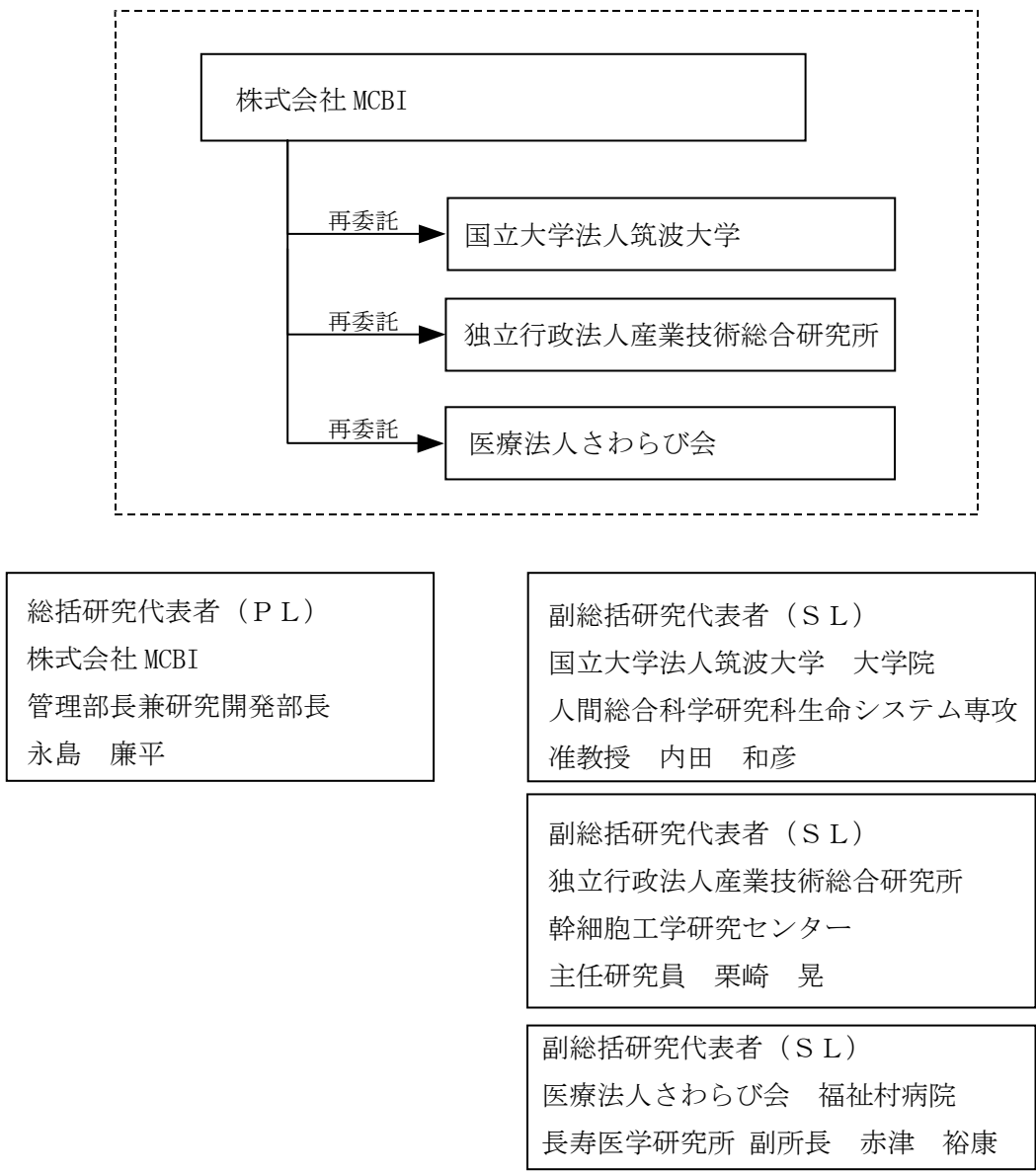
3-2. 患者の臨床情報を記録した電子カルテとの連携のための技術開発

電子カルテの情報と LC-MS ロボットの情報を組み合わせ、医師に対して自動的に診断(推定される確率)の情報を出力できる組込みソフトウェアの構築を目指す。

1－2 研究体制

1－2－1 研究組織及び管理体制

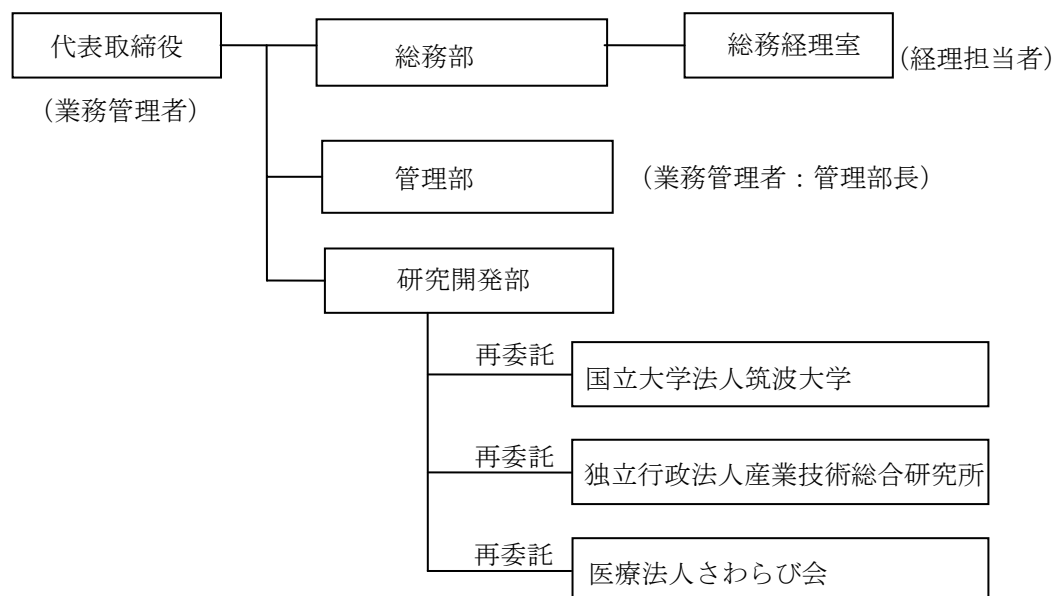
1) 研究組織（全体）



2) 管理体制

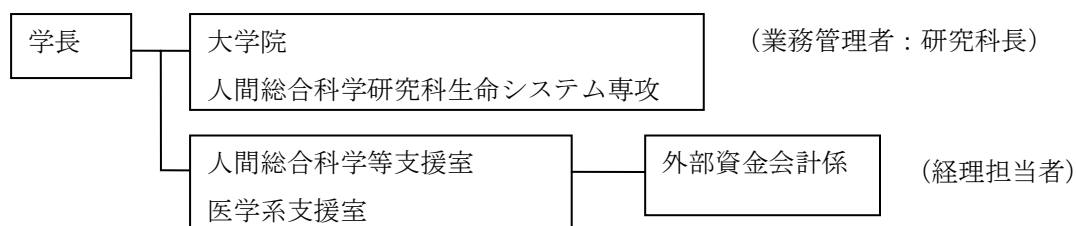
①事業管理機関

[株式会社 MCBI]

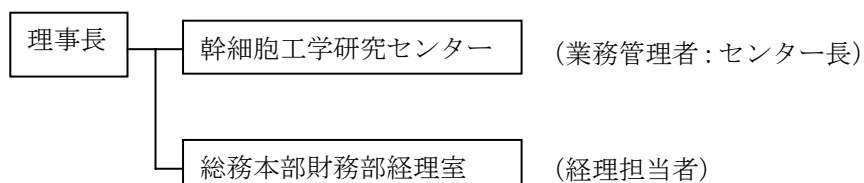


② 再委託先

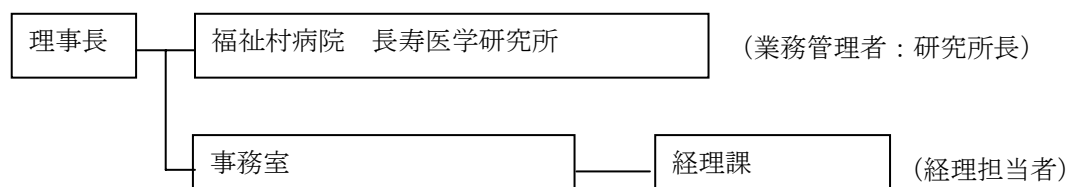
[国立大学法人筑波大学]



[独立行政法人産業技術総合研究所]



[医療法人さわらび会]



1－2－2 管理員及び研究員（以下の表の実施内容（番号）は1－3 成果概要を参照）

【事業管理機関】株式会社 MCBI

① 管理員

氏名	所属・役職
永島 廉平	管理部長兼研究開発部長
酒井 弘枝	総務部総務経理室長

② 研究員

氏名	所属・役職
永島 廉平（再）	管理部長兼研究開発部長
目野 浩二	研究開発部・室長
鈴木 秀昭	研究開発部・室長 兼 事業部・部長代理
ディサヤケ・サント・クマ	研究開発部・研究員
是永 龍巳	研究開発部・研究員
川頭 信之	研究開発部・研究員

【再委託先】（研究員）

国立大学法人筑波大学

氏名	所属・役職
内田 和彦	大学院 人間総合科学研究科生命システム専攻 准教授

独立行政法人産業技術総合研究所

氏名	所属・役職
栗崎 晃	幹細胞工学研究センター 主任研究員

医療法人さわらび会

氏名	所属・役職
赤津 裕康	福祉村病院 長寿医学研究所 副所長

1－2－3 経理担当者及び業務管理者の所属、氏名

（事業管理機関）

株式会社 MCBI

（経理担当者）	総務部総務経理室長	酒井 弘枝
（業務管理者）	代表取締役	内田 和彦
	管理部長	永島 廉平

（再委託先）

国立大学法人筑波大学

（経理担当者）	人間総合科学等支援室	医学系支援室	外部資金会計係長	菅谷 哲宏
---------	------------	--------	----------	-------

(業務管理者) 大学院 人間総合科学研究科 研究科長 五十殿 利治

独立行政法人産業技術総合研究所

(経理担当者) 総務本部財務部経理室 出納責任者 経理室長 井佐 好雄

(業務管理者) 幹細胞工学研究センター長 浅島 誠

医療法人さわらび会

(経理担当者) 事務室 経理課 長寿医学研究所事務 園田尚子

(業務管理者) 福祉村病院長寿医学研究所長 奥田 研爾

1-2-4 他からの指導・協力者

研究開発推進委員会 委員

氏名	所属・役職	備考
永島 廉平 内田 和彦	株式会社 MCBI 管理部長兼研究開発部長 国立大学法人筑波大学 大学院 人間総合科学研究科生命システム専攻 准教授	委 PL SL
栗崎 晃	独立行政法人産業技術総合研究所 幹細胞工学研究センター 主任研究員	SL
赤津 裕康	医療法人さわらび会 福祉村病院 長寿医学研究所 副所長	SL
朝田 隆	国立大学法人筑波大学 大学院 人間総合科学研究科疾患制御医学専攻 教授	アドバイザー
高野 純	株式会社島津製作所 経営戦略室兼基盤技術研究所	アドバイザー(旅費のみ)

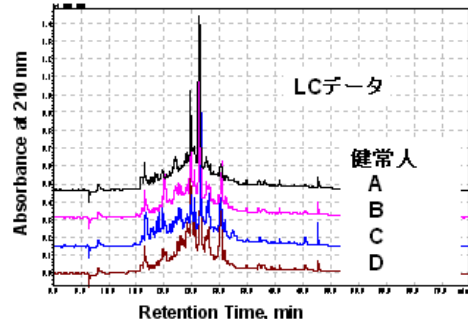
1-3 成果概要

「1-1-2 研究目的及び目標」に記載した研究項目番号にしたがって記述する。

平成 22 年度事業を開始する前の研究開発

1-1. LC におけるデータの再現性の確保

純水および測定対象とする生体試料をサンプルとして LC 装置に負荷し、期待外の異常が何ら生じないこと、および繰り返し試行でのバラツキが少なく一定範囲内であることを確認する作業を自動的に行う組込みソフトウェアのためのプログラムを構築し、実データの処理を行って期待通りの確認を行うことができた (図 1)。



1-2. MS における質量分析データの正確な取扱いの保証

本研究開発の組込みソフトウェアに関する国際規格は、①タンパク質・ペプチド特定のための retention time (RT) を限定できること、②質量許容誤差範囲内に MS データが収まることである。

LC-MS 装置から得られる全てのイオンの強度の和からなる TIC からバックグラウンド信号を除き、各溶出分画 (RT と 1:1 に対応) の最大信号を溶出分画に対してプロットしたものを base peak chromatogram (BPC) と呼ぶが、BPC の分析によって、MS データの正確な取扱いの保証が可能である。標準合成ペプチドを用いて LC-MS 装置の RT および質量に関するパフォーマンスを検証するプログラムを構築し、実データについて質量分析データの正確な取扱いの保証を行うことができた (図 2)。

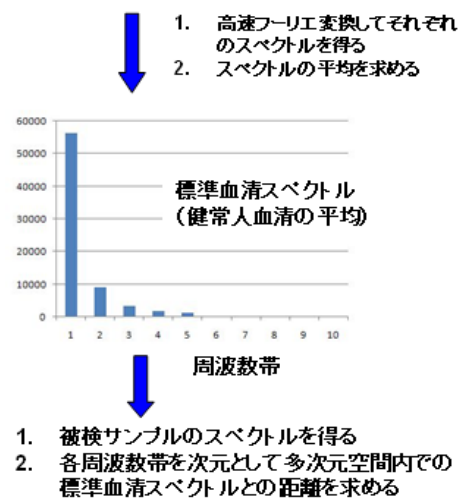


図1 血清サンプルのLCデータの合否判定

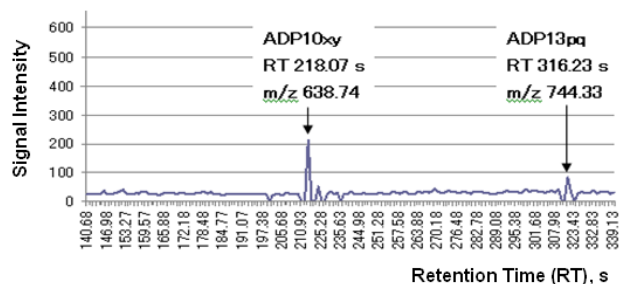


図2 MSにおける質量分析データのBPCによる保証

1-3. 標準サンプルを用いた LC-MS ロボットのパフォーマンスの検証

シームレスに LC-MS としてのパフォーマンスを検証できるプログラムを有する組込みソフトウェアを以下のように構築し、実データについて期待通りの動作を行うことを確認した。

- 1) 複数の健康人血清についての LC-MS 測定データ (mzXML 形式ファイル) を入力する。
- 2) データを内部標準によって標準化し、ピーク強度が一定以上かつ連続する一定数分画以上で出現するピークからなる「標準化 m/z -RT データ」を得る。
- 3) 上記データについてユークリッド平方距離の閾値 D2t の下でクラスタリングを行う。
- 4) 複数の健康人血清についての標準化 m/z -RT データに共通するクラスターをランドマークとして、

重心の m/z 値および RT 値を登録する。これを「登録標準ランドマーク」と呼ぶ。

5) 被検患者血清の LC-MS 測定データを入力してステップ 2)、3) を行い、得られた各クラスターの重心が登録標準ランドマークに一定の m/z 値および RT 値の範囲内で一致するかどうか調べる。

6) 上記 5) における一致率が一定の閾値以上である場合に合格とする。

2-2. サンプルから得られる情報の可視化とこれによる比較解析機能の構築

これまでに開発した組込みソフトウェアは、対象サンプルについての①試験の操作条件、②得られた生データなどの一次的情報を蓄積して必要時にはいつでも引き出せるようにし、理解に便利のように可視化を可能にした。これらはツリー図による可視化を行った。データ解析の部分では、異なる患者群について別々の色づけを行ったドット図、あるいはバーグラフを用意して比較解析を容易にした（図 3）。

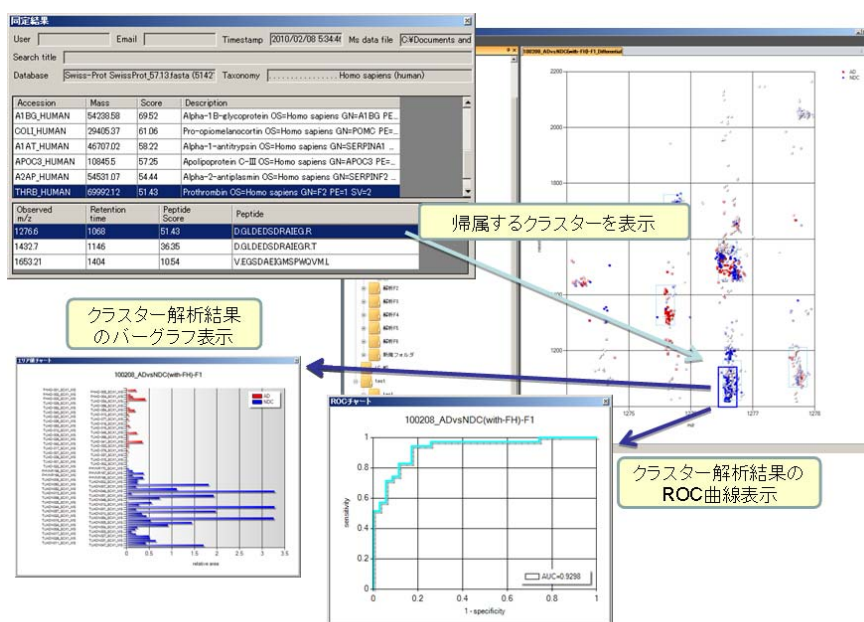


図3. 情報の可視化と比較解析機能および複数のサンプルに関する統計解析機能の付与

2-3. 複数のサンプルに関する統計解析機能の付与

LC-MS ロボットによる疾患の診断では検査結果が数値で表わされるが、ROC 曲線解析における感度と特異度は陽性と陰性の境界値（カットオフ値）をいくつに設定するかによって変わってくる。様々なカットオフ値における感度を縦軸に、1-特異度を横軸にプロットした曲線が ROC 曲線である。ここで感度とは疾患がある患者のうち検査結果が陽性であった率、特異度とは疾患のない患者のうち、検査結果が陰性であった率を意味する。診断法としては、疾患があるとき検査が陽性である程度（感度）が高いほうが望ましく、かつ、疾患がなければ検査が陰性である程度（特異度）も高いほうが望ましい。診断の有用性の指標としては ROC 曲線の下面積（area under the curve, AUC : 最大値は 1）が用いられる。本組込みソフトウェアにはこのような ROC 曲線による解析機能を付与した（図 3）。

平成 22 年度事業における研究開発

「平成 22 年度戦略的基盤技術高度化支援事業「血液診断バイオマーカーのための定量比較 LC-MS ロボットにおける組込みソフトウェアの開発」に係る委託業務実施計画書」(22fy 予備費)の成果概要を示す。

2. LC-MS ロボットの高性能化および機能の向上

2-1. 組込みソフトウェアによるデータ処理時間の短縮

現在のデータ処理は、毎回分析条件にしたがってマニュアルでパラメタを設定している。そこで、どのようなパラメタの組合せが繁用されているかを調査し、その結果から複数の繁用メニューを用意することにより、パラメタ設定に掛かる時間を短縮した。また、LC の分画が速く出すぎて、質量分析が間に合わないことがある。この場合、質量分析のスピードに調和して LC 分画が溶出するように組込みソフトウェアが制御できるものとした。得られた情報の信頼性が疑われるために、LC-MS ロボットによる解析のやり直しが必要なことがある。解析のやり直しが起きないようにするために、すでに 1-1~1-3 で構築した信頼性向上策を活用した。これらの手段を用いて、目標の一日あたり 500・1,000 サンプルの処理を可能にした。

2-2. サンプルから得られる情報の可視化とこれによる比較解析機能の構築（とくにサンプルの属性および臨床情報の蓄積と可視化）

組込みソフトウェアによって、対象サンプルについての属性（血清か血漿か、採取日など）、臨床情報（患者の年齢、性、診断名、検査結果およびその時間的要素など）について一次的情報を蓄積して必要時にいつでも引き出せるようにした。また、サンプルから得られる情報のユーザーの理解を支援するために可視化した。サンプル情報（属性、臨床情報）についてはツリー図による可視化を行った。

3. LC-MS ロボットの情報を利用する機関の情報システムとの連携をサポートする技術の構築

3-1. 医療機関への情報提示方法の構築

医療機関への LC-MS ロボットからの情報の提示は医療機関における診断を支援することを目的とする。各疾患カテゴリーにおいて、可能な診断のリストがあるとして、それぞれの診断の可能性（推定される診断確率）を与えることができるような情報提示の方法を構築することが理想である。そこで、認知症および肝疾患に関してこれまでに得たデータおよび情報の解析を行った。本研究開発では、特定の疾患をモデルとして、そのアルゴリズムを構築することを目標としたが、認知症および肝疾患に関して「推定される診断確率」を与えるアルゴリズムを構築することができた。

3-2. 患者の臨床情報を記録した電子カルテとの連携のための技術開発

LC-MS ロボットの構築は医療機関における診断を支援することを目的とする。患者の臨床情報を記録した電子カルテからの情報と LC-MS ロボットからの情報を組み合わせることによって、医師に対して自動的に診断の可能性（推定される確率）の情報を出力できるような組込みソフトウェアを構築することが理想である。そこで、認知症および肝疾患に関してこれまでに得た電子カルテに記載されている臨床的データおよび情報を LC-MS ロボットからの情報と組み合わせ解析を行った。本研究開発では、認知症および肝疾患をモデルとして、「推定される診断確率」を与えるアルゴリズムの構築ができた。

第2章 本論

「平成22年度戦略的基盤技術高度化支援事業「血液診断バイオマーカーのための定量比較 LC-MS ロボットにおける組込みソフトウェアの開発」に係る委託業務実施計画書」(22fy 予備費)の成果を以下に記載する。項目番号は本実施計画書にしたがい、大項目を①、②のような丸数字で表す。

① 組込みソフトウェアによるデータ処理時間の短縮

①-1. パラメタの繁用メニューの設定

現在のデータ処理は、毎回分析条件にしたがってマニュアルでパラメタを設定している。そこで、以下のパラメタのそれぞれについて2つのオプションを設定することでパラメタ設定に掛かる時間を短縮した。

1) LC へのサンプル量：25 μ L、50 μ L (通常は 50 μ L)

2) solvent A および solvent B のうち、solvent B が 5%から 25%に達するまでの時間：30 分、40 分、任意 (X 分とする) (通常は 30 分)

①-2. 質量分析に調和した LC 分画溶出速度の制御

LC の分画が速く出すぎて、質量分析が間に合わない場合に備えて、上述②-1 の 2)における「solvent B が 5%から 25%に達するまでの時間」を任意 (X 分) に変えることができるような組込みソフトウェアとした。

①-3. すでに構築した信頼性向上策の活用

1-1～1-3 で構築した信頼性向上策については **1-3 成果概要**に記載した。

以上の手段を用いて、目標の一日あたり 500-1,000 サンプルの処理が可能となる見通しを得た。

② サンプルから得られる情報の可視化とこれによる比較解析機能の構築 (とくにサンプルの属性および臨床情報の蓄積と可視化)

データ入力と比較解析のための組み込みソフトウェアをプログラム作成した。組込みソフトウェアによって、対象サンプルについての属性 (血清か血漿か、採取日など)、臨床情報 (患者の年齢、性、診断名、検査結果およびその時間的要素など) のような一次的情報を csv 形式で蓄積して必要時にはいつでも引き出せるようにした。また、理解に便利のように可視化した。サンプル情報 (属性、臨床情報) についてはツリー図による可視化を行った。

比較解析としては本研究開発の目的から ROC 曲線の解析と散布図の描出を行うようプログラム作成を行った。

③ 医療機関への情報提示方法の構築

医療機関への LC-MS ロボットからの情報の提示は医療機関における診断を支援することを目的とする。各疾患カテゴリーにおいて、可能な診断のリストがあるとして、それぞれの診断の可能性 (推定される診断確率) を与えることができるような情報提示の方法を構築することが理想である。そこで、認知症および肝疾患に関してこれまでに得たデータおよび情報の解析を行った。本研究開発では、特定の疾患をモデルとして、そのアルゴリズムを構築することを目標としたが、認知症および肝疾患に関して

「推定される診断確率」を与えるアルゴリズムを構築することができた。

③-1. 検討した解析手法の種類と原理

測定対象となるパラメタがどのように診断確率に寄与するかを調べる手法には以下の三つの方法が考えられる。ロジスティック回帰は、次の項目「③-3. 各解析手法の適用結果」から判断して、今回のプロジェクトの目的に最適な方法である。

1) ロジスティック回帰

この方法は各パラメタの係数をデータセットから得て、二つの疾患カテゴリーにおける患者ごとの診断確率を与えることができる。ロジスティック回帰の正答率は以下の式(1)によって定義する。判別は被験者の二つのカテゴリー（例えば MCI と Normal）のうちのどちらに属するかをロジスティック回帰式から推定することによって行われる。被験者のカテゴリーを i （例えば MCI）とし、ロジスティック回帰式から得られた診断確率が 0.5 以上のとき i と正しく診断されたとみなす。カテゴリー i の総被験者数を N_i 、正しく i と診断された被験者数を C_i とする。

$$\text{正答率} = C_i / N_i \quad (1)$$

2) 判別分析

この方法は、データ点が二つのカテゴリーのどちらに属するかを「マハラノビスの汎距離」の概念を用いて判別する方法である。本研究開発では線形判別分析を用いた。

正答率はロジスティック回帰の項で記載した式(1)で定義される。

誤答率 (P_{1e} 、 P_{2e}) は、カテゴリー1 の中で 2 と判定された個数を E_1 、2 の中で 1 と判定された個数を E_2 とするとき、

$$P_{1e} = E_1 / N_1 ; P_{2e} = E_2 / N_2 \quad (2)$$

判別分析ではデータ全体の誤判別の確率は得られるが、新たなサンプルがどちらにどの大きさの確率で属するかは求められない。その点でロジスティック回帰に劣る。

3) 主成分分析

この方法は相関するパラメタをまとめた互いに直交する多次元の合成パラメタ（主成分）を作り、各主成分への元のパラメタの影響の大きさを「因子負荷量」から検討するのに便利である。因子負荷量と「寄与率」は臨床検査値や本研究開発で対象にしている診断マーカーの意義を比較評価するのに適している。さらに、「主成分得点」を用いることによって、そのような多次元の主成分の数値から異なる疾患群や時点間の分別も可能である。

③-2. 解析対象とした LC-MS 分析によって得られた診断マーカーのデータ

③-2-1. 診断マーカー測定の対象とした試料

以下の被験者から得た血清を用いた。すなわち、認知症（軽度認知障害－MCI, mild cognitive impairment、アルツハイマー病－AD, Alzheimer disease）および認知症患者と性、年齢においてマッチさせた健常人のセット（NDC, non-demented control）と、肝疾患（慢性肝炎－CH, chronic hepatitis、肝硬変－LC, liver cirrhosis、肝がん－HCC, hepatocellular carcinoma）および対応する

健常人のセット (Normal) である。

③-2-2. 診断マーカーの測定方法

血清 25 μ l からタンパク質の分解を防ぐとともにペプチドを抽出した。ついで、2次元 LC による分画を行い、MALDI-TOF MS で分画の測定を行った。被験者カテゴリー間 (例えば疾患-健常人間) で差の認められた MS ピークについては MS/MS によるペプチドの同定を行った。さらに、MS ピーク面積は標準ペプチドの MS ピーク面積との比算出し、これを診断マーカーの血清中濃度とした。

③-3. 各解析手法の適用結果

③-3-1. ロジスティック回帰

ロジスティック回帰分析は MedCalc for Windows, version 9, 2007, MedCalc Software を用いて実行した。このプログラムは Newton-Raphson 法によっている。

③-3-1-1. 認知症

③-3-1-1-1. MCI 35 例と NDC 53 例の比較

AUC 値 0.6 以上の診断マーカーの組み合わせを trial and error によって求めたところ、以下の診断マーカーがいずれも p 値<0.05 であり、その組み合わせに基づく回帰式による正答率が MCI 群 68.6%、NDC 群 96.2%に上昇した。

ついでながら、NDC (性・年齢においてマッチした非認知症被験者) をほとんど 100%といえる高率で正しく NDC と判定したことは、これらの診断マーカーの意義とロジスティック回帰法の有用性を示すものである。

③-3-1-1-2. AD 68 例と NDC 53 例の比較

測定対象の診断マーカー18種のうち AUC 値が 0.6 以上の診断マーカーで正答率が両群について 70%以上の回帰式を与えたものは、AD1015 (AUC=0.751)と AD1046 (AUC=0.675)であった。

これらの診断マーカーによる回帰式を求めたところ、係数と標準誤差は以下のとおりであり、正答率は AD 群 73.5%、NDC 群 71.7%であった。

③-3-1-1-3. AD 68 例と MCI 35 例の比較

測定対象の診断マーカー18種のうち AUC 値が 0.6 以上の診断マーカーで AD における正答率が 85%以上の回帰式を与えたものは、AD1015 (AUC=0.732)、AD1048 (AUC=0.625)および ADPEP109315 (AUC=0.668)であった。

これらの診断マーカーによる回帰式を求めたところ、係数と標準誤差は以下のとおりであり、正答率は AD 群 85.3%、MCI 群 62.9%であった。

③-3-1-2. 肝疾患

③-3-1-2-1. HCC 43 例と Normal (健常人) 21 例の比較

測定対象の診断マーカー16種のうち AUC 値が 0.7 以上の診断マーカーで両群における正答率が 90%以上の回帰式を与えたものは、marker_31 (AUC=0.901)と marker_56 (AUC=0.826)であった。

これらの診断マーカーによる回帰式を求めたところ、係数と標準誤差は以下のとおりであり、正答率は HCC 群 97.7%、Normal 群 90.5%であった。

③-3-1-2-2. HCC 43 例と CH 50 例の比較

測定対象の診断マーカー11種のうち AUC 値が 0.7 以上の診断マーカーで両群における正答率が 70%以上の回帰式を与えたものは、marker_26 (AUC=0.743)、marker_36 (AUC=0.728)および marker_55 (AUC=0.719)であった。

これらの診断マーカーによる回帰式を求めたところ、係数と標準誤差は以下のとおりであり、正答率は HCC 群 74.4%、CH 群 80.0%であった。

③-3-1-2-3. HCC 43 例と LC 44 例の比較

測定対象の診断マーカー9種のうち AUC 値が 0.7 以上の診断マーカーで両群における正答率が 70%以上の回帰式を与えたものは、marker_60 (AUC=0.738)と marker_26 (AUC=0.68)であった。

これらの診断マーカーによる回帰式を求めたところ、係数と標準誤差は以下のとおりであり、正答率は HCC 群 72.7%、LC 群 74.4%であった。

③-3-2. 判別分析

以下に、ロジスティック回帰の場合と同一の診断マーカーの組み合わせを用いて判別分析を行った結果をロジスティック回帰による結果と比較して記載する。次の 2 点に注目して評価した。すなわち、(1)正答率の比較、(2)各被験者の判別分析による判定がロジスティック回帰からの判定と異なる場合について、a)判別分析のみが誤答を与えた被験者数、b)判別分析のみが正答を与えた被験者数。

③-3-2-1. 認知症

(1)正答率の比較

表 2 は認知症の場合の各比較群対について正答率を示したものである。

表 2 ロジスティック回帰と判別分析の正答率の比較（認知症）

疾患群	合計	ロジスティック回帰分析				判別分析			
		判定→	NDC	MCI	正答率(%)	判定→	NDC	MCI	正答率(%)
NDC群	53		51	2	96.2		46	7	86.8
MCI群	35		11	24	68.6		12	23	65.7
疾患群	合計	判定→	NDC	AD	正答率(%)	判定→	NDC	AD	正答率(%)
NDC群	53		38	15	71.7		40	13	75.5
AD群	68		18	50	73.5		31	37	54.4
疾患群	合計	判定→	MCI	AD	正答率(%)	判定→	MCI	AD	正答率(%)
MCI群	35		22	13	62.9		14	21	40.0
AD群	68		10	58	85.3		10	58	85.3

表 2 から正答率の点では、概ね、ロジスティック回帰のほうが優れていることが分かった。

(2) 両手法による判定が異なる場合の評価

両手法による判定が異なる場合についての結果より、判別分析では誤答率が高いことからロジス

ティック回帰のほうが信頼できると考えられる。

③-3-2-2. 肝疾患

(1) 正答率の比較

表 3 は肝疾患の場合の各比較群対について正答率を示したものである。

表 3 ロジスティック回帰と判別分析の正答率の比較（肝疾患）

		ロジスティック回帰分析				判別分析			
疾患群	合計	判定→	Normal	HCC	正答率 (%)	判定→	Normal	HCC	正答率(%)
Normal群	21		19	2	90.5		20	1	90.5
HCC群	43		1	42	97.7		11	32	97.7
疾患群	合計	判定→	CH	HCC	正答率 (%)	判定→	CH	HCC	正答率(%)
CH群	50		40	10	80.0		40	10	80.0
HCC群	43		11	32	74.4		13	30	69.8
疾患群	合計	判定→	LC	HCC	正答率 (%)	判定→	LC	HCC	正答率(%)
LC群	44		32	12	72.7		32	12	72.7
HCC群	43		11	32	74.4		10	33	74.7

表 3 の結果によれば、両手法による正答率にほとんど差はなかった。

(2) 両手法による判定が異なる場合の評価

両手法による判定が異なる場合についての結果では HCC 群の患者 43 例中 10 例を Normal と判定している点で判別分析は肝疾患においても信頼性の点で劣ると評価せざるを得ない。

③-3-3. 主成分分析

主成分分析の威力を端的に示すのは、この分析によって得られる主成分得点を多次元空間にプロットしたとき、比較すべき二つのカテゴリーが分離して描出されることである。以下に認知症と肝疾患についてその例を述べる。

③-3-3-1. 認知症

AD 68 例と MCI 35 例の診断マーカー AD1015、AD1048、ADPEP109315（これらの診断マーカーに関する情報は③-3-1-1-3 を参照）の血清中濃度について主成分分析を行い、主成分得点を 3 次元プロットした。これらのマーカーについて 3 次元空間における一定の分離が認められた。これらの診断マーカーはロジスティック回帰において、正答率の高い組み合わせとして選択されたものである。

③-3-3-2. 肝疾患

CH 50 例と HCC 43 例の診断マーカー marker_26、marker_36、marker_55（これらの診断マーカーに関する情報は③-3-1-2-2 を参照）の血清中濃度について主成分分析を行い、3 次元プロットした。前項の認知症におけると同様、3 次元空間における一定の分離が認められた。これらの診断マーカーもロジスティック回帰において、正答率の高い組み合わせとして選択されたものである。

③-3-4. 解析手法に関する結論ーロジスティック回帰の選択

ロジスティック回帰と判別分析の基本的相違は、③-1 に述べたように、前者が診断確率を数値として与えることができるのに対し、後者はこれを与えることができない点にある。その上に判別分析では認知症の場合に誤答率が一般に高く、肝疾患においても信頼性の点で劣ると評価せざるを得なかった。

以上から、本項「③医療機関への情報提示方法の構築」においては、ロジスティック回帰を情報提示のための手法として選択することを結論とする。

④ 患者の臨床情報を記録した電子カルテとの連携のための技術開発

LC-MS ロボットの構築は医療機関における診断を支援することを目的とする。患者の臨床情報を記録した電子カルテからの情報と LC-MS 分析の結果からの情報を組み合わせることによって、医師に対して自動的に診断確率に関する情報を出力できるような組込みソフトウェアを構築することが理想である。そこで、認知症および肝疾患に関してこれまでに得た電子カルテに記載されている臨床的情報を LC-MS 分析によって得られた診断マーカーに関する情報と組み合わせ、解析を行った。本研究開発では、特定の疾患をモデルとして、そのアルゴリズムを構築することを目標とした結果、認知症および肝疾患に関して「推定される診断確率」を与えるアルゴリズムを構築することができた。③-3-4. に述べたように、診断マーカーのみでなく、臨床検査値をパラメタとした場合にも、ロジスティック回帰を情報提示のための手法として選択することとして検討を行った。

④-1. ロジスティック回帰プログラムの適用結果

本項では、認知症（③-3-1-1）および肝疾患（③-3-1-2）のロジスティック回帰分析において選択された診断マーカーの組み合わせを固定し、そのデータに臨床検査値のデータを加えて解析した。臨床検査項目が複数の場合は、正答率の改善に寄与する項目を選んだ。

④-1-1. 異常値の取扱い

本研究開発では、①多数の被験者から得られたデータ（これを基礎データと呼ぶ）についてロジスティック回帰分析を行い、②その結果から回帰式を得て、新規被験者の各パラメタの測定値をこの回帰式に代入して診断確率を求める。一方、診断マーカー、臨床検査値のいずれも基礎データ中に異常値が出現することは免れない。

今回用いた認知症の基礎データは認知機能検査である MMSE 値を含んでいた。MMSE 値は 30 を上限とし、正常値が 25-30 の認知機能スコアであるが、NDC 被験者 21 例中 1 例が異常に低い MMSE 値 (= 10) を示していた。この値について Smirnov の棄却検定を行ったところ、棄却すべきもの（危険率 $p < 0.01$ ）と判定された。この結果から、基礎データに組み入れるパラメタの異常値については診断マーカー、臨床検査値のいかなるものを問わず棄却検定を行い、その結果が強度に（例えば $p < 0.01$ で）有意であれば棄却しておくべきものと判断された。

④-1-2. 認知症

認知症被験者において多くの症例に検査値が存在したのは MMSE であり、しかも、MMSE 値の

ある症例はいずれの群も③において診断マーカーのみで解析した場合より少なかった。そこで、解析対象を MMSE 値のある被験者に絞り、MMSE データのない場合とこれを加えた場合を比較し、臨床情報を解析に加えることの意義を評価した。

④-1-2-1. MCI 33 例と NDC 20 例の比較

ROC 曲線の AUC 値は以下のとおりで、MMSE の AUC 値は 0.75 であった。

診断マーカー	AD1031	AD1042	AD1046	AD1049	MMSE
AUC値	0.736	0.792	0.752	0.779	0.754

MMSE をロジスティック回帰式に加える前と加えた後の比較の結果、正答率に MMSE を加えたことによる変化はなかった。これは、MMSE が他のパラメタ（診断マーカー）の値と関連していたためと考えられる。

④-1-2-2. AD 64 例と NDC 20 例の比較

ROC 曲線の AUC 値は以下のとおりで、MMSE の AUC 値は 0.977 と非常に高い値であった。

診断マーカー	AD1015	AD1046	MMSE
AUC値	0.798	0.709	0.977

この場合は MMSE が非常に寄与していて、NDC 群の正答率は 40%から 95%へと飛躍的に上昇した。

④-1-2-3 AD 64 例と MCI 33 例の比較

ROC 曲線の AUC 値は以下のとおりで、MMSE の AUC 値は 0.927 と高い値であった。

診断マーカー	AD1015	AD1048	ADPEP10	MMSE
AUC値	0.719	0.627	0.681	0.927

MMSE を加えたことによって MCI 群の正答率が 52%から 82%へと上昇し、AD 群のそれも 84%から 91%へと上昇した。

④-1-3. 肝疾患

肝疾患において最も多くの被験者で測定された臨床検査値は AST、ALT、albumin、bilirubin、platelet の 5 項目であった。そこで、これらの項目の測定値がすべて揃っている被験者を対象として、診断マーカーのみの場合とこれに臨床検査値のデータを加えた場合を比較した。

④-1-3-1. HCC 42 例と Normal 21 例の比較

ROC 曲線の AUC 値は以下のとおりであった。

診断マーカー	marker_31	marker_56	AST	ALT	Albumin	Bilirubin	Platelet
AUC値	0.899	0.822	0.932	0.883	0.976	0.766	0.957

臨床検査値のうち、正答率の改善に寄与したのは albumin であり、これを加えることによって、いずれの群も正答率が 100%になった。

④-1-3-2. HCC 42 例と CH 48 例の比較

ROC 曲線の AUC 値は以下のとおりであった。

診断マーカー	marker_26	marker_36	marker_55	AST	ALT	Albumin	Bilirubin	Platelet
AUC値	0.755	0.738	0.713	0.545	0.618	0.848	0.746	0.88

臨床検査値のうち、正答率の改善に寄与したのは platelet であり、これを加えることによって、

いずれの群においても正答率の改善が認められた。

④-1-3-3. HCC 42 例と LC 42 例の比較

この場合は、いずれの臨床検査値も AUC 値が 0.5 に近い値に留まり、両群の判別に寄与しなかった。

④-1-4. 臨床検査値を加えることの意義

臨床検査値を診断マーカーの組み合わせに加えると、多くの場合に診断の正答率が改善される結果となった。これは定量比較 LC-MS ロボットの出力にとって重要な知見と考えられる。

④-2. 診断確率の医療機関への提示のためのプログラム

(1) 疑われる疾患対（例えば、AD と MCI）についてのロジスティック回帰の結果から得られた係数の組み合わせに関する情報を用いて新規の被験者の診断確率を算出し、これを医療機関に提示するプログラムによる画面を構築した。

(2) さらに、診断確率を医療機関に提示するためのプログラムを作成した。本プログラムの動作はサンプルデータにより確認した。

第3章 総括

1. 研究開発成果

本研究開発の最終目標は、LC-MS ロボットと医療機関の情報システムとの連携によって医師による客観的で正確な診断を支援することである。これにより、健康・医療の質が向上する（図 4）。波及効果としては、客観的で正確な診断が行われることによって、患者の疾患に必ずしも有効ではない医薬品の投与による副作用などの患者の不利益がなくなり、同時に医療費の無駄もなくなる。製薬産業においては、客観的かつ正しい診断を受けた患者のみを対象にした新薬の臨床試験を実施することができ、ここでも精度の高い試験を行うことができる。その結果、多大の時間と経費の節減が可能となる。



図 4

血液診断バイオマーカーのための定量比較 LC-MS ロボットにおける組込みソフトウェアの開発によって健康・医療の質の向上が期待できる

2. 研究開発後の課題

2-1. 成果全体のまとめと現在のレベル

以上、各項にわたって述べた成果は、本プロジェクトの研究開発対象である「血液診断バイオマーカーのための定量比較 LC-MS ロボットにおける組込みソフトウェア」のいわば、パーツあるいはモジュールである。これらのパーツあるいはモジュールをまとめて LC-MS 装置に組み込んではいじめて目的の定量比較 LC-MS ロボットが完成する。本研究開発によって、これらのパーツあるいはモジュールが現段階ではほぼ完成したことになる。

2-2. 研究開発後の課題

研究開発上の今後の課題は、以下の 2 段階に分けることができる。

第 1 段階は、図最終章-2 の左側、LC-MS 本体+システム PC と、右側、外部 PC による制御の「両者を接続する機構」の開発である。この機構は PC ソフトウェアによって構築することが可能であるが、動作が本体部分のシステム PC による制御の仕方により異なるので、メーカーおよび製品ごとに異なる仕様の機構としなければならない。

第 2 段階は、左側のシステム PC と右側の外部 PC を一つに合体させ、対象の LC-MS 装置に直接組み込めるソフトウェアを開発することである。これは第 1 段階の成果を活用して開発することが可能と考えられる。第 2 段階も、対象の LC-MS 装置とシステム PC 部がメーカーや製品により仕様が異なることが想定される。複数の代表的装置システムを対象とすることになる。

次の課題は産業財産権にかかわる。本研究開発の成果からの特許出願の対象としては、大まかに以下がある。4 は今後の研究開発が必要であるが、1~3 はすでに開発済みである。

1. 入力と比較解析のためのプログラムー画面構成を含む
2. ロジスティック回帰を求めるためのプログラム
3. 診断確率を医療機関に提示するためのプログラムー画面構成を含む
4. LC-MS 本体+システム PC と外部 PC による制御の両者を接続する機構

以上。