

201238003A

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(難病関係研究分野)

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24年度 総括研究報告書

研究代表者 梅澤 明弘

平成25(2013)年4月

厚生労働科学研究費補助金  
再生医療実用化研究事業

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24年度 総括研究報告書

研究代表者 梅澤 明弘

平成25（2013）年4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. 日本人ゲノムのプロファイル作成 .....	31
秦 健一郎	
2. 次世代シーケンサーを用いた 先天性サイトメガロウイルス感染症の病態解析 .....	39
中村 浩幸	
3. 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 .....	44
服部 正平	
III. 研究成果の刊行に関する一覧表 .....	47
VI. 研究成果の刊行物・別刷り .....	51

# I. 総括研究報告書

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）  
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
研究分担者：深見 真紀	独立行政法人国立成育医療研究センター・部長
服部 正平	東京大学・教授
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
秦 健一郎	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・医長
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・医長
堀川 玲子	独立行政法人国立成育医療研究センター・医長
大矢 幸弘	独立行政法人国立成育医療研究センター・医長

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因子を解明する事を目的とする。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 1,450 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 500 症例の解析では、25%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)

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  
(インフォマティックス)

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)

#### 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図る。

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 法の導入を試みる。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。

#### ②長鎖ゲノム DNA ライブラリーを簡便作製プロトコルの検討

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法により微量ゲノム 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する条件の検討を行う。

#### 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試みる。

#### 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一形態であり、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する。

### 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、データ供出体制と基本的なデータ解析体制を整えた。

昨年度まで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。大量データを内部で処理する体制を強化すべく、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、ヴァリエーション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ことを今年度の目標とした。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。

(添付資料 1)

## 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についての解析状況を表 1 (次世代シーケンサー解析)、表 2 (アレイ解析) にまとめた。これまでに様々な遺伝性疾患を含む合計 1,450 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 500 症例の解析では、25% 程度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 232 検体のライブラリーを作製し、いずれも適切な平均インサート鎖長のライブラリーが出来ていることを確認した。結果例を図 2 に示した。

##### 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 56 検体のライブラリーを作製した。作製したすべてのライブラリーにおいて、従来の手動操作で作製した場合と同様のライブラリーができることを確認した。結果例を図 4 に示した。また、ライブラリー増幅に必要な PCR サイクル数を検討し、イルミナ社推奨は 15 サイクルであるのに対し、9 サイクルで十分量のライブラリーを増幅できることを確認した (図 5)。

3) HaloPlex ターゲットエンリッチメントシステム (図 6) はカスタムエクソーム濃縮解析のための実験プロトコールであり、ヒトゲノムにおける任意のゲノム領域 (合計<500Kb) をキ

ャプチャー対象として選ぶことができる。225 ngのゲノムDNAを出発材料としてライブラリー作製を行った。制限酵素によるゲノムDNAの断片化した後、設計したプローブとハイブリダイゼーションし、ターゲットDNAの捕捉をした。ライブラリー作製後、Bioanalyzer High Sensitivity KitにてDNA断片が225~525 bpに分布していることを確認した。アジレント自動化システムによりさらに効率的にライブラリーを作製する体制を整えた。

### ②長鎖ゲノムDNAライブラリーを簡便作製プロトコルの検討

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る (タグメンテーションと呼ばれる、図7)。従来法に比べて所用時間が短く (従来法6時間に対し、90分)、必要とするゲノムDNA量が僅かであることがこの方法の長所である。従来よりも長いインサート鎖長が得られるようなトランスポゾン活性酵素量を検討した。酵素量を通常の30%および10%に減らした反応の結果を図8に示した。酵素量100%ではインサート鎖長平均は200bp程度であるのに対し、30%量では200~1000bpに10%量では1400~2200bpにピークを観察した。酵素量10%でタグメンテーション反応を行い、PCR増幅(8サイクル)後、Pippin Prep (sage science社) 1.5%カセットにて1000~1500bpを切り出すことで、目的鎖長を有するライブラリーを作製した (図9)。

### ③メイトペアライブラリー作製プロトコルの検討

Mate-Paired Library Preparation 5500SOLiD System キット (Life Technologies, 図10) によりメイトペアライブラリーを作製した。5 ugのゲノムDNAをHydroShearにて約2 kbに切断した断片を実験に供した。ライブラリー作製工程を図9に示した。アガロースゲル電気泳動にてサイズセレクションを行った後の収量は0.843 ug、アダプター付加後は0.725ug、DNA環状化後は42 ugであった。Trial PCRの結果より、PCRのサイクル数を12とした。Bioanalyzerで波形を確認したところ、ややサイズの分布が広がったため、さらにサイズセレクトを行い、

最終的に平均鎖長262 bpのライブラリーを得た (図11)。

### 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

FFPE由来gDNAを出発材料として、1) Nextera DNA Preparation kit, 2) TruSeq DNA Preparation kitによるライブラリー作製条件を検討した。FFPE由来total RNAを出発材料として、3) TruSeq RNA Preparation Kit, 4) Ovation RNA-Seq FFPE Systemによるライブラリー作製条件を検討した。

Nextera DNA Preparation kitを用いて、FFPE由来ゲノムDNA 10 ngおよびコントロールゲノムDNA 10 ngからライブラリーを作製した。ライブラリー増幅に必要なPCRサイクル数は、コントロールゲノムDNAでは7サイクルであったのに対し、FFPE由来ゲノムDNAは18サイクルを要した。インサート鎖長はどちらも約200 bpであった (図12)。

TruSeq DNA Preparation kitを用いて、FFPE由来ゲノムDNA 200 ng, コントロールゲノムDNA 300 ngおよび30 ngからライブラリーを作製した。FFPE由来ゲノムDNAは既に断片化されていたので、断片化処理は省略した。コントロールゲノムDNAについては30 ngを出発材料とした場合でもPCR x5サイクルで増幅が認められたのに対し、FFPE由来ゲノムDNAではPCRサイクル数をx24サイクルまで増やしてもライブラリーは増幅されなかった (図13)。

TruSeq RNA Preparation kitを用いて、FFPE検体由来total RNA 1 ugを出発材料としてライブラリー作製を試みた。Bioanalyzer RNA nano kitでFFPE由来RNAの波形を確認したところ、50~1000 bpにピークが見られ、一部は4000 bp付近までの幅広い分布を示していた。このようにFFPE由来RNAは既に断片化されているため、断片化工程を1分と短くした (または省略した)。PCRサイクル数も検討したが増幅が見られず、ライブラリーの作製は困難であった (図14)。

Ovation RNA-Seq FFPE System (NuGEN, 図15) を用いてFFPE由来total RNAからcDNAを合成し、さらにTruSeq RNA Preparation Kit (illumina)にて平滑末端化以降の操作を行い、ライブラリー作製を行った (図16)。FFPE由来total RNA 185 ng、およびコントロールとした血液由来total RNA 180 ngを出発材料としてcDNA合成を行った。cDNA収量は、FFPE由



来では 1.5 ug、コントロールでは 1.3 ug であった。Bioanalyzer で波形を確認したところ、コントロールとした血液由来 cDNA では 100~1500 bp までの幅広い分布が見られたが、FFPE 由来 cDNA では 50~250 bp に分布し、150 bp 付近にピークがあるシャープな波形を示した。最終的に作製したライブラリーのインサート鎖長もこの cDNA サイズを反映して、コントロールは 200~1000 bp に分布、FFPE 由来サンプルでは 200~500 bp の分布と小さかったが、どちらも 12 サイクルの PCR で十分な増幅が見られた。

## 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 (図 17)。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。ライブラリー作製結果を図 18 に示した。

### 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 (図 19)。

### 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 (図 20)。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)。実際にマウス ES 細胞由来ゲノム DNA 100 ng を出発材料として作製した PBAT ライブラリー収量を図 20 に示した。

## 2. 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・リシーケンシング

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・リシーケンシング

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・リシー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 によるアラインメントを行い、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)。解析結果例を図 24・図 25 に示した。

#### 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

### 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自動化プロトコール評価

アジレント社 NGS ライブラリー作製自動化装置の導入により、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の多くを自動化することで、ライブラリー作製効率が大幅に向上した (一度の実験で 24-96 検体についてライブラリー作製が可能)。ライブラリー作製工程においては、インサート鎖長・濃度測定を行う品質チェックが必須だが、LabChipGX (Caliper 社) の導入により、この工程の効率も向上した。

エクソーム濃縮前のゲノムライブラリーの収量は手動作製時と比較しても全く遜色無かったが、自動化プロトコールではエクソーム濃縮後の収量がやや劣る傾向が見られ、その後の工程に進むためにライブラリー濃縮 (高濃度化) が必要な場合に度々遭遇した。そこで、エクソーム濃縮後の PCR サイクル数を 1~2 サイクル増やすことで、手動作製時と同等の収量を得た。

TruSeq RNA Sample Preparation 法の自動化プロトコールはインサートサイズ・収量とも手動作製時と遜色無かった。

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

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のはこのキットは 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。今後、このライブラリーについてシーケンズデータを取得・解析することで、更なる評価を加える予定である。

#### ⑤エピゲノム解析用 NGS ライブラリー作製プロトコールの構築

NEBNext ChIP-Seq Library Prep kit (NEB) のプロトコールでは、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ることを確認した。

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。PBAT データをモデルに構築したデータ解析が RRBS データにも使用可能である。

## 2. 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、リシーケンシング、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。それにより、昨年度まで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。

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り、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。

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り、今後も併用していく予定である。

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になっている。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る。USB 3.0 の導入を進めているが、今後はギガビット・イーサネット等も必要になると考えている。

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ケ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

昨年度に確立した 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。ゲノム解析拠点として、今後もますます増大すると予想されるデータ量に対応すべく、ハードウェアの強化も含め、解析インフラの整備、開発を進める。

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 1,450 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 500 症例の解析では、25%程度が確定診断に至った。

## F. 研究発表

### 1. 論文発表

#### 原著論文

Higuchi A, Ling QD, Hsu ST, Umezawa A. Biomimetic cell culture proteins as extracellular matrices for stem cell differentiation. *Chem Rev.* 2012 Aug 8;112(8):4507-4540.

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.

#### 日本語総説

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

服部正平: 第15回プロバイオティクス学会''  
'Genomics of Human Microbiome Using  
Next-Generation Sequencers' (招待講演) (平成24  
年6月)

服部正平: 第16回腸内細菌学会 'ヒト常在細菌  
叢研究の新しい道標' (招待講演) (平成24年6  
月)

服部正平: 第12回日本抗加齢医学会 '日本人腸  
内マイクロバイオームの特徴' (招待講演) (平  
成24年6月)

服部正平: 日本遺伝学会第84回大会公開市民講  
座 '人間の体内に共生する細菌の網羅的なゲノ  
ム解析から細菌とヒトの共生を読み解く' (招  
待講演) (平成24年9月)

服部正平: 第44回小児感染症学会 'ヒトマイク  
ロバイオーム研究の現状と展望' (招待講演)  
(平成24年11月)

服部正平: 第9回日本消化管学会 'ヒトマイク  
ロバイオームの全体像を読み解く' (招待講演)  
(平成25年1月)

服部正平: 第46回日本無菌生物ノートバイオロ  
ジー学会 'ヒトマイクロバイオームのメタゲノ  
ム科学' (招待講演) (平成25年1月)

#### (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

服部正平他: Cold Spring Harbor Meeting: The  
Biology of Genomes 'Metagenomics of Japanese  
gut microbiomes' (ポスター発表) (平成24年5  
月)

ヒト人工多能性幹細胞を用いた神経幹細胞へ  
の実験的 HCMV 感染系の確立. 中村浩幸, 廖  
華南, 南佳ほり, 阿久津英憲, 梅澤明弘, 井  
上直樹, 藤原成悦. 第60回日本ウイルス学会  
学術集会 (2012年11月14日, 大阪)

## 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

厚生労働省科学研究費補助金 難治性疾患克服研究事業 「本邦における反復胎状奇胎症例の実態把握と確定診断法の開発に関する研究」平成 22-23 年度総合研究報告書 研究代表

厚生労働省科学研究費補助金 成育疾患克服等次世代育成基盤研究事業 「不育症治療に関する再評価と新たなる治療法の開発に関する研究（主任研究者 齋藤滋）」平成 20-22 年度総合研究報告書 分担研究

厚生労働省科学研究費補助金 子ども家庭総合研究事業 「生殖補助医療の医療技術の標準化、安全性の確保と生殖補助医療により生まれた児の長期予後の検証に関する研究（主任研究者 吉村泰典）」平成 19-24 年度総合研究報告書 分担研究

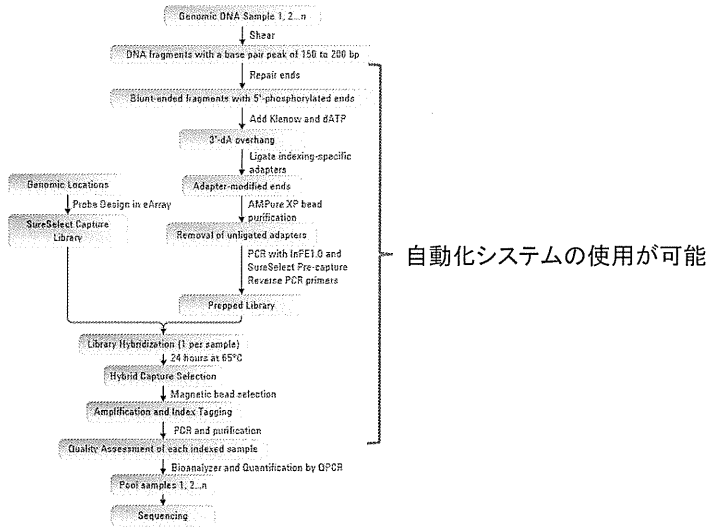


図1: エクソーム濃縮ライブラリー作成ワークフロー(SureSelect)

No.	発測値(ng/uL)	原液(ng/uL)	総量(30μL) ng	for 750ng
1	3.76	37.61	1128.4	19.94
2	11.32	113.22	3396.7	6.62
3	5.13	51.25	1537.6	14.63
4	11.53	115.27	3458.0	6.51
5	5.14	51.35	1540.6	14.60
6	5.65	56.52	1695.5	13.27
7	5.09	50.85	1525.5	14.75
8	3.82	38.17	1145.0	19.65
9	8.03	80.31	2409.4	9.34
10	3.82	38.17	1145.1	19.65
11	3.91	39.13	1174.0	19.16
12	3.74	37.39	1121.6	20.06
13	5.49	54.93	1648.0	13.65
14	7.92	79.20	2376.1	9.47
15	7.07	70.74	2122.1	10.60
16	4.60	46.01	1380.4	16.30
17	4.68	46.79	1403.8	16.03
18	8.67	86.72	2601.6	8.65
19	9.66	96.65	2896.6	7.77
20	4.53	45.33	1359.8	16.55
21	4.21	42.12	1263.5	17.81
22	6.55	65.53	1965.8	11.45
23	5.88	58.81	1764.3	12.75
24	4.66	46.64	1399.1	16.08
25	6.62	66.25	1987.5	11.32
26	14.37	143.71	4311.3	5.22
27	10.14	101.39	3041.7	7.40
28	3.88	38.83	1165.0	19.31
29	6.55	65.46	1963.9	11.46
30	6.92	69.16	2074.8	10.84
31	9.12	91.20	2736.1	8.22
32	7.36	73.57	2207.2	10.19

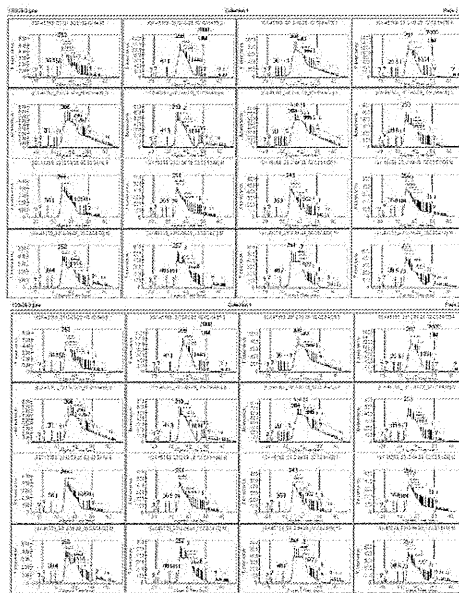


図2:エクソーム濃縮ライブラリーの作製(SureSelect)

A. Exome濃縮前、PCR 5 cycles後のサイズチェックと定量結果(LabChipGX)



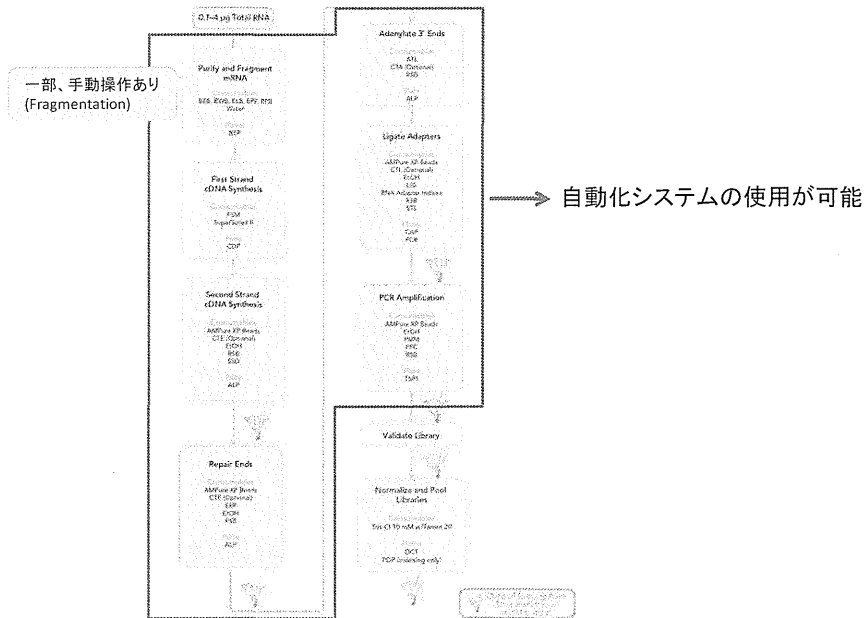


図3: イルミナ TruSeq RNAライブラリー作成ワークフロー

sample	From (bp)	To (bp)	Area	% of Total	Average Size (bp)	Size distribution in CV (%)	Conc. (ng/ul)	Molarity (nmol/l)	総量(30µL) ng
1	200	1209	1072.9	94	397	38.4	39.00	170.7	1170.0
2	210	1145	766.3	82	404	38.4	32.45	139.4	973.5

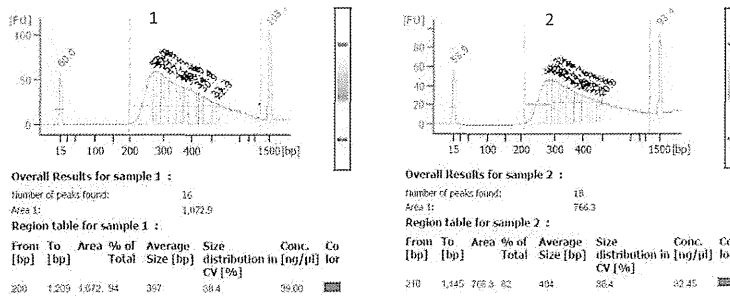


図4: TruSeq RNAライブラリーの作製結果

Fragmentation time 3 min, PCR 15 cyclesで作製したライブラリーをBioanalyzer DNA1000で測定

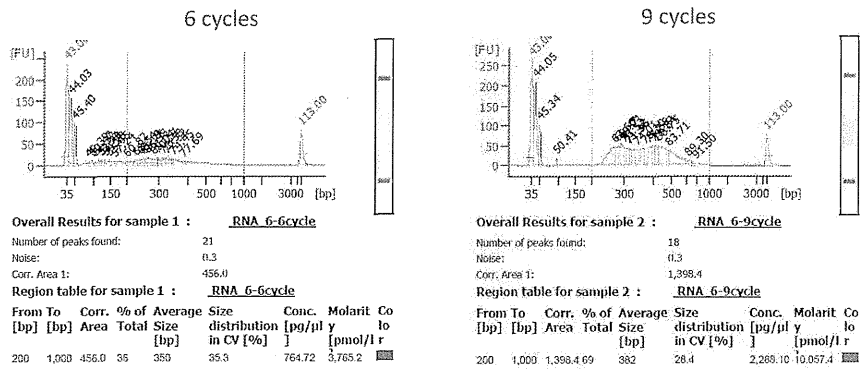


図5:TruSeq RNAライブラリー PCR cycle検討結果  
 Bioanalyzer HighSensitivityで測定

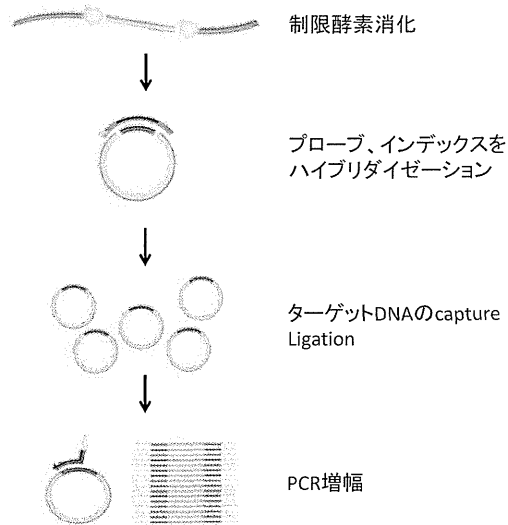


図6:HaloPlex ワークフロー

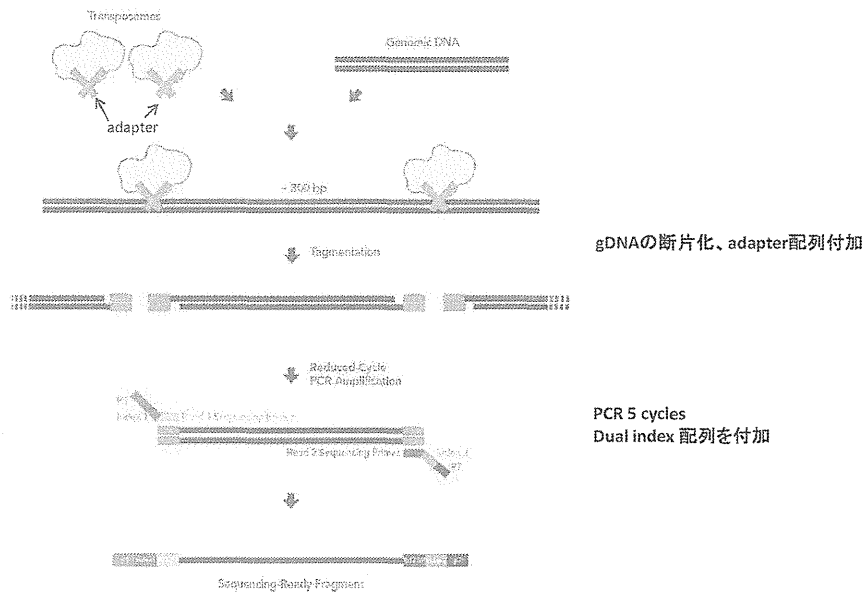


図7:Nextera ライブラリー作製ワークフロー

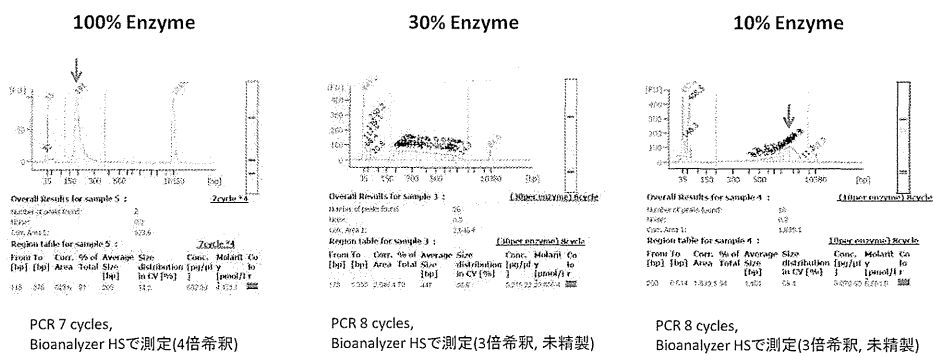


図8:Nextera DNA ライブラリー作製結果(1)

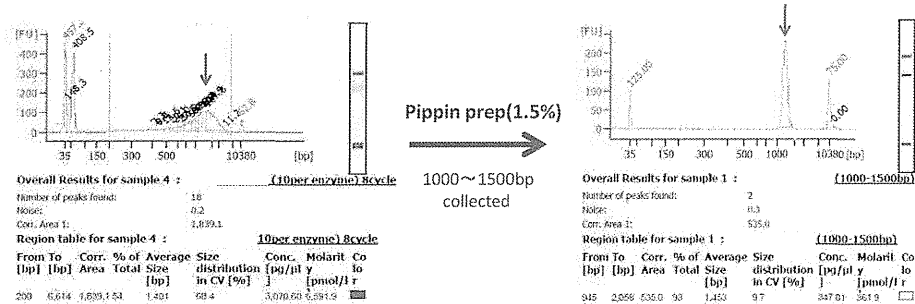


図9:Nextera DNA ライブラリー作製結果(2)

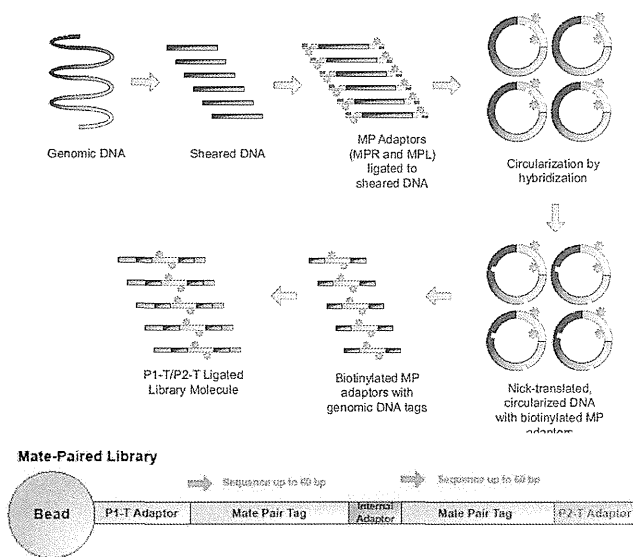


図10: Mate pair ライブラリー作製ワークフロー

5500 SOLiD™ Mate-Paired Library Preparation Kit