

明な脳室拡大を認め、先天性水頭症が疑われた（図 1）。標的遺伝子群パネル解析の結果、L1CAM exon7, c:956G>T, p:G319V を同定し、Sanger シークエンス法で遺伝子異常が確定された（図 2）。本家系において両親の同部位を検索した結果、いずれも変異を認めず、患児で認められた異常は de novo 変異であると確認された（図 2）。

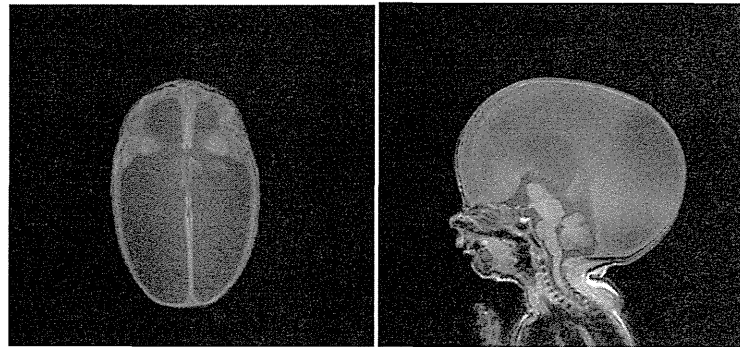


図 1 : 頭部 MRI 画像

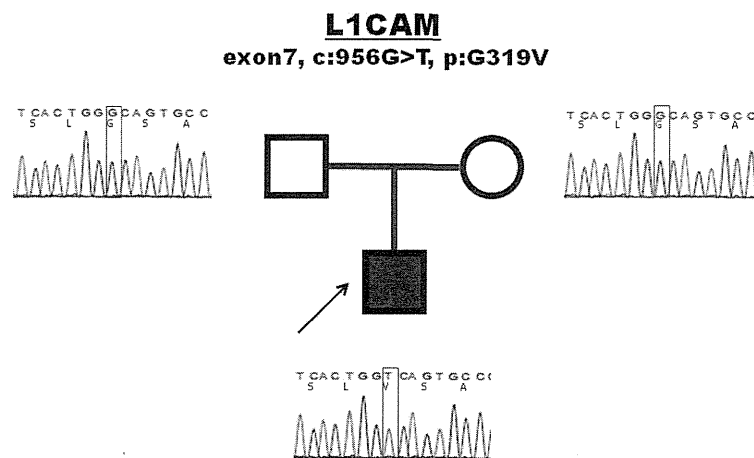


図 2 : 家系図とシークエンス解析結果 (Sanger シークエンス法)

## 2. 全エクソーム解析

神経チーム全体では 16 家系 43 検体のシークエンスデータの取得を行った（表 1）。その中で、データ解析が先行して進められた 11 家系 28 検体において、報告書作成の段階で疾病の責任遺伝子あるいは候補遺伝子の同定に至ったのは 4 家系であり、解析を行った全家系における候補遺伝子同定率は 36.4%であった。

（代表的症例 2）

症例は家系内男児に先天性水頭症発症の家族歴があり、X 連鎖性先天性水頭症を疑い、L1CAM 遺伝子解析を実施したが、明らかな異常が認められなかった男児である。高度の水頭症と発達障害を認めた。全エクソーム解析の結果、X 染色体上の遺伝子 X にミスセンス異常が認められ、

母親は同部位のヘテロ接合体を有することが示唆された。Sanger シークエンス法の結果、遺伝子異常が確定された。同定された候補遺伝子は、文献上、先天性中枢神経奇形症候群の発症との関連性の報告が無い新規遺伝子であり、現在、家系内の他のメンバーの解析を継続して実施中であり、先天性水頭症発症との関連性を注意深く考察中である。

## D. 考察

先天性中枢神経奇形症候群の病態解明において、その原因遺伝子の同定は重要な位置を占める。現在までにその原因遺伝子として、多数の遺伝子異常が報告されており、これら研究によって当該領域の研究は大きく進展した。しかし一方で、依然として多くの症例、とりわけ家族

歴が明確でない孤発例の多くはその原因遺伝子が不明なままであり、その病態も十分に解明されていない。このような原因遺伝子が不明な症例の遺伝子解析に関しては、従来法（Sanger シークエンス法）を用いて多数の候補遺伝子のシークエンス解析を個別に行うことは、時間的・労力的・コスト的に大きな負担を要する解析であり、このことが新たな遺伝子異常検索を困難にしてきた。しかし、次世代シークエンス法を用いることで、多数の標的遺伝子の検索を一度に短時間で実施できるようになったのみならず、全エクソンレベルでの解析、さらに全ゲノムレベルでの解析も実現可能となった。

24年度は、まず23年度に着手した284遺伝子の標的遺伝子群パネル解析を本格的に実施した。従来から実施してきた個別の遺伝子解析では原因遺伝子が不明であった症例を対象とした解析の結果、その原因遺伝子が判明し、確定診断に至る家系が得られた。以上の結果は、原因不明の先天性中枢神経奇形症候群家系の原因遺伝子検索における、標的遺伝子群パネル解析の有用性を証明した結果であると考えられる。しかし一方で、原因遺伝子の同定率は現時点では約10%程度で、予想に反して必ずしも高くはないことが判明した。標的遺伝子群パネル解析は、解析対象家系の原因遺伝子とそのパネル中に存在する場合は、一度のシークエンス解析で多数の候補遺伝子の中から原因遺伝子を同定することが可能であり、効率的に原因遺伝子の同定に成功する手法である。しかし、原因が全く未知の疾患・家系の場合、その原因遺伝子がパネル中に存在しない場合も十分想定され、そのような場合は限定された数の遺伝子の検索では、原因遺伝子を同定することは困難であると考えられる。よって、標的遺伝子群パネル解析は未知の原因遺伝子探索には必ずしも効率的な手法ではなく、その特性から、1つの病気に複数の原因遺伝子が存在する場合に、その遺伝子型を決定する際に使用することに適した解析アプローチであると考えられた。

一方、全エクソーム解析に関しては、詳細な解析を進めた家系数は少ないながらも、原因遺伝子あるいは候補遺伝子の同定は36.4%で成功しており、未知の原因遺伝子探索には非常に有用な手法であることが示唆された。現在、標的遺伝子群パネル解析と全エクソーム解析にコスト面では大きな差は無くなりつつあり、今後、未知の原因遺伝子探索には全エクソーム解析を主体に実施していくことが適当であると判断され、最終年度の25年度は更に解析を進めていく予定である。

## E. 結論

次世代シークエンス法を用いた先天性中枢神経奇形症候群の原因遺伝子検索を本格的に開始

した。標的遺伝子群パネル解析および全エクソーム解析のいずれにおいても原因遺伝子同定に成功した家系が存在するが、その効率は全エクソーム解析のほうが高く、原因不明の先天性中枢神経奇形症候群家系の原因遺伝子検索においては、全エクソーム解析がより効率的な解析手法であることが示唆された。25年度は全エクソーム解析を主体、さらに研究を進めていく予定である。

## F. 研究発表

### 1. 論文発表

- 1) Takenouchi T, Nakazawa M, Kanemura Y, Shimozato S, Yamasaki M, Takahashi T, Kosaki K. Hydrocephalus with Hirschsprung Disease: Severe End of X-linked Hydrocephalus Spectrum. *Am J Med Genet A* 158A(4):812-815, 2012
- 2) Kanematsu D, Kanemura Y: Cellular properties of mesenchymal cells derived from the decidua of human term placenta and their applications for regenerative medicine. *Biomaterials for stem cell therapy: State of Art and Vision for the Future* (Edited by Loredana De Bartolo, Augustinus Bader) P.228-248, CRC Press, FL, USA, 2013
- 3) Itoh K, Pooh R, Kanemura Y, Yamasaki M, Fushiki S. Hypoplasia of the spinal cord in a case of fetal akinesia/arthrogryposis sequences. *Neuropathol Appl Neurobiol*, in press

### 2. 学会発表

- 1) 金村米博：iPS細胞研究を応用した難治性脳形成障害症の分子病態解析と新規治療技術開発の可能性。第54回日本小児神経学会総会、2012年5月17日；札幌
- 2) 根岸豊、服部文子、安藤直樹、宮冬樹、角田達彦、北井征宏、荒井洋、岡本伸彦、加藤光広、山崎麻美、金村米博、小崎健次郎、齋藤伸治：網羅的遺伝子解析により難治性てんかん患児に同定された複合ヘテロSCN2A遺伝子変異。日本人類遺伝学会第57回大会、2012年10月27日；東京
- 3) 金村米博：iPS細胞を応用した小児神経難病の研究。第5回南大阪遺伝診療研究会、2012年9月28日；大阪

## G. 知的財産権の出願・登録状況（予定を含む。）

1. 特許取得  
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（難病関係研究分野）  
分担研究報告書

次世代シーケンサーを用いた骨系統疾患の原因遺伝子群の網羅的解析系の設計

分担研究者 小崎 里華  
独立行政法人国立成育医療研究センター 器官病態系内科部遺伝診療科 医長

### 研究要旨

本研究は次世代シーケンサーを用いることにより、臨床診断が困難である先天奇形症候群を遺伝子診断により原因を特定し、合併症を回避することによって質の高い診療を提供することを目的としている。初年度は、遺伝子診断を行う臨床的意義が高いと思われる疾患を選定し、ターゲット・エンリッチメントのためのシステムを設計した。

今年度は約 300 個の骨系統疾患および関連する遺伝子の蛋白質コーディング全領域を網羅的に解析する系を設計し解析体制を整備した。

### A. 研究目的

本研究は次世代シーケンサーを用いることにより、臨床診断が困難である先天奇形症候群を遺伝子診断により原因を特定し、合併症を回避することによって質の高い診療を提供することを目的としている。今年度は、臨床診断が困難であり遺伝子診断を行う臨床的意義が高いと思われる骨系統疾患の遺伝子診断を目的としてターゲット・エンリッチメントのためのシステムを設計した。

### B. 研究方法

既知の骨系統疾患の原因遺伝子と、既知疾患原因遺伝子が属する分子パスウェイの上流・下流の遺伝子を併せて300個程度の遺伝子をリストアップした。これらの遺伝子を標準的な登録番号（カリフォルニア大学サンタクルツ校ゲノムデータベース UCSD ID）に変換した患者ゲノムDNAからこれらの遺伝子に対応する領域を効率的に回収するためのオリゴヌクレオチドアレイを設計した。具体的にはAgilent社のカスタムマイクロアレイを用いることとし、同社のマイクロアレイ設計ソフトウェアである「eArray」を使用した。

eArrayで自動設計したカスタムキットにはGC含量が高い領域や反復配列の混入を避けるために、複数のエクソン領域でプローブが未設計であった。そこで、プローブの配列とUCSCのエクソン位置情報を比較し、可能な限り未設計領域へ新たにプローブを追加設計し、Agilent社に合成を依頼した。

疾患原因遺伝子は、

Warman ML, Cormier-Daire V, Hall C, Krakow D, Lachman R, LeMerrer M, Mortier G, Mundlos S, Nishimura G, Rimoin DL, Robertson S, Savarirayan R, Sillence D, Spranger J, Unger S, Zabel B, Superti-Furga A. 2011. Nosology and classification of genetic skeletal disorders: 2010 revision. Am J Med Genet Part A 155:943?968.

によった。

### ②患者由来ゲノムDNAの前処理

解析検体は、インターカレーション法による精密なDNA濃度測定を行い、Covaris社製超音波破碎機を用い、至適条件下で、約150bpへ断片化する。これをベックマン社AMPure磁性ビーズを用いて精製しSureSelectターゲットエンリッチメントライブラリーとハイブリダイズさせ、各検体の解析対象となる骨系統疾患原因遺伝子領域を濃縮した。各検体はインデックスタグを付加し増幅し、イルミナ社製次世代シーケンサーMiSeqを用いて塩基配列データを創出した。品質チェックを経て参照ゲノムDNA配列上へのマッピングを行い、検体毎に遺伝子変異を同定した。本遺伝子解析研究計画は、研究開始に先立ち各研究施設での倫理審査で承認をうけた。また本研究は、各検体をご提供下さった患者、ご家族の同意下のもと実施した。

### ③次世代シーケンサーによる解析

デスクトップ型次世代シーケンサーであるMiSeqを併用した。

### ④変異の同定のためのデータの可視化

患者検体から得られた DNA 配列データをヒトの標準的なゲノム配列に整列（アラインメント）し、アラインメントのあとのデータ解析にはグラフィカルユーザーインターフェースを有するプログラム NextGENe および varChifter を使用した。

（倫理面への配慮）

個人情報保護に関する法律を踏まえて研究を実施した。遺伝子解析に際しては文部科学省・厚生労働省・経済産業省「ヒトゲノム・遺伝子解析研究に関する倫理指針」を遵守し、倫理委員会の承認下に研究を実施した。

## C. 研究結果

### ①解析系の設計

約300個の骨系統疾患原因遺伝子および関連する遺伝子の蛋白質コーディング全領域を網羅的に解析する系を設計できた。これら遺伝子の解析体制を整備し、具体的な遺伝子解析に着手した。

②次世代シーケンサーからの粗配列の出力ターゲット濃縮（アジレント社SureSelect）とデスクトップ型シーケンサー（イルミナ社MiSeq）を用いて300個の程度の対象疾患と関連する可能性のある遺伝子を網羅的に解析した。次世代シーケンサーによる変異探索では変異の検出におよそ平均深度50程度のリードが必要とされているが、患者当たり平均深度60を得る事ができた。

### ③粗配列のアラインメント

次世代シーケンサーからはDNA配列と塩基毎の精度がfastqフォーマットで出力される。そこで、得られたfastqファイルをヒト参照配列にSoftGenetics社のソフトウェアおよびオープンソースソフトウェアであるbwa (<http://bio-bwa.sourceforge.net/>) によりアラインメントした。

④変異探索と原因候補遺伝子のリストアップアラインメントの結果をSoftGenetics社のソフトウェアおよびオープンソースプログラムであるvarChifterにより可視化して解析した。campomelic dysplasiaの原因遺伝子であるSOX9遺伝子の変異により異なる疾患であるStickler症候群の表現型が生じることなどの新発見が得られ、既知遺伝子の上流・下流の遺伝子を併せて解析するアプローチの有効性が示された。

## D. 考察

今年度は、次世代シーケンサーを用いて骨系統疾患遺伝子群の変異解析系を完成した。臨床症状から、特定の疾患が疑われ、当該疾患の原因とされる遺伝子にナンセンス変異やフレームシフト変異など、明らかに病的意義がある変化については、解釈に困難を要することは無い。しかし、鑑別疾患に含まれる複数の疾患の原因遺伝子にミスセンス変異が複数個同定された場合には、病的意義の解釈が容易ではない。

特に骨系統疾患の原因である細胞外マトリックスタンパク等はコード領域が大きく、疾患と直接に関係のないvariantも多く含まれるため、遺伝子診断のみで診断がつかない場合が少なくない。X線写真の読影結果を踏まえ、最終的な診断を決定する必要がある。日本人におけるアミノ酸置換のレパートリーを蓄積し、研究者間で共有してゆく必要があると考えられた。また、臨床診断の重要性が改めて認識された。

## E. 結論

本年度の成果によりカスタムキットとMiSeqにより、300個の骨系統疾患および関連する可能性のある多数の遺伝子を網羅的に解析する事が可能であると証明できた。

## F. 研究発表

### 1. 論文発表

- 1) Takenouchi T, Nishina S, Kosaki R, Torii C, Furukawa R, Takahashi T, Kosaki K. Concurrent deletion of BMP4 and OTX2 genes, two master genes in ophthalmogenesis. *Eur J Med Genet.* 56(1):50-3.2013
- 2) Takenouchi T, Yagihashi T, Tsuchiya H, Torii C, Hayashi K, Kosaki R, Saitoh S, Takahashi T, Kosaki K. Tissue-limited ring chromosome 18 mosaicism as a cause of Pitt-Hopkins syndrome. *Am J Med Genet A.* 158(A)(10):2621-3.2012
- 3) Takenouchi T, Okuno H, Kosaki R, Ariyasu D, Torii C, Momoshima S, Harada N, Yoshihashi H, Takahashi T, Awazu M, Kosaki K. Microduplication of Xq24 and Hartsfield syndrome with holoprosencephaly, ectrodactyly, and clefting. *Am J Med Genet A.* 158(A)(10):2537-41.2012
- 4) Kosaki R, Nagao K, Kameyama K, Suzuki M, Fujii K, Miyashita T. Heterozygous tandem duplication within the PTCH1 gene results in nevoid basal cell carcinoma syndrome. *Am J Med Genet A.* 158(7):1724-28.2012
- 5) Yagihashi T, Kosaki K, Okamoto N, Mizuno S, Kurosawa K, Takahashi T, Sato Y, Kosaki R, Age-dependent change in behavioral feature in Rubinstein-Taybi syndrome. *Congenit Anom (Kyoto).* 52(2):82-6.2012
- 6) Miyazaki O, Nishimura G, Sago H, Horiuchi T, Hayashi S, Kosaki R. Prenatal diagnosis of fetal skeletal dysplasia with 3D CT. *Pediatr Radiol.* 42(7):842-52.2012
- 7) Kosaki R, Kaneko T, Torii C, Kosaki K. EEC syndrome-like phenotype in a patient with an IRF6 mutation. *Am J Med Genet A.* 158A(5):1219-20.2012
- 8) Nishina S, Kosaki R, Yagihashi T, Azuma N, Okamoto N, Hatsukawa Y, Kurosawa K, Yamane T, Mizuno S, Tsuzuki K, Kosaki K. Ophthalmic features of CHARGE syndrome with CHD7 mutations. *Am J Med Genet A.* 158A(3):514-8.2012
- 9) Okamoto N, Hayashi S, Masui A, Kosaki R, Oguri I, Hasegawa T, Imoto I, Makita Y, Hata A, Moriyama K, Inazawa J. Deletion at chromosome 10p11.23-p12.1 defines characteristic phenotypes with marked midface retrusion. *J Hum Genet.* 57(3):191-6.2012
- 10) Kasahara M, Sakamoto S, Kanazawa H, Karaki C, Kakiuchi T, Shigeta T, Fukuda A, Kosaki R, Nakazawa A, Ishige M, Nagao M, Shigematsu Y, Yorifuji T, Naiki Y, Horikawa R. Living-donor liver transplantation for propionic acidemia. *Pediatr Transplant.* 16(3):230-4.2012

## 2. 学会発表

- 1) Shimizu A, Torii C, Suzuki N, Mutai, J, Kudoh H, Kosaki R, Mmatsunaga T, Kosaki K. Rapid and efficient mutation in the hundreds of target genes by bench-top next generation sequencer with custom target capture method. American Society of Human Genetics, 2012
- 2) 岡田 朋美 大柴 葉子 佐々木 愛子 谷口 公介 杉林 里佳 住江 正大 和田 誠司 柿島 裕樹 小須賀 基通 小崎 里華 小澤 伸晃 左合 治彦: 羊水検査において FISH 法 (AneuVysion) と G-band 法で異なった核型結果が得られた 2 症例 第 57 回 日本人類遺伝学会 2012.10.26
- 3) 藤田 秀樹 小崎 里華: 当院で経験した未診断の発達遅滞・先天性多発奇形症例のアレイ CGH 解析(続報) 第 57 回 日本人類遺伝学会 2012.10.26
- 4) 和田 友香 花井 彩江 佐々木 愛子 藤田 秀樹 小崎 里華: 近四倍体と 18 トリソミーのモザイクの 1 女児例 第 57 回 日本人類遺伝学会 2012.10.26
- 5) 鳥居千春 丸岡亮 清水厚志 小崎里華 小崎健次郎: 次世代シーケンサーを用いた先天奇形症候群の網羅的診断 第 52 回 日本先天異常学会 2012. 7.6
- 6) 藤田秀樹 小崎里華 : アレイ CGH で 1q24-q 25 領域に欠失を認めた精神運動発達遅滞児についての検討 第 52 回 日本先天異常学会 2012. 7. 6
- 7) 三須久美子 桐林和代 佐谷秀行 鳥居千春 小崎里華 小崎健次郎: 神経線維腫症 1 型の遺伝様式に関する患者家族の誤解の類型化 第 36 回 日本遺伝カウンセリング学会 2012 6.10
- 8) 岡田朋美 佐々木愛子 黒田くみ子 上田 英梨子 江川真希子 杉林里桂 住江正大 李紅蓮 藤田秀樹 小崎里華 左合治彦: 成育医療研究センターにおける周産期遺伝カウンセリング体制 第 36 回 日本遺伝カウンセリング学会 2012 6.9
- 9) 武内俊樹 下郷幸子 山崎麻美 小崎里華 小崎健次郎 高橋高雄: LICAM 変異により発症した先天性水頭症と Hirshsprung 病の合併例 第 54 回 日本小児神経学会総会 2012.4.19
- 10) 柳橋達彦 小崎健次郎 小崎里華 吉橋博史 井原正博 高橋孝雄: Williams 症候群における欠失範囲の大きさと発達遅滞の重症度との関係 第 115 回 日本小児学会学術集会 2012.4.22
- 11) 一宮優子 石黒精 中館尚也 前川貴伸 小崎里華 藤田秀樹 阪井裕一 : TPO 受容体作動薬が有効であった 3p12-13 欠失の難治性慢性 ITP 第 115 回 日本小児学会学術集会 2012.4.22

## G. 知的財産権の出願・登録状況(予定を含む。)

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（難病関係研究分野）  
分担研究報告書

次世代シーケンサーによる解析に関する研究

分担研究者 工藤 純 慶應義塾大学医学部遺伝子医学研究室 教授

研究要旨

次世代シーケンサーを用いた全ゲノム解読あるいは全翻訳領域解読（いわゆるエキソーム解析）は、未知の疾患原因遺伝子を探索する研究手段としてはほぼ確立されており、我々も既に遺伝性疾患の新規原因遺伝子の探索にシーケンシングからデータの解析、検証まで一括して実践している。本研究においては、以上のゲノムワイドな解析のウェットとソフトの両面における精度のさらなる向上を目指すために、そのノウハウを生かして、疾患毎の様々な状況に応じて、次世代シーケンサーを活用した遺伝子解析システムを構築し、実践することを目指す。近い将来日常的な遺伝子診断において次世代シーケンサーを活用するためには、経済的かつ簡便なシステムの構築が望まれるため、現在利用できる手段の中から、最適の手段を比較検討した。特にこれまで主に使用されて来たベイト RNA とのハイブリダイゼーションによりエキソンのみを濃縮する手法（SureSelect 等）を用いたエキソーム解析の結果からは、解読が困難な難読領域が存在することが明らかになってきており、本研究では最近開発された HaloPlex ターゲットエンリッチメントシステムを用いた難読領域の増幅を試みた。

研究協力者

清水厚志（慶應義塾大学医学部分子生物学教室）  
鳥居千春（慶應義塾大学医学部臨床遺伝学センター）  
小崎健次郎（慶應義塾大学医学部臨床遺伝学センター）

A. 研究目的

疾患毎の様々状況に応じて、次世代シーケンサーを活用した遺伝子解析システムを構築し、実践することを目指す。近い将来日常的な遺伝子診断において次世代シーケンサーを活用するためには、経済的かつ簡便なシステムの構築が必要であり、そのことを念頭に置いて、実用的なシステムを構築する。

B. 研究方法

1) 標的領域遺伝子配列の増幅

これまでに実践したエキソムキャプチャー（SureSelect：アジレント社）における難読領域の中で、ベイトが設計されていても厚みが×10未満のシーケンスデータしか得られていなかった難読領域 724kb の中から 57nt 以上の長さの難読領域 3,845 カ所計 497kb を標的として HaloPlex ターゲットエンリッチメントシステム（アジレント社）のプロブライブラリーを作成し、ヒト NA18987 の全ゲノム DNA から標的領域のみを増幅した。

2) パーソナル次世代シーケンサーによるシーケンシング

HaloPlex システムを用いて増幅した DNA 断片を MiSeq（イルミナ社）を用いて解読した。解析にあたっては個人情報の保護に関する法律を踏まえ、文部科学省・厚生労働省・経済産業省「ヒトゲノム・遺伝子解析研究に関する倫理指針」、文部科学省・厚生労働省「疫学研究

指針」を遵守し、倫理委員会の承認の下に研究を実施した。

C. 研究結果

これまでに構築したエキソムキャプチャー（SureSelect：アジレント社）によるエキソーム解析のシステムを用いて、変異解析を行った。Bwa・GATK, Picard, SnpEff・SomaticSniper, MuTect等のオープンソースのプログラム組み合わせによるパイプラインを作成した。先天異常症候群のデノボ変異例の検出、常染色体劣性遺伝病の変異の検出などから、システムの構築が確認された。これまでに実践したエキソーム解析における難読領域の大部分は GC 含量が 60～80%の GC リッチな配列であった。全ゲノム DNA から標的領域のみを増幅し、回収した DNA 断片をパーソナル次世代シーケンサーMiSeqを用いて解読したところ、GC リッチの標的領域のうち 60%以上について、×10 以上の厚みで解読に成功した。

D. 考察

我々が構築したエキソーム解析のシステムが、未知の疾患原因遺伝子の探索において有効に働くことが実証された。

次世代シーケンサーを用いた遺伝子解析システムのうち、全翻訳領域解読（エキソーム解析）や、候補遺伝子群を一括して解析するためには、

エキソソングキャプチャー (SureSelect) を用いる手法が、優れていると考えられる。ただし、エキソソングキャプチャーでは、5-10%の難読領域が発生するため、その部分のみに照準を絞った補助的手段として HaloPlex ターゲットエンリッチメントシステムの併用は、考慮する価値があると思われる。

## E. 結論

我々が構築したエキソソーム解析のシステムは、有効に働くことが実証された。また、次世代シーケンサーを用いた遺伝子解析システムにおいて、エキソソングキャプチャー (SureSelect) における難読領域を解読する手法として HaloPlex システムの併用は、現実的であると考えた。

## F. 研究発表

### 1. 論文発表

Sasaki, T., Niizeki, H., Shimizu, A., Shiohama, A., Hirakiyama, A., Okuyama, T., Seki, A., Kabashima, K., Otsuka, A., Ishiko, A., Tanese, K., Miyakawa, S., Sakabe, J., Kuwahara, M., Amagai, M., Okano, H., Suematsu, M., Kudoh, J. Identification of mutations in the prostaglandin transporter gene *SLCO2A1* and its phenotype-genotype correlation in Japanese patients with pachydermoperiostosis. *J. Dermatol. Sci.*, **68**(1):36-44 (2012).

### 2. 学会発表

工藤 純、ゲノム解析による遺伝性疾患の原因解明、第 16 回九州基礎皮膚科研究会 特別講演、福岡、2012 年 11 月 17 日

## G. 知的財産権の出願・登録状況 (予定を含む。)

### 1. 特許取得

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（難病関係研究分野）  
分担研究報告書

次世代シーケンサーによる先天異常症患者ゲノム変異探索システムの改良

分担研究者 清水厚志  
岩手医科大学 いわて東北メディカル・メガバンク機構 生体情報解析部門 特命教授

**研究要旨**

次世代シーケンサー（NGS）の技術革新と解析技術の発展を受け、大規模なヒト多型データの収集が世界規模で進められている。さらに、小型のNGSが登場した事や、コマンドラインを必要としない市販ソフトが登場したこともあり、NGSを利用する研究への敷居は明らかに低くなってきた。一方で、疾患解析のために公開されている多型データをどのように選択し、利用するか、また市販ソフトの精度はコマンドラインで構築された解析システムとどの程度の差があるかといった、ビックデータを直接解析しなければ得られない情報は少ない。そこで、本年度は変異探索システムの改良を進めるとともに、公開多型データの比較検討や、構築したシステムと市販ソフトの比較を行った。さらに、全ゲノム/全エクソーム解析では原因を同定できない疾患への応用に向けて、発現変動遺伝子同定システムの構築を開始した。

**A. 研究目的**

次世代シーケンサー（NGS）の技術革新を受け、1000 Genomes Project (1KG) や、NHLBI GO Exome Sequencing Project (ESP) によるヒト多型データの収集が行われた。その結果、旧来のSNPsの定義である「頻度1%以上」を大幅に下回るレア多型が個人ゲノム解読の過程で多数同定されることが判明した。これらのレア多型の大部分は本研究事業の中で我々が現時点で対象としている遺伝性の先天異常症の原因には直接関与しないことが考えられる。そこで、本年度はすでに構築済みである変異探索システムを改良すると共に、複数の患者間で共通の変異、あるいは保因者である患者両親の多型情報をインターセクトすることで候補変異を抽出するシステムを構築した。また、従来のゲノムDNA解析（全ゲノム解析、エクソーム解析）に依存した手法では疾患の原因であると同定できない変異を、将来的に患者特異的iPS細胞から同定することを目標に、次世代シーケンサーによる遺伝子発現解析（RNA-seq）のデータを複数検体間で比較した際に遺伝子発現量が増加、あるいは減少している遺伝子群を同定するシステムを構築した。

**B. 研究方法**

**1. 変異探索システムの改良**

昨年度までに主にアカデミックフリーのソフトウェアを利用して、

- i) NGS配列データ (fastq) から校正済みゲノムアライメントファイル (bam) の作成
- ii) ヒト参照配列との比較による多型の抽出
- iii) 得られた個人の多型情報と公開多型データベース (dbSNPs, 1KG) の比較による変異候補遺伝子の抽出

を行うシステムを構築した（清水厚志、細胞工学. 2011）。しかし、この手法によっても個人ごとに300ヶ所以上のレア多型/変異が検出され、単一遺伝子疾患と考えられる患者1名分のデータから原因遺伝子を同定することが不可能な症例が多数あった。一方で、公開多型データベースの登録情報すべてをフィルターとして使用すると、数千人に1アレルというごく稀な多型/変異まで疾患候補変異から除外してしまうため、公開多型データの健常人ボランティアの中に偶然患者と同じ変異を持っていた場合に擬陰性が生じる可能性があった。また、頻度の高い一塩基多型 (SNPs) に隣接して変異が生じた場合に、既存の手法ではそれぞれを独立の変異として処理してしまい、異なるアレル上の正確なアミノ酸置換を再現できない問題もあった。

そこで、本年度は、

- i) 複数の公開多型データベースを統合した最新の既知多型リストへの更新
- ii) 頻度情報を考慮したフィルターの作成
- iii) 連続してレファレンスと異なる場合の同一アレル内多型同定機能の追加
- iv) 複数の検体間の共通変異抽出機能の追加を行った。

また、コマンドライン入力に慣れていない医師、研究者が解析研究を行う状況を考慮し、市販ソフトウェアでは評価の高いCLC Genomics workbench (CLC) と我々が構築した解析システムとの比較検討を行った。

**2. 発現変動遺伝子同定システムの構築**

ゲノムDNA解析によって同定された変異のうち、ナンセンス変異やエクソン-イントロージャクション等の変異を疾患の原因と推定することは比較的容易である。一方で、イントロンや



プロモーター等の制御領域の変異はゲノムDNA解析ではその影響を推定することがほぼ不可能である。そこで、本事業分担研究者の赤松らにより、将来的に疾患患者特異的iPS細胞の樹立が可能となるという予測のもと、分化誘導させた患者特異的細胞とコントロール細胞の遺伝子発現量を次世代シーケンサーのデータから比較できる解析システムを構築した。

(倫理面への配慮)

個人情報保護に関する法律を踏まえて研究を実施した。遺伝子解析に際しては文部科学省・厚生労働省・経済産業省「ヒトゲノム・遺伝子解析研究に関する倫理指針」を遵守し、倫理委員会の承認下に研究を実施した。

## C. 研究結果

### 1. 変異探索システムの改良

#### 1.1. 既知多型データベースからの多型情報の入手と既知多型リストの更新

dbSNP137, ESP6500, 1KG low coverage, 1KG日本人エクソームの多型データを統合し、既知多型リストを更新した。特に日本人エクソームのデータに関しては我々が構築した多型抽出システムを使用して、同じ処理/同じ精度の多型リストを独自に作成した。その結果、昨年度の既知多型リスト (48,455,489ヶ所) におよそ400万ヶ所を追加できた (52,659,387ヶ所)。

#### 1.2. 頻度情報を考慮したフィルターの作成

それぞれの公開多型データに付随した頻度情報 (MAF; Minor allele frequency) から  $MAF > 0.01$  ( $> 1\%$ ) の多型のみを抽出した高頻度多型フィルターを作成し、効果を検証した (表1: 後述)。

#### 1.3. 連続塩基置換のアレル同定

構築した変異探索システムにはGATK (Genome Analysis tool kit; <http://www.broadinstitute.org/gatk/>) を利用しているが、GATKで抽出された塩基置換リストにはアレル情報が欠失している。そこで、抽出された塩基置換がゲノム位置上で連続していた場合に、構成済みアライメントデータに含まれる配列情報から、それらの塩基が同一アレル由来か判定するシステムを構築した。HapMapの検体 (NA18943) を用いて解析したところ、CACNA1Hのコーディングエクソン上に同一アレル由来の連続塩基置換を同定した (図1)。

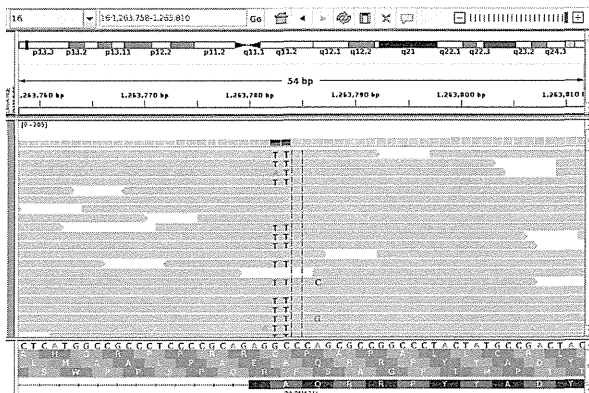


図1: NA18943で同定された連続塩基置換

#### 1.4. 複数の検体間での共通変異の抽出

変異探索システムで候補として上がったゲノム上の変異から対応する遺伝子のリストを作成して、患者内でのみ変異が見つかった遺伝子を抽出する機能を追加した。

#### 1.5. 変異探索システムの検証

1.4.までの追加機能の効果を検証するため、すでに原因遺伝子が同定されている常染色体劣性遺伝子疾患患者エクソーム解析データで検証を行った (表1)。

表1: 原因遺伝子の候補リスト中の順位\*

A:	B:	C:	D:
完全新規	MAF <	MAF <	MAF <
dbSNP137	0.01	0.01	0.01
JPT	dbSNP137	dbSNP137	dbSNP137
Exome	JPT	JPT	
1KG	Exome	Exome	
ESP6500	1KG		
	ESP6500		
2位	6位	10位	131位

\*HLA除く

現在構築したシステムは擬陰性をできるだけ排除するために、シーケンズ精度が比較的低い塩基由来の候補部位も残すフィルターに設定している ( $QV \geq 20$ ,  $Depth \geq 4$ )。そこで、IGV (Integrative Genomics Viewer) を用い、目視にて表1 A-Cの上位遺伝子群の候補部位を確認すると、候補遺伝子の大多数はシーケンズ精度が今回設定した精度の下限値であった。また、原因遺伝子以外の十分なシーケンズ精度持つ変異候補は異なる家系であるにもかかわらず同一の塩基置換であることがわかった。

#### 1.6. 市販ソフトウェアとの比較

1.5.の検証で我々のシステムにより効果的に原因遺伝子を同定できることが確認できた。しかし、本システムはLinux上で動くコマンドラインベースで構築されているため、一定レベルのコンピュータスキルを必要とする。そこで、コマンドライン入力に慣れていない医師、研究者が市販のソフトウェア (CLC) を使用した際にどの程度の精度で変異を同定することができるか、主にアライメントの機能について比較検討した。同一個人のエクソーム解析の配列データを本システム、CLC、bowtie2でレファレンス配列にマッピングし、特にアライメントの結果が大きく異なると予想された繰り返し配列を含む indel (insertion-deletion) の領域をアライメントデータから抽出した。本システムではindel周辺も正確にアライメントできていた (図2上

段)のに対し、CLCでは大半が考慮されずにトリミングされており(図2中段)、比較のため使用したbowtie2ではアライメントが正確には行われず、indelではなく多数のSNPsとして同定された(図2下段)。

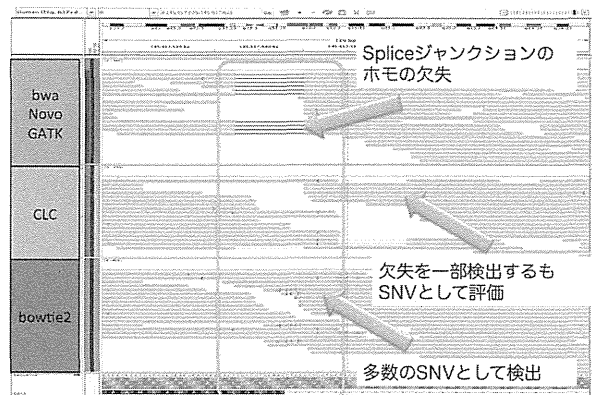


図2：複数のアライメントソフトとの比較

## 2. 発現変動遺伝子同定システムの構築

現時点で疾患患者特異的iPS細胞のデータを手に入っていないため、Human Body Map 2.0 からヒト組織別RNA-seqの配列データ

(<http://www.ebi.ac.uk/arrayexpress/experiments/EMTAB-513/>)を入手し、システム構築のモデルデータとして、以下のシステムを構築した。

- i) 得られた配列データに含まれるアダプター配列とクオリティ値が低い塩基の除去
  - ii) ヒトレファレンス配列への遺伝子構造を考慮したマッピング
  - iii) 遺伝子毎の発現量の推定
  - iv) 得られた予測発現量の差分処理
  - v) 組織間で発現量が有意に異なる遺伝子群の抽出と描画
  - vi) Gene Ontology 抽出
  - vii) 既知の遺伝子ネットワーク上への描画
- 差分処理iv)までをLinux上で処理し、その後の処理をPC上のR packageで処理することで効率的にネットワーク解析まで行うことができた。

## D. 考察

今回の結果から、変異探索には最新のdbSNPs version 137だけでは十分ではなく、日本人データを我々の変異探索システムと同じ処理をした独自の参照多型データの追加が有効であった(表1: D->C)。

一方で、MAFフィルターの設定を<0.01としても極端に候補遺伝子は増加せず(表1: A->B)、頻度情報が保証できない多型は擬陰性を避けるために少なくとも劣性遺伝子疾患に関しては使用しないほうが良いことが示唆された。

また、我々の探索システムで原因遺伝子以外にも十分なシークエンス精度をもつ候補遺伝子が複数抽出された。これらの擬陽性の候補では、異なる家系の患者にもかかわらず共通の塩基置換が抽出されており、多型情報の不足が原因と考えられる。来年度以降、日本人全ゲノム/全

エクソーム解析から得られた日本人標準SNPsが複数の機関から公開されることが予定されているが、解析手法、精度などが統一されていないため、配列データが公開されない以上は1KGの日本人エクソーム以上の効果がどの程度効果が見込めるかは疑わしい。

我々が開発したシステムはLinuxなどのコンピュータスキルを必要とするが、GUI(Graphical User Interface)で操作できる市販ソフトウェアも発売されている。そこで、本年度は主にアライメント処理について比較したが、現時点の精度は我々のシステムに及ばなかった。しかし、ユーザーからの情報をフィードバックしながらソフトウェアの開発は継続されており、今後の改善を考慮すると、疾患特異的パネルなどのような比較的データ量が少ない解析に関しては、我々の開発、作成したファイルと組み合わせることで十分な精度が担保できると判断できた。

さらに本年度は、変異解析システムに加えて、疾患特異的iPSの性状解析のための発現解析システムを構築した。現時点では組織別データを用いたパイロット開発であったが、発現変動遺伝子群の抽出まで行う事ができた。来年度以降に実際のiPS細胞のデータを利用する事ができた場合は解析系をさらにiPSに特化した形に改良する。

## E. 結論

本年度改良した変異探索システムは既知の疾患関連変異、および新規変異を同定するのに有効であった。特に連続塩基置換は既報のシステムでは対応されていなかったが、今回、存在することが確認できたため、既知多型リストについて再度精査し、変異検出精度の向上を図る。

市販ソフト(CLIC)との比較ではアライメント部分に加え、フィルタリングについてもまだ十分な精度とは言えないレベルであった。ただし、今後の定期的なアップデートにより改良が見込めること、また、単独での使用ではなく、我々がカスタマイズしたデータを提供することで十分な精度が見込めると判断できた。

発現変動遺伝子同定システムについては初期パイプラインが完成したので、来年度以降に本研究事業の分担者から提供可能なデータを手入れし、実データを元に解析パイプラインの改良を進める。

分担研究者の松永らの研究結果や、既報の遺伝子性疾患患者のゲノム解析の結果から、ゲノムDNAの解析のみではすべての患者の発症原因を同定することが困難なことがわかっている。疾患の原因であると同定できない変異のうち、ミスセンス変異や、タンパク質複合体を形成する複数の遺伝子が関与するものも多数含まれていると考えられるが、遺伝子発現量に影響を与えている制御領域上のDNA変異、あるいは後天的なゲノム修飾(エピゲノム)なども関与することが考えられる。そこで、将来的にはゲノム、トランスクリプトーム、エピゲノムの三層にわ

たる複合オミックス解析の結果から疾患の原因遺伝子や危険因子を同定するようなシステムの構築を目指す。

## F. 研究発表

### 1. 論文発表

- 1) 清水厚志: 次世代シーケンサーによる発現変動遺伝子の同定. *Medical Science Digest*: 38(13): 587-588, 2012.

### 2. 学会発表

- 1) 清水厚志, 古家育子, 佐々木貴史, 工藤純: 慶應義塾大学医学部内での受託解析システムの構築. 第2回次世代シーケンサー現場の会大会. 2012 5/26 大阪.
- 2) 清水厚志: Miseq による既知疾患遺伝子群のターゲットリシーケンシング. 2012 イルミナ次世代シーケンサーユーザーフォーラム. 2012 6/20 東京 (招待講演)
- 3) 清水厚志: 比べてみよう CLC Genomics Workbench とオープンソース; 市販ソフトとアカデミックフリーソフトの共存. CLCbio ユーザーミーティング 2012. 2012 6/26 東京 (招待講演)
- 4) 鳥居千春, 丸岡亮, 清水厚志, 小崎里華, 小崎健次郎: 次世代シーケンサーを用いた先天奇形症候群の網羅的診断. 第52回日本先天異常学会学術集会. 2012 7/16-18 東京.
- 5) 清水厚志: ゲノムプロジェクトからパーソナルゲノムまでの10年. 次世代シーケンシングが切り開く医学研究と展望 京都大学 iPS 細胞研究所ゲノム・エピゲノム解析コアファシリティ. 2012 7/31 京都 (招待講演).
- 6) Shimizu A, Torii C, Suzuki N, Mutai H, Kudoh J, Kosaki R, Matsunaga T, Kosaki K. Rapid and efficient mutation detection in the hundreds of target genes by bench-top next generation sequencer with custom target capture method. The American Society of Human Genetics 62th Annual Meeting, San Francisco, USA (Nov. 2012) 11/6-11
- 7) 鈴木直大, 務台英樹, 鳥居千春, 清水厚志, 宮冬樹, 難波一徳, 工藤純, 小崎健次郎, 松永達雄: カスタムターゲットリシーケンスによる難聴関連遺伝子の変異探索. 日本人類遺伝学会. 第57回大会, 2012 11/20-23 東京.
- 8) 清水厚志. 遺伝性疾患患者ゲノムの Exome および Targeted resequencing 解析の実例. 第35回日本分子生物学会年会アジレントテクノロジーランチョンセミナー. 2012 12/13 福岡 (招待講演).

## G. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得  
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（難病関係研究分野）  
分担研究報告書

次世代シーケンサーを用いたバイオインフォマティクス解析による  
先天性異常症の遺伝要因の解明に関する研究

研究分担者 宮 冬樹

独立行政法人理化学研究所ゲノム医科学研究センター情報解析研究チーム・研究員

研究要旨

次世代シーケンサーのデータから真の疾患関連変異候補を抽出するため、解析パイプラインの更なる改良を行った。その解析パイプラインを用いて、変異箇所が既知のHapMapサンプルについて、我々の実験系で次世代シーケンサーにて変異を探索し精度を検証した結果、偽陽性率（変異が存在しないのに変異を誤って検出してしまう率）・偽陰性率（変異が存在するのに検出できない率）とも極めて0に近い正確な精度を有することが確認できた。これにより、疾患関連候補の変異検出の段階は完成したと言え、家系サンプルを用いての最終的な疾患原因変異同定と機能解析という次のステップへ移行できるようになった。

A. 研究目的

膨大な次世代シーケンスデータから疾患関連変異を探し出す際の大きな課題となるのは多数の候補変異（大部分は疾患と無関係の変異またはコールエラーによる疑陽性）が生じる中から真の疾患関連変異を導き出す工程にある。昨年度までで次世代シーケンスデータから疾患関連候補変異を抽出するまでの解析パイプラインは完成したが、さらにこれを進展させ、より疑陽性率を下げて高精度に、かつ見逃しがないように偽陰性率も下げるためにパラメータの最適化や各種プログラムの改良を施すこととした。また、追加で各種既報変異データベースや疾患データベース等の情報を取り込み、疾患との関わりが予測される変異の抽出法を組み込むこととした。また実際の研究班での実験環境で出力された結果データを元に検討を行い、精度を確認することとした。

B. 研究方法

次世代シーケンサーを用いた疾患関連候補遺伝子同定までの一連の解析の流れの概略を図1に示した。昨年度に解析パイプラインの基本は完成したが、図1中の\*1～\*5の部分にさらなる改良を加えた。

まず、図1\*1の部分において、研究班のシーケンスデータはpaired-end法というシーケンスライブラリを両方向から読む手法をとっているが、この両方向からの2つのシーケンスリードが的確なゲノム上の位置にマッピングされているかどうかを調査し、サンプルごとの全リードの平均値と分散を算出し、ペアとなるリードが全ペアリード長の平均値±2SD（標準偏差）に存在するリードのみを選択することにした。この範囲外にあるリードは多くがマッピングエラーのリードであると推測されるため除外することとした。

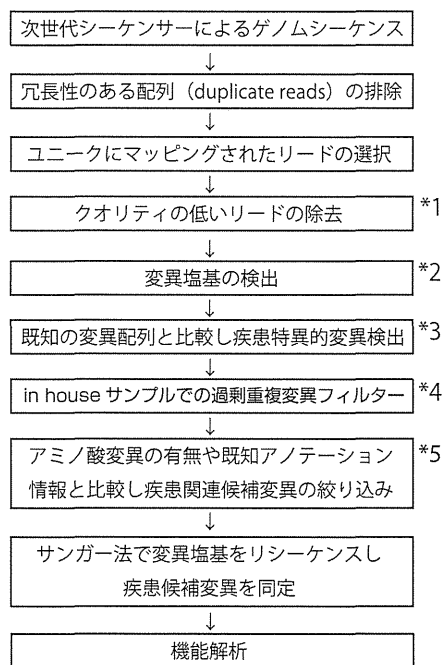


図1. 次世代シーケンスデータ解析の流れ

次に図1\*2の部分において、各種変異検出のパラメータの最適化を行った。その一例として、次世代シーケンスにて読まれたある同一箇所のアレルの頻度が偏っていてもheteroとコールされる場合があり、そのような頻度が偏っている場所はエラー率が高いことが明らかとなっている（データ未掲載）ことから、同一箇所でも25%以上リフレレンスと異なる塩基がコールされる場合のみ変異として出力するように変更した。

次世代シーケンスデータから偽陽性ではない

真の変異を抽出した場合でも、一般集団で有する common variant であれば先天性疾患の原因遺伝子である可能性は極めて低く、レアな変異のみに絞り込むことは極めて重要である。図 1\*3の部分に該当する既知変異のフィルターについては、昨年度も使用した dbSNP と 1000 Genomes Project のデータを最新のデータに更新し、さらに新たに 6500 人のシーケンスデータを有する Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>) を追加することとした。さらに重要な追加点として、一般的なフィルターでは変異が登録されている染色体上の位置情報のみからフィルターにかける場合が多いが、疾患要因変異と登録変異は同じ場所でも塩基が異なる可能性も十分ある。そのことから、我々のフィルターでは塩基の一致についても調査し、データベースに染色体上の同一位置で変異が登録されていても塩基が異なる場合はフィルターで除外しないように改変した。また、一般集団でもわずかながら発症はしなくとも原因遺伝子変異の保因者である可能性もあることから、フィルターを 2 種類作製し、1 つは既報のものを全て除く「Filter set 1」、もう一つはアレル頻度 0.001 を閾値として一般集団でも有する変異を残す緩めの「Filter set 2」として作製した。さらに図 1\*5 の追加の一部として、既知の疾患関連変異として知られている変異で、一般集団におけるアレル頻度が 0.01 未満の変異は Filter において除外せずに「Filter set 2」において残すように改変した。これは、先天性疾患のみならず、他の疾患の原因として知られている変異が、実は先天性疾患の表現型とリンクしている可能性を考慮したことによる。データベースとしては世界最大のヒトメンデル病疾患データベースである HGMD Professional と、NCBI の OMIM データベース、dbSNP の登録変異のうち疾患との関与（示唆）が報告されている pathogenic SNP の 3 種を用い、合計で 75,221 塩基を疾患関連変異としてフィルターに用いた。

また、データベース上の既知の SNV にはないが、我々の実験系で読まれたデータ中で多数の検体が染色体上の同一箇所の変異を保有していた場合、その変異はプラットフォーム依存的な偽陽性である可能性が大きい。そこで図 1\*4 において、我々の検体データ (in house data) で、多数の検体が同一変異を保有する場合（異なる家系で 4 検体以上同一箇所の変異を有する場合）にはフィルターで除外するようにした。

以上のような改良を加えたパイプラインについて、実際に研究班で行っている実験系で次世代シーケンスを行い、その精度を検証することとした。検証には、HapMap Project で用いられた変異が既知の日本人サンプルを用い、変異同定率・一致率を調査した。

また、作製した改良解析パイプラインを用いて、先天性神経疾患検体の次世代シーケンス実験を行い、累計約 160 名の exome 変異探索を行った。

(倫理面への配慮)

個人情報保護に関する法律を踏まえて研究を実施した。遺伝子解析に際しては文部科学省・厚

生労働省・経済産業省「ヒトゲノム・遺伝子解析研究に関する倫理指針」を遵守し、倫理委員会の承認下に研究を実施した。

### C. 研究結果

HapMap Project でも変異探索が行われた日本人サンプルについて、研究班の実験系を用いて次世代シーケンスを行い、作製した改良解析パイプラインにて変異検出を行ったところ、HapMap で genotype データがある箇所と我々のシーケンスデータがある箇所との間で 3,129 箇所の重複があった。その 3,129 箇所の genotype の一致率について調べたところ、HapMap データと我々の次世代シーケンスデータで 16 箇所だけ不一致であった（一致率 99.49%。図 2）。

自分の call	HapMap と一致	HapMap と不一致	concordance rate
ref と同一 homo	2,170	11	99.50%
Hetero SNV	459	2	99.57%
Homo SNV	484	3	99.38%
total	3,113	16	99.49%

図 2. 我々の次世代シーケンスデータと HapMap データとの比較

この不一致であった 16 箇所について、真の genotype は何であることを確認するため、サンガーシーケンス法にて再シーケンスを行ったところ、実際は 16 箇所全てにおいて我々の genotype call が合っていた（HapMap データが誤っていた。図 3）。すなわち HapMap の他の genotype データが全て合っているとしたり、我々の call は極めて 100% 近く正確であることが確認できた。

SNV call	position	rsID	Ref	CONGS	HapMap	sanger seq
Homo	chr6: 133,827,354	rs2277083	A	G	A/G	G
Homo	chr9: 123152241	rs2297456	A	C	A/C	C
Hetero	chr8: 144,887,734	rs11783725	G	A/G	A	A/G
Hetero	chr9: 139,328,369	rs10781543	T	C/T	T	C/T
Hetero	chr17: 7,579,472	rs1042522	G	C/G	G	C/G
Ref	chr1: 22,176,959	rs2290499	C	C	T	C
Ref	chr1: 206,574,659	rs1754497	C	C	G	C
Ref	chr3: 12,573,336	rs299641	A	A	G	A
Ref	chr4: 4,861,780	rs10213286	C	C	T	C
Ref	chr6: 151,413,828	rs2295734	C	C	A	C
Ref	chr7: 103,206,013	rs2711839	T	T	A	T
Ref	chr7: 103,244,981	rs362681	A	A	G	A
Ref	chr16: 2,143,849	rs2854581	C	C	G	C
Ref	chr16: 2,162,979	rs2855349	G	G	A	G
Ref	chr19: 41,034,912	rs1461008	G	G	A/G	G
Ref	chr20: 62,038,757	rs3746364	C	C	C/T	C

図 3. 我々のデータと HapMap データで不一致だった箇所のサンガーシーケンス結果。「CONGS」の列が我々の genotype call 結果。

一方、データベースに掲載されている既知の変異を統合した Filter set は約 4,700 万箇所 (Filter set 1) または約 2,800 万箇所 (Filter set 2) となった

(図4)。1000 Genomes Project の報告では exon 領域に存在する 1%以上の頻度の変異は同プロジェクトの結果データにおいて 99.8%網羅されていると推測されている (*Nature* 491, 56-65, 2012) ことから、他のデータベースをさらに統合させた我々の Filter set においてはそれ以上の割合で、exon 領域に存在し一般集団が保有する common variant を疾患関連候補から除外できるものと考えられる。

	Filter set 1 (unique bases)	Filter set 2 (unique bases)
dbSNP build137	45,953,783 (All)	10,138,587 (MAF > 0.001)
1000G exome (1,092 individual)	714,058 (All)	299,491 (MAF > 0.001)
1000G whole (1,092 individual)	37,957,691 (All)	25,552,088 (MAF > 0.001)
ESP6500 European American (4,300 individual)	1,872,496 (All)	350,350 (MAF > 0.001)
ESP6500 African American (2,203 individual)	1,872,496 (All)	667,967 (MAF > 0.001)
<b>Total unique bases</b>	<b>46,967,933</b> (47,162,884 allelic)	<b>28,077,587</b> (28,119,539 allelic)

Filter set 2からは上記SNVからPathogenic SNV [MAF < 0.01] (HGMD + OMIM + dbSNP\_pathogenic) の75,221 unique basesを追加で除外し (Filterにかけない)、最終結果に追加。

図4. 既知のデータベース情報を統合した Filter set

作製したパイプラインで神経系先天性疾患献体約 160 を解析した結果の概要を図5に示した。神経系の 284 遺伝子をターゲットとした神経系標的カスタム exome (プローブ設計領域 1.6M) では塩基辺りの平均読み深度 (depth) が 1001.9、全 exon を対象とした All exome (プローブ設計領域 51M) の平均 depth は 98.8 となっており、SNV を call する際の閾値とした depth 10 を超えている割合はそれぞれ 99.7%と 98.0%となり、双方とも十分なデータ量が得られており、データのクオリティに関しても問題なかった。最終的に疾患原因候補変異となるアミノ酸変異を伴う変異または終始コドンの入る変異の数は、神経系カスタム exome の場合が 1 サンプル当たりの平均で 5 個、All exome の場合の平均数は 172 個であった。ここからさらに家系サンプルや遺伝形式情報などを元に絞り込み、最終的な疾患の原因となっている変異を同定する工程を現在進めている (疾病ごとの詳細な進捗状況については研究分担者の各報告を参照。)

	神経系遺伝子カスタム exome	All exome
キャプチャープローブ設計領域長	1.6 M bases	51 M bases
シーケンスターゲット領域長 (キャプチャー領域±100bp)	4.4 M bases	88 M bases
シーケンス手法	101 bp paired-end, 24 multiplex / lane	101 bp paired-end, 6 multiplex / lane
シーケンスデータ平均リード長	2.11 G bases	7.88 G bases
ゲノムマッピング率	99.30 %	99.68 %
冗長除去リード率	11.62 %	4.84 %
ユニークにマッピングされ pair read間の距離が平均±2SDのリード中、ターゲット領域に含まれるリード長 [率]	1.25 G bases [74.66 %]	6.14 G bases [87.75 %]
プローブ設計領域平均読み深度	1001.9	98.8
シーケンスターゲット領域平均読み深度	811.9	68.8
プローブ設計領域の読み深度10以上 (SNV call可能場所) の割合	99.7 %	98.0 %
フィルター前の各サンプルで callされる SNV の平均数	1,862	68,526
一般集団で既知の SNV を除去後 SNV 平均数	29	980
アミノ酸変異を伴う、または終始コドンの入る変異の各サンプル平均数	5	172

## D. 考察

SNV callの解析パイプラインの改良により、より疾患関連候補遺伝子を絞り込めるようになった。SNVが既知のHapMapプロジェクトで用いられたサンプルを我々の実験系で調査した結果、16個不一致があったが、サンガーシーケンスの結果、我々のcallの方が全て合っていたことから、他のHapMapデータが正確だと仮定した場合、偽陽性率は0であると共に、偽陰性（callがある場所においてSNVを発見できない確率）も今回の結果の場合0であり、極めて正確かつ見落としもないSNV callパイプラインが完成したものと考えられる。

現在この解析パイプラインを用いて神経系先天性疾患をはじめとした疾患のデータを解析しており、疾患原因変異をほぼ1つに同定できている患者数も一定数存在する。ただし、全 exome の場合は、絞り込み後の最終候補で一人当たり172個の疾患関連候補が残ることから、家系サンプルがある程度揃っている検体を用い、絞り込む必要があり、家系内の遺伝形質を考慮した検体収集が非常に重要であると考えられる。一方で、神経系遺伝子に絞ったカスタム exome の場合は、最終的に残る候補遺伝子数は少なく、神経系の重要な変異が検出された場合の絞り込みはわりと容易いが、最終候補遺伝子が残らない場合もあり、今回の1.6Mの神経系遺伝子セットには含まれない遺伝子の変異が疾患原因遺伝子であることも考えられる。コストパフォーマンスと想定されるターゲット遺伝子とを考慮し、神経系遺伝子カスタム exome か全 exome かを選択し実験を進める必要性があると考えられる。近年のコストと解析技術の進展を考えると、全 exome の方がややメリットは大きいようにも考えられる。

現在明らかになっている疾患関連候補遺伝子について、今後サンガーシーケンス法による確認実験や機能解析を進め、さらなる検証を行っていくと共に、さらに検体を増やし追加実験を行っていくことで、原因遺伝子変異の確定、疾患メカニズムの解明、診断システムの整備等を進めていく予定である。

## E. 結論

本プロジェクトの先天性疾患の原因遺伝子探索およびその解明への達成に向け、次世代シーケンサーを用いたデータから、疾患原因候補遺伝子変異を抽出するまでの解析パイプラインが完成し、一つの段階を終了できた。ここからは原因候補変異からそれぞれの検体ごとに原因変異を同定し、そのメカニズムを解明する次の段階に入ることになる。引き続き必要性に応じ解析パイプラインを改良しつつ、候補の中から原因を同定、そして解明を目指し、実験・解析を進めていく。

## F. 研究発表

### 1. 論文発表

- 1) Tateno C, Miya F, Wake K, Kataoka M, Ishida Y, Yamasaki C, Kakuni M, Wisse E, Verheyen F, noue K, Sato K, Kudo A, Arii S, Itamoto T,

Asahara T, Tsunoda T, Yoshizato K; Morphological and microarray analyses of human hepatocytes from xenogeneic host livers. *Lab. Invest.* **93**, 54-71 (2013). (本論文は Global Medical Discovery [ISSN 1929-8536] の Key Scientific Articles にも掲載された。 <http://globalmedicaldiscovery.com/key-scientific-articles/morphological-and-microarray-analyses-of-human-hepatocytes-from-xenogeneic-host-livers/>)

- 2) Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, Hosono N, Kubo M, Miya F, Arai Y, Takahashi H, Shirakihara T, Nagasaki M, Shibuya T, Nakano K, Watanabe-Makino K, Tanaka H, Nakamura H, Kusuda J, Ojima H, Shimada K, Okusaka T, Ueno M, Shigekawa Y, Kawakami Y, Arihiro K, Ohdan H, Gotoh K, Ishikawa O, Ariizumi S, Yamamoto M, Yamada T, Chayama K, Kosuge T, Yamaue H, Kamatani N, Miyano S, Nakagama H, Nakamura Y, Tsunoda T, Shibata T, Nakagawa H; Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760-764 (2012).

### 2. 学会発表

- 1) Miya F, Morizono T, Abe T, Kubo M, Tsunoda T; Massive genome-wide eQTL analysis in Japanese population. The 35<sup>th</sup> Annual Meeting of the Molecular Biology Society of Japan, Fukuoka (2012).
- 2) Okada Y, Miya F, Koike M, Tomisato S, Kanematsu D, Kanemura Y, Kohda K, Yamanaka S, Yuzaki M, Uchiyama Y, Ikeda E, Okano H; Incompletely reprogrammed human iPSCs form glioma-like tumors through genomic instability during neural differentiation. The 35<sup>th</sup> Annual Meeting of the Molecular Biology Society of Japan, Fukuoka (2012).
- 3) 宮 冬樹, 阿部 哲雄, 森園 隆, 秋山 真太郎, 久保 充明, 角田 達彦; 次世代シーケンサーを用いた日本人における全ゲノム網羅的 eQTL 解析, 日本人類遺伝学会 第57回大会, 東京 (大会最優秀ポスター賞 受賞) (2012).
- 4) 鈴木 直大, 務台 英樹, 鳥居 千春, 清水 厚志, 宮 冬樹, 難波 一徳, 工藤 純, 小崎 健次郎, 松永 達雄; カスタムターゲットリシーケンスによる難聴関連遺伝子の変異探索, 日本人類遺伝学会 第57回大会, 東京 (2012).
- 5) 根岸 豊, 服部 文子, 安藤 直樹, 宮 冬樹, 角田 達彦, 北井 征宏, 荒井 洋, 岡本 伸彦, 加藤 光広, 山崎 麻美, 金村 米博, 小崎 健

次郎, 齋藤 伸治; 網羅的遺伝子解析により  
難治性てんかん患児に同定された複合ヘテロ  
SCN2A遺伝子変異, 日本人類遺伝学会 第57回  
大会, 東京 (2012).

## G. 知的財産権の出願・登録状況

### 1. 特許取得

METHOD FOR SELECTING IPS CELL CLONE,  
AND METHOD FOR SELECTING GENE USED  
IN METHOD FOR SELECTING SAME (iPS 細胞  
クローンの選択方法、及びその選択方法に用い  
る遺伝子の選択方法), 特許出願, 世界知的所  
有権機関(WIPO) 国際公開番号: WO 2012/115270,  
日本出願番号: 特願 2011-040979, アメリカ出  
願番号: 201161466298P, 出願人: Yohei Okada  
(岡田 洋平), Hideyuki Okano (岡野 栄之),  
Tatsuhiko Tsunoda (角田 達彦), Fuyuki Miya  
(宮 冬樹), (2012)

ヒト肝細胞を有するマウスの処置方法. 日本国  
特許取得, PATENT NUMBER 特許第 5025173 号 (出  
願番号 2006-181621, 公開番号: 2007-037542, 登  
録日: 2012.6.29), 国際出願 (PCT) 済 出願番  
号: PCT/JP2006/313077, 出願人: 向谷 知世, 吉  
里 勝利, 舩本 法生, 片岡 美穂, 角田 達彦,  
宮 冬樹, (2012)

### 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（難病関係研究分野）  
分担研究報告書

次世代シーケンサーの臨床診断への応用

分担研究者 黒澤 健司  
地方独立行政法人神奈川県立病院機構 神奈川県立こども医療センター遺伝科 部長

研究要旨

次世代シーケンサーの臨床応用を目的に、臨床検体を用いて遺伝的異質性の高い奇形症候群の遺伝子診断を試みた。LA-PCR とキット化されている custom amplicon との 2 つの方法による target-enrichment で解析を進めた。LA-PCR では Kabuki 症候群 32 例中に、MLL2 変異を 22 例に検出し、KDM6A 変異を 1 例に検出した。このうち 1 例は、変異モザイクで、現在まで MLL2 変異モザイクは報告がない。Custom amplicon では、HaloPlex (Agilent) を使い、Marfan 症候群類縁疾患・Ehlers-Danlos 症候群パネル (MES パネル: 14 症例解析) と Ciliopathy 疾患群・Treacher Collins 症候群混合パネル (Cil パネル: 15 症例解析) を作成し、MES パネルは増幅解析が得られたターゲット領域数は 615 (416,587bp) で、20 reads 以上の領域が 83.12%、100 reads 以上が 52.73% を達成した。病原性変異検出は、Marfan 症候群 6 例中で FBN1 が 2 例、TGFB2 が 1 例、Ehlers-Danlos 症候群 6 例中で COL5A1 が 3 例、COL3A1 が 1 例に検出した。病原性 variant を検出しなかった症例の臨床像から CNV の関与を想定し、アレイ CGH でのスクリーニングを行ったところ、1 例で COL5A1 を含む 9q34.3 に 3.2Mb 重複と微細な欠失を伴う複雑なゲノム再構成を検出した。Cil パネルでは、Joubert 症候群 6 例中 1 例に TMEM67 の複合ヘテロ変異、1 例に CEP290 の nonsense 変異、1 例に CC2D2A の nonsense 変異をそれぞれヘテロのみで検出した。Bardet-Biedl 症候群 1 例に BBS12 の複合ヘテロ変異を、頭蓋外胚葉異形成症 1 例に IFT122 の複合ヘテロ変異を検出した。Treacher Collins 症候群 (TCS) 類縁疾患 (4 例) では、2 例で TCOF1 の missense 変異を検出 (うち 1 例は Nager 症候群) し、variant を検出しない 2 症例に対して GATK で CNV 変換による再評価を行ったところ、1 例に EFTUD2 の一部を含む 17q21.31 の欠失を、1 例に POLRD1 の欠失を認めた。今後の課題として、検査フローの確立、正確な臨床情報と両親の遺伝情報が不可欠であること、次世代シーケンサーの解析能力の向上やライブラリー作成キットの性能の向上など、があげられた。

研究協力者

成戸卓也

神奈川県立こども医療センター遺伝科

井田一美 同遺伝科

黒田友紀子 同遺伝科

大橋育子 同遺伝科

A. 研究目的

難病・がんの多くはゲノム・遺伝子の変異を原因とするため、その診断解析に DNA シーケンス技術を用いることは必須である。従来、Sanger 法がその中心をなしてきたし、今後も最終的には Sanger 法での確認の必要性は変わらない。しかし、解析対象遺伝子の数が増せばその労力は膨大な作業になり、特に難病の中心をなす遺伝性疾患では、解析遺伝子を構成する exon 数は多く、同一遺伝子であっても患者ごとに変異の位置が異なる。さらに遺伝的異質性の高い疾患解析では複数の遺伝子について解析が必要となり、これらを総合すると膨大な作業量となる。次世代シーケンサーは、研究としての活用は勿論であるが、臨床応用としての遺伝子診断にも極めて有用である。現在、国内での次世代シーケンサーを用いた研究は、Exome 解析など網羅的ゲ

ノムの解析が中心で、臨床応用に関する研究は乏しい。解析に必要なパイプラインや解析キットは開発されたものの、実際の臨床診断での活用例が少ないことが理由の一つに挙げられる。Exome 解析を臨床診断に持ち込むことは、現時点では労力やコストの面などから有用とは思われない。しかし、上述の臨床診断技術としては極めて有用な技術と思われる。今回我々は、遺伝性疾患が受診患者の半数以上を占める小児病院において次世代シーケンサーをどのような形で活用していくかその臨床応用の実践についてまとめた。

B. 研究方法

臨床診断で次世代シーケンサーを用いる場合、診断に基づいた対象領域 (遺伝子) は既に絞られている。いわゆる Target-enrichment によるゲ

ノム DNA 処理が基本である。この Target-enrichment の方法は、PCR、custom amplicon、hybrid-capture の 3 つの方法に分類できる。我々の施設では、Long and accurate PCR (LA-PCR) 産物、およびキットを用いた custom amplicon での疾患パネルによるライブラリー作成で解析を進めた。

#### 1) 対象遺伝子 LA-PCR 産物を用いたライブラリー作成による Kabuki 症候群の遺伝子診断

Kabuki 症候群は、特異顔貌、骨格異常、精神遅滞を伴う先天奇形症候群で、12q13.12 にマップされる MLL2 の機能不全が原因である。MLL2 遺伝子は全長約 36kb、54 エクソンからなり、これまでの報告変異は特定のエクソンに集中する傾向はあるものの、診断には全エクソンのシーケンス解析が必要であった。したがって従来のキャピラリーシーケンスでは相当の労力を要していた。イントロンも含めた MLL2 遺伝子変異を高精度かつハイスループットで解析することを目的として解析を進めた。対象は臨床症状の組み合わせから Kabuki 症候群と診断された 32 例で、このうち、3 例は既に Sanger 法で変異を確認していた。方法は、MLL2 遺伝子領域を 5 つに分割し (MLL2\_1a : 7,813bp、MLL2\_2 : 6,156bp、MLL2\_3 : 7,287bp、MLL2\_4a : 7,197bp、MLL2\_5 : 8,268bp) LA PCR を行った。LA PCR の産物 38,192bp を Nextera DNA Sample Prep Kit にてサンプルを処理した。研究途中に Kabuki 症候群第 2 の責任遺伝子 KDM6A が明らかとなり、KDM6A についても同様のシステムを構築し、MLL2 と同時解析を可能にした。32 例は 3 回に分けて Illumina MiSeq でシーケンスを行った。3 回目の解析では、KDM6A を同時解析とし、前 2 回で変異を検出しなかった 7 例も含んだ。各解析はサンプル調整 3 日間、シーケンス 30 時間、解析 5 時間の 5 日間で完了することができた。Nextera Index Kit でインデックスを付加し、10 サンプル (1 回目)、16 サンプル (2 回目)、13 サンプル (3 回目) と多検体同時解析を行った。ソフトウェアは MiSeq に内蔵されている MiSeq Reporter を用いた。

#### 2) Custom amplicon による疾患パネルを用いた奇形症候群遺伝子診断

HaloPlex (Agilent) は、制限酵素による断片化、ライブラリー化されたターゲット領域のプロープのハイブリダイズ、ターゲット領域の精製、PCR によるターゲット断片の増幅、といった一連の処理により複数のターゲット領域 DNA 断片のライブラリーを作成するキットである。この方法のメリットは、全てがキット化されワークフローが明瞭であり、Array capture で必要とされるゲノム DNA の断片化が不要なことにある。限られた解析機器でライブラリーを作成する臨床検査では、極めて有用と考えられる。また、対象遺伝子選択とオリゴプローブ設計も全てメーカー開設のインターネット上で自動化されているため、自由度が高く、遺伝的異質性が

高い小児の先天異常疾患の遺伝子診断では有用である。今回我々は、この HaloPlex (Agilent) を用いて、サンプル DNA のライブラリー作成を行った。

#### a. Marfan 症候群、および Marfan 症候群類縁疾患・Ehlers-Danlos 症候群の疾患パネルによる遺伝子解析

Marfan 症候群は、大動脈基部拡張、特徴的骨格、水晶体亜脱臼などを特徴とする遺伝病で、FBN1 遺伝子の変異を原因とする。診断基準を満たす症例では患者の 7-9 割で FBN1 の変異を検出するが、1-3 割では変異を検出できない。この理由として、共通した臨床像を呈する類縁疾患が複数あり、臨床上鑑別が困難なことがあげられる。したがって、類縁疾患も合わせて網羅的に解析を進めることが理想であるが、FBN1 だけでも exon が 65 個存在し、類似の血管病変を呈する一連の Ehlers-Danlos 症候群責任遺伝子も合わせると数十遺伝子数百エクソンの解析が必要となり、従来の Sanger 法では現実的ではない。そこで我々は、上述の HaloPlex で独自の疾患パネルを作成し、解析を進めた。対象は、文書によるインフォームドコンセントを得た 15 症例で、Marfan 症候群 6 例 (全例 High resolution melting (HRM) 法で FBN1 全 exon スクリーニングを行い病原性が疑われる融解曲線を認めなかった)、Ehlers-Danlos 症候群 6 例 (3 例は HRM 法で COL3A1 スクリーニングを行い病原性を疑われる融解曲線を認めなかった)、Stickler 症候群 1 例、Loeyz-Dietz 症候群 1 例、Cutis laxa 症候群 1 例であった。疾患特異的パネル作成は、Agilent 社開設の専用ウェブサイト SureDesign を経由して作成した。ターゲット領域数は 927 (141,840bp) で、設計によるカバー率は 99.0%、対象疾患とターゲット 29 遺伝子を【表 1】にまとめた。

#### b. Ciliopathy 疾患群および Treacher Collins 症候群混合パネル解析

Ciliopathy は絨毛疾患の総称で、具体的には Bardet-Biedl 症候群、Joubert 症候群、Nephronophthisis、Alstrom 症候群、Meckel 症候群、OFD (oral facial digital) 症候群、Jeune 症候群、頭蓋外胚葉異形成症 (Sensenbrenner 症候群) などが含まれる。それぞれの疾患がさらに遺伝的異質性が高く、Bardet-Biedl 症候群だけでも少なくとも BBS1 から BBS16 までの遺伝子を原因とすると考えられている。しかも、時に上記の Ciliopathy 疾患群の間で臨床症状が重なるために、解析は網羅的である必要があるが、従来の Sanger 法でのスクリーニングは現実的ではない。HRM スクリーニングは安価であるが、サンプル処理の労力は少なくなく、また PCR 条件を設定することも困難を伴う。こうした背景から、この疾患群は海外ではいち早く次世代シーケンサーでのパネル式の解析が取り入れられた (殆どは hybrid capture による)。今回、我々は、上記 Marfan 症候群類縁疾患・Ehlers-Danlos 症候群での経験をもとに、HaloPlex を用いて Ciliopathy

パネルを設計し、解析を試みた。しかし、発生頻度が極めて低く、結果として需要が少ない遺伝子検査を臨床検査のフローに乗せることは難しい。解析規格サンプル数（15）を満たすために、より有効にパネルを設計することが求められる。その結果、Ciliopathy 疾患群とは関連がないものの、遺伝的異質性が最近注目されつつある Treacher Collins 症候群を対象疾患として両疾患群混合パネルを作成した。対象は、臨床症状の組み合わせから診断がなされた 14 例で、Joubert 症候群（6 例）、Bardet-Biedl 症候群（2 例）、頭蓋外胚葉異形成症（2 例）、および Treacher Collins 症候群（TCS）類縁疾患（4 例、うち 1 例は Nager 症候群）であった。Ciliopathy+TCS 混合パネル設計は、Marfan 症候群類縁疾患パネル作成と同様で、ターゲット領域数 1454（398,456bp）、設計上のカバー率 99.4%、ターゲット遺伝子数は 75 であった。

### 3) データ解析ならびに変異検出後の Sanger シーケンス、アレイ CGH 解析

得られたデータの解析は機器付属の Miseq Reporter を用いた。Ciliopathy+TCS パネル（Cil パネル）解析では、Genome Analysis Toolkit（GATK）によるデータ解析も行った。Variant call としてスコアの高い変異は、Sanger シーケンスにより確認した。また、得られた bam file を Integrative Genomics Viewer（IGV）で hg19 上に可視化し、確認した。Variant の病原性については、Human Gene Mutation Database（HGMD）で既知変異と比較参照を行った。Marfan 症候群類縁疾患・Ehlers-Danlos 症候群パネル解析で、variant コールが得られない Ehlers-Danlos 症候群典型 2 症例については、アレイ CGH で CNV 検索を試みた。Treacher Collins 症候群類縁疾患症例で、GATK での解析で CNV が想定される症例についても同様にアレイ CGH での確認をおこなった。

（倫理面への配慮）

解析はこども医療センター倫理委員会での承認を得た。全ての解析において文書による同意を親権者から得て行った。

## C. 研究結果

### 1) 対象遺伝子 LA-PCR 産物を用いたライブラリー作成による Kabuki 症候群の遺伝子診断

1 回目および 2 回目の解析で、既知の SNP を 27 個、未報告の SNP を 19 個見出した。平均カバレッジは 2385 回（1 回目）、1,652 回（2 回目）であった。exon 中で SNP データベース中に報告のない変異を Sanger シーケンスと比較して一致することを確認した。全 32 例中 23 例（MLL2 変異 22 例、KDM6A 変異 1 例）に変異を検出し、9 例は変異がなかった。MLL2 変異の種類としては Frame shift 6 例、Nonsense 12 例、Missense 4 例であった。既知 3 症例のうち 1 例には Sanger シーケンスで見落とされた真の病因変異を確認した。次世代シーケンサーで variant の割合が

30%であったが、サンガーシーケンスでは 25%未満であるモザイク例が 1 例あった。

### 2) Custom amplicon による疾患パネルでの奇形症候群遺伝子診断

#### a. Marfan 症候群、および Marfan 症候群類縁疾患・Ehlers-Danlos 症候群の疾患パネルによる遺伝子解析

実際に増幅解析が得られたターゲット領域数は 615（416,587bp）で、20 reads 以上の領域が 83.12%であった。少なくとも 100 reads 以上が 52.73%であった。

Marfan 症候群 6 例中 2 例に FBN1 病原性変異を、1 例に TGFBR2 病原性変異を検出した。この FBN1 変異 2 症例は HRM スクリーニングで見落とされており、特に 1 例（MS-018; MES-12）は同 exon にコントロール DNA でも正常 SNP が存在したために、融解曲線で明瞭に区別されなかったことが原因であった。Ehlers-Danlos 症候群 6 例中 3 例に COL5A1 の病原性変異を、1 例に COL3A1 病原性変異を検出した。また、病原性 variant を検出しなかった 2 症例は、発達遅滞（軽度）も合併したことから CNV の関与を想定し、アレイ CGH（Agilent 60k）でのスクリーニングを行ったところ、1 例（MES-02）で、COL5A1 を含む 9q34.3 に 3.2Mb 重複と微細な欠失を伴う複雑なゲノム再構成を検出した。Stickler 症候群、Loeys-Dietz 症候群、Cutis laxa 症候群では病原性 variant を検出しなかった。

#### b. Ciliopathy 疾患群および Treacher Collins 症候群混合パネル解析

Joubert 症候群 6 例中 1 例に TMEM67 の複合ヘテロ変異（Cil-13）、1 例（Cil-11）に CEP290 の nonsense 変異を（もう一方の変異は未確認）、1 例（Cil-10）に CC2D2A の nonsense 変異を（もう一方の変異は不明）、Bardet-Biedl 症候群 2 例のうち 1 例（Cil-03）に BBS12 の複合ヘテロ変異を確認、頭蓋外胚葉異形成症 2 例のうち 1 例（Cil-07;同胞例）に IFT122 の複合ヘテロ変異を検出した。Treacher Collins 症候群（TCS）類縁疾患（4 例）では、2 例で TCOF1 の missense 変異を検出（うち 1 例は Nager 症候群（Cil-09））したが、他 2 症例で変異を認めないため、GATK で CNV 変換による再評価を行ったところ、1 例に EFTUD2 の一部を含む 17q21.31 の 37kb の欠失を、1 例に POLRD1 を含む 13q12.2 の 470kb の欠失をそれぞれ認めた。いずれも、文献的に認められる臨床症状と矛盾しなかった。

## D. 考察

次世代シーケンサーの臨床応用を目的に、実際の臨床検体を用いて遺伝的異質性の高い奇形症候群の遺伝子診断を試みた。診断的検査として導入するためには、ワークフローがプロトコル化し、得られた結果の再現性が高く、多検体処理（High-throughput）が可能であることなどが求められる。次世代シーケンサーはその解析能力から多検体処理は得意であるものの、診断的

検査の流れに乗せるには、DNA 処理とデータ解析の工夫は不可欠である。今回、2 つの Target-enrichment の方法によるライブラリー作成から、次世代シーケンサーでの遺伝子診断を試みた。

#### 1) 対象遺伝子 LA-PCR 産物を用いたライブラリー作成による Kabuki 症候群の遺伝子診断

Nextera を用いた LA-PCR と組み合わせた PCR 産物のシーケンス解析は、多検体処理、再現性、フローのプロトコル化、いずれの側面も満たすことが可能であった。さらにこれまで報告のない MLL2 変異モザイクも明瞭に検出できた。このモザイク検出は Sanger シーケンスでは見落とされる可能性があり、次世代シーケンスの特性（高いリード数）を明確に示したものであった。また、研究途中で明らかになった Kabuki 症候群第 2 の責任遺伝子 KDM6A の解析も迅速に対応可能であり、LA-PCR による DNA 処理・ライブラリー作成の柔軟性も示すことができた。今後の課題は、サンプルインデックスの工夫によっては、1 サンプル解析のコストをさらに下げることがあげられる。

#### 2) Custom amplicon による疾患パネルを用いた奇形症候群遺伝子診断

HaloPlex (Agilent) を用いた独自の疾患パネルによる解析を、Marfan 症候群類縁疾患・Ehler-Danlos 症候群 (MES パネル)、および Ciliopathy 疾患群・Treacher Collins 症候群 (Cil パネル) で試みた。MES パネルでの variant 検出率は高く、極めて有用であった。FBN1 では、HRM で多型にマスクされ、見落とされた病原性変異の検出が可能であった。HRM も high-throughput で高い精度が特徴であるが、非病原性 SNP が対象サンプルあるいはコントロールサンプルに含まれると、その鑑別には Sanger シーケンスを用いる他なく、SNP が多くなると作業量も増えてくる。さらに、HRM では適切な PCR プライマーの設計や一定した PCR 条件の設定も不可欠であり、これらを対象疾患に合わせて設計すると膨大な作業量になる (FBN1 (全 65 exon) や COL3A1 (全 51 exon)、COL5A1 (全 66 exon)、COL5A2 (全 54 exon))。したがって検出精度に加え、コストと労力でも、パネルを用いた解析が HRM を凌ぐ可能性も出てくる。また、今回の解析から、本来小さな variant の検出を目的とした custom amplicon でも、データ解析法 (GATK) の工夫により CNV 評価が可能であり、variant が検出されない場合の再評価の方法が明らかにされた。Custom amplicon の CNV 変換による EFTUD2 と POLRD1 の欠失の検出は、これまで報告がない最初の例である。

今回、診断検査として次世代シーケンサーを用いる場合の課題もいくつか提示された。一つは、こうしたパネルを用いた網羅的解析では、検出された variant と臨床診断との整合性の検討が不可欠なために、解析サイドと臨床サイドで密接な情報共有が重要なことである。複数の疾

患を一つのパネルで行う場合には、明らかに臨床症状 (診断) と乖離する variant も多数検出される。データ解析以前に、こうした variant を症状から絞り込むことは重要である。解析を進める上で、データの共有に関するシステムも課題と思われる。第 2 は、遺伝性疾患の遺伝子診断では、遺伝様式と症状から、変異の意義検証として両親解析は必須であるが、今回の解析でもそのことは明らかとなった。特に digenic 変異も含まれる Ciliopathy のような極めて遺伝的異質性が高い疾患では、複数の遺伝子に複数の variant が検出されるため、病原性変異の確定には慎重な検討が必要である。High-throughput を目指すものの、この大きなジレンマを克服するには、データ解析などを整備する必要はあるかもしれない。第 3 は、こうした custom amplicon の Target-enrichment は、常に 100% ではなく、限界を有していることである。今回の MES パネルでも、臨床的に信頼がおける reads 数 100 以上の領域は 52.73% で、read 数 20 でようやく 80% を超えるにとどまった。高い read 数で均一にターゲット領域をカバーするキットの工夫は必要である。今後、次世代シーケンサーの性能の向上 (総 read 数の増加) により、どれだけ厚みが期待できるか検討する必要もある。遺伝性疾患の変異の多様性に合わせたワークフローを検討すべきである。

## E. 結論

次世代シーケンサーの臨床応用を目指し、遺伝性疾患の解析を target-enrichment の方法別に試みた。特定遺伝子に絞り込まれる場合には、LA-PCR が極めて有用で、低頻度モザイクの検出も可能で、柔軟な対応が達成できた。Custom amplicon による疾患別パネルでは、膨大な variant を整理するために、正確な臨床情報、両親の遺伝情報が不可欠であった。同様変異スクリーニングで優れる HRM 以上の効果を期待できることが分かった。次世代シーケンサーの解析能力の向上や、ライブラリー作成キットの性能の向上なども今後の課題としてあげられた。

## F. 研究発表

### 1. 論文発表

- 1) 石川 亜貴、榎本 啓典、古谷 憲孝、室谷 浩二、朝倉 由美、安達 昌功、黒澤 健司 CHARGE 症候群 26 例の臨床的検討 日児誌 2012;116:1357-1364.
- 2) Tachibana Y, Aida N, Enomoto K, Iai M, Kurosawa K. A case of Sjögren-Larsson syndrome with minimal MR imaging findings facilitated by proton spectroscopy. *Pediatr Radiol*. 2012;42:380-382.
- 3) Kurosawa K, Tanoshima-Takei M, Yamamoto T, Ishikawa H, Masuno M, Tanaka Y, Yamanaka M. Sirenomelia with a de novo balanced translocation 46,X,t(X;16)(p11.2;p12.3). *Cong Anom* 2012;52:106-110.
- 4) Enomoto K, Kishitani Y, Tominaga M, Ishikawa