

図11: Mate pair ライブラリー作製結果

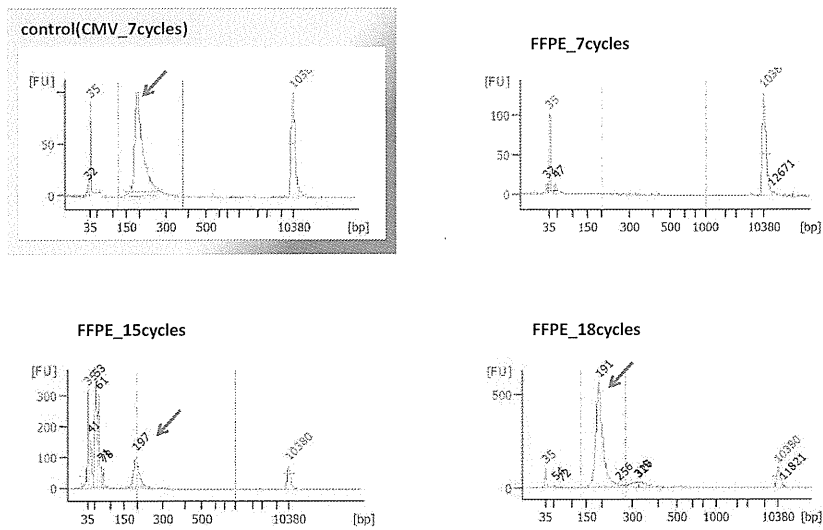


図12: FFPE由来gDNAライブラリー作製結果(Nextera\_1st)

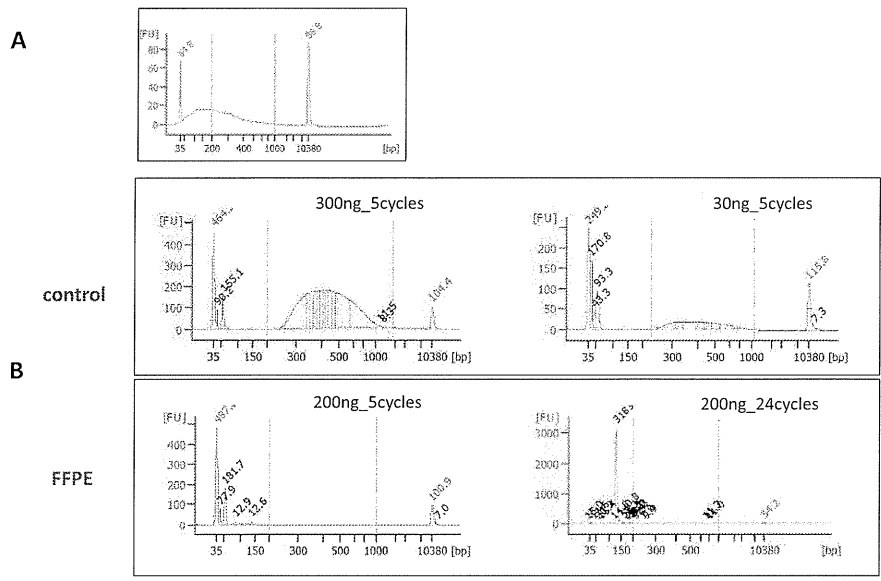


図13:FFPE由来gDNAライブラリー作製結果(TruSeq DNA)  
 A. FFPE由来ゲノムDNA (すでに断片化されている)  
 B. ライブラリー作製結果 (FFPE由来サンプルでは、アダプターダイマーのみが増幅)

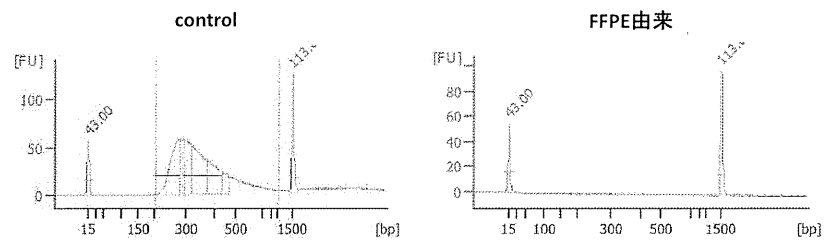


図14:FFPE由来ライブラリー作製結果(TruSeq RNA)  
 PCR 15cycles, Bioanalyzer DNA1000測定結果

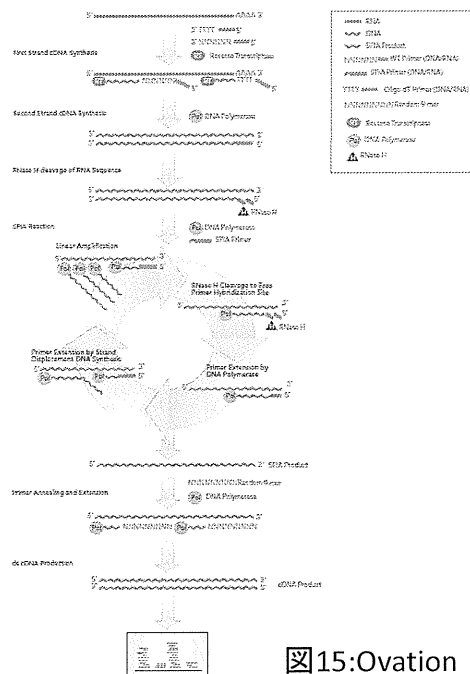


図15:Ovation RNA-Seq  
FFPE ワークフロー

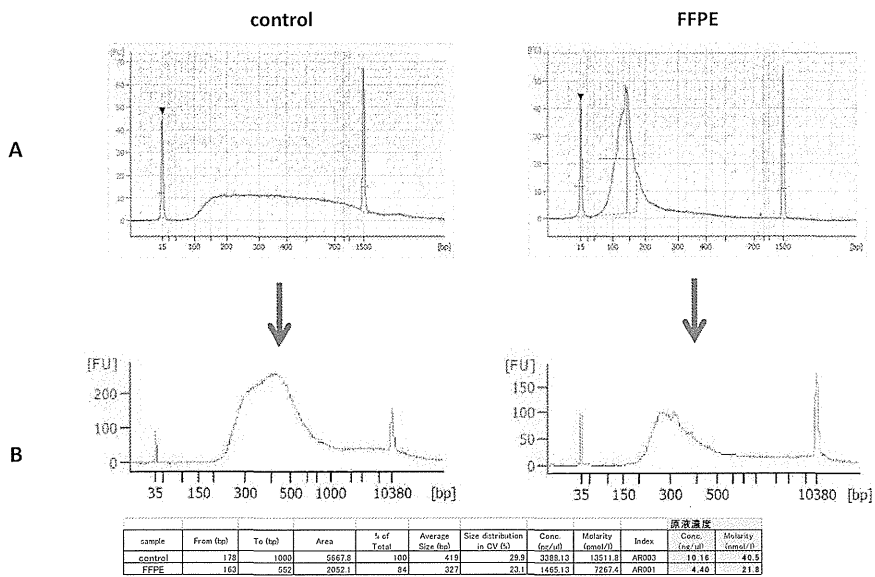


図16:FFPE由来ライブラリー作製結果(Ovation&TruSeq)

- A. cDNA合成結果(Ovation FFPE System)
- B. DNAライブラリー作製結果(TruSeq RNA)

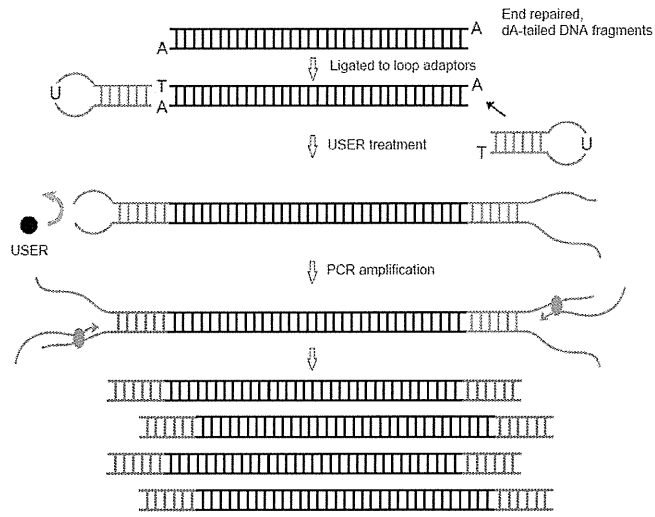


図17:ChIP Seq library 作製ワークフロー(NEB)

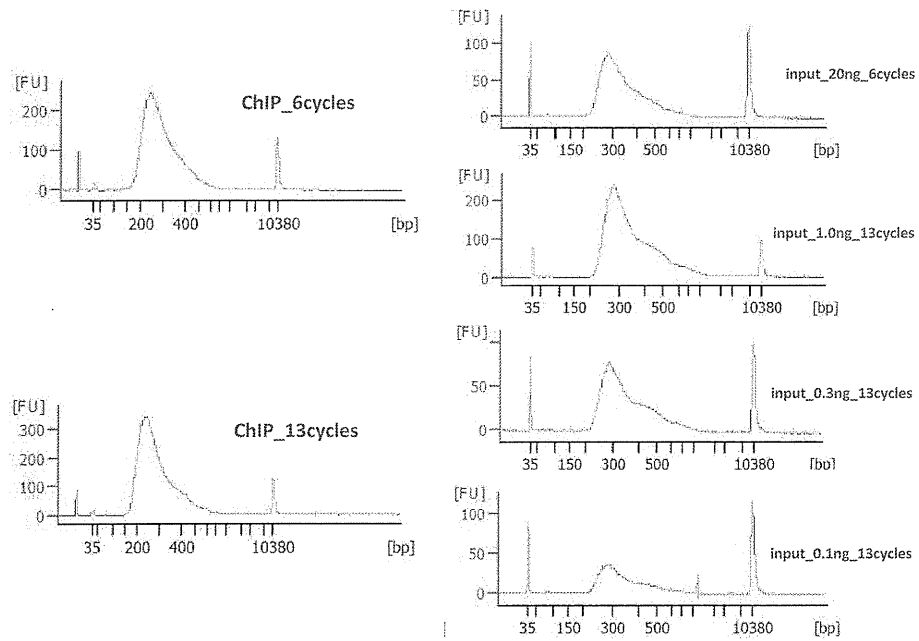


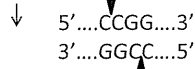
図18:ChIP Seqライブラリー作製結果(NEB Next)

- Start material; 100 ng genome DNA



- MspI digestion

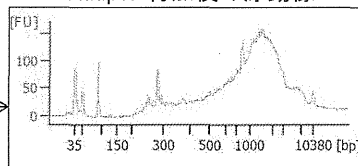
**Recognition Site:**



- Gap filling and A-tailing
- SPRI bead clean-up
- Multiplexed adaptor ligation
- Bisulfite conversion
- Amplification PCR
- Final SPRI bead clean-up

	Sample 1	Sample 2
Start gDNA 量	100 ng	100 ng
Adaptor ligation後のDNA量	84.4 ng	80.4 ng
Bisulfite conversion後のDNA量	35.0 ng	35.2 ng
Amplification後のDNA量	62.1 ng	69.3 ng

Adaptor付加後の泳動像



Library泳動像

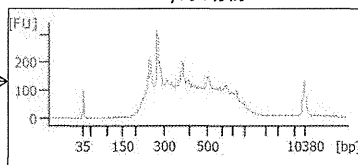


図19 : Protocol for gel-free multiplexed reduced representation bisulfite sequencing (mRRBS)

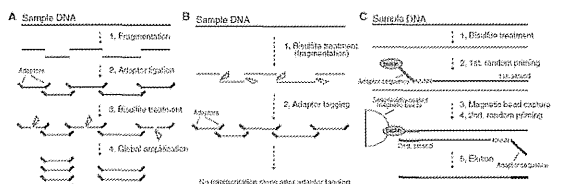


Figure 1. WGBS and PBAT. (A) Schematic of the conventional WGBS protocols. Bisulfite treatment induces adaptor tagging, thereby leading to bisulfite-induced fragmentation of adaptor-tagged template DNAs. (B) Schematic of PBAT strategy. Bisulfite treatment precedes adaptor tagging, thereby circumventing bisulfite-induced fragmentation of adaptor-tagged template DNAs. (C) Random priming-modified PBAT method. Two rounds of random priming on bisulfite-treated DNA generate directionally adaptor-tagged template DNAs.

Miura et al NAR2012

PBAT法ライブラリー収量結果

PBAT samples: NCHm001-m004 (2012/8/1)

Sample name	ng/ul	volume (ul)	yield (ng)	yield (%)	nM	pM	copies	copies/ul	10 <sup>9</sup> copies 相当 (ul)	4lanes 相当 (ul)
1~	0.140	20	2.80	2.80%	7.07E-01	707	8.52E+08	4.26E+08	2.3	9.4
2~	0.067	20	1.33	1.33%	3.39E-01	338	4.06E+08	2.03E+08	4.9	19.7
3~	0.210	20	4.21	4.21%	1.06E+00	1060	1.28E+09	6.40E+08	1.6	6.3
4~	0.470	20	9.40	9.40%	2.38E+00	2380	2.88E+09	1.43E+09	0.7	2.8
1+	0.141	20	2.83	2.83%	7.15E-01	715	8.60E+08	4.30E+08	2.3	9.3
2+	0.081	20	1.61	1.61%	4.03E-01	407	4.90E+08	2.45E+08	4.1	16.3
3+	0.192	20	3.84	3.84%	7.70E-01	770	9.20E+08	4.60E+08	2.2	8.8
4+	0.132	20	2.69	2.69%	6.81E-01	681	8.20E+08	4.10E+08	2.4	9.6
			100 ng						1.66fmol 相当	[1 lanes] [4 lanes]

average insert size 200bp (total 320 bp)での計算と思われる

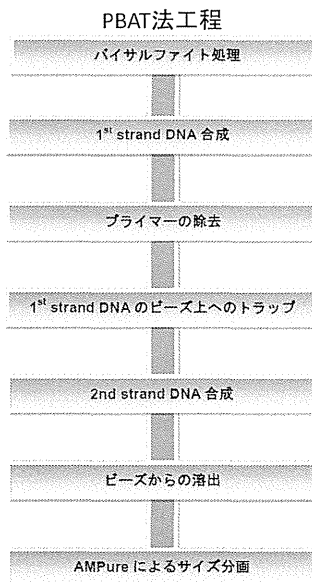


図20 : post-bisulfite adaptor tagging (PBAT)法

解析ソフトウェア	バージョン	本パイプラインにおける用途
FastQC	0.10.1	CASAVAから出力された配列データの品質確認
Trimmomatic	0.22	配列データのトリミングまたはクリーニング
BWA	0.6.2	リファレンス配列へのアラインメントおよびマッピング
Bowtie2	2.0.6	リファレンス配列へのアラインメントおよびマッピング
TopHat2	2.0.7	スプライシングを考慮したリファレンスへのマッピング
SAMtools	0.1.18	マッピング情報を保持するSAMおよびBAMファイルの管理
Picard	1.83	マッピング情報を保持するSAMおよびBAMファイルの管理
GATK	2.3.9	ヴァリアント検出およびフィルタリング、遺伝子型決定
snpEff	3.1	ヴァリアントのアノテーションおよび効果予測
IGVTools	2.1.24	VCF形式のファイルのインデックス作成
BMap	2012.09	PBAT法による配列データのマッピング

図21：解析パイプラインに組み込んだフリーソフトウェア一覧  
2012年2月現在、パイプラインに組み込まれているソフトウェアの名称、バージョン、その用途を示した。

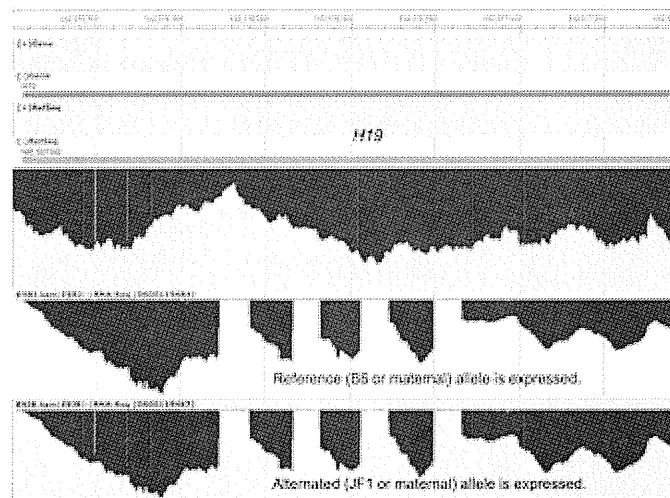


図23：全ゲノム・リシークエンシング解析例

母性発現を示すマウスH19における全ゲノム・リシークエンシング(上段)、B5xJFおよびJFxB5のRNA-seq(中段および下段)の結果をGenome Explorerで表示。紺色の帯の太さはそのサイトにおける相対的なカバレッジを示す。リファレンスに対して検出された多型は、塩基の種類によって赤、緑、青、橙色で、リード数に応じた長さで表示されている。

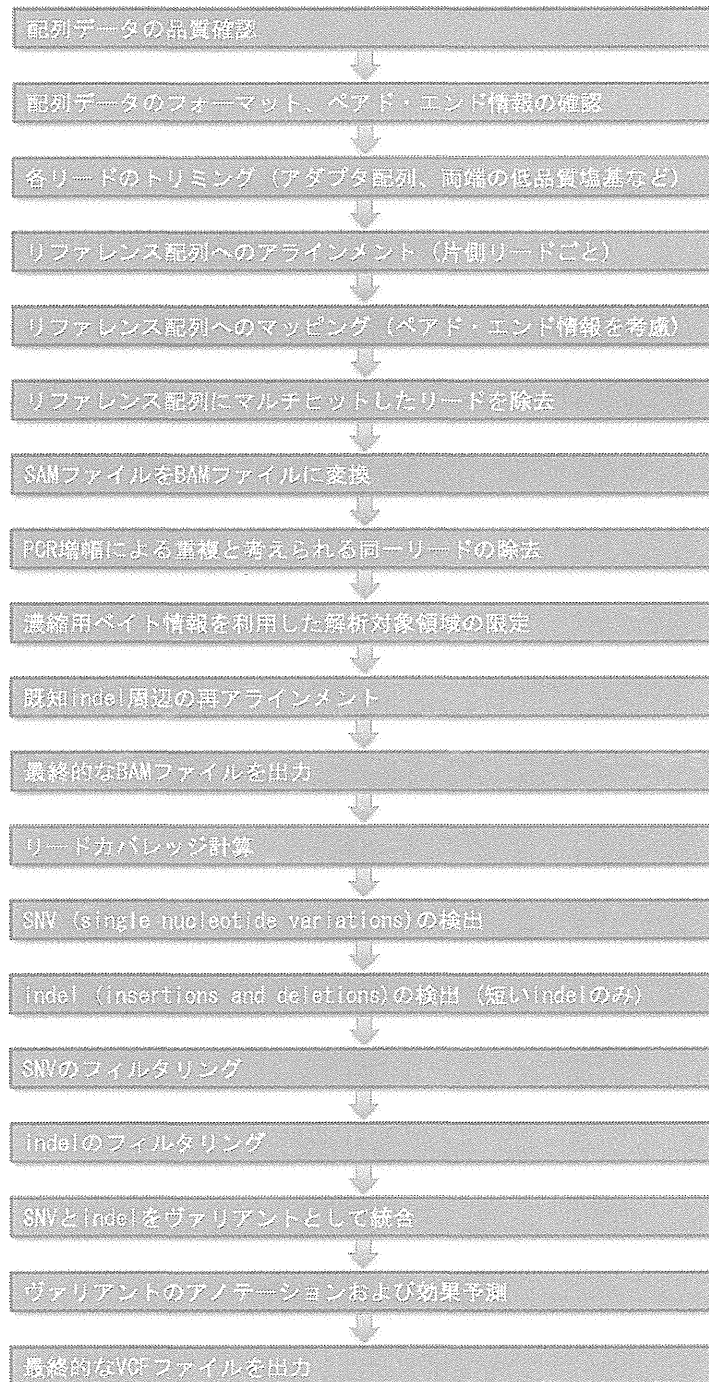


図22: エクソーム解析における処理の流れ

リードのトリミングまでは全ての解析において共通の処理であり、ターゲット・リシークエンシングまたは全ゲノム・リシークエンシングにおいてはエクソーム解析とほぼ同じ処理で解析を行うことができる。

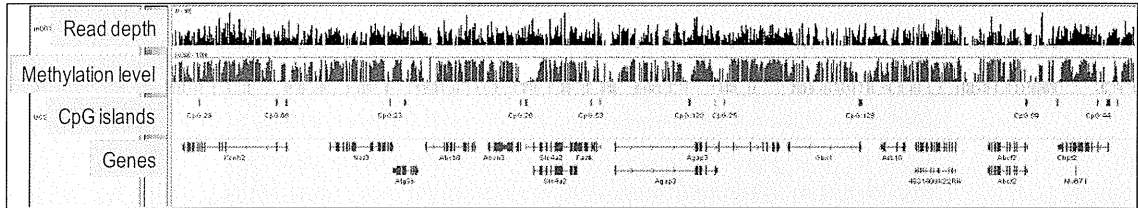


図24 : PBATデータ解析結果例 (マウスES細胞)



これまでの主な解析のまとめ	解析数	陽性者数	同定された異常の詳細
次世代シーケンス			
カスタムアンプリコンシーケンス			
性分化疾患	95	5	1例で既知病的変異、4例で候補変異を同定
カスタムターゲットエンリッチメント (sureselect)			
性成熟疾患	65	8	5例で既知病的変異、3例で候補変異を同定
SRS・原因不明成長障害	96	25	4例で既知病的変異、21例で候補変異を同定
性分化疾患	42	21	6例で既知病的変異、15例で候補変異を同定
カスタムターゲットエンリッチメント(haloplex)			
性分化疾患	95	準備中	
先天性甲状腺機能低下症	92	結果解析中	
SRS・原因不明成長障害	10	結果解析中	
性成熟疾患	25	結果解析中	1例で既知病的変異を同定
upd(14)mat症候群	12	結果解析中	
プラダーウィリー症候群	2	結果解析中	
エクソーム			
性分化疾患	20	結果解析中	
先天性甲状腺機能低下症	1家系	結果解析中	同定された未知変異の機能解析中
MODY	1家系	結果解析中	
upd(14)pat/upd(14)mat症候群	24	結果解析中	
腎機能異常	1家系	結果解析中	
稀な合併症を伴う先天奇形症候群	4	結果解析中	
MFB	30	結果解析中	
KUH	20	結果解析中	
免疫不全症候群	2家系	結果解析中	
ベックウィズヴィーデマン症候群	15	結果解析中	

表1: 解析検体リスト(次世代シーケンス)

これまでの主な解析のまとめ	解析数	陽性者数	同定された異常の詳細
450K Methylation			
偽性副甲状腺機能低下症Ib	2	2	疾患責任エピ変異(確認)
upd(14)pat/upd(14)mat症候群	7	7	疾患責任エピ変異(確認)
シルバーラッセル症候群	35	35	疾患責任エピ変異(確認)
プラダーウィリー症候群	3	3	疾患責任エピ変異(確認)
ベックウィズヴィーデマン症候群	7	7	疾患責任エピ変異(確認)
インプリンティング疾患疑い症例	10	4	疾患責任エピ変異(確認)
先天奇形症候群・精神発達遅延	15	結果解析中	
胎児発育遅延	20	結果解析中	
妊娠高血圧症候群	15	結果解析中	
日本人正常分娩集団	40		
SNP BeadChip			
日本人正常分娩集団	400		
インプリンティング疾患疑い症例	10	結果解析中	
原因不明先天奇形・精神発達遅延	20	結果解析中	
アレイCGH			
全ゲノムCGH			
性分化疾患	47	4	4例で病的欠失変異候補を同定
内分泌異常を伴う先天奇形症候群	14	4	3例で病的欠失変異候補を同定
偽性副甲状腺機能低下症Ib	1	1	機能未知領域に3Mbの欠失を同定
upd(14)pat/upd(14)mat症候群	7	4	14番染色体微小欠失を同定
シルバーラッセル症候群	22	0	
アロマターゼ過剰症	6	3	4例で病的変異候補を同定
染色体特定領域カスタムCGH			
upd(14)pat/upd(14)mat症候群	16	7	6例で染色体微小欠失、1例で微小重複を同定
プラダーウィリー症候群	13	0	
シルバーラッセル症候群	80	1	1例で病的変異候補を同定
ベックウィズヴィーデマン症候群	1	0	
IMAGE症候群	4	0	
片側肥大	4	1	1例で病的変異候補を同定

表2: 解析検体リスト (アレイ)

## II. 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）

分担研究課題「日本人ゲノムのプロフィール作成」

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

研究要旨：

標準的な日本人がもつゲノム多型を明らかにし、日本人一般集団を代表する対照リファレンスとして利用可能なデータセットを作成することを目的として研究をすすめている。本年度は、特にゲノムのコピー数が増減したゲノム構造変異 CNV (Copy Number Variation) についての情報収集を行った。コピー数の変化は、個人や人口集団間の差として塩基配列置換よりも高頻度で存在し、病因や疾患感受性とも関連することが明らかになっている。妊娠生殖歴に問題のなかった女性集団を対象として大規模遺伝子配列解析技術のひとつである SNP アレイ解析技術を駆使し、全ゲノムを網羅的に解析した日本人 CNV マップを作成した。収集された CNV は、日本人一般集団に保存された CNV であり、さらにヒトの発育や生殖に影響のない CNV と考えられる。本邦における小児科・産科領域の疾患研究での対照リファレンスとして有用なデータセットが確立されたと考えている。

A. 研究目的

標準的な日本人がもつ正常変異としてのゲノム構造変異を明らかにし、将来の小児科・産科領域における疾患研究の対照リファレンスとして利用可能な日本人ゲノムのデータセットを作成することを目的として研究をすすめている。本年度は、特にゲノムのコピー数の変化する CNV (Copy Number Variation) についての情報収集を行った。

近年、ヒトのゲノム解析が急速にすすみ、従来の遺伝学的手法では解析は困難であった欠失や重複といったコピー数の変化を含むゲノム構造異常が個人や集団間のゲノム情報の変化として、塩基配列置換よりも高頻度で存在し、様々な疾患と関連することが明らかになった。本年度は、新しい大規模遺伝解析技術のひとつである SNP タイピングを行うアレイ解析を駆使しゲノム構造変異の解析をめざした。より信頼性の高い解析スキームを構築し、日本人集団における CNV の位置と頻度のデータ収集を収集し、CNV マップを作成した。対象として妊

娠生殖歴に問題のない女性集団を用いたことから、日本人に存在する CNV のデータというだけでなく、日本人集団に保存されている CNV のデータであり、さらにヒトの発育や生殖に影響のない CNV を抽出することができたと考える。本解析により、今後の先天奇形症候群、生殖発生異常等の疾患にたいする対照リファレンスとして利用可能な CNV データセットが確立されたと考えている。

B. 研究方法

マイクロ・アレイの解析データからゲノム構造異常の発見パイプラインを検討し、ゲノム構造変異を網羅的に検索した。対象として明らかな先天疾患および不妊症を呈しておらず正常な満期産を経験した妊娠生殖歴に問題のない日本人女性サンプル 412 人を用いた。

サンプルは、2009 年より多施設共同研究として研究参加者から収集した血液を使用した。血液から QIASymphony SP システムを用いて DNA を抽出し、picogreen 定量キット

(Invitrogen) により定量、濃度の調整を行い以下の解析に供した。

マイクロ・アレイ解析は、イルミナ社の最新の稠密な SNP アレイチップである BeadChip (HumanOmni 2.5M)を用い、同社 iScan システムによってスキャンしてデータを収集した。本チップ上には、ゲノムに均等に配置されたおよそ 240 万のプローブが存在しており、1k 塩基毎のゲノムのコピー数についての情報が得られるものと期待される。(図 1: 解析の概要)

iScan システムによって収集されたスキャンデータは、アレイ上の 2 種類の蛍光についてそれぞれの蛍光強度のデータである。各プローブ位置にある SNP の遺伝子型、蛍光強度の量のデータを当該研究室の KaryoStudio、GenomeStudio ソフトウェア (イルミナ) をインストールしたワークステーション上で、各種アルゴリズムを用いて CNV を推定した。(図 2: CNV 検出) マイクロ・アレイ解析のデータから CNV を推定する際には、アレイ特有のエラーやノイズから生じるアーチファクトや偽陽性の問題がある。各アルゴリズムにおける解析パラメーターの検討と cnvPartition および隠れマルコフモデルを用いた PennCNV などコピー数の変化領域を同定する複数のアルゴリズムを利用し、それらのデータを統合する解析パイプラインを検討した。さらに、日本人集団に保存維持されたゲノム構造異常の稠密な解析を行い、自作スクリプトから、比較的頻度の高い CNV の位置と頻度情報のデータセットを抽出し CommonCNV として検討した。検出に使用するためワークステーションに導入したソフトウェアの一覧を表 1 に示す。

#### (倫理面への配慮)

本研究は、ヒト由来のゲノム解析・遺伝子解析研究であるためヒトゲノム・遺伝子解析研究に関する倫理指針をはじめとする関連法令と指針を遵守して計画され、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、採取時に各医療

機関で匿名化が実施された。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. SNP アレイ・データからコピー数多型データセットの収集・解析

複数の解析ソフトウェアを使用し (表 1)、アレイ解析データから全ゲノムにわたる CNV の解析フローを検討した。(図 3)

工程 1: 正常な妊娠分娩歴を有する日本人女性から血液を収集した。
工程 2: DNA を抽出し、SNP アレイ (Illumina 2.5M-8S beads chip) によりデータの収集。
工程 3: GenomeStudio プログラム上で、遺伝子方判定と蛍光強度データをもとにイルミナ社推奨の cnvPartition による CNV を検出
工程 4: 隠れマルコフモデルアルゴリズムを用いた PennCNV による解析を行い、より確からしいコピー数多型を検出 (図 3)
工程 5: 対象サンプルの高頻度に観察される CNV を日本人正常 common CNV として抽出

アレイから得られたデータからより確実な CNV を検出するために、複数のアルゴリズムと図 3 で示したパラメーターとを組み合わせた検出フローを確立した。複数のアルゴリズムを用いたコピー数多型は両者で共通して検出された CNV を採用するため、両者で予測された CNV の領域が 80%以上重複するものを検出 CNV として採用した。(図 4)

#### 2. 検出したコピー数多型データ

412 人のサンプルから、上記のフィルターをパスした 19,243 個の CNV が検出された。(図 5) 1 サンプル当たり検出個数は、平均 46.7 CNV (7 から 181 CNVs / サンプル) であり、CNV の平均的なサイズはおよそ 5 kb であった。ほとんどの CNV は、複数のサンプルで検出されている。検出された CNV 領域の重複をまとめた CNVR (copy number variable region) は、1,483 領域 (209.6Mb、ゲノムの 6~7%) であり、3,360 の遺伝子が影響を受けると予想された。

### 3. 日本人の”common CNVs”

対象集団に比較的頻度の高い CNV について検討した。対象の 5%以上 (21 人以上) に共通に観察された CNV を日本人の “common CNV” と定義した。図 6 に、19 番染色体で観察された CNV とその頻度を例示する。前述のように、複数のサンプルに重複して観察されるコピー数多型は多いが、common CNV はさらに 180 領域に絞られ、観察された CNVR の 1 割程度となった。(図 7)

### 4. ”common CNVs” の特徴

Common CNV は、110 個の染色体領域の重複と推定される Gain と 70 個の欠失と推定される Loss からなっていた。欠失は転写量やさらにもう一方アレルに遺伝子配列変異がある場合など、より表現型への影響が大きいと考えられる。欠失領域は、重複領域よりも遺伝子領域が含まれない傾向が確認された。(表 2)

アレイ解析による日本人一般集団の網羅的 CNV 解析は過去に 3 報の報告がある。(表 3) 過去の報告と比較し現在までに当研究計画で得られた CNV データは、より稠密なプローブが配置されたアレイを使用しており、さらに個々のサンプルの健康状態についての情報がそろっている。3 報中の 2 報では、健康状態の情報が不詳である HapMap を用いており、我々の対象は特に生殖に関する表現型の情報がそろっておりより一般集団としての品質が高い集団での解析と考えられた。当研究は、これまでの報告より多数の CNV 領域を発見しており、その一部は既知の報告にある CNV の Database である DGV (Database of Genomic Variants) にも登録のないものであった。

### D. 考察

ゲノム構造異常のひとつである CNV (Copy Number Variation コピー数多型) は、ゲノム情報の量的な変化をもたらす CNV に含まれるか断端位置によってさまざまな遺伝子の遺伝子発現量や遺伝子機能の破壊といった影響が起りうる。しかしながら、CNV はヒトの正常

多型としても存在し、ヒトの発育や生殖への影響が乏しいものが多数であると考えられている。本研究での対象サンプルは、出産、分娩に問題のない日本人集団であることに大きな特徴がある。取得された CNV が次世代への伝達に問題がなく、日本人集団に保存された正常変異である構造多型の収集につながる。さらに、異常妊娠歴を持たない成人から収集した検体を用いたことで、生殖発生異常等の疾患の対照リファレンスとしても利用可能な類例のない貴重なデータセットが作成されたと考えている。今年度の検討によって、本データセットが臨床検体に存在した病因欠失の詳細検討にも役立つことも見出されている。

染色体異常と疾患との関連は、多数報告されているものの、その発症機構にはまだまだ不明な点がある。たとえば、精神疾患の感受性やかかりやすさにかかわる CNV のなかには、健康集団に同じ CNV が発見され浸透率が低い変異であるのか、正常変異であるのか不明なものが報告されている。さらに、Williams 症候群をはじめとする染色体異常では、親世代にある染色体構造変異あるいは塩基配列の特徴が子の世代での病因染色体異常の新たな発生と関連するものが示唆されている。健康や生殖に大きな問題のない一般集団のデータを収集し、そのデータセットを整備することは将来の病因変異に役立つだけでなく、こうした基礎的な研究への資料にもなりうる。

ヒトのゲノムは多様であり、CNV をはじめとしたゲノム構造変異については、これまで配列情報から推定されたよりもはるかに多くの個人差や人種間の差があることが明らかになっている。大規模遺伝子解析技術の進展によって、各個人のゲノムレベルでの解析が可能となった現在、対照として適切な一般集団のゲノム情報の重要性が高まっている。これまで、ヒト・ゲノムに関する公開データは欧米人に由来するものに偏ってきたが、CNV の人種差の存在は個々の人種あるいは人口集団でのデータ整備の必要性を示唆している。東アジアの人口集団については、近年、東アジアの人口集団について、中国を中心にアジア人ゲノムに関する

報告が盛んになってきている。これらの研究によって、同じ東アジア人でも多様なゲノムが存在することが明らかになっており、日本人集団に特化した情報の収集は本邦におけるゲノム解析に非常に重要な基礎情報である。我々の CNV についてのデータセットは多方面における応用が期待される。今後、これらのデータのより効果的な公開方法を検討し、あらたなデータ公開体制をめざす。

## E. 結論

本年度、他分担課題の遂行にも不可欠な SNP ジェノタイピング・アレイ解析によるデータ取得ならびコピー数多型 (CNV) 解析から、標準的な日本人 CNV データベースを確立した。本研究では、イルミナ社 BeadChip (HumanOmni 2.5M) を、同社 iScan システムを用いたデータ取得系から、効率的で高い信頼性のある CNV の探索手法を検討した。これらは、今後の疾患罹患者群、不妊、不育といった生殖発育異常群の解析にも応用できる。

すでに 412 検体についてアレイ解析とその CNV データセットを取得した。さらに、このデータセットを利用し、病因変異の効率的な探索に応用できた例も経験している。今年度の経験は、アレイ解析データを活用した CNV 解析に役立つと考えられ、今後このデータをさらに整備しひろく公開することをめざす。

## F. 研究発表

### 1. 論文発表

なし

### 2. 学会発表

開山麻美、徐じゅひょん、小須賀基通、木田和宏、秦健一郎、右田王介、青木洋子、松原洋一、奥山虎之

「広範囲の欠失とミスセンスの複合ヘテロ接合体遺伝子変異を持つ異染色性白質ジストロフィーの出生前遺伝子診断」

第 54 回日本先天代謝異常学会 / 第 11 回アジア

先天代謝異常症シンポジウム 2012 年 11 月 16 日 (金) 岐阜 (じゅうろくプラザ)

右田 王介、前原 佳代子、岡村 浩司、西濱 啓一郎、中林 一彦、秦 健一郎 1

「日本人集団に共通するコピー数多型」

第 35 回日本分子生物学会 2012 年 12 月 11 日

(火) 福岡 (福岡国際会議場)

## H. 知的財産権の出願・登録状況

なし

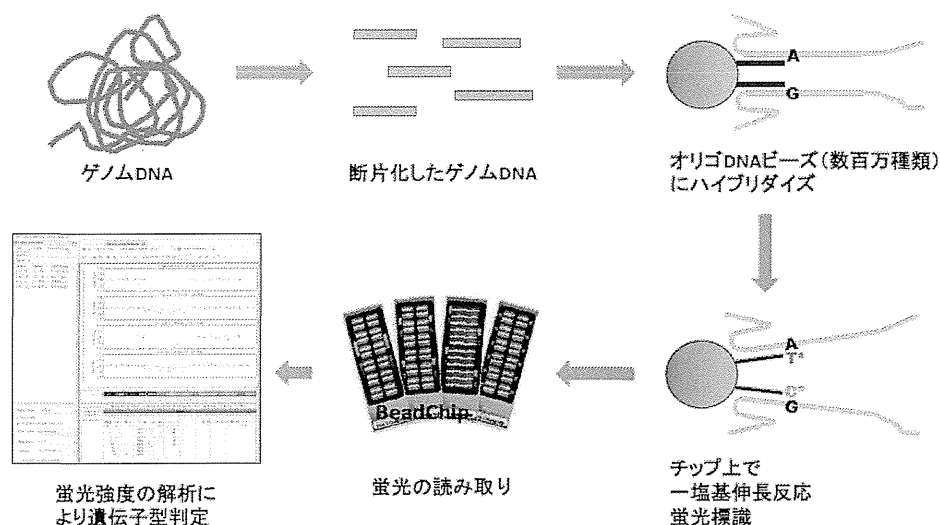
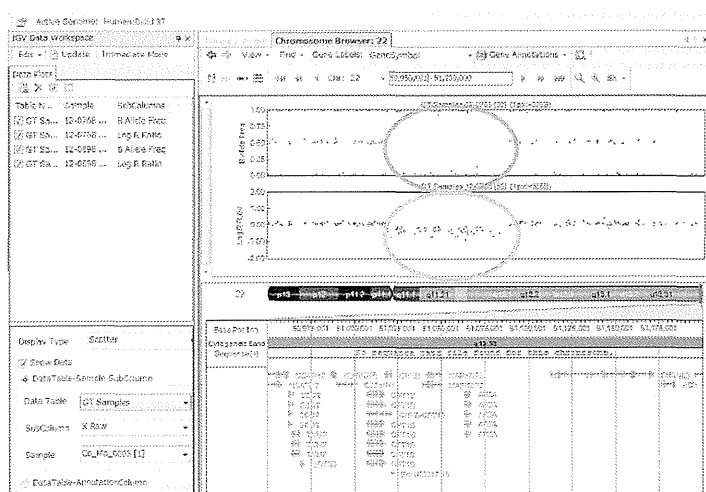


図1: BeadChipによる解析の概要

特定のSNPに対応するオリゴDNAプローブがビーズに結合しており、そのようなオリゴDNA結合ビーズ(数十~数百万種類)がチップ上に配されている。断片化したゲノムをオリゴDNAプローブとハイブリダイズさせ、個々のビーズでSNP部位の一塩基伸長反応を行う。取り込まれた塩基を二種類の蛍光で標識し、専用スキャナー iScan で各ビーズにおける蛍光強度を測定する。データ解析により個々のSNP部位の遺伝子型を判定する。



B Allele Frequency:  
蛍光強度から推定された  
B Allele (reference alleleと違う  
配列)の頻度。

Log R Ratio:  
プローブの蛍光強度。

図2: SNPアレイデータからのCNV検索

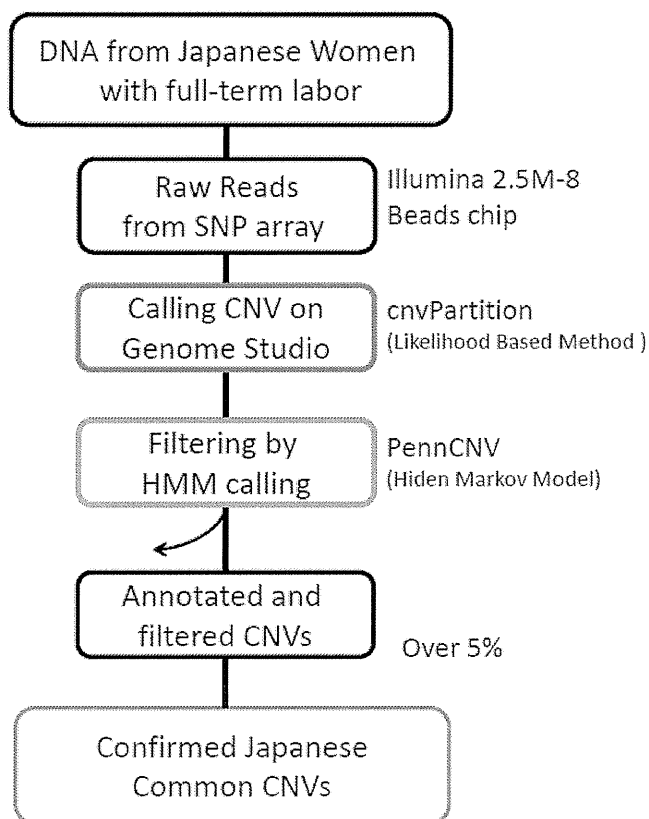
専用スキャナーで得られた各ビーズ(プローブ)における蛍光強度は、個々のSNP部位の遺伝子型(B Alleleの頻度情報として図示)と、その位置における蛍光強度とが解析される。ここでは、KaryoStudioの表示を示す。



cnvPartition
Confidence Threshold: 60
Minimum probe count: 3
pennCNV
Confidence Threshold: 0
Minimum SNP count: 3

### 図3: 解析処理の流れ

本研究では、正常な妊娠分娩歴を有する日本人女性を解析の対象とした。SNPアレイ (Illumina 2.5M-8S beads chip) のデータを、2つの異なるアルゴリズムによって解析し、さらに対象サンプルの5%以上に高頻度に観察されるCNVを日本人正常CommonCNVとして検討を加えた。



### 表1: 導入ソフトウェア

データ解析ツール	バージョン
<MS-Windows System>	
• KaryoStudio	Ver 1.4
• GenomeStudio	V2011.1
- CNVPartition	Ver 3.2.0
- PennCNV	Ver 2009Aug27
• Cygwin	1.5
- Perl	V5.10.1
<Linux System>	
• CentOS	6.2
- Perl	V5.10.1
- PennCNV	Ver 2009Aug27

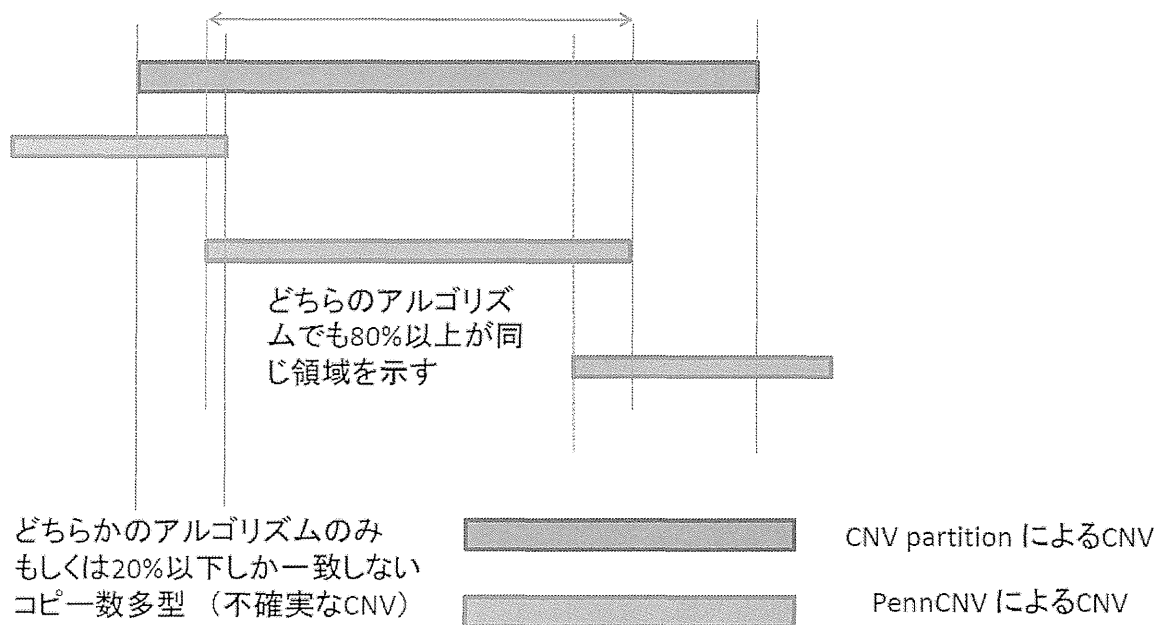
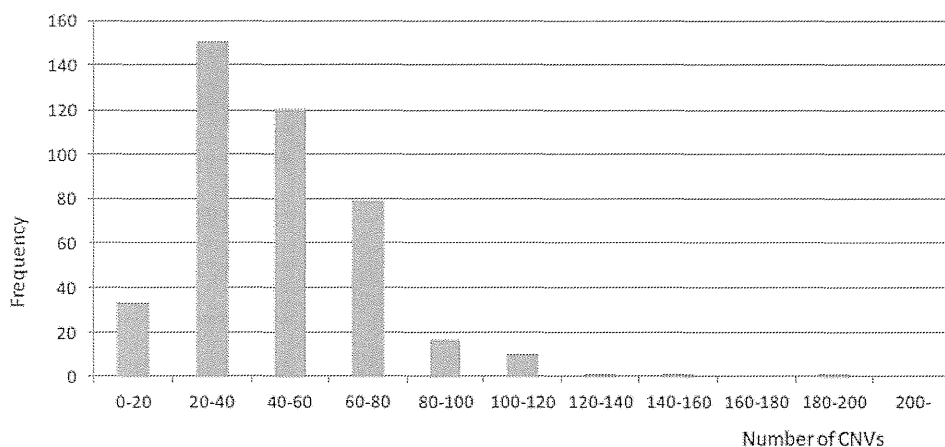
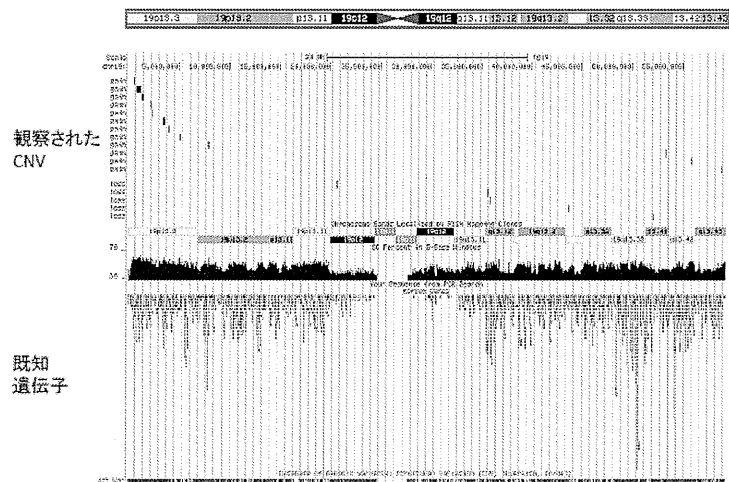


図4: 2つのアルゴリズムによるコピー数多型領域の判定



- 図5: 検出されたCNV
- 19,243 候補CNV / 412人
  - 1,483 領域 (209.6Mb)
  - 3,360 遺伝子
  - 一人あたりに観察されたCNV数(平均46.7)

図6: 頻度 5% 以上の CNV マップ:  
Chromosome 19



19番染色体に観察されたCommon CNVを示す。  
上段に観察されたCNVの位置、中断に遺伝子配列の位置を示す。最下段は、頻度情報を加味したすべてのCNV位置情報を赤色の線がGain、黒色の線がLossとして示した。412人中の5%から21人以上で観察されたCNVを上段に示すCommonCNVとした。

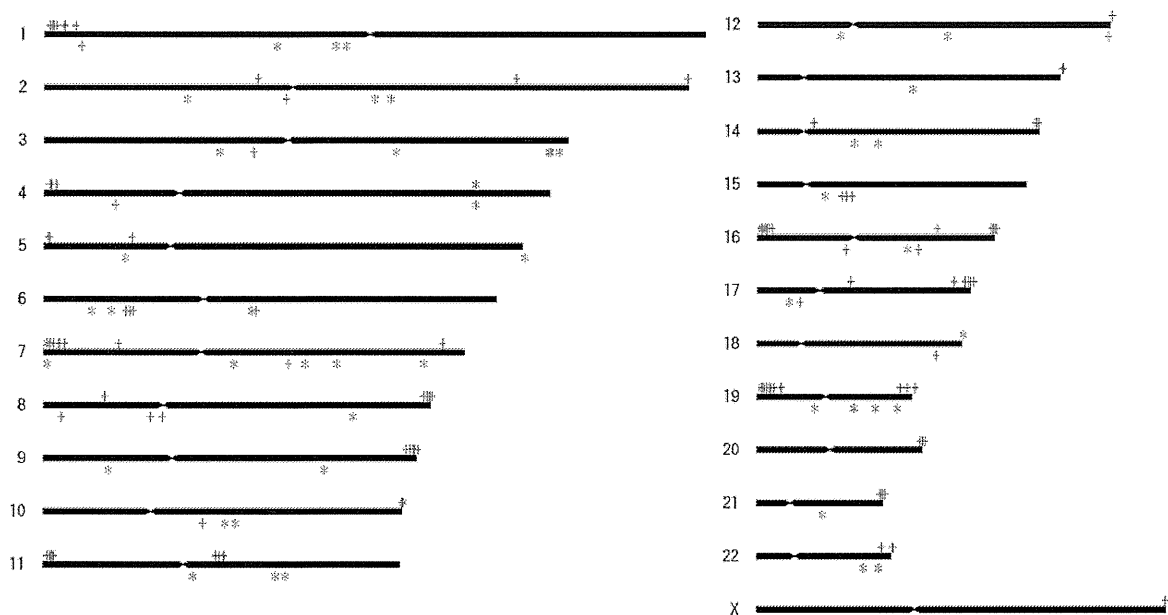
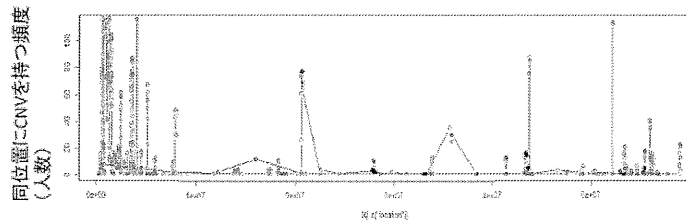


図7: 日本人コピー数多型(頻度5%以上)の染色体上の位置

SNPアレイ解析から確認されたcommon CNV 位置を示す。  
赤字がLossを、青字がGainを示す。

## 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

研究分担者 中村浩幸 国立成育医療研究センター研究所母児感染研究部

研究要旨：先天性サイトメガロウイルス（CMV）感染症の病態解明を目的として、CMV 感染ヒト細胞を対象に次世代シーケンサーを用いた網羅的発現解析を行った。その結果、蛋白質コード領域をはじめ、多様な配列領域において転写産物が同定され、それらの発現動態も明らかとなった。また、transcript variant に関する解析から、CMV 感染によって発現変動する新規 isoform の存在も示唆された。さらに、Gene ontology 解析から、当該疾患の病態解析を進める上で着目すべき機能遺伝子群に関する情報が得られた。これらの結果より、次世代シーケンサーを用いた網羅的発現解析は、当該疾患の病態解析を進める上で有用な情報を提供し得る解析法と考えられた。

### A. 研究目的

本研究の目的は、次世代シーケンサーを用いた網羅的発現解析によって、サイトメガロウイルス(Cytomegalovirus; CMV)が引き起こす先天性CMV感染症における病態形成のメカニズムを明らかにすることである。

先天性CMV感染症は、先天感染症として最も頻度が高く、本邦では全出生の約0.31%、年間3,300人ほどの先天性CMV感染児が発生すると推定される。近年は、若年層の妊娠可能女性における抗CMV抗体保有率が低下傾向にあることから、当該疾患の発生頻度は今後さらに増加することが危惧されている。先天性CMV感染児の約1割を占める症候性感染児は、脳障害をはじめ種々の臓器障害を呈し、重篤な経過をたどるケースが多い。一方、先天性CMV感染児の約9割を占める無症候性感染児は、一般に予後良好とされるものの、約10～15%の無症候性感染児は精神発達遅滞や難聴などを遅発性に発症するとされる。このように、先天性CMV感染症は、神経学的合併症など長期間にわたる後遺症をとともなう場合があり、社会的インパクトの極めて大きい疾患と言えるが、その病態形成機構については不明な点が多い。

CMVは多種類のヒト細胞に感染することが知られており、このことが当該疾患の多様な病態と深く関連すると推測される。そのため、ヒ

ト細胞を用いたCMV感染モデルは、当該疾患における病態を理解する上で有用と考えられる。本研究は、ヒト細胞を用いたCMV感染モデルにおいて、次世代シーケンサーを用いた網羅的発現解析を行い、その知見をもとに当該疾患の病態形成メカニズムの理解を深めることを目的とする。本年度は、ヒト神経幹細胞を用いたCMV感染モデルにおいて網羅的発現解析を行った。

### B. 研究方法

#### 1) ヒト神経幹細胞株を用いたCMV感染モデル

ヒト線維芽細胞株 MRC5 由来の人工多能性幹(induced pluripotent stem; iPS)細胞を用いて、Smad シグナル 2 段階阻害法により神経幹細胞(NSC-iPS)を作製した。CMV 液は、CMV をヒト不死化線維芽細胞 BJ1 に接種し増殖させた後の培養上清より調整し、CMV 感染実験に供した。

#### 2) RNA-seq 解析：

RNA-seq ライブラリーは、1.0  $\mu$ g の total RNA を用いて、TruSeq RNA Sample Preparation V2 法により作製した。ライブラリー定量には、Kapa Biosystems 社 Library Quantification キットを用いた。シーケンシングデータの取得は、HiSeq1000 シーケンサーで行い、データ解析ソ