

201331003A

厚生労働科学研究費補助金  
(難病・がん等の疾患分野の医療の実用化研究事業)

小児科・産科領域疾患の大規模遺伝子配列解析による  
病因解明とゲノム解析拠点整備

平成25年度 総括・分担研究報告書

研究代表者 梅澤 明弘

平成26(2014)年3月

厚生労働科学研究費補助金  
(難病・がん等の疾患分野の医療の実用化研究事業)

小児科・産科領域疾患の大規模遺伝子配列解析による  
病因解明とゲノム解析拠点整備

平成25年度 総括・分担研究報告書

研究代表者 梅澤 明弘

平成26(2014)年3月

# 目 次

I. 総括研究報告書	
小児科・産科領域疾患の大規模遺伝子配列解析による 病因解明とゲノム解析拠点整備 .....	3
梅澤 明弘	
II. 分担研究報告書	
1. 小児科・産科領域疾患の大規模遺伝子配列解析による 病因解明とゲノム解析拠点整備 .....	13
秦 健一郎 他	
2. 大規模シーケンスのデータ解析技術の開発および 常在細菌叢ゲノムの解析に関する研究 .....	27
服部 正平	
III. 研究成果の刊行に関する一覧表 .....	31
VI. 研究成果の刊行物・別刷り .....	35

# I. 総括研究報告書

小児科・産科領域疾患の大規模遺伝子配列解析による病因解明とゲノム解析拠点整備  
 (H23－実用化(難病)－一般－003)

研究代表者 梅澤 明弘 独立行政法人国立成育医療研究センター研究所・副所長

研究要旨：

成育疾患、特に、小児科・産科領域の難治性疾患や稀少疾患に対し、次世代シーケンサー等の大量配列解析装置を用いた網羅的な遺伝子解析を行い、関連遺伝因子を解明する事を目的とする。一昨年度・昨年度に引き続き、本研究事業の遂行に不可欠な次世代シーケンサーによるデータの取得とバイオインフォマティックなデータ解析系を構築・改良と、臨床的診断を目的とした遺伝子解析支援を並行して進めた。リシーケンシング解析およびトランスクリプトーム解析については、自動化装置導入によるスループットの大幅向上を達成すると共に、データ解析のためのパイプラインを独自に構築し、データ解析の迅速化と柔軟な応用改変を可能とした。また疾患エピゲノム解析のためのライブラリー作製実験系を整備し、データ解析系の構築にも着手した。次世代シーケンサー・アレイを用いた迅速遺伝子診断システムにより、これまでに様々な遺伝性疾患を含む合計 2,200 症例の解析を行い、そのうち特に臨床診断を目的とした約 600 症例の解析では、23%程度が確定診断に至った。

研究分担者 氏名	村島 温子	独立行政法人国立成育医療研究センター・部長
深見 真紀	独立行政法人国立成育医療研究センター・部長	中澤 温子
服部 正平	東京大学・教授	新関 寛徳
松本 健治	独立行政法人国立成育医療研究センター・部長	堀川 玲子
小野寺 雅史	独立行政法人国立成育医療研究センター・部長	大矢 幸弘
奥山 虎之	独立行政法人国立成育医療研究センター・部長	連携研究者（成育医療研究センター）：
秦 健一郎	独立行政法人国立成育医療研究センター・部長	周産期病態研究部
東 範行	独立行政法人国立成育医療研究センター・部長	周産期ゲノミクス研究室長
藤原 成悦	独立行政法人国立成育医療研究センター・部長	中林一彦
中村 浩幸	独立行政法人国立成育医療研究センター・部長	（ライブラリー構築・シーケンシング）
田上 昭人	独立行政法人国立成育医療研究センター・部長	システム発生再生医学研究部
高田 修治	独立行政法人国立成育医療研究センター・部長	組織工学研究室長
		岡村浩司
		（インフォマティクス）
		周産期病態研究部・胎児発育研究室長
		前原佳代子
		（SNP アレイ解析）

## A. 研究目的

次世代シーケンサー等の大規模配列解析装置を駆使し、小児科・産科領域の難治性疾患や稀少疾患の関連遺伝因子を解明する事を目的とする。多くの小児先天性疾患、異常妊娠は、稀少性に加え、点変異や微細欠失、多因子、de novo 変異、エピゲノム変異の背景があると推測され、次世代シーケンサーやマイクロアレイ技術による網羅的配列解析が必須かつ極めて有効と期待される。

## B. 研究方法

### 1. 対象疾患

新生児食物蛋白誘発胃腸炎、壊死性腸炎 (NEC)、炎症性腸疾患 (分担研究 野村、松本) :

キャンボメリックディスプラジア (分担 浅原、高田) :

慢性活動性 EB ウイルス感染症、EB ウイルス関連 T/NK リンパ増殖性疾患 (分担 藤原、中村) :

胆道閉鎖症 (分担 田上) :

反復胞状奇胎、異常妊娠、リファレンスデータ整備 (分担 秦) :

原発性免疫不全症 (分担 小野寺、奥山) :

造血幹細胞遺伝子治療後のプロウイルス挿入部位の網羅的解析 (分担 小野寺、奥山) :

小児眼疾患 (分担 東) :

先天性内分泌疾患 (分担 堀川) :

リンパ管腫 (分担 藤野) :

難治性発作性気道閉塞障害 (分担 大矢)

### 2. 次世代シーケンサー解析用ライブラリー作製技術系の構築 (効率化・微量化・新規技術導入)

#### ①ライブラリー作製自動化・効率化

次世代シーケンサー (next generation sequencer, NGS) ライブラリー作製自動化装置を導入し、エクソーム濃縮ライブラリーならびに RNA-Seq ライブラリー作製工程の効率化を行った。現在、一名の専任者がライブラリー作製と NGS 操作全般を行う体制で、年間最大 500 サンプル程度のエクソーム解析が可能である。

また、限られた数の標的遺伝子群を対象としたエクソーム解析を実施する場合に有効な方法である Haloplex 法を導入した。この方法では、従来のカスタムエクソーム濃縮法と比較して、より少ない量のゲノム DNA から、より短時間でライブラリーを構築できる。比較的少数

(数十から百程度) の遺伝子変異解析を、比較的少数の症例 (数百例程度) を行うのに有効である。また、エクソーム解析でカバーされないプロモーター領域等の任意の領域の解析もデザイン可能であるという利点がある。

#### ②長鎖ゲノム DNA ライブラリーを簡便作製するプロトコルの確立

次世代シーケンサーによるゲノム構造解析においては、インサート鎖長が短い (300 bp 程度) 通常の NGS ライブラリーよりも、インサート平均鎖長が長いゲノムライブラリーからデータを取得する方が有利であるが、通常のアダプター付加法では長鎖ライブラリーの作製効率は高くない。そこでトランスポセーズ組換えにより微量ゲノム DNA からインサート平均鎖長 2-5kb 程度のゲノムライブラリーを作製した。

#### ③メイトペアライブラリー作製プロトコルの検討

ゲノム構造解析やゲノム配列の新規アセンブリー (de novo assembly) を目的とする NGS 解析では、メイトペアライブラリー法がよく用いられる。この方法は通常の NGS ゲノムライブラリー作製法と比べて工程が長い。インサート長が長い (>10kb) ライブラリーの構築も原理的には可能であるが、各実験工程の条件を最適化することで初めて可能となる。本年度はそのための準備として、まずはインサート長 2-3kb のメイトペアライブラリーの作製を行った。

#### ④FFPE 由来 gDNA および total RNA からの NGS ライブラリー作製条件の検討

ホルマリン固定パラフィン包埋 (Formalin-Fixed Paraffin-Embedded: FFPE) 組織は、最も一般的にアーカイブされる組織標本の形態であり、FFPE 組織が入手可能な唯一の臨床検体である場合にも頻繁に遭遇する。一方で、そこから得られる核酸試料 (ゲノム DNA・total RNA) は断片化や化学修飾を受けており、ライブラリー作製が困難な場合が多い。FFPE 由来 gDNA および total RNA からの NGS ライブラリー作製条件の検討を行う。

#### ⑤エピゲノム解析用 NGS ライブラリー作製プロトコルの構築

DNA メチル化やヒストン修飾などのエピゲノム解析により疾患部位のエピゲノム動態を解析することは、疾患の病態解明の基盤情報となる。疾患検体のエピゲノム解析 (DNA メチ

ル化解析・ヒストン修飾解析)の実施に必要なプロトコールとして、1)クロマチン免疫沈降 DNA ライブラリー (ChIP-seq ライブラリー)、2) RRBS (reduced representation bisulfite sequencing) ライブラリー、3) PBAT (post-bisulfite adaptor tagging) ライブラリーの作製系を構築した。

### 3. 次世代シーケンサーデータ解析系の構築：改良および新規解析系導入

生命科学のさまざまな分野に普及しつつある次世代シーケンサーは、単に DNA 塩基配列を決定するにとどまらず、遺伝子発現、DNA メチル化、クロマチン、ヌクレオソームの解析や、塩基から染色体構造にわたる変異検出、疾患原因遺伝子の検出など多岐にわたり、医学の分野においても研究手法を一変させつつある。本拠点においては昨年度に Illumina HiSeq 1000 および MiSeq を導入し、その後共通機器として HiSeq2500 を導入した。また、パイオインフォマティクスの面では、配列データの品質確認、トリミングから、リファレンス配列へのマッピング、ヴァリエント検出等までを自動で行うパイプラインを構築し、二次解析以降の効率化を図るとともに、頻度情報やアミノ酸置換によるタンパク質への影響予測など、病因変異解析に有用な情報を付加し、利便性を図った。

#### (倫理面への配慮)

本研究は、ヒトゲノム・遺伝子解析研究に関する倫理指針に基づいて計画され、本センターおよび連携研究者の所属する機関の倫理委員会の承認を得て行っている。すべての検体は、書面で患者本人もしくは代諾者からインフォームドコンセントを得た後に採取され、各医療機関で匿名化された。網羅的遺伝子解析については、説明書・同意書を用いて同意を取得した。

## C. 研究結果

### 1. 疾患検体群を対象とした次世代シーケンサー解析・アレイ解析

本事業に参加している分担研究者から提供された検体と、全国 45 か所の研究協力機関から寄せられたヒト疾患検体を中心に、合計約 2,200 症例の解析を行った。そのうち特に臨床診断を目的とした 600 症例の解析では、約 23% が確定診断に至った。

### 2. 次世代シーケンサー解析用ライブラリー作製技術系の構築 (効率化・微量化・新規技術導入)

#### ①ライブラリー作製自動化・効率化

##### 1) エクソーム濃縮ライブラリー (SureSelect) 作製自動化

使用するゲノム DNA 濃度は picogreen 定量キット (Invitrogen) により定量し、1-2 ug をゲノムライブラリー作製に供した。

SureSelectXT all exon 80Mb (もしくは 51Mb) キットを用いてライブラリーを作製した。ライブラリー作製工程を図 1 に示した。大部分の工程において自動化システムを使用した。DNA ライブラリーの評価 (サイズ確認・濃度推定) には Bioanalyzer (Agilent) もしくは LabChip GX (Caliper 社) を用いた。自動化装置を用いてこれまでに 481 検体のライブラリーを作製し、いずれも再現性良く安定して適切な平均インサート鎖長のライブラリーが作製できた。

##### 2) TruSeq RNA ライブラリー作製の自動化

TruSeq RNA Sample Preparation v2 自動化対応プロトコール (図 3) を導入し、これまでに 272 検体のライブラリーを作製した。

3) HaloPlex ターゲットエンリッチメントシステムは、任意のゲノム領域合計<500Kb を標的としてカバーできる。数百 ng のゲノム DNA を出発材料として安定して再現性良くライブラリー作製が可能であった。また、アジレント自動化システムを併用することで、さらに効率的にライブラリーを作製する体制が構築できた。

#### ②長鎖ゲノム DNA ライブラリーを簡便作製プロトコールの確立

30ng のゲノム DNA を出発材料とし、Nextera DNA Sample Preparation Kit (イルミナ社) を用いてライブラリーを調整した。この試薬系においては、トランスポゾン活性酵素によるトランスポゾン転移反応を応用してゲノム DNA 断片化・アダプター配列付加をワンステップで行うことが可能であり、従来法に比べて所用時間が約四分の一、必要ゲノム DNA 量が約十分の一と僅かであることがこの方法の長所である。酵素量 10%量でインサート鎖長 1400~2200bp にピークを観察することを確認し、酵素量 10% でライブラリーを作製した。

#### ③メイトペアライブラリー作製プロトコールの確立

Mate-Paired Library Preparation 5500SOLiD

System キットを用い、メイトペアライブラリーを作製した。5 ug のゲノム DNA を出発材料とし、最終的に平均鎖長 250 前後の bp のライブラリーを安定して 作製可能となった。

#### ④FFPE 由来 gDNA および total RNA からの NGS ライブラリー作製条件の検討

ホルマリン固定後のパラフィン包埋試料 (FFPE) から、レーザーマイクロダイセクション等により、gDNA あるいは total RNA を出発材料として、ライブラリー作製を行った。一般的に gDNA の偏りの無い増幅とライブラリー作製が困難であったのに対し、RNA の増幅は容易であった。今後これらの NGS 解析を進め、品質を検証する。

#### ⑤エピゲノム解析用 NGS ライブラリー作製プロトコルの構築

##### 1) ChIP-seq ライブラリー作製

特定の転写因子に対する抗体を用いて実施したクロマチン免疫沈降法 (ChIP) により得られた免疫沈降 DNA を出発材料として、NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (New England Biolabs) を用いてライブラリーを作製した。大阪大学木村宏博士との共同研究により、高品質の特異的モノクローナル抗体を用い、現在  $10^7$  個弱の細胞から数種類の ChIP とライブラリー作製が可能である。

##### 2) RRBS (reduced representation bisulfite sequencing) ライブラリーの作製

全ゲノムメチル化パターンの解読 (WGBS 法あるいは PBAT 法) は、解析 (特に配列解読) に要する費用が膨大なのに対し、Reduced representation bisulfite sequencing (RRBS) 法は、より少ない費用で、ほぼゲノム全域にわたって DNA メチル化解析することを目的としている。制限酵素 MspI を利用してゲノム DNA を断片化することで DNA メチル化の標的となりうる領域を濃縮し、bisulfite 配列解読を行い、メチル化率を解析する。現在、いくつかの DNA メチル化異常症例を対象に、RRBS 法による DNA メチル化解析を行っている。

##### 3) PBAT (post-bisulfite adaptor tagging) ライブラリーの作製

バイサルファイト変換反応後にアダプターを付加するため、DNA の損失が少なく、少量のゲノム DNA で解析可能なのが PBAT 法の特徴である 100 ng (2 万個程度の細胞に相当) のゲノム DNA を用いれば、PCR 増幅無しで十分なライブラリーを得ることが可能であった。

#### 3. 次世代シーケンサーデータ解析系の構築：自動化・効率化および新規解析系導入

##### ①リーシケンシング解析

本研究班独自に、次世代シーケンサーデータ解析のためのパイプラインを構築した。本パイプラインは Linux 等、UNIX ライクな環境の CUI で動作する。コマンド入力に馴染みのないユーザでも利用し易いように、特に設定変更の必要がなければコマンド 1 つで動作するように設計した。すでに当研究所内で得られた 150 検体以上の全エクソンエクソーム解析データに対しての解析実績がある。

##### ②全ゲノム・リシークエンシング

本研究で頻用されるモデルマウスとその亜種系統を用い、両者の比較では全ゲノムに渡って平均 150 bp に 1 ヶ所程度の多型検出に成功した。

##### ③トランスクリプトームデータ解析

###### i) 二次解析 (マッピング)

RNA-seq ライブラリーのシーケンスデータ (fastq ファイル) のマッピング工程をパイプライン化した。

###### ii) Avadis NGS を用いた三次解析

分担研究者らの利便性のために、商業的に入手できる Avadis NGS ソフトウェアによる解析環境を整備した。複数サンプル間での遺伝子発現量比較、スプライスアイソフォーム毎の発現量比較、新規アイソフォームの検出、SNP 情報を用いたアレル別発現解析、gene ontology 解析、パスウェイ解析などに利用していただいた。

##### ④ChIP-Seq データ解析

特定の DNA 結合タンパクのゲノム標的部位の同定や、ヒストン修飾のゲノムワイド解析を目的に、ChIP-Seq データ解析系を構築した。

##### ⑤DNA メチル化解析

バイサルファイト処理後にアダプター付加を行う PBAT 法で得られた配列データを解析するための専用パイプラインを構築し、ヒトゲノムおよびマウスゲノムのリファレンスに含まれる全 CpG サイトの DNA メチル化率を算出できる体制を整備した。DNA メチル化異常を伴うサンプルの解析を進めている。

#### D. 考察

##### 1. 次世代シーケンサー解析用ライブラリー作製技術系の構築



### ①自動化プロトコル評価

アジレント社 NGS ライブラリー作製自動化装置を導入し、SureSelect エクソーム濃縮ライブラリー、TruSeq RNA-Seq ライブラリー、Haloplex ライブラリーの作製工程を半自動化し、ライブラリー作製効率が大幅に向上した。現在は、一度の作業で 24-96 検体を処理することが可能である。また、LabChipGX (Caliper 社) の導入により、インサート長と濃度測定等の工程も効率化した。

自動化プロトコルでは、全エクソン配列解析に供するためのゲノム DNA 濃縮が主導と比較してやや劣る傾向が見られるが、濃縮後の PCR サイクル数を標準プロトコルより 2 サイクル増やして対応した。

### ②NGS ゲノムライブラリー作製における Nextera DNA Sample Preparation Kit の有効性確認

Nextera DNA Sample Preparation Kit (illumina) は必要 DNA 量が 50 ng と少なく、また、ライブラリー作製の所要時間も約 90 分と短いのが特徴である。今回、1000 bp 以上のインサート鎖長を有するライブラリーの作製を目指したが、gDNA サンプルによって tagmentation 効率に差が見られ、DNA 量、DNA 転移酵素量、反応時間を複数条件検討する必要があった。今のところ、他の条件を固定し、DNA 転移酵素量を複数条件試すのが最善と考えている。断片化 gDNA サイズを 4~5 kb に調節することは可能であったが、その後の PCR 反応における高分子領域 (2 kb 以上) の増幅が困難であった。本キットで作製できるライブラリーのインサート鎖長は 2-3 kb 程度が上限と思われる。

### ③FFPE 由来 gDNA/total RNA からの NGS ライブラリー作製

Nextera DNA Preparation kit による FFPE 由来 gDNA のライブラリー化に成功した。FFPE 組織切片から抽出したゲノム DNA 量は少量である場合も多いことから、微量ゲノム DNA からのライブラリー作製が可能なこのキットは FFPE 検体の NGS 解析に極めて有効である。

FFPE 由来 total RNA のライブラリー化には、Ovation RNA-Seq FFPE System による cDNA 増幅が有効であった。このシステムでは増幅バイアスが少なく各転写産物の相対的な発現量が保持される酵素系が採用されており、100~200 ng の total RNA から ug オーダーの cDNA を得ることができる。本キットを用いて FFPE

由来 total RNA から cDNA を合成し、さらに TruSeq RNA Preparation kit にて平滑末端化以降の操作を行うことで、ライブラリー化に成功した。

### ④エピゲノム解析用 NGS ライブラリー作製プロトコルの構築

NEBNext ChIP-Seq Library Prep kit (NEB) のプロトコルでは、10 ng の免疫沈降 DNA を出発材料として、x18 サイクルの PCR 増幅を経てライブラリーを作成するよう推奨されているが、10 ng の免疫沈降 DNA を使用した場合は x6 PCR サイクルで十分であること、0.1 ng の免疫沈降 DNA を使用した場合でも x13 PCR サイクルで十分量のライブラリーが得られた。

二つの異なる DNA メチル化解析プロトコル (PBAT 法ならびに RRBS 法) を確立した。両者はそれぞれ異なる長所を有しており、取得したいデータ内容、検体数、入手可能なゲノム DNA 量などにより、使い分けが可能である。バイオインフォマティク観点からは、PBAT データからモデルを構築し、RRBS データ解析に転用することができた。

### 2. 次世代シーケンサーデータ解析系の構築 :

本研究事業の支援により、エクソーム解析、リシークエンシング、RNA-seq 等の独自の解析系 (パイプライン) を構築した。これにより、初年度は外部委託が必要だったバイオインフォマティクスを自前で実行することが可能となった。この結果、解析が早くなる (通常外部委託では、まとまった数の検体がそろってまで発注をかけないため) のみならず、解析系の細かな調整や変更が可能となったため、RNA-seq によるアレル別発現解析や PBAT 法のデータ解析など、特殊な解析手法を独自に確立することができた。

このように、独自性の高い、使いやすいバイオインフォマティクス体制を構築するに伴い、これまで気づきにくかった問題点も表面化してきた。すなわち、高精度に変異・多型候補の検出が可能になると共に、標準配列にうまくマッピングできない変異・多型候補の検出がわかってきた。これは主に、標準配列が多数の日本人集団を基に得られたものでないことが原因と考えられる。当研究事業の最大の目標である、本邦の医療への貢献の為には、日本人の標準配列情報が必要である。しかも、当研究事業の場合は、小児科・産科領域に利用可能な標準配列が必要であるため、例えば妊娠分娩歴の明らか

な日本人の標準配列情報が必要と考えられる。

申請時、解析症例数を合計 500 例としていたが、早い段階で解析体制の立ち上げに成功したことが功を奏し、最終的に約 2,200 症例の解析を終了させることができた。成育疾患のゲノム解析拠点体制を整備するというもっとも主要な目的は、十分達成することができたと考える。また、本研究事業では、ゲノム異常（新規変異同定）のみならず、エピゲノム異常（DNA メチル化異常）等の解析を行った研究成果・論文に、学術的特徴があると考えられる。国際的意義としては、国際学術誌への投稿・国際共同研究・海外学会での発表に加え、本研究事業で培われた解析技術を基に発展的に国際ヒトエピゲノムコンソーシアムへの参加など、様々な国際貢献も行った。社会的意義としては、約 600 症例の遺伝子検査的解析を行い、その四分の一で確定診断に至ることができたことが挙げられる。

今後は、次世代シーケンサーの急速な普及に伴い、臨床の現場からの遺伝子検査としての遺伝子解析の要望が高まると予想される。その一方で、今後も本邦で医療として持続可能な遺伝子検査体制を構築していくことは、喫緊の課題であると考えられる。具体的には、遺伝子検査の拠点集約化、厳密な適応基準の作成、結果返却に伴う学術的解釈あるいは倫理的対応に関する支援体制、等の整備が挙げられる。

申請書にも記載したように、本研究班は遺伝子検査としての配列解析も積極的に受けている。その結果、約半数の解析が確定診断を目的とした遺伝子検査的解析であり、その約四分の一で分子遺伝学的確定診断がついた。非常に効率の良い遺伝子診断が可能であった背景には、本研究事業では、分担研究者らのスクリーニングを経た検体のみを解析したことが挙げられる。

## E. 結論

小児科・産科領域の原因不明疾患に対する遺伝因子の解明を目的とし、次世代シーケンサー等の大量配列解析装置を用いた様々な解析体制を整備した。特に本事業の特徴として、ゲノム解析に加え、エピゲノム解析体制も構築した。エクソーム解析には半自動化装置を導入し、スループットの大幅向上を達成した。また、独自のパイプライン構築によりデータ解析迅速化すると共に、変異によるタンパク質への影響予測、1000 人ゲノムや独自の日本人標準配列情報に基づく多型頻度など、様々な情報を付加し、病因変異抽出のために使いやすい解析結果

出力を可能とした。また、疾患エピゲノム解析のための準備を進め、網羅的 DNA メチル化解析、ヒストン修飾解析を行った。次世代シーケンサー・アレイを用いた迅速遺伝子診断システムにより、これまでに様々な遺伝性疾患を含む合計約 2,200 症例の解析を行い、そのうち特に臨床診断を目的とした 600 症例の解析では、約 23%が確定診断に至った。本研究事業により、様々な高度の遺伝子診断技術の提供が可能となり、今後はこれらの技術の均てん化すなわち、遺伝子解析支援体制の為のより具体的な運用体制の検証が急務であると考えられる。

## F. 健康危険情報

なし

## G. 研究発表

### 1. 論文発表

Nakamura K, Aizawa K, Nakabayashi K, Kato N, Yamauchi J, Hata K, Tanoue A : DNA methyltransferase inhibitor zebularine inhibits human hepatic carcinoma cells proliferation and induces apoptosis. *PLoS One*. 2013;8(1):e54036.

Iglesias-Platas I, Martin-Trujillo A, Cirillo D, Court F, Guillaumet-Adkins A, Camprubi C, Bourc'his D, Hata K, Feil R, Tartaglia G, Arnaud P, Monk D : Characterization of novel paternal ncRNAs at the Plag1 locus, including Hymai, predicted to interact with regulators of active chromatin. *PLoS One*. 2012;7(6):e38907.

Higashimoto K, Nakabayashi K, Yatsuki H, Yoshinaga H, Jozaki K, Okada J, Watanabe Y, Aoki A, Shiozaki A, Saito S, Koide K, Mukai T, Hata K, Soejima H : Aberrant methylation of H19-DMR acquired after implantation was dissimilar in soma versus placenta of patients with Beckwith-Wiedemann syndrome. *Am J Med Genet A*. 2012;158A(7):1670-1675.

Kobayashi H, Sakurai T, Sato S, Nakabayashi K, Hata K, Kono T : Imprinted DNA methylation reprogramming during early mouse embryogenesis at the Gpr1-Zdbf2 locus is linked to long cis-intergenic transcription. *FEBS Lett*. 2012;586(6):827-833.

Nakanishi MO, Hayakawa K, Nakabayashi K, Hata K, Shiota K, Tanaka S : Trophoblast-specific DNA methylation occurs after the segregation of the trophoblast and inner cell mass in the

mouse periimplantation embryo. Epigenetics. 2012 Feb;7(2):173-182.

Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T: Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet. 2012 ;8(1):e1002440.

秦健一郎：妊娠維持機構とその破綻-ジェネティクス・エピジェネティクスの視点から-。HORMONE FRONTIER IN GYNECOLOGY 2012;19:17-22

右田王介, 中林一彦：次世代シーケンサー・アレイ技術を用いた遺伝子診断。小児科診療 2013;76:395-399

## 2. 学会発表

### (招待講演)

秦健一郎：合併症妊娠におけるゲノム解析。第28回日本糖尿病・妊娠学会年次学術集会，東京，2012.11.17

秦健一郎：次世代遺伝子解析技術を用いた生殖発生異常のゲノム研究。第30回日本受精着床学会学術講演会，大阪，2012.8.31

秦健一郎：最先端テクノロジーを駆使したヒト周産期疾患ゲノム解析への取り組み。第19回遺伝子診断学会，千葉，2012.7.27

秦健一郎：ヒト発生異常のゲノム解析・エピゲノム解析。東京大学大学院農学部応用動物科学セミナー，東京，2012.7.20

秦健一郎：ジェネティクスを越えてエピジェネティクスへ-周産期のエピジェネティクス-。遺伝医学合同学術集会(第35回日本遺伝カウンセリング学会学術集会，第18回日本遺伝子診療学会大会，第17回日本家族性腫瘍学会学術集会)，京都，2011.6.18

秦健一郎：異常妊娠のエピジェネティクス。日本繁殖学会国際シンポジウム，盛岡，2011.9.19

### (一般演題)

Nakabayashi K, Trujillo A.M, Tayama C, Romanelli V, Lapunzina P, Kagami M, Soejima H, Ogata H, Court F, Monk D, Hata K: DNA methylation analysis of reciprocal genomewide UPDs to define imprinted differentially methylated regions in the human genome. The American

Society of Human Genetics 62<sup>nd</sup> Annual Meeting, San Francisco, 2012.11.8

久須美真紀, 中林一彦, 前原佳代子, 田嶋敦, 副島英伸, 吉原一, 秦健一郎：生殖補助医療技術が出生時のDNAメチル化状態に及ぼす影響の検討。日本人類遺伝学会第57回大会，東京，2012.10.27

中林一彦, 田山千春, Trujillo Alex Martin, 岡村浩司, 緒方勤, 副島英伸, Monk David, 秦健一郎：全ゲノム片親性ダイソミー症例のDNAメチル化解析によるヒトインプリントーム解明。日本人類遺伝学会第57回大会，東京，2012.10.26

鏡雅代, 古庄知己, 中林一彦, 松岡健太郎, 松原圭子, 福家智子, 深見真紀, 緒方勤：14番染色体インプリンティング領域メチル化制御機構の解明：MEG3-DMRエピ変異症例の解析から。日本人類遺伝学会第57回大会，東京，2012.10.27

岩田圭子, 中林一彦, 松崎秀夫, 中村和彦, 秦健一郎, 森則夫：自閉症死後脳縫線核メチル化状態の網羅的解析。第39回日本脳科学会，福岡，2012.10.6

秦健一郎：発生とエピジェネティクス。第30回日本受精着床学会，大阪，2012.8.31

久須美真紀, 神通寛子, 秦健一郎, 千田将：妊娠母獣の栄養状態に伴う新生仔レプチン受容体遺伝子領域のエピジェネティックな変化。第64回日本産科婦人科学会，兵庫，2012.4.15

直井裕子, 鳥巢弘道, 久須美真紀, 諸隈誠一, 山田崇弘, 水上尚典, 和氣徳夫, 左合治彦, 秦健一郎：異常妊娠のエピゲノム解析。第64回日本産科婦人科学会高得点演題，兵庫，2012.4.13

由良洋文, 桜庭喜行, 森下和代, 関沢明彦, 秦健一郎, 北川道弘：母体血から分離・濃縮された赤芽球の機器検出。第15回胎児遺伝子診断研究会，東京，2012.2.4

## H. 知的財産権の出願・登録状況

### 1. 取得特許

なし

### 2. 実用新案登録

なし

### 3. その他

なし

## II. 分担研究報告書

厚生労働科学研究費補助金（難病・がん等の疾患分野の医療の実用化研究事業）  
分担研究報告書

小児科・産科領域疾患の大規模遺伝子配列解析による  
病因解明とゲノム解析拠点整備（H23－実用化(難病)－一般－003）

研究分担者	氏名	秦 健一郎	独立行政法人国立成育医療研究センター・部長
		深見 真紀	独立行政法人国立成育医療研究センター・部長
		松本 健治	独立行政法人国立成育医療研究センター・部長
		小野寺雅史	独立行政法人国立成育医療研究センター・部長
		奥山 虎之	独立行政法人国立成育医療研究センター・部長
		東 範行	独立行政法人国立成育医療研究センター・部長
		藤原 成悦	独立行政法人国立成育医療研究センター・部長
		中村 浩幸	独立行政法人国立成育医療研究センター・室長
		田上 昭人	独立行政法人国立成育医療研究センター・部長
		高田 修治	独立行政法人国立成育医療研究センター・部長
		村島 温子	独立行政法人国立成育医療研究センター・部長
		中澤 温子	独立行政法人国立成育医療研究センター・部長
		新関 寛徳	独立行政法人国立成育医療研究センター・部長
		堀川 玲子	独立行政法人国立成育医療研究センター・部長
		大矢 幸弘	独立行政法人国立成育医療研究センター・部長

研究要旨：

次世代シーケンサー等の大規模配列解析装置を駆使し、小児科・産科領域の難治性疾患や稀少疾患の関連遺伝子を解明する事を目的とする。一昨年度・昨年度に引き続き、本研究事業の遂行に不可欠な次世代シーケンサーによるデータ取得系ならびにデータ解析系を構築・改良することを目的として研究を実施した。リシーケンシング解析およびトランスクリプトーム解析については、自動化装置導入によるスループットの大幅向上を達成し、独自のパイプライン構築によるデータ解析迅速化の基盤を確立した。また疾患エピゲノム解析のためのライブラリー作製実験系を整備し、データ解析系の構築にも着手した。次世代シーケンサー・アレイを用いた迅速遺伝子診断システムにより、これまでに様々な遺伝性疾患を含む合計 2,200 症例の解析を行い、そのうち特に臨床診断を目的とした約 600 症例の解析では、23%程度が確定診断に至った。

連携研究者（成育医療研究センター）：

周産期病態研究部

周産期ゲノミクス研究室長

中林一彦

(ライブラリー構築・シーケンシング)

周産期病態研究部・胎児発育研究室長

前原佳代子

(SNP アレイ解析)

システム発生再生医学研究部

組織工学研究室長

岡村浩司

(インフォマティクス)

## A. 研究目的

次世代シーケンサー等の大規模配列解析装置を駆使し、小児科・産科領域の難治性疾患や稀少疾患の関連遺伝因子を解明する事を目的とする。多くの小児先天性疾患、異常妊娠は、稀少性に加え、点変異や微細欠失、多因子、de novo 変異、エピゲノム変異の背景があると推測され、次世代シーケンサーやマイクロアレイ技術による網羅的配列解析が必須かつ極めて有効と期待される。

## B. 研究方法

### 1. 対象疾患

新生児食物蛋白誘発胃腸炎、壊死性腸炎 (NEC)、炎症性腸疾患 (分担研究 野村、松本) :

キャンボメリックディスプラジア (分担 浅原、高田) :

慢性活動性 EB ウイルス感染症、EB ウイルス関連 T/NK リンパ増殖性疾患 (分担 藤原、中村) :

胆道閉鎖症 (分担 田上) :

反復胎状奇胎、異常妊娠、リファレンスデータ整備 (分担 秦) :

原発性免疫不全症 (分担 小野寺、奥山) :

造血幹細胞遺伝子治療後のプロウイルス挿入部位の網羅的解析 (分担 小野寺、奥山) :

小児眼疾患 (分担 東) :

先天性内分泌疾患 (分担 堀川) :

リンパ管腫 (分担 藤野) :

難治性発作性気道閉塞障害 (分担 大矢)

### 2. 次世代シーケンサー解析用ライブラリー作製技術系の構築 (効率化・微量化・新規技術導入)

H23・H24 年度の二年間で、全エクソーム解析、カスタムエクソーム解析、RNA-seq 解析、微量ゲノム DNA からのゲノムライブラリー構築系 (Nextera システム)、Mate-pair ライブラリー構築系、FFPE 由来 DNA および RNA からのライブラリー構築系、CHIP-seq ライブラリー構築、Bisulfite-seq ライブラリー (RRBS 法ならびに PBAT 法) 構築など、ゲノム解析拠点として機関内外の共同研究者の多様なニーズに答えるべく積極的に新規技術導入に努めてきた。

今年度は次のライブラリー作製系の導入を試みた。

① IonAmpliseq 法による簡便かつ安価なカス

タムリシーケンシング法の導入

② 微量 DNA の in vitro transcription による増幅ならびにライブラリー化

### 3. 次世代シーケンサーデータ解析系の構築：改良および新規解析系導入

生命科学のさまざまな分野に普及しつつある次世代シーケンサーは、単に DNA 塩基配列を決定するにとどまらず、遺伝子発現、DNA メチル化、クロマチン、ヌクレオソームの解析や、塩基から染色体構造にわたる変異検出、疾患原因遺伝子の検出など多岐にわたり、医学の分野においても研究手法を一変させつつある。本拠点においては平成 23 年度に Illumina HiSeq 1000 および MiSeq を導入し、データ供出体制と基本的なデータ解析体制を整えた。平成 25 年度には HiSeq1000 を 1500 にアップグレードし、更に HiSeq2500 を導入することでスループットが大幅に向上すると共に、必要データ量に応じたシーケンシング規模の調節が可能となった。

初年度に構築したデータ解析体制のもとでは、特に多検体についてリシーケンシング解析を行う場合は、解析を外部委託することが多かった。そこで大量データを内部で処理する体制を強化すべく、昨年度中に配列データの品質確認、トリミングから、リファレンス配列へのマッピング、ヴァリアント検出等までを自動で行うパイプラインを構築し、二次解析以降の効率化に成功した。しかしながら現状のパイプラインは BAM および VCF 形式のファイルの出力にとどまっておき、例えば疾患責任変異絞り込みのためのアノテーション付与 (1000 人ゲノムプロジェクトにおける日本人アレル頻度情報、OMIM 情報など) や複数サンプルデータの統合などの工程を自動化するには至っていない。実際のユーザからの要望を聞きながら、引き続きデータ加工パイプラインの整備を進めていく予定である。

今年度はエクソーム解析データから疾患責任変異を効率的に絞り込むためのスクリプトの整備を進めた。さらに、RNA-seq データ、エピゲノム解析データ (CHIP-seq データならびにバイサルファイトシーケンシングデータ) の解析系の構築にも着手した。

#### (倫理面への配慮)

本研究は、ヒトゲノム・遺伝子解析研究に関する倫理指針に基づいて計画され、本センター

および連携研究者の所属する機関の倫理委員会の承認を得て行っている。すべての検体は、書面で患者本人もしくは代諾者からインフォームドコンセントを得た後に採取され、各医療機関で匿名化された。網羅的遺伝子解析については、説明書・同意書を用いて同意を取得した。

## C. 研究結果

### 1. 疾患検体群を対象とした次世代シーケンサー解析・アレイ解析

本事業に参加している分担研究者から様々な遺伝性疾患を含む合計2,200症例の解析を行い、そのうち特に臨床診断を目的とした約600症例の解析では、23%程度が確定診断に至った。分担研究者らとの共同研究も進め、その一例として、遺伝性掌蹠角化症の1つである長島型掌蹠角化症（乳児～幼児期から手の平・足の裏の皮膚が紅みをおびて分厚く固くなる疾患。1977年に日本で初めて報告され、原因は不明であった）の原因遺伝子変異が、SERPINB7の変異であることを突き止めた。

これらの解析過程に必須の正常コントロールデータ（対照群データ）は、公開されているデータベース等の情報を利用している報告がほとんどである。しかし、公開されているバイオリソースのほぼ全てが、提供者の妊娠分娩歴に関する情報は付加されておらず、成育疾患の対照群として適切かどうかを判断することができない。そこで本研究で収集した「妊娠分娩歴および出生児に特段の異常を認めない日本人女性」候補変異の絞り込みに決定的な役割を果たす。染色体異常等をはじめとするヒトゲノムの量的異常が、不妊・不育・先天異常等の生殖発生の問題にかかわることは広く知られている。近年ヒトゲノムには、従来の細胞生物学412人の網羅的一塩基多型データを収集し、その遺伝学的背景を明らかにした。その結果、対象に共通して存在するコピー数減少の領域に30の遺伝子を観察した。これらの遺伝子は、周産期の異常との関連が「低い」ことが予測された。

### 2. 次世代シーケンサー解析用ライブラリー作製技術系の構築（別添資料参照）

エピゲノム解析用 NGS ライブラリー作製プロトコルの構築

#### 1) ChIP-seq ライブラリー作製

特定の転写因子に対する抗体を用いて実施し

たクロマチン免疫沈降法（ChIP）により得られた免疫沈降 DNA を出発材料として、NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (New England Biolabs) を用いてライブラリーを作製した。ChIP DNA は Qubit ds HS kit でも測定できないほど微量であったため、input DNA と出発量をそろえることができない。そのため、適当量を用いて ChIP DNA ライブラリーを作製し、増幅に必要な PCR サイクルを見出した後に、同等の増幅量が得られるような input DNA の出発量を決定した。ChIP DNA ライブラリーの増幅に必要な PCR サイクルはサンプルによって異なり、6 サイクルで十分な場合もあれば、13 サイクルまで必要とする場合もあった。input DNA 量の出発量を 0.1～20 ng の間で検討し、増幅程度が ChIP DNA と同等となる濃度を検討した。

#### 2) RRBS (reduced representation bisulfite sequencing) ライブラリーの作製

全ゲノムメチル化パターンの解読（WGBS 法あるいは PBAT 法）は、解析（特に配列解読）に要する費用が膨大であり、複数のサンプルのメチル化パターンの比較には、限界がある。

#### Reduced representation bisulfite sequencing

(RRBS) 法は、より少ない費用で、多くのサンプルのメチル化率を、ほぼゲノム全域にわたって解析することを目的としている。DNA メチル化のほとんどが CpG の 2 塩基上の Cytosine で起こることより、RRBS 法は、制限酵素 MspI を利用しゲノム DNA を断片化し、メチル化を受けやすい領域を配列特異的に絞って (reduced representation) 配列解読を行い、メチル化率を解明する。より多くのサンプルを区別して、一度にシーケンサーで配列解析できるよう、multiplexed adaptor を各サンプルに付加し、かつ、これまで時間と労力を要していたゲル切り出しの工程をビーズ精製に代替し、一度に多くのサンプルを調整する方法を発表した論文 (Boyle et al. Genome Biology 2012, 13:R9) の方法を採用した。この論文で報告されているのと同程度の効率でライブラリーが作製できることを確認した。

#### 3) PBAT (post-bisulfite adaptor tagging) ライブラリーの作製

通常的全ゲノムバイサルファイトシーケンシング法 (whole genome bisulfite sequencing, WGBS) では、ゲノム DNA をバイサルファイト変換する前にアダプターを付加するため、バイサルファイト変換反応時に、アダプター付加

済み DNA の多くが断片化されてしまう。それに対しバイサルファイト変換反応後にアダプターを付加するのが PBAT 法の特徴である。また、アダプタープライマーがビオチン化されており、ストレプトアビジンとの高親和性を利用し、アダプター付加した DNA のみを効率よく回収できるよう工夫されている。このような理由により、PBAT 法に必要なゲノム DNA 量は少量であり、ヒトサンプルの場合は、100 ng のゲノム DNA を用いれば、PCR 増幅無しで十分量のライブラリーを得ることが可能である (WGBS では PCR 増幅は不可欠)。

## 次世代シーケンサーデータ解析系の構築：

### 自動化・効率化および新規解析系導入

#### ①リーシケンシング解析

配列データ解析に利用可能な既存ソフトウェアは多数あるが、大学等研究機関により開発される場合、たいていはオープンソースなフリーソフトウェアであり、他の研究グループによる利用や検証を容易にしている。本研究では、他の研究グループや配列解析受託企業などが使用しているソフトウェアを調査し、フリーでありかつ広く用いられ、十分な検証が行われているソフトウェアを選び出し、B シェルスクリプトでパイプライン (多数の既存あるいは新規プログラムを連結してデータをシームレスに自動処理するソフトウェア) を構築した。その際、必要に応じて最低限の処理を行う Perl または C による独自プログラムも作成して組み込んだ。用いたソフトウェア名およびバージョンは別にまとめた (図 21)。主なプラットフォームとしては CentOS 6 を用いたが、Fedora 16、Red Hat Enterprise Linux 4、OS X Mountain Lion、さらには Cygwin 1.7 においても動作確認を行っている。

本パイプラインは Linux 等、UNIX ライクな環境の CUI で動作する。コマンド入力に馴染みのないユーザでも利用し易いように、特に設定変更の必要がなければコマンド 1 つで動作するように設計した。本研究ではペアド・エンド・データのみを扱った。

適当な場所にプロジェクト名を付けたディレクトリを作成し、そこに CASAVA から出力された GNU ZIP 形式で圧縮された一連の FASTQ ファイルを保存する。必要に応じて B シェルスクリプト内にあらかじめ書き込まれている変数の値を書き換えることができるが、実行するとまず圧縮ファイルを解凍し、両端を示す R1 および R2 ごとに順番に結合する。

FastQC により品質確認を行い、処理はそのまま次に進むが、ユーザはここで得られた配列の全体的なクオリティを視覚的に確認することができる。独自プログラムにより、FASTQ 形式および R1 と R2 の対応確認も行い、データのコピーが不完全であった等、問題があった場合は処理を止める。配列は、インサート長が短い場合にはアダプター配列を含むことがあり、また両端のクオリティの低い塩基がマッピングの妨げになることがある。これらは Trimmomatic により除去し、さらに RNA-seq でポリ A 由来の配列を除去したい場合は独自プログラムで対応した。不要となったサイズの大きな配列データは再圧縮し、BWA、Bowtie2、TopHat2 などを用いたマッピング処理へと移行する。

#### 1) エクソーム解析およびターゲット・リーシケンシング

トリミングされた配列データを、BWA を用いてリファレンス配列へアラインメントおよびマッピングし、SAM 形式のファイルを出力する。シングルヒットのみに限定したい場合は、独自プログラムを用いてこのファイルを書き換える。SAMtools を用いてその後の処理に便利な BAM 形式に変換し、PCR 増幅によって得られたと考えられる同一配列はこの段階で Picard MarkDuplicates によって除去し、最終的な BAM 形式のファイルが得られる。ヴァリアント検出は GenomeAnalysisTK (GATK) を中心に組み立てた。まず、濃縮を行ったベイト配列情報から解析する領域を限定する。ヒトゲノムへのマッピングの場合には、既知 indel 周辺の再アラインメントを行う。カバレッジ計算を行い、ヴァリアントは SNV および indel ごとに分けて検出、フィルタリングを行ってから統合する。最後に snpEff を実行してヴァリアントのアノテーションおよび効果予測を加えた VCF 形式のテキストファイルを出力する (図 22)。このパイプラインは、当研究所内で得られた 50 検体以上の全エクソームエクソーム解析データに対しての解析実績がある。

#### 2. 全ゲノム・リーシケンシング

ベイト配列情報を用いて解析領域を限定しないこと、またデータ量の違いはあるが、エクソーム解析とほぼ同じパイプラインで全ゲノム・リーシケンシングも可能となっている。マウスリファレンス配列とマウス亜種系統との比較では全ゲノムに渡って平均 150 bp に 1 ヶ所程度の多型検出に成功している (図 23)



に例を示す)。

## ②トランスクリプトームデータ解析

### i) 二次解析 (マッピング)

RNA-seq ライブラリーのシーケンスデータ (fastq ファイル) についてリードトリミングを行った後、TopHat2.0 および Bowtie2 によるアライメントを行い、BAM ファイルを作成した。リファレンスとしてトランスクリプトームおよびゲノムの両データを用意し、TopHat2 および Bowtie2 を用いて、ゲノムおよびトランスクリプトームへのマッピングを行った。詳細は割愛するが、これらの工程をパイプライン化した。

### ii) Avadis NGS を用いた三次解析

前述のように作成した BAM ファイルを Avadis NGS ソフトウェアに読み込み、一連のデータ解析 (複数サンプル間での遺伝子発現量比較、スプライスアイソフォーム毎の発現量比較、新規愛想フォームの検出、SNP 情報を用いたアレル別発現解析、gene ontology 解析、パスウェイ解析などを実施した)。

## ③ChIP-Seq データ解析

特定の DNA 結合タンパクのゲノム標的部位の同定やヒストン修飾のゲノムワイド解析を実施する体制を整備すべく、ChIP-Seq データ解析系の構築を目指した。

ChIP-seq ライブラリーの配列データを BWA あるいは Bowtie2 でマッピングし、BAM ファイルを作成した。BAM ファイルを Avadis NGS に読み込み、MACS, PICS を含む 4 種類のピーク検出アルゴリズムをテストした。

## ④DNA メチル化解析

バイサルファイト処理後にアダプター付加を行う PBAT 法で得られた配列データを後述の手順で処理する解析パイプラインを構築し、ヒトゲノムおよびマウスゲノムのリファレンスに含まれる全 CpG サイトの DNA メチル化率を算出できる体制を整備した。トリミング後の配列データは Bmap (東大・三浦らが開発) により、配列情報が変換された特別なリファレンス配列 (バイサルファイト変換配列) にマップする。その結果を利用して、CpG サイトと対になっている相補的な CpG サイトのメチル化状態は同一であるとの仮定のもとにメチル化、非メチル化分子数を集計する。マウス ES 細胞について PBAT ライブラリーを作製 HiSeq1000 で 80 Gb の配列データを取得し、ゲノムワイドにメチル化率を算出した結果例を図 20 に示した。

## D. 考察

### 次世代シーケンサーデータ解析系の構築：

エクソーム解析、リシークエンシング、RNA-seq 等の解析手法を綿密に調査し、それぞれに対し独自のパイプラインを構築した。当初は外部委託が必要だったデータ解析工程を、研究所内でより短時間で実行することが可能となった。またこれらの調査が、RNA-seq によるアレル別発現解析や PBAT 法のデータ解析など、エピジェネティクスに関連する独自性の高い解析手法の確立にも役立った。Avadis-NGS のような Window OS 対応の GUI ソフトウェアを使ったデータ解析も、ソフトウェア機能と解析目的が合致した場合は極めて有効であった。

主として CentOS などの Linux で環境を構築しているため、ディスクフォーマットの互換性が一つの大きな問題となっていた。これまでは Linux、OS X、Windows で読み書き可能な FAT32 を使えば良かったが、4 ギガバイトを越えるファイルを扱うことができず、OS X や Windows で利用しにくい ext4 形式を利用せざるを得ない。この場合、パイプライン処理後に LAN を経由したデータのコピーが必要になり、ビッグデータとも呼ぶことができる次世代シーケンサーのデータの転送には非常に時間がかかり、作業効率化の大きな妨げとなっていた。これらの問題は、今年度中に研究所内ネットワークのデータ転送速度を改善したこと、ネットワーク上に大容量データ保管用サーバーを設置したことで大幅に改善された。

高精度でヴァリアントの検出が可能になると、エクソーム解析等のリシークエンシングではリファレンスにマッピングできないヴァリアントの検出が問題になってくる。当拠点では日本人ゲノムを扱うことが多く、全体的な傾向として日本人ゲノムがリファレンスに対しどの程度異なっているのかは医学的な観点から非常に興味深い点である。シーケンシング技術の向上 (スループット増大・コスト低下) に伴い、全ゲノムデータの取得がより現実的になること、ゲノム *de novo* アセンブリの需要が高まることが予想される。

## E. 結論

NGS ライブラリー作製実験系 (エクソーム濃縮ライブラリー・RNA-seq ライブラリー) については、自動化・効率化を図り、スループット

トの大幅向上と微量検体・FFPE由来検体からのライブラリー作製に成功した。また、ゲノム構造解析、エピゲノム解析を指向した複数のNGSライブラリー作製プロトコルを確立した。配列データ量増大に伴いデータ解析の迅速化の必要性が高まる中、リシーケンシング、RNA-seqデータ解析パイプラインを独自に構築した。さらに、エピゲノム研究に向けたアレル別発現解析、DNAメチル化解析パイプラインも構築した。

次世代シーケンサー・アレイを用いた迅速遺伝子診断システムにより、これまでに様々な遺伝性疾患を含む合計2,200症例の解析を行い、そのうち特に臨床診断を目的とした約600症例の解析では、23%程度が確定診断に至った。これまで原因未知とされた疾患の原因遺伝子変異同定と併せ、今後の成育疾患解析に必須の正常妊婦コントロールデータの整備も進め、本邦成育疾患の遺伝学的解析の実用化均てん化に資する基盤的知見を得ると共に、解析体制を構築した。

## F. 研究発表

### 1. 論文発表

#### 原著論文

Nakamura K, Aizawa K, Nakabayashi K, Kato N, Yamauchi J, Hata K, Tanoue A : DNA methyltransferase inhibitor zebularine inhibits human hepatic carcinoma cells proliferation and induces apoptosis. *PLoS One*. 2013;8(1):e54036.

Hiraoka D, Yoshida W, Abe K, Wakeda H, Hata K, Ikebukuro K. *Anal Chem* : Development of a Method To Measure DNA Methylation Levels by Using Methyl CpG-Binding Protein and Luciferase-Fused Zinc Finger Protein. in press 2012

Iglesias-Platas I, Martin-Trujillo A, Cirillo D, Court F, Guillaumet-Adkins A, Camprubi C, Bourc'his D, Hata K, Feil R, Tartaglia G, Arnaud P, Monk D : Characterization of novel paternal ncRNAs at the *Plagl1* locus, including *Hymai*, predicted to interact with regulators of active chromatin. *PLoS One*. 2012;7(6):e38907.

Higashimoto K, Nakabayashi K, Yatsuki H, Yoshinaga H, Jozaki K, Okada J, Watanabe Y, Aoki A, Shiozaki A, Saito S, Koide K, Mukai T, Hata K,

Soejima H : Aberrant methylation of H19-DMR acquired after implantation was dissimilar in soma versus placenta of patients with Beckwith-Wiedemann syndrome. *Am J Med Genet A*. 2012;158A(7):1670-1675.

Kobayashi H, Sakurai T, Sato S, Nakabayashi K, Hata K, Kono T : Imprinted DNA methylation reprogramming during early mouse embryogenesis at the *Gpr1-Zdbf2* locus is linked to long cis-intergenic transcription. *FEBS Lett*. 2012 23;586(6):827-833.

Nakanishi MO, Hayakawa K, Nakabayashi K, Hata K, Shiota K, Tanaka S : Trophoblast-specific DNA methylation occurs after the segregation of the trophectoderm and inner cell mass in the mouse periimplantation embryo. *Epigenetics*. 2012 Feb;7(2):173-182.

Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T : Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet*. 2012 ;8(1):e1002440.

## G. 知的財産権の出願・登録状況

### 1. 取得特許

なし

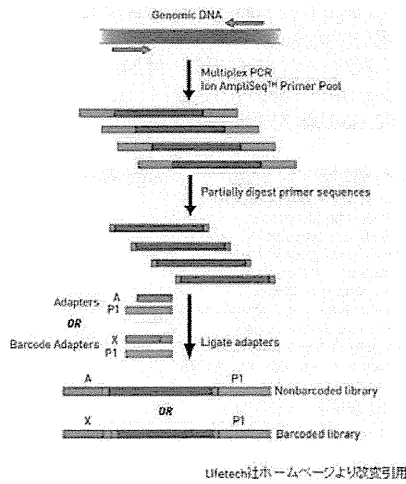
### 2. 実用新案登録

なし

### 3. その他

なし

## Ion-Ampliseq (LifeTech 社)

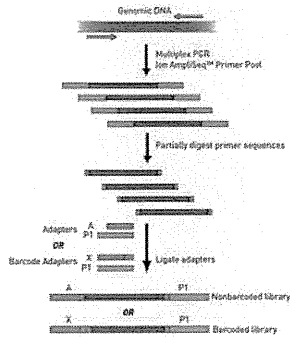


## Ion AmpliSeq Custom Oligos価格

アンプリコン数	ターゲットサイズ	価格
12	2.4 Kb	9,000 円
24	4.8 Kb	18,000 円
48	9.6 Kb	36,000 円
96	19.2 Kb	72,000 円
192	38.4 Kb	144,000 円
384	76.8 Kb	288,000 円
768	153.6 Kb	576,000 円
1536	307.2 Kb	1,152,000 円

2250 or 9000反応分  
750円 / アンプリコン

## Ion-Ampliseq (LifeTech 社)



## SCID原因遺伝子

**ADA**      **ORAI1**  
**AK2**      **PNP**  
**CD247**    **PRKDC**  
**CD3D**    **PTPRC**  
**CD3E**    **RAG1**  
**CD3G**    **RAG2**  
**CIITA**    **RFX5**  
**DCLRE1C** **RFXANK**  
**IKZF1**    **RFXAP**  
**IL2RG**    **STIM1**  
**IL7R**     **TAP1**  
**JAK3**     **TAP2**  
**LIG4**     **TAPBP**  
**NHEJ1**    **ZAP70**

(28 genes)

## カスタムマルチプレックスプライマーのデザイン: Ion AmpliSeq Designer Website <https://www.ampliseq.com/browse.action>

Gene symbolリストを入力すると2-48時間後に結果が出力される

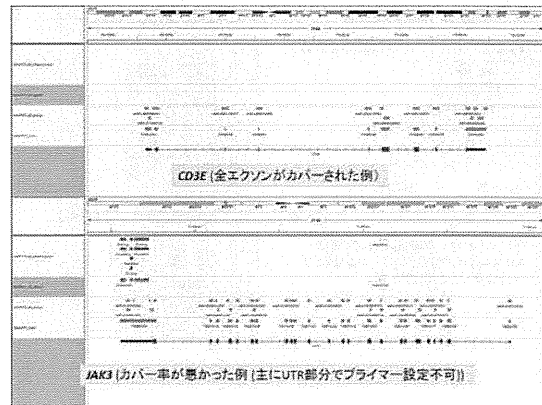
Target	Barcode	ID	Status	Gene Type	Target Size	Amplicons	Primer	Coverage (%)	Date	Primer Barcode
ADA	ADA001	ADA	Ready to use	Standard CDS	24	10	5'	98.10	2017-01-15	1
AK2	AK2001	AK2	Ready to use	Standard CDS	35	10	5'	98.10	2017-01-15	2
CD247	CD247001	CD247	Ready to use	Standard CDS	9	10	5'	98.10	2017-01-15	3
CD3D	CD3D001	CD3D	Ready to use	Standard CDS	35	10	5'	98.10	2017-01-15	4
CD3E	CD3E001	CD3E	Ready to use	Standard CDS	1	10	5'	98.10	2017-01-15	5
CD3G	CD3G001	CD3G	Ready to use	Standard CDS	24	10	5'	98.10	2017-01-15	6
CIITA	CIITA001	CIITA	Ready to use	Standard CDS	31	10	5'	98.10	2017-01-15	7

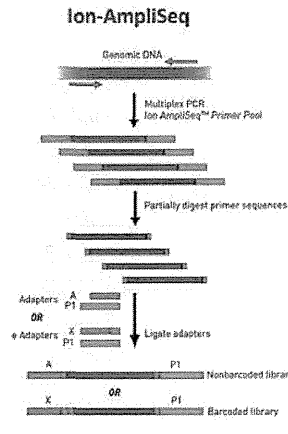
## カスタムマルチプレックスプライマーのデザイン: Ion AmpliSeq Designer Website <https://www.ampliseq.com/login/login.action>

Gene symbolリストを入力すると2-48時間後に結果が出力される

Target	Barcode	ID	Status	Gene Type	Target Size	Amplicons	Primer	Coverage (%)	Date	Primer Barcode
ADA	ADA001	ADA	Ready to use	Standard CDS	24	10	5'	98.10	2017-01-15	1
AK2	AK2001	AK2	Ready to use	Standard CDS	35	10	5'	98.10	2017-01-15	2
CD247	CD247001	CD247	Ready to use	Standard CDS	9	10	5'	98.10	2017-01-15	3
CD3D	CD3D001	CD3D	Ready to use	Standard CDS	35	10	5'	98.10	2017-01-15	4
CD3E	CD3E001	CD3E	Ready to use	Standard CDS	1	10	5'	98.10	2017-01-15	5
CD3G	CD3G001	CD3G	Ready to use	Standard CDS	24	10	5'	98.10	2017-01-15	6
CIITA	CIITA001	CIITA	Ready to use	Standard CDS	31	10	5'	98.10	2017-01-15	7

**SCID 28遺伝子 CDS**      target size    64.8 kb [537 amplicons]    coverage 98.56%  
**SCID 28遺伝子 CDS & UTR**    target size    112.1 kb [753 amplicons]    coverage 88.26%





1. 実験概要  
2. 実験手順  
3. 実験結果

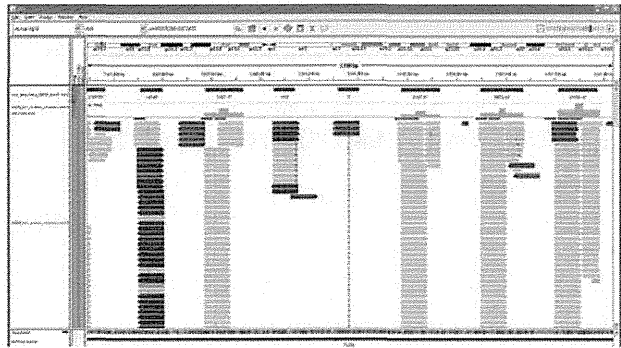
通常プロトコール イルミナアダプター付加

- 1検体について
- ↓ Pool 1 oligo (462 primer pairs, total 50kb) を用いて multiplex PCR
- ↓ ライブラリー化 (アダプター付加)
- ↓ HiSeq paired-end (100 bp x2) readで 400 Mbの配列を取得 (⇒ x8000)

↓ 標的領域50Kbについて on-target率を計算

% of reads within regions: 91.72  
 % of reads within 100bp: 91.83  
 % of reads within 200bp: 91.83  
 % of reads within 500bp: 91.83  
 % of reads within 1000bp: 91.83  
 % of reads beyond 1000bp: 8.17  
 Average coverage over all target regions: 5291.61  
 Total number of reads in sample overlapping with targeted regions: 3781990  
 Total number of reads in sample: 4123270

on-target率 91.72%  
 coverage 5291.6



- 1検体について
- ↓ Pool 1 oligo (462 primer pairs, total 50kb) を用いて multiplex PCR
- ↓ ライブラリー化 (アダプター付加)
- ↓ HiSeq paired-end (100 bp x2) readで 400 Mbの配列を取得 (⇒ x8000)

↓ 標的領域50Kbについて on-target率を計算

% of reads within regions: 91.72  
 % of reads within 100bp: 91.83  
 % of reads within 200bp: 91.83  
 % of reads within 500bp: 91.83  
 % of reads within 1000bp: 91.83  
 % of reads beyond 1000bp: 8.17  
 Average coverage over all target regions: 5291.61  
 Total number of reads in sample overlapping with targeted regions: 3781990  
 Total number of reads in sample: 4123270

on-target率 91.72%  
 平均coverage 5291.6

Coverage worst top 40

Rank	Region	Coverage	Count
1	chr1:115024211-115024272	154	210
2	chr2:30300000-30300042	154	210
3	chr17:69582000-69582042	154	210
4	chr17:69582000-69582042	154	210
5	chr15:45449000-45449042	154	210
6	chr15:45449000-45449042	154	210
7	chr11:83000000-83000042	154	210
8	chr10:44400000-44400042	154	210
9	chr10:44400000-44400042	154	210
10	chr22:42400000-42400042	154	210
11	chr22:42400000-42400042	154	210
12	chr14:30000000-30000042	154	210
13	chr14:30000000-30000042	154	210
14	chr10:10000000-10000042	154	210
15	chr10:10000000-10000042	154	210
16	chr10:10000000-10000042	154	210
17	chr10:10000000-10000042	154	210
18	chr10:10000000-10000042	154	210
19	chr10:10000000-10000042	154	210
20	chr10:10000000-10000042	154	210
21	chr10:10000000-10000042	154	210
22	chr10:10000000-10000042	154	210
23	chr10:10000000-10000042	154	210
24	chr10:10000000-10000042	154	210
25	chr10:10000000-10000042	154	210
26	chr10:10000000-10000042	154	210
27	chr10:10000000-10000042	154	210
28	chr10:10000000-10000042	154	210
29	chr10:10000000-10000042	154	210
30	chr10:10000000-10000042	154	210
31	chr10:10000000-10000042	154	210
32	chr10:10000000-10000042	154	210
33	chr10:10000000-10000042	154	210
34	chr10:10000000-10000042	154	210
35	chr10:10000000-10000042	154	210
36	chr10:10000000-10000042	154	210
37	chr10:10000000-10000042	154	210
38	chr10:10000000-10000042	154	210
39	chr10:10000000-10000042	154	210
40	chr10:10000000-10000042	154	210

- 1検体について
  - ↓ Pool 1 oligo (462 primer pairs, total 50kb) を用いて multiplex PCR
  - ↓ ライブラリー化 (アダプター付加)
  - ↓ HiSeq paired-end (100 bp x2) readで 400 Mbの配列を取得 (⇒ x8000)
- ↓ 標的領域50Kbについて on-target率を計算
- % of reads within regions: 91.72  
 % of reads within 100bp: 91.83  
 % of reads within 200bp: 91.83  
 % of reads within 500bp: 91.83  
 % of reads within 1000bp: 91.83  
 % of reads beyond 1000bp: 8.17  
 Average coverage over all target regions: 5291.61  
 Total number of reads in sample overlapping with targeted regions: 3781990  
 Total number of reads in sample: 4123270
- on-target率 91.72%  
 平均coverage 5291.6

Coverage worst top 40

Rank	Region	Coverage	Count
1	chr1:115024211-115024272	154	210
2	chr2:30300000-30300042	154	210
3	chr17:69582000-69582042	154	210
4	chr17:69582000-69582042	154	210
5	chr15:45449000-45449042	154	210
6	chr15:45449000-45449042	154	210
7	chr11:83000000-83000042	154	210
8	chr10:44400000-44400042	154	210
9	chr10:44400000-44400042	154	210
10	chr22:42400000-42400042	154	210
11	chr22:42400000-42400042	154	210
12	chr14:30000000-30000042	154	210
13	chr14:30000000-30000042	154	210
14	chr10:10000000-10000042	154	210
15	chr10:10000000-10000042	154	210
16	chr10:10000000-10000042	154	210
17	chr10:10000000-10000042	154	210
18	chr10:10000000-10000042	154	210
19	chr10:10000000-10000042	154	210
20	chr10:10000000-10000042	154	210
21	chr10:10000000-10000042	154	210
22	chr10:10000000-10000042	154	210
23	chr10:10000000-10000042	154	210
24	chr10:10000000-10000042	154	210
25	chr10:10000000-10000042	154	210
26	chr10:10000000-10000042	154	210
27	chr10:10000000-10000042	154	210
28	chr10:10000000-10000042	154	210
29	chr10:10000000-10000042	154	210
30	chr10:10000000-10000042	154	210
31	chr10:10000000-10000042	154	210
32	chr10:10000000-10000042	154	210
33	chr10:10000000-10000042	154	210
34	chr10:10000000-10000042	154	210
35	chr10:10000000-10000042	154	210
36	chr10:10000000-10000042	154	210
37	chr10:10000000-10000042	154	210
38	chr10:10000000-10000042	154	210
39	chr10:10000000-10000042	154	210
40	chr10:10000000-10000042	154	210

- 1検体について
  - ↓ Pool 1 oligo (462 primer pairs, total 50kb) を用いて multiplex PCR
  - ↓ ライブラリー化 (アダプター付加)
  - ↓ HiSeq paired-end (100 bp x2) readで 400 Mbの配列を取得 (⇒ x8000)
- ↓ 標的領域50Kbについて on-target率を計算
- % of reads within regions: 91.72  
 % of reads within 100bp: 91.83  
 % of reads within 200bp: 91.83  
 % of reads within 500bp: 91.83  
 % of reads within 1000bp: 91.83  
 % of reads beyond 1000bp: 8.17  
 Average coverage over all target regions: 5291.61  
 Total number of reads in sample overlapping with targeted regions: 3781990  
 Total number of reads in sample: 4123270
- on-target率 91.72%  
 平均coverage 5291.6

Coverage worst top 40

Rank	Region	Coverage	Count
1	chr1:115024211-115024272	154	210
2	chr2:30300000-30300042	154	210
3	chr17:69582000-69582042	154	210
4	chr17:69582000-69582042	154	210
5	chr15:45449000-45449042	154	210
6	chr15:45449000-45449042	154	210
7	chr11:83000000-83000042	154	210
8	chr10:44400000-44400042	154	210
9	chr10:44400000-44400042	154	210
10	chr22:42400000-42400042	154	210
11	chr22:42400000-42400042	154	210
12	chr14:30000000-30000042	154	210
13	chr14:30000000-30000042	154	210
14	chr10:10000000-10000042	154	210
15	chr10:10000000-10000042	154	210
16	chr10:10000000-10000042	154	210
17	chr10:10000000-10000042	154	210
18	chr10:10000000-10000042	154	210
19	chr10:10000000-10000042	154	210
20	chr10:10000000-10000042	154	210
21	chr10:10000000-10000042	154	210
22	chr10:10000000-10000042	154	210
23	chr10:10000000-10000042	154	210
24	chr10:10000000-10000042	154	210
25	chr10:10000000-10000042	154	210
26	chr10:10000000-10000042	154	210
27	chr10:10000000-10000042	154	210
28	chr10:10000000-10000042	154	210
29	chr10:10000000-10000042	154	210
30	chr10:10000000-10000042	154	210
31	chr10:10000000-10000042	154	210
32	chr10:10000000-10000042	154	210
33	chr10:10000000-10000042	154	210
34	chr10:10000000-10000042	154	210
35	chr10:10000000-10000042	154	210
36	chr10:10000000-10000042	154	210
37	chr10:10000000-10000042	154	210
38	chr10:10000000-10000042	154	210
39	chr10:10000000-10000042	154	210
40	chr10:10000000-10000042	154	210

平均Coverage >20を基準とした場合:  
 405/426 (95.3%)のアンプリコン領域で  
 十分なデータが取得できた。

平均Coverage >20を基準とした場合:  
 405/426 (95.3%)のアンプリコン領域で  
 十分なデータが取得できた。

幾つかの小規模変異解析プロジェクトを  
 試し、評価予定