

FFPE がん組織における 病原性微生物シグネチャーを 検出するための PathoChip アレイの開発と活用

著者

Erle S. Robertson
Department of Microbiology
University of Pennsylvania
Philadelphia, PA

Josh Zhiyong Wang
Agilent Technologies
Santa Clara, CA

概要

有益な微生物と有害な微生物のバランスが、人間の健康に重要な影響を与えていることがますます認識されつつあります。このバランスは代謝や免疫反応に影響を与える可能性がある一方、感染性の病原体（HPV や *H. pylori* など）はがん発症の最も高いリスク要因のひとつでもあります。これらの知見から、ヒト腫瘍組織中の数千のウイルス、病原性細菌 / 真菌、寄生虫をスクリーニングすることで、がんの進行や治療への応答性などに対して微生物バランスがどのような役割を持つのかを解明する手がかりが得られるものと期待されます。本アプリケーションノートでは、ホルマリン固定パラフィン包埋（FFPE）腫瘍組織を用いて、数千種類のウイルスや病原性微生物の DNA および RNA を検出することが可能な PathoChip アレイの開発について紹介します。このアッセイには、病原体の検出のための上流のサンプル調製手順と下流のデータ解析手順が含まれています。さらに、トリプルネガティブ乳がんや口腔 / 中咽頭扁平上皮がんを含む多様ながん組織において、本アッセイを用いることで明確な微生物シグネチャーを検出できることを示しました。これらの結果から PathoChip は、解析が難しいことも多い FFPE サンプルであっても網羅的に病原体の情報を得るために活用可能であることが示されました。

はじめに

正常なヒトのマイクロバイームは、何千もの微生物種から構成されています。消化管、皮膚、気道、免疫系など、マイクロバイームが常在することが知られている組織に焦点を当てた研究が盛んに行われています。また、ウイルス、細菌、寄生虫などによる感染症は、肝臓、胃、子宮頸部、血液など様々な組織のがんの主要な原因となる可能性があります。

微生物を検出する方法としては、従来の培養法をはじめとして多くの方法があります。しかし、特定の疾患に関連する感染性微生物を効率的に同定するためには、微生物の遺伝子情報に基づくメタゲノム解析ツールがますます必要とされています。PCR 増幅後に 16s rRNA の塩基配列を決定する方法は細菌種を特定するための一般的な方法ですが、ウイルスや真核微生物には対応していません。また、サンプルからの全 DNA のショットガンシーケンスは偏りのない検出が可能ですが、宿主であるヒト DNA 由来のリードが大量にシーケンスされるため、効率が著しく低下するという問題を抱えています。

一方で、DNA マイクロアレイは、大量のサンプルから幅広い種類の微生物を迅速かつ経済的にスクリーニングできるので、この分野でも使われ始めています。この方法は市販のソリューションとしても存在しますが、欠点としては、それぞれ異なる（場合によっては重複する）微生物のサブセットを対象としているため、いずれも全体像を把握することはできない点が挙げられます。

本アプリケーションノートでは、Agilent SurePrint マイクロアレイ技術に基づく PathoChip アレイ (図 1) の開発について紹介します。PathoChip アレイには、公的に利用可能なすべての既知のウイルス配列と数百の病原性細菌、真菌、原生生物、寄生虫に対するプローブが含まれており、経済的なフォーマットで微生物病原体を幅広くカバーしています。それぞれのゲノムで可能な限り複数の領域をターゲットとするプローブを使用することで、検出力を向上させています。さらに、PathoChip のプローブは既知のターゲットに対する配列で開発されていますが、ウイルスのファミリー内およびファミリー間で保存されている配列に対するプローブをも組み込んでいるので、PathoChip で新しい系統や新しい微生物を発見できる可能性もあります。本アプリケーションノートでは、FFPE 腫瘍サンプルで DNA・RNA ターゲットの同時検出も含め、微生物をプロファイリングするためのワークフローについても説明しています。

また、PathoChip アレイを用いることでトリプルネガティブ乳がんおよび口腔・中咽頭扁平上皮がんの数百の FFPE サンプルを解析し、ウイルス・微生物シグネチャーの検出を行うことができました。

材料と方法

マイクロアレイのデザイン

ゲノム、遺伝子、核酸配列のデータベースである National Center for Biotechnology Information (NCBI) に対し、すべての分類のウイルスアノテーション、文献検索 (<http://www.ncbi.nlm.nih.gov/pubmed>) および Web リソース (<http://www.niaid.nih.gov/Emerging and Re-emerging Infectious Diseases, Category A, B, and C Priority Pathogens>) によってまとめられた原核生物および真核生物のヒト病原体リストのアクセッション番号をクエリーとして使用し、配列情報を得ました。この結果得られた配列は、それぞれ長さ約 500 万から 1,000 万ヌクレオチド (nts) の 58 の“仮想的な染色体”に分割されたメタゲノムとしてアセンブルされました。文献 6 に記載しているように、このメタゲノム中のユニークな領域と保存された領域の両方を同定し、設計対象領域に含めました。

メタゲノム中のユニークな領域と保存領域に対するプローブ配列 (最大 60 nts) はアジレントのアレイ比較ゲノムハイブリダイゼーション (aCGH) のアルゴリズムで設計されました。ハイブリダイゼーションのバックグラウンドノイズを低減するために、これらの配列からヒトゲノム配列とのクロスハイブリダイゼーションの可能性が低いものが選択されました。

病原体ゲノムのユニークな領域や保存領域に対して、あるいは原核生物または真核生物の病原体配列に対してマッピングされるプローブは、もとの配列に対して使用可能なプローブが 10 個未満の場合にはそのままマイクロアレイデザインに追加されました。もとの配列に対して 10 個以上のプローブが作成された場合、以下の基準でプローブをフィルタリングしました：① 最大 20 プローブ/ゲノム、② プローブ間の最小間隔 100 bp、③ ゲノム全体をカバーし、均等に分布すること。発がん性ウイルスについては、プローブ数の制限を設けず、ウイルスゲノム全体をカバーするよう飽和したタイリングのプローブセットを作成しました⁶。

Agilent SurePrint マイクロアレイ 8x60K フォーマットは、1 スライドあたり 8 サンプルをアッセイできます。まず、プローブの性能を評価するために、PathoChip v2a と v2b の 2 種類のアレイデザインを作成し、それぞれユニークな領域と保存領域をターゲットとする 60,000 プローブを搭載しました。PathoChip v2a および v2b からパフォーマンスの高いプローブを組み合わせて PathoChip v3 を作成し、ユニークな領域に対する 37,704 プローブと保存された領域に対する 23,627 プローブを搭載しました。これにより、すべての発がん性・病原性ウイルスを網羅することができました。本アプリケーションノートに記載されている様々ながんの FFPE 組織の解析には、この PathoChip v3 を使用しました。

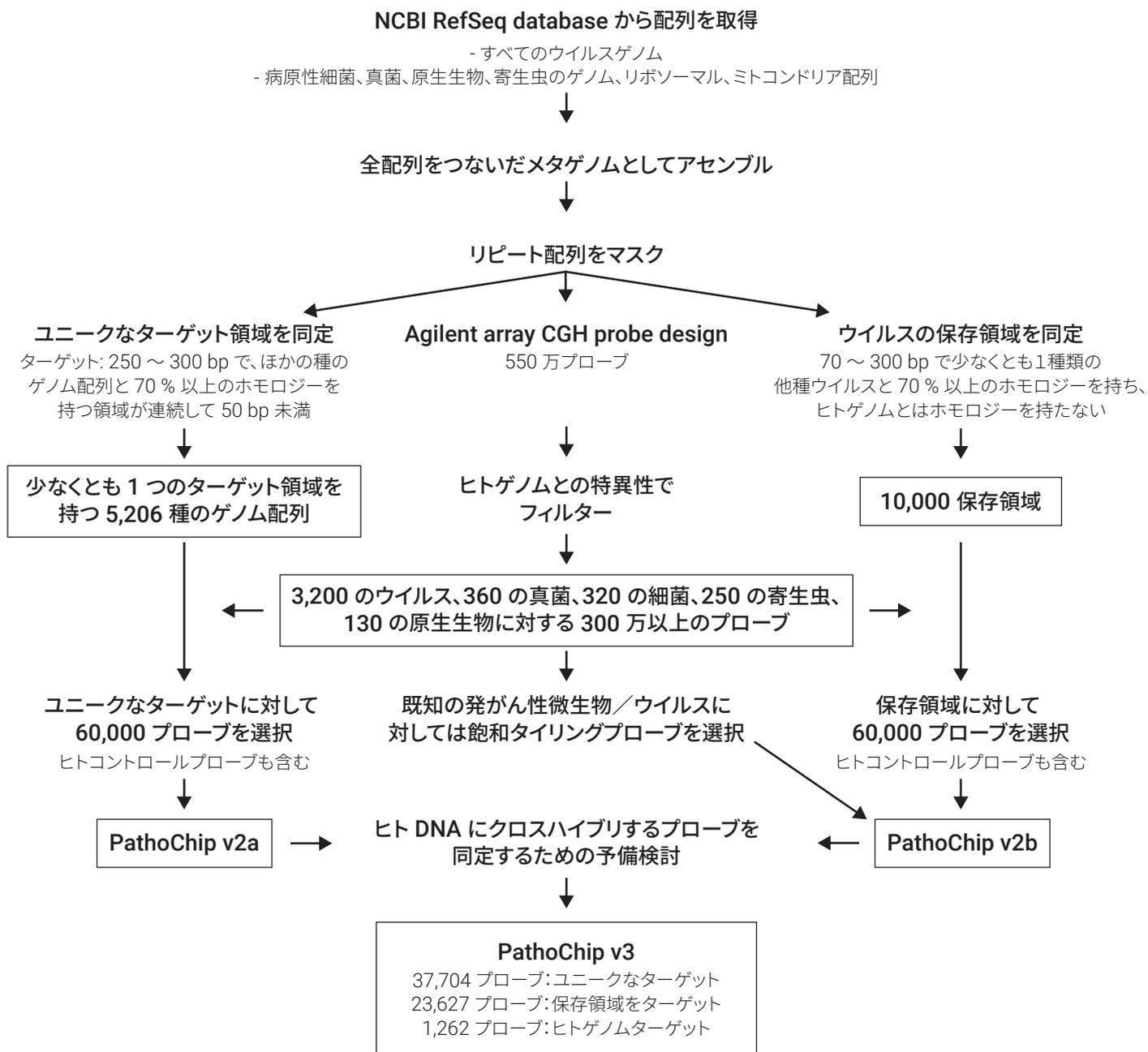


図 1. PathoChip のデザイン戦略とデザインの改良

サンプル

すべてのサンプルは Abramson Cancer Center の Tumor Tissue and Biospecimen Bank から入手しました。病理医が症例経過を確認し、腫瘍の種類およびがん細胞の境界を確認しました。隣接する正常組織がある場合、連続した切片をヘマトキシリン・エオジン (HE) 染色してがん領域が明確にわかる非帯電スライドをテンプレートとして、腫瘍部分の切り出しを行いました。

口腔・中咽頭扁平上皮がんでは、匿名化された 100 検体の FFPE サンプル (OCSCC と総称) を非帯電スライドガラス上の 10 μ m の切片として提供されました。マッチドコントロールと非マッチドコントロール各 20 のサンプルも合わせて提供されました⁷。ここで「マッチドコントロール」とは、本研究に含まれる 20 人のがん患者から得られ、がんに隣接する臨床的に正常な部位としています。「非マッチドコントロール」は、健常者から採取した口腔内組織 (口蓋垂) です。

また、トリプルネガティブ乳がんでは、匿名化された 100 検体の FFPE サンプルを非帯電スライドガラス上の 10 μ m 切片の形で受け取りました。17 のマッチドコントロールと 20 の非マッチドコントロールは、パラフィンロールとして提供されました⁸。マッチドコントロールは、がん組織と同じ患者で、隣接する非がん乳房組織から得られたものです。非マッチドコントロールは、健常人から得た乳房組織です。

DNA/RNA の抽出、増幅、標識、ハイブリダイゼーション

標準的なアジレントの CGH アッセイでは、出発材料としてゲノム DNA を使用しますが、ウイルスゲノムは DNA または RNA のいずれかであるため、プロトコルの変更を試みました。具体的には、FFPE サンプルから AllPrep DNA/RNA FFPE kit (QIAGEN) を用いて DNA と RNA を続けて抽出しました。核酸の品質確認として、A260/280 比、収量、アガロースゲル電気泳動によるサイズ分布などを評価しました。一部の FFPE RNA サンプルでは部分的に分解が見られましたが、ほとんどのサンプルで RNA の断片サイズと抽出収量は十分であり、total RNA を直接 cDNA 合成に使用することが可能でした。

テストサンプルとしてゲノム DNA 50 ng と total RNA 50 ng をあわせて TransPlex WTA-2 kit (Sigma-Aldrich) のインプットとして使用しました。このキットでは、ゲノム DNA と cDNA を同時に増幅することが可能です。増幅産物は QIAquick PCR purification kit (QIAGEN) で精製しました。精製した増幅産物 2 μ g を SureTag labeling kit (Agilent) を用いて、テストサンプルとして Cy3 で標識しました。リファレンスサンプルとして、ウイルス陰性の参照ゲノム DNA (BJAB 細胞株) 50 ng を TransPlex WTA-2 で増幅し、精製した増幅産物 2 μ g を Agilent SureTag labeling kit で Cy5 で標識しました。このリファレンスサンプルは、プローブとヒト DNA とのクロスハイブリダイゼーションを検出するためのコントロールとして使用されています。

Cy3 および Cy5 で標識した DNA を SureTag labeling kit に含まれるスピニングカラムで精製しました。標識 DNA の色素取り込み率を測定し、ハイブリダイゼーションのために混合しました。アジレントの推奨条件に従い、ハイブリダイゼーションオープンを用いてハイブリダイゼーション温度 65 °C、回転数 20 rpm、40 時間で標識した DNA を PathoChip v3 8x60K アレイにハイブリダイゼーションを行いました。アレイは標準的な手順で洗浄し、Agilent SureScan マイクロアレイスキャナ C または D (型番 G2565AA & G4900DA) でスキャンしました。

マイクロアレイデータ解析

スキャンしたマイクロアレイ画像は、Agilent Feature Extraction ソフトウェアを使用して解析し、各フィーチャーについてピクセルの平均シグナル強度を計算し、バックグラウンドを差し引きしました。

QC プロセスの一環として、高いバックグラウンドや傷、または品質管理 (QC) の閾値を超えるその他の技術的アーチファクトの影響を受けたアレイに注意を払うために、スキャン画像をマニュアルで確認しました。

Cy3 および Cy5 のシグナル強度を Partek Genomics Suite (Partek Inc) にインポートし、解析を行いました。ヒト遺伝子コントロールプローブの平均強度を、ハイブリダイゼーションしたテスト DNA サンプル (Cy3 標識) とヒトリファレンス DNA サンプル (Cy5 標識) について算出しました。これに基づき、Cy5 ヒトリファレンス DNA シグナル平均を Cy3 シグナル平均に補正するための係数が決定されました。次に、すべての PathoChip プローブの Cy5 強度に係数を乗じ、色素間の補正をしました。各プローブについて Cy3/Cy5 比および Cy3 - Cy5 差分を計算し、続くデュアルチャンネル解析パイプライン、シングルチャンネル解析パイプラインのインプットとしました。

Accession average (AccAvg) は、1 つのゲノム配列に対する全プローブにおける Cy3 または Cy5 強度の平均値として、またゲノム配列のシグナル (AccSig) は AccAvg (Cy3) - AccAvg (Cy5) とそれぞれ定義しました。

Partek に実装されている Model-based analysis of tiling arrays (MAT) を用いて、各腫瘍サンプルのプローブシグナル (Cy3 - Cy5) のスライドウィンドウ解析を行いました。MAT のパラメータは、p 値のカットオフを 0.99、ウィンドウサイズを 5,000 bp、最少陽性プローブ数を 5、廃棄値 (discard value) を 0% としました。候補領域は、MAT スコア 30 ~ 300、300 ~ 3,000、>3,000 に分類しました。Partek の分散分析 (ANOVA) ツールを用いて、すべての腫瘍サンプルをテスト条件のレプリケートとして、また一緒にハイブリダイズしたヒトリファレンス DNA をコントロール条件のレプリケートとして、多重検定補正によるペア t 検定を実施しました。

AccAvg (Cy3) vs AccAvg (Cy5) を用いて各ゲノム単位で、また Cy3 vs Cy5 強度を用いて個々のプローブ単位で比較しました。有意性の閾値は、False discovery rate<0.05、fold difference >2 で設定しました。

全腫瘍サンプルの AccSig またはプローブシグナルの標準偏差を計算し、集団の平均より 2 以上標準偏差が大きい値についてフィルタリングを行うことで、各ゲノムおよびプローブレベルでのはずれ値解析も行いました⁶。

また、一部のデータ解析では、正規化とデータ解析のために R-program も使用しました⁷。係数は、ヒトプローブに対する緑と赤のチャンネルのシグナルを用いて計算されており、ヒトプローブに対する緑のシグナルの和 / 赤のシグナルの和の比率と定義しました。これらの係数は、そのほかのすべてのプローブの補正シグナルを得るために使用しました。

ヒトプローブを除くすべてのプローブについて、正規化シグナルは、緑色シグナルの log2 変換値 / 係数で補正した赤色シグナル (log2 g - log2 scale factor * r) となります。がんサンプルとコントロール群 (マッチドおよび非マッチドコントロール) を比較し、がんサンプルで陽性シグナルを示すプローブを選択するため、あるいはマッチドまたは非マッチドコントロールでがんサンプルと比較して、有意差のあるプローブを選択するために、補正したシグナルに対して t 検定を行いました。有意差のカットオフは、log2 fold change > 0.5、補正済み p 値 < 0.05 としました。これらの補正済み p 値は、Benjamini-Hochberg 法を用いた多重検定のために得られたものであり、この補正済み p 値のカットオフの下では、コントロールサンプルにおいてどのプローブも有意であると検出されました。

コントロールサンプル群で、多重比較補正なしの名目上の p 値が 0.05 未満のトップのものを、がんサンプル群で有意なプローブと比較するために提示しました。

有病率は、がんサンプルとコントロールサンプルにおけるシグネチャーの検出率に基づいて計算しました。

結果と考察

マイクロアレイデザインの最適化

アセンブルされたメタゲノムには、4,200 以上のウイルス、バクテリア、真菌、寄生虫の計 5,206 のゲノム配列が含まれ、58 本の仮想的な染色体 (合計 4 億 4,890 万 bp) にまとめられました。このメタゲノムに対して約 550 万のプローブが、CGH アプリケーション用のアジレントのカスタムプローブデザインアルゴリズムによって設計されました。これらのプローブのうち 300 万個以上は、ヒトゲノム配列とのクロスハイブリダイゼーションのリスクが低いと予測されました。

選択された病原体のユニークな領域に対応するプローブのサブセットは PathoChip v2a マイクロアレイに搭載され、少なくとも 2 種類以上のウイルス間の保存領域に対応する別のサブセットは PathoChip v2b アレイに搭載されました⁶。

PathoChip v2b で強化された特徴は、ヒトのがんと密接に関連することが知られている病原体に対応する 22 のゲノム配列に対しては配列全長にタイリングした 2,085 個のプローブが搭載されていることです。

アジレントのヒトリファレンス DNA を用いたパイロットアッセイでは、プローブのシグナル強度の中央値が、PathoChip v2a ではヒト配列に対するプローブで 750 蛍光単位 (RFU) を超え、非ヒト・ユニークプローブで約 17 RFU、そして PathoChip v2b の非ヒト・保存プローブでは 120 RFU でした¹。全体としてこれらのアッセイにより、150 RFU 以上の蛍光値で、ヒト DNA とクロスハイブリダイズする可能性が高い 6,360 プローブが同定されました。これらのプローブはバックグラウンドが高くなる原因になり得るため、ユニークな領域と保存領域に対するプローブサブセットを組み合わせた PathoChip v3 デザインを作成する際に、検討対象から除外されました。

興味深いことに、Agilent Female リファレンス DNA では Epstein-Barr ウイルス (EBV; ヒトヘルペスウイルス 4) に対するプローブで高いシグナル強度が認められました。この Male および Female の SureTag リファレンス DNA を調製するために使用している細胞株は、EBV ウイルスを感染させて樹立したものであることが製造元により確認されました。Female リファレンスの Cy5 チャンネルで EBV プローブシグナルがノーマライズされていない場合、EBV 偽陽性となる可能性があります。そのため、使用するリファレンス DNA をウイルスフリーでヒト B 細胞由来の DNA に切り替えました。このリファレンスは、数々の厳しい検出工程を経て、ヒトウイルスに対するシグナルが検出されないことが確認されました。

独自のラベリングワークフローの開発

ゲノム DNA (gDNA) のみを使用する標準的な CGH ワークフローとは異なり、PathoChip アッセイでは gDNA と total RNA の両方を使用します。多くの市販の DNA および RNA 精製キットは、FFPE サンプルを処理することができますが、QIAGEN AllPrep DNA/RNA FFPE キットでは、同じ FFPE サンプルから gDNA と total RNA を効率的に抽出できることがわかりました。このキットにより、微生物の中では核酸抽出が困難な真菌やグラム陽性菌、グラム陰性菌からも核酸を抽出することに成功しました。本キットは、*Saccharomyces cerevisiae*、*Bacillus cereus*、*Escherichia coli* の培養からも gDNA と RNA を効率よく抽出することができました（データ未掲載）。

考慮した点として、微生物によってはゲノムのコピー数が低い可能性もあるため、検出可能なコピー数にするために DNA と RNA を増幅する必要のある点が挙げられます。PathoChip スクリーニング技術では、TransPlex WTA-2 amplification kit による増幅ステップを利用しており、ゲノムコピー数の少ない微生物やウイルスの検出が可能です。このような PathoChip アレイにより、複数の腫瘍サンプルから様々な微生物の存在を迅速かつ高感度にスクリーニングすることが可能です。

初期のテストでは、ヒトゲノム DNA 15 ng のバックグラウンドに対して、ヒトアデノウイルス 5 型、JC ポリオーマウイルス、BK ポリオーマウイルス DNA を、それぞれ絶対コピー数 1 万から 10 の範囲で添加しました。TransPlex で増幅後、すべてのコピー数においてヒトアデノウイルス 5 型が PathoChip プロンプで検出され、ポリオーマウイルスのプロンプは 100 ゲノムコピー以上でバックグラウンドを上回るシグナルが検出されました（図 2A）。次に、アデノウイルス 5 型を含む細胞株から抽出した DNA と RS ウイルス (Respiratory syncytial virus) を含む RNA を混合し、TransPlex WTA-2 で同時に増幅しました。両ウイルスに対するプロンプは、強力かつ特異的なシグナルを示しました。このことから、TransPlex WTA-2 は、ゲノム DNA の存在下でも堅牢に逆転写を行うことが可能で、gDNA と cDNA のターゲットが共に増幅されることが示されました⁶。

以上により、PathoChip を用いた手法（改変 CGH ワークフローに基づく [図 2B]）で、腫瘍組織において真核生物および原核生物の病原体から核酸を抽出し、検出できることが実証されました。

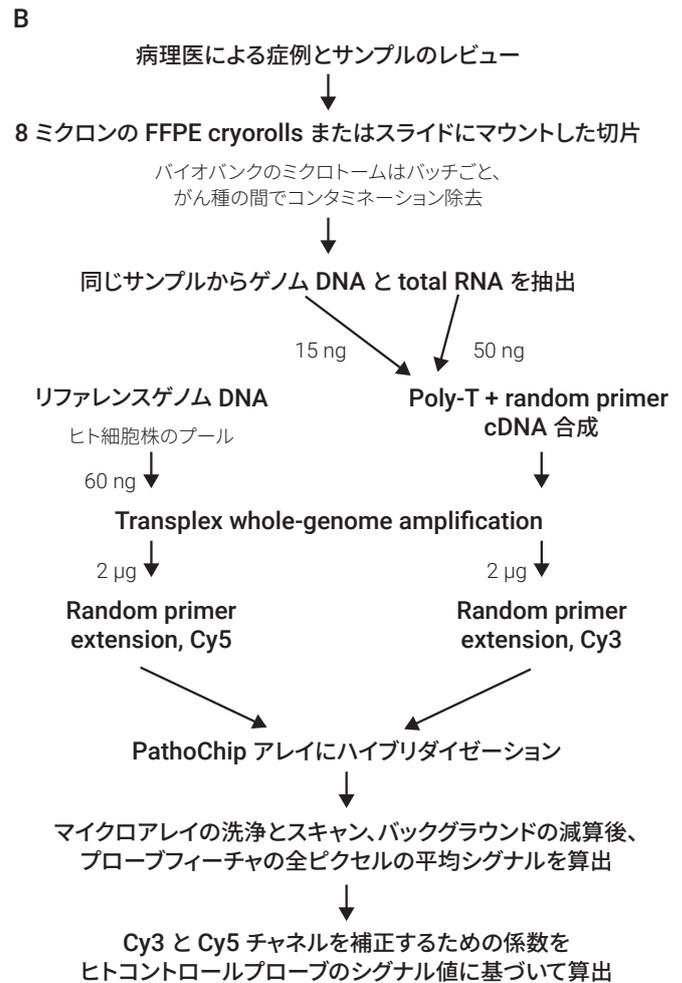
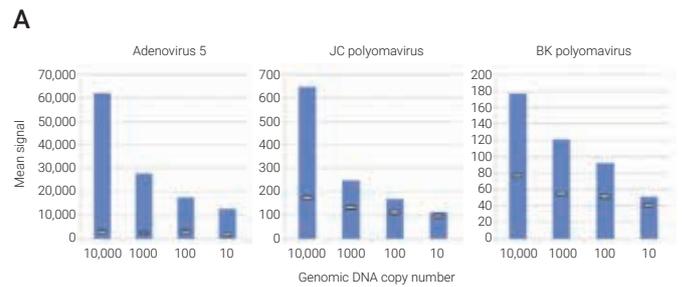


図 2. PathoChip を用いたラベリングワークフローの開発。
 (A) ウイルスゲノムの検出感度。3 種のウイルスについて、サンプルあたり 10,000 コピーから 10 コピーのウイルスゲノムをヒトゲノム DNA にスパイクインした際の検出感度を検討した。青いバーはテストサンプルにハイブリダイズしたウイルスを示す全プロンプの Cy3 シグナルの平均値を、白い線はコントロールサンプル（ヒト DNA のみ）の平均 Cy5 シグナル値をそれぞれ示す。
 (B) コントロール病原体 DNA を検出する PathoChip のワークフロー。

OSCC サンプルを用いた PathoChip アレイのデータ解析戦略の開発

発がん性ウイルスは、宿主の腫瘍の中で大きなゲノム再構成や欠失を起こすことがあります。さらには、ウイルスゲノムに多型があるため、新しい病原体の検出は、単一のプローブのシグナルに依存することも考えられます。したがって、スクリーニングプロジェクトで予想される3つの主要な「ヒット」を検出するためには、いくつかのレベルでのデータ解析が必要となります(図3)。Accession signal (AccSig) は、単一のゲノム配列 (Accession) の全プローブの平均値をヒト DNA のクロスハイブリダイゼーションの値で補正したものです。AccSig は、1つのゲノム配列に対する大多数のプローブのシグナル値に基づいて、一般的な病原体を検出するスクリーニングを行うために計算された値です。

プローブのスライディングウィンドウから MAT (model-based analysis of tiling arrays) スコアを計算し、プローブの生物種由来には関係なく局所的に高シグナルの領域を検出しました。個々のプローブレベルで多重検定補正を行った t 検定を採用し、腫瘍で一貫してバックグラウンドよりシグナルが高いプローブを同定しました。外れ値解析は、スクリーニングしたサンプルの中で1つ(または数個)の腫瘍サンプルのみで高いシグナルを持つプローブに対して行いました。

この解析方法を評価するために、OSCC 腫瘍 100 サンプルのスクリーニングプロジェクトのデータを使用しました。HPV16 の AccSig は p16 の病理報告⁶と一致し(図4)、p16 (+) 腫瘍のうち 80% で AccSig 値 100 以上を示しました。HPV16 AccSig が低い、もしくは 0 である 8 サンプルの p16 (+) 腫瘍のうち、4 サンプルは HPV16 プローブのサブセットで高いシグナルを示すか、HPV26 または HPV92 で有意に高い AccSig 値が得られました。スライディングウィンドウ解析でも、AccSig 結果が再現され、HPV16 ゲノムの完全長を検出する場合と、部分的に検出する場合の違いが明らかになりました⁶。

個々のプローブレベルでの解析も、候補を検出するのに有用であることが示されました。ほとんどの HPV16 プローブは、腫瘍サンプル全体において t 検定で有意にバックグラウンドより高いシグナルで検出されました(表1)。HPV16 は OSCC では非常に一般的に検出されるウイルスゲノムであることから、予想と一致する結果と言えます。また、多くの HPV16 プローブは外れ値テストもパスし、シグナルは一貫してバックグラウンドより高いことがわかりましたが、シグナル強度の幅はサンプル間で広いこともわかりました(そのため、外れ値も含まれます)。一方、HPV18 と HPV26 については t 検定で有意になるプローブは少なく、このがん種においてこれらのゲノムの出現率はるかに低いことを反映しています。

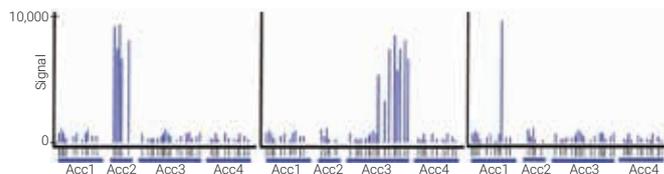


図3. PathoChip を用いた3つの解析戦略を示すモデルデータ。3つの仮説腫瘍サンプルにハイブリダイゼーションした後、4つのゲノム配列 (Acc) に対する個々のプローブ (x 軸) のシグナルをプロットしている。Acc2 の全プローブは腫瘍 1 (左) で高シグナルを示すので、この候補はテストサンプルのゲノム配列ごとの全プローブ平均とコントロールサンプルのそれを比較することで検出できるはずである。Acc3 プローブのサブセットは腫瘍 2 (中央) で高いシグナルを示すが、おそらく株の配列の違いかゲノムの部分的欠失により Acc3 の全プローブの平均シグナルは低くなり、検出が難しくなる。この場合、局所プローブシグナルのスライディングウィンドウ解析は、アクセッションアノテーションに偏らず、候補同定の感度が高い可能性がある。腫瘍 3 (右) では Acc1 の1つのプローブのみが高シグナルを示すため、PathoChip が特異的にターゲットとしている生物ではないものの、1つもしくは数個のプローブと配列相動性を有する生物を検出するために、個々のプローブの性能のみに基づく第3層の解析が必要となる。

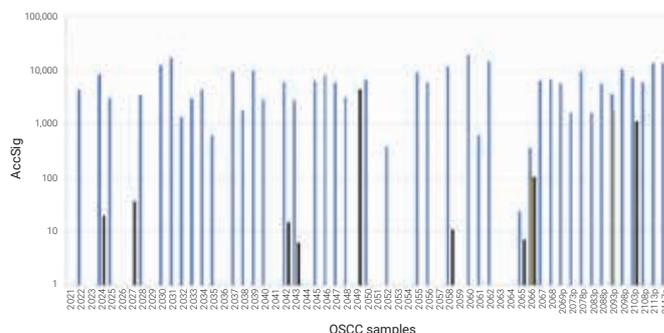


図4. 腫瘍中の HPV の Accession Average 解析。HPV16 (青)、HPV18 (オレンジ)、HPV26 (黒) の Accession Signal (AccSig) を、個別 (2021 ~ 2068) またはプール (2069p ~ 2117p) でアッセイした口腔扁平上皮がん (OSCC) 100 検体の PathoChip 結果から算出した。

表1. ヒトパピローマウイルス検出のための個々のプローブ解析

Probe	No. of probes		
	HPV16	HPV18	HPV26
Total probes	68	85	13
Specific probes	67	84	11
Pass t test	64	11	4
Pass outlier test	65	66	9
Conserved probes	1	1	2
Pass t test	1	1	0
Pass outlier test	1	0	2

しかしながら外れ値解析により、少数の陽性サンプルにおいて AccSig または MAT スコアにより HPV18 または HPV28 が検出された比較的多くのプローブが簡便に同定されました。これらの稀な候補について、いくつかのプローブは、シグナル値は低いものの、サンプル全体を通してバックグラウンドを超える一貫したシグナルを示したことから、t 検定で有意となりました。これは存在するゲノムコピー数によると考えられ、驚くことではありません。このことは、初期段階の同定方法にかかわらず、フォローアップ検証の候補を検討する際には、アルゴリズムの出力スコアを分析するだけでなく、プローブレベルのハイブリダイゼーションシグナルを調べる必要があることを示しています。

PathoChip アレイを用いた口腔咽頭がんおよび口腔扁平上皮がんに関連する微生物シグネチャーの検出

先のデータ解析⁶では、各サンプルについて MAT スコアが 3,000 以上のメタゲノム領域を集計しています。各領域の個々のプローブは、ゲノム上の位置に従ってプローブシグナルをプロットしました。この解析により、病原性口腔細菌を含む他の多くの生物が検出され、OSCC には HPV 以外にも様々な微生物が関与している可能性が示されました。

PathoChip を用いて、病理学的に定義された口腔扁平上皮がん (OCSCC) 患者由来の 100 検体の FFPE サンプルと、がんに関連する正常部位サンプル (マッチドコントロール) 20 検体、健康者の口腔組織 (uvula) サンプル (非マッチドコントロール) 20 検体のスクリーニングを行いました。このスクリーニングでは、腫瘍組織と関連するウイルスおよび微生物の特徴的なシグネチャーの同定を目的としています⁷。本研究で分析したサンプルは、舌、舌根、扁桃、口底、頬から採取されたがん検体で、最も数が多いのは中咽頭がんでした。

最適化したワークフローに従って、サンプルから DNA と RNA の両方を抽出し、全ゲノムおよびトランスクリプトーム増幅 (ここでは WGTA としています) を行い、蛍光標識し、PatoChip とのハイブリダイゼーションを行いました。

OCSCC に関連するウイルスのシグネチャー

本研究では、がん検体およびコントロール検体に関連する RNA および DNA ウイルスを同定しました (図 5)。スクリーニングされた OCSCC 検体では、パピローマウイルス科に属するウイルス配列が最も高いハイブリダイゼーションシグナルを示し、次いでヘルペスウイルス科、ポックスウイルス科、レトロウイルス科、ポリオーマウイルス科のウイルスが認められました (図 5A)。これらのウイルスファミリーに属するウイルスシグネチャーは、スクリーニングされた 100 の OCSCC サンプルにおいて 75% 以上のケースで検出されました。興味深いことに、パピローマウイルス科は 98% のケースで検出されました。マッチドコントロール、非マッチドコントロールと比較して OCSCC サンプルではパピローマウイルス科でより高いハイブリダイゼーションシグナルが検出されました⁹。

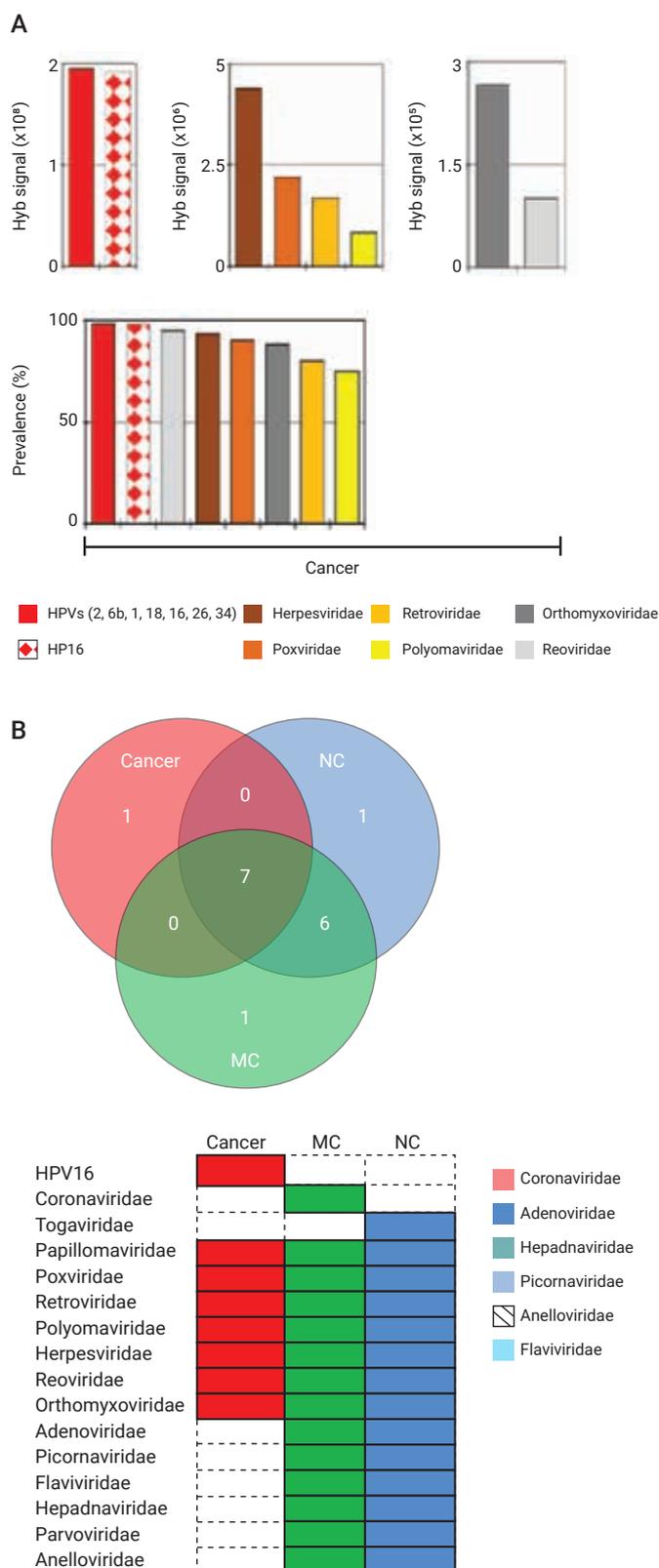


図 5. 口腔がんおよびコントロールサンプルで検出されたウイルスシグネチャー。(A) 口腔がん 100 検体の PathoChip スクリーニングでハイブリダイゼーションシグナル (g-r > 30) が検出されたウイルスシグネチャーを示し、ハイブリダイゼーションシグナルの減少 (ゲノム配列ごとの全プローブの加重スコア合計) と有病率に従ってランク付けした。(B) ウイルスファミリーの様々な分子シグネチャーとがんおよびコントロールとの関連を、ベン図および色付きバーで示す。

重要なことは、HPV16 が高いハイブリダイゼーションシグナルと有病率（98 %）の両方で検出されたのは OCSCC サンプルだけであったことです（図 5A、図 5B）。レオウイルス科、ヘルペスウイルス科、ポックスウイルス科、オルソミクソウイルス科、レトロウイルス科およびポリオーマウイルス科のシグネチャーは、OCSCC サンプルで高い有病率とコントロールより 100 ~ 1000 倍高いハイブリダイゼーションシグナルとして検出されました（図 5A）。注目すべきは、コロナウイルス科、ピコルナウイルス科、アデノウイルス科、アネロウイルス科、ヘパドナウイルス科、フラビウイルス科のウイルスシグネチャーが、非 HPV16 パピローマウイルス科のシグネチャーとともにコントロール群で有意かつ明確に検出された点です⁷。これらのデータは、OCSCC 組織と比較した場合、ウイルスのシグネチャーが有意に変化していることを示しています。

OCSCC に関連する細菌のシグネチャー

図 6A は、OCSCC、マッチドコントロールおよび非マッチドコントロールサンプルで見つかった様々な細菌シグネチャーを示しています。このベン図は、我々の発見を要約したもので、13 の属を代表する細菌シグネチャーが OCSCC サンプルと関連することが判明しましたが、マッチドコントロールや非マッチドコントロールとは関連しないことが示されました。これらには、プロテオバクテリアの 11 属、アクチノバクテリアとファーミキューテスの各 1 属が含まれていました⁷。*Proteobacteria Brevundimonas* と *Actinobacteria Mobiluncus* が最も多く（98 %）、プロテオバクテリアの大腸菌属とブレブンディモナス属のプロープがそれぞれ 88 % と 98 % のがん症例で検出され、高いハイブリダイゼーションシグナルが見られました。OCSCC サンプルから検出されたアクチノバクテリアプロープも高いハイブリダイゼーションシグナルを示し、最も高かったのはアルカナバクテリア属のプロープでした。ウイルスの場合と同様に、細菌の微生物シグネチャーは、通常のシグネチャーと比較すると、OCSCC において有意に変化しており、差はより顕著でした。マッチドコントロール、あるいは非マッチドコントロールサンプルでは、アクチノマイセス属、モビルンカス属、マイコバクテリウム属の細菌が検出されました。重要なこととしては、コントロールサンプルで検出された細菌シグネチャーのほとんどは、正常な口腔内細菌叢のものであることです。

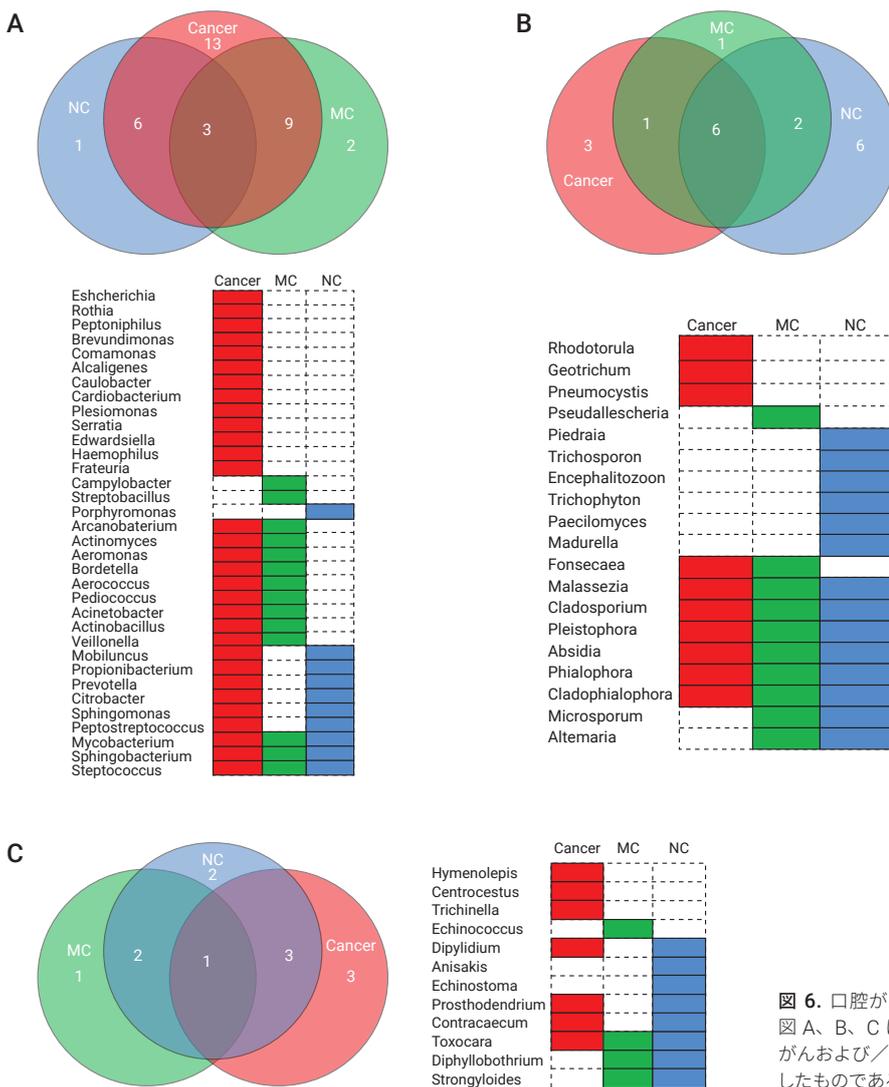


図 6. 口腔がんサンプルで検出された細菌、真菌、寄生虫のシグネチャー。図 A、B、C は、異なる細菌、真菌、寄生虫属の分子シグネチャーと、口腔がんおよび/またはコントロールとの関連を、ベン図およびカラーバーで表したものである。

OCSCC に関連する真菌および寄生虫のシグネチャー

ベン図に、OCSCC、マッチドおよび非マッチドコントロールで共有される真菌シグネチャーと特異的な真菌シグネチャーを示しています。注目点としては、ロドトルラ属、ゲオトリクム属およびニューモシスチス属の3つの真菌のシグネチャーがOCSCCに特異的に関連している点です(図6B)。フォンセカエア属、マラセチア属、プレストフォラ属、ロドトルラ属、クラドフィアロフォラ属、クラドスポリウム属の分子シグネチャーがスクリーニングしたすべてのOCSCCサンプルで検出されました。ニューモシスチス属は93%のがんサンプルで検出され、ゲオトリクム属、フィアロフォラ属、アブシディア属、プレボテラ属のシグネチャーはスクリーニングしたがん症例の75%以上で検出されました⁷。OCSCCの真菌バイオームには、コントロールの口腔内サンプルと比較して、有意な変化が観察されました。一方、図6Cのベン図は、がんおよびコントロールサンプルと寄生虫シグネチャーの関連性をまとめたものです。ヒメノレピス属、セントロセストス属、およびトリヒナ属のシグネチャーはOCSCCとのみ関連することがわかりました。エキノコックス属のシグネチャーはマッチドコントロールサンプルにのみ関連し、アニサキス属とエキノストーマ属のシグネチャーは非マッチドコントロールサンプルにのみ関連していました。このように、がん、マッチドコントロール、非マッチドコントロールは、それぞれ異なるシグネチャーで区別されることがわかりました。

トリプルネガティブ乳がん(TNBC)に関連する微生物シグネチャーの検出を目的としたPathoChipの使用

乳がんは最も一般的ながんのひとつであり、特定のホルモン受容体や増殖因子受容体の有無によって分類されます。乳がんの中で最も進行が速いのは、エストロゲン、プロゲステロン、HER2の3つの受容体がいずれも発現していないトリプルネガティブ乳がん(Triple negative breast cancer: TNBC)で、その理由としてはホルモン治療が適用できないことがあげられます。乳がんの進行には、遺伝的要因、環境要因、生活習慣が関与していますが、乳がんのいくつかの研究では、ヘルペスウイルス、ポリオーマウイルス、パピローマウイルス、レトロウイルスとの関連性が示されています⁹。

合計100検体のTNBCサンプルと、17のマッチドコントロール、20の非マッチドコントロールについて、PathoChipを使用してスクリーニングを行いました。すべてのサンプルはFFPE由来です(方法参照)。スクリーニングした100のTNBCサンプルのうち、40検体はサンプルごとにスクリーニングを行い、60検体は1反応あたり5検体(RNA/DNA各10ng)のプールでスクリーニングを行い、合計で52アレイを100検体のTNBCサンプルのスクリーニングに使用しました。コントロールサンプルについては、17のマッチドコントロールと20の非マッチドコントロールをそれぞれでプールし、各セットに4つのアレイを使用しました。

PathoChipのスクリーニングで、特定のプローブについてCy5(r)シグナルよりも高くCy3(g)シグナルが検出された場合、そのプローブは陽性とみなしました。特定の生物種のプローブについて、検出可能なハイブリダイゼーションシグナル($g-r > 30$)が検出され、マッチドまたは非マッチドコントロールと比較してがんサンプルで有意に高い値を示す場合に、がんサンプルと関連していると判断しました。また、複数の検出手法により、陽性プローブを慎重に評価しました。これらの手法(表2)には、Accession outlier、Accession t-test、Specific probe outliers、Specific probe t-test、Conserved probe outlier、Conserved probe t-test、Model based analysis for tiling Arraysが含まれています。

レポートの際には、PathoChip上の特定のプローブで検出された特定のウイルスや微生物の名前をリストアップします。ただし、特定のプローブによる検出は、その生物そのものを検出しているのではなく、近縁種を検出している可能性があることにも注意が必要です。これはTNBCサンプルが、特定のウイルスや微生物に対するプローブセット全体でハイブリダイゼーションシグナルに幅があったり、セットの中のいくつかのプローブでハイブリダイゼーションシグナルを示さないことがあったりした場合に特に考える必要があります。また、その特定の腫瘍でゲノム領域の一部が欠失していたり、その株に変異がおきていたりする可能性もあります。

保存領域に対するプローブのうち、ヘルペスウイルス科、レトロウイルス科、パラポックスウイルス科、ポリオーマウイルス科、パピローマウイルス科ファミリーに属するウイルスのシグネチャーが検出されました⁸。例えば、ヘルペスウイルス科では、ヒトサイトメガロウイルス(HCMV; Human Cytomegalovirus)、ヒトヘルペスウイルス (Human Herpesvirus, HHV1; Herpes simplex type 1)、カポジ肉腫関連ウイルス (KSHV; Kaposi sarcoma herpes virus)、エプスタイン・バーウイルス (EBV/HHV4; Epstein-Barr virus / Human Herpesvirus 4) のプローブが乳がんサンプルのそれぞれ 92 %、65 %、96 %、78 % から有意に検出されました。また、パピローマウイルス科については、特異的なプローブにより Human Papillomavirus (HPV) 6b、HPV18、HPV2、HPV16 が乳がんサンプルの 78.8 %、75 %、84.6 %、78.8 % でそれぞれ検出されました。また、特異的プローブにより、GB、C、B 型肝炎ウイルスのシグナルがそれぞれ 82.7 %、90.4 %、86.5 % のがんサンプルで検出されました。興味深いことに、これらのウイルスに対する特異的プローブがすべて検出されたわけではありませんでした⁸。これは、プローブ領域と同じ配列を持つ類似生物の断片が存在する、生物の DNA 断片が統合されているなど、いくつかの可能性が考えられます。

ハイブリダイゼーションシグナルを考慮せずに、有病率順にウイルスのプローブを並べると、ヘパドナウイルスとフラビウイルスが 86.5 %、パラポックスウイルスが 83.3 %、ヘルペスウイルスが 83.2 %、レトロウイルス (79.6 %)、パピローマウイルス (79.3 %) となりました。しかし、ハイブリダイゼーションシグナルが高い順に並べると、腫瘍サンプル全体でヘルペスウイルスのプローブが最もシグナルが高く、次いでパラポックスウイルス、フラビウイルス、ポリオーマウイルス、レトロウイルス、ヘパドナウイルス、パピローマウイルスの順で高いハイブリダイゼーションシグナルを示しました (表 2)。また、細菌、真菌、寄生虫についても同様のデータが作成されました⁸。

表 2. トリプルネガティブ乳がん 100 検体から検出された、ハイブリダイゼーションシグナル (ゲノム配列ごとの全プローブのハイブリダイゼーションシグナルの合計として算出) とウイルスプローブの有病率。候補を検出した方法を以下に示す。

AO: Accession outlier、AT : Accession t-test、SO : Specific probe outliers、ST: Specific probe t-test CO : Conserved probe outlier、CT: Conserved probe t-test、MAT: Model based analysis for tiling array.

(a) Associated viral agent	Detection methods	Percent detected (%)	Probe sum/ accession
Human herpesvirus 5/HCMV	AO, SO, ST, CO, MAT	92	14332000
Human herpesvirus 8/KSHV	AO, SO, MAT	96	12119800
Simian virus 40	AO, SO, MAT	75	8113970
Hepatitis C virus genotype 1	SO, CO, MAT	90	7199330
Human T-lymphotropic virus 2	AO, SO, CO, MAT	88	7040500
Orf virus	CO, MAT	75	6422460
Pseudocowpox virus	AO, SO, CO, MAT	90	5037880
Human herpesvirus 4/EBV	AO, SO, CO, MAT	79	5024970
Bovine papular stomatitis virus	AO, SO, CO, MAT	85	4214040
Okra mosaic virus	AO, SO, CO, MAT	75	3435060
Human papillomavirus 2	SO, MAT	85	3361460
Human T-lymphotropic virus 1	AO, SO, CO, MAT	83	2745990
Hepatitis B virus	AO, SO, CO, MAT	87	2621640
Human herpesvirus 1	AO, SO, CO, MAT	65	2319570
Human papillomavirus type 16	SO, MAT	79	1651350
Moloney murine leukemia virus	SO, CO, MAT	58	1587600
Merkel cell polyomavirus	AO, SO, MAT	90	1551830
Mouse mammary tumor virus	AO, SO, MAT	79	1464980
Human papillomavirus type 6b	AO, SO, MAT	79	1271950
Human papillomavirus 18	SO, CT, MAT	75	1184610
JC polyomavirus	AO, CO, SO, MAT	77	755288
Hepatitis GB virus A	SO, MAT	83	749098
Fujinami sarcoma virus	SO, CO, MAT	90	691071

がんサンプルで検出されたウイルスの保存領域プローブと特異的プローブ、およびがんサンプルでの細菌、真菌、寄生虫のプローブの概要を表 3 および表 4 に示します⁸。トリプルネガティブ乳がんサンプルで検出されたウイルス、細菌、真菌、寄生虫のシグネチャーは、マッチドおよび非マッチドコントロールと比較して、がんサンプルと有意に関連していることがわかりました ($p < 0.05$)⁸。

表 3. PathoChip によるトリプルネガティブ乳がん検体のスクリーニングで検出されたウイルスプローブシグネチャーの数

	Retroviridae				Polyomaviridae				Herpesviridae			
	MMTV	MMLV	HTLV1	HTLV2	FSV	SV40	JC	MCPV	HCMV	EBV	KSHV	HHV1
Total probes	31	24	41	86	8	41	42	62	299	235	259	22
Specific	31	15	37	84	5	41	40	62	275	149	256	15
Outlier	1	4	24	43	4	25	12	27	139	67	132	7
t-test	30	0	0	0	0	0	0	0	1	0	0	0
Conserved	0	9	4	2	3	0	2	0	24	86	3	7
Outlier	0	2	3	2	2	0	1	0	15	2	3	3
t-test	0	0	0	0	0	0	0	0	0	0	0	0

	Papillomaviridae				Hepadnaviridae	Flaviviridae		Poxviridae		
	HPV16	HPV18	HPV6b	HPV2	HBV	HCV-1	HepGB virus A	BPSV	PCP	ORF
Total probes	68	85	91	92	49	121	14	109	105	111
Specific	67	84	90	92	47	119	14	12	12	13
Outlier	19	28	37	49	25	72	7	1	3	1
t-test	0	0	0	0	0	0	0	0	1	0
Conserved	1	1	1	0	2	2	0	97	93	98
Outlier	0	0	0	0	2	0	0	74	80	76
t-test	0	1	0	0	0	0	0	0	0	0

表 4. PathoChip によるトリプルネガティブ乳がん検体のスクリーニングで検出された細菌、真菌、寄生虫のプローブシグネチャーの数

Microbial signatures	Total no. of probes in the Chip	Total no. of probes detected	Detection in triple negative breast tumors	Type of agent
Members	Specific	Specific	Percent positive	Organism
<i>Arcanobacterium</i>	4	4	75	Bacteria
<i>Brevundimonas</i>	3	3	73	Bacteria
<i>Sphingobacteria</i>	5	5	67	Bacteria
<i>Providencia</i>	1	1	67	Bacteria
<i>Prevotella</i>	2	2	67	Bacteria
<i>Brucella</i>	10	10	65	Bacteria
<i>Escherichia</i>	13	10	64	Bacteria
<i>Actinomyces</i>	4	4	52	Bacteria
<i>Mobiluncus</i>	4	4	50	Bacteria
<i>Propionibacteria</i>	2	2	50	Bacteria
<i>Geobacillus</i>	2	1	44	Bacteria
<i>Rothia</i>	3	3	40	Bacteria
<i>Peptinophilus</i>	2	2	39	Bacteria
<i>Capnocytophaga</i>	1	1	37	Bacteria

Microbial signatures	Total no. of probes in the Chip	Total no. of probes detected	Detection in triple negative breast tumors	Type of agent
Members	Specific	Specific	Percent positive	Organism
<i>Pleistophora</i>	8	8	98	Fungi
<i>Piedra</i>	6	6	90	Fungi
<i>Foncecaea</i>	3	3	89	Fungi
<i>Phialophora</i>	4	4	87	Fungi
<i>Paecilomyces</i>	4	4	69	Fungi

<i>Trichuris</i>	7	7	96	Parasite
<i>Toxocara</i>	1	1	62	Parasite
<i>Leishmania</i>	6	5	60	Parasite
<i>Babesia</i>	2	2	56	Parasite
<i>Thelazia</i>	1	1	40	Parasite
<i>Paragonimus</i>	3	2	15	Parasite

結論

高度にマルチプレックス化したメタゲノムアッセイにおいて、ヒトゲノムの圧倒的なバックグラウンドから、サイズの小さな非ヒトゲノムを検出する能力は、いくつかの要因に依存します。それらの要因には、核酸の抽出と回収、ターゲットサイズとコピー数、増幅効率（WTA を使用した場合）、特異的プローブの性能などが含まれます。PathoChip の開発にあたっては、アッセイの感度を向上させるためにいくつかの改良を加えました。具体的には、1 つのゲノム配列に対して複数のプローブを含めることと、さまざまなレベルのデータ解析から得られた候補を統合することがあげられ、スクリーニングプロジェクトにおける病原体検出の最適化を可能にしました。

飽和したタイリングのプローブセット、RNA および DNA の検出を組み合わせた PathoChip は、既知の発がん性病原体やその他の微生物のスクリーニングを強化することができます。したがって、PathoChip アッセイにより、複数の微生物の同時感染の頻度や、発がんプロセスやその他の疾患の進行との相関関係を包括的に評価することが可能になります。

PathoChip によるスクリーニングプロジェクトでは、検出のシグナル強度、階層的クラスタリングを含む複数の解析ストラテジーによる検出、およびサンプル集団全体における検出率によって優先順位を付けた候補リストを作成することができます。これらの結果を、宿主の範囲、組織特異性、一般集団における有病率など、ウイルスや病原性微生物に関するアノテーションと組み合わせることで、どの病原体がさらなる解析に値するかを決定するのに役立ちます。これらの結果は、PCR またはキャプチャープローブベースのハイブリダイゼーションと次世代シーケンサーの組み合わせによってフォローアップ調査することもできます。このアプローチにより、特定のがんや疾患について、様々な程度で寄与する微生物の有望なシグネチャーが得られると期待されます。

References

1. de Martel C, Ferlay J, Franceschi S, Vignat J, Bray F, Forman D, Plummer M. “Global burden of cancers attributable to infections in 2008: a review and synthetic analysis.” *Lancet Oncol.* 13:607–615 (2012).
2. Brodie EL, Desantis TZ, Joyner DC, Baek SM, Larsen JT, Andersen GL, Hazen TC, Richardson PM, Herman DJ, Tokunaga TK, Wan JM, Firestone MK. “Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation.” *Appl. Environ. Microbiol.* 72:6288–6298, (2006).
3. Wong CW, Heng CL, Wan Yee L, Soh SW, Kartasasmita CB, Simoes EA, Hibberd ML, Sung WK, Miller LD. “Optimization and clinical validation of a pathogen detection microarray.” *Genome Biol.* 8:R93, (2007).
4. Chen EC, Miller SA, DeRisi JL, Chiu CY. “Using a pan-viral microarray assay (Virochip) to screen clinical samples for viral pathogens.” *J. Vis. Exp.* 50:2536 (2011).
5. Tu Q, Yu H, He Z, Deng Y, Wu L, Van Nostard JD, Zhou A, Voodeckers J, Qin Y, Hemme CL, Shi Z, Xue K, Yaun T, Wang A, Zhou J. “GeoChip 4: a functional gene array- based high throughput environmental technology for microbial community analysis.” *Mol. Ecol. Resour.* 14:914–928, (2014).
6. Baldwin, D. A., Feldman, M., Alwine, J. C. & Robertson, E. S. “Metagenomic assay for identification of microbial pathogens in tumor tissues.” *MBio* 5. e01714–01714, (2014).
7. Banerjee S, Tian T, Wei Z, Peck KN, Shih N, Chalian AA, O’Malley BW, Weinstein GS, Feldman MD, Alwine J, Robertson ES. “Microbial Signatures Associated with Oropharyngeal and Oral Squamous Cell Carcinomas.” *Sci Rep.* 7(1):4036 (2017).
8. Banerjee S, Wei Z, Tan F, Peck KN, Shih N, Feldman M, Rebbeck TR, Alwine JC, Robertson ES. “Distinct microbiological signatures associated with triple negative breast cancer.” *Sci Rep.* 5:15162, (2015).
9. Shiovitz S, Korde LA. “Genetics of breast cancer: a topic in evolution.” *Ann Oncol.* 20, (2015).

[お問い合わせ窓口]

アジレント・テクノロジー株式会社

本社 / 〒192-8510 東京都八王子市高倉町9-1

●カスタマコンタクトセンター 0120-477-111

mail : email_japan@agilent.com

※仕様は予告なく変更する場合があります。

※本資料掲載の製品は全て研究用です。

診断目的にご利用いただくことはできません。

G240619

<http://www.agilent.com/chem/genomics:jp>

© Agilent Technologies, Inc. 2020

本書の一部または全部を画面による事前の許可なしに複製、
改変、翻訳することは、著作権法で認められている場合を除き、
法律で禁止されています。

Printed in Japan, March, 2024

5994-1968JAJP