

東京女子医大における患者情報収集による証の確立について

研究分担者 木村容子 東京女子医科大学東洋医学研究所准教授

研究要旨

患者情報問診データを用いて漢方治療の科学的エビデンスを創出し、そのエビデンスを基に、漢方の特性を活かした診療支援システムを構築することを目的として、平成 22 年度から 24 年度にまで本研究を行った。

平成 22 年度には、患者の自覚症状を的確かつ効率的に問診することを目的として、頭痛を主訴に来院した患者について、問診項目間の相関ルールについて解析を行った。その結果、頭痛を主訴に来院した患者については、「こり」に関する問診項目を簡素化することができると考えられた。

平成 23 年度には、冷えを主訴に来院し、加味逍遙散、当帰芍薬散、温経湯のいずれかを投与された患者を対象として、初診時の問診項目に基づく患者タイプを分析した。

平成 24 年度には、前年度の研究を基に、当帰芍薬散および加味逍遙散の有効例と無効例の問診データを比較することで、各処方「証」を検討した。

A. 研究目的

漢方治療の問診票には数百にも及ぶ多数の質問項目があるが、どのような自覚症状の違いによって、処方の違いが生じているのかなど、問診項目の整合性や関連性などを研究した報告は調べた範囲ではみられない。

このため、まず、初診時の問診項目の検討から問診項目の簡素化を図った。そして、冷えの患者を対象として、治療前の自覚症状の特徴を分析した。さらに治療効果との関連を解析することにより、冷えの治療における各処方の鑑別、すなわち「証」の確立を目的とした。

B. 研究方法

初診時の問診票では、主症状のほか、随伴症状、食欲、便通、排尿、月経、体質など 370 項目にわたる質問に対して記入してもらった。

問診および治療前後の冷えを、「頻度（なし、まれに、ときどき、ほぼいつも、いつも）」および「程度（なし、わずかに、少し、かなり、非常に）」を各 5 段階に分類し、患者自身により治療前と治療 1 ヶ月時点の 2 回行い有効性を判定した。

（倫理面への配慮）

本研究は「ヘルシンキ宣言」ならびに「疫学研究に関する倫理指針」を遵守し行った。

C. 研究結果

頭痛を主訴とした患者の場合、「こり」については、「こり（首筋）」、「目の疲れ」や「冷え（足）」の問診項目があれば、「こり（肩）」という一項目に減らして、問診の簡素化を行うことも可能と考えられた。

2008年10月から2011年12月までに冷えを主訴として当院を受診した患者のうち、加味逍遙散が41名、当帰芍薬散が71名、温経湯が21名に処方された。冷えの部位については、加味逍遙散、当帰芍薬散および温経湯のすべてにおいて、「足」が約80%、「手」が48-64%と頻度が高かった。逆に、「全身の冷え」を訴えている割合は少なかった(18-32%)。また、その他の寒熱の症状については、加味逍遙散、当帰芍薬散および温経湯のすべてで「寒がり」(48-64%)の症状の訴えが多かった。

さらに、2008年10月から2012年9月までに来院した冷えを訴える患者のうち、当帰芍薬散または加味逍遙散を投与した188名を対象とした。当帰芍薬散を投与した92名(男性0人、女性92人、中央値37歳、範囲17-63歳)のうち、冷えの改善群が66名(72%)、非改善群が26名(28%)であった。また、加味逍遙散を投与した96名(男性3人、女性93人、中央値40歳、範囲22-72歳)のうち、冷えの改善群が67名(70%)、非改善群が29名(30%)であった。

当帰芍薬散が有効な冷えの最適モデルとして、腹部の冷え(オッズ比5.0, 95%信頼区間1.3-19)、めまい(7.7, 1.4-42)、目のかすみ

(16, 1.2-200)、のぼせ(5.6, 1.2-26)を訴えるが、怒りっぽさ(0.11, 0.019-0.66)や耳鳴(0.025, 0.001-0.68)がみられないことが挙げられた($p < 0.001$)。一方、加味逍遙散が有効な冷えを予測する因子として、冷えは全身(0.099, 0.032-0.31)ではなく、発作性の発汗(14, 1.2-150)がみられるが、立ちくらみ(0.21, 0.067-0.65)を訴えないことが選択された($p < 0.001$)。モデル適合度検定は、当帰芍薬散に関して $p = 0.87$ 、加味逍遙散に関して $p = 0.83$ であり、いずれも適合しないとは言えない(図1参照)。

D. 考察

漢方治療では、西洋医学的な検査では明らかな異常がみられない場合でも、患者の自覚症状があるときは治療対象となる。また、同じ主訴で来院しても、陰陽虚実、気血水、五臓などの漢方医学における病態概念を総括した「証」が異なれば、漢方処方も自ずと違うものになる。この意味で、患者の自覚症状をいかに正確に把握するかは、「証」を診断する上で大切になる。

今回、多数例の問診データの解析により、冷え症の患者に対し、当帰芍薬散証か加味逍遙散証を鑑別する上で役立つ情報を検討できたと考える。

E. 結論

問診データを解析することは、患者が治療前にどのような自覚症状の特徴を有しているかが分かり、漢方治療における処方鑑別に役立つものと考えられた。

今後、症例を重ね、さらに治療効果との関連を解析することにより、エビデンスレベルの高い知見、頻用処方「証」の確立をめざす。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

木村容子，田中彰，佐藤弘：当帰芍薬散および加味逍遙散が有効な冷えについての検討，日本東洋医学雑誌（掲載予定）。

2. 学会発表

なし

H. 知的財産権の出願・登録状況

1. 特許取得

該当なし

2. 実用新案登録

該当なし

3. その他

なし

図1 冷え改善の予測因子の最適モデル

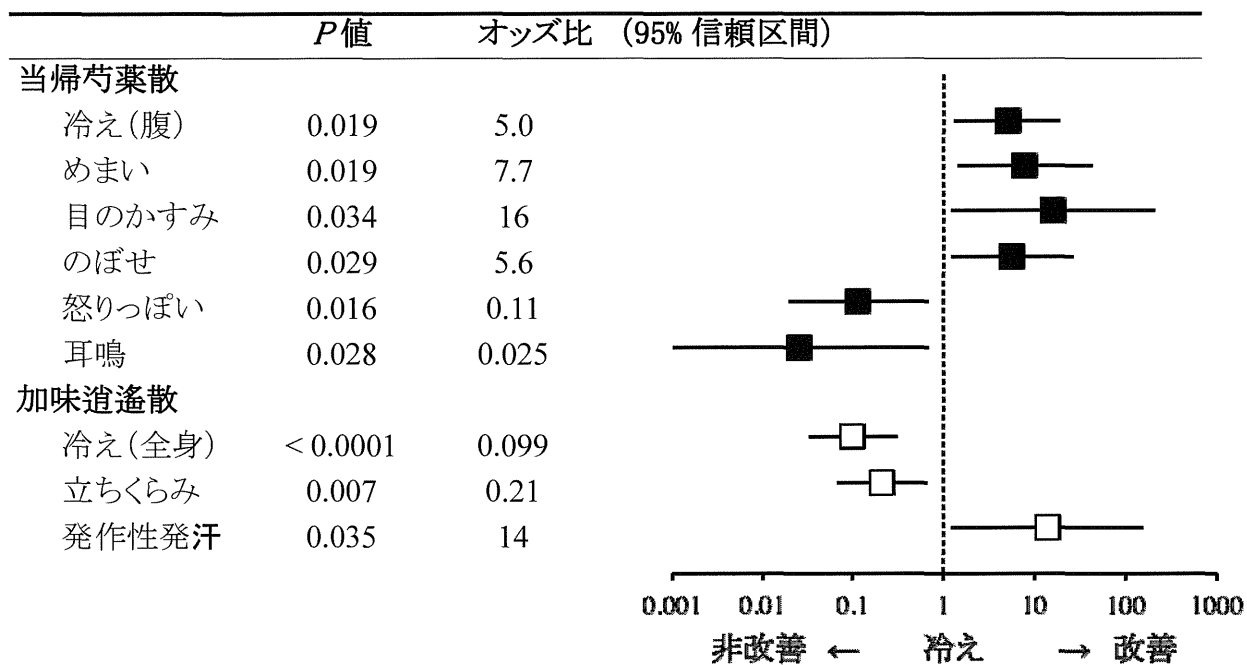


図1 冷え改善の予測因子の最適モデル

当帰芍薬散に関する冷え改善の予測因子として、腹部の冷え、めまい、目のかすみ、のぼせは有意な陽性所見であり、怒りっぽい、耳鳴りは陰性所見であった。一方、加味逍遙散に関する陽性所見は発作性発汗のみであり、全身の冷え、立ちくらみは陰性所見として選択された。

Ⅲ. 会 議 議 事 録

平成 22～24 年度厚生労働科学研究費補助金

『漢方の特性を利用したエビデンス創出と適正使用支援システムの構築』

報告会 概要

日時：2012 年 11 月 29 日（木） 17 時～19 時

場所：慶應義塾大学 医学部 新教育研究棟 2 階講堂

参加者：

会議室参加（順不同）：渡辺賢治・井元清哉・山口類・片山琴絵・美馬秀樹・植松すみれ・並木隆雄・多田浩貴・吉田一郎・吉野鉄大・宗形佳織

1. 基調講演「医療情報科学の未来」：井元清哉

臨床シーケンスの時代に入り、膨大なデータ解析が必要とされる時代が来ている。ゲノム解析においては種々の会社からシーケンサーが続々と発売され、価格競争が起こり、比較的安価になりつつある。近年ではこれらの解析データを用いて治療に繋げる試みが各国でされており、東京大学医科学附属病院でも推進事業として進められている。今後はパーソナルゲノム情報を如何に医療に結びつけるか、その有効性をいかに示すかという事が課題となっていくであろう。

2. 厚生労働省科学研究費

1) 研究の経緯：渡辺賢治

医師の 90%以上が漢方薬を処方しているという報告の一方で、実際に漢方理論に基づいて処方されているとは言い難い現状がある。漢方医学的診断「証」の科学的・統計学的解明および RTC に不向きな漢方治療の臨床試験を新しい手法によりエビデンス構築を行い、漢方診療支援システムの構築および漢方治療の標準化を目指す。

2) 問診から見る証の予測：片山琴絵

問診システムから得られたデータに基づき、診療支援ツールとして「証」をリコメンドする事を目的としている。虚実について 6 施設の比較を行った結果、施設間で虚実診断に相違が認められた。またランダムフォレストを用いた虚実予測において、慶應のモデルで慶應を判別した際には高い中率であったものの、他施設を判別した際には精度が落ちていく事が明らかとなった。その要因として、重要視される問診項目が施設間で差がある事お

よび重要視の程度が異なる事が示唆された。

3) 診療支援ツールの開発：美馬秀樹

診療中に医師が画面を見る事ができる時間は限られており、データ解析結果の見える化は必須である。その為のツールとしてこれまでもシラバスの見える化で実績のある MIMA サーチを使用し、漢方診断支援を行う事を目的としている。過去の臨床エビデンスの効率的な検索、類似性・関連性の効率的な分析、可視化による経験値の「見える化」による暗黙知の開示、これらにより迅速な意志決定を促し漢方診断支援となる。

4) 漢方診断改良の科学的な取り組み：並木隆雄

問診システムから得られたデータから約 30 年前に寺澤捷年らにより研究・提唱された気血水スコアの再検討を行う事を計画している。また、舌診はこれまで医師の主観や経験に大きく依存していたが、千葉大学で開発した舌撮影解析システム (TIAS) が舌診の科学化、診療、教育に寄与する事が期待されている。この TIAS を問診システムと併用する事で客観データの追加による診断支援のさらなる強化に貢献するものと考えられる。

3. 総合討論

Q1) MIMA サーチは漢方のみならず西洋医学でも使えるシステムか。

A1) 処理の大変さは多少異なるが、大きく変わらずに使えると考えられる。

Q2) 多剤併用には対応しているか。

A2) 現行では対応していないが、将来的には必要と考えられる。

Q3) 西洋的な所見と漢方薬の効果との相関は見る事ができるか。

A3) どのデータを取らないかも重要。シンプルな系で進める。

Q4) 脈診の科学化は困難か。

A4) データは取れてもその病理的解釈が非常に困難。

Q5) 将来的には自由記載カルテや患者の音声のような情報源から MIMA サーチへ繋がっていく事ができるか。

A5) 言葉には多数の表現があるものがあり、データの取り逃しに繋がる可能性がある。また、言葉には多数の意味をもつ物もあり、その場に適合した意味を取り出す事が難しい。その為、あらかじめ設定された問診項目に対する○×形式でデータを取得する事が最も効率よく解析できると考えられる。

Q6) 舌撮影解析システムをスマートフォンや iPad に応用していく事ができるか。

A6) 画像解析には光源の問題が大きく関与しており、さまざまな試行錯誤の結果、球形にする事で安定した一定の条件を保って画像解析ができるような仕組みになっている。そのため、条件を均一にできないスマートフォンやiPadでは医療情報となる精度の画像解析は望めない。また形状に関しても胖大、苔の範囲や厚さの判断は非常に困難であり、ハードルは高い。

平成 22～24 年度厚労科研費

「漢方の特性を利用したエビデンス創出と適正使用支援システムの構築」 報告会（2012. 11. 29. 17:00～19:00）議事録

会議室参加（順不同）：渡辺賢治・井元清哉・山口類・片山琴絵・美馬秀樹・植松すみれ・並木隆雄・多田浩貴・吉田一郎・吉野鉄大・宗形佳織

渡辺：皆様お忙しい中お集まり頂きましてありがとうございます。慶應大学で漢方をやっております、渡辺と申します。今日の会は厚労省の科研費の医療技術実用化総合研究事業というもので平成 22 年から 24 年度、今年で最後という事でやって参りました「漢方の特性を利用したエビデンス創出と適正使用支援システムの構築」ということでございます。漢方という物の特性とかという話は後ほど私の方から話しをさせていただきますけれども、この研究の組織自体が、漢方というのは一口でいうと非常に個別化医療という事にして、これから個別化医療に向かっていくであろうという事から共同研究をお願いしたのが、これからお話いただく井元先生の教授の宮野悟先生です。宮野先生と井元先生は九州大学の数学科の先生でして、今医科研のヒトゲノム解析センターという所でゲノムの解析をしているという先生でして、まさに数学者と漢方の医者と、後ほどお話を頂く美馬先生というスーパーコンピューターサイエンティストというコラボでできたというものでございます。どうしてそのようなものが繋がるのかという事はだんだん話を聞くうちに分かる、紐解けるのではないかと思いますけれども、今日は初めに基調講演として「医療情報科学の未来」という事で、まさにゲノム時代になると個別化医療というのが避けては通れないものという事になりますので、ゲノムを中心として医療情報科学の未来という事でお話を頂きます。それでは井元先生お願いいたします。

井元：ご紹介ありがとうございました、井元でございます。今ご紹介いただきましたように私は九州大学の数学科の出身でしてずっと学部修士博士と数学科におりまして空想の世界で幸せに生きていたのですが、2001 年に九州大学の方から今所属している東京大学の医科学研究所ヒトゲノム解析センターの宮野先生の所に参りまして、一転空想の世界から現実の世界に“突き落とされた”という表現が正しいかと思います。私は統計学を専門にしているのですが、リアルなデータがこんなにも複雑なのかという事を思い知らされる、今でも思い知らされる毎日です。（スライド 1）そんなこんなで渡辺先生と縁ができて漢方の事をやってまいりましたけれど、今日は「医療情報科学の未来」という事で私が本当に専門としているゲノムサイエンス、ゲノムデータサイエンスという観点からこの医療情報科学において今どういう状況であ

って、どういう問題がこれから生じてくるのか、我々はどういうふうに対応していかなければいけないのかという事について少しお話できればというように思います。どうぞよろしく願いいたします。(スライド 2) 最初にちょっと「退屈な話」という事を書いたのですが、少しだけ生物学の事を、私は数学科なので本当はこういう事が専門ではないのですが少しバックグラウンドをお話したいと思います。私たちの身体を作っているものはもちろん細胞なので、細胞の中には核という物がございます。核の中には DNA と呼ばれるものがあって、これが遺伝情報を持っているという事でございます。この上には遺伝子と呼ばれる領域がヒトだとだいたい 2 万カ所程度あって、そこからたんぱく質が作られます。たんぱく質は我々の身体を作っているという事になります。そういう情報を総じてゲノムとっております。次に書いてありますエピゲノムというのは少し耳慣れない言葉かと思うのですが、例えていうならば、私たちの皮膚は皮膚であるのですね、肝臓は肝臓であるのです。心臓は心臓である。でも基となっているゲノム、DNA 情報というのは同じ。何か違うのでしょうか。ひとつはエピゲノムというものが違います。どういうことかという例えて言うならば DNA の情報が書かれている本があったとする。エピゲノムというのはホッチキスです。本をバシヤ、バシヤと閉じている部分があります。心臓の DNA 情報はここの部分が開きません、でも肝臓では開きます。皮膚では、また別の部分、ここが開きません。でも他の所だと開きます。こうやって使い分けているのですね。そういったものがエピゲノムの持っている情報だと思っていただければ結構です。どこの部分が使えて、どこの部分が使えないのか、そういう物です。たんぱく質を作るというプロセスにおいてはまずトランスクリプトームというのが大切な情報を持っています。どういう事かというと、DNA の上に遺伝子が載っています。遺伝子からたんぱく質を作る、作る時に一旦メッセンジャーRNA という物質に変換されてそのメッセンジャーRNA がたんぱく質に翻訳される。そのメッセンジャーRNA を含む一般に RNA をトランスクリプトームといいます。そういう訳でたんぱく質の前の段階なのですね。そこからたんぱく質が作られて私たちの身体を作る、もしくはいろいろな反応を起こすという事になります。いろいろ DNA であったりゲノムであったりエピゲノムであったりトランスクリプトームであったりは、今まで我々は別々の装置を使ってそれぞれ計測していました。ちまちまという表現はちょっと悪いかもかもしれませんが、それぞれ計測していました。(スライド 3) 今日お話するシークエンサーという物を使うとこのようないろいろな情報が何でもとれます。もちろん実験の手間でもあるのですけれども、シークエンサーという機械を使えば DNA の、ゲノムの情報もエピゲノムの情報もトランスクリプトーム、どういった遺伝子がどれくらいのメッセンジャーRNA を作っているかという事も計測できます。DNA の情報ですと、どこに他の人との変異が、違いがあるのかという配列の情報が分かります。エピゲノムで言うと僕があなたと違ってここの部分が使えないという事が分かります。全部シークエンサーで分かります。ではシークエンサー、シークエンスと簡単に言いましたけれども、

ちょっとここで整理をいたしましょう。(スライド 4) 生物の DNA 情報を読み取る装置の事を「シーケンサー」と言います。DNA 情報というのは ACGT の 4 つの塩基で作られております。その 4 つの塩基をコンピューターが読めるようにする事を「シーケンスする」といいます。コンピューターで読めるという事が大切な事です。(スライド 5) これが一番最先端のシーケンサーのマシンです。illumina という会社が作っている Hi-Seq 2500 という物です。非常に高価なマシンで一台当たり 1 億円します。高いですけども日本で 2500 はあまり無いと聞いています。このひとつ前の 2000 というものだと 100 台以上日本にございます。稼働しているのは数台。なぜかと言うのは後で言います。これは標準モードと Fast ランモードという 2 つのモードに切り替えができて、コストとかそういうのが違う訳なのでですけども、600 G base、600 G 個の塩基を 11 日間で読み取る事ができます。ヒトの DNA というのは 3 G base です、使えるのは、11 日間で 200 倍できます。ヒト 1 人の DNA 情報をこの機械で完全に読み取る事ができます。完全にというとちょっと言葉が悪いですが、読み取る事ができます。Fast ランモードだと 1 日で 180 G base 読めます。Base は塩基です。180 G base 読めます。ヒトは 3Gb なのでヒト 1 人十分な量が得られます。コストが問題です。機械が高かったです、1 億円。ランニングコストはいかがでしょうか。1 回、11 日間この機械を回すのに 360 万円の試薬代とかが必要になります。Fast ランモードだと 1 日で 110 万円消費します。10 日間回すと 1,100 万円です。1 年間フル稼働しますと、1 年間、2 年間、3 年間くらいで私の生涯年収を超えます。それぐらいの高価な物ですけども、こういう機械を使うと私のゲノムが分かります。ただ、高価です。まだお金持ちの研究室にあって、お金持ちの先生方がいっぱい部下を引き連れて使っている。そういうような現状です。(スライド 6) シーケンサーの性能というのは日進月歩という以上に発展しています。これは縦軸が単位時間あたりに読み取る事ができる塩基の数の Log スケール、横軸は year です。Log スケールで線形に増えていく、これは実際は指数的に増えている。このように感じてシーケンサーの性能というのは高まっています。これはデータの的にはちょっと古い、2010 年の物ですので今はもっと増えています。(スライド 7) もっと安くなる。今 illumina の機械高かったです。機械が高い、ランニングコストが高い。まだまだ普通には使えません。安くなります。半導体シーケンサー、これは Ion Proton、発売されました。先程の illumina の機械は 1 億円でしたが、全部で 3,000 万円になります。なんだかセールストークみたいですね。残念ながら Ion Proton からは何ももらっていないです、ちょっともらってもいいかなと思うくらいですが、残念ですがもらっていません。Ion Proton のランニングコストは今の所ホールゲノム、ヒト 1 人の情報を全部読むには足りないのですけれども約 10 万円、たんぱく質をコードしている遺伝子と呼ばれる領域だけ読むのだったら十分です。来年発売されるとされている P II チップという物だったらホールゲノム、ゲノム全部、3Gb 全部読む事ができます。ただ、読む時にももちろんエラーが入ります。だいたい illumina のシーケンサーで 1%程度、こっ

の Ion Proton の方はよく分からないのですけれども、もうちょっと高いのではないかと聞いています。エラーが入るので 1 回だけサーッと読んでお終いという訳にはまいりません。それからひとつずつが長く読める訳でもありません。どうやって読むかというと、ブツブツに短い断片で読んでいくのです。これは後でお話しますが、そうすると何をしなければならないかということ、エラーが入ってくるし、短くしか読めない、だから何回も何回も何回も読んで積み上げて繋げるという作業をしないといけないのです。そういう訳で先程の Hi-Seq でも本当の 3Gb よりももっとたくさん読まなければならない。600Gb でも良いし、Fast ランの 180 G base 読んでやってやっとな十分な量という事になります。他に来年使えるのではないかとされている Oxford Nanopore テクノロジーという所が発売するシーケンサー、あまり詳しいことは言えませんが、やっぱり 1,000 \$ 以下で発売して、これがシーケンサーで USB のようになっていてパソコンに刺して使うシーケンサーで、15 分で読めるそうです。早い、コストも安い。こうなってくるとパーソナルユースというのはちょっと言い過ぎかもしれないですけど、私のゲノムが読める可能性が出てきます。まだまだ私のゲノムを読むのに 100 万、200 万、300 万というのはちょっと高いかなとは思いますが、こういうコストで読めるようになっていきます。「とっていた」ということで、これは全部聞いた話なのですけれども、もっと安くなる。(スライド 8) 来年 Genia テクノロジー、これはベンチャーの会社なのですけれども注釈付きで、この会社の CEO が言っていたのですが、100 \$ 以下で 1 時間以内に 1 人のゲノムを決定する為に必要なデータが出ます。今、大切な事を言ったのですけれども、1 人のゲノムを決定するのに十分なデータが出てくる。それは解析しないとももちろん解析しない個人のゲノムは分からないのですけれども、それに十分なデータが出てくる訳です。(スライド 9) そういう訳でまたデータ処理が必要になってくるのですけれども、2つの言葉だけ覚えていただければ良いかと思えます。この読み方は「40 カケ」と読みます。あとはショートリードです。40 カケというのは何かというと先程申し上げたように短い断片でしか読み込みません。それで積み上げます、何回も何回も読みます、積み上げの高さです。全てのゲノム領域を見た時に平均で 40 カケくらい、40 個くらい積み上がっている、平均値です。それくらいを目標に読みます。ショートリードというのは何かと申しますと先程短い断片で読みますと申し上げました。短い断片で読むというのはどういう事かと申しますと、DNA を断片化します。どのくらいの長さかということ 200~250 くらい。その両端、150 を読みます。という訳でデータとしてあるのはこの部分だけなのです。(スライド 10) どういう事をするかということこの短い断片をまず最初に繋げないといけない。何も見ずに繋げるのは非常にハードワークなのでリファレンスゲノムを使います。国際ヒトゲノム計画で作ったリファレンス配列というものがあります。ただしそれは 1 人の代表選手なので、自分のゲノムとは実は厳密には違います。という訳であてはまらない所はもちろんリードの中にエラーが入ってくるという状況で、リファレンス配列にアライメ

ントしていきます。この作業というのは言い換えるならば 10 億個のピースのあるジグソーパズルを、はまらないピースが入っている、もしくはピースに傷が入っている物もありながらジグソーパズルを作るような作業。ではどうしてもこのジグソーパズルにアライメントできないパーツはどうするのか、というとそういうものはかき集めて繋ぎ合わせる、要するにのりしろみたいな感じですね。ここここは同じだからくつつくだらうね、ここここは・・・とどんどんどんどん繋ぎ合わせていきます。すると新しい配列ができる。これはもしかするとその人にしかない配列かもしれない。もしくは他にもあるかもしれない。分からないですけれども、こういう事をやります。これはジグソーパズルに例えると完成の図が分からない。そして、繋がるかどうか分からない、組み合わせるかどうか分からない、何の絵ができるか分からないという数万ピースのジグソーパズルを作る訳です。これは人の手ではできない。だけど、こういうことがもしできるとリファレンス配列がございします。その上に張り付いた、もちろんこれはエラーが入っているのですよ、変異も入っています。こういうフレーターが得られます。こういう中でリファレンス配列とは違う配列が多くリードで出てくると、ここはこの人独自の変異なのではないの？もしくは多様性と言った方がいいかもしれないですけど、これだけだったら多様性と言えるかもしれないですけど、病気と関連するのだったら病変といえるかもしれない、とこういった所が分かります。こういう所が病気と関連しているのではないかという疑問がでてきます。(スライド 11) もちろんこういう短い、一文字の変異もございしますし、もう少し例えばある人はこれがリファレンス配列だとします、ごっそりこの部分がなくなっているような人、ガンとかに多いのですが、同じ所が 2 回繰り返されている、こういうものもあります。もしくは例えば 1 番染色体内に 3 番染色体のある部分が挿入されている、ガンとかに多いのですが、そういう挿入などの変異がこういう物を見る事で分かってきます。では実際どのようなサイズのデータであるかという事を説明いたします。(スライド 12) これはがんゲノムの解析の例ですけれどもがん患者さん 1 人の方のがん細胞からデータを取ります。それだけではがん細胞特異的に起こっている変異を見つけないわけですからそれだけでは分からないので、その人の正常細胞をとってきて比較します。そうしないとリファレンス配列に比べてその人が元々持っている変異がある。その変異に新たに変異が加わってがんになる。このがんになる原因となる変異を見つけないのですね。その比較をする。がんの組織からデータを取る、正常組織からデータを取る。シークエンサーではき出すデータというのはこれくらいのサイズ、600GB、450GB。差があるのはなぜかというところががん細胞では 40 カケで読んでいます。ただ正常細胞の方はがん細胞に比べると変異が少ないだろうという事で 30 カケで読んでいます。30 カケくらいになるように読んでいます。本当に 30 カケになっているかというのはデータを解析して初めて分かります。では、リファレンス配列に配分してみましよう。どれくらい時間がかかるのか。これはもちろんスパコンを使うわけですが、1,000 コア使います。33 時間かかります。本当はも

う少し早くなるかとは思うのですが、並列化する部分が抜かしている部分が若干残っていてその部分の並列化に時間がかかっているのですね。そういう部分がございます。シングルコアだと 600 時間かかります。パソコンでやると 600 時間かかります。最終的に出てくる積み上げた配列はだいたいこのようなサイズになります。という事は 1 人の患者さんトータルで 1.3TB のデータが出てきます。1000 人分では 1.3PB になります。このデータを使ってガン細胞と正常細胞を比較する事によって、全ゲノム領域では 8 万個くらいの違う所が出てきます。タンパク質のコード領域に絞ると 1400 あります。違う所がでてきています。その中で意味があるものを探します。この人が元々持っている違いはどのくらいかという、これもけっこう出てきます。これも解析のパラメーターの設定で非常に強いシグナルの物だけ拾うともっともっと減るので、これは結構広めに拾っています。これから統計解析に持っていく。 (スライド 13) 使っているスパコンというのはこういうもので、ヒトゲノム解析センターではスパコンが設置されていてハードディスクは 3.3 ペタバイトございます。計算能力はだいたい 22,000 コア、225 テラフロップスというあまり耳慣れない単位のデータボリュームです。これだけでは分からないので、みなさん大きく見られているかと思えます。(スライド 14) これぐらいの程度がございます。これは実はライフサイエンスに特化した、ライフサイエンス専用のマシンです。このマシンで例えば、宇宙の誕生のシミュレーションをやりたいという方が申請を出したとしてもそれは許可できません。ライフサイエンス専用のマシンです。ライフサイエンス専用のマシンとしては日本で一番大きいマシンです。2009 年に shirokane1 と shirokane2 という 2 つの大きなシステムになっていまして、shirokane1 が先にスタートしていて shirokane2 が今年の 1 月にスタートしました。shirokane1 の性能は当時は世界 69 位だったのですが、だんだん下がってきて Top500 になるくらい、shirokane2 は今年 1 月からで、単独の性能としては世界 183 位。ご時世ですのでトップとなるとあまり批判される所なので、適度な所かなと我々は思っています。実際はこの 2 つを合算して使っていますので、十分というにはまだまだ全然駄目なのですが、いろいろな計算ができるなという所です。(スライド 15) という訳でこういうシーケンサーの発展、こんな発展をベースにして、臨床シーケンスの時代になりました。臨床シーケンスとは何かというとシーケンシングして、データ解析して、その結果を患者さんに返す。診断、治療の方針をそのデータを基にして決める、そういうものです。世界ではもう始まっています。これは 2009 年から 2011 年、世界でたぶん初めて臨床シーケンスの結果を本当に患者さんに返している。まあハッピーな結論になったという事だと思うのですが、アメリカのウィスコンシン医科大学でタンパク質のコードネームだけでも、読んで、原因遺伝子を見つけて、その治療に利用する事ができた。非常にいい話です。今年の 2 月ですけれどもカナダ Ontario Institute for Cancer Research という所は国際がんゲノムコンソーシアムという大きなコンソーシアムがあるので、そこの結果を使って臨床シーケンスをやりま

という事をアナウンスしました。トロントの Sick Children というホスピタルは先程の Ion Proton を最初 4 台入れて臨床シーケンスを始めるとアナウンスしました。彼らが言うにはここ数年のうちに 1 年間で 1 万人の全ゲノムシーケンスをやると言っています。本当はこれは可能です。患者さんさえアグリーメントをとれば可能です。もちろんインフラが許すかどうかは別問題ですが、可能な数です。もちろんこういうちょっと「えっ！」というようなプロジェクトもあります。5 万人住んでいる島を全体シーケンスする。家系図から病歴から全部入ります。企業が買って売るといいかもしれないですね。日本はどうか、日本はまだ臨床シーケンス始まっていない。(スライド 16) 医科研は始めます。今月アナウンスしました。医科研は医科研付属病院、うちの研究所には病院がございませぬ。東大には 2 つ病院があります。ひとつは医学部付属病院もうひとつは医科研付属病院。我々の所には病院があります。この病院では 2001 年から遺伝子カウンセリングをやっているわけです。我々ヒトゲノム解析センター、既にゲノム解析、国際がん研のコンソーシアムで既に作っています。経験があります。先端医療研究センター、ここには大腸がん、血液がん、既に遺伝子カウンセリングを行って遺伝子検査の治療の実績がございませぬ。この 3 つの拠点、センターが一緒になってこの医科研の中でゲノム情報を拠点とした個別化医療をやりますという事を今月 12 日に発表しました。という訳で、我々はもう後には引けなくなりました。アナウンスすると僕は聞いていなかったんで、ちょうど僕は韓国に行っている時にアナウンスがあつて、帰ったあとでうちの宮野先生にどういう事ですかと聞いたのですが、もう発表しちゃいましたという事なので、我々はやるしかないです。ただ、今そつと書いてあるのですけれどももお金はありませぬ。そこがちょっと痛いところなのですけれども、やる気はあるのですけれども、いかんせんお金がない。(スライド 17) という訳で、こういうバックグラウンドがあつて、これから我々に何が待っているかという、ゲノムビッグデータとの格闘、何をやらなくてはいけないか、もちろんデータ解析の方法を考える。方法を考えるだけではお話になりませぬ。インフラが必要です。スパコン使う、クラウドコンピューター、ストレージこういうものを併せて考えていかなくてはいけない。ただ解析しても全く意味がないです。結果を臨床のお医者さんと一緒に議論して、もしくはバイオロジストと議論して治療を決められるようなものにしていかなければならない。その為には結果の見える化、可視化がものすごく大切になっていきます。ものすごく大切になっていきます。これがないと何も分からないのと全く同じになってしまいます。で、どういう事が問題になりどういう事をやらなければならないかという、臨床情報、お医者さんの見立てと病歴、患者情報、究極のパーソナル情報なのですがそれをいかに結びつけて医療に結びつけるかという事が大切です。今シーケンサーを使っている人達は生物学者の人達が多いです、やはり研究の為です。われわれは研究の為ではなく医療の為に使います。これはすごいプレッシャーなのですけれども、もうスタートしています。最後に統計学的にこれが非常に興味があるというか、もうマスト

事項だと思うのですが、パーソナルゲノムを使うともちろん治療は個別化されます、個別化された治療の有効性をどう証明するか。これ、先日 NHK スペシャルでがんワクチンの話があったのですが、やっぱり個別化された治療をどうやって有効性を証明するか。漢方もそうです。個別化された治療をどうやって有効性を証明するか。これは今から考えなくてはいけない、命題だと思います。というわけでちょっと専門的になりましたけれども、私の話はこれで終わりになります。どうもありがとうございました。

渡辺：井元先生どうもありがとうございました。専門用語が出たのでなかなか理解しにくい所もあったかと思えますけれども、最後の方のスライドは宮野先生がしゃべり始めると興奮し始めて怒り始めるところなのですね。要するに日本の問題は情報はとるのだけれども、それが解析ができない。医療情報全てそうなのですね。例えば慶應病院であればひとりの患者さんの採血の結果なんかはものすごい何年間かの蓄積があるのだけれども、有効に活用されないという中で、今度ゲノムになるとデータがものすごい量になってですね、データをとる事は簡単なのですが、MRI とか CT とか全てそうなのですから、とるのは簡単な時代になったのですが、それをフィードバックする為にはマイニングしなければいけないと。そのマイニングできる情報のプロがないと。その辺りが宮野先生の顔が真っ赤になるところなのですから、せっかくなので質問があれば受け付けたいと思います。いかがでしょうか。

質問者：PFI の長谷川と申します。計算処理っていろいろとやられていると思うのですが、計算処理の一番のボトルネックはどのへんにあるのですか。対応性なのか、コンピュータの CPU パワーなのか、どの辺にボトルネックがあつて。例えばアルゴリズム的に分散させて解析するような事というのはできないのでしょうか。

井元：基本的にデータパラレルなので、その部分のパフォーマンスはできます。その部分はそのように苦労はしていません。ただ結果の妥当性を見るのに我々だけではまったく話にならないのでお医者さん達、遺伝カウンセリングの方々、モレキュラー・バイオロジーやられている方々、遺伝学やられている方々に擦り寄っていきながら、少しずつしか進まない、そうすると。データ解析だけで話しをすると話は早いのですが、そういう所でスピードがなかなか出ないというのはありますね。あと、データ解析の方向自体がなかなかデータ自体が大きいので、そこから少しだけサンプリングしてきてうまい結果が出て、じゃあ全体でどうかというと全然駄目だったり、そういうのはちょっと苦労します。

質問者：ありがとうございました。

渡辺：他いかがでしょうか。よろしいでしょうか。では井元先生どうもありがとうございました。

井元：ありがとうございました。

渡辺：ではこれから厚労科研費の報告会に移りたいと思います。(スライド 1) まずどのような思いでこの研究を始めたかというイントロの所を私から話をさせて頂いて、その後で実際のデータをお話いただいて、最後に共同研究の千葉大の並木先生から舌診の測定器の話をしてもらいます。(スライド 2) まず漢方というのは、これはみなさんご存じのように葛根湯という薬なのですね、有名な薬なのですが。葛根湯と一口にいうと簡単なのですが、生薬のレベルでいうと7つ入っています。これは葛です。シナモンとか芍薬とか棗とかこういったものが7つ入っている。この7つが非常に複雑な成分で構成されているというのが漢方というようにお考えください。(スライド 3) これを実際に3次元のHPLCというもので見ると、この測定だと低分子の物しか見られないのですが、それでもこれだけ複雑な成分があるという中で、ここに加えて高分子成分というものが入ってくると非常に複雑な物であるという事がご理解いただけるかと思います。(スライド 4) こういった複雑な物を飲むという訳なのですが、当然生体というものも複雑系な訳ですね。よく学生に講義していて勘違いしているのは、漢方薬と西洋薬はどう違うのだと、そうすると西洋の薬というのはターゲットが決まっています。そこだけに作用すると。それは全くの大嘘ですね。例えば、高脂血症のコレステロールを下げる薬を飲むと筋肉の痛みが出ると。肝臓だけに働くのであればそのような副作用は出るわけがないのですが、生体というのはシステムで動いているのでどこかにミサイルが撃ち込まれると、そこから必ずシステムが動くのです。そういう事で複雑系に対して複雑なもの、さらに漢方という複雑なものをかけると、非常に反応が複雑になると。それを極めて単純化しようとしたのが近代科学という事になります。(スライド 5) 今は漢方薬というのは医者9割が日常診療に用いていると。医者の数が28万人とすると実に26万人くらいが漢方薬を毎日使っているという現状であります。ただその多くの医師というのが漢方的な考えではなくて、この病名にはこの漢方薬というような使い方をしているというのが現状なのですけれども、なぜかという漢方の考え方というのは非常に経験的なのです。漢方の診断を証、あかしという字を書いて証といいます。これが極めて経験値に基づくもので科学的ではないという事であまり勉強しながらないという事があります。そういうドクターでも漢方というのはこういうふうに取り組むのだよという事が目に見えて分かれば使いやすくなるという事で、そういうのをエビデンスというのですが、臨床的なエビデンスというものを示したいという事で申請をしたという事です。こうい

ったエビデンスが分かればこういった勉強をしていない医者でももっと効率のいい効果的な安全性の高い漢方の使い方ができるという事を目指しています。(スライド 6) 今、日本東洋医学会という学会にはエビデンスレポートという臨床研究の結果というのが 360 くらい集まっています。全て英語に訳されていてコクランライブラリーという物に収められているのですけれども、2005 年に調べた時点では極めてお寒い状態で、無作為化比較試験というのが非常に数が少なかったのですね。無作為化比較試験というのは実は西洋医学の中ではゴールドスタンダードといわれる研究方法です。(スライド 7) どういう研究方法かというのは後でお話しますが、なぜ漢方に無作為化比較試験が少ないかという事なのですが、まず漢方というのは個別化の治療です。血圧が高いというので誰もが同じ漢方薬を飲めば良いのではなくて個人個人違うという所で個別化であります。それから患者さんの主観を重視する。ですから評価の方法は血圧を測れば良いという事ではなくて、患者さん側の状態がどうなったかという事を評価しなければいけない。証というものを元にして治療方針を決定しているのだけれどもどうしても臨床の間では分かりやすい研究というのは西洋の病名を使う、体系が違うものを組み合わせなければいけないという所あたりがつらいという所です。もうひとつは医療用として 30 年以上保険が適応されていますので、今更その臨床研究してどうするのという声も実際にはあります。これは西洋医学の中では漢方の世界ではどちらかというとい偉い先生がこんな事を言っているとかなんな事を基にして治療を構築するのですが、それではいかんと。同じような症例を比較したような研究の方がもっとレベルが高いと。(スライド 8) もっとレベルが高いのは振り分けをちゃんとする、だけでも振り分けの時に医者もその振り分けの方法を知っている、この患者さんにはこれを出そうとか分かっているような研究デザイン。もっと上になると医者も患者もどの薬を飲んでいるか分からないというようなのを無作為化比較試験といいます。それをさらに集めて複数の論文を解析するのをメタアナリシスというような言い方で言います。これはその複数の論文の解析なのでこの臨床研究の中では最高水準のものが無作為化比較試験であるということになります。(スライド 9) これはですね、問題発言になるかもしれませんが、非常に莫大な時間と労力と費用を要するのですね。だいたいひとつの薬の承認の為に何百億円が必要です。研究精度を上げれば上げるほど多施設、これは実は人間が仮説を作るのですね。(スライド 10) うちの会社の薬は B 社の会社の薬よりも効くに決まっている。それを検証しようというので、1 万人集めて、それを 2 つに分けてどっちがいいかという事を決めると。非常に莫大なお金をかける割に 1 つの仮説を証明するだけと。その 1 つの仮説は人間がたてたもの。ここに問題がある訳ですね。今中国では先程井元先生からお話があった中にも島というのが出てきましたけれども、中国の中にもある地域を決めて 110 万人分の情報をずっと何十年もかけてトレースする。そうすると本当にたばこを吸っている人が肺ガンになるかという事を 110 万人の中でコントロールのスタディが組めると。ではたばこを吸って人ではなくて、脂肪をたくさん食べる人

はどうなのかという別の見方もできる。それをどんどんコンピューターがその仮説を作って検証していくという事が現実的に起こりつつあります。よく日常診療であるのは、医者というのはどんどんマニュアル化していきましてアメリカのこんなデータだと、この薬は 50%効くんですけど、10%くらいは副作用があって、そのうちの半分 5%くらいは死んでしまうんです。あなたどうしますか、と患者さんに決めさせるのですね。すると患者さんはそんな事言われても困りません。自分はどうすればいいのでしょうか。患者さんが知りたいのは自分の情報なのです。けれども医者が提示するのは集団としてのエビデンスという事でここで患者さんに選ばせるという非常に無理があります。(スライド 11) それでどうすれば良いのかという事なのですが、莫大な費用をかけた動物実験的なものではなく、日常診療の中からデータを集める。それをデータマイニングという方法で意義のある事を見つけるという事を考えた訳です。(スライド 12) データマイニング、ここにいらっしゃる方にはもう新しくないのですが意味のある事を掘り起こすというのがマイニングの意味なので大量のデータから意味のあるものを掘り起こすというのがデータマイニングになります。(スライド 13) 平成 22 年からこのようなメンバーでやってきて、(スライド 14) 組織としては 7 つの大学と 4 つの病院、日本を代表する漢方の病院でデータを集めると、臨床からデータを集めてそれを宮野先生の研究室と美馬先生の研究室によって解析をしていただくという事をやってまいりました。(スライド 15) 簡単にシステムの紹介をします。システムはデモ機が今ありますので後でいじっていただければと思います。これ、今タッチパネルの据え置きモニターもしくは iPad になります。それで患者さんが自由に見られると。(スライド 16) 従来は問診票というのはこのような感じですね。いちいち細かい所はみなくても良いのですが、このような複雑な物をいちいち目で見ながら次はどのようなだけと○をする。(スライド 17) それをタッチパネルの上でやるのですが、(スライド 18) 特に症状の程度を VAS スケールというのがあって、0 から 100 まで実際には数値でいうと 0 から 100 なのですが、これに置き換えられるように自分の症状の程度を自分で入力すると、なぜこれが必要なのかというと、私は医者なので良く分かっていますが、医者のカルテほど当てにならないものはないと。例えば患者さんが、「頭痛はどうなりましたか?」「ちょっと良くなりました。」カルテ「ちょっと良くなった」。どれくらい良くなったのかは全然分からないのですよ。もう少し良くなったですとか、そういう表現で医者は、(スライド 19) これは症例報告なのですが現実には医者の学会発表ではこういうのがまかり通っていてですね、患者さんがこんな事言っていたよという事を適当に引いてしまうのですね。それで発表すると。だけれど患者さんの思いとはちょっと違うのかなと思う時もある訳です。(スライド 20) それで毎回の症状の変化を打ってもらくと、その症状の変化が分かる。それを診療で「ちょっと頭痛が良くなりましたね。」とか会話のきっかけにもなるというようなものがこのシステムであります。(スライド 21) 医者側は西洋の病名と漢方の診断、それから処方の入力という事を行っており

ます。それでは漢方の診断というのはちゃんと標準化されているのかという疑問はあるかと思えますけれどもこれははっきり言ってバラバラです。(スライド 22) このバラバラなものを統一しようという試みが WHO の ICD 国際疾病分類というのがありまして、この中に伝統医学を入れるという事でその中で日本の漢方の標準化を図るという事を逆にそれを利用してやっております。そこで漢方の専門家によって違うのですけれども、なにせ医者の方の 9 割が漢方を使うという事になりますと、26 万人の医者でもちゃんと診断ができるという非常に分かりやすい必要があるという事で、非常にロジカルに分かりやすい体系にしているという事になっています。(スライド 23) 今 ICD11 は Beta 版というのでできておりまして一般公開されておりますので、ICD11 の Beta と入力すると見られます。まあ見たところで英語で、訳分からないものが並んでいるので、あるという事は御確認いただけるのですけれども中身は読まれない事をお勧めします。定義とか読んでもさっぱり分からないと思いますので。ただ 1900 年から 115 年の歴史の中で今度 2015 年に初めて伝統医学が入るという画期的な年になるというのが我々にとっては意義であります。(スライド 24) 本システムの目指すものとしては、漢方処方ごとの個別化のエビデンスの構築と、証というものを標準化して証というものがどういうものかという事を分かりやすく標準化できるといった事、それから漢方の治療の標準化と、こういった事を行って漢方の治療の支援システム、要するに漢方の深い勉強をしなくても漢方を使われている先生方が少しでもレベルアップできるような治療ができるという事を狙っています。(スライド 25) これがなぜ iPad の上に載っかっているかという事、これは患者さんが病院に来る度に入力するのではなくて、自宅にいても症状の変化を入力できるという事を考えたのですけれども、残念ながら今 7 大学の中では大学のセキュリティが非常に厳しくて iPad を、無線を使って待合室から診察室までの 3 m の距離を飛んでいるという状況なのですけれども、これがクラウドに収められるようなものが将来的にはできるだろうという事を考えています。(スライド 26) 今お話したのは医療のレベルの問題なのですけれどもこういった個別化の主観的な医療を必要としているのは医療の世界だけではなくて我々はヘルスと呼んでいる未病の領域といいますか、半健康の方々の支援にもなるという事を考えていまして、そういう方々に対していろいろな健康情報を提供するという事を考えております。(スライド 27) 今まで診療情報というのは医者側のだけだったのですが、患者さんが中心になって患者さんから情報提供してもらおう。医療というよりはヘルスの方になればなるほど医者というよりは患者さんが中心のものになりますので、患者さんが主役の個別化ヘルス情報の時代が来ないかなという事を将来的には見据えているという事でございます。少し長くなりましたけれど、イントロはこのようなところで終わらせていただきたいと思います。どうもありがとうございました。

それでは続いてお話いただいて最後に総合討論という事で質問を受け付けたいと思います。

片山先生は宮野研の中で中心になって我々のデータを解析していただいた先生です。宮野先生