

分担研究課題： 不明症例の病理検体からの新規病原体検索

研究分担者 片野晴隆 国立感染症研究所・感染病理部

研究協力者 福本 瞳、佐藤由子、高橋健太、保科しほ、中島典子、長谷川秀樹（国立感染症研究所・感染病理部）、都築慎也、佐藤典子、望月眞、峰宗太郎、松下竹次（国立国際医療研究センター）、黒田 誠、関塚剛史（国立感染症研究所・病原体ゲノム解析研究センター）、李 天成（国立感染症研究所・ウイルス第2部）、鈴木哲朗（浜松医科大学）、梁 明秀（横浜市大医学部）

研究要旨

原因不明であった感染症症例の病理検体を対象に、網羅的ウイルス遺伝子検出法を用いて、既知の病原体遺伝子を検出すること、および、次世代シーケンサーを用いて、未知の病原体遺伝子を検出することを目的とした。2013年に国立感染症研究所感染病理部にコンサルトされた23例の不明疾患症例につき、163種類のウイルスを同時に網羅的に検出可能な **multivirus real-time PCR** を用い、原因ウイルスの同定を試みた。その結果、8例（35%）で原因ウイルスが同定され、**multivirus real-time PCR** が次世代シーケンサーを行う前のスクリーニングに有効であることが示唆された。また、**multivirus real-time PCR** でも有意なウイルスが検出されなかった3症例の病理検体につき、次世代シーケンサーでの解析を行ったが、有意なウイルスは検出されなかった。一方、2012年に心筋炎患者から次世代シーケンサーを用いて検出された **trichodysplasia-spinulosa associated polyomavirus (TSV)** の解析を行った。当該患者の心筋組織から TSV の全長遺伝子のクローニングに成功し、その全塩基配列を決定し、世界で3番目の株として GenBank に登録した。

A. 研究目的

次世代シーケンサーは感染症検体における微生物遺伝子を検出する、極めて有力なツールであり、データベースに登録されている既知の病原体遺伝子配列の他に、未知の病原体の遺伝子配列も解読することが可能で、しかも、その感度はPCRとほぼ同等である。近年、発見された新しいウイルスの多くは次世代シーケンサーを用いて発見されている（Merkel cell polyomavirus, human polyomavirus 6と7）。しかし、次世代シーケンサーの解析には高額なランニングコストがかかる上、一度のランで極めて多くの遺伝子情報が得られることから、その解析には高度なバイオインフォマティクスの知識と技術を要し、長い時間がかかる。このため、現在ではすべての検体につき、次世代シーケンサーの解析を行うことは現実的でない。次世代シーケンサ

ーの解析にふさわしい症例、サンプルであるかどうかは、多くの臨床情報を集め、患者の状態、症状、緊急性、サンプルの種類、状態、量等から総合的に判断する必要がある。特に、十分な微生物学的検索がなされているかが、重要な点であるが、では、どのような検索が事前になされていればよいか、は基準がない。

国立感染症研究所感染病理部には臨床で感染症が疑われながら、医療機関や地方衛生研究所で病原体が同定できなかった不明感染症例や、これまで感染症が疑われながらも、原因が明らかでない疾患の病理検体が毎年、数十検体、コンサルテーションとして送られてきている（感染病理部リファレンス症例）。われわれはこれらの症例につき、組織形態学的検索や免疫組織化学、PCRなど、様々な手法を駆使して、病原微生物の核酸や蛋白の同定を試みてき

た。また、感染病理部では、160種類以上のウイルスを網羅的に検出できるReal time PCR法(multivirus real-time PCR)を開発し、これらの疾患の原因の解明に努めているが、依然として原因不明のままである症例が多い。本研究では感染病理部に送付された原因不明の感染症疾患の病理検体を対象に、次世代シーケンサーで全遺伝子を解読することで、未知、あるいは、既知の病原体遺伝子の検出を試みた。

また、われわれは2012年に0歳児の心筋炎サンプルから、trichodysplasia-spinulosa associated polyomavirus (TSV)の核酸断片を、次世代シーケンサーにより検出した。TSVは2010年にヨーロッパで発見された新しいヒトポリオーマウイルスであり、8番目のヒトポリオーマウイルスということからヒトポリオーマウイルス8 (human polyomavirus 8, HPyV8)とも呼ばれる。今年度は、本症例から、TSVの全長遺伝子をクローニングし、全遺伝子配列を決定した。

B. 研究方法

1) 臨床検体

全国の医療機関または地方衛生研究所から感染病理部に送付された不明感染症疾患の病理検体を使用した。これにはホルマリン固定パラフィン包埋標本が含まれる。

2) 核酸抽出

ホルマリン固定パラフィン包埋標本からの核酸抽出は Qiagen DNeasy FFPE kit (DNA)および Invitrogen PureLink FFPE RNA extraction kit (RNA)を用いた。凍結サンプル、髄液、血清、鼻咽頭ぬぐい液、尿、便等からの核酸抽出は Qiagen DNeasy kit (DNA), RNeasy Plus kit (RNA)を用いた。

3) real-time PCR によるウイルスの網羅的検出法

ヒトに病原性を持つと考えられる 163 種類のウイルスを 96 穴プレート上で一度に検出できる real-time (RT-)PCR システムを本研究室で独自に開発し、これにより RNA および DNA サンプルからウイルスの検出を試みた (Katano H et al. J Med Virol 2011)。個々のウイルスを検出する定量的 PCR は MX3005P (ストラタジーン社)、または ABI Prism 7900HT (アプライド・バイオシステムズ社)を用いて行った。

4) 次世代シーケンサーによる遺伝子解析 (研究協力者 黒田らによる): RNA ~10 ng をもとに ScriptSeq V2 RNA-seq library preparation kit にて網羅配列解読用のライブラリーを作成した。調整した

ライブラリーを ベンチトップ型次世代シーケンサー MiSeq にて 150 mer x 150 mer のペアエンド法で解読した。検出されたリードは megablast 法にて NCBI nt データベースに配列照合し、MEGAN 5 にて各リードを生物種ごとに分類した。ヒト遺伝子を除き、既知のウイルスゲノムに相同性のある配列を抽出した。

5) TSV 遺伝子の増幅と全長遺伝子のクローニング

心筋凍結組織から抽出した DNA を鋳型に Toyobo KOD-FX を用い、long PCR を行った。増幅された TSV の全長遺伝子は pCR-Blunt にクローニングした。塩基配列の決定は Dye Terminator 法により、通常の DNA シーケンサーを用いて行った。TSV を特異的に検出する real-time PCR では、TSV の VP1 領域である nt 1,841-1,923 (GenBank GU989205)を標的とする Taqman PCR を行った。

(倫理面への配慮)

本研究計画は国立感染症研究所・ヒトを対象とする医学研究倫理審査委員会にて承認済である (H25/7/30 No.417)。試料提供者の個人情報、検体を提出する医療機関において削除され、試料には患者 ID がつけられた状態で感染研に送付される。個人を特定するための対応。表は医療機関が保管する (連結可能匿名化)。したがって、検査実施者が試料提供者個人を特定することはできない。

C. 研究結果

1) 感染病理部リファレンス症例における既知、または未知の病原体遺伝子の検索

2013 年に感染病理部に不明感染症例として送られてきた病理検体のうち、multivirus real-time PCR で解析した症例は 23 例であり、このうちの 8 例 (35%)で、本検査系により病原ウイルスを同定することができた (表 1)。検出されたウイルスの内訳は単純ヘルペスウイルス 2 型、水痘帯状疱疹ウイルス、コクサッキーウイルス A6 型、コクサッキーウイルス B3 型 (2 例)、日本脳炎ウイルスであった。multivirus real-time PCR では、TTV やヒト内因性レトロウイルスの遺伝子が検出される例が少なくないが、これらのウイルス遺伝子はヒトの健常者や健常部位からも検出されるものであり、有意ではない。上記、ウイルスの多くは病理組織(生検、剖検)において、免疫組織学的に確認されたもので、病理学的に原因と考えられるウイルスである。

さらに、multivirus real-time PCR で有意なウイルスが検出されなかった症例 3 例については病理検体から抽出した RNA について、次世代シーケンサーの解析を行った (表 2)。検索の結果、有意なウイルスを検出することができなかった。表 2 に示す第 2 例では Avian leukosis virus の遺伝子断片が検出されているが、これは、逆転写酵素に含まれるものであり、非特異的なものであることが判明している。また、3 例目は EBV が関連することがすでに分かっており、EBV 以外のウイルスの関与を期待したが、有意なウイルスは検出されなかった。

2) 心筋炎から検出された TSV の検討

2012 年に感染病理部にリファレンス症例として解析依頼のあった、0 歳児の心筋炎症例から次世代シーケンサーの解析により TSV の遺伝子断片が検出された。今年度はこの心筋の凍結検体から TSV の遺伝子を増幅し、その全長遺伝子の配列決定を行った。心筋から抽出した DNA からは long PCR により 5.2 kpb におよぶ TSV の全長遺伝子の増幅が可能であった (図 1)。PCR 産物を pCR-Blunt ベクターにクローニングし、TSV 遺伝子の全長遺伝子配列を決定した。シーケンスした遺伝子配列は既報告の 2 つの TSV と 99% の相同性があり、他のヒトポリオーマウイルスとの相同性は、HPyV9 と 52%、その他は 50% 以下であり、本遺伝子配列が TSV であることが確認された。TSV-TMC 株 (Tokyo MyoCarditis が由来) と命名し、GenBank に登録した (accession no. AB873001)。TSV-TMC 株は世界で 3 番目に報告された TSV の全長遺伝子配列となった。

本症例は剖検例であり、脳を除くほぼすべての臓器が採取されていることから、臓器別の TSV 量を、TSV 遺伝子の特異的に検出する real-time PCR により測定した。その結果、心臓に TSV が濃縮されていることが明らかになった (図 3)。さらに、TSV-TMC 株から、VP1 の遺伝子を増幅し、現在、VP1 に対するモノクローナル抗体を作製している (研究協力者 梁らによる)。

D. 考 察

不明感染症例の一般的な定義は存在せず、病院や医療機関の多くの医師は、不明感染症例とされる症例を日常的に経験している。その中には、軽症のため、十分な微生物検査がなされる前に軽快し、原因不明とされた症例もあれば、重症例でさまざまな微生物検査がなされたにもかかわらず、原因が不明で、

不明感染症例とされている症例もあるであろう。地方衛生研究所等から微生物検索依頼がなされた症例は、原則として疑わしい微生物に対する検索が終了しているものである。次世代シーケンサーの解析は、サンプルに含まれるすべての遺伝子情報を解読するものであり、病因微生物の遺伝子が含まれる場合には、かなり高い確率で検出可能である。しかし、不明感染症例のすべての症例を解析することは予算的にも、時間的にも不可能である。不明感染症例の、どの検体を次世代シーケンサーの解析を行うかについての一般的な選択基準は存在しないが、目安となる検査やスクリーニング法が存在すれば、効率的な検査が可能かも知れない。本研究でわれわれが用いた multivirus real-time PCR は次世代シーケンサーよりも手軽に、低予算で、多くのウイルスを網羅的にスクリーニングできるシステムであり、次世代シーケンサーに進む前の検査としては極めて有用なスクリーニング法と考えられる。

われわれが病理の検体を検索対象としている理由は 2 つある。一つは病理検体はすでに組織学的診断が付いているものがほとんどであり、明らかな感染症以外の疾患を除外することができる点である。組織学的検索により病変部が確実に含まれているかどうか分かる点は、不明感染症の検索には極めて有利である。もう一つの理由は、multivirus real-time PCR や次世代シーケンサーで検出された微生物が、果たして病因微生物であるかどうかを、病理検体に戻って、病理組織学的に検索が可能である点である。血液などの体液サンプルでは、病変部位を反映した結果が得られず、最終的な結論が出ないことが多いのに対し、病理検体は病変部を直接、組織学的に免疫組織化学や in situ hybridization などで検索することにより、検出微生物が病変部と関連があったかどうか、結論が得られる確率が高い。病原体遺伝子を検出する網羅的微生物検出法を利用した際には、PCR も次世代シーケンサーも、極めて感度の高い方法であるために、有意でない微生物遺伝子がしばしば検出される。病理検体以外のサンプルではこの判定がむずかしく、今後、多くのサンプルを検討することにより、検出された微生物が病因微生物であるかどうかを判定する方法、適切なサンプルを選択する方法が次第に明らかになっていくことが期待される。

TSV はこれまでのところ、顔面皮膚の高角化症である trichodysplasia-spinulosa 以外には疾患との関連は明らかにされていない。今回の心筋炎の症例で

は心臓に TSV が濃縮されていること、心臓から TSV の全長遺伝子が検出されていること、また、Ori を挟んだプライマーで PCR を行ったところ、増幅が可能であったこと（データは示していない）から、心臓には環状の TSV が存在していたことが示され、これは心臓で、TSV の複製が起きていた可能性を示唆している。これまで、TSV の報告例はほとんどが皮膚に限定されており、心臓から検出されたとする報告はない。心筋炎の原因ウイルスはエンテロウイルスやヘルペスウイルスなどが知られるが、感染病理部のリファレンス例でも、多くの心筋炎の症例が原因微生物不明であり、TSV は新たな心筋炎の原因ウイルスであるかも知れない。今後、作成中の抗体を用いた組織学的検索の結果が待たれるところである。

E. 結論

不明感染症例の病理組織検体を対象に、multivirus real-time PCR を用い、原因ウイルスの同定を試み、検索した症例の 35% で原因ウイルスを同定した。Multivirus real-time PCR は次世代シーケンサーを行う検体のスクリーニング法として有効な手段である。また、3 症例の病理検体につき、次世代シーケンサーでの解析を行ったが、有意なウイルスは検出されなかった。心筋炎症例から TSV の全長遺

伝子のクローニングし、全塩基配列を決定した。

F. 健康危険情報 特になし

G. 研究発表

1) 論文発表 なし

2) 学会発表

- (1) 片野晴隆、佐藤由子、中島典子、福本瞳、鈴木忠樹、黒田誠、長谷川秀樹 病理検体からの不明病原体検出法の最先端 ワークショップ 「感染病理学の新展開」第 102 回 日本病理学会総会. 札幌。2013.4.
- (2) 中島典子、片野晴隆 定量的 PCR によるウイルスの網羅的検出法と病理検体への応用 シンポジウム 3 病原体の新しい診断法 第 18 回日本神経感染症学会総会学術集会 宮崎 2013 年 10 月

H. 知的財産権の出願・登録状況（予定を含む） 該当なし

表1 2013年にMultivirus real-time PCRで検索した病理検体の結果。HSV: herpes simplex virus, VZV: varicella zoster virus, Cox: coxsackievirus, JEV: Japanese encephalitis virus.

| 臨床診断 | 検索症例数 | ウイルス検出症例数 | 検出ウイルス |
|---------|-------|-----------|------------------------|
| 脳炎、髄膜炎 | 9 | 4 | HSV-2, VZV, CoxA6, JEV |
| 心筋炎 | 7 | 3 | CoxA2, B3 |
| リンパ節腫脹 | 2 | 0 | — |
| 血液貪食症候群 | 2 | 0 | — |
| その他 | 3 | 1 | VZV |
| 合計 | 23 | 8 | — |

表2 2013年に次世代シーケンサーの解析を行った病理検体。SFTSV: Severe fever with thrombocytopenia syndrome virus, EBV: Epstein-Barr virus.

| No. | 疾患名 | 年齢、性別 | 組織 | 結果 |
|-----|-------------|---------|--------------------------|---|
| 1 | 急性心筋炎 | 10歳代 女性 | 剖検組織 | 有意な微生物遺伝子は検出されない。 |
| 2 | 亜急性壊死性リンパ節炎 | 50歳代 男性 | ダニに咬傷後の所属リンパ節腫大。SFTSV陰性。 | 有意な微生物遺伝子は検出されない。 (Avian leukosis virus →逆転写酵素由来) |
| 3 | 形質芽細胞リンパ腫 | 60歳代 男性 | EBV陽性。 | EBVのみ。 |

表3 TSV-TMC株と既報告のTSVおよび他のヒトポリオーマウイルスとの相同性。TSV: trichodysplasia-spinulosa associated polyomavirus, HPyV: human polyomavirus.

| Virus gene (GenBank Accession No.) | Homology (%) |
|--------------------------------------|--------------|
| TSV (JQ723730) | 99.4 |
| TSV (GU989205) | 99.4 |
| HPyV9 (NC_015150) | 52.0 |
| KI polyomavirus (NC_009238) | 49.88 |
| JC virus (PLYCG) | 49.53 |
| BK virus (PLYCGAS) | 49.39 |
| Merkel cell polyomavirus (NC_010277) | 49.34 |
| HPyV7 (NC_014407) | 47.70 |
| HPyV6 (NC_014406) | 47.55 |

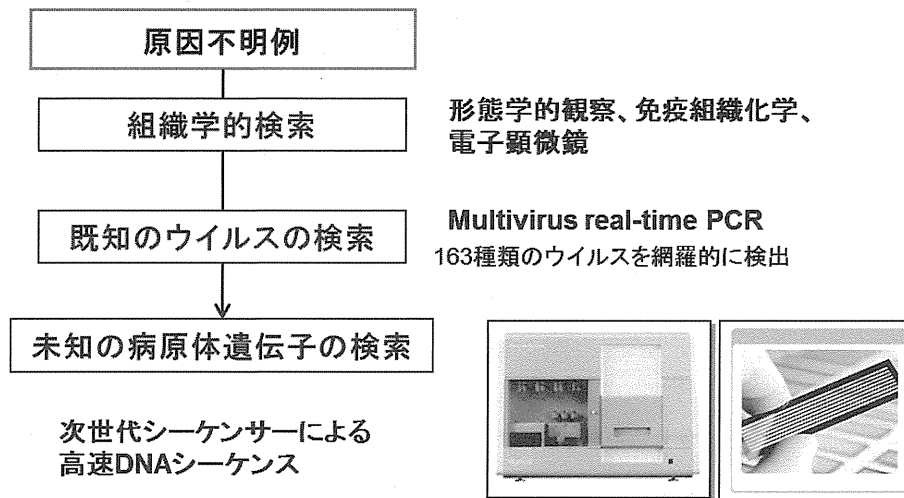


図1 本研究における原因不明症例の病原微生物の検出の概要

原因不明症例の病理検体は、まず、HE染色などの組織学的検索がなされ、明らかな感染症以外の疾患は除外される。また、検体に病変部が含まれるかどうか、組織学的に検索される。組織学的に感染症が疑われる検体では、考えられる微生物を検出する免疫染色や特殊染色を行なう。必要により、電子顕微鏡でウイルス粒子などを観察する。組織学的にウイルス感染が疑われるものの、病理組織学的な検索で原因ウイルスを同定できないサンプルは、multivirus real-time PCRによりウイルスの網羅的検索を行う。Multivirus real time PCRは、multiplex Taqman real time PCRを応用した網羅的ウイルス検出キットで、感染病理部で独自に開発したものである(Katano et al. J Med Virol 2011 83:322-330)。各ウイルスの probe-primer セットを 96 穴プレートに配し、Fam と Hex の二つの蛍光標識プローブを使用することで、ひとつの well で 2 種類のウイルスを検出できるよう設計されている。また、同一プレート上で定量線を作成し、各ウイルス量がだまかに計測可能である。各 probe-primer セットは約 100bp のウイルス核酸断片を検出できるよう設計されており、パラフィン切片から抽出した核酸など、断片化した核酸にも対応可能である。Multivirus real-time PCR でも、原因ウイルスの同定に至らない検体が次世代シーケンサーの解析対象となる。

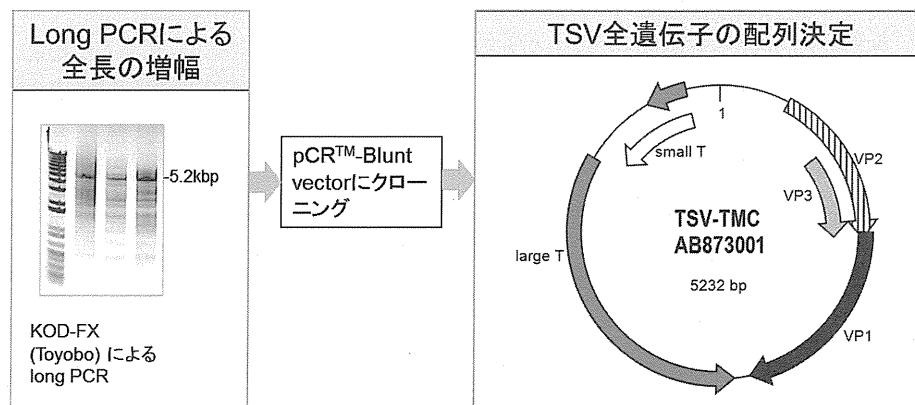


図2 TSVの全長遺伝子のクローニングと塩基配列の決定

心筋組織から抽出したDNAより long PCR により、TSV 遺伝子の全長 (5.2 kbp) の増幅が可能であった (左図)。PCR 産物を pCR-Blunt vector にクローニングし、TSV 遺伝子の全長の塩基配列を決定した。塩基配列は TSV-TMC 株として、GenBank に登録されている (accession no. AB873001)。右図には TSV-TMC 株の遺伝子構造を示す。既報告の TSV と遺伝子構造は変わらない。

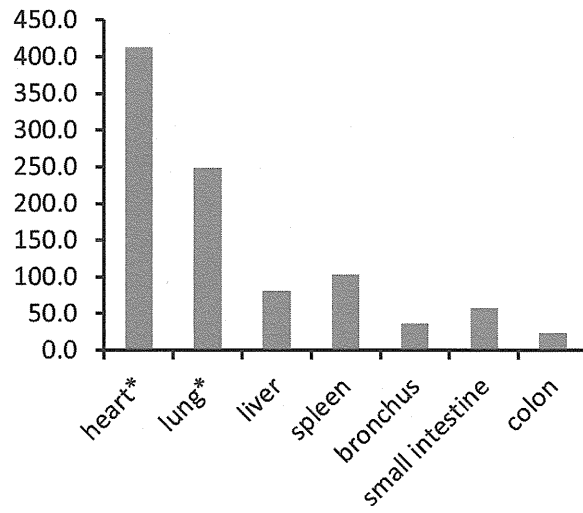


図3 TSV が検出された心筋炎患者の各臓器における TSV 量の定量。

TSV の VP1 を特異的に検出する real-time PCR により、各臓器から抽出した DNA を検索した。縦軸は DNA 100ng あたりの TSV コピー数を示す。心臓が最も TSV 量の高い臓器であり、心筋炎との関連が示唆される。心臓*と肺*以外はパラフィン切片による検索である。

迅速・網羅的病原体ゲノム解析法を基盤とした感染症対策ネットワーク構築に関する研究
～病原体網羅遺伝子解析を基盤にしたプロテオーム解析による抗原解析と新規病原体検査法の
開発

研究分担者 梁 明秀 横浜市立大学医学部微生物学
研究協力者 松永智子 横浜市立大学医学部微生物学

研究要旨 近年、次世代核酸シーケンサを用いた新興・再興感染症に対するメタゲノム解析が進歩し、新たな病原体の同定や病因の解明が進んでいる。本研究プロジェクトでは、網羅遺伝子解析を基盤とした病原体のプロテオーム解析を実施し、それを基盤とした病原体に対する新規検査診断法の開発を目指している。網羅的遺伝子情報に基づくウイルス抗原の解析やウイルスタンパク質の機能解析等について、コムギ無細胞タンパク質合成系を用いたプロテオミクスを活用する。本年度は新規病原体 *Trichodysplasia spinulosa-associated polyomavirus* (TSV) の VP1 タンパク質を、コムギ無細胞系を用いて作製し、本抗原に対するモノクローナル抗体の作製を試みた。本抗体はウエスタンブロットにおいて VP1 タンパク質およびウイルス様粒子を特異的に認識した。本抗体を用いることで TSV に対する新たな検査法や診断法の開発に結びつくものと考えられる。

A. 研究目的

感染症の疑いのある不明疾患やバイオテロ、新興・再興感染症などによるアウトブレイク対策のための迅速・網羅的病原体解析法を基盤とした感染症対策ネットワークシステムの構築が重要である。一方で、感染症の疑いのある不明疾患等のための迅速・網羅的病原体解析法として、ウイルス抗原の検出や血清中の抗ウイルス抗体の測定法を整備することが必要となる。

次世代シーケンシングの進歩に伴ってメタゲノム研究の分野は大きく発展し、不明感染症の病原体由来のゲノム断片を多数検出することが可能である。しかしながら、疾患検体に存在するゲノム断片のみでは、当該病原体の疾患病因との関与について確定することは難しく、核酸検査と平行して疾患臓器における病原体抗原の存在および宿主血清中の病原体特異的抗体の存在を証明すべきである。そのためには、病原体網羅遺伝子解析を基盤にしたプロテオーム解析による抗原解析と新規病原体検査法の開発が必須である。

近年、新興・再興感染症に対する抗原・抗体診断法の開発が進んでいる。しかしながら、従来の手

法は、大腸菌や培養細胞への遺伝子導入によりウイルスタンパク質の合成が基盤であり、細胞毒性が強く、かつ可溶性の低いウイルス抗原タンパク質の作製には不向きであった。今回、我々は、コムギ無細胞タンパク質合成系を用いて新規病原体の可溶化全長タンパク質を作製し、これを抗原として用いることで、免疫学的診断に利用可能な高品質のモノクローナル抗体の作製を行った。

B. 研究方法

1. コムギ無細胞系による TSV ウウイルスタンパク質の合成

Trichodysplasia spinulosa-associated polyomavirus (TSV) がコードする VP1 遺伝子を PCR 法を用いて増幅し、無細胞タンパク質発現ベクター pEU-bls-S1 (bls; biotin ligation site GLNDIFEAQKIEWHE, S1: linker sequence LHPPPRIS) に導入した。作製した pEU ベクターを鋳型に SPu primer 及び AODA2303 primer を用いて PCR 法により転写鋳型を作製し、SP6 polymerase を用いた転写反応により mRNA を合成、続いてコムギ無細胞合成系・重層法によりタンパ

ク質合成を行った。タンパク質のビオチン化は、下層にコムギ無細胞系で合成した biotin ligase 1 μ l (~50ng/ μ l) および 終濃度 0.5 μ M Biotin を加えることにより行った。免疫用抗原の作製においては、His タグを付加した VP1 タンパク質を、ウイルスタンパク質の可溶性を亢進させるために界面活性剤である Brij35 (0.5%) 存在下にて合成後、Ni-sepharose ビーズにタンパク質を吸着させた。カラムを 8 M の尿素を含む洗浄液にて 2 回洗浄した後、500mM イミダゾールバッファーを用いて精製タンパク質を抽出した。

2. モノクローナル抗体の作製

コムギ無細胞系を用いて合成したウイルス抗原タンパク質に keyhole limpet hemocyanin(KLH)を共有結合させたものを免疫源とした。これを 6 週齢の Balb/c マウスに foot pad法で2週間置きに1ヶ月間免疫した。免疫後、マウス脾細胞を採取し、マウスミエロマ細胞である SP2/O と PEG 法で融合しハイブリドーマを作製した。

3. ELISA 法

96 well plate に終濃度で 50 ng/well の抗原を一晩コートした。反応後、バッファー液を除きブロッキング剤を加えて室温で 1 時間静置した。その後 PBS で 3 回洗浄後、各倍希釈したハイブリドーマ上清を加え室温で 1 時間反応させた。次に PBS で 3 回洗浄後 HRP 標識抗マウス IgG 抗体を室温で 1 時間反応させた。これを PBS で 3 回洗浄後発色基質を加え、30 分後 1 N 硫酸で反応を停止させマイクロプレートリーダーで吸光度を測定した。

(倫理面への配慮)

本研究において、遺伝子組換え実験を用いることから、研究者が所属する機関の組換え DNA 実験安全委員会、バイオセーフティ委員会、動物実験委員会、医学研究倫理委員会等の承認・認可を得て実験を行った。また、本年度はヒト検体を使用した実験を実施していないが、臨床サンプルの解析

及びデータの公表にあたっては、倫理委員会の規則にのっとり、当該患者（感染者）の同意を得る予定である。

C. 研究結果

1. TSV-VP1 タンパク質の合成

コムギ無細胞タンパク質合成系を用いて作製した TSV-VP1 タンパク質は、概ね不溶化タンパク質であった。しかしながら、翻訳反応液中に界面活性剤の Brij35 またはリポソームを添加することで、可溶性率が顕著に亢進した。これらの可溶性タンパク質を精製し、モノクローナル抗体作製のための抗原とした。

2. TSV-VP1 モノクローナル抗体の作製

上記にて作製した抗原タンパク質を Balb/c マウスに foot pad 法で 2 週間置きに 2 回免疫した。1 ヶ月後マウスの脾細胞を採取し、マウスミエロマ細胞である SP2/O と PEG 法で融合し 48 種類のハイブリドーマを得た。1 次スクリーニングとして ELISA 法を用いて活性を確認した結果、48 種類のうち、18 種類のハイブリドーマ上清が活性保持していることが明らかとなった。

3. ウェスタンブロット解析による抗原認識の確認

次に上記にて選択した 18 種類のハイブリドーマ上清がウェスタンブロット解析に使用できるか否かについて検討した。VP1 タンパク質が、培養細胞において発現がほとんど見られなかったため、リコンビナント VP1 タンパク質を抗原として用いた。その結果 18 種類すべてのハイブリドーマ上清がウェスタンブロット解析において TSV-VP1 タンパク質を認識できることが確認された。また、VP1 を有するウイルス様粒子を用いた解析においても同様に VP1 タンパク質を検出することができた。現在本抗体の免疫組織化学染色に応用可能かどうかについて検討中である。

D. 考察

本研究において、合成した可溶性 TSV-VP1 タンパク質は、機能および構造が保持されている状態であると考えられることから、生体内で誘導される構造を認識する抗体の検出も可能であると示唆される。また、全長タンパク質を用いることで、様々なエピトープに対する抗体が検出されることが期待できる。異なるエピトープを認識する抗体を取得することで、ELISA やイムノクロマト法などの免疫学的アッセイ法の開発に結びつくものと考えられる。

E. 結論

コムギ無細胞タンパク質合成システムを活用して作製した全長 TSV-VP1 タンパク質の精製に成功した。また、本抗原に対するモノクローナル抗体を作製した。今後は、免疫組織化学や免疫沈降法などに有用な抗体クローンを選別することで、TSV の診断や病態解析に貢献できる抗体が得られるものと考えられる。

G. 研究発表

1. 論文発表

- (1) Kimura H, Yoshizumi M, Ishii H, Oishi K, Ryo A. Cytokine production and signaling pathways in respiratory virus infection. *Front Microbiol.* 2013 Sep 17;4:276.
- (2) Kushibuchi I, Kobayashi M, Kusaka T, Tsukagoshi H, Ryo A, Yoshida A, Ishii H, Saraya T, Kurai D, Yamamoto N, Kanou K, Saitoh M, Noda M, Kuroda M, Morita Y, Kozawa K, Oishi K, Tashiro M, Kimura H. Molecular evolution of attachment glycoprotein (G) gene in human respiratory syncytial virus detected in Japan 2008-2011. *Infect Genet Evol.* 2013 Aug;18:168-73.
- (3) Ishigami T, Abe K, Aoki I, Minegishi S, Ryo A, Matsunaga S, Matsuoka K, Takeda H, Sawasaki T, Umemura S, Endo Y. Anti-interleukin-5 and multiple autoantibodies are associated with human atherosclerotic diseases and serum interleukin-5 levels. *FASEB J.* 2013 Sep;27(9):3437-45.

- (4) Kiyota N, Kushibuchi I, Kobayashi M, Tsukagoshi H, Ryo A, Nishimura K, Hirata-Saito A, Harada S, Arakawa M, Kozawa K, Noda M, Kimura H. Genetic analysis of the VP4/VP2 coding region in human rhinovirus species C in patients with acute respiratory infection in Japan. *J Med Microbiol.* 2013 Apr ; 62(Pt 4):610-7.

2. 学会発表等

なし

H. 知的財産権の出願・登録状況

なし

研究成果の刊行に関する一覧表

雑誌

| 発表者氏名 | 論文タイトル名 | 発表誌名 | 巻・号 | ページ | 出版年 |
|---|---|----------------------|--------|---------|-------|
| Fumihiko Takeuchi, Tsuyoshi Sekizuka, Akifumi Yamashita, Yumiko Ogasawara, Katsumi Mizuta, and Makoto Kuroda. | MePIC, Metagenomic Pathogen Identification for Clinical Specimens. | Jpn. J. Infect. Dis. | 67 (1) | 62-65. | 2014 |
| Senchi K, Matsunaga S, Hasegawa H, Kimura H, Ryo A. | Development of oligomannose-coated liposome-based nasal vaccine against human parainfluenza virus type 3. | Front in Microbiol. | 4: | 346 | 2013. |
| Kimura H, Yoshizumi M, Ishii H, Oishi K, Ryo A. | Cytokine production and signaling pathways in respiratory virus infection. | Front Microbiol | 4 | 276 | 2013 |
| Kushibuchi I, Kobayashi M, Kusaka T, Tsukagoshi H, Ryo A, Yoshida A, Ishii H, Saraya T, Kurai D, Yamamoto N, Kanou K, Saitoh M, Noda M, Kuroda M, Morita Y, Kozawa K, Oishi K, Tashiro M, Kimura H. | Molecular evolution of attachment glycoprotein (G) gene in human respiratory syncytial virus detected in Japan 2008-2011. | Infect Genet Evol | 18 | 168-73 | 2013 |
| Ishigami T, Abe K, Aoki I, Minegishi S, Ryo A, Matsunaga S, Matsuoka K, Takeda H, Sawasaki T, Umemura S, Endo Y. | Anti-interleukin-5 and multiple autoantibodies are associated with human atherosclerotic diseases and serum interleukin-5 levels. | FASEB J. | 27(9) | 3437-45 | 2013 |

| | | | | | |
|--|--|----------------------|----------|---------|------|
| Kiyota N, Kushibuchi I, Kobayashi M, Tsukagoshi H, Ryo A, Nishimura K, Hirata-Saito A, Harada S, Arakawa M, Kozawa K, Noda M, Kimura H. | Genetic analysis of the VP4/VP2 coding region in human rhinovirus species C in patients with acute respiratory infection in Japan. | J Med Microbiol. | 62(Pt 4) | 610-7 | 2013 |
| Kobayashi M, Takayama I, Kageyama T, Tsukagoshi H, Saitoh M, Ishioka T, Yokota Y, Kimura H, Tashiro M, Kozawa K. | A new reassortant swine influenza A (H1N2) virus derived from A (H1N1) pdm09 virus isolated from swine. | Emerg Infect Dis. | 19(12): | 1972-4 | 2013 |
| Yamazaki M, Sugai K, Kobayashi Y, Kaburagi Y, Murashita K, Saito N, Niino H, Imagawa T, Tsukagoshi H, Kimura H. | A child case of hypocomplementemic urticarial vasculitis due to Coxsackievirus type A9. | J Med Microbiol Case | 1(1): | 1-5 | 2014 |
| Miyaji Y, Kobayashi M, Sugai K, Tsukagoshi H, Niwa S, Fujitsuka-Nozawa A, Noda M, Kozawa K, Yamazaki F, Mori M, Yokota S, Kimura H. | Respiratory severity and virus profiles in Japanese children with acute respiratory illness. | Microbiol Immunol. | 57(12): | 811-821 | 2013 |
| Kiyota N, Kobayashi M, Tsukagoshi H, Ryo A, Harada S, Kusaka T, Obuchi M, Shimojo N, Noda M, Kimura H. | Genetic analysis of human rhinovirus species A to C detected in patients with acute respiratory infection in Kumamoto prefecture, Japan 2011–2012, | Infect Mol Evol. | 21: | 90-102 | 2013 |

| | | | | | |
|---|---|---------------------------|---------|-------------|------|
| Abe M, Tahara M, Yamaguchi H, Kanou K, Shirato K, Kawase M, Noda M, Kimura H, Matsuyama S, Fukuhara H, Mizuta K, Maenaka K, Ami Y, Esumi M, Kato A, Takeda M. | TMPRSS2 is an activating protease for respiratory parainfluenza viruses. | J Virol. | 87(21): | 11930-11935 | 2013 |
| Matsuda S, Nakamura M, Hirano E, Kiyota N, Omura T, Suzuki Y, Noda M, Kimura H. | Characteristics of human metapneumovirus infection prevailing in hospital wards housing patients with severe disabilities. | Jpn J Infect Dis. | 66(3): | 195-200 | 2013 |
| Tsukagoshi H, Yokoi H, Kobayashi M, Kushibuchi I, Okamoto-Nakagawa R, Yoshida A, Morita A, Noda M, Yamamoto N, Sugai K, Oishi K, Kozawa K, Kuroda M, Shirabe K, Kimura H. | Genetic analysis of attachment glycoprotein (G) gene in new genotype ON1 of human respiratory syncytial virus detected in Japan. | Microbiol Immunol. | 57(9): | 655-659 | 2013 |
| Ishioka T, Yamada Y, Kimura H, Yoshizumi M, Tsukagoshi H, Kozawa K, Maruyama K, Hayashi Y, Kato M. | Elevated macrophage inflammatory protein 1 α and interleukin-17 production in an experimental asthma model infected with respiratory syncytial virus. | Int Arch Allergy Immunol. | 161(2): | 129-137 | 2013 |
| Mizuta K, Yamakawa T, Nagasawa H, Itagaki T, Katsushima F, Katsushima Y, Shimizue Y, Ito S, Aokia Y, Ikeda T, Abiko C, Kuroda M, Noda M, Kimura H, Ahiko T. | Epidemic myalgia associated with human parechovirus type 3 infection among adults occurs during an outbreak among children: Findings from Yamagata, Japan, in 2011. | J Clin Virol. | 58(1): | 188-193 | 2013 |

| | | | | | |
|--|--|-------------------|-----------|-----------|------|
| Saraya T, Mikoshiha M, Kamiyama H, Yoshizumi M, Tsuchida S, Tsukagoshi H, Ishioka T, Terada M, Tanabe E, Tomioka C, Ishii H, Kimura H, Kozawa K, Shiohara T, Takizawa T, Goto T. | Evidence for reactivation of human herpes virus 6 in generalized lymphadenopathy in a patient with drug induced hypersensitivity syndrome. | J Clin Microbiol. | 51(6): | 1979-1982 | 2013 |
| Mizuta K, Abiko C, Aoki Y, Ikeda T, Matsuzaki Y, Itagaki T, Katsushima F, Katsushima Y, Noda M, Kimura H, Ahiko | T. Seasonal patterns of respiratory syncytial virus, influenza A virus, human metapneumovirus, and parainfluenza virus type 3 infections based on virus isolation data between 2004 and 2011 in Yamagata, Japan. | Jpn J Infect Dis. | 66(2): | 140-145 | 2013 |
| Kiyota N, Kushibuchi I, Kobayashi M, Tsukagoshi H, Ryo A, Nishimura K, Hirata-Saito A, Harada S, Arakawa M, Kozawa K, Noda M, Kimura H. | Genetic analysis of VP4/VP2 coding region in human rhinovirus species C detected from the patients with acute respiratory infection in Japan. | J Med Microbiol. | 62, | 610-617 | 2013 |
| Seki E, Yoshizumi M, Tanaka R, Ryo A, Ishioka T, Tsukagoshi H, Kozawa K, Okayama Y, Goya T, Kimura H. | Cytokine profiles, signaling pathways, and effects of fluticasone propionate in respiratory syncytial virus-infected human fetal lung fibroblasts. | Cell Biol Int. | in press. | | 2014 |
| Nishi M, Sakai Y, Akutsu H, Nagashima Y, Quinn G, Masui S, Kimura H, Perrem K, Umezawa A, Yamamoto N, Lee SW, Ryo A. | Induction of cells with cancer stem-cell properties from non-tumorigenic human mammary epithelial cells by defined reprogramming factors. | Oncogene | in press. | | 2014 |

| | | | | | |
|--|---|-------------------|--------|----------|------|
| Abiko C, Mizuta K, Aoki Y, Ikeda T, Itagaki T, Noda M, Kimura H, Ahiko T. | An outbreak of parainfluenza virus type 4 infections among children with acute respiratory infections, in the 2011-12 winter season in Yamagata, Japan. | Jpn J Infect Dis. | 66(1): | 76-78 | 2013 |
| Nakamura M, Hirano E, Ishiguro F, Mizuta K, Noda M, Tanaka R, Tsukagoshi H, Kimura H. | Molecular epidemiology of human metapneumovirus from 2005 to 2011 in Fukui, Japan. | Jpn J Infect Dis. | 66(1): | 56-59 | 2013 |
| Kobayashi M, Tsukagoshi H, Ishioka T, Mizuta K, Noda M, Morita Y, Ryo A, Kozawa K, Kimura H. | Seroepidemiology of saffold cardiovirus (SAFV) genotype 3 in Japan. | J Infect. | 66(2): | 191-193 | 2013 |
| Okada S, Hasegawa S, Hasegawa H, Ainai A, Atsuta R, Ikemoto K, Sasaki K, Toda S, Shirabe K, Takahara M, Harada S, Morishima T, Ichiyama T. | Analysis of bronchoalveolar lavage fluid in a mouse model of bronchial asthma and H1N1 2009 infection. | Cytokine. | 63(2): | 194-200. | 2013 |
| | | | | | |

学会発表一覧表

| 発表者氏名 | 発表タイトル名 | 学会名 | 開催年月日 | 開催地 |
|---|---|--|-----------------|-----------------|
| Miyaji Y, Kobayashi M, Sugai K, Tsu-kagoshi H, Niwa S, Fujitsuka-Nozawa A, Noda M, Kozawa K, Yamazaki F, Mori M, Yokota S, Kimura H. | Severity of respiratory signs and symptoms and virus profiles in Japanese children with acute respiratory illness. | Asia Pacific Congress of Asthma, Allergy, and Clinical Immunology (AP-CAACI) | Nov 14-17, 2013 | Taipei, Taiwan. |
| Obuchi M, Hatasaki K, Tsubata S, Kaneda H, Shinozaki K, Tsuji T, Kasei M, Konishi M, Inasaki N, Obara-Nagoya M, Horimoto E, Sata T and Takizawa T. | Detection of Resistant Influenza A(H3N2) Virus in Children Treated with Neuraminidase Inhibitors Using a Next-Generation DNA Sequencer. | 53rd Interscience Conference on Antimicrobial Agents and Chemotherapy. | 2013年9月 | 米国デンバー |
| 小渕正次、畑崎喜芳、津幡眞一、金田尚、篠崎健太郎、辻隆男、紺井正春、小西道雄、稲崎倫子、名古屋(小原)真弓、堀元栄詞、佐多徹太郎、滝澤剛則。 | ノイラミニダーゼ阻害薬投与小児患者における薬剤耐性A(H3N2)インフルエンザウイルスの検出。 | 第61回日本ウイルス学会学術集会 | 2013年11月 | 神戸 |
| 片野晴隆、佐藤由子、中島典子、福本瞳、鈴木忠樹、黒田誠、長谷川秀樹 | 病理検体からの不明病原体検出法の最先端ワークショップ「感染病理学の新展開」 | 第102回日本病理学会総会。 | 2013.4. | 札幌 |
| 中島典子、片野晴隆 | 定量的PCRによるウイルスの網羅的検出法と病理検体への応用シンポジウム3 病原体の新しい診断法 | 第18回日本神経感染症学会総会学術集会 | 2013年10月 | 宮崎 |

Short Communication

MePIC, Metagenomic Pathogen Identification for Clinical Specimens

Fumihiko Takeuchi¹, Tsuyoshi Sekizuka¹, Akifumi Yamashita¹,
Yumiko Ogasawara¹, Katsumi Mizuta², and Makoto Kuroda^{1*}

¹*Pathogen Genomics Center, National Institute of Infectious Diseases, Tokyo 162-8640; and*

²*Department of Microbiology, Yamagata Prefectural Institute of Public Health,
Yamagata 990-0031, Japan*

(Received July 10, 2013. Accepted September 26, 2013)

SUMMARY: Next-generation DNA sequencing technologies have led to a new method of identifying the causative agents of infectious diseases. The analysis comprises three steps. First, DNA/RNA is extracted and extensively sequenced from a specimen that includes the pathogen, human tissue and commensal microorganisms. Second, the sequenced reads are matched with a database of known sequences, and the organisms from which the individual reads were derived are inferred. Last, the percentages of the organisms' genomic sequences in the specimen (i.e., the metagenome) are estimated, and the pathogen is identified. The first and last steps have become easy due to the development of benchtop sequencers and metagenomic software. To facilitate the middle step, which requires computational resources and skill, we developed a cloud-computing pipeline, MePIC: "Metagenomic Pathogen Identification for Clinical Specimens." In the pipeline, unnecessary bases are trimmed off the reads, and human reads are removed. For the remaining reads, similar sequences are searched in the database of known nucleotide sequences. The search is drastically sped up by using a cloud-computing system. The webpage interface can be used easily by clinicians and epidemiologists. We believe that the use of the MePIC pipeline will promote metagenomic pathogen identification and improve the understanding of infectious diseases.

Next-generation DNA sequencing technologies have led to a new method of identifying the causative agent of infectious diseases in hospitalized patients and during outbreaks (1,2). By directly sequencing millions of DNA/RNA molecules in a specimen and matching the sequences to those in a database, pathogens can be inferred. The analysis comprises three steps. First, the nucleotide sequences of the specimen, which includes the pathogen, human tissue and commensal microorganisms, are read using a next-generation sequencer. Second, from bioinformatic processing of the reads, the organisms from which the individual reads were derived are inferred. Last, the percentages of the organisms' genomic sequences in the specimen are estimated, and the pathogen is identified. Although the first and last steps have become easy due to the development of benchtop sequencers and metagenomic software, the middle step still requires computational resources and bioinformatic skill. To facilitate the middle step, we developed a cloud-computing pipeline that is easy and fast.

The prototype of the pipeline has been used in our metagenomic search for pathogens in various clinical cases. We have reported metagenomic analyses of clinical specimens that successfully identified *Francisella*

tularensis in an abscess as a pathogen (3), *Streptococcus* spp. in a lymph node as a possible causative candidate of Kawasaki disease (4) and heterogeneity of the 2009 pandemic influenza A virus (A/H1N1/2009) in the lung (5). In an outbreak of 22 adults with myalgia, the majority were infected with human parechovirus type 3, which typically causes disease in young children (6). In recent food poisoning outbreaks that were due to raw fish consumption, a flounder parasite *Kudoa septempunctata* was discovered as the causative agent (7). In all cases, the pathogen identification was primarily due to metagenomic analyses using next-generation sequencers.

In the workflow of metagenomic pathogen identification using MePIC, the first step is performed by the user. DNA and/or RNA is extracted from a specimen, such as sputum, feces, an abscess or blood, and a library of DNA/cDNA is prepared for sequencing. The library is sequenced using a benchtop next-generation sequencer, and the sequenced reads are uploaded to the MePIC pipeline via a secure internet connection. The pipeline accepts input files in FASTQ format, which is the standard for next-generation sequencing analysis. When using Illumina MiSeq sequencers (San Diego, Calif., USA).

In the second step, the uploaded reads are processed by the MePIC pipeline (Fig. 1). Unnecessary adapter sequences and low quality bases are trimmed off the reads using the fastq-mcf program in the ea-utils package (<http://code.google.com/p/ea-utils/>). Human-derived reads are detected through comparisons with the human genome using the BWA (8) program; the

*Corresponding author: Mailing address: Pathogen Genomics Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan. Tel: +81-3-5285-1111, Fax: +81-3-5285-1166, E-mail: makokuro@niid.go.jp

Register a new analysis

1. target read files

single end | paired end

File1: File2:

2. reads trimming

2.1. trimming of adaptor portions

exec trimming

adaptor sequence: TruSeq

percent occurrence: 0.0 error percent: 10 min reads length: 0

min clip length: min QV: 15 window size: 1

method: remove_adaptor_portions

2.2. trimming of LQ regions

exec trimming

min QV: 12 window size: 10 min reads length: 0

3. screening

exec screening

matching algorithm: bwasw

DB: Homo sapiens Build: hg19

E-value(blast only): 1e-08

screening rule: remove reads pairs matching to DB

4. reads classification

program: megablast

reference DB: NCBI NT

split number: 4

blast options

E-value: 0.01

max hit number per read: 1

max alignment per hit: 1

filter off:

output format: TEXT

other options:

*set by command line style of blast (bwa)

Register Cancel

Fig. 1. Screenshot of the MePIC pipeline. In box 1, the user specifies the next-generation sequencer reads for upload. In box 2, details are set for trimming adaptor sequences and low quality bases from the reads. In box 3, criteria are set for the exclusion of human reads. In box 4, the user chooses the program for searching the database of known sequences.

number of human reads are counted, but the reads themselves are removed from the downstream analysis. For each of the remaining reads, similar sequences are searched in the database of all known nucleotide sequences (NCBI nt) using the MEGABLAST (9) or BWA program (10). Based on the information of the database sequences that match with the read, we can infer the gene (e.g., virulence gene) and organism (e.g., *Escherichia coli*) that the read is derived from.

The run time of the pipeline is primarily allotted to searching the database of known sequences. The required time can be drastically shortened by splitting the job and running in parallel using a cloud-computing system or local server. Respectively, it takes 10 h for one

core of 2.67 GHz and 6 min for 100 cores to perform MEGABLAST search against the nt nucleotide database (as of year 2013) for one million reads of length 200 bp. The run time varies according to the sample source and condition. In blood samples, >90% of reads are derived from human and removed in the preprocessing step, and accordingly the time for database search of the remaining reads is reduced. Human derived reads are less in sputum samples (60%) or normal feces (~0%), which accordingly demands more database search time.

In the final step of the workflow, the user downloads the database search result to a local PC, including the reads annotated with the organism and gene function. To summarize the taxonomic and functional informa-



Fig. 2. (Color online) Analytic results of a case where Saffold virus was detected in a pharyngeal specimen. DNA and RNA were extracted from the specimen and sequenced. (A) Taxonomic view of MEGAN software. Saffold virus was detected in the RNA-seq reads. In this case, 0.4% of the reads were derived from the virus, which was sufficient to detect the virus in this patient. (B) Reads mapped to the reference genome of Saffold virus. The horizontal axis represents the position on the viral genome, and the vertical axis represents the abundance of mapped reads. The reads cover the whole 8 kb genome. Colored bars indicate the nucleotides where the patient's strain and the reference strain differ. The plot was made using Genome Jack software (<http://www.mss.co.jp/businessfield/bioinformatics/solution/products/genomejack/>).

tion over all reads, a metagenome browser can be used, such as MEGAN (11), which is one useful free software program (Fig. 2). The existence and quantity of pathogenic organisms and virulence genes are inferable from the number of detected reads, which is proportional to the number of the corresponding nucleotide sequences in the original specimen.

We developed a simple metagenomic analysis pipeline for removing ambiguous and host-derived short reads and rapidly identifying disease-causing pathogens in hospitalized patients and during outbreaks. The MePIC pipeline has a webpage interface that can be used easily by clinicians and epidemiologists, who do not have bioinformatic skill. The locally required equipment includes a benchtop next-generation DNA sequencer and a desktop PC for viewing the results. The adoption of cloud computing for metagenomic pathogen identification was proposed in the PathSeq software (12), which required bioinformatic skill for cloud computing. The MePIC pipeline, in contrast, manages the computational aspects in the background. The sequence similarity search of the database is the most computationally demanding step of metagenomic studies, and one solu-

tion is to thin the database. Using such an approach, the MetaPhlan system (13) can speedily identify bacterial and archaeal organisms. However, we opted to maintain the entire database and rely on augmenting the computational power to hasten the analysis because clinical applications require finer taxonomic distinction: for example, the distinction of enterohemorrhagic and commensal *E. coli* is critical. Within the broad possible applications of metagenomics, our pipeline is tailored for clinical use.

Metagenomic pathogen identification using next-generation sequencers surpasses conventional detection systems in sensitivity. The approach of directly sequencing nucleotides of a specimen is particularly powerful for unculturable or slow-growth pathogens (e.g., *Mycobacterium*). Whereas conventional PCR-based detection can miss new variants of a known pathogen due to mismatches of pre-designed primer sets, the de novo DNA/RNA sequencing approach overcomes this limitation. Metagenomic analysis can also identify a causal agent that was not known to be pathogenic (7). As for quantitative sensitivity, the metagenomic approach has been shown to be comparable to RT-PCR in

virus detection (2).

The major drawback of metagenomic pathogen identification is the cost of next-generation sequencers and reagents. Although the sequencers remain expensive, their versatile clinical and research utility (not restricted to infectious diseases) is pushing their widespread implementation in research institutes and hospitals. The rapidly decreasing reagent cost has reached approximately \$100 for one million reads, which would be appropriate for pathogen identification.

The current methodology of metagenomic pathogen identification is based on sequence matches with known pathogenic species/strains. To enable detection of *unknown* pathogens, an abundant dataset of “disease cases” and “normal flora controls” is necessary. If the number of reads of an organism (which is proportional to the amount of its DNA in the specimen) is much larger in cases than controls, infection by the organism can be suspected. The development of such pathogen discovery will require the accumulation of metagenomic data for disease-causing and normal flora and the invention of analytic tools. We believe that the use of the MePIC pipeline will promote metagenomic pathogen identification and improve the understanding of infectious diseases.

The source code for installing on a local server is available from the authors upon request. The website of the pipeline is <https://mepic.nih.go.jp/>. The sequence reads of the pharyngeal specimen that included the Saffold virus are available from the DDBJ Sequence Read Archive under accession number DRA000973.

Acknowledgments This work was supported by a grant for Research on Emerging and Re-emerging Infectious Diseases (H23 Shinko-Ippan-007, H25 Shinko-Ippan-015) from the Ministry of Health, Labour and Welfare, Japan, and was also supported by a grant-in-aid for Scientific Research (C) from the Japan Society for the Promotion of Science (Grant No. 23590527). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The development of MePIC pipeline

was partially supported by Illumina Inc.

Conflict of interest None to declare.

REFERENCES

1. Tang, P. and Chiu, C. (2010): Metagenomics for the discovery of novel human viruses. *Future Microbiol.*, 5, 177–189.
2. Chan, J.Z-M., Pallen, M.J., Oppenheim, B., et al. (2012): Genome sequencing in clinical microbiology. *Nat. Biotechnol.*, 30, 1068–1071.
3. Kuroda, M., Sekizuka, T., Shinya, F., et al. (2012): Detection of a possible bioterrorism agent, *Francisella* sp., in a clinical specimen using next-generation direct DNA sequencing. *J. Clin. Microbiol.*, 50, 1810–1812.
4. Katano, H., Sato, S., Sekizuka, T., et al. (2012): Pathogenic characterization of a cervical lymph node derived from a patient with Kawasaki disease. *Int. J. Clin. Exp. Pathol.*, 5, 814–823.
5. Kuroda, M., Katano, H., Nakajima, N., et al. (2010): Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by *de novo* sequencing using a next-generation DNA sequencer. *PLoS ONE*, 5, e10256.
6. Mizuta, K., Kuroda, M., Kurimura, M., et al. (2012): Epidemic myalgia in adults associated with human parechovirus type 3 infection, Yamagata, Japan, 2008. *Emerg. Infect. Dis.*, 18, 1787–1793.
7. Kawai, T., Sekizuka, T., Yahata, Y., et al. (2012): Identification of *Kudoa septempunctata* as the causative agent of novel food poisoning outbreaks in Japan by consumption of *Paralichthys olivaceus* in raw fish. *Clin. Infect. Dis.*, 54, 1046–1052.
8. Li, H. and Durbin, R. (2009): Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.
9. Morgulis, A., Coulouris, G., Raytselis, Y., et al. (2008): Database indexing for production MegaBLAST searches. *Bioinformatics*, 24, 1757–1764.
10. Altschul, S.F., Gish, W., Miller, W., et al. (1990): Basic local alignment search tool. *J. Mol. Biol.*, 215, 403–410.
11. Huson, D.H., Mitra, S., Ruscheweyh, H-J., et al. (2011): Integrative analysis of environmental sequences using MEGAN4. *Genome Res.*, 21, 1552–1560.
12. Kostic, A.D., Ojesina, A.I., Pedamallu, C.S., et al. (2011): PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat. Biotechnol.*, 29, 393–396.
13. Segata, N., Waldron, L., Ballarini, A., et al. (2012): Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods*, 9, 811–814.

