

ゲノム変異と疾病の連関を解析するバイオインフォマティクス 技術の開発

研究分担者 氏名 高木利久 所属 東京大学

研究要旨：本研究は、次世代シーケンサー等の大規模配列解析装置から生産されるマイクロバイオームを含めたゲノム配列について、ゲノム変異や疾病の連関を情報学的に解析するためのバイオインフォマティクス技術を開発する。

A. 研究目的

本研究は、次世代シーケンサーによって生産される種々の小児科・産科領域の難治性疾患や稀少疾患の患者のエキソームや腸内細菌叢等のマイクロバイオームを対象に、その効率的で高速なバイオインフォマティクス手法を開発し、ゲノム変異やこれらと疾患との関連性を解明する。

B. 研究方法

本研究では、分担者である同一機関所属の服部が次世代シーケンサーから生産されるマイクロバイオームを含む大量のゲノム情報のバイオインフォマティクス解析手法の開発を行い、データ品質の管理、疾病とゲノム変異の相関や菌叢構造の相違等について、高速で精度の高い解析パイプラインの構築を行う。

（倫理面への配慮）

分担者である同一機関所属の服部の場合と全く同様な配慮となっている。すなわち、本研究で得られる個人の常在菌のゲノム／遺伝子情報は、個人に固有の情報であるため、法令等により保護される個人情報として取り扱う。そのため、被験者に対して文書と面接による十分な説明（研究の目的、被験者が特定できない匿名化、個人名が特定されない情報公開、同意の途中撤回または継続の中止の容認等）が行われ、インフォームドコンセントが得られた被験者のサンプルを収集する国立成育医療研究センター等の研究者が所属する機関の「ヒトゲノム倫理審査委員会」の承認を得たサンプルのみを解析対象とする。

C. 研究結果

本年度では、分担者の服部が生産した、健常人（18名）からのヒト腸内細菌叢の16Sデータとメタゲノムデータについて、配列データの精度チェック、相同性検索、遺伝子発見等のバイオインフォマティクスの解析条件の最適化をはかった。

シーケンサーから得られたメタゲノムデータからのヒト等の真核生物由来の配列及び夾雑多重リードを除去することによってユニークな高品質メタゲノムリードを収集する品質評価解析ラインを構築した。ついで、各サンプルのメタゲノムリードをNewblerでアセンブリを試行・調整を行い、18サンプルから合計1,800 Mb（平均100 Mb/サンプル）のユニーク配列データを取得した。ついで、分担者らが2006年に開発したMetaGeneを用いた遺伝子予測を行った。MetaGeneはコドン使用頻度とリボソーム結合部位配列等を指標に、分類が不明の細菌種（＝原核生物）ゲノム配列から遺伝子領域を類推するメタゲノムデータ用の遺伝子予測プログラムである。この解析から、合計約450万の遺伝子を同定した。さらに、配列類似度 $\geq 95\%$ でのクラスタリングによって、約230万個のユニークな遺伝子を抽出した。

16S解析では、服部が生産した24名の健常人から得られた3,000データ/サンプルの16S（V1-V2）配列を用いて情報学的解析を行った。まず、16S配列の両端のプライマー配列の有無、公的データベース中の既知16S配列とのアラインメントのカバー率、GSシーケンサーに備わっている塩基精度数の3つを指標とした配列の品質管理システムの開発を行った。これにより、得られる全リード数の60～70%が高品質データとして収集できた。

D. 考察

メタゲノム配列と 16S 配列に含まれるヒト由来リード、夾雑多重リード、低精度リード、キメラリード等の次世代シーケンサーから生産される副産物データをほぼ完全に除去する品質管理システムを構築した。また、メタゲノムデータのアセンブリと MetaGene を用いた解析によって約 230 万個のユニークな遺伝子を抽出できた。

E. 結論

次世代シーケンサーから生産される大量の粗配列データに含まれる低品質または夾雑データを効率的に除去するシステムを構築した。また、Newbler を用いたアセンブリ及び MetaGene を用いた遺伝子予測プロセスを確立した。

F. 研究発表

1. 論文発表

なし。

2. 学会発表

なし。

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。

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
		松本 健治	独立行政法人国立成育医療研究センター・部長
		野村伊知郎	独立行政法人国立成育医療研究センター・上級研究員
		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・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
		浅原 弘嗣	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
		藤野 明浩	独立行政法人国立成育医療研究センター・医員
		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：

反復胎状奇胎を中心とする異常妊娠や先天性奇形症候群の遺伝要因の解明を目指す。本年度は、本分担課題ならびに本研究事業の他の分担研究者の分担課題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することを目的として研究を実施した。本事業においては最も需要が高いエクソーム解析ならびにRNA-seq解析についてイルミナ社 HiSeq1000 システムを用いたデータ取得系を確立し、更にエクソーム解析データからSNP/InDel リストを抽出する解析パイプラインを構築した。これらの解析系を用いて、当初の計画を上回る116症例についてエクソーム解析データを取得した（全エクソン解析20例・部分エクソン解析96例）。また、分娩歴と出生児に特段の異常を認めなかった症例の、大規模収集を開始した。

連携研究者

周産期病態研究部・周産期ゲノミクス研究室長
中林一彦

周産期病態研究部・胎児発育研究室長
前原佳代子

反復胎状奇胎を中心とする異常妊娠や先天性奇形症候群の遺伝要因の解明を目指す。本年度は、本分担課題ならびに本研究事業の他の分担研究者の分担課題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することを目的として研究を実施した。

A. 研究目的

B. 研究方法

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)

1. エクソーム濃縮ライブラリー作製

使用するゲノム DNA の濃度は picogreen 定量キット (Invitrogen) により定量し、1-3 ug をゲノムライブラリー作製に供した。

アジレント社 SureSelectXT all exon 50Mb キットならびにカスタムキット (特定の遺伝子群・ゲノム領域を濃縮対象とした SureSelect カスタムキット)、イルミナ社 TruSeq Exome Enrichment キットを用いてエクソームライブラリーを作製した。各プロトコールの実験フローを図 1 に示した。DNA ライブラリーの評価 (サイズ確認・濃度推定) にはアジレント社バイオアナライザーを用いた。

エクソーム濃縮ライブラリーの定量にはアジレント社 QPCR NGS ライブラリー定量キットならびに Kapa Biosystems 社 Library Quantification キットを用いた。ABI HT7500 あるいは HT7900 を用いて定量データを取得した。

2. RNA-seq ライブラリー作製

疾患組織由来の RNA を次世代シーケンサーによるトランスクリプトーム解析に供することで、疾患の病態解明が期待できる。本研究班

においてもいくつかの分担研究班でトランスクリプトーム解析を実施する計画があることから、RNA-seq ライブラリー作製プロトコールを確立することを目指した。イルミナ社 TruSeq RNA Sample Preparation V2 キットならびに Epicentre ScriptSeq V2 RNA-Seq Library Preparation キットを用いた。図 2 に両プロトコールの概要を示した。

3. イルミナ社次世代シーケンサーシステムによる配列データ取得

イルミナ社 cBot Cluster Generation システムを用いた DNA 増幅によりフローセル上にクロノラルクラスターを形成させた。イルミナ社 HiSeq1000 システムにより大量並列シーケンシング (Paired-end read 100bp X2) を実施した。シーケンシング試薬には TruSeq SBS Kit v3-HS キットを用いた。

4. データ解析

配列データ解析を内部で実施するためのワークステーションとして次のスペックのものを導入した。

CPU: Intel Core i7-2600 3.4GHz 4core

メモリ: 16GB

HDD: 2.5TB x 4 台 (RAID10) = 実効 5TB

OS: CentOS Linux 6.0

このワークステーションに、配列データ解析に必要な配列情報 (ヒトリファレンスゲノム配列 hg19 version, マウスリファレンスゲノム配列 mm9 version, Refseq gene 配列ファイルなど) および配列解析用ソフトウェア (BWA, Picard, GATK など、図 3) をインストールした。

(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. 次世代シーケンサーによるデータ取得系の確立

①エクソーム濃縮ライブラリーの作製と定量

疾患における原因変異同定のためにエクソーム解析系を構築した。アジレント社 SureSelect ならびにイルミナ社 TrueSeq エクソーム濃縮プロトコルの二つについてライブラリー作製が可能な体制を整備した。

i) SureSelect プロトコール

4 検体についての結果を図 4 に示した。Exome 濃縮前と濃縮後の DNA ライブラリーのサイズ分布をバイオアナライザー 2100 で解析し、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。

ii) TrueSeq Exome enrichment プロトコール

イルミナ社 Truseq DNA Sample Prep キットを用いて 4 検体についてゲノムライブラリーを作製し、Truseq Exome Enrichment キットを用いてエクソーム濃縮を行った。4 検体のゲノムライブラリーをプールしてエクソーム濃縮工程を実施した。結果例は図 5 に示した。こちらについても最終的には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。

iii) エクソーム濃縮ライブラリーの定量解析

得られたエクソーム濃縮ライブラリーの配列データを次世代シーケンサーで取得する際は、同一レーン内に複数のサンプルを混合する(混合した各検体の配列はアダプター配列中に含まれる index 配列の違いによりデータ解析時に分類が可能)。この時に混合する各ライブラリーのモル濃度を均一にすることが肝要であり、定量 PCR による高精度な濃度測定が必要となる。アジレント社 QPCR NGS ライブラリ定量キットならびに Kapa Biosystems 社 Library Quantification キットを用いて定量 PCR 解析を行った。結果例は図 6 に示した。定量の正確性についてはどちらも遜色無かったが、操作の簡便性やランニングコストを考慮し、今後の定量

解析には後者キットを採用するに至った。また、定量 PCR 結果 (HT7500 あるいは HT7900 出力結果) をコピー&ペーストすれば、個々のサンプルの定量値や補正倍率を計算・グラフ表示できるエクセルシートを独自に作成し、作業を効率化した (図 7)。

iv) 自動化装置導入によるライブラリー作製スループット向上

本年度備品として次世代シーケンサーライブラリー作製自動化装置ベックマン SPRIWorks を導入し、ゲノムライブラリー作製工程の大幅な効率化に成功した。

②RNA-seq ライブラリーの作製・定量

i) TruSeq RNA Sample Preparation V2 法

1.0 μg の total RNA を出発材料としてライブラリー作製を行い、狙い通りのインサート長のライブラリーが出来ることが確認できた (図 8)。100 ng からの total RNA を出発材料としても同様の結果が得られた。その後のライブラリーの定量には前述の Kapa Biosystems 社 Library Quantification キットを用いた。

ii) ScriptSeq V2 RNA-Seq Library Preparation 法

1.0 μg の total RNA を Epicentre Ribo-Zero Gold キットで処理し、ribosomal RNA 除去を行った。次に ScriptSeq V2 RNA-Seq Library Preparation キットを用いてライブラリー作製を試みた。この方法では期待されるインサート長のライブラリーが得られなかった (図 8)。

③HiSeq1000 システムによる大量配列データ取得

当センター共通機器として HiSeq1000 システムが導入され、今年度 11 月末より稼働可能となった。今年度はシステムのテストランを含めて 5 回のラン (paired-end read, 101x2 cycle sequencing run) を実施し、各ランにおいてシステムのスペック通りの配列データ (300Gb 前後) の取得に成功した。図 9・10 にデータクオリティーチェック結果の例を示した。

なお今年度末にイルミナ社 MiSeq システムの導入も完了したことで、大規模解析には

HiSeq1000 システム、小規模迅速解析には MiSeq システム、という使い分けが可能な体制が整備された。

④その他の NGS ライブラリー作製プロトコール着手状況

i) 微量ゲノム DNA を出発材料とした次世代シーケンサー用ゲノムライブラリー構築法の検討

ゲノムライブラリー作製のための通常のアダプター付加プロトコールには最低でも 1 μ g 程度のゲノム DNA が出発材料として必要であり、量が限られた検体からライブラリーを作製するのは困難である。イルミナ社 Nextera DNA サンプル調製キットはトランスポゼース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)。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(50ng)であることがこの方法の長所である。臨床検体を対象としたゲノムライブラリー作製にはこの方法が有効であると考えられることから、ゲノム DNA および PCR 産物(5kb)を出発材料として Nextera DNA サンプル調製キットによるライブラリー調製を行った。ゲノム DNA を鋳型とした場合はキット推奨の実験条件でライブラリー作製がうまくいくのに対し、PCR 産物を対象とした場合はトランスポゼース反応の効率がゲノム DNA のときに比べて低く、シーケンシング可能なインサート長のライブラリーを得るには反応時間を 10 倍程度に延長する必要があることが判明した(例えば DNA の GC 含量など塩基配列特徴がタグメンテーション効率に影響を及ぼす可能性が考えられる)。またいずれのタイプの鋳型においてもキット推奨量(50ng)以上のゲノム DNA を反応に用いるとタグメンテーション効率が低下すること、より少量(25ng 程度)の DNA を出発材料とすると良好な結果が得られることが判明した。以上の様に、条件を最適化すればこの方法がゲノム DNA・PCR 産物の両方に適応できることを

確認した。

ii) エピゲノム解析プロトコール

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。ChIP-seq 法ならびに RRBS 法 (Reduced Representation Bisulfite Sequencing) が実施可能な体制を構築すべく、テストサンプルでライブラリー作製を開始している。次世代シーケンサーによる DNA メチル化・ヒストン修飾解析で優れた実績を挙げている国内外施設から入手したプロトコールを参照している。

2. 次世代シーケンサーデータ解析系の確立

次世代シーケンサーによる産出される生データは、大量の 100bp 程度の短い DNA 断片の解読情報(リードと呼ばれる)で、我々が使用しているイルミナ社 HiSeq1000 の場合は、30 億リード(塩基配列では 300G 塩基対)のデータが算出される。この大量のリード情報を生物学的に意味のある情報へ変換するために、バイオインフォマティクス解析系を構築した。今年度は、当研究グループ内でデータ解析を実施するためのコンピュータシステムの整備、ならびにエクソーム解析用のデータ処理パイプラインを構築することを優先的に実施した。

①エクソームデータ解析パイプラインの構築

予備的に複数の解析ソフトウェアを試用した上で、以下のような解析パイプラインを構築するに至った(図 1 1)。

- | |
|---|
| 工程 1 : BWA によるリード情報のリファレンスゲノムへのマッピング |
| 工程 2 : Picard による duplicate リード (PCR 増幅に起因すると考えられる同一配列) の除去 (GATK 使用) |
| 工程 3 : 1000 Genomes 情報に基づいた既知 InDel 周辺の再アライメントおよび塩基スコアのキャリブレーション |
| 工程 4 : GATK によるリードカバレッジ算出 |
| 工程 5 : エクソーム解析ターゲット領域内での SNV (SNP)、InDel の検出 |
| 工程 6 : snpEff による Effect 解析
SNV (SNP)・InDel の効果予測とアノテーション付与 |

SNV: single nucleotide variation
SNP: single nucleotide polymorphisms
InDel: insertion and deletion

工程 1 補足 (BWA による配列マッピング)
BWA (Burrows-Wheeler Aligner) はリード情報をリファレンスゲノムへアライメントするためのフリーソフトウェアであり、現時点で最も使用されている。http://bio-bwa.sourceforge.net/ より version0.6.1 を取得した。アライメント後のファイル形式変換や圧縮、ソート、インデックス作成には、フリーソフトウェア SAMtools (Tools for alignments in the SAM format) を使用した。
http://samtools.sourceforge.net/ より version0.1.18 を取得した。BWA の実行コマンドは、ラッパーとして Perl スクリプト中に組み込み実行した。BWA を実行するための Perl スクリプトを図 1 2 に示す。

工程 2 補足 (Picard による重複リード除去)
Duplicate (重複) リードは、PCR 増幅のバイアスによって引き起こされる全く同じゲノム位置の同じ配列のリードを意味する。Duplicate リードは、SNV (SNP) 検出でミスコールを引き起こす要因となりうるため、取り除く必要がある。Picard は、http://picard.sourceforge.net/ より version1.61 を取得した。Picard 実行コマンドを含む Perl スクリプトによりアライメントファイル (サンプル名.sort.rm.bam) を作成した。

工程 3 - 5 補足 (GATK によるデータ処理)
GATK (Genome Analysis Toolkit) は、次世代シーケンサーでのデータ解析を行うにあたり、様々な機能を提供しているフリーウェアパッケージである。GATK は
http://www.broadinstitute.org/gsa/wiki/index.php/Main_Page より version1.4-21 を取得した。詳細は省くが、以下の各工程で GATK 実行コマンドを含む Perl スクリプトを使用した。

100bp 程度の短いリードをアライメントする際、アライメント箇所に挿入(insertion)・欠失(deletion)が存在するとミスアライメントが生じやすくなる。この問題を解決するため、

GATK ソフトウェアを用いて 1000 Genomes Project (http://www.1000genomes.org/) で報告されている既知 InDel 周辺の再アライメントおよび塩基スコアのキャリブレーションを行った。この操作で、再アライメントおよびキャリブレーションを行ったアライメントファイル (サンプル名.realn.recal.bam) を作成した (工程 3)。

GATK の DepthOfCoverage 機能を用いて、エクソームターゲット領域のリードカバレッジ (read depth of coverage) の計算および集計を行った (工程 4)。結果例 (エクセルによる集計) を図 1 3 に示す。更に GATK を用いてアライメントファイル (サンプル名.realn.recal.bam) からの SNV (SNP), InDel 変異の検出 (コール) を行った。この操作により、検出結果ファイル (ファイル名.snv_indel.filt.vcf) を得た。VCF (Variant Call Format) の詳細については、後述する。

工程 6 補足 (snpeffによる Effect 解析)

GATK により検出された SNV (SNP) ・ InDel 変異についてフリーソフトウェア snpeff (SNP effect predictor) により、効果予測 (NON_SYNONYMOUS_CODING や FRAME_SHIFT 等) とアノテーション (遺伝子名や ID 等) 付けを行った。snpeff は、
http://snpeff.sourceforge.net/ より version2.0.5d を取得した。なお snpeff による効果予測・アノテーション付与には主に ENSEMBL データベース、GRCh37.65 が利用されている。最終的な結果は、エクセル (.xlsx), VCF, HTML ファイルとしてまとめた。

②SNV/InDel リスト様式の作成

i) SNV/InDel リスト (Excel 形式・VCF 形式)

フィルタをパスした SNV, InDel の効果予測とアノテーションを transcript (転写産物 accession number) 単位でまとめた。Transcript 単位のため、同じ SNV, InDel が複数の異なる transcript 毎に冗長に出力される場合があるが、一つの SNV/InDel の効果が transcript 毎に異なる可能性があることからこのようなまとめを作成し、ファイル名を**サンプル名.snv_indel.pass.annot.xlsx**と定めた。

また、SNV, InDel の効果予測とアノテーションを遺伝子単位 (またはゲノム領域) でまとめたファイルも作成した。全 SNP/InDel をまとめた表のファイル名を**サンプル名.snv_indel.genes.xlsx**、フィルタをパスした

SNV/InDel をまとめた表のファイル名は、**サンプル名.snv_indel.pass.genes.xlsx** とした。

さらに全ての SNV, InDel の効果予測とアノテーションの結果を VCF (Variant Call Format) でもまとめた (ファイル名は**サンプル名.snv_indel.annot.vcf** とした)。VCF はタブ区切りのテキストファイルのため、エクセルで確認することができる。VCF は更に変異を記述するフォーマットとしてスタンダードなものとなっており、変異のフィルタリングや他ツールへの取り込み等に用いることができる。各ファイル形式を図 1 4 - 図 1 7 に示した。

ii) SNV, InDel 統計情報 (HTML 形式)

検出された SNV, InDel の様々な統計情報 (効果予測、カバレッジ、アミノ酸置換等) を HTML 形式でまとめた。これらのファイルはウェブブラウザで確認することができる。

フィルタをパスした SNV, InDel のみの統計情報ファイル名を**サンプル名.snv_indel.pass.summary.html** とし、全 SNV, InDel の統計情報ファイルを**サンプル名.snv_indel.summary.html** と定めた。ファイル形式例を図 1 8 に、snpEff による変異効果予測データ書式とブラウザ出力例を図 1 9 - 2 1 に示した。

以上のように、バイオインフォマティクスを駆使して、エクソーム解析データを 6 段階の工程からなる解析パイプラインで処理することにより、データ品質チェック、マッピング、SNP/InDel 検出、アノテーション付けまでを実施する体制を整えた。また最終的な結果はインフォマティクスの専門知識を持たないウェット系研究者が扱いやすいような形式 (主に Excel 形式と HTML 形式) でまとめることとし、その様式を定めた。

3. 疾患検体群を対象としたエクソーム解析・RNA-seq 解析実施状況

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、上記のデータ取得系・データ解析系を用いて、エクソーム解析・RNA-seq 解析を開始した。エクソーム解析の進行状況を次表にまとめた。表中の数字は各プロジェクト・各段階の検体数を示している。

		→				
エクソーム解析進行状況		前処理済	gDNAライブラリー作製済	Exome濃縮ライブラリー作製済	シーケンシング済	データ解析 (SNPリスト作成) 済
		成育医療研究センター内	深見_A	52		
深見_B			2			48
深見_C	15		20			
深見_D	15					
高田	20					
秦				8		12
その他				4	8	
外部		12		12		

深見 A・B・C プロジェクトならびに高田プロジェクトでは SureSelect カスタムキット (対象領域 187kb から 1490kb) を使用しており、その他は全エクソン解析。配列データ取得を終えている検体を青字表示した。

RNA-seq 解析については、40 検体についてライブラリーを作成し、そのうち 24 検体についてはシーケンシングデータを既に取得している。

D. 考察

1. エクソーム濃縮ライブラリー

エクソーム解析において繁用されている二

社のキットのいずれでもライブラリー調製がうまく行くことを確認した。両キットには濃縮対象領域 (3' 非翻訳領域を対象とするか否か)、価格、プロトコルの煩雑さ (ゲノムライブラリー調製時にゲル切り出し工程が必要か否か、複数サンプルを混合してエクソーム濃縮を行えるか否か)、などについて相違があり、一概に優劣は付けがたい。両プロトコルを確立したことで、今後のエクソーム解析の対象となる疾患群の特徴 (予想される変異のタイプや遺伝様式ならびに解析規模など) に応じて二つのキットを使い分けることが可能となった。なお、コーディングエクソン領域における SNP 同定の効率は両プロトコルで重大な差は無いことが 2011 年に Nature Biotechnology に報告されている。

2. RNA-seq ライブラリー

実験系の確立を試みた二つの方法のうちの一つである TruSeq RNA Sample Preparation V2 法では、これまでに 40 検体程度についてライブラリー作製を終えており、常に安定した結果が得られている。取得したデータの一部をモデルとして用いることで、RNA-seq データ解析系の構築にも着手している。

これに対してもう一つの方法である ScriptSeq V2 RNA-Seq Library Preparation 法では、これまでに 2 回試行したものの、今のところ期待通りのライブラリーが得られていない。TruSeq 法と Epicentre 法の大きな違いは、最終的に取得した配列データの転写の向き (+ strand あるいは - strand) の識別が出来るか否かで、これは Epicentre 法でのみ可能となっている。通常解析目的では TruSeq 法が十分に役割を果たすのに対し、例えばエクソン-イントロン構造を持たない non-coding RNA の同定などの目的を持ってデータ解析を行う場合は転写方向情報が重要となる。将来的にそのような目的での解析が必要になる可能性を考慮し、Epicentre 法のトラブルシューティングを継続する予定である。

3. 大量配列データ解析系

次世代シーケンサーデータの解析に最適化したワークステーションを構築し、エクソームデータ解析系を短時間で構築するに至った。これまでに取得したエクソーム解析データ (全エクソンデータ 20 検体分および部分的エクソンデータ 96 検体分) について、SNP/Indel 検出までが完了している。現在までに個々の検体について得られた SNP/Indel リストには疾患とは無関係な SNP が多数含まれていると考えられ、疾患の原因変異をその中から効率よく絞り込む方法を確立していくことが今後の課題である。当分担研究班で取得したデータを各分担研究者に提供し、意見交換することでエクソームデータ解析系の完成度をより高めたい。

解析スループットを考慮し、本年度は Linux OS で稼働する Character User Interface (コマンド入力主体) のオープンソフトウェアを用いた解析系の構築に重点を置いた。インフォマティクス知識があまり無い分担研究者がデータ閲覧・解析を実施できるよう、Graphical User Interface の解析ソフトウェア (アジレント社 Avadis NGS など) の導入も進めており、NGS ライブラリー作製だけでなくデータ解析においても多様なニーズに応じられる体制の構築を目指している。

E. 結論

本年度 (9 月開始) は本分担課題ならびに他分担課題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した。本事業においては最も需要が高いエクソーム解析ならびに RNA-seq 解析についてイルミナ社 HiSeq1000 システムを用いたデータ取得系を確立し、更にエクソーム解析データから SNP/InDel リストを抽出する解析パイプラインを構築した。これらの解析系を用いて、計画を上回る 116 症例についてエクソーム解析データを取得した (全エクソン解析 20 例・部分エクソン解析 96 例)。解析系の構築が順調に進んだことから、来年度以降はより大規模なエクソーム解析が実施できる可能性

が高い。また今年度の経験は、次世代シーケンサー技術を基盤とするトランスクリプトーム解析やエピゲノム解析系の確立にも役立つと考えられる。

F. 研究発表

1. 論文発表

原著論文

Kobayashi H, Sakurai T, Sato S, Nakabayashi K, Hata K, Kono T. A cis-linkage between large-range Gpr1-Zdbf2 intergenic transcription and imprinted DNA methylation at Gpr1 locus during early embryogenesis. *FEBS Letters* (in press)

Fuke-Sato T, Yamazawa K, Nakabayashi K, Matsubara K, Matsuoka K, Hasegawa T, Dobashi K, Ogata T (2012) Mosaic upd(7)mat in a patient with Silver-Russell syndrome. *Am J Med Genet A* 158A:465-468.

Kobayashi H, Sakurai T, Takahashi N, Fukuda A, Obata Y, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T (2012) Mouse gametic DNA methylomes show the role of intragenic DNA methylation in the establishment of oocyte-specific heritable marks. *PLoS Genet* 8:e1002440.

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

O'Doherty AM, Rutledge CE, Sato S, Thakur A, Lees-Murdock DJ, Hata K, Walsh CP (2011) DNA methylation plays an important role in promoter choice and protein production at the mouse Dnmt3L

locus. *Dev Biol* 356:411-420.

Sato S, Yoshida W, Soejima H, Nakabayashi K, Hata K (2011) Methylation dynamics of IG-DMR and Gtl2-DMR during murine embryonic and placental development. *Genomics* 98:120-127. 2011.

Nakabayashi K, Trujillo AM, Tayama C, Camprubi C, Yoshida W, Lapunzina P, Sanchez A, Soejima H, Aburatani H, Nagae G, Ogata T, Hata K, Monk D (2011) Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes. *Hum Mol Genet* 20:3188-3197.

Tomizawa S, Kobayashi H, Watanabe T, Andrews S, Hata K, Kelsey G, Sasaki H (2011) Dynamics stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* 138:811-820.

日本語総説

秦健一郎：胎盤形成のエピジェネティクス
臨床産婦人科 第65巻第3号 243-246 2011

久須美真紀、秦健一郎：エピジェネティックな変異と発生異常
産婦人科の実際 第60巻第10号 1423-1430 2011

2. 学会発表

(招待講演)

秦健一郎：ヒト異常妊娠のエピジェネティクス
大阪大学蛋白質研究所セミナー「疾患におけるエピゲノム異常の分子機構」大阪、2011.11.18

秦健一郎：早産・異常妊娠のゲノム解析とエピゲノム解析
第5回日本早産予防研究会 東京、2011.12.3

(一般演題)

山口裕子、鳥巢弘道、中林一彦、久須美真紀、
前原佳代子、菅原直子、秦健一郎：胎盤特異的
エピゲノム異常に起因する胎児発育不全 日
本人類遺伝学会第56回大会千葉、2011.11.10

田山千春、Alex Martin Trujillo, Pablo Lapunzina、
副島英伸、緒方勤、秦健一郎、David Monk、
中林一彦：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新規インプリン
ト遺伝子の同定
第34回日本分子生物学会年会 横浜2011.12.15

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

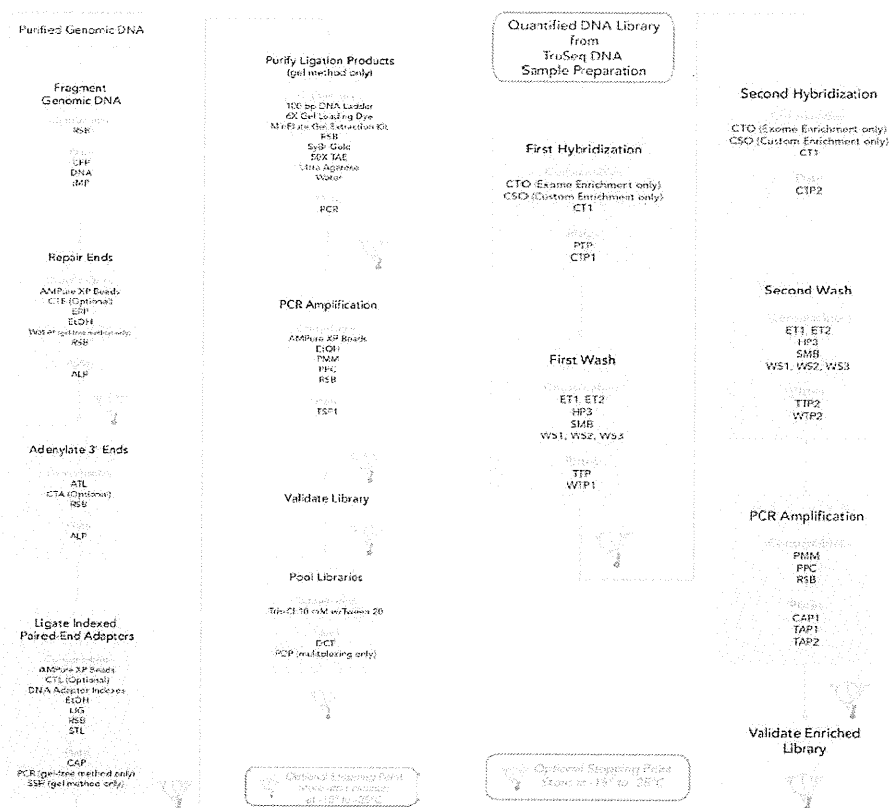
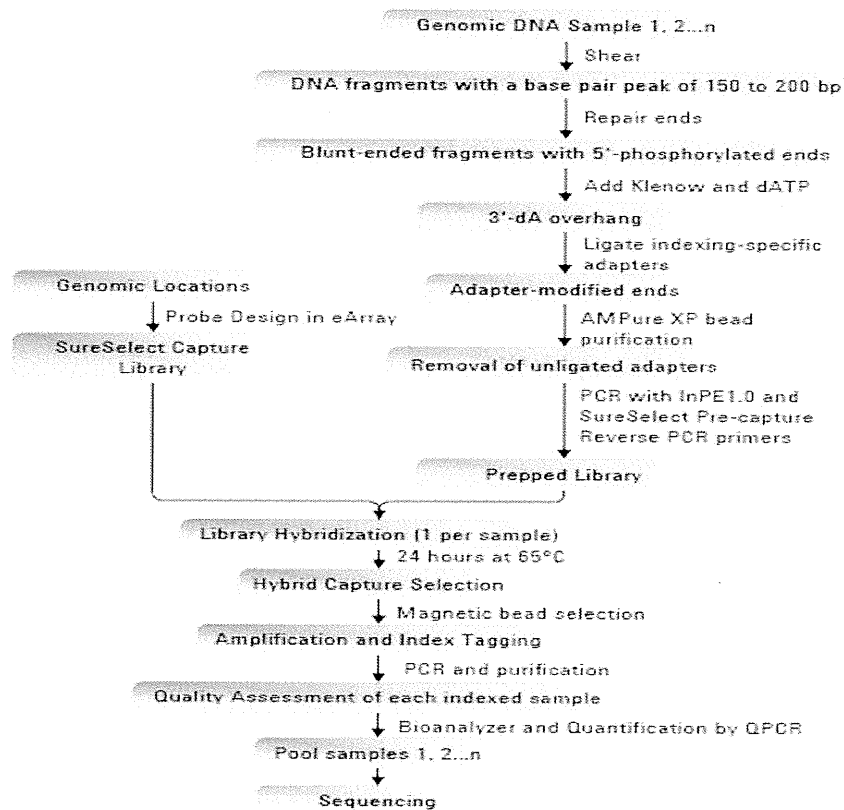
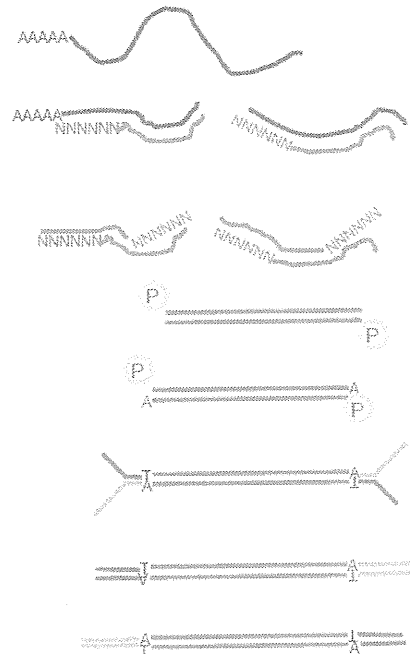
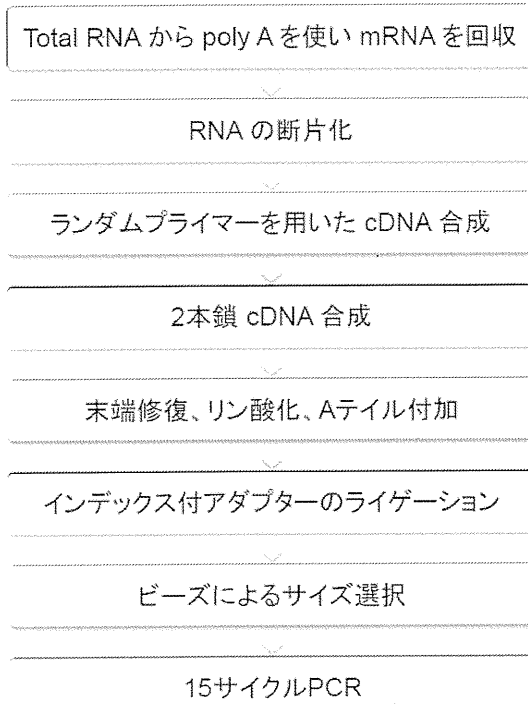


図1: エクソーム濃縮ライブラリー作製の実験フロー

アジレント社SureSelectプロトコール(上)とイルミナ社TruSeq Exome Enrichmentプロトコール(下)

A



B

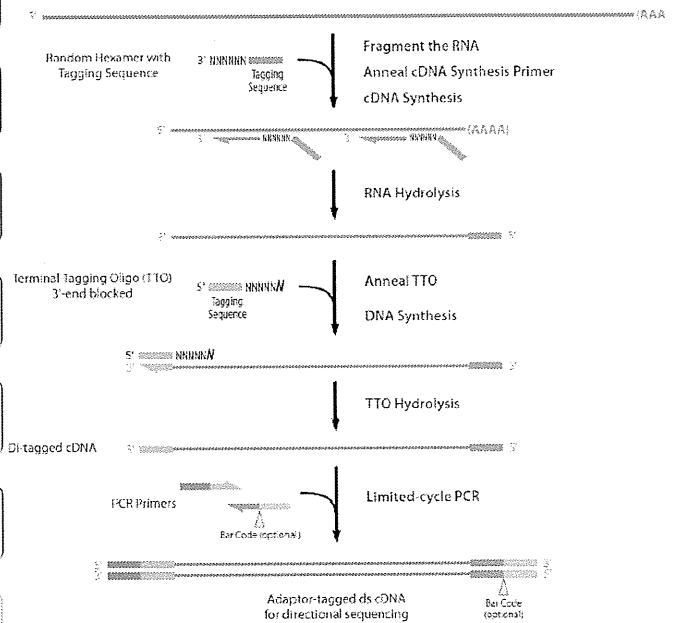
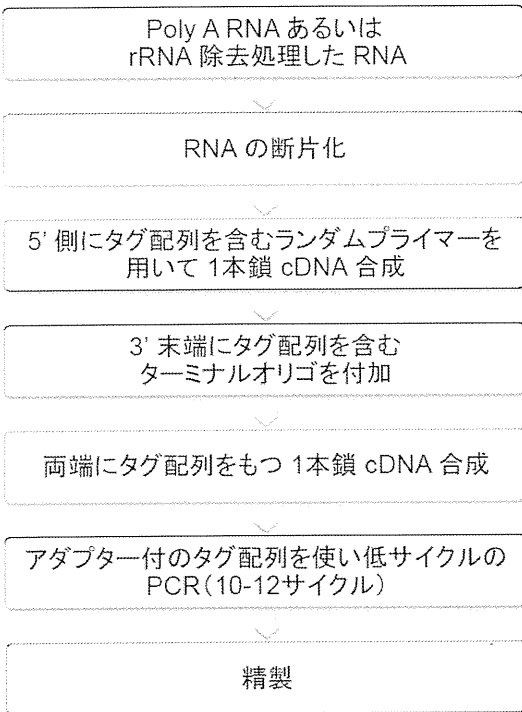


図2: RNA-seqライブラリー作製工程

TruSeq RNA Sample Preparation プロトコール概要 (A)

Epicentre Script-Seq v2 プロトコール概要 (B)

http://www.illumina.com/redirect/webinar/2011_ILMN_RNA-Seq_Session4.html
より抜粋

データ解析ツール	バージョン	ライセンス	解析項目
FastQC	fastqc_v0.10.0	GPLv3	QC
FastX	fastx_toolkit_0.0.13	GPLv3	
bwa	bwa-0.5.9	GPLv3	Reseq, ChIP-seq
picard	picard-tools-1.54	Apache License V2.0 MIT License	
igvtools	igvtools_1.5.9	GNU LGPL license	
samtools	samtools-0.1.18	BSD License MIT License	
break dancer	breakdancer-1.1_2011_02_21	GPLv3	
gtf2gff.pl	v1.0.0	Indiana University	
quick annotater	v1.0.0	Amelieff Co. Ltd Science Design LLC	
macs	MACS-1.4.1.tar.gz	The Artistic License	ChIP-seq
bowtie	bowtie-0.12.7	The Artistic License	RNA-seq
tophat	tophat-1.3.3	The Artistic License	
cufflinks	cufflinks-1.1.0	Boost Software License	

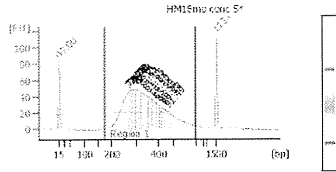
図3:
大量配列データ解析用ワークステーションに導入した主なソフトウェア

A

PCR 6cycles濃縮後、5倍希釈

Name	From (bp)	To (bp)	Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [ng/ μ l]	Color	Molarity [nmol/l]	peak top (bp)	5 \times Conc. (ng/ μ l)	μ l	HS	500ng(μ l)	D.W.
HM18mo	176	700	926.8	95	354	28.1	31.85	16711680	148.8	280	156.25	9.7	1545	8.14	0.26
HM19mo	155	700	936.9	89	345	30.2	30.59	16711680	193.3	278	191.85	12.9	2354	2.53	0.87
HM21mo	170	700	879.8	90	346	30.1	30.31	16711680	176.1	275	181.55	8.2	1489	2.75	0.85
HM25mo	155	700	884.3	95	342	26.8	36.15	16711680	197.8	289	195.75	12.8	2506	2.55	0.85

DNA1000

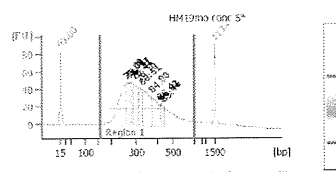


Overall Results for sample 3 : HM18mo conc 5%

Number of peaks found: 17
 Area 1: 926.8

Region table for sample 3 : HM18mo conc 5%

From [bp]	To [bp]	Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [ng/ μ l]	Co lo
176	700	926.8	95	354	28.1	31.85	

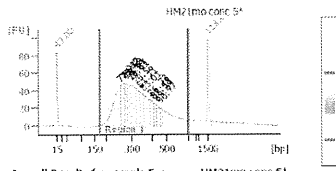


Overall Results for sample 4 : HM19mo conc 5%

Number of peaks found: 13
 Area 1: 936.9

Region table for sample 4 : HM19mo conc 5%

From [bp]	To [bp]	Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [ng/ μ l]	Co lo
155	700	936.9	89	345	30.2	30.59	

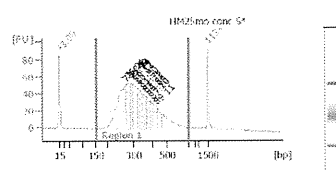


Overall Results for sample 5 : HM21mo conc 5%

Number of peaks found: 15
 Area 1: 879.8

Region table for sample 5 : HM21mo conc 5%

From [bp]	To [bp]	Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [ng/ μ l]	Co lo
170	700	879.8	90	346	30.1	30.31	



Overall Results for sample 6 : HM25mo conc 5%

Number of peaks found: 14
 Area 1: 884.3

Region table for sample 6 : HM25mo conc 5%

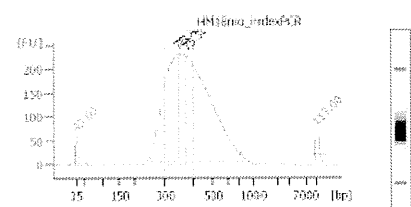
From [bp]	To [bp]	Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [ng/ μ l]	Co lo
155	700	884.3	95	342	26.8	36.15	

B

hvb-index PCR 10cycles

Name	From (bp)	To (bp)	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [ng/ μ l]	Molarity [nmol/l]	Color	peak top (bp)	pg/30ul (\pm)	fmol/30ul (\pm)
HM18mo	300	400	2,299.20	48	350	8	3,000.28	12,970.40	12930121	353	90068	90
HM19mo	300	400	1,136.7	50	350	8.1	1,488.83	6,442.30	12930121	352	44658	45
HM21mo	300	400	2,105.18	48	350	8.1	2,790.58	12,962.80	12930121	352	83712	84
HM25mo	300	400	668.7	50	351	8	1,036.53	4,174.92	12930121	353	31098	31

High Sensitivity

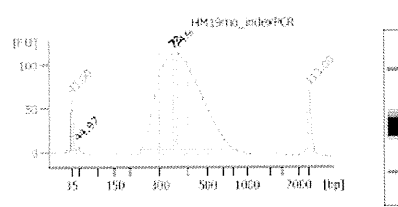


Overall Results for sample 1 : HM18mo_indexPCR

Number of peaks found: 3
 Area 1: 6.2
 Corr. Area 1: 2,299.2

Region table for sample 1 : HM18mo_indexPCR

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/ μ l]	Molarity [nmol/l]	Co lo
300	400	2,299.2	48	350	8	3,000.28	12,970.4	

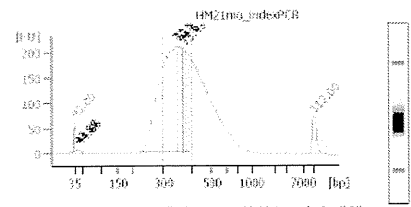


Overall Results for sample 2 : HM19mo_indexPCR

Number of peaks found: 4
 Area 1: 0.2
 Corr. Area 1: 1,136.7

Region table for sample 2 : HM19mo_indexPCR

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/ μ l]	Molarity [nmol/l]	Co lo
300	400	1,136.7	50	350	8.1	1,488.83	6,442.3	

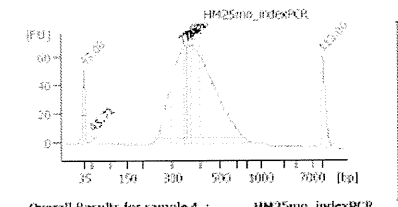


Overall Results for sample 3 : HM21mo_indexPCR

Number of peaks found: 6
 Area 1: 0.3
 Corr. Area 1: 2,105.1

Region table for sample 3 : HM21mo_indexPCR

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/ μ l]	Molarity [nmol/l]	Co lo
300	400	2,105.1	48	350	8.1	2,790.58	12,962.8	



Overall Results for sample 4 : HM25mo_indexPCR

Number of peaks found: 5
 Area 1: 0.3
 Corr. Area 1: 668.7

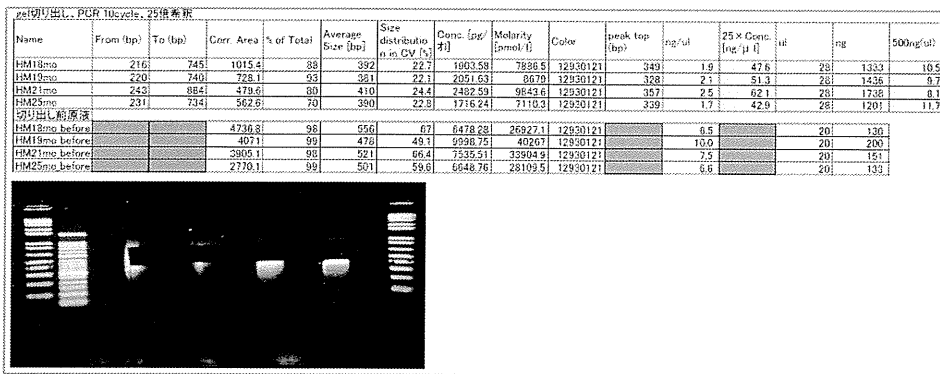
Region table for sample 4 : HM25mo_indexPCR

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/ μ l]	Molarity [nmol/l]	Co lo
300	400	668.7	50	351	8	1,036.53	4,174.9	

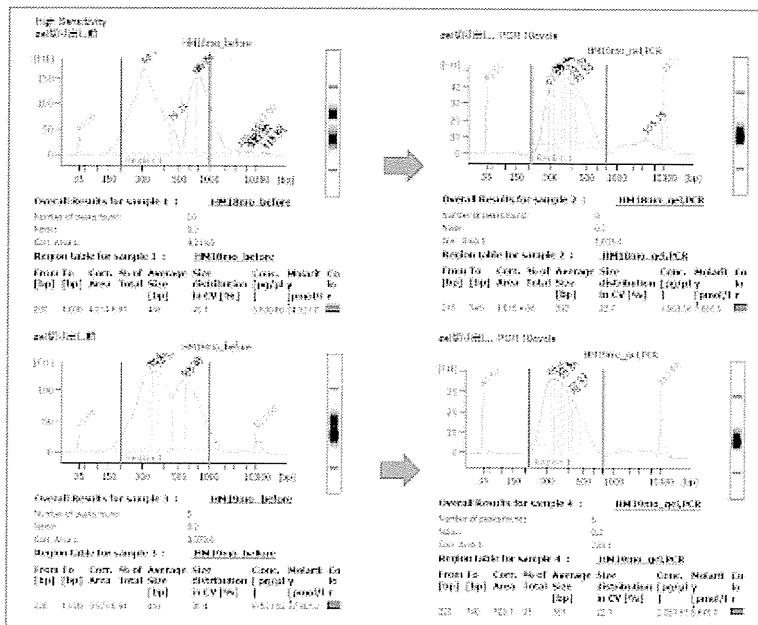
図4: SureSelect法によるエクソーム濃縮ライブラリーの作製

A. Exome濃縮前、PCR 6cycles後のサイズチェックと定量的結果(バイオアナラザーDNA1000アッセイ)
 B. Exome濃縮後、PCR 10cycles後のサイズチェック結果(バイオアナラザー High Sensitivityアッセイ)

A



B



C

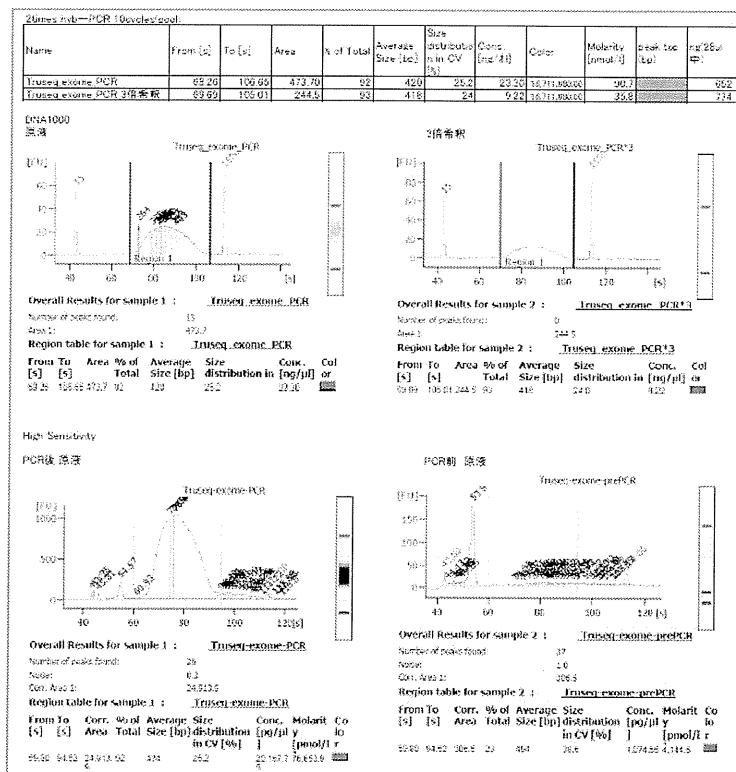


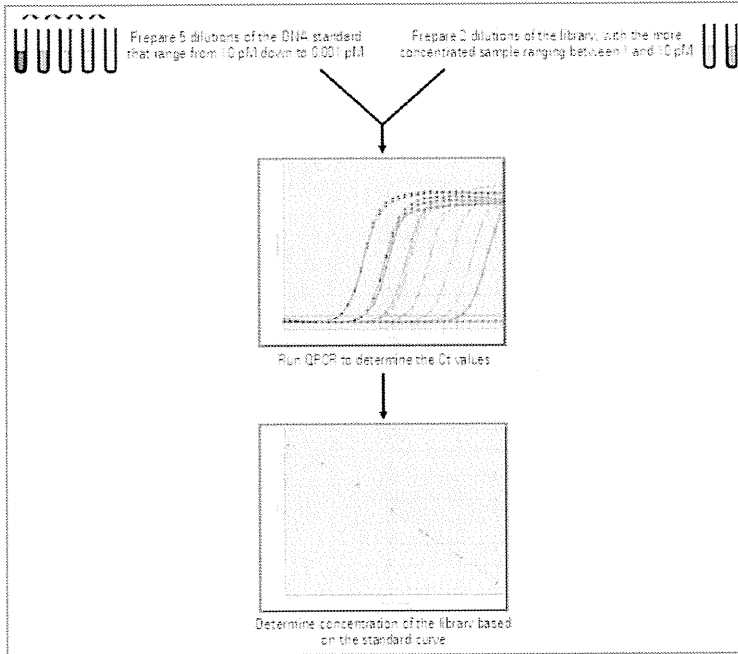
図5: TruSeq法によるエクソーム濃縮ライブラリーの作製

ゲル切り出し工程 (A)

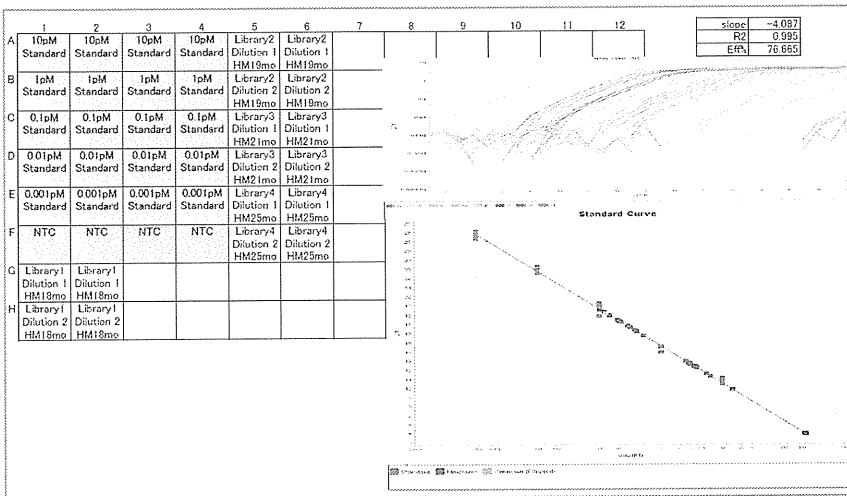
ゲノムライブラリー作製工程。PCR 10cycles後のDNAサンプルをAgilent2100バイオアナラザ High Sensitivityアッセイによりサイズチェックを行った結果 (B)

エクソーム濃縮工程。サンプルプール、Exome濃縮、PCR 10cycles後のDNAサンプルをAgilent2100バイオアナラザ DNA1000、High Sensitivityアッセイによりサイズチェックを行った (C)

A



B



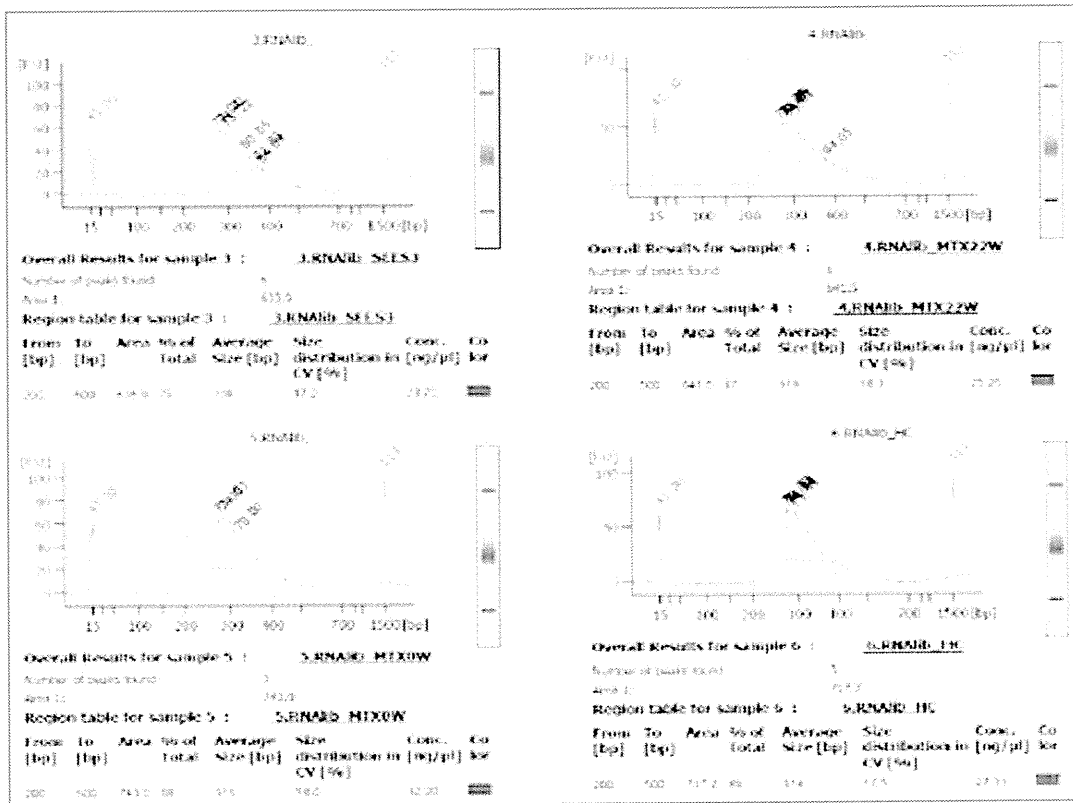
C

Library sample	10pM-10pM		希釈液1 (100倍希釈)				Dilution 1				Dilution 2				
	From (bp)	To (bp)	Conc. (pg/μl)	Molarity (pmol/μl)	希釈率	Library (ul)	Dilution buffer (ul)	Molarity (pmol/μl)	希釈率	Library (ul)	Dilution buffer (ul)	Molarity (pmol/μl)	Dilution 1 Library (ul)	Dilution 2 Library (ul)	Molarity (pmol/μl)
HM18mo	300	400	3.600	12.970	12.970	2.0	198.0	129.7	129.7	2.0	257.4	1.0	2.0	18.0	0.1
HM18mo	300	400	1.439	6.442	6.442	2.0	198.0	64.4	64.4	2.0	126.8	1.0	2.0	18.0	0.1
HM18mo-10	300	400	2.791	12.083	12.083	2.0	198.0	120.8	120.8	2.0	239.7	1.0	2.0	18.0	0.1
HM25mo	300	400	1.037	4.475	4.475	2.0	198.0	44.7	44.7	2.0	87.5	1.0	2.0	18.0	0.1

Name	CT	CT Mean	CT SD	Quantity	Quantity Mean	Quantity SD	Automatic Ct Threshold	Ct Threshold	Automatic Baseline	Baseline Start	Baseline End	誤差	原液pM	
HM18mo	12.15727	12.37558	0.308728	3.841349	3.422512	0.592326	TRUE	0.056408	TRUE	3	9	1pM	3.4	44391
HM18mo	12.58368	12.37558	0.308728	3.003674	3.422512	0.592326	TRUE	0.056408	TRUE	3	9			
HM18mo-10	15.69301	15.99043	0.420614	0.524016	0.449407	0.105513	TRUE	0.056408	TRUE	3	12	0.1pM	4.5	58290
HM18mo-10	16.28785	15.99043	0.420614	0.374798	0.449407	0.105513	TRUE	0.056408	TRUE	3	13			
HM13mo	12.23142	12.28384	0.074137	3.684191	3.578528	0.149428	TRUE	0.056408	TRUE	3	9	1pM	3.6	23054
HM15mo	12.33626	12.28384	0.074137	3.472867	3.578528	0.149428	TRUE	0.056408	TRUE	3	9			
HM18mo-10	17.14869	16.85904	0.409631	0.230761	0.275291	0.062976	TRUE	0.056408	TRUE	3	14	0.1pM	2.8	17735
HM18mo-10	16.56939	16.85904	0.409631	0.319822	0.275291	0.062976	TRUE	0.056408	TRUE	3	13			
HM21mo	11.27298	11.40234	0.182948	6.322042	5.893273	0.606371	TRUE	0.056408	TRUE	3	8	1pM	5.9	71207
HM21mo	11.5317	11.40234	0.182948	5.464504	5.893273	0.606371	TRUE	0.056408	TRUE	3	8			
HM21mo-10	16.15431	16.23051	0.107762	0.404085	0.387461	0.02351	TRUE	0.056408	TRUE	3	13	0.1pM	3.9	46516
HM21mo-10	16.30671	16.23051	0.107762	0.370837	0.387461	0.02351	TRUE	0.056408	TRUE	3	13			
HM25mo	4.940919	4.914378	0.037535	223.9813	227.3812	4.808167	TRUE	0.056408	TRUE	1	2	130pM	1.7	22738
HM25mo	4.887837	4.914378	0.037535	230.7811	227.3812	4.808167	TRUE	0.056408	TRUE	1	2			
HM25mo-10	9.742766	9.759569	0.023763	14.97185	14.83145	0.198564	TRUE	0.056408	TRUE	3	6	13pM	1.1	14831
HM25mo-10	9.776373	9.759569	0.023763	14.69104	14.83145	0.198564	TRUE	0.056408	TRUE	3	6			

図6: エクソーム濃縮ライブラリーのqPCRによる定量
概念図(A)および結果例(B,C)

A



B

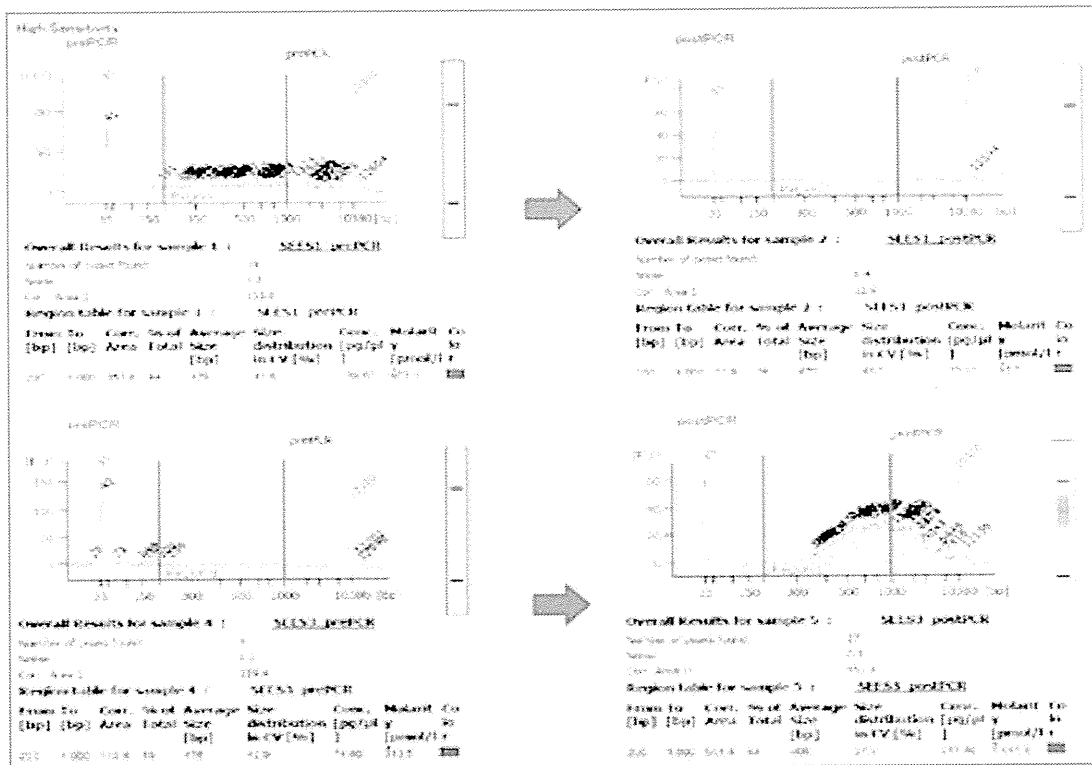


図8: RNA-seqライブラリーの作製

TruSeq RNA Sample PreparationによるRNA-seqライブラリー作製例(A)
 Epicentre Script-Seq v2プロトコールによるRNA-seqライブラリー結果(B)

Run Folder: D:\illumina\HiSeqTemp\111122_SNL123_0077_AD0E90ACXX

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads PF (M)	Perfect Reads (M)	<=3 Err (M)	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	48	742 +/- 106	95.1 +/- 1.2	0.183 / 0.228	194.75	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	3809 +/- 230	77.7 +/- 1.0
2	48	1095 +/- 84	87.6 +/- 3.9	0.193 / 0.213	252.06	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	3046 +/- 242	76.2 +/- 0.9
3	48	1099 +/- 98	88.2 +/- 3.4	0.175 / 0.209	259.77	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	3456 +/- 259	75.4 +/- 0.8
4 (C)	48	960 +/- 62	95.4 +/- 0.8	0.161 / 0.219	174.16	197.06	453.17	100	82.1 +/- 0.3	0.32 +/- 0.16	0.13 +/- 0.07	0.27 +/- 0.12	0.25 +/- 0.16	7384 +/- 103	78.1 +/- 0.8
5	48	763 +/- 110	94.4 +/- 1.4	0.182 / 0.235	193.82	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	3930 +/- 195	78.2 +/- 0.8
6	48	1017 +/- 103	89.8 +/- 3.5	0.173 / 0.233	251.59	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	3760 +/- 205	76.4 +/- 0.7
7	48	1024 +/- 101	89.5 +/- 3.7	0.173 / 0.226	252.32	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	3925 +/- 229	76.5 +/- 0.9
8	48	965 +/- 119	91.1 +/- 2.6	0.173 / 0.225	242.10	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	3028 +/- 294	76.8 +/- 0.8

Read 2 (I)

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads PF (M)	Perfect Reads (M)	<=3 Err (M)	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	48	742 +/- 106	95.1 +/- 1.2	0.000 / 0.000	194.75	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	8700 +/- 335	0.0 +/- 0.0
2	48	1095 +/- 84	87.6 +/- 3.9	0.000 / 0.000	252.06	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	7732 +/- 259	0.0 +/- 0.0
3	48	1099 +/- 98	88.2 +/- 3.4	0.000 / 0.000	259.77	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	7478 +/- 232	0.0 +/- 0.0
4 (C)	48	960 +/- 62	95.4 +/- 0.8	0.000 / 0.000	174.16	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	106 +/- 127	0.0 +/- 0.0
5	48	763 +/- 110	94.4 +/- 1.4	0.000 / 0.000	193.82	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	2298 +/- 87	0.0 +/- 0.0
6	48	1017 +/- 103	89.8 +/- 3.5	0.000 / 0.000	251.59	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	1806 +/- 158	0.0 +/- 0.0
7	48	1024 +/- 101	89.5 +/- 3.7	0.000 / 0.000	252.32	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	1862 +/- 155	0.0 +/- 0.0
8	48	965 +/- 119	91.1 +/- 2.6	0.000 / 0.000	242.10	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	1826 +/- 143	0.0 +/- 0.0

Read 3

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads PF (M)	Perfect Reads (M)	<=3 Err (M)	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	48	742 +/- 106	95.1 +/- 1.2	0.183 / 0.191	194.75	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	7513 +/- 224	80.7 +/- 1.6
2	48	1095 +/- 84	87.6 +/- 3.9	0.174 / 0.184	252.06	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	6781 +/- 239	79.5 +/- 1.0
3	48	1099 +/- 98	88.2 +/- 3.4	0.166 / 0.182	259.77	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	7181 +/- 259	78.7 +/- 0.8
4 (C)	48	960 +/- 62	95.4 +/- 0.8	0.157 / 0.173	174.16	174.16	155.48	100	91.0 +/- 0.6	0.44 +/- 0.13	0.19 +/- 0.11	0.34 +/- 0.12	0.44 +/- 0.12	6209 +/- 253	79.3 +/- 0.8
5	48	763 +/- 110	94.4 +/- 1.4	0.186 / 0.187	193.82	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	7654 +/- 200	79.8 +/- 0.5
6	48	1017 +/- 103	89.8 +/- 3.5	0.116 / 0.132	251.59	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	7439 +/- 289	80.0 +/- 1.1
7	48	1024 +/- 101	89.5 +/- 3.7	0.158 / 0.181	252.32	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	7432 +/- 281	80.1 +/- 0.8
8	48	965 +/- 119	91.1 +/- 2.6	0.154 / 0.182	242.10	0.00	0.00	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	7670 +/- 238	82.2 +/- 0.8

Q_≥30のPass Filterの値

Lane1: 92.6% Lane3: 86.4% Lane5: 91.9% Lane7: 86.0%
 Lane2: 85.6% Lane4: 89.5% Lane6: 86.1% Lane8: 86.7% Ave.: 87.8%

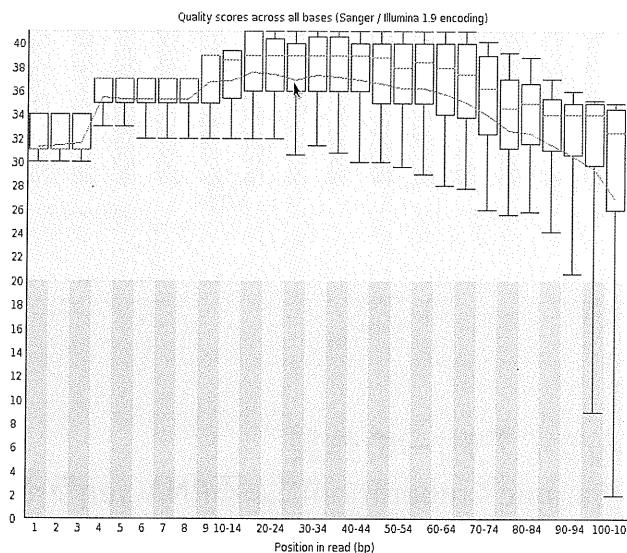
FastQC Report

金 18 11 2011
 lane1_Undetermined_L001_R1.fastq

Summary

- Basic Statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per base GC content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Kmer Content

Per base sequence quality



Produced by FastQC (version 0.10.0)

図9: HiSeq1000による大量並列シーケンシング