

筑波大学

博士（医学）学位論文

次世代シーケンサーを用いた先天性
神経疾患原因変異の探索と同定

2015

筑波大学

宮 冬樹

目次

I.	要旨	1
II.	略語	3
III.	緒言	5
IV.	背景と目的	8
V.	対象と方法	11
	1. 対象検体	11
	2. 神経系遺伝子カスタムターゲットリシーケンス用プローブの作成	12
	3. 神経系遺伝子カスタムターゲットリシーケンスの実験手法	12
	4. 全 exon をターゲットとした WES の実験手法	12
	5. NGS データの解析プラットフォームの構築とその variant の call 方法	13
	6. NGS で call された variant からの疾患原因変異候補の絞り込み方法	14
	7. NGS の call 評価のための比較用 SNP chip タイピング	15
	8. NGS の call の確認のための Sanger シーケンス法	16
	9. スプライシングバリエーションの確認用 RT-PCR 実験	16
	10. TUBA1A の強制発現実験と免疫組織化学、微小管の密度測定実験	16
	11. 繊維芽細胞を用いた tubulin の解重合実験	17
	12. CDS 補完用カスタムリシーケンス (CCCS) 用プローブのデザイン	17
	13. CCCS の実験手法	18
	14. CCCS の解析手法	18
	15. ヒト遺伝子アノテーションとタンパク質ドメインアノテーションの収集	19
	16. 統計解析	19
VI.	結果	20
	1. 神経系遺伝子ターゲットリシーケンスの NGS call 評価	20
	2. 神経系遺伝子ターゲットリシーケンスでの疾患原因変異探索結果	20
	2-1. 神経系遺伝子ターゲットリシーケンスのデータ量と原因変異同定率	20
	2-2. <i>ACTB</i> 遺伝子変異による Baraitser-Winter 症候群 (BRWS)	21
	2-3. <i>ASPM</i> 遺伝子変異による小頭症 (MCPH5)	21
	2-4. <i>PTPN11</i> 遺伝子変異による脳梁欠損症 (Noonan 症候群)	23
	2-5. <i>DYRK1A</i> 遺伝子変異による大頭症	23
	2-6. <i>AKT3</i> 遺伝子変異による大頭症/巨脳症 (MCAP)	24

2-7.	<i>CASK</i> 遺伝子変異による知能障害と小頭、橋小脳低形成 (MICPCH)	25
2-8.	<i>GABRD</i> 遺伝子変異による Rett 症候群様疾患	26
3.	WES の NGS call 評価	27
4.	WES での疾患原因変異探索結果	27
4-1.	WES のデータ量と原因変異同定率	27
4-2.	<i>KIF1A</i> 遺伝子変異による進行性神経変性症	28
4-3.	<i>NF1A</i> 遺伝子変異による脳梁低形成と尿路異常症	29
4-4.	<i>TUBA1A</i> 遺伝子変異による水無脳症様皮質形成異常	30
5.	これまでの各種 WES プラットフォームでの CDS 領域カバー率の評価	32
6.	CCCS の CDS 領域カスタムデザイン	32
7.	CCCS のデータ量と call の評価	33
8.	WES と CCCS の統合による CDS カバー率	34
9.	WES と CCCS の統合解析により明らかになった疾患原因変異	34
VII.	考察	36
VIII.	引用文献	40
IX.	図表	49
X.	謝辞	92

要旨

2005 年後半に誕生した次世代シーケンサー（NGS）は、ここ 10 年の間に指数関数的に解読可能量と解読速度を増大させ、かつて 10 年以上を要したヒト全ゲノムシーケンスも数日程度で低コストに行えるようになり、ゲノム解析の研究を加速度的に進展させている。疾患研究においても、遺伝性と想定される疾患検体を対象とし、exon 領域を選択的にターゲットとした全 exome シーケンス（WES）手法を用いた研究を中心に、多数の疾患原因遺伝子変異が発見されてきている。

しかしながら、次世代シーケンス解析はまだ新しい手法ゆえにゴールドスタンダードと呼べる確立された手法が無く、如何にして多数 call される variant の中から効率的に疾患原因変異を同定するか、偽陽性と偽陰性を如何にして減らすか、その call 方法の各種パラメーターはどのように設定するのが最善なのか、そして WES を行っても疾患原因変異が見つけれない場合の原因はどこにあるのかという最大の課題がある。WES は遺伝性疾患の原因変異同定に非常に有用ではあるが、実際の WES を行った際の疾患原因変異同定率は世界的に約 25%程度と見積もられている。これらの課題の解決策を模索し、新たな解析手法を開発し、実際の疾患検体を用いて解析を行い、これまで以上の精度と感度で疾患原因変異を同定することが本研究の主目的である。

本研究の対象は先天性の遺伝性疾患と想定される神経疾患（診断名としては小頭症、滑脳症、多小脳回、脳梁欠損、脳梁低形成、小脳低形成、水頭症、大頭症、大脳皮質形成異常、Angelman 症候群、Dandy-Walker 症候群、Phaces 症候群等）の患者およびその家族を対象とした。方法としては、最初に NGS データを解析するプログラムの作製と解析時の各種閾値等のパラメーターの検証を、実際の NGS データを用いて実施した。次に、臨床での応用を考慮に入れ、これまでの既知の論文やデータベース情報を元に、候補を含めた神経系疾患関連遺伝子情報を収集し、神経系遺伝子の 284 遺伝子をターゲットとしたカスタムリシーケンス用プローブを作製し、各種神経系疾患検体で原因変異探索を行った。次に、神経系遺伝子カスタムターゲットリシーケンスでは疾患原因変異が同定できなかった検体および新規検体を対象に全 exon をターゲットとした WES を実施し疾患原因変異を探索した。さらに、WES でも原因変異が未同定であった検体を対象に、WES ではカバーしきれないタンパク質コード領域（CDS 領域）を洗い出し、WES で用いた手法とは別の選択的環状化法と呼ばれる手法で CDS 領域シーケンスを補完するカスタムリシーケンス用プローブを作製して WES 単独では探索困難な疾患原因変異の探索を実施した。

結果として、神経系遺伝子カスタムリシーケンスでは 119 検体中 17 検体(同定率 14.3%)、

WES では 43 家系 (138 検体) 中 20 家系 (同定率 46.5%) で確定的な疾患原因変異を同定することができた。この WES の同定率は世界的に報告されている平均約 25%と比較すると高めであり、その要因は、既知の一般集団頻度情報データベースの整備が進み、それを多数解析に取り入れたことと、閾値等のパラメーターを精査し、偽陽性率と偽陰性率を極力減らすように調整した解析プログラムによるところが大きいと考えられる。

今回発見した変異には Rett 症候群様症状を引き起こす新規疾患原因遺伝子としての *GABRD* 遺伝子や、これまでゲノム上の領域が脳梁低形成等の責任領域候補として知られていながら遺伝子が確定的ではなかった *NFIA* 遺伝子を 1 塩基レベルの変異の発見により確定したものが含まれていた。また、特に神経系疾患の場合は表現型が多岐に渡り、様々な疾患と重複する表現型 (例えばてんかんや精神・運動発達遅滞など) があり、臨床症状だけでは確定診断が非常に難しい場合が多くある上に、症状が典型例とは異なる場合も多い。例えば *ACTB* 変異が見つかった Baraitser-Winter 症候群例では既報の症例よりも形成異常の度合いが軽度で頭囲が平均より大きいなどの特徴があったり、Noonan 症候群の原因遺伝子である *PTPN11* に変異が見つかった患者ではこれまで *PTPN11* 変異症例では報告がない脳梁欠損が見られたり等、これらは NGS で網羅的に探索しなければ確定診断や原因遺伝子の同定が難しく、非典型例の確定診断のためにも NGS は非常に有用であると言える。

WES でも候補すら何も残らない検体が本研究の検体でも約 20.9%存在しており、この原因変異未発見家系検体の原因変異が WES で読まれていない CDS 領域に存在する可能性を考え、本研究では CDS 未読領域を補完するカスタムリシーケンスプローブを作製し統合解析を行った。これまでの WES では 10%程度の CDS 領域がシーケンスできておらず、これを補完するカスタムリシーケンス用プローブを作製することで WES と統合した際の CDS カバー率をプローブのデザイン段階で 98.5%、実際のシーケンスでも最大 97.4%まで上げることができる手法を確立した。この統合手法を用い、WES では何も候補が残らなかった 7 家系について調査を行い、少なくとも 1 家系で新規の疾患原因変異を同定することができた。本手法は WES で原因変異が見つからない場合に適用可能な選択肢の一つとして非常に有用であろうと考えられる。これらの結果は累計 9 報の論文雑誌でも報告を行った。

本論文研究では、神経系遺伝子カスタムターゲットリシーケンス、WES、そして WES で読まれていない CDS 領域を補完する形のカスタムリシーケンスという 3 種の実験手法とバイオインフォマティクスの解析手法を統合して、多数の神経系疾患原因遺伝子の新規変異の同定と、新たな実験・解析手法の開発を行った。今後さらに各種の NGS ハードウェアの改良およびバイオインフォマティクスを駆使したソフトウェアの改良により、一層多くの新規遺伝子を含めた疾患原因変異が同定され、更なる疾患メカニズムの解明や治療へと繋がっていくことが期待され、本研究はそのための実験・解析手法の道標を示した。

略語

- ABR . . . auditory brain-stem response (聴性脳幹反応)
- ALT . . . alanine transaminase (アラニンアミノトランスフェラーゼ)
- AST . . . aspartate transaminase (アスパラギン酸トランスアミナーゼ)
- bp . . . base pairs (塩基対長)
- BRWS . . . Baraitser-Winter syndrome
- bwa . . . Burrows-Wheeler Aligner (マッピングソフト名)
- CCCS . . . complementary custom CDS sequencing (CDS 未読領域補完用カスタム
リシーケンス)
- CDS . . . coding sequence (タンパク質コード配列)
- CGH . . . comparative genomic hybridization (染色体構造異常検査用アレイ試験)
- CNV . . . copy number variation (コピー数多型、コピー数多様性)
- CT . . . computed tomography (コンピュータ断層撮影)
- DNA . . . deoxyribonucleic acid
- EBV . . . Epstein-Barr virus
- EDTA . . . ethylenediaminetetraacetic acid
- EEG . . . electroencephalography (脳波検査)
- GC . . . guanine and cytosine (塩基配列の G と C の総称、用例は GC-content (含量) 等)
- GH . . . growth hormone (成長ホルモン)
- GWAS . . . genome-wide association study (ゲノムワイド関連解析)
- indel . . . insertions and deletions (挿入配列と欠失配列の総称)
- LCL . . . lymphoblastoid cell line (リンパ球由来不死化細胞株)
- MAF . . . minor allele frequency (マイナーアレル頻度)
- MCAP . . . Megalencephaly capillary malformation (巨脳症に血管腫、過成長、合指症
などを伴う症候群)
- MCPH . . . autosomal recessive primary microcephaly (常染色体劣性先天性小頭症)
- MICPCH . . . microcephaly with pontine and cerebellar hypoplasia (小頭症と橋および
小脳低形成)
- MRI . . . magnetic resonance imaging
- ncRNA . . . non-coding RNA (非タンパク質コード RNA)
- NGS . . . next generation sequencer (次世代シーケンサー)

PBS . . . phosphate buffered saline (リン酸バッファー)
PCR . . . polymerase chain reaction
RNA . . . ribonucleic acid
SNP . . . single nucleotide polymorphism (一塩基多型)
SNV . . . single nucleotide variants (リファレンス配列と異なる 1 塩基多様性配列)
UTR . . . untranslated region (非翻訳領域)
VCUG . . . voiding cystourethrogram (排尿時膀胱尿道造影)
VEP . . . visual evoked potentials (視覚誘発電位)
VUR . . . vesicoureteral reflex (膀胱尿管逆流現象)
WES . . . whole-exome sequencing (全エクソームシーケンス)
WT . . . wild type (野生型)

緒言

13年間に及ぶヒトゲノムプロジェクトにより、2003年にヒトゲノムの99%の領域の配列について99.9%の精度で解読終了が宣言され¹⁻³、ヒトのゲノム解析研究における手法は大きく変容した。個々の遺伝子を対象とするだけでなく、全ゲノムレベルでの網羅的解析へと研究の対象を拡大することが可能となった。それ以前にも、例えば現在よりは対象数は相当少ないもののマイクロアレイによる網羅的発現解析⁴やHapMapプロジェクト⁵に代表されるようなゲノムワイドな一塩基多型 (single nucleotide polymorphism, SNP) 解析などは存在したが、リファレンスゲノムの完成によって、より高精度に、より詳細に、ヒトゲノムを研究・解析・理解することが可能となった。疾患研究に目を向けると、2002年から多因子疾患 (心筋梗塞、糖尿病、リウマチなど) について疾患群と健常群の間で頻度に有意差のあるSNPを探索し、そのSNPが存在あるいは影響を及ぼす疾患に関わる遺伝子を調査し疾患のメカニズムの解明や治療ターゲットの発見を目指すゲノムワイド関連解析 (genome-wide association study, GWAS)^{6,7}が世界的に広く行なわれるようになったが、ヒトゲノム配列が完成し、その連鎖不平衡の詳細が研究されて明らかになったことにより、同時に連鎖が起こる一定領域内にタグSNPと呼ばれる代表となるSNPをより高精度に設定できるようになり、少数(数十万)のSNPをジェノタイピングすることでより効率的にGWASが行えるようになった。一方でメンデル遺伝病とも言われる単一遺伝子疾患については、1980年代後半から1990年代に連鎖解析に基づくポジショナルクローニング法によって、ハンチントン病⁸や筋ジストロフィー⁹、嚢胞性線維症¹⁰などの原因遺伝子が次々に明らかとされたが、リファレンスヒトゲノム配列の完成が単一遺伝子疾患の解析でも精度や解析速度の向上に貢献していることは言うまでもない。しかしながら、主に多因子疾患あるいは低い浸透率疾患を対象としたGWASでは多数(数百から数千あるいはそれ以上)の疾患群と健常群の検体が必要なこと、一方で単一遺伝子疾患でも連鎖解析から有意な領域を同定するためには十分な一定数以上の大家系などの検体が必要なことや、実験・解析が容易ではないことなど、ヒトゲノム配列が解読されても各種の課題が存在した。

そんな中、2005年末に次世代シーケンサー (next generation sequencer, NGS) が開発され、再びゲノム研究に大きなブレイクスルーが到来した¹¹。それまでDNAシーケンスに用いられていたサンガーシーケンス法に比べ一度に膨大なシーケンスを産出することが可能なNGSは、ここ10年の間にさらに改良を重ねられ、いわゆるムーアの法則と言われるコンピュータのトランジスタ数が約1.5年おきに約2倍ずつ増加するという速度を遥かに超える速度で指数関数的にその解読可能量を増大させ、2005年に最初に登場した454 Life Science社

製 Genome Sequencer System GS20 という NGS が 20 Mb / run であったのと比較し、2015 年現在の最新の Illumina 社製 HiSeq 4000 という NGS では 1 回 3.5 日の稼働で 1.5 Tb / run ものシーケンスデータを産出することが可能となっており、これは 10 年で 75,000 倍もの増加量である（ムーアの法則では 10 年で約 102 倍）。この 1.5 Tb というシーケンス量は 1 回で約 12 人分の全ゲノムのシーケンス解析に相当する。それと同時にシーケンスコストも格段に下がり（図 1）、一人の全ゲノム当たりのシーケンスコストは場合によっては約 1,000 ドルに届きつつある。ヒトゲノムプロジェクトでかかった費用が約 30 億ドル以上と算定されることに比べると正に隔世の感がある。

この NGS は疾患研究の世界にもブレイクスルーを起こしつつある。例えばこれまでの GWAS 研究においては、主にヨーロッパ系の白人人種に頻度の高い SNP の情報を基にジェノタイピング用の SNP チップが開発されているため、日本人を対象に同じ SNP チップを使ってタイピングすると、GWAS 研究の対象にできる SNP は約半分以下しか得られない場合が多い。これは SNP 塩基のリファレンス配列と異なるマイナーアレル頻度（minor allele frequency, MAF）が人種ごとに異なる場合が多数存在し、SNP チップの設計元となったヨーロッパ系白人人種集団と日本人集団とで大きく MAF が異なる SNP については日本人集団で SNP が検出できない（存在しない）ことが最大の要因である。また同時に日本人集団特有の SNP がこれまでの一般的な SNP チップには載っておらず、まだ多数存在する日本人特有の SNP を見逃していたことにもなる。日本人集団の SNP をより詳細に解析できるようにするため、東北大学を中心とした東北メディカル・メガバンク機構では日本人 1,070 人分の全ゲノムを NGS でシーケンスし、そのデータから日本人集団での SNP を詳細に探索し、日本人の SNP 解析に最適化した 66 万カ所の SNP 解析が可能なジャポニカアレイ SNP チップを 2015 年に作製した¹²。同様に他の人種でも NGS の結果を元にしてその人種の SNP に最適化した SNP チップ等を設計することで、より効率的に見落としを少なく GWAS が実施されるようになっていくことが期待される。一方で、NGS は単一遺伝子疾患を対象とした研究手法にも大きな変貌をもたらしている。解読量の増大と低コスト化が進んだことで、大家系の連鎖解析を経ずともゲノムを一気にシーケンスすることで変異を探索することが可能となり、少なくとも発症者とその両親のトリオ検体があれば疾患原因変異を同定することができる場合も多く存在するようになった。特に単一遺伝性疾患は既知原因変異の 90% 以上が遺伝子の exon 上に存在する（アメリカ国立生物工学情報センター（NCBI）の変異データベース ClinVar の 2015 年 10 月現在の情報に基づく）ことから、全ゲノム中の約 1~2% の全 exon 領域を選択的に NGS でシーケンスすることで疾患原因変異を探索するという戦略が 2010 年頃から積極的に行われるようになり、その手法は全エクソームシーケンス（whole-exome sequencing, WES）解析と呼ばれる^{13,14}。あるいは少数の既知の疾患原因遺伝

子の exon のみを選択的にシーケンスし、原因遺伝子探索を行うターゲットリシーケンス (targeted resequencing) という手法も用いられる場合がある¹⁵。この WES あるいはターゲットリシーケンスにより同定された疾患原因変異が既に数百を超える単位で報告されており、新規原因遺伝子も多数明らかになっている¹⁶。

しかしながら、本格的な活用が始まってまだ間もない NGS による疾患解析ゆえにまだ多くの課題もある。本研究では、各種の先天性神経疾患を対象とし、NGS を用いた疾患原因変異探索を行うと共に、NGS 解析の課題点を挙げ、その対応策・解決策を考案し、新たな疾患原因変異探索手法を開発し、疾患原因変異の同定を試みた。21 世紀の疾患研究は、如何にビッグデータと言われる膨大なデータからバイオインフォマティクスの手法と生物医学的知見を統合させ、疾患の原因変異探索を行うと共に、疾患メカニズムの解明・薬剤ターゲットの探索、そして治療に結びつけていくかが最重要のキーであると考えられ、本研究はその統合解析により得られる実例の実証を行うと共に、疾患研究のさらなる進展に寄与することを目指したものである。

背景と目的

遺伝性疾患に対する NGS を用いた変異解析は、効率性の面からタンパク質コード配列 (CDS) 領域である全 exon (約 50 Mb) をターゲットとした WES^{13,14} か、あるいは既知または候補を含む疾患関連遺伝子をターゲットとしたターゲットリシーケンス¹⁵ が主に実施されている。その手法は何種類か存在するが、ターゲット領域となるゲノム上の DNA を選択的に捕捉しシーケンスするという点では共通しており、大きく以下の3つに分類される。1つ目は、ターゲット領域に相補的な合成 DNA または合成 RNA でできたキャプチャープローブによってターゲット領域のゲノム DNA とプローブを結合させ、その結合した DNA を回収後にシーケンスするキャプチャーハイブリダイゼーション法 (capture hybridization method)、2つ目は複数の制限酵素カクテルによりゲノム DNA を断片化後、両側の制限酵素処理末端配列に重複し一部ターゲット領域配列を含む特異的プローブ (molecular inversion probe; MIP と呼ばれる) を結合させ、そのプローブから末端間の DNA を合成し環状の DNA を作成し、その環状 DNA をキャプチャー後にシーケンスする選択的環状化法 (selective circularization method)、3つ目は末端に NGS 用アダプター配列を付けたプライマーにより一度に複数の PCR を行い、増幅産物を NGS でシーケンスする PCR 増幅法 (PCR amplification method) の3種である (図 2)^{15,17,18}。このうち最も一般的に使用されるのはキャプチャーハイブリダイゼーション法であり、その理由は3種のうち最も多くの領域の捕捉が一度に可能であるためで、CDS 領域のみならず遺伝子の UTR 領域や非タンパク質コード機能性 RNA (non-coding RNA, ncRNA) なども含めた最大 100 Mb 程度のキャプチャーが一度に可能なためである。しかしながら、キャプチャーハイブリダイゼーション法では特異的なプローブが設計できない領域が存在する場合や、塩基配列の GC 含量が高い領域、リピート配列を多く含む領域は解読が難しいという幾つかの問題点もある¹⁹。一方で、選択的環状化法や PCR 増幅法は解読できる量が最大でも 10 Mb 程度が限界の場合が多く、WES にはあまり向かないが、コストが安く済む場合もあり、キャプチャーハイブリダイゼーション法では読みにくいゲノム領域も読めることもある。また選択的環状化法や PCR 増幅法の方がキャプチャーハイブリダイゼーション法よりターゲット領域の捕捉時の特異性が高い場合もある (図 3) など、場合によって手法を使い分けることが望ましい。

NGS は各解読リードが 100~150 bases 前後のショートリードの NGS (Illumina 社 HiSeq シリーズ等)と、各解読リードが数万 bases にも及ぶロングリードの NGS (Pacific Biosciences 社 PacBio RS II など) があるが、コストパフォーマンスとシーケンスエラーの少なさから、現状はショートリードの NGS が使われる場合が多い (ただし最近ではロングリードの NGS

の性能が向上してきており、非モデル生物を対象とする場合は特に *in silico* でのアッセンブリが容易なロングリードシーケンサが有用であることは特筆しておく)。いずれの NGS の場合でもヒトのシーケンスデータを扱う場合の基本的な解析手法はほぼ同一であるが、これ以降は本研究でも用いたショートリードの NGS について主に論ずる。

NGS を用いた基本的な疾患原因変異探索手法は以下の通りである。まずヒト疾患検体 DNA から NGS によって得られたリードを、ヒトリファレンス配列にマッピングする。マッピングソフトはこれまでに複数作製されているが、bwa (Burrows-Wheeler Aligner)²⁰ というソフトウェアが最も一般的に使用されマッピング率も非常に優秀である。その後、一般健常集団に見られる SNP や variant (リファレンス配列と異なり頻度は低いが同様のアレルを持つ健常者がいる多様性のある塩基) を除外し、タンパク質コード領域にあり、アミノ酸変化を伴うダメージが生じると考えられる variant (1 塩基の多様性配列 (single nucleotide variants, SNV) および挿入配列 (insertions) と欠失配列 (deletions) (挿入と欠失の総称を indels と呼ぶ)) に絞り込み、家系情報由来の想定遺伝形質と辻褃が合う variant を探し、最終的にその変異が疾患原因変異かどうかを同定する。流れとしては以上だが、そこには多数の課題が存在する。NGS の解析手法はまだゴールドスタンダードと呼べる確立された方法が存在せず、如何にして多数 call される variant の中から効率的に疾患原因変異を同定するか、偽陽性と偽陰性を如何にして減らすか、その call 方法の各種パラメーターはどのように設定するのが最善なのか、そして WES を行っても疾患原因変異が見つけれない場合の原因はどこにあるのかという最大の課題がある。WES は遺伝性疾患の原因変異同定に非常に有用ではあるが¹⁶、実際の WES を行った際の疾患原因変異同定率は世界的に約 25%程度と見積もられている^{21,22}。これらの課題の解決策を模索し、これまで以上の精度と感度で疾患原因変異を同定することが本研究の目的の一つである。

本研究では研究対象を先天性神経疾患とした。先天性疾患は生活習慣病のような環境要因を含む多因子疾患と異なり単一遺伝子が原因である場合がほとんどであり、また浸透率 (ある遺伝子変異を持っている場合の実際に発病する率) を 100%と仮定した場合、患者と同一 (homozygote あるいは heterozygote の型が同一) の variant 保因者が一般健常者集団で存在した段階で、その variant を疾患原因変異候補から除外できるので、絞り込みが容易かつ確実性が増すという利点がある。本研究のグループは、2011 年に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) において「分野横断型全国コンソーシアムによる先天異常症の遺伝要因の解明と遺伝子診断ネットワークの形成」(慶應義塾大学 小崎健次郎 教授代表) という課題名の元、全国先天性疾患解析コンソーシアムを立ち上げ、本論文の研究はそのコンソーシアム内で実施された研究が主となる。各研究グループメンバーと役割は図 4 の通りであり、本論文著者は神経系先天性

疾患グループの NGS の実験とその解析を主に担当した。

本研究の第一義の目的は先天性疾患原因変異探索とその同定であり、将来的に臨床の現場でのスクリーニングにも応用できる技術の開発を目指した。合わせて今後継続的に利用可能な NGS を用いた包括的な解析システムの構築も目的とした。さらに、WES だけでは原因変異が同定できない場合の更なる探索手法の開発、新規原因遺伝子変異同定からの機能解析により疾患メカニズムの解明をさらに進めること、薬剤ターゲットを探索すること等、今後の疾患治療へと繋がる知見を得ることも目指した。

対象と方法

研究の全体的な方法の流れとして、まず NGS の解析プラットフォームの構築（最初にサーバー構築とプログラムの作製、次いで NGS シーケンス後の call の精度の検証、パラメーターの検証）を行った。なお、NGS 用の計算機サーバーは 2.4 GHz の CPU コア数 32 個、メモリ 256 GB、ハードディスク 72 TB のサーバーをメインで用いた。NGS でのシーケンスに関しては、まず神経系の既知および候補遺伝子を選び出し、カスタムリシーケンス用キャプチャープローブを作成し、先天性神経疾患家系検体に対して NGS による神経系遺伝子のシーケンスを行い、疾患原因遺伝子変異探索を行った。次に、ターゲットリシーケンスで疾患原因変異が見つからなかった家系および別の新規家系に対して WES を実施した。さらに、WES でも候補が見つからなかった家系について、その原因を探索し、追加の WES で読めなかった領域を補完するためのカスタムリシーケンス用プローブを作製後、疾患原因変異探索とその同定を行った。

1. 対象検体

対象とした先天性神経疾患は小頭症、滑脳症、多小脳回、脳梁欠損、脳梁低形成、小脳低形成、水頭症、大頭症、大脳皮質形成異常、Angelman 症候群、Dandy-Walker 症候群、Phaces 症候群等とした。またその家系検体は神経系遺伝子のカスタムリシーケンスの場合は基本的には患者のみ（ただし Sanger シーケンスによる確認時には両親も確認）、そして WES およびその CDS 補完用カスタムリシーケンスの場合は最低でも患者とその親のトリオ以上の人数がいる家系を対象とした。

なお、本研究は理化学研究所および各所の共同研究機関の倫理委員会の承認を得て行われ、患者（またはその保護者）、およびその親族提供者からのインフォームドコンセントを書面でいただき、全て承認された計画の元で実験および解析を実施した。

検体は末梢血とし、Qiagen 社の QIAamp DNA Blood Midi Kit あるいは同等品でキットのプロトコル通り DNA を精製し、それを NGS 用のシーケンスサンプルとした。

また、一部、検証用として、HapMap プロジェクト⁵や 1000 Genomes Project²³で使用されたリンパ球由来の細胞株（EBV-transformed lymphoblastoid cell line, LCL）で日本人由来の検体 (ID: NA18943) を Coriell 社より購入し使用した。細胞は RPMI 1640 培地にウシ血清 10%、Penicillin/Streptomycin 1%を加えた培地で 37°C、CO₂ 濃度 5%にて培養した。数継代後、細胞培養液を 1500 rpm、5 分にて遠心し、上清を破棄後、PBS で一回細胞を洗浄し、その細胞から Qiagen 社の Tissue & Blood Kit にて細胞 DNA をプロトコル通り精製し、NGS 検体とした。

さらに、ある特定遺伝子のヒト組織内での遺伝子発現確認用として、TaKaRa Bio 社よりヒト大脳、胎児脳、脊髄由来の total RNA を購入し実験に用いた。

2. 神経系遺伝子カスタムターゲットリシーケンス用プローブの作成

ターゲットとする遺伝子は、1) 前述の本研究での対象疾患の原因遺伝子として既に知られている遺伝子群、2) 本研究での対象疾患の原因候補として考えられている遺伝子（大きな構造欠失や転座が該当遺伝子上や周囲に知られている遺伝子等）、3) 神経系の関連疾患の原因遺伝子あるいはその候補遺伝子、4) 既知疾患原因遺伝子の遺伝子ネットワーク上に存在する遺伝子、の4種から我々が独自にキュレートし、284 遺伝子を選別した（表1にその遺伝子リストを示した）。選別した遺伝子の CDS 領域に対し、Agilent Technologies 社の SureDesign ソフトウェアを用いてプローブを設計し、Agilent Technologies 社によって SureSelect カスタムキャプチャープローブを合成、作製した。

3. 神経系遺伝子カスタムターゲットリシーケンスの実験手法

検体の血液由来 DNA 3 µg 分を合計 130 µl になるように LowTE buffer（10 mM Tris-HCl pH 8.0, 0.1 mM EDTA 溶液）に溶解させ、Covaris 社の Covaris Shearing System にてピークが 150 ~200 bp となるように機器のプロトコル通り DNA を断片化させた。断片化後の DNA から SureSelect XT Target Enrichment System for Illumina Paired-End Sequencing Library Kit（Agilent 社）と上記で設計した神経系遺伝子カスタムターゲットリシーケンス用 SureSelect プローブとを用いて、キットのプロトコル通り NGS 用シーケンスライブラリを調製した。その後、Illumina 社の Genome Analyzer IIX または同社の HiSeq 2000 NGS にて 76-bp または 101-bp の paired-end 法にてプロトコル通りシーケンスを実施した。Genome Analyzer IIX NGS を用いる場合は1レーン当たり異なる index 配列を付けた6検体のプールにて、HiSeq 2000 を用いる場合は異なる index 配列を付けた1レーン当たり24検体のプールにてシーケンスを実施した。

4. 全 exon をターゲットとした WES の実験手法

検体の血液由来 DNA 3 µg 分を元に、上記の神経系カスタムターゲットリシーケンス用の場合と同様に DNA の断片化を行った。WES 用のキャプチャープローブは SureSelect V4 または V5（Agilent Technologies 社）を使用し、そのプローブと SureSelect XT Target Enrichment System for Illumina Paired-End Sequencing Library Kit（Agilent Technologies 社）を用いて、プロトコル通り断片化後の DNA から NGS 用シーケンスライブラリを作製した。その後、Illumina 社の HiSeq 2000 NGS にて 101-bp の paired-end 法にてプロトコル通りシーケンスを

実施した。WES の場合は 1 レーン当たり異なる index 配列を付けた 6 検体のプールにてシーケンスを実施した。

5. NGS データの解析プラットフォームの構築とその variant の call 方法

解析パイプラインの概要を図 5 に示した。NGS によって読まれた read は、まず bwa ソフトウェア²⁰ (<http://bio-bwa.sourceforge.net>) にてヒトゲノムリファレンス配列にマッピングした。ヒトリファレンス配列は、ヒトゲノム build 37/hg19 のゲノム配列に、decoy 配列と呼ばれる配列がさらに追加された 1000 Genomes Project²³ のリファレンスとしても使用された hs37d5.fa という配列を用いた (以下の ftp サイトよりダウンロード可能: ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz)。bwa は bwa mem というオプションが追加された新しいバージョン (version 0.7.10 以降) を用いた。そのコマンドラインは以下のように実施した。

```
> bwa mem -M hs37d5.fa ssample1_forward.fastq.gz sample1_reverse.fastq.gz > mapped_read.sam
```

次に、マッピングされた read に対して、picard ソフトウェア (<http://broadinstitute.github.io/picard/>) にて染色体の位置順に並べ替えつつ、sam フォーマットファイルを bam フォーマットに以下のコマンドで変換した。

```
> java -jar picard.jar SortSam SORT_ORDER=coordinate INPUT=mapped_read.sam OUTPUT=mapped_read_sorted.bam
```

マッピングされた read に対して、picard ソフトウェアにて PCR エラーが多く含まれることが知られる完全同一重複リード (duplication read) を以下のようなコマンドで除外した。

```
> java -Dsnappy.disable=true -jar picard.jar MarkDuplicates I=mapped_read_sorted.bam O=mapped_read_sorted_remDup.bam METRICS_FILE=jeter.metrics REMOVE_DUPLICATES=true ASSUME_SORTED=true VALIDATION_STRINGENCY=LENIENT
```

次に、ゲノム上にユニークにマッピングされている read のみを samtools ソフトウェア²⁴ (<http://samtools.sourceforge.net>) にて以下のコマンドで抜き出した。

```
> samtools view -bq 1 mapped_read_sorted_remDup.bam > mapped_read_sorted_remDup_uniq.bam
```

続いて、自作プログラムにて全 paired-end (読まれたシーケンスライブラリ由来の read のペアとなる forward 側と reverse 側の read) の平均距離を算出し、その ± 2 SD (standard deviation) 以内に存在する read だけを選出した。その後、プローブが設計されたターゲット領域の ± 100 bp 以内にオーバーラップする read のみに絞り込んだ。

以上のように抜き出したターゲット領域上の read に対して、samtools にて次の 2 つのコマンドで 2 段階に SNV と indel を call した。

```
> samtools pileup -Q 20 -Bvcf human_reference_build37.fa ontarget_reads.bam >
pileup_result_step1.txt
```

```
> samtools.pl varFilter -Q 10 -d 1 -D 10000 pileup_result_step1.txt > pileup_result_step2.txt
```

さらに SNV に関しては、samtools で call された SNV の consensus score が 20 以上、かつ SNP quality score が 20 以上、かつ read depth が 10 以上の SNV に絞り込んだ。また、ゲノム上の同一箇所にもたがる read 塩基におけるアレル頻度が、0.25 以下あるいは 0.75 以上の箇所を homozygote として、またその頻度が 0.25 より大きく、かつ 0.75 より小さい箇所を heterozygote として call し、samtools 上での call とその allele 頻度が一致しなかった場合は NGS で読めなかった場所として call なしとした (図 6)。

一方、indel に関しては、GATK ソフトウェア²⁵ (<https://www.broadinstitute.org/gatk/>) を用いて、samtools で SNV と indel が call された領域上の read について再度 local realignment という再マッピング処理を行った。これは SNV や indel が call された領域はマッピングエラーが起こりやすいため、local realignment によって、より正確にマッピングし直すためである。local realignment 後は SNV の場合と同様に samtools によって再度 call した。indel についても consensus score が 20 以上、かつ SNP quality score が 20 以上、かつ read depth が 10 以上で read depth が 10 以上に絞り込み、ゲノム上の同一箇所にもたがる read 塩基におけるアレル頻度が、0.25 以下あるいは 0.75 以上の箇所を homozygote として、またその頻度が 0.25 より大きく、かつ 0.75 より小さい箇所を heterozygote として call し、samtools 上での call とその allele 頻度が一致しなかった場合は NGS で読めなかった場所として call なしとした。

6. NGS で call された variant からの疾患原因変異候補の絞り込み方法

call された variant から先天性疾患原因変異を探すため、各種データベースから情報を収集し、一般健常者集団でアレル頻度が 0.001 以上 (ただし homozygote の保因者がいる場合はアレル頻度に関わらず) の保因者がいずれかのデータベースに存在する variant を除外することとした。用いたデータベースは以下の通りである。

- ・ dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>)
- ・ 1000 Genomes Project (<http://www.1000genomes.org>)
- ・ NHLBI Exome Sequencing Project ESP6500 (<http://evs.gs.washington.edu/EVS/>)
- ・ ExAC (<http://exac.broadinstitute.org>)
- ・ 京都大学 HGVD (<http://www.genome.med.kyoto-u.ac.jp/SnpDB/>)
- ・ 東北メディカル・メガバンク (<http://ijgvd.megabank.tohoku.ac.jp>)

上記データベース情報をすべて統合し、フィルターとして用いた。ただし、以下の 2 種のヒト疾患変異データベースに乗っているアレルで、頻度が 0.01 よりも低い variant は、その

疾病情報を付加しつつを除外せずに残した。

- ・ NCBI ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar/>)
- ・ HGMD (<http://www.biobase-international.com/product/hgmd>)

また、これまでに実施した自分らの NGS の in-house データを統合し、in-house 内の健常者で homozygote の保因者がいる SNV は除外し、heterozygote に関しても 10 人以上保因者がいる SNV は common variant とし疾患原因候補から除外した。

次に上記フィルターを通して残った variant に対してアノテーション (variant の存在する遺伝子名や、遺伝子上での場所情報、アミノ酸の変化情報など) を annovar²⁶ ツール (<http://annovar.openbioinformatics.org/en/latest/>) を用いて付加し、アミノ酸変化を伴う nonsynonymous 変異 (missense 変異、nonsense 変異、あるいは splice site 領域に存在する variant) を選び出した。

最後に、variant のダメージ予測 (タンパク質の重要なドメイン等に存在しないかや、立体構造の変化によるダメージ予測、進化的保存性から見たダメージ予測等) の情報を付加した。これは SIFT²⁷ や Polyphen2²⁸ などの多数のダメージ予測プログラムを統合して実行できる dbNSFP^{29,30} (<https://sites.google.com/site/jpopgen/dbNSFP>) というツールを使用し付加した。ただしそれらのダメージ予測結果は付随的な情報としてのみ使用し、ダメージ情報による候補 variant の除外や選別は行わなかった。

7. NGS の call 評価のための比較用 SNP chip タイピング

NGS の call 評価のために、一部の検体を用いて SNP chip にてタイピングを実施し、NGS の call と比較した。SNP chip は 951,117 箇所のジェノタイピングが可能な Illumina 社の HumanOmniExpress Exome chip を用い、その実験手法はすべてプロトコル通り実施した。SNP chip は triallele (3 種の allele の SNP が存在する場所) や indel が存在する場所では call エラーが起こりやすいため、1) 既知のデータベースで triallele が存在する場所、2) 我々の in-house の NGS call に triallele が存在する場所、3) 既知のデータベースで indel が存在する場所、4) 自身の in-house の NGS call に indel が存在する場所、の 4 種のいずれかに該当する箇所は SNP chip との比較対象領域から除外して比較した。

また、神経系遺伝子カスタムリシーケンスの評価の場合のみは、日本人由来の LCL 細胞 (ID: NA18943) を我々の神経系遺伝子カスタムリシーケンスで実施した上で、そのデータとダウンロードした HapMap プロジェクトで実施された同一検体のデータ (ftp://ftp.ncbi.nlm.nih.gov/hapmap/genotypes/2010-08_phaseII+III/) とを比較することで評価した。

8. NGS の call の確認のための Sanger シーケンス法

まず検体 DNA 1 ng 分をテンプレートとして、確認したい塩基を含むように設計した primer を用いて PCR を実施し、PCR 増副産物を得た。primer 配列は表 3、表 7、表 13 等の該当箇所に記載した。PCR はまず 94 度で 2 分間変性ステップを行い、その後 94 度で 30 秒、60 度で 30 秒、72 度で 30 秒というステップを 35 サイクル繰り返したのち、最後に 72 度で 5 分間追加の伸長時間を入れ、その後 4 度で保温というプログラムで実施した。PCR の *Taq* Polymerase は Promega 社の GoTaq Green Master Mix を使用した。PCR 産物は 2%アガロースゲルで電気泳動後、目的の長さのバンドが目視できることを確認し、その PCR 産物のバンドをアガロースゲルから切り出し、Promega 社の Wizard SV Gel and PCR Clean-Up System キットにて精製し、Sanger シーケンス用 DNA 検体溶液とした。Sanger シーケンスは Life Technologies 社の Applied Biosystems 3730xl DNA アナライザを用いてプロトコル通り実施した。サンガーシーケンスの波形データは 4Peaks ソフトウェア (<http://nucleobytes.com/4peaks/>) を用いて目視でも確認した。

9. スプライシングバリエントの確認用 RT-PCR 実験

Nuclear factor I/A (*NFIA*) 遺伝子に関しては、4 種のスプライシングバリエントが知られている。どのバリエントがヒト脳内で発現しているかを確認するため、RT-PCR 実験を行った。胎児脳由来、成人大脳由来、成人小脳由来、脊髄由来のそれぞれの total RNA 1 μ g を逆転写酵素 SuperScript III (Life Technologies 社) とランダムプライマー ($pd(N)_6$) を用いてプロトコル通り逆転写して cDNA (complementary DNA) を合成し、そのうち 1/100 量を RT-PCR に用いた。プライマーは forward 側が intron を跨るように設計し、forward 側の配列が 5'-CAT TAC AGG ACC CAG AGC AAG -3'、reverse 側の配列が 5'-GCT GCA ACT TTT ATC CCA GGT -3'のプライマーを用いた。PCR 反応条件は上記の Sanger シーケンス法と同一とした。

10. TUBA1A の強制発現実験と免疫組織化学、微小管の密度測定実験

TUBA1A 変異の機能解析のため、全長の *TUBA1A* を購入した胎盤由来の cDNA をテンプレートに PCR (F primer: TAAGCGGCCCGCCATGCGTGAGTGATCTCCATCCAC、R primer: TAACTGCAGGTATTCCTCTCCTTCTTCCTCACCCTC) にて増幅させ、*Not* I と *Pst* I の制限酵素を用いて pCMV-4A ベクター (Agilent Technologies 社) にクローニングした。この発現用ベクターは C 末端に FLAG (DYKDDDDK) エピトープを付加したタンパクを発現するようになっている。変異型 *TUBA1A* は PCR を用いて野生型を改変して作製し、全ての発現用ベクター配列は Sanger シーケンスにて間違いがないことを確認した。トランスフェクションは COS7 細胞へ行い、Lipofectamine 2000 (Life Technologies 社) 試薬を用いてプロトコル

通り実施した。トランスフェクションの 24 時間後に氷冷したメタノールで固定し、goat polyclonal anti-DYKDDDDK (FLAG) 抗体 (1 : 5,000、Novus Biologicals 社) または mouse monoclonal anti- α -tubulin 抗体 (1 : 250、SantaCruz 社) の 1 次抗体と、donkey anti-goat IgG Alexa Fluor 594 (1 : 1000、LifeTechnologies 社) または donkey anti-mouse IgG Alexa Fluor 488 (1 : 1000、LifeTechnologies 社) にて染色した。染色された FLAG または α -tubulin の密度測定は Kumar らの方法³¹と同様に行い、ImageJ ソフトウェアと KBI プラグインを用いて、対象の細胞のある面積中に占める対象の染色された微小管の面積の割合の比率にて算出した。

11. 繊維芽細胞を用いた tubulin の解重合実験

TUBA1A 変異の機能解析のため、TUBA1A 変異 (p.R64W および p.C25F) を持つ各患者と変異を持たない健常者の皮膚生検由来の繊維芽細胞を培養して実験に用いた。解重合実験は Poirier らの方法³²と同様に実施した。簡潔に書くと、10%ウシ血清と Penicillin/Streptomycin 1%を添加した IMDM 培地で 3 継代培養後、カバーガラススライド上で培養し、それを氷上に 0、5、10、15、または 20 分間静置後、細胞を固定した。再重合の実験では氷上で 20 分間静置後に 37 度に移し 15 分間静置してから細胞を固定した。細胞は mouse monoclonal anti- α -tubulin 抗体 (1 : 100、SantaCruz 社) の 1 次抗体と、donkey anti-mouse IgG Alexa Fluor 488 (1 : 1,000、LifeTechnologies 社) にて染色した。解重合は顕微鏡下で各時間における細胞 100 個当たりの解重合している細胞数を計測して割合を算出した。

12. CDS 補完用カスタムリシーケンス (CCCS) 用プローブのデザイン

WES で読まれていなかった CDS 領域についてカスタムでリシーケンス用プローブをデザインした。デザインは WES 用のキャプチャープローブである SureSelect V5 を用いて日本人由来の LCL 細胞である NA18943 検体を NGS にて読んだデータに基づいて設計し、NA18943 検体の CDS 領域の depth (読み深度) が 15 未満の領域をターゲットとした。なお、CDS 領域かどうかのアノテーションデータはヒトリファレンスゲノム build 37.3 の NCBI RefSeq データベースの遺伝子情報から判別した。上記の CDS の読み深度が浅い領域について、Agilent Technologies 社の HaloPlex テクノロジー (その手法は SureSelect がキャプチャーハイブリダイゼーション法であるのに対し HaloPlex は選択的環状化法を用いている) を使ってプローブをデザインすることとした。プローブのデザインは Agilent Technologies 社の SureDesign ソフトウェア (<https://earray.chem.agilent.com/suredesign/?locale=ja>) を用いたが、以下の 3 ステップを経てデザインを行った。パラメーターの詳細は上記 WEB サイトに記載されている。

- 1) ターゲット領域に対し”Maximize Specificity”というパラメーターにてプローブをデザインし、その後、独自にデザインされたプローブを用いて NGS から読まれることが予想

される read をシミュレーションし、そのシミュレートされた paired-end の read を BLAT³³ にてヒトゲノムにマッピングした。その paired-end の read の距離が 1,000 bp 以内で、かつ最小のミスマッチ数でマッピングされる場所が一意に決定できる read を産出すると予想されるプローブのみを残した。

- 2) 1) の結果、ターゲット領域の中でまだ予想シーケンス領域に含まれていない領域について、さらに SureDesign ソフトウェア上で、“Maximize Specificity”に追加して“Optimize for Fragmented Samples”オプションを選択してプローブをデザインした（このオプションを選択すると短い増幅産物も許すという条件でより多くのプローブがデザインできるようになる）。その後、1) の場合と同様に NGS から読まれることが予想される read をシミュレーションし、その read を BLAT にてマッピングし、ユニークにマッピングできる優良なプローブのみを残した。
- 3) 2) の結果、ターゲット領域の中でまだ予想シーケンス領域に含まれていない領域について、さらに SureDesign ソフトウェア上で、“Maximize Coverage”と“Optimize for Fragmented Samples”オプションを選択してプローブをデザインした（“Maximize Specificity”を“Maximize Coverage”に変更するとプローブの領域特異性がかなり落ちるが設計される数が増える。したがって次の BLAT により実際は大部分が捨てられるが一部はユニークにマッピングされる）。その後、1), 2) の場合と同様に NGS から読まれることが予想される read をシミュレーションし、その read を BLAT にてマッピングし、ユニークにマッピングできる優良なプローブのみを残した。

最後に 1), 2), 3)のそれぞれのステップで設計できたプローブを合算し、Agilent Technologies 社にて実際のプローブを合成した。

13. CCCS の実験手法

初期量 225 ng のゲノム DNA を検体として、上記で作製したカスタムプローブと、HaloPlex Target Enrichment System For Illumina Sequencing Kit（Agilent Technologies 社）を用いて NGS 用シーケンスライブラリをプロトコル通り作製し、Illumina 社の NGS HiSeq 2000 にて 100-bp paired-end にてシーケンスを実施した。シーケンス時は 1 レーン当たり異なる index 配列を付けた 24 検体のプールにてシーケンスを実施した。

14. CCCS の解析手法

解析パイプラインの概要を図 7 に示した。大部分は WES および神経系遺伝子カスタムリシーケンスの場合と同様だが、以下の 3 点が異なる。一つ目は、選択的環状化法を用いた短い増幅産物由来の read は、読まれた read の後半にターゲット配列に続いて NGS 用のアダプ

ター配列が突き抜けて読まれる場合があるので、Cutadapt ソフトウェア (<http://cutadapt.readthedocs.org/en/latest/index.html#>) を用いて余分なアダプター配列を除去した。そのコマンドは以下の通りである。forward 側の read に対しては、

```
> cutadapt -overlap 5 10 -m 50 -a AGATCGGAAGAGCACACGTCTGAACTCCAG -o  
output.fastq input.fastq
```

にて、reverse 側に関しては、

```
> cutadapt -overlap 5 10 -m 50 -a AGATCGGAAGAGCGTCGTGTAGGGAAAGAG -o  
output.fastq input.fastq
```

にて実行した。二つ目は、duplication read の除去を実施しなかった。これは、選択的環状化法では PCR にて大量に同一 amplicon を増幅させるため、大部分の read が duplicate となるためである。三つ目は、最初の bwa にてリファレンスゲノムへのマッピングがうまくできなかった read と、paired-end の read 間の距離が 1,000 bp 以内に存在せずに除去された read、ターゲット領域にマッピングされずに除去された read に関して、BLAT にて再度マッピングをし直した点である。BLAT の条件はデフォルト (5 base までのミスマッチを許す) で実施し、paired-end の read の距離が 1000 bp 以内で、かつ最小のミスマッチ数でマッピングされる場所が一意に決定できる read を再度リカバーした。

15. ヒト遺伝子アノテーションとタンパク質ドメインアノテーションの収集

ヒトの遺伝子情報 (タンパク質コード領域や遺伝子名等) は NCBI のビルドバージョン build 37.3/hg19 の seq_gene.md というファイルから入手した (ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/mapview/seq_gene.md.gz)。

ゲノムのビルドが異なる情報を build 37.3/hg19 に揃えるためには、UCSC の liftOver ツールを用いた (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>)。

タンパク質のドメイン情報は EMBL-EBI の InterPro データベース (<http://www.ebi.ac.uk/interpro/>) より入手し、ドメイン領域の作図には MyDomains ツール (<http://prosite.expasy.org/cgi-bin/prosite/mydomains/>) を用いた。

16. 統計解析

One-way ANOVA は実施後、post-hoc test (事後検定) として Tukey's test を実施した。Fisher's exact test (Fisher の正確確率検定) は多重検定補正として Bonferroni 補正を実施した。統計解析は Linux の R ソフトウェアにて実施し、統計的有意差の基準の閾値は $P = 0.05$ 未満を有意差ありとした。

結果

1. 神経系遺伝子ターゲットリシーケンスの NGS call 評価

神経系遺伝子 284 遺伝子 (表 1) のターゲットリシーケンスを評価用の LCL 細胞 (ID: NA18943) を用いて実施し、その得られた NGS call と、HapMap データベースに存在する同検体の SNP chip を用いたジェノタイプデータとを比較して評価を実施した。NGS call と HapMap データの双方でジェノタイプデータが存在する箇所が 3,129 箇所存在し、そのうち 16 箇所は call されたジェノタイプデータが不一致であった (一致率 = $((3,192 - 16) / 3,192) \times 100 = 99.50\%$ 、表 2)。この不一致だった 16 箇所について、我々の NGS call データと HapMap データのいずれが合っているのかを確認するため、Sanger シーケンスによって確認実験を行った。その結果、16 箇所全てについて我々の NGS call が正しかった (表 3)。この結果から、他の HapMap データと一致していた箇所が正しい call だと仮定した場合、我々の NGS の call は疑陽性率、偽陰性率共に 0 となり、少なくとも疑陽性率は $1 / 3,192 = 0.031\%$ 未満、偽陰性率は $1 / 957$ (HapMap と比較可能だった領域のうちリファレンスと異なる SNP 数) = 0.104% 未満であると言える。

2. 神経系遺伝子ターゲットリシーケンスでの疾患原因変異探索結果

2-1. 神経系遺伝子ターゲットリシーケンスのデータ量と原因変異同定率

神経系遺伝子ターゲットリシーケンスは、小頭症家系 28 検体、皮質形成異常家系 22 検体、水頭症家系 11 検体、小脳形成異常症家系 11 検体、Angelman 症候群家系 11 検体、脳梁欠損症家系 9 検体、脊髄髄膜瘤家系 6 検体、大頭症家系 4 検体、てんかん家系 3 検体、等、正常コントロールの LCL 細胞 1 検体を含めて 120 検体を実施した。1 検体当たりの平均 read 量は 1.65 Gb で、リファレンスゲノムへのマッピング成功率は 99.17%、ターゲット領域 (probe 設計領域とその前後 100 bp) である 1.63 Mb の平均 depth (読み深度) は 819.79、probe 設計領域における depth 10 以上の割合 (base call に必要な最低限の depth を超えていた割合) は 99.63% であった。一般集団が保因する既知の variant を除いた後に残った exon 領域でアミノ酸変化を伴う nonsynonymous な variant の数は 1 検体当たりの平均で 8.4 個であった。

解析結果として、正常コントロール以外の 119 検体に関して、疾患原因変異を調査した結果を表 4 に示した。既知の変異および既知の疾患原因遺伝子上に変異が存在するか、あるいは遺伝子機能のアノテーション情報から見つかった変異が疾患原因変異であることが確定的な検体は 14 検体で全体の 11.8% であった。また、確定的ではないが、過去の機能解析の文献等から見つかった変異が疾患原因であることが強く疑われる検体が 3 検体あった。

上記2種の場合をリシーケンスによる疾患原因変異同定成功症例とすると、全体の14.3%が原因変異を同定できたということになる。また、表4に示した通り、一般集団の保因 variant を除外後に何らかの候補となりうる variant が残った検体が39.5%、何も候補が残らなかった検体が46.2%という結果であった。以下に、疾患原因変異が同定できた、あるいは原因変異として極めて有力な変異を同定できた症例をいくつか記載する。

2-2. *ACTB* 遺伝子変異による Baraitser-Winter 症候群 (BRWS)

患者は3歳1ヶ月の女児で近親婚ではない健常な日本人の両親の第一子として生まれた。女児は妊娠37週目に通常の出産で生まれ、出生体重は2,844 gで身長は48.3 cm、頭囲は35.2 cmであった。彼女の成長過程は著しい遅れが見られ、頸座は生後6ヶ月、寝返りは9ヶ月、座位は18ヶ月、独歩は認められず、最近、補助ありの歩行が可能となった。また有意語に乏しく、彼女の発達指数 (global developmental quotient) は2歳児の段階で39であった。自閉症症状は見られなかった。1歳2ヶ月の段階で発達の遅れと頭囲の拡大を指摘され受診に至った。MRI画像より厚脳回症を認め(図8)、脳波検査 (EEG) で棘波を認めた。また、特異顔貌 (内眼角贅皮、耳介低位、眼瞼下垂、アーチ状の眉毛など) を認めた。2歳6ヶ月の段階で、彼女の身長は86.3 cm (-0.8 SD)、体重は12.3 kg (平均内)、頭囲は50 cm (+1.2 SD) であった。Gバンド染色による染色体異常や、CGHアレイによるゲノム構造異常は見受けられなかった。

神経系遺伝子ターゲットリシーケンスの結果、*ACTB* (actin, beta; GeneID: 60, RefSeqID: NM_001101.3) 遺伝子に、c.733G>A, p.G245S の heterozygous な missense 変異を同定した。この変異は既存のどのデータベースにも登録がなく、in-house 検体にも見られなかった。SIFT²⁷によるダメージ予測スコアは0 (damaging) で、MutationAssessor³⁴による予測でもHigh (ダメージ) であった。両親のSangerシーケンスの結果、両親には変異が認められず、*de novo* 変異であることが確認された。症状と遺伝子から、この女児は Baraitser-Winter 症候群 (BRWS, OMIM #243310)^{35,36} と診断された。BRWSは各種の先天性異常と精神遅滞を有し、眼瞼下垂を含む特徴的な顔貌と神経細胞異常を伴う疾患として Baraitser と Winter により提唱された疾患³⁵で、2012年にその原因遺伝子の一つとして *ACTB* が発見された³⁷。本症例は同一遺伝子内の新規変異パターンによる同一疾患として論文雑誌にも報告した³⁸。

2-3. *ASPM* 遺伝子変異による小頭症 (MCPH5)

患者は8歳の男児で近親婚ではない健常な日本人の両親の第四子として生まれた(図9a)。男児は妊娠39週目に通常の出産で生まれ、出生体重は3,160 g (-1.0 SD) で身長は48 cm (-0.9 SD)、頭囲は29.5 cm (-2.7 SD) であった。男児の他の3人の兄弟は健常で、最年長の兄に

は小児期欠伸てんかんが見られた。頸座は生後4ヶ月、座位は20ヶ月で、その後顕著な成長の遅れが見られ、痙性つい麻痺が見られた。発達指数は3歳児の段階で30であった。4歳の段階で日常的な強直発作が見られ、carbamazepine、valproate、levetiracetam、lamotrigineのいずれの薬剤にも不応性であった。15週齢の時点(図9b)と5歳児の時点(図9cおよびd)でのMRIを図9に示したが、側脳室の拡大と厚脳回が認められた。8歳児の段階で体重は15.1 kg (-2.3 SD) で身長は114 cm (-2.2 SD)、頭囲は41 cm (-7.6 SD) であった。男児の筋肉は上肢で硬直と軽微な痙縮が見られ、下肢では重度の痙縮が見られた。男児には独歩不能や有意語の欠乏等の重症心身障害が見られた。8歳児の段階でのEEGでは右前頭葉にfrequent spikeとslow waveの複合を認めた。聴性脳幹反応(auditory brain-stem response, ABR)と血液検査結果には異常は見られなかった。

神経系遺伝子ターゲットリシーケンスの結果、最初のcallの段階で1,907個のSNVと227個のindelが検出された。既知の一般集団が保因するvariantや、synonymous variantを除外した結果、最終的に4個のSNVと1個のindelに絞られ、そのうちの2個は*ASPM*(abnormal spindle-like microcephaly associated、GeneID: 259266、RefSeqID: NM_018136)遺伝子上に存在する遺伝子上に存在するheterozygousなvariantであった(図9e)。常染色体劣性先天性小頭症(autosomal recessive primary microcephaly, MCPH)は4つの診断基準があり、小頭症であること(通常は頭囲が3SD以下)、知的障害があること、てんかんを除く他の神経系の異常が見られないこと、顔面奇形や他の器官の奇形が見られないこと、の4点で、これまでに*MCPH1*、*WDR62* (MCPH2)、*CDK5RAP2* (MCPH3)、*CASC5* (MCPH4)、*ASPM* (MCPH5)、*CENPJ* (MCPH6)、*STIL* (MCPH7)、*CEP135* (MCPH8)、*CEP152* (MCPH9)、*ZNF335* (MCPH10)、*PHCI* (MCPH11)の10遺伝子が原因遺伝子として知られている^{39,40}。*ASPM*は生前および生後の神経新生に関与していることが知られており、*ASPM*の変異は脳のサイズの減少を引き起こす³⁹。*ASPM*はMCPHの中でも最も主要な原因遺伝子で、およそ40%のMCPHが*ASPM*の変異によるものと推定されている⁴¹。*ASPM*変異により発症する小頭症であるMCPH5 (OMIM #608716)は中度から重度の知的障害と、中度の運動発達障害、正常あるいは軽度の大脳皮質奇形が知られている⁴⁰。今回発見された*ASPM*変異はexon 11のc.3055C>T (p.R1019*)とexon18のc.6750delT (p.F2250Lfs*10)であり、双方ともSangerシーケンスで確認された(図9f)。前者の変異は母親が保因者で、後者の変異は父親が保因者のcompound heterozygoteであった。前者のc.3055C>T (p.R1019*)変異はdbSNPデータベースにもID: rs199422153として登録されている既報の疾患原因変異であり⁴¹、後者の欠失変異はデータベースに一切登録のない新規の変異であった。MutationAssessor³⁴によるダメージ予測でも双方ともHigh(ダメージ)であった。原因遺伝子としてよく知られている遺伝子であること、変異は一方がnonsense変異で、もう一方もframeshiftが起こる欠失変異で

あったこと（重要なドメインと考えられる calmodulin-binding IQ ドメインに停止コドンが引き起こされること）、表現型が一般的な疾患の表現型と合致することから、上記 *ASPM* 変異による MCPH であるという確定診断に至った。*ASPM* による変異で小頭症と同時に厚脳回が見られるケースは稀であることを含め、本変異は論文雑誌にも報告した⁴²。

2-4. *PTPN11* 遺伝子変異による脳梁欠損症（Noonan 症候群）

患者は 5 歳男児で近親婚ではない健常な日本人の両親の第一子として生まれた。男児は妊娠 40 週目に通常の出産で生まれ、出生体重は 2,930 g で身長は 47 cm、頭囲は 30 cm であった。新生児の段階で難聴が明らかになった。頸座は生後 4 ヶ月、寝返りは 6 ヶ月で見られた。初期より発達障害が見られ、座位は 14 ヶ月で、独歩は 3 歳児の段階で確認され、その段階で身長は 90.7 cm (-1.8 SD)、体重は 14.3 kg (-0.4 SD)、頭囲は 48.3 cm (-1.1 SD) であった。4 歳児の段階での MRI を図 10 に示したが、脳梁欠損症であることがわかる。染色体構造異常検査結果は正常であった。男児の形成異常症状、成長障害、脳梁欠損から Mowat-Wilson 症候群が当初疑われたが、その原因遺伝子である *ZFH1B* (*ZEB2*) には Sanger シーケンスで変異が見つからなかった。

神経系遺伝子ターゲットリシーケンスの結果、*PTPN11* 遺伝子 (protein tyrosine phosphatase, non-receptor type 11, Gene ID: 5781, RefSeqID: NM_002834.3 の exon 3 に c.188A>G (p.Y63C) の heterozygous な missense 変異を発見した。Sanger シーケンスの結果、両親はこの変異を保有しておらず、*de novo* 変異であることが確認された。この変異は同一変異が Noonan 症候群 (OMIM #163950) の原因遺伝子として報告されており⁴³、難聴症状や発達障害等の表現型は疾患と合致していたが、*PTPN11* 変異による脳梁欠損は過去に報告がなく (Noonan 症候群の他の原因遺伝子では稀に脳梁欠損が知られている)、*PTPN11* 変異による Noonan 症候群の新規の表現型として論文雑誌にも報告した³⁸。

2-5. *DYRK1A* 遺伝子変異による大頭症

患者は 7 歳女児で近親婚ではない健常な日本人の両親の第二子として生まれた。女児は妊娠 39 週目に誘発分娩による出産で生まれ、出生体重は 3,250 g で身長は 50.9 cm、頭囲は 34 cm であった。女児は摂食量が少なく、早期の乳児期から著しい成長障害が見られた。頸座と寝返りは生後 10 ヶ月で見られ、座位は 12 ヶ月で確認された。独歩は確認されておらず、精神発達障害も重度であり、会話語も確認されていない。経管栄養も必要であった。また強度の弱視も確認され、30 cm 以内のものしか認識できなかった。自傷行動、癇癩、口蓋振動による音声チックも見られた。しばしば意味のない笑いも見られた。女児は自閉症スペクトラム障害と診断された。彼女の発達指数は視覚障害のため適切に評価できなかった。

た。女兒には発熱性痙攣やてんかんの発作は見られなかった。身体的特徴所見として、前頭隆起、両眼隔離症、眼振、内眼角贅皮、鞍鼻、両側の耳介低位、短い人中、高口蓋、下向きの口角、小顎症を含む形成異常が見られた。聴覚および腹部検査結果は問題なかった。女兒の現在の体重は 14.6 kg (-1.8 SD) で身長は 103.5 cm (-3.1 SD)、頭囲は 52 cm (+0.6 SD) であり、相対的大頭症 (relative macrocephaly) と診断された。脳の CT と MRI の所見は正常であった (図 11)。網膜の異常や視神経形成異常も確認されなかった。心臓と腹部のエコーも正常であった。脳波検査 (EEG) でもてんかん性脳波を確認できなかった。染色体 G バンド検査も正常で、CGH アレイによる染色体構造変化も確認されなかった。

神経系遺伝子カスタムリシーケンスの結果、*DYRK1A* (dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A、GeneID: 1859、RefSeqID: NM_001396.3) 遺伝子の exon 11 に c.1699C>T (p.Q567*) の heterozygous な nonsense 変異を同定した。*DYRK1A* は神経分化や神経新生を通して脳の成長に関与していることが知られており、その領域はダウン症の責任領域でもある。*DYRK1A* 遺伝子異常 (OMIM #614104) あるいはその領域の異常は一般的に小頭症を発症することが知られているが、それ以外に重度の精神発達遅滞、てんかん、成長遅滞、摂食障害、本疾患で見られたような特徴的顔貌も知られている⁴⁴⁻⁴⁹。*DYRK1A* 遺伝子異常症例に多く見られる小頭症ではなく、本症例では大頭症症状が見られる理由は不明だが、重度の精神発達遅滞等の症状が合致することから、*DYRK1A* が本症例の責任遺伝子と見られ、視覚障害が見られるという新規の知見も併せて論文雑誌にも報告した³⁸。

2-6. *AKT3* 遺伝子変異による大頭症/巨脳症

患者は生後 4 ヶ月の男児で、最後の MRI と血液採取をした生後 4 ヶ月の時点から 1 ヶ月後の生後 5 ヶ月の段階で自宅にて睡眠時に突然死した。近親婚ではない健常な日本人の両親の第二子として生まれた。男児は妊娠 39 週目に吸引分娩により出生し、出生体重は 3,464 g (+1.0 SD) で身長は 58 cm (+4.3 SD)、頭囲は 38.5 cm (+3.7 SD) で過成長を伴う頭囲拡大を認めた。生後 5 日目の MRI では両側シルビウス裂近傍に多小脳回を認めた (図 12 a)。生後 4 ヶ月時の MRI (図 12 b) では脳室拡大を認めなかったが、小脳の肥大による後頭蓋窩の狭小化、脳梁の肥大を認めた。生後から頭囲は拡大し、生後 3 ヶ月の時点で 4 SD を超えていた (図 12 c)。発達遅滞も重度で、哺乳児の咽せも確認されていた。啼泣時に前額部に血管腫様の淡い発赤が認められた。Megalencephaly capillary malformation (MCAP、巨脳症に過成長、血管腫、合指症などを伴う症候群) を臨床的に疑い経過観察していたが生後 5 ヶ月の時点で突然死した。死亡時の CT では硬膜下血腫など新たな病変は認められず、小脳の肥大との関連性が示唆された。

生後 4 ヶ月の時点で採取していた血液検体を用いて神経系遺伝子カスタムリシーケンスを実施した結果、*AKT3* (v-akt murine thymoma viral oncogene homolog 3、GeneID: 10000、RefSeq ID: NM_005465.4) 遺伝子に c.686A>G (p.N229S) の heterozygous な missense 変異変異を同定した。両親も含め Sanger シーケンスによる確認の結果、両親はこの変異を保因しておらず、*de novo* 変異であることが確認された (図 12 d)。MCAP は 2012 年に PI3K-AKT-mTOR シグナル伝達系に参与する *PIK3R2*、*AKT3*、*PIK3CA* が責任遺伝子であることが明らかにされ⁵⁰、既知の遺伝形質が常染色体優性遺伝であること、臨床診断の基準⁵¹⁻⁵⁴ からも *AKT3* 変異による MCAP であると確定診断に至った。巨脳症により水頭症や小脳扁桃ヘルニアを来すことがあり、突然死の報告も散見されることから、本症例のように臨床診断が難しい場合に早期の遺伝子診断が外科的介入 (大孔部減圧術や水頭症手術) の適応決定の一助となる可能性がある指摘を含め、本結果は論文雑誌にも報告した⁵⁵。

2-7. *CASK* 遺伝子変異による知能障害と小頭、橋小脳低形成 (MICPCH)

患者は 13 歳女兒で近親婚ではない健常な日本人の両親の第一子として生まれた。女兒は満期妊娠の通常出産で生まれ、出生体重は 2,918 g で身長は 50 cm、頭囲は 33 cm (-0.1 SD) であった。生後 1 ヶ月の段階で喘鳴、成長障害 (頭囲が 33 cm (-3.1 SD) という小頭症所見) が見られたため受診に至った。脳の MRI で小脳と橋の低形成が確認され、視神経萎縮も見られた。頸座は生後 3 ヶ月で見られたが、その後は重度の成長の遅れが見られた。生後 6 ヶ月ぐらいから四肢の筋緊張亢進が確認された。乳児期の EEG ではてんかん性放電は確認されなかった (図 13 a)。13 歳の段階で、身長は 137.0 cm (-3.0 SD)、体重は 15.0 kg (-3.8 SD)、頭囲は 41.5 cm (13 歳の段階で健常な生後 5 ヶ月の平均に当たる) で重度の小頭症であった。卵型顔、大きな眼、広い鼻梁、上顎前突などの特徴的な顔貌が見られた。女兒は四肢と体幹に硬直と進行性の胸部の変形、脊椎側湾症、両側の膝と肘に関節拘縮が見られた。彼女は単独の座位が不能で、言語の理解も不能であった。MRI では重篤な橋および小脳の形成不全が確認された一方で脳梁は保持されていた (図 13 b)。てんかん性攣縮は、EEG によるてんかん性放電の確認を含めて、生後 3 年 8 ヶ月の時点から見られるようになった。11 歳の時点で特に両側前頭部に頻繁に多発的な棘徐波 (hypsarrhythmia と呼ばれる波形ではない) が見られた (図 13 c - e)。てんかん性攣縮の発作は、sodium valproate、zonisamide、clobazam、lamotrigine などの各種の抗てんかん薬では効果が見られなかった。通常の染色体検査では異常は見受けられなかった。

神経系遺伝子カスタムリシーケンスの結果、*CASK* (calcium/calmodulin-dependent serine protein kinase、GeneID: 8573、RefSeqID: NM_003688.3) 遺伝子の exon 21 に c.1896dupC (p.C633fs*2) の 2 塩基 C による heterozygous な重複変異を同定した。Sanger シーケンスに

より、両親はこの変異を保因しておらず、患者のみの *de novo* 変異であることが確認された。*CASK* 遺伝子は小頭症と橋および小脳低形成 (microcephaly with pontine and cerebellar hypoplasia、MICPCH、OMIM #300749) の責任遺伝子として知られており^{56,57}、本性例の女兒は *CASK* 変異による知能障害を伴う MICPCH であることが確定診断された。

本症例のように、生後 1 年以降の遅発性てんかん性攣縮は珍しく、乳児期 (生後 1 年未満) に発症し、てんかん性攣縮と *hypsarrhythmia* を特徴とする West syndrome (infantile spasms とも呼ばれる) と、本症例の遅発性てんかん性攣縮は区別されている^{58,59}。現段階で知る限りでは、*CASK* に変異を持ち、てんかん性攣縮が見られた患者はこれまでに本症例を含めて 9 例報告されている⁶⁰⁻⁶⁴ (表 5)。これまでに報告された症例を見ると、女兒の場合は男児よりも発作発症年齢が遅めの傾向が見られた。発作発症年齢を含め、新規領域の *CASK* 変異症例として本結果を論文雑誌でも報告した⁶⁴。

2-8. *GABRD* 遺伝子変異による Rett 症候群様疾患

患者は 12 歳女兒で近親婚ではない健常な日本人の両親の第二子として生まれた。女兒は妊娠 41 週目に誘発分娩による出産で生まれ、出生体重は 2,920 g で身長は 48 cm であった。女兒の成長は乳児期から著しい遅れが見られ、筋緊張低下も見られた。生後 6 ヶ月時の筋生検では特に異常は確認されなかった。座位は生後 4 年目に確認され、独歩は未だに確認されていない。精神遅滞も重度で、会話後は見られない。知能指数は 9 歳児の段階で 12 であった。彼女には手揉み動作や歯ざしりなど常同的行為が見られた。意図的な手動作は見られない。異常な呼吸パターンは確認されていない。EEG では後頭部に high voltage slow spike と混合波を認めた。彼女は症状から Rett 症候群 (OMIM #312750) と推定された。熱痙攣やてんかん発作は確認されていない。現在の彼女の身長は 137 cm (-3.4 SD)、体重は 35 kg (-2.1 SD)、頭囲は 51 cm (-1.8 SD) であった。脳 CT や MRI、代謝機能試験、G バンド染色体試験、CGH アレイによる構造染色体異常検査では異常は見受けられなかった。

神経系遺伝子カスタムリシーケンスの結果、*GABRD* (gamma-aminobutyric acid (GABA) A receptor delta、GeneID: 2563、RefSeqID: NM_000815.4) 遺伝子の exon 5 に c.498G>A (p.M166I) と c.499G>A (p.D167N) の隣り合う 2 塩基で 2 つのアミノ酸が変わる heterozygous な missense 変異を同定した。Sanger シーケンスにより、両親はこの変異を保因しておらず、患者のみの *de novo* 変異であることが確認された。既知の Rett 症候群の責任遺伝子 (*MECP2*) や Angelman 症候群、Pitt-Hopkins 症候群の責任遺伝子群には変異が見つからなかった。*GABRD* は主要な抑制系神経伝達物質である GABA のレセプターとなるイオンチャネルのサブユニットで⁶⁵、*GABRD* を含む他の GABA レセプターである *GABRA1*、*GABRB3*、*GABRG2*、*GABRD*

はてんかんと関連が示唆されており、*GABRD* の存在する 1 番染色体の 1p36 領域の欠損はてんかん発作を発症するケースがいくつか報告されている⁶⁶。また *GABRB3* の heterozygous な機能喪失は Angelman 症候群との関連も示唆されている⁶⁷。本症例は初めて *GABRD* 遺伝子 (変異) が Rett 症候群症状を引き起こす可能性を示唆した結果であり、本結果は論文雑誌でも報告した³⁸。

3. WES の NGS call 評価

WES を実施した検体と同一検体 7 検体を Illumina 社の HumanOmniExpress Exome という SNP chip でジェノタイプングし、WES のデータと比較し評価を行った。LCL 細胞の NA18943 検体だけは、WES の capture probe のプラットフォームを SureSelect V4 と V5 の 2 種で実施したので、それぞれの WES の結果と比較した (合計 7 検体 8 種の WES の SNP chip データの比較)。NGS call と SNP chip データの双方でジェノタイプデータが存在する箇所が 1 検体当たり平均で 243,573 箇所存在し、その一致率は平均で 99.95%であった (表 6)。一致率を調べた 8 種のデータの中で最も不一致数の多かった NA18943 検体の SureSelect V4 のデータの 140 箇所について、我々の NGS call データと SNP chip データのいずれが合っているのかを確認するため、Sanger シーケンスによって確認実験を行った (表 7)。その結果、140 箇所中 77 箇所は我々の NGS call が正しかった (表 7 および 8)。この結果から、他の SNP chip と一致していた箇所が正しい call だと仮定した場合、我々の WES の call の疑陽性率は 0.021% (50/243,573)、偽陰性率は 0.064% (13/20,375) であった (表 8)。なお、Sanger シーケンスがうまくできなかった箇所は我々の call が間違いであるという保守的な算出方法の結果である。

4. WES での疾患原因変異探索結果

4-1. WES のデータ量と原因変異同定率

WES は、皮質形成異常 13 家系 39 検体、小頭症 11 家系 35 検体、脳梁欠損症 4 家系 12 検体、水頭症 3 家系 9 検体、等、正常コントロールの LCL 細胞 1 検体を含めて 44 家系 139 検体 (50 人の患者と、その家族 88 人と、正常コントロール LCL の 1 検体) を実施した。1 検体当たりの平均 read 量は 6.50 Gb で、リファレンスゲノムへのマッピング成功率は 99.74%、ターゲット領域 (probe 設計領域とその前後 100 bp) である約 89 Mb の平均 depth (読み深度) は 53.15、probe 設計領域における depth 10 以上の割合 (base call に必要な最低限の depth を超えていた割合) は 97.35%であった。一般集団が保因する既知の variant を除き、家系の表現型の遺伝形質に合致し、exon 領域でアミノ酸変化を伴う nonsynonymous な variant の数は 1 検体当たりの平均で 3.28 個であった。

解析結果として、正常コントロール以外の 43 家系（138 検体）に関して、疾患原因変異を調査した結果を表 9 に示した。既知の変異および既知の疾患原因遺伝子上に変異が存在するか、あるいは遺伝子機能のアノテーション情報から見つかった変異が疾患原因変異であることが確定的な検体は 13 検体で全体の 30.2%であった。また確定的ではないが、過去の機能解析の文献等から見つかった変異が疾患原因であることが強く疑われる検体が 7 検体あった。上記 2 種の場合をリシーケンスによる疾患原因変異同定成功症例とすると、全体の 46.5%が原因変異を同定できたということになる。また、表 9 に示した通り、一般集団の保因 variant を除外後に何らかの候補となりうる variant が残った検体が 32.6%、何も候補が残らなかった検体が 20.9%という結果であった。以下に、疾患原因変異が同定できた、あるいは極めて原因変異として有力な変異を同定できた症例をいくつか記載する。

4-2. *KIF1A* 変異による進行性神経変性症

患者は 8 歳男児で近親婚ではない健常な日本人の両親の第三子として生まれた。妊娠 40 週目に通常出産で出生した。出生体重は 3,700 g で、身長は 51 cm、頭囲は 34 cm であった。乳児期より筋緊張低下と発達の遅れが見られた。また、トランスアミナーゼ上昇も見られた。患者の乳児期のアスパラギン酸アミノトランスフェラーゼ（AST）の値は 77 - 170 IU/L（正常値は<30 IU/L）、アラニンアミノトランスフェラーゼ（ALT）は 58 - 206 IU/L（正常値は<30 IU/L）であった。G バンド染色体試験結果では異常は見られなかった。生後 8 ヶ月の時点の MRI で、脳室拡大、脳梁低形成、小脳虫部低形成を認めた（図 14）。頸座 18 ヶ月、2 歳で寝返り、3 歳で座位保持が確認された。3 歳児の段階で身長は-3 SD 以下で、成長ホルモン分泌不全があり、GH 治療開始となる。4 歳時に全身痙攣があった。EEG で diffuse small spikes の脳波異常が見られたが、投薬なしで経過観察となった。頭囲は 6 歳時で 47.5 cm（-2.4 SD）であった。8 歳時に支えると少し座れるようになった。下肢腱反射亢進、痙性、左股関節脱臼、眼振、追視不良、視神経萎縮、視覚誘発電位（VEP）反応不良が見られた。有意語がなく、重度精神遅滞の状態である。睡眠時無呼吸、神経因性膀胱、慢性便秘も認めた。

本患者と両親の WES を実施した結果、家系の表現型から遺伝形質に合致する変異として *KIF1A*（kinesin family member 1A、GeneID: 547、RefSeqID: NM_004321.6）遺伝子の exon 3 に c.296C>T（p.T99M）の heterozygous な missense 変異を同定した。Sanger シーケンスにより、両親はこの変異を保因しておらず、患者のみの *de novo* 変異であることが確認された。

KIF1A はシナプス小胞を軸索上で輸送するモータータンパク質であり⁶⁸、*Kif1a* KO マウスは軸索や神経細胞の脱落を起し致死性であることが知られている⁶⁹。2011 年に Hamdan らが今回見つかった変異と完全に同一の変異で、同じ疾患の表現型を報告しており⁷⁰、本症例は 2 報目の *KIF1A* 変異による新規の進行性神経変性症と考えられる疾患として論文雑誌

にも報告した⁷¹。なお、我々の報告の後、2015年にも *KIF1A* 変異による同様の3例目となる症例が報告されている⁷²。

4-3. *NFIA* 変異による脳梁低形成と尿路異常症

患者は5歳男児で近親婚ではない健常な日本人の両親の第一子として生まれた。妊娠28週目で脳梁形成不全が認められ、妊娠41週目に頭囲拡大と妊娠期間の延長を理由に帝王切開にて出生した。産後5分の段階のApgarスコアは9点で、出生体重は3,180g (+0.4 SD) で頭囲は38.2 cm (+3.3 SD) であった。特にその他の奇形は認められなかった。生後3日目のMRIにより、大脳半球間の嚢胞と、脳室拡大、脳梁低形成が確認されたが、生後の状態は良好であった。頸座は生後4ヶ月、独歩は1年3ヶ月で確認された。有意語は生後2年1ヶ月で確認され、若干の遅れが見られた。4歳の段階でTanaka-Binet intelligence Scale法で知能指数は75であった。生後11ヶ月時にEEGにより前頭部で鋭波が確認されたが、現在までてんかん発作は確認されていない。4歳時の再度のMRIで、前頭葉に多小脳回が確認され、大脳半球間の嚢胞と、大脳縦裂、脳室には以前と変化は見られなかった(図15 a および b)。5歳時の段階での脊髄のMRIには異常が見られなかったが、膀胱拡張と水腎が見られた。5歳時の排尿時膀胱尿道造影(VCUG)ではgrade IVの膀胱尿管逆流現象(VUR)が確認された(図15 c)。頭囲は56.1 cm (+3.8 SD) で、進行性の頭囲拡大は認められなかった。軽度の大頭症と、広額、薄上唇という僅かながらの顔面形成異常が見られた。

本患者と両親のWESを実施した結果、家系の表現型から遺伝形質に合致する変異として唯一、*NFIA* (nuclear factor I/A, GeneID: 4774, RefSeqID: NM_001134673.3) 遺伝子の exon 8 に heterozygous な c.1094delC (p.P365Hfs*32) の1塩基Cの欠失変異を同定した(図16)。Sanger シーケンスにより、両親はこの変異を保因しておらず、患者のみの *de novo* 変異であることが確認された(図16 d)。この欠失によりアミノ酸114個が欠けると予想された。*NFIA* は複数の isoform が存在し、isoform 2だけが後半の exon をタンパク質コード領域から欠く short type の *NFIA* を産生するが、この変異があると双方共にアミノ酸の欠失が起こると予想された(図16 e)。どの isoform が脳内で発現しているかを健常者の胎児脳、大脳、小脳、脊髄のRNAを用いてRT-PCRによって確認した結果、short type の isoform 2の発現量が最も少なく、それ以外の long type の isoform の *NFIA* の発現量が多く、少なくとも脳内では long type の *NFIA* がより多く存在することが確認された(図16 f)。

これまで9例の *NFIA* 遺伝子を含む周辺領域に大きな変異(6報は欠失、2報は転座、1報は非遺伝子領域の欠失)があることにより本症例と同様の脳梁低形成と尿路異常症が発症したという報告⁷³⁻⁷⁹が存在し、*NFIA* 遺伝子が責任遺伝子であろうという推測はされてい

たが、本症例は初めて 1 塩基レベルで *NFL1A* が責任遺伝子であることを実証した。本結果は論文雑誌にも報告した⁸⁰。

4-4. *TUBA1A* 変異による水無脳症様皮質形成異常

患者は 2 家系の 2 名おり、一人目は 3 歳女児で、健康な近親婚ではない両親の第二子として、妊娠 37 週目に帝王切開にて出生した。その家系の第一子目の長女は健康で正常な成長が見られた。患者は妊娠 28 週目の超音波検査で成長障害、小頭症、脳室拡大が確認された。彼女の Apgar スコアは出生 1 分後が 3 点で、5 分後が 5 点であった。彼女は自発呼吸ができず、人工呼吸器が必要であった。出生児の身長は 44 cm (-1.8 SD)、体重は 2,116 g (-2.0 SD)、頭囲は 29.6 cm (-2.4 SD) であった。彼女は小頭症、小眼症、離れた眼、小顎症が見られた。四肢の攣縮と麻痺も確認された。眼科検査で視神経形成不全も確認された。妊娠 28 週と生後 6 日目の MRI で、薄脳実質、脳幹・小脳・脳梁形成不全が確認された (図 17 a - d)。各種感染性試験は陰性であった。染色体試験結果は 46 本の XX 型で正常であった。生後、強直間代てんかん発作が見られた。EEG において、てんかん発作時にデルタ波と背景脳波の極度の微弱化が確認された。フェノバルビタール薬剤により部分的に発作が抑制された。1 歳時に反復性誤嚥性肺炎により器官切開を行った。その際に胃瘻造設と噴門形成術も実施された。中枢性尿崩症、下垂体機能低下症のためトリクロルメチアジド、ヒドロコルチゾン、レボチロキシンの投与も必要であった。頭囲拡大のため、生後 7 ヶ月の時点で脳室腹腔短絡術 (シャント術) が実施された。3 歳児の現段階で四肢の痙性麻痺で、周囲の環境に対する反応はない。

二人目の患者は 2 歳男児で、上記家系とは別の健康な近親婚ではない両親の第二子として、妊娠 39 週目に通常分娩出生した。胎児期から脳室拡大が確認されていた。その家系の第一子目の長女は健康で正常な成長が見られた。男児の Apgar スコアは出生 1 分後が 8 点で、5 分後が 9 点であった。出生児の身長は 49.5 cm (+0.3 SD)、体重は 2,792 g (-0.9 SD)、頭囲は 33 cm (-0.2 SD) であった。生後の MRI で脳回欠損が確認された (図 17 e - h)。男児は自発呼吸と飲み込みが可能であった。生後 8 ヶ月でてんかん発作が見られたが、バルプロ酸ナトリウムとゾニサミドでコントロールが可能であった。生後 2 歳時に痙性麻痺が見られ、寝返りができない状態であった。アイコンタクトや会話は不可能であったが、男児の名前を呼ぶと声に応じて笑顔が見られた。

本 2 例の患者とその双方の両親の WES を実施した結果、家系の表現型から遺伝形質に合致する変異として 3 個ずつ候補が残ったが、その両患者に共通の遺伝子として、1 例目の患者からは *TUBA1A* (tublin alpha 1a, GeneID: 7846, RefSeqID: NM_006009.3) 遺伝子の exon 2 に heterozygous な c.190C>T (p.R64W) の missense 変異が、2 例目の患者からは同遺伝子の

exon 2 に heterozygous な c.74G>T (p.C25F) の missense 変異が同定された (図 18 a)。Sanger シーケンスにより、両家系とも両親はこの変異を保因しておらず、患者のみの *de novo* 変異であることが確認された (図 18 b)。他の残り候補は神経系とは無関係な遺伝子であった。上記の2つの変異は両方とも PolyPhen-2 と SIFT によるダメージ予測でタンパク機能にダメージを受けるという予測結果であった。両変異共に種間のアミノ酸保存性の高い場所に位置しており (図 18 c)、TUBA1A のタンパク質内では両変異共に N 末端側に位置した (図 18 d)。

TUBA1A は tubulin という細胞骨格や細胞分裂、神経細胞の伸長や神経伝達物質の輸送などに重要な機能を果たす微小管の構成タンパク質の主要な構成要素で⁸¹、*TUBA1A* は α -tubulin タンパク質をコードし、神経細胞においても発現している。tubulin は主に α -tubulin と β -tubulin のヘテロ二量体 (ヘテロダイマー) から成り、細胞中の微小管は主に α -tubulin と β -tubulin のヘテロダイマーが 13 個繋がったプロトフィラメントという構造で管を構成している^{82,83} (図 19 a および b)。脳の形成・発達にも重要であり、tubulin 遺伝子群 (*TUBA1A*⁸⁴、*TUBB2B*⁸⁵、*TUBB3*^{32,86}、*TUBB5*⁸⁷、*TUBB4A*⁸⁸、*TUBB2A*⁸⁹、*TUBG1*⁹⁰) の変異は神経系の構造や形成異常を引き起こすことが知られている。その形成異常は脳回欠損や厚脳回、多小脳回、小頭症などを含む^{83,87,91}。TUBA1A の変異だけでも小脳低形成を伴う脳回欠損、大脳基底核の形成異常、脳梁の薄化あるいは欠損、先天性小頭症、脳室拡大、海馬や脳幹の形成異常、シルビウス裂周囲の多小脳回、多小脳回様皮質形成異常症、精神遅滞、両側麻痺、四肢麻痺、てんかんなどが報告されており^{31,92-96}、その表現型は多岐にわたる。

今回同定された変異箇所のアミノ酸である R64 (N 末端から 64 番目のアルギニン) と C25 (N 末端から 25 番目のシステイン) は立体構造内で側方に位置し (図 19 a および b)、微小管構造内での相互作用に関与していると予測された。R64 は α -tubulin の H1'-S2 ループと呼ばれる箇所に位置し (図 19 c)、tubulin の構造の構成や維持に重要な役割を果たす領域と見られ、この箇所の変異は構造の破綻に繋がる可能性がある。一方、C25 は α -tubulin の helix H1 と H1-H1'ループと呼ばれる境界上に位置し (図 19 c)、tubulin 管内の内側に面しており、H1-H1'ループは H1'-S2 ループと H2-S3 ループと呼ばれる側鎖相互作用の補足的機能を有すると過去に報告されている⁸²ことから、この C25 の変異も tubulin の構造や機能に支障をもたらす可能性がある。我々は COS7 細胞に野生型 (wild type、WT) の TUBA1A と、今回見つかった変異 2 種の p.R64W と p.C25F、そして過去に脳回欠損の原因変異であることが報告⁹⁷されている p.R402C の合計 4 種を強制発現させて tubulin の密度を調査した。その結果、tubulin の存在密度が WT に比べ変異型では下がっていることが確認された (図 19 d および e)。また、tubulin は重合と解離や脱重合を常に行っており、その平衡が tubulin の構造の機能や維持に極めて重要であることから、変異型の解重合の速度が機能不全に関連している

可能性が考えられた。そこで以前論文報告された方法と同様に、氷冷時の解重合の速度を調査した結果、変異型は野生型に比べ、有意に早い時間で解重合が起こっていた（図 19 f および g）。以上のことから、*TUBA1A* 変異による α -tubulin 変異型は野生型に比べ tubulin の脱解離が起こりやすく、それに伴い tubulin の密度（量）が低く、そのことが神経組織内の tubulin の機能不全に繋がり疾患を発症していることが想定された。本変異と実験結果は論文雑誌にも報告した⁹⁸。

5. これまでの各種 WES プラットフォームでの CDS 領域のカバー率の評価

表 9 に示したように本研究で WES を実施しても疾患原因変異候補が何も候補が残らなかった家系が 43 家系中 9 家系（20.9%）存在した。原因変異未発見の理由は様々考えられるが、単一遺伝性疾患は既知原因変異の 90%以上が遺伝子の exon 上に存在することを鑑みると、WES で未発見の変異がまだ exon 上に存在する可能性が考えられた。そこで、本研究で WES にて使用した SureSelect V4 および V5 で exon の CDS 領域をどの程度カバーできているのかを確認することとした。我々のデータに加え、一般的に市販され使われている他の WES のプラットフォーム（NimbleGen (Roche 社) の SeqCap と Illumina 社の TruSeq の 2 種）についても、Clark らの論文¹⁹ データを使用して同様に CDS カバー率の解析を実施した。

結果を表 10 に示したが、プローブのデザイン段階では我々が使用した SureSelectV4 で 93.9%、SureSelect V5 では 96.5%の CDS 領域をカバーしていた。また、実際に NGS の call に必要な depth 10 以上の read が得られていた CDS 領域は今回の我々の実験データでは SureSelect V4（38 検体の平均）で 90.6%、SureSelect V5（104 検体の平均）で 93.9%であった。他のプラットフォームでもほぼ同様の結果であった（表 10）。以上のことから、本研究で使用した SureSelect を含め、ほとんどのプラットフォームでは CDS 領域の 7~10%程度がシーケンスで読み逃されていることが明らかとなった。

6. CCCS の CDS 領域カスタムデザイン

WESでのCDS領域の読み逃しされている領域に未発見の疾患原因変異がある可能性を考え、WESで読まれていない領域にカスタムのプローブを設計し追加シーケンスすることにした（これを「CDS補完用カスタムリシーケンス」、Complementary Custom CDS Sequencingの略で「CCCS」と呼ぶこととした）。キャプチャープローブは選択的環状化法を用いるHaloPlex（Agilent社）で設計することにした。これはWESで用いたSureSelectのキャプチャーハイブリダイゼーション法と手法や原理が異なるため、WESでデザインできなかった領域もカバーできる可能性があると考えたためである。デザインはコントロールのLCL細胞（NA18943）検体のSureSelect V5プローブを用いたWESデータにおいて、depthが15未

満の CDS 領域 (2,171,214 bp) をターゲットして設計を試みた。デザイン用閾値の基準 depth を NGS call の閾値である 10 ではなく、それより大きい 15 にしたのはデータのばらつきを考慮してのことである。結果的に基準を満たすキャプチャプローブとして 120,383 プローブがデザインできた (デザイン方法の詳細は本論文の「方法」部分に記載した)。これは CDS 領域の 3.3 Mb (全 CDS の 9.8%) をカバーしており、ターゲットにした 2,171,214 bp のうち 1,472,963 bp (67.8%) がデザインしたプローブでカバーできた (デザインされた HaloPlex プローブのさらなる個別詳細情報は発表済み論文 Miya *et al.*⁹⁹ の Supplementary Table S3 に公表した)。

7. CCCS のデータ量と call の評価

WES で疾患原因候補変異が残らなかった 9 家系のうち、CCCS のデザインに使用した SureSelect V5 と同じプラットフォームで実施した 7 家系 (27 検体) およびデザインに使用した LCL 細胞 (NA18943 検体) の合計 28 検体について CCCS を実施した。28 検体の平均で 1.74 Gb のデータ量が得られ、そのうち 0.97 Gb がターゲット領域に paired-end でマッピングできた。さらに、リファレンスゲノムにマッピングができなかった read、paired-end 間の距離が 1,000 bp 以内に存在せずに除去された read、ターゲット領域にマッピングされずに除去された read について BLAT にて再マッピングを実施し (図 7 と「方法」参照)、追加で 106 Mb のリード (結果的に約 11% 増量) をリカバーすることができた。合計で 1 検体当たりの平均で 1.07 Gb が call 対象になった。ターゲット CDS 領域の平均 depth は 194.6 であった (図 20 a)。ターゲット CDS 領域の depth 10 以上の割合 (NGS call 対象領域) は 85.6% であった。その他の集計値を合わせて結果を表 11 にまとめた。

CCCS を実施した検体と同一検体 2 検体を Illumina 社の HumanOmniExpress Exome SNP chip でジェノタイピングし、CCCS のデータと比較し評価を行った。NGS call と SNP chip データの双方でジェノタイプデータが存在する箇所が 2 検体の合計で 26,168 箇所存在し、その一致率は平均で 99.85% であった (表 12)。CCCS データと SNP chip データで call が不一致だった全 40 箇所について、我々の NGS call データと SNP chip データのいずれが合っているのかを確認するため、Sanger シーケンスによって確認実験を行った (表 13)。その結果、40 箇所中 24 箇所は我々の NGS call が正しかった (表 13 および表 14)。この結果から、他の SNP chip と一致していた箇所が正しい call だと仮定した場合、CCCS の call の疑陽性率は 0.038% (10 / 26,168)、偽陰性率は 0.27% (6 / 2,223) であった (表 14)。なお、Sanger シーケンスがうまくできなかった箇所は我々の call が間違いであるという保守的な算出方法の結果である。また、BLAT によってリカバーされた read によって NGS call が追加された箇

所についてのみを別途調査したところ、SNP chip と NGS の双方で call が存在する場所が 260 箇所あり、その全てについて双方の call が一致していた（一致率 100%）。

8. WES と CCCS の統合による CDS カバー率

上述の通り、プローブデザイン段階での CDS 領域のカバー率は本研究で用いた WES の SureSelect V5 が 96.5%、CCCS が 9.8%であった。この両者を統合することで、重複なしの CDS カバー率はデザイン段階で 98.5%となった。NGS で call 可能な閾値の depth が 10 以上ある CDS 領域の割合は今回 CCCS を実施した 28 検体の平均で WES が 93.7% (31.7/33.8 Mb) で、WES と CCCS を組み合わせることでその平均値が 97.1% (32.8/33.8 Mb) まで上昇した（図 20 b および表 15）。すなわち、WES に CCCS を追加統合することで、シーケンスが得られた CDS 領域が 1 検体当たりの平均で約 110 万 base 増加した。CDS 領域で depth が 10 以上ある塩基の割合は 28 検体同士で非常に似ており、その SD は WES で 0.37%、WES と CCCS の統合時で 0.15%であった。さらにその検体間の類似性を確認するため、(a) 性染色体を除く CDS の全ターゲット領域 (31,295,571 bases) の WES データ、(b) そのうち CCCS でターゲットとした LCL 細胞 (NA18943) で depth が 15 未満だった CDS 領域 (1,368,357 bases) の WES データ、(c) (b) と同領域の CCCS データ、のそれぞれの塩基について確認したところ、その平均 depth はそれぞれ (a) 75.1、(b) 4.1、(c) 268.2 で、その Spearman の相関係数を全 28 検体の組み合わせ (378 種の組み合わせ) について確認したところ、それぞれ (a) 0.945 ± 0.003 、(b) 0.920 ± 0.005 、(c) 0.928 ± 0.032 (平均値 \pm SD) という結果で、全検体とも各塩基の depth が非常に類似していることが確認できた（図 21）。また、WES と CCCS の統合解析による各染色体の CDS 領域の call が得られた割合を調べたところ、半分以上の染色体で 98%以上の CDS 領域で call が得られていた一方で、Y 染色体だけ際立って call が得られた割合が 40.9%と低かった（図 20 c）。

9. WES と CCCS の統合解析により明らかになった疾患原因変異

上述の通り WES で何も疾患原因変異候補が残らなかった 7 家系 (27 検体) について CCCS を実施し、そのうち少なくとも 1 家系について、ほぼ確定的に疾患原因変異と見られる変異を新たに同定することができた。

その家系は小頭症家系で、家系図を図 22 a に示した。患者は家系図に示すように 4 人の子供のうち 6 歳の次女（第二子、II-2）と 3 歳の長男（第三子、II-3）で、近親婚ではない健常な日本人の両親の子として生まれた。一人目の患者（II-2）の女兒は妊娠 40 週目に自然分娩によって出生した。出生体重は 2,228 g (-0.4 SD) で頭囲は 29 cm (-3.2 SD) であった。小頭症であること以外、特にその他の奇形は認められなかった（図 22 b）。若干の言語

獲得の遅れが見られたが、6歳時点での知能と運動機能は正常値であった。6歳時での頭囲は42 cm (-6.6 SD)であった。二人目の患者 (II-3) の男児は妊娠40週目に自然分娩によって出生した。出生体重は2,626 gで頭囲は29 cm (-3.2 SD)であった。3歳時点での頭囲は42 cm (-4.9 SD)で、小頭症であること以外に他の奇形異常は見当たらず、他の精神や運動機能の成長にも特に遅れは見られなかった。

両親 (図22 aの家系図のI-1 およびI-2) と二人の小頭症患者 (II-2 およびII-3) についてWESおよびCCCSを実施し、その二種を統合解析することで、最終的に2個のheterozygousな変異に絞り込むことができた (図22 c)。その変異はASPM (abnormal spindle-like microcephaly associated、GeneID: 259266、RefSeqID: NM_018136) 遺伝子の exon 18 のc.8098C>T (p.R2700*) と、同遺伝子の exon 27 のc.10168C>T (p.R3390*) のstop codonが入る2つのnonsense変異であった。ゲノム遺伝子上の位置とタンパク質上の位置を図22 dと図22 eに示した。後者のc.10168C>T (p.R3390*) 変異の箇所はWESではdepthが浅く、他の全てのWES検体でもcallが得られていない領域であり、CCCSを実施して初めて発見することができた変異であった。WESとCCCSを実施した両親と患者と併せ、健常な他の2人の子を含めてASPMの変異箇所をSangerシーケンスにて確認した結果、両親が一つずつの変異保因者で、患者の子がcompound heterozygoteの状態に変異を2つ保因していた (図22 f)。また、健常な他の二人の子はどちらの変異も有していなかった (図22 f)。

ASPMは、神経系カスタムリシーケンスデータで発見された小頭症患者の部分でも述べた通り、小頭症の原因遺伝子の中で最も原因遺伝子として頻度の高い遺伝子であり⁴¹、正常な神経発生や神経移動¹⁰⁰、大脳サイズの規定³⁹、有糸分裂紡錘体の機能^{39,101,102}に重要な役割を担っていることが知られている。また、図22 eにこれまで論文で報告されてきたASPMの変異箇所を図示したが、タンパク質上の変異位置とその小頭症の表現型の重症度には特に関連がないことが知られており^{103,104}、例えば今回見つかった変異周辺のC末端側のp.Y3353*¹⁰⁴や、p.R3354*¹⁰⁵なども小頭症原因変異として報告されている。以上のことから、本小頭症患者は上記のcompound heterozygousなASPM変異による小頭症であると確定診断された。本結果は論文雑誌でも報告を行った⁹⁹。

考察

本研究での NGS を用いた神経系遺伝子カスタムリシーケンス及び WES によって多数の検体の疾患原因遺伝子を同定することができた。古典的な染色体試験や CGH アレイ等のゲノム構造異常試験では異常が明らかにできなかった検体でも高感度に原因変異の検出が可能であった。また、代表的な原因遺伝子の Sanger シーケンスでは時間がかかるだけでなく、見逃しも発生してしまう場合もあり、迅速性やコスト面からの効率性という面でも NGS を用いた原因変異探索は極めて有用であると言える。また、特に神経系疾患の場合は表現型が多岐に渡り、様々な疾患と重複する表現型（例えばてんかんや精神・運動発達遅滞など）があり、臨床症状だけでは確定診断が非常に難しい場合が多くある。さらに、候補となる疾患名が存在しても、症状が典型例とは異なる場合も多い。例えば *ACTB* 変異が見つかった BRWS 症例では既報の BRWS よりも形成異常の度合いが軽度で、頭囲が平均より大きいなどの特徴があった。また、神経系遺伝子カスタムリシーケンスで見つかった小頭症の *ASPM* 変異を有する患者では厚脳回も見られるという稀な症状があったり、Noonan 症候群の原因遺伝子である *PTPN11* に変異が見つかった患者ではこれまで *PTPN11* 変異患者では報告がない脳梁欠損が見られたり等、これらは NGS で網羅的に変異を持つ遺伝子を探索しなければ確定診断や原因遺伝子の同定が難しく、非典型例の確定診断のためにも NGS は非常に有用であると言える。

今回の 284 遺伝子を対象とした神経系遺伝子カスタムリシーケンスでは、119 検体中 14 検体で確定的な原因変異を同定することができ、その全体に占める同定成功率は 11.8%であった。本論文では記載しなかったが、別の共同研究で実施した難聴を対象とした既知の難聴関連遺伝子の約 200 遺伝子のカスタムリシーケンスの場合は同定成功率が約 50%程度であったことと比較すると、神経系遺伝子の場合の同定成功率は低めであった。NGS 解析による疾患原因変異の同定成功率は、その対象とする疾患や、対象の家系の選択方法等によって大きく異なることが予想されるが、先天性神経系疾患の場合は未だに原因遺伝子が未確定な遺伝子に原因が存在する可能性や、同一疾患でも原因遺伝子が非常に多岐に渡るものが多いため、カスタムリシーケンスのプロンプト作製遺伝子リスト内で網羅され切れておらず同定成功率が低めだったと考えられる。今後、さらに多数の先天性神経疾患の原因遺伝子が同定された後に、改めてそれらの新規遺伝子も網羅したカスタムリシーケンスを設計することで同定成功率を上げることは可能であろう。少なくとも現時点で先天性神経疾患に対しては 11.8%程度（難聴なら約 50%）であれば少数遺伝子のみをターゲットとしたリシーケンスでも原因変異を同定でき、少数の遺伝子をターゲットとしたリシーケンスは

WES に比べてコストが安く済み（現時点で約 1 検体 3 万円程度）、HiSeq2000 のような大型 NGS ではなく MiSeq などの小型 NGS でも十分探索が可能でシーケンスにかかる時間も短くできる（最短 2 日以内）ことから、現時点あるいは今後の改良版カスタムリシーケンスを臨床診断の現場で活用することも現実的であると考えられる。実際に一部機関では臨床での応用が始まっており、今後さらに改良を進めながら活用が広まっていくと考えられる。

神経系遺伝子カスタムリシーケンスの同定成功率が低めであったため、疾患原因変異の同定を目的とした基礎研究を主体とした本研究では、基本的にその後の検体を全て WES で実施する戦略を取った。WES での同定成功率は 46.5%であり、これは世界的に報告されている平均約 25%^{21,22}と比較すると高めであった。これは既知の一般集団頻度情報データベースの整備が進み、それを多数解析に取り入れたことと、閾値等のパラメーターを精査し、偽陽性率と偽陰性率を極力減らすように調整した解析プログラムによるところが大きいと考えられる。また、何らかの候補が残った家系が 32.6%あり、それらの変異を詳細に機能解析することで更なる新規疾患原因遺伝子の発見に繋がる可能性が大いにある。現在、WES 解析によって全ゲノムの遺伝子コード領域の塩基配列を網羅的に短時間でスクリーニングすることが可能となったが、疾患原因遺伝子として知られていない遺伝子に候補が見つかった場合に、どのように迅速に機能解析を行い、どのようにその変異の存在する遺伝子が疾患原因遺伝子なのかを調べていく手法面は今後の大きな課題である。例えば、近年開発された簡便で迅速な遺伝子配列改変手法である CRISPR/Cas-9 システム¹⁰⁶の手法などを用いて迅速に候補変異の機能解析用のスクリーニングも行えるようになっていくことが期待される。

WES でも候補すら何も残らない検体が本研究の検体でも約 20.9%存在しており、この原因変異未発見家系検体の原因変異が WES で読まれていない CDS 領域に存在する可能性を考え、CDS 未読領域を補完するカスタムリシーケンスプローブを作製し CCCS を実施した。その結果、WES では何も候補が残らなかった 7 家系中、少なくとも 1 家系で新規の疾患原因変異を同定することができた。また、CDS 領域のシーケンスカバー率は既存のどのプラットフォームよりも高い 97.4%に達することができ、これは既存の WES 単独に比べ約 4%から 7%多かった。各既存の WES プラットフォームは CDS のカバー領域が若干ずつ異なるので、既存の WES を 2 種組み合わせることで CDS 領域のカバー率を上げる方法も考えられるが、その場合は当然コストが単独の WES の 2 倍かかってしまうことになる。一方、本研究で用いた CCCS は未読 CDS 領域のみ（約 2 Mb 以下）をターゲットとするのでプローブ数が少なく済み、必要シーケンス量も少なく済むのでコストも抑えられる（現時点で神経系カスタムリシーケンスと同程度の約 1 検体 3 万円程度）。今回は SureSelect V5 という WES 用プローブを用いた WES データを元に CCCS のプローブを設計したが、別の WES プラットフ

フォームを使用している場合でも同様のストラテジーで CDS 領域を補完するプローブを設計することで CDS 領域カバー率を上げ、新規疾患原因変異を探索が可能と考えられ、WES で候補が何も残らなかった場合に、CCCS を追加で実施することは広く適用可能な選択肢の一つとして非常に有用であろうと考えられる。

CDS のカバー率は WES と CCCS の組み合わせで 97.4%まで達することができ、ほとんどの染色体の CDS 領域で 98%を超えていたが、Y 染色体の CDS 領域だけは 40.9%とそのカバー率が極端に低かった (図 20 c)。この理由としては、Y 染色体には pseudoautosomal regions (PAR) と呼ばれる X 染色体に非常に類似した配列があることと、male-specific regions of Y chromosome (MSY)¹⁰⁷ と呼ばれる多くの繰り返し塩基配列があることにより、重複のないユニークなプローブを設計することが難しく、マッピングも難しいことが原因と考えられる。もしプローブのデザインやマッピングの厳密性を緩めたとしたら、Y 染色体の CDS カバー率や他の染色体のカバー率も多少はさらに増加すると予想されるが、その閾値の取り方は NGS call の正確性とカバー率とのトレードオフの関係になると考えられる。また、今回実施した検体間ではゲノム上の領域ごとの NGS の depth が非常に類似していた (図 21 等) ことから、同じプラットフォームでデータを収集する限りは、1 つの検体データだけを元にしたプローブのデザインで、全ての検体で流用してほぼ同様にカバー率を上昇させることができると言える。また、その depth の類似性を調査した際の相関係数は Pearson の相関係数の方が Spearman の相関係数よりもやや低く、特に WES の depth が低い領域のデータで差が顕著であった (図 21)。その理由は、depth が 0 の領域が多いと depth の全体の分布が正規分布から大きく外れていくことと、depth が浅い領域ほどその depth のブレ (偏差) が大きくなるために Pearson の相関係数の値が小さくなりやすい一方で、Spearman の相関係数は順位相関によるものであるため、そのようなデータに堅牢性が高いためと考えられた。

最後に、WES でも CCCS の追加解析でも疾患原因変異 (候補も含めて) が見つからない検体が存在することの原因と解決法について考察したい。原因変異未同定の原因の可能性としては、一般的にターゲットとして調査されないアミノ酸変化を伴わない CDS 領域の変化 (synonymous 変異) が原因である可能性¹⁰⁸ や、WES および CCCS ではターゲットとしていない UTR 領域やイントロンや非遺伝子領域 (non-coding RNA や promoter 領域等) に原因変異がある可能性、疾患の浸透率が低い可能性、疾患が遺伝性 (Mendelian disease) ではない可能性、マッピングや exon キャプチャーが難しい領域 (リピート配列領域や極度に GC 含量が高い領域等) に原因変異がある可能性、NGS では検出が難しい長距離の indel が原因である可能性、単一遺伝子ではなく複数の遺伝子の変異が原因で疾患が発症している可能性などが考えられる。現状では、synonymous 変異や浸透率が低い疾患の原因遺伝子を同定することは非常にチャレンジングなことではあるが、GWAS のように exome でも検体が多

数ある場合には、case 群と control 群に分けて遺伝子単位で有意に疾患群に変異が多い遺伝子を探査する手法 (SKAT¹⁰⁹⁻¹¹¹ 等) が探査に有効かもしれない。また、ミトコンドリア DNA は DNA のコピー数が通常の染色体 DNA より多く、通常の染色体と一緒にプローブをデザインすると NGS データのバランスが崩れ、通常の染色体からのデータ比が減少し効率が悪化するため WES ではターゲットにされないが、ミトコンドリア DNA も神経疾患に関与するという報告もあり^{112,113}、WES で通常の染色体からのデータ比をほとんど減少させずにミトコンドリアをキャプチャーする特殊な方法も考案されている¹¹⁴ ので、今後ミトコンドリア DNA もターゲットに加えることで原因変異同定率が多少向上するかもしれない。NGS で検出が難しい大きなサイズの indel が疾患原因である場合もこれまで数多く報告されており、それらの多くは CGH アレイで発見されている。同一検体に対し WES と CGH アレイの双方を実施することが理想的ではあるが、近年、CGH ほどの感度までは達していないが WES からも大きなサイズのコピー数多型 (copy number variation、CNV) を検出する解析手法がいくつか開発されており^{115,116}、今後さらに精度が増していけば WES 単独でも疾患原因となっている CNV (indel) を同時に検出していく手法がスタンダードになっていくことが期待される。また、未だにプローブの設計やマッピングが困難でカバーできていない残りの CDS 領域は、次世代シーケンサーが今後も改良され、第 3 世代、あるいはさらにその先の世代のシーケンサー¹¹⁷ で read 長が伸びていくことで、解読が可能となっていくことが期待される。実際に Pacific Biosciences 社の NGS である PacBio RS II では 2015 年現在で試薬の改良により、以前より低エラー率で平均 read 長が 10 kb 以上でシーケンス可能となってきたため、今後の更なる精度や読み長の改良と低コスト化が期待される。

結論として、本論文研究では、神経系遺伝子カスタムターゲットリシーケンス、WES、そして WES で読まれていない CDS 領域を補完する形のカスタムリシーケンス (CCCS) という 3 種の実験手法とバイオインフォマティクスの解析手法を統合して、多数の神経系疾患原因遺伝子変異の同定に成功した。新規の表現型を示す原因変異や新規原因遺伝子も複数発見し、一部は機能解析により疾患発症メカニズムとの関連性も示唆できた。それら新規遺伝子群およびネットワーク上の遺伝子群は新たな薬剤ターゲットとなりうるものが期待できる。また、WES だけでは原因変異が同定できない場合の新たな探索手法を開発し、実際に WES で原因変異が未同定であった検体の変異同定にも成功した。この手法は今後の検体のみならず、過去に実施した検体も含めての疾患原因変異同定スクリーニングに活用できると考えられる。本研究を足掛かりの一つとしつつ、更なる各種の NGS ハードウェアの改良およびバイオインフォマティクスを駆使したソフトウェアの改良により、今後より一層多くの新規遺伝子を含めた疾患原因変異が同定され、更なる疾患メカニズムの解明や治療へと繋がっていくことを期待したい。

引用文献

- 1 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
- 2 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351 (2001).
- 3 Consortium., I. H. G. S. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
- 4 Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470 (1995).
- 5 International HapMap, C. The International HapMap Project. *Nature* **426**, 789-796 (2003).
- 6 Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650-654 (2002).
- 7 Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7-24 (2012).
- 8 Gusella, J. F. *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234-238 (1983).
- 9 Murray, J. M. *et al.* Linkage relationship of a cloned DNA sequence on the short arm of the X chromosome to Duchenne muscular dystrophy. *Nature* **300**, 69-71 (1982).
- 10 Tsui, L. C. *et al.* Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* **230**, 1054-1057 (1985).
- 11 Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11**, 31-46, (2010).
- 12 Kawai, Y. *et al.* Japonica array: improved genotype imputation by designing a population-specific SNP array with 1070 Japanese individuals. *J. Hum. Genet.* **In press** doi: 10.1038/jhg.2015.68 (2015).
- 13 Biesecker, L. G. Exome sequencing makes medical genomics a reality. *Nat. Genet.* **42**, 13-14 (2010).
- 14 Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745-755 (2011).
- 15 Mamanova, L. *et al.* Target-enrichment strategies for next-generation sequencing. *Nat. Methods* **7**, 111-118 (2010).
- 16 Zhang, X. Exome sequencing greatly expedites the progressive research of Mendelian

- diseases. *Front. Med.* **8**, 42-57 (2014).
- 17 Hedges, D. J. *et al.* Comparison of three targeted enrichment strategies on the SOLiD sequencing platform. *PLoS ONE* **6**, e18595 (2011).
- 18 Mertes, F. *et al.* Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief. Funct. Genomics* **10**, 374-386 (2011).
- 19 Clark, M. J. *et al.* Performance comparison of exome DNA sequencing technologies. *Nat. Biotechnol.* **29**, 908-914 (2011).
- 20 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
- 21 Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **369**, 1502-1511 (2013).
- 22 Atwal, P. S. *et al.* Clinical whole-exome sequencing: are we there yet? *Genet. Med.* (2014).
- 23 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
- 24 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
- 25 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303 (2010).
- 26 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
- 27 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073-1081 (2009).
- 28 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248-249 (2010).
- 29 Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* **32**, 894-899 (2011).
- 30 Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* **34**, E2393-2402 (2013).
- 31 Kumar, R. A. *et al.* TUBA1A mutations cause wide spectrum lissencephaly (smooth brain) and suggest that multiple neuronal migration pathways converge on alpha tubulins. *Hum. Mol. Genet.* **19**, 2817-2827 (2010).
- 32 Poirier, K. *et al.* Mutations in the neuronal ss-tubulin subunit TUBB3 result in malformation

- of cortical development and neuronal migration defects. *Hum. Mol. Genet.* **19**, 4462-4473 (2010).
- 33 Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656-664 (2002).
- 34 Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
- 35 Baraitser, M. & Winter, R. M. Iris coloboma, ptosis, hypertelorism, and mental retardation: a new syndrome. *J. Med. Genet.* **25**, 41-43 (1988).
- 36 Verloes, A. *et al.* Baraitser-Winter cerebrofrontofacial syndrome: delineation of the spectrum in 42 cases. *Eur. J. Hum. Genet.* **23**, 292-301 (2015).
- 37 Riviere, J. B. *et al.* De novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. *Nat. Genet.* **44**, 440-444 (2012).
- 38 Okamoto, N. *et al.* Targeted next-generation sequencing in the diagnosis of neurodevelopmental disorders. *Clin. Genet.* **88**, 288-292 (2014).
- 39 Bond, J. *et al.* ASPM is a major determinant of cerebral cortical size. *Nat. Genet.* **32**, 316-320 (2002).
- 40 Passemard, S. *et al.* Expanding the clinical and neuroradiologic phenotype of primary microcephaly due to ASPM mutations. *Neurology* **73**, 962-969 (2009).
- 41 Nicholas, A. K. *et al.* The molecular landscape of ASPM mutations in primary microcephaly. *J. Med. Genet.* **46**, 249-253 (2009).
- 42 Nakamura, K. *et al.* Primary microcephaly with anterior predominant pachygyria caused by novel compound heterozygous mutations in ASPM. *Pediatr. Neurol.* **52**, e7-8 (2015).
- 43 Tartaglia, M. *et al.* Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat. Genet.* **29**, 465-468 (2001).
- 44 Yamamoto, T. *et al.* Clinical manifestations of the deletion of Down syndrome critical region including DYRK1A and KCNJ6. *Am. J. Med. Genet. A* **155A**, 113-119 (2011).
- 45 van Bon, B. W. *et al.* Intragenic deletion in DYRK1A leads to mental retardation and primary microcephaly. *Clinical genetics* **79**, 296-299 (2011).
- 46 Valetto, A. *et al.* Molecular cytogenetic characterization of an interstitial deletion of chromosome 21 (21q22.13q22.3) in a patient with dysmorphic features, intellectual disability and severe generalized epilepsy. *Eur. J. Med. Genet.* **55**, 362-366 (2012).
- 47 Oegema, R. *et al.* Distinctive Phenotypic Abnormalities Associated with Submicroscopic 21q22 Deletion Including DYRK1A. *Mol. Syndromol.* **1**, 113-120 (2010).
- 48 Fujita, H. *et al.* Microdeletion of the Down syndrome critical region at 21q22. *Am. J. Med.*

- Genet. A* **152A**, 950-953 (2010).
- 49 Courcet, J. B. *et al.* The DYRK1A gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. *J. Med. Genet.* **49**, 731-736 (2012).
- 50 Riviere, J. B. *et al.* De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat. Genet.* **44**, 934-940 (2012).
- 51 Franceschini, P. *et al.* Macrocephaly-Cutis marmorata telangiectatica congenita without cutis marmorata? *Am. J. Med. Genet.* **90**, 265-269 (2000).
- 52 Robertson, S. P., Gattas, M., Rogers, M. & Ades, L. C. Macrocephaly--cutis marmorata telangiectatica congenita: report of five patients and a review of the literature. *Clin. Dysmorphol.* **9**, 1-9 (2000).
- 53 Wright, D. R. *et al.* The misnomer "macrocephaly-cutis marmorata telangiectatica congenita syndrome": report of 12 new cases and support for revising the name to macrocephaly-capillary malformations. *Arch. Dermatol.* **145**, 287-293 (2009).
- 54 Martinez-Glez, V. *et al.* Macrocephaly-capillary malformation: Analysis of 13 patients and review of the diagnostic criteria. *Am. J. Med. Genet. A* **152A**, 3101-3106 (2010).
- 55 Harada, A. *et al.* Sudden death in a case of megalencephaly capillary malformation associated with a de novo mutation in AKT3. *Childs Nerv. Syst.* **31**, 465-471 (2015).
- 56 Hayashi, S. *et al.* The CASK gene harbored in a deletion detected by array-CGH as a potential candidate for a gene causative of X-linked dominant mental retardation. *Am. J. Med. Genet. A* **146A**, 2145-2151 (2008).
- 57 Najm, J. *et al.* Mutations of CASK cause an X-linked brain malformation phenotype with microcephaly and hypoplasia of the brainstem and cerebellum. *Nat Genet* **40**, 1065-1067 (2008).
- 58 Auvin, S., Lamblin, M. D., Pandit, F., Vallee, L. & Bouvet-Mourcia, A. Infantile epileptic encephalopathy with late-onset spasms: report of 19 patients. *Epilepsia* **51**, 1290-1296 (2010).
- 59 Ishikawa, N., Kobayashi, Y., Fujii, Y., Tajima, G. & Kobayashi, M. Ictal electroencephalography and electromyography features in symptomatic infantile epileptic encephalopathy with late-onset spasms. *Neuropediatrics* **45**, 36-41 (2014).
- 60 Burglen, L. *et al.* Spectrum of pontocerebellar hypoplasia in 13 girls and boys with CASK mutations: confirmation of a recognizable phenotype and first description of a male mosaic patient. *Orphanet J. Rare Dis.* **7**, 18 (2012).

- 61 Saitsu, H. *et al.* CASK aberrations in male patients with Ohtahara syndrome and cerebellar hypoplasia. *Epilepsia* **53**, 1441-1449 (2012).
- 62 Takanashi, J. *et al.* Clinical and radiological features of Japanese patients with a severe phenotype due to CASK mutations. *Am. J. Med. Genet. A* **158A**, 3112-3118 (2012).
- 63 Michaud, J. L. *et al.* The genetic landscape of infantile spasms. *Hum. Mol. Genet.* **23**, 4846-4858 (2014).
- 64 Nakamura, K. *et al.* Primary Microcephaly With Anterior Predominant Pachygyria Caused by Novel Compound Heterozygous Mutations in ASPM. *Pediatr. Neurol.* **52**, e7-8 (2015).
- 65 Windpassinger, C., Kroisel, P. M., Wagner, K. & Petek, E. The human gamma-aminobutyric acid A receptor delta (GABRD) gene: molecular characterisation and tissue-specific expression. *Gene* **292**, 25-31 (2002).
- 66 Rosenfeld, J. A. *et al.* Refinement of causative genes in monosomy 1p36 through clinical and molecular cytogenetic characterization of small interstitial deletions. *Am. J. Med. Genet. A* **152A**, 1951-1959 (2010).
- 67 DeLorey, T. M. & Olsen, R. W. GABA and epileptogenesis: comparing gabrb3 gene-deficient mice with Angelman syndrome in man. *Epilepsy Res.* **36**, 123-132 (1999).
- 68 Okada, Y., Yamazaki, H., Sekine-Aizawa, Y. & Hirokawa, N. The neuron-specific kinesin superfamily protein KIF1A is a unique monomeric motor for anterograde axonal transport of synaptic vesicle precursors. *Cell* **81**, 769-780 (1995).
- 69 Yonekawa, Y. *et al.* Defect in synaptic vesicle precursor transport and neuronal cell death in KIF1A motor protein-deficient mice. *J. Cell Biol.* **141**, 431-441 (1998).
- 70 Hamdan, F. F. *et al.* Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **88**, 306-316 (2011).
- 71 Okamoto, N. *et al.* KIF1A mutation in a patient with progressive neurodegeneration. *J. Hum. Genet.* **59**, 639-641 (2014).
- 72 Lee, J. R. *et al.* De novo mutations in the motor domain of KIF1A cause cognitive impairment, spastic paraparesis, axonal neuropathy, and cerebellar atrophy. *Hum. Mutat.* **36**, 69-78 (2015).
- 73 Campbell, C. G., Wang, H. & Hunter, G. W. Interstitial microdeletion of chromosome 1p in two siblings. *Am. J. Med. Genet.* **111**, 289-294 (2002).
- 74 Shanske, A. L., Edelmann, L., Kardon, N. B., Gosset, P. & Levy, B. Detection of an interstitial deletion of 2q21-22 by high resolution comparative genomic hybridization in a

- child with multiple congenital anomalies and an apparent balanced translocation. *Am. J. Med. Genet. A* **131**, 29-35 (2004).
- 75 Lu, W. *et al.* NFIA haploinsufficiency is associated with a CNS malformation syndrome and urinary tract defects. *PLoS Genet.* **3**, e80 (2007).
- 76 Koehler, U. *et al.* A novel 1p31.3p32.2 deletion involving the NFIA gene detected by array CGH in a patient with macrocephaly and hypoplasia of the corpus callosum. *Eur. J. Pediatr.* **169**, 463-468 (2010).
- 77 Chen, C. P. *et al.* Chromosome 1p32-p31 deletion syndrome: prenatal diagnosis by array comparative genomic hybridization using uncultured amniocytes and association with NFIA haploinsufficiency, ventriculomegaly, corpus callosum hypogenesis, abnormal external genitalia, and intrauterine growth restriction. *Taiwan J. Obstet. Gynecol.* **50**, 345-352 (2011).
- 78 Ji, J., Salamon, N. & Quintero-Rivera, F. Microdeletion of 1p32-p31 involving NFIA in a patient with hypoplastic corpus callosum, ventriculomegaly, seizures and urinary tract defects. *Eur. J. Med. Genet.* **57**, 267-268 (2014).
- 79 Rao, A. *et al.* An intragenic deletion of the NFIA gene in a patient with a hypoplastic corpus callosum, craniofacial abnormalities and urinary tract defects. *Eur. J. Med. Genet.* **57**, 65-70 (2014).
- 80 Negishi, Y. *et al.* Truncating mutation in NFIA causes brain malformation and urinary tract defects. *Hum. Genome Var.* **2** (2015).
- 81 Jackson, A. P. Diversifying microtubules in brain development. *Nat. Genet.* **41**, 638-640 (2009).
- 82 Sui, H. & Downing, K. H. Structural basis of interprotofilament interaction and lateral deformation of microtubules. *Structure* **18**, 1022-1031 (2010).
- 83 Tischfield, M. A., Cederquist, G. Y., Gupta, M. L., Jr. & Engle, E. C. Phenotypic spectrum of the tubulin-related disorders and functional implications of disease-causing mutations. *Curr. Opin. Genet. Dev.* **21**, 286-294 (2011).
- 84 Keays, D. A. *et al.* Mutations in alpha-tubulin cause abnormal neuronal migration in mice and lissencephaly in humans. *Cell* **128**, 45-57 (2007).
- 85 Jaglin, X. H. *et al.* Mutations in the beta-tubulin gene TUBB2B result in asymmetrical polymicrogyria. *Nat. Genet.* **41**, 746-752 (2009).
- 86 Tischfield, M. A. *et al.* Human TUBB3 mutations perturb microtubule dynamics, kinesin interactions, and axon guidance. *Cell* **140**, 74-87 (2010).

- 87 Breuss, M. *et al.* Mutations in the beta-tubulin gene TUBB5 cause microcephaly with structural brain abnormalities. *Cell Rep.* **2**, 1554-1562 (2012).
- 88 Simons, C. *et al.* A de novo mutation in the beta-tubulin gene TUBB4A results in the leukoencephalopathy hypomyelination with atrophy of the basal ganglia and cerebellum. *Am. J. Hum. Genet.* **92**, 767-773 (2013).
- 89 Cushion, T. D. *et al.* De novo mutations in the beta-tubulin gene TUBB2A cause simplified gyral patterning and infantile-onset epilepsy. *Am. J. Hum. Genet.* **94**, 634-641 (2014).
- 90 Poirier, K. *et al.* Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly. *Nat. Genet.* **45**, 639-647 (2013).
- 91 Jaglin, X. H. & Chelly, J. Tubulin-related cortical dysgeneses: microtubule dysfunction underlying neuronal migration defects. *Trends Genet.* **25**, 555-566 (2009).
- 92 Jansen, A. C. *et al.* TUBA1A mutations: from isolated lissencephaly to familial polymicrogyria. *Neurology* **76**, 988-992 (2011).
- 93 Cushion, T. D. *et al.* Overlapping cortical malformations and mutations in TUBB2B and TUBA1A. *Brain* **136**, 536-548 (2013).
- 94 Poirier, K. *et al.* Expanding the spectrum of TUBA1A-related cortical dysgenesis to Polymicrogyria. *European journal of human genetics : Eur. J. Hum. Genet.* **21**, 381-385 (2013).
- 95 Zanni, G. *et al.* Description of a novel TUBA1A mutation in Arg-390 associated with asymmetrical polymicrogyria and mid-hindbrain dysgenesis. *Eur. J. Paediatr. Neurol.* **17**, 361-365 (2013).
- 96 Fry, A. E., Cushion, T. D. & Pilz, D. T. The genetics of lissencephaly. *Am. J. Med. Genet. C Semin. Med. Genet.* **166c**, 198-210 (2014).
- 97 Bahi-Buisson, N. *et al.* Refinement of cortical dysgeneses spectrum associated with TUBA1A mutations. *J. Med. Genet.* **45**, 647-653 (2008).
- 98 Yokoi, S. *et al.* TUBA1A mutation can cause a hydranencephaly-like severe form of cortical dysgenesis. *Sci. Rep.* **in press** (2015).
- 99 Miya, F. *et al.* A combination of targeted enrichment methodologies for whole-exome sequencing reveals novel pathogenic mutations. *Sci. Rep.* **5**, 9331 (2015).
- 100 Buchman, J. J., Durak, O. & Tsai, L. H. ASPM regulates Wnt signaling pathway activity in the developing brain. *Genes Dev.* **25**, 1909-1914 (2011).
- 101 Fish, J. L., Kosodo, Y., Enard, W., Paabo, S. & Huttner, W. B. Aspm specifically maintains symmetric proliferative divisions of neuroepithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**,

- 10438-10443 (2006).
- 102 van der Voet, M. *et al.* NuMA-related LIN-5, ASPM-1, calmodulin and dynein promote meiotic spindle rotation independently of cortical LIN-5/GPR/Galpha. *Nat. Cell Biol.* **11**, 269-277 (2009).
- 103 Bond, J. *et al.* Protein-truncating mutations in ASPM cause variable reduction in brain size. *Am. J. Hum. Genet.* **73**, 1170-1177 (2003).
- 104 Gul, A. *et al.* Novel protein-truncating mutations in the ASPM gene in families with autosomal recessive primary microcephaly. *J. Neurogenet.* **21**, 153-163 (2007).
- 105 Halsall, S., Nicholas, A. K., Thornton, G., Martin, H. & Geoffrey Woods, C. Critical consequences of finding three pathogenic mutations in an individual with recessive disease. *J. Med. Genet.* **47**, 769-770 (2010).
- 106 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).
- 107 Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825-837 (2003).
- 108 Supek, F., Minana, B., Valcarcel, J., Gabaldon, T. & Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324-1335 (2014).
- 109 Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82-93 (2011).
- 110 Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* **91**, 224-237 (2012).
- 111 Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D. & Lin, X. Sequence kernel association tests for the combined effect of rare and common variants. *Am. J. Hum. Genet.* **92**, 841-853 (2013).
- 112 Fernandez-Vizarra, E. *et al.* Impaired complex III assembly associated with BCS1L gene mutations in isolated mitochondrial encephalopathy. *Hum. Mol. Genet.* **16**, 1241-1252 (2007).
- 113 Samuels, D. C. *et al.* Finding the lost treasures in exome sequencing data. *Trends Genet.* **29**, 593-599 (2013).
- 114 Falk, M. J. *et al.* Mitochondrial disease genetic diagnostics: optimized whole-exome analysis for all MitoCarta nuclear genes and the mitochondrial genome. *Discov. Med.* **14**, 389-399 (2012).

- 115 de Ligt, J. *et al.* Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum. Mutat.* **34**, 1439-1448 (2013).
- 116 Guo, Y. *et al.* Comparative study of exome copy number variation estimation tools using array comparative genomic hybridization as control. *Biomed. Res. Int.* **2013**, 915636 (2013).
- 117 Huddleston, J. *et al.* Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res.* **24**, 688-696 (2014).

图表

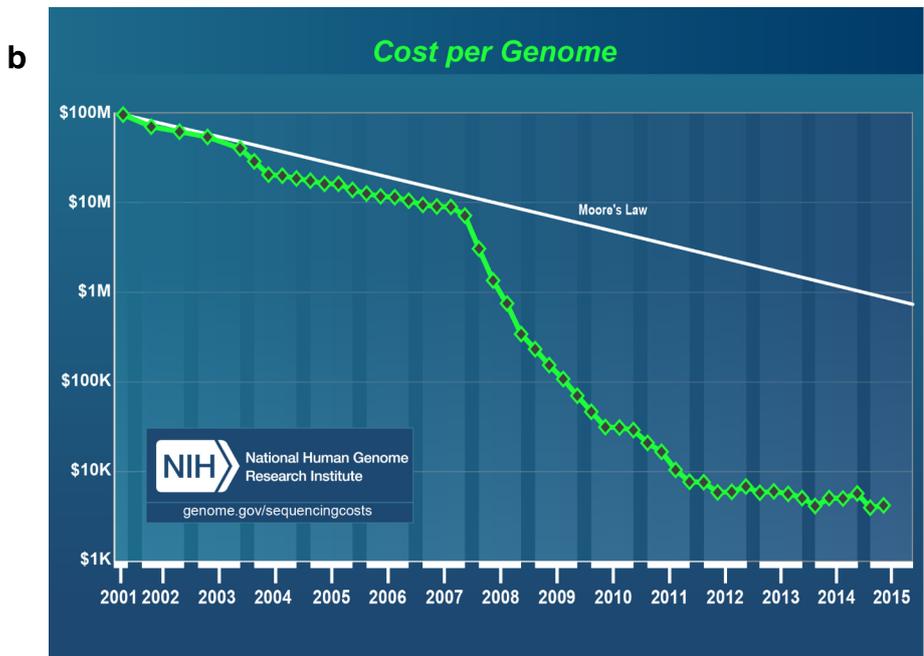
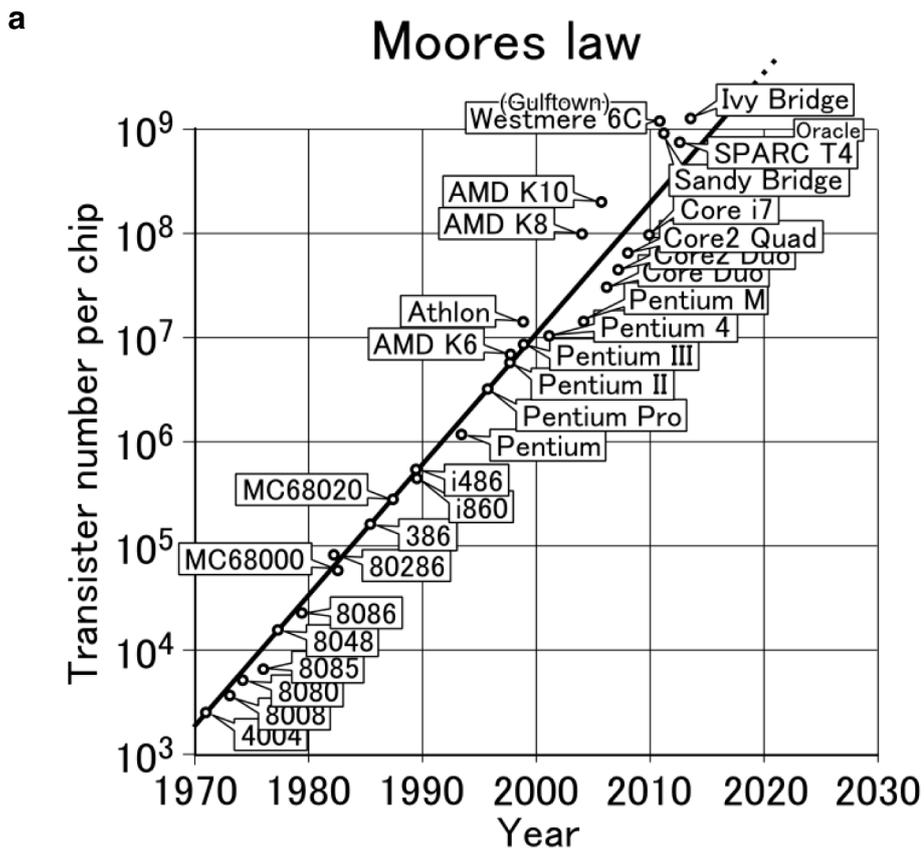


図 1. ムーアの法則と次世代シーケンサーのコストの変遷

(a) 年代ごとのコンピュータのチップ上のトランジスタの数の変遷。その数は1年半ごとに約2倍になっており、式にすると n 年後のトランジスタ数 p は $p = 2^{n/1.5}$ となる。

(b) 次世代シーケンサーの全ゲノムシーケンス当たりにかかるコストの変遷。ムーアの法則の逆数と比較すると遙かにその割合を超えてシーケンスコストが下がっていることがわかる。

(図は National Human Genome Research Institute (NHGRI) のサイト : <http://www.genome.gov/sequencingcosts/> より引用)

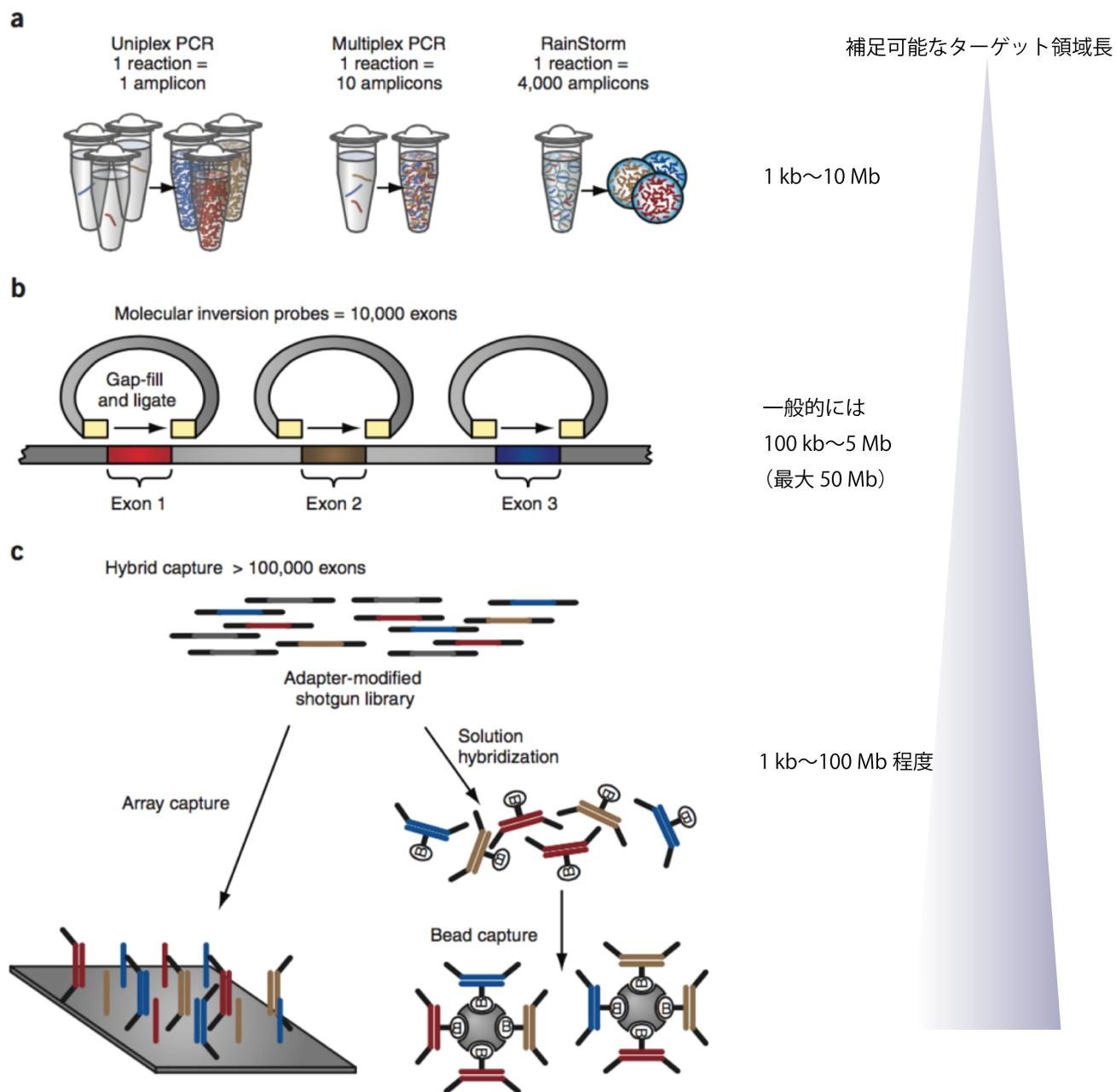


図2. ターゲット領域のゲノム DNA を選択的に補足するための3つの主な手法

(a) PCR 増幅法。1本の tube 内で各々あるいは一度に PCR で増幅を行い、それを NGS にてシーケンスする。

(b) 選択的環状化法。シーケンスしたい領域を挟んで環状構造が作られるようにプローブを設計し、その後、環状化した DNA だけを回収し、増幅、NGS でのシーケンスへと進める。

(c) ハイブリッドキャプチャー法。合成プローブによって、アレイスライドなどの固相上あるいは液体内でターゲット領域をハイブリダイゼーションによってキャプチャーし、回収後に増幅後、NGS にてシーケンスを行う。現在は一般的に液体内でキャプチャーする方が一般的。

図は参考文献 15 および 16 より引用し一部を改変した。

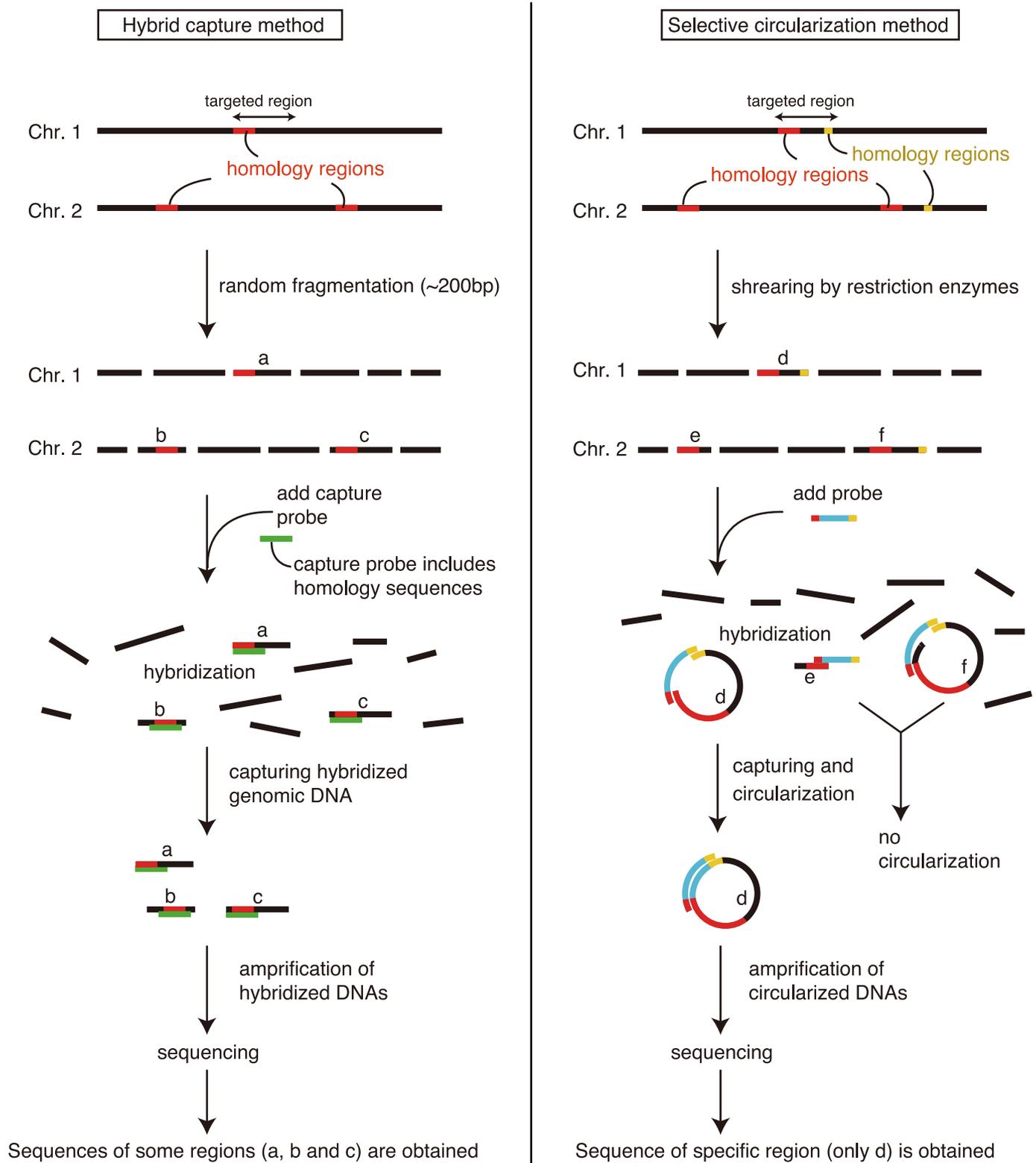


図3. ハイブリッドキャプチャー法に比べ選択的環状化法の方が優れている場合の例

左がハイブリッドキャプチャー法で、右が選択的環状化法の概略図を示し、一部過程は簡略化のため省略している。この図の例では、ターゲット領域に配列が類似している領域 (a に対する b および c、d に対する e および f) が存在する。ハイブリッドキャプチャー法では配列のホモロジーの高い b や c の配列まで同時にキャプチャーされてしまう場合がある。一方、選択的環状化法では、プローブと制限酵素の末端が一致しなければならず、さらに2箇所環状化構造を取れる条件を満たさなければ環状化されないという選択がかかるため、この図では d の特異的な領域しか補足されないというように、ハイブリッドキャプチャー法に比べ高い選択性が保たれる場合がある。



図 4. 全国先天性疾患解析コンソーシアムのメンバーと役割、およびその目的

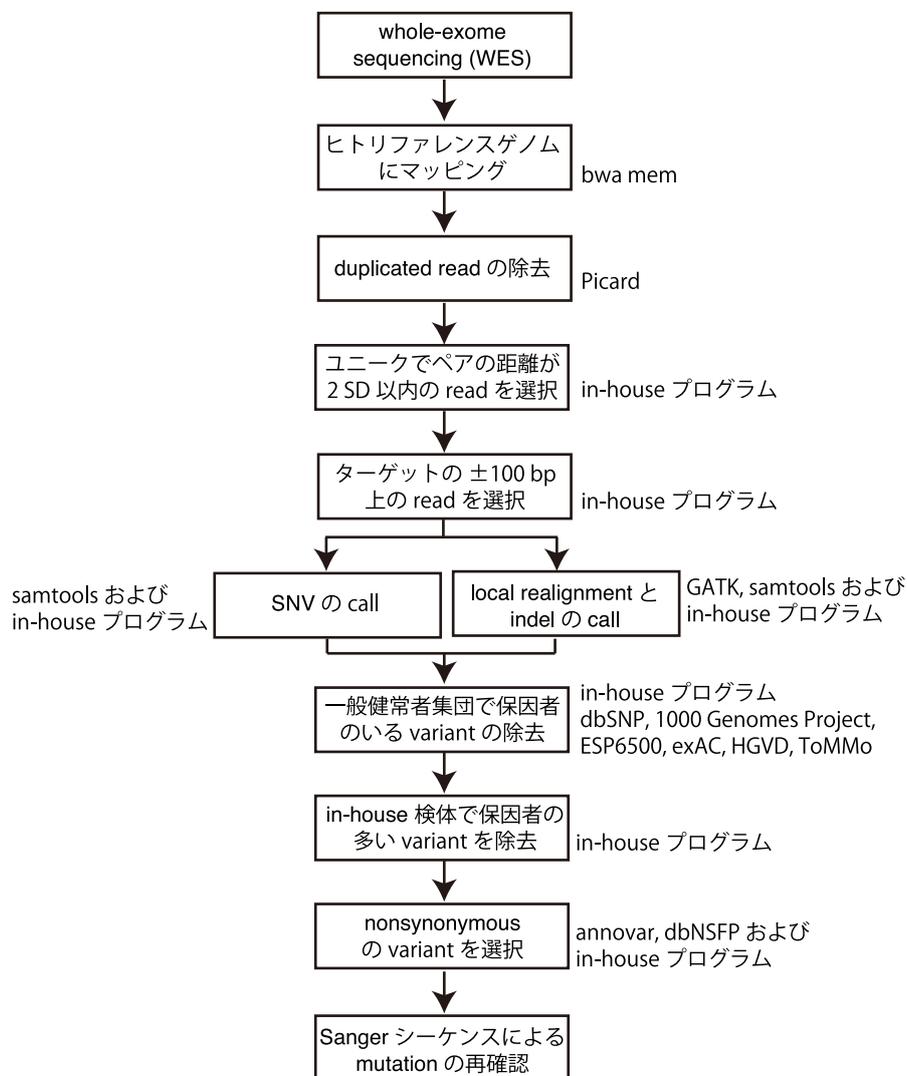


図 5. NGS データの解析パイプラインの概要
 枠外には用いたツール等を記載した。

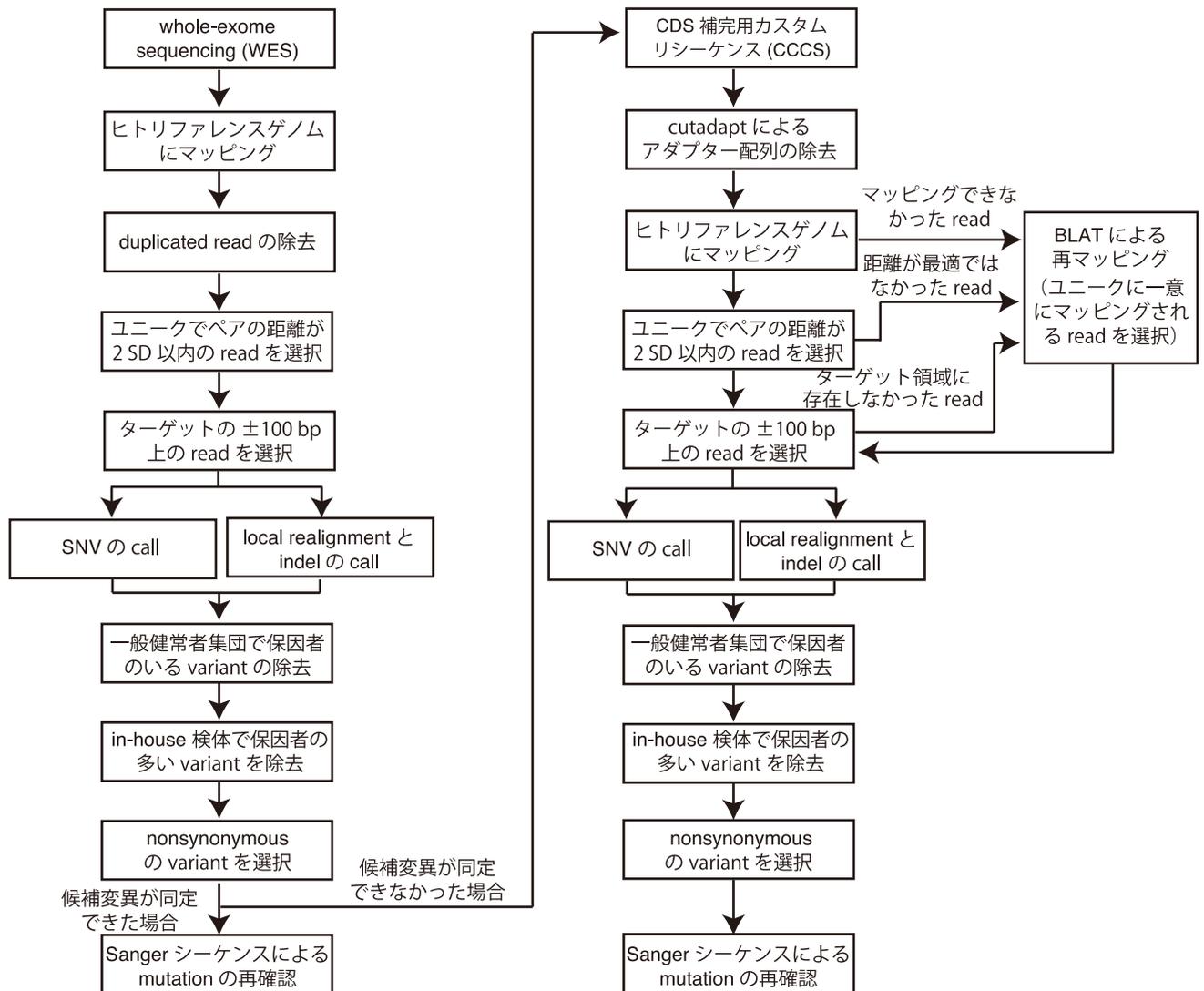


図 7. CCCS を含めた NGS データの解析パイプラインの概要

WES で疾患原因候補変異が見つからなかった検体に関しては、WES に続いて CCCS を実施した。

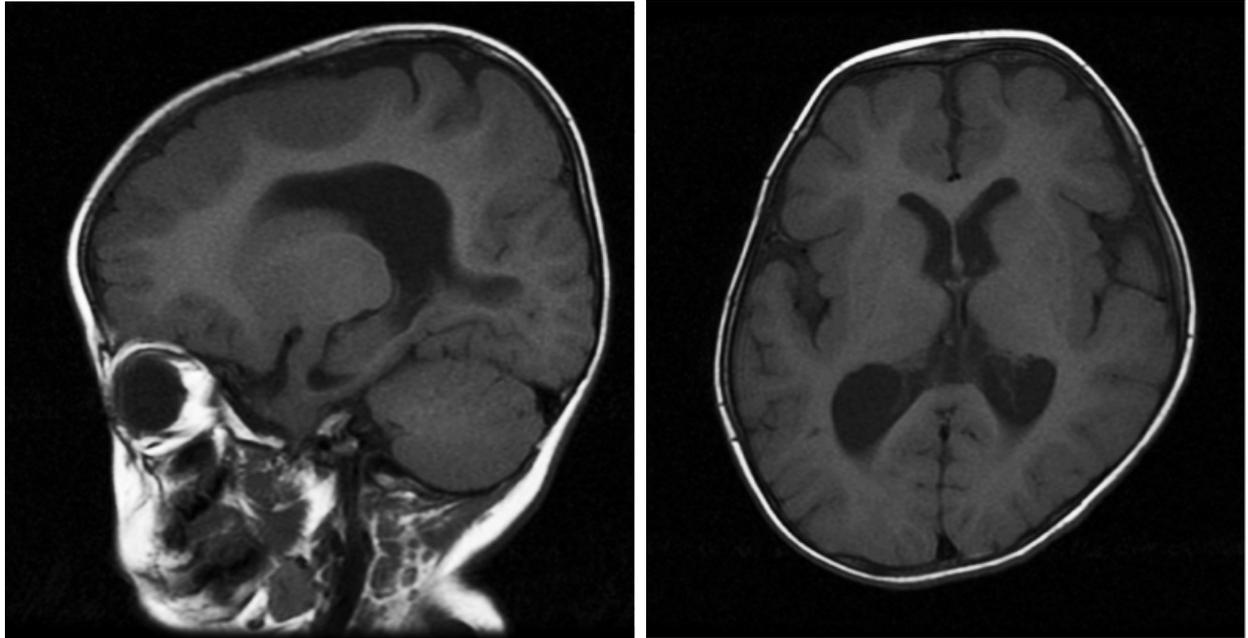


図 8. ACTB 遺伝子に変異が認められた Baraitser-Winter 症候群 (BRWS) 患者の MRI 画像
側脳室の拡大と白質容積の減少、前頭葉の厚脳回が認められる。

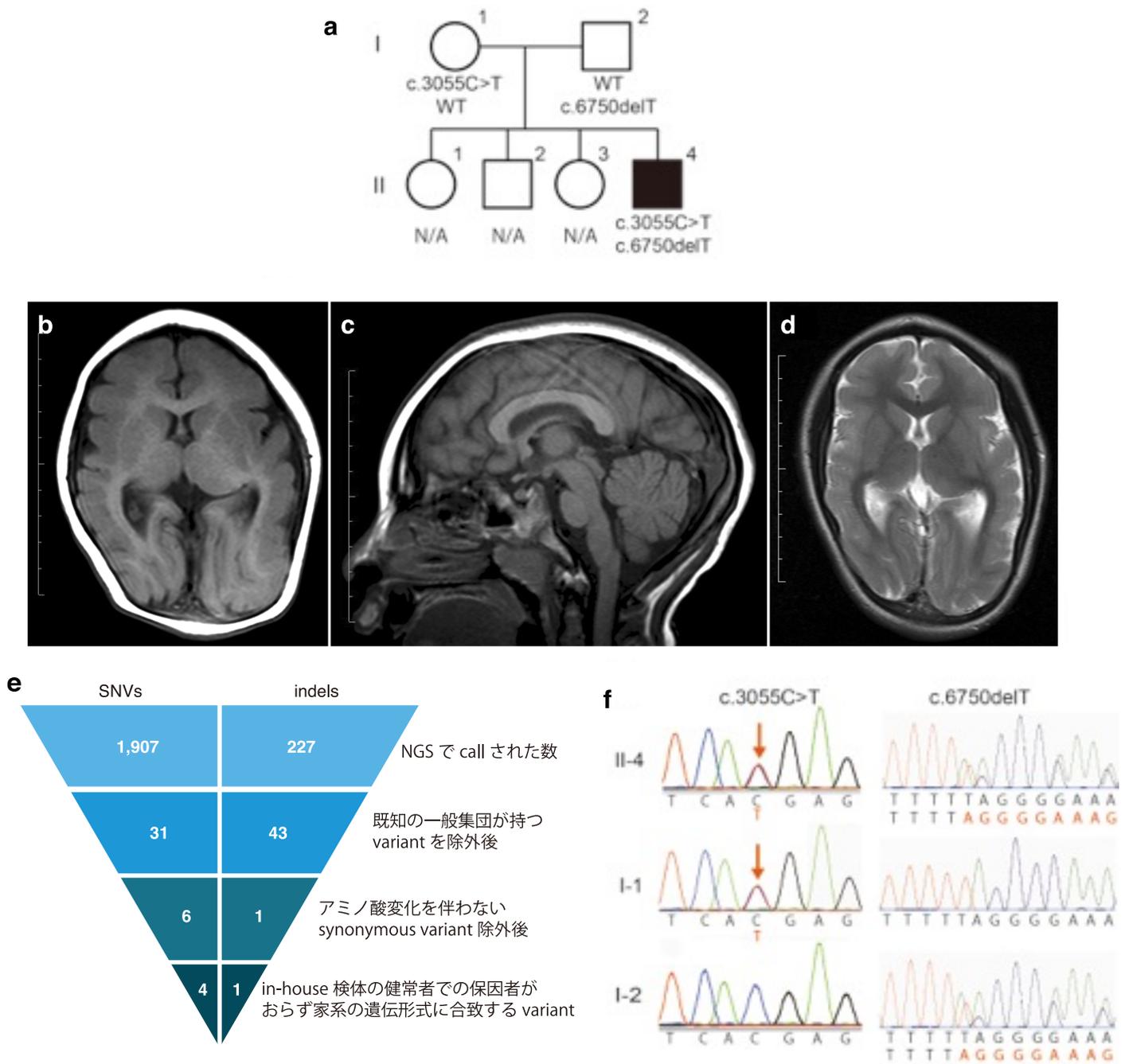


図 9. ASPM 遺伝子に変異が認められた小頭症患者

(a) 患者の家系図。(b-d) MRI にて側脳室の拡大と厚脳回が認められた。(b) は 15 週齢時点、(c) と (d) は 5 歳児の時点での MRI。
 (e) NGS による神経系遺伝子ターゲットリシーケンスでの絞り込み。(f) 見つかった変異の Sanger シーケンスによる確認実験結果。

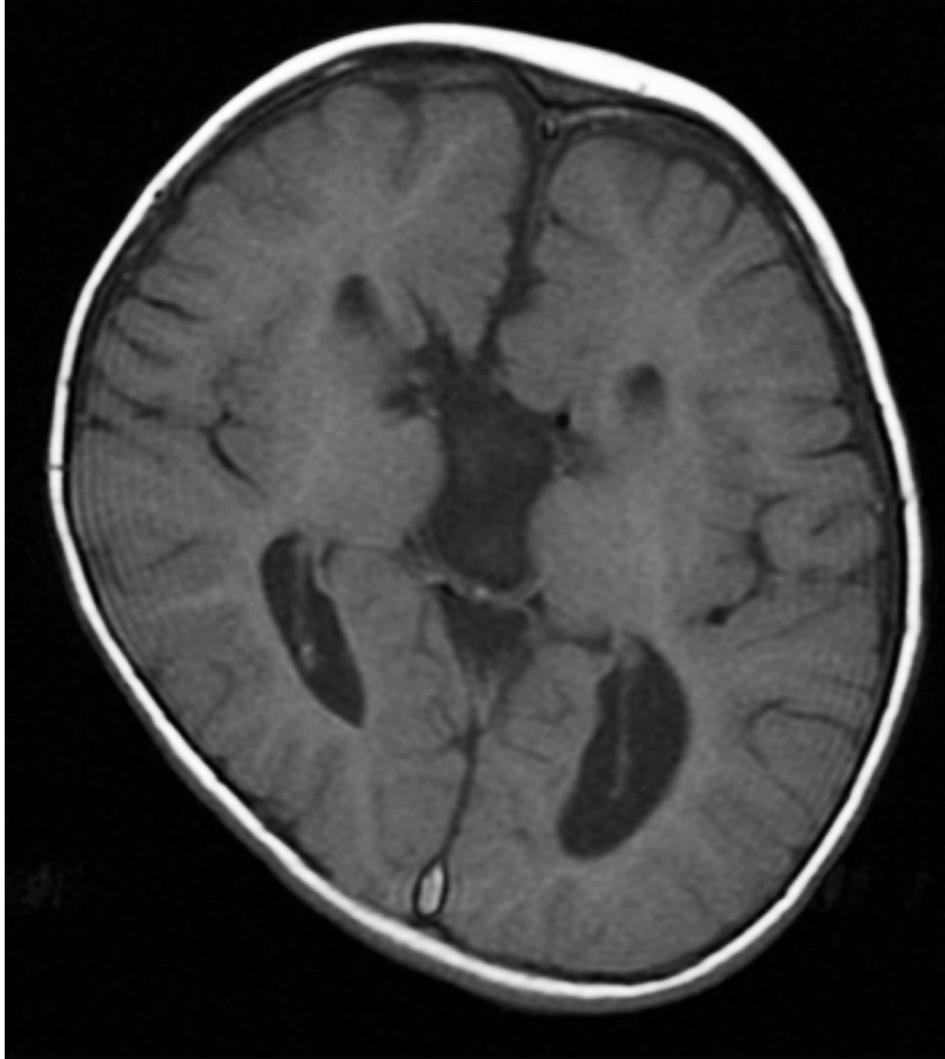


図 10. PTPN11 遺伝子に変異が認められた Noonan 症候群患者の MRI 画像
脳梁欠損が認められる。

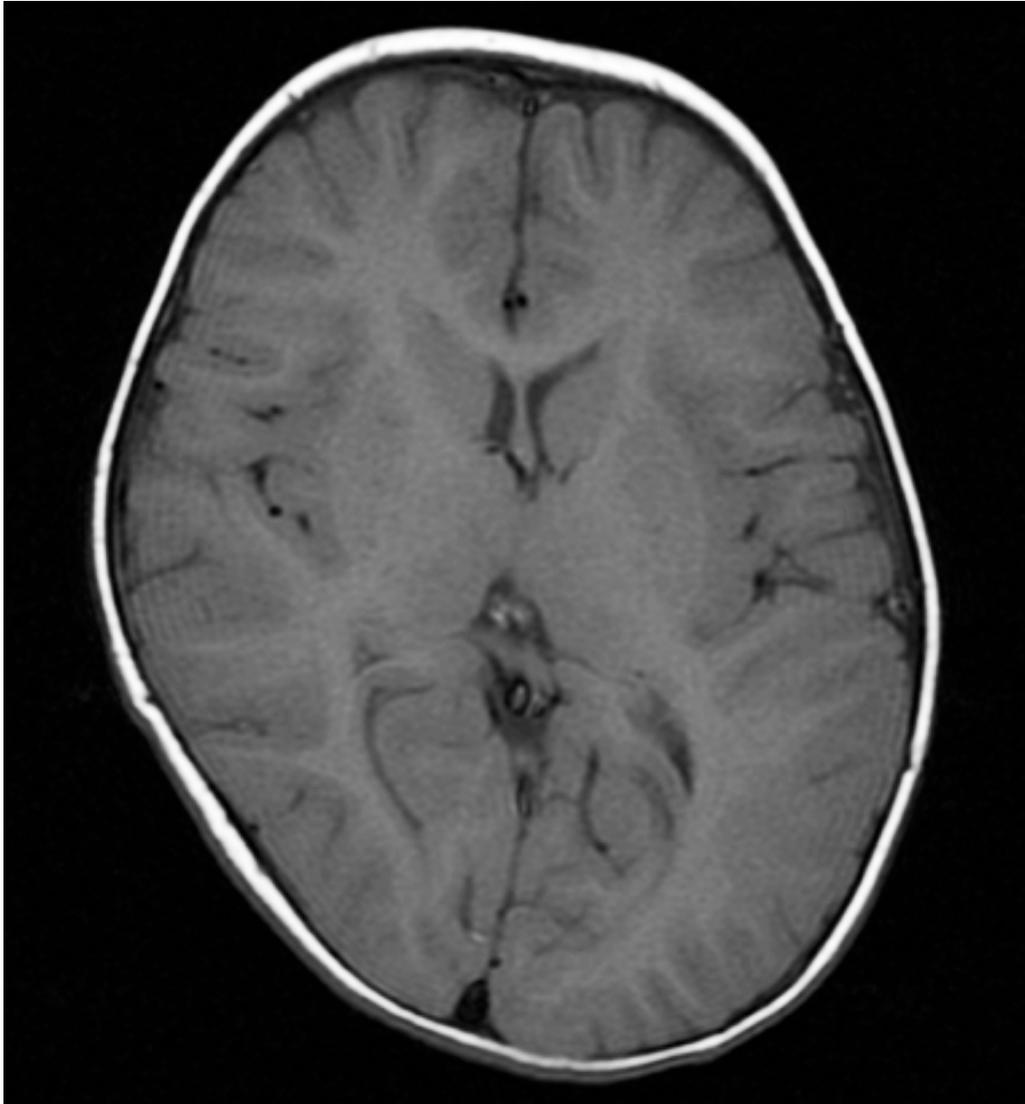


図 11. DYRK1A 遺伝子に変異が認められた相対的大頭症患者の MRI 画像
MRI からは特に異常は見受けられなかった。

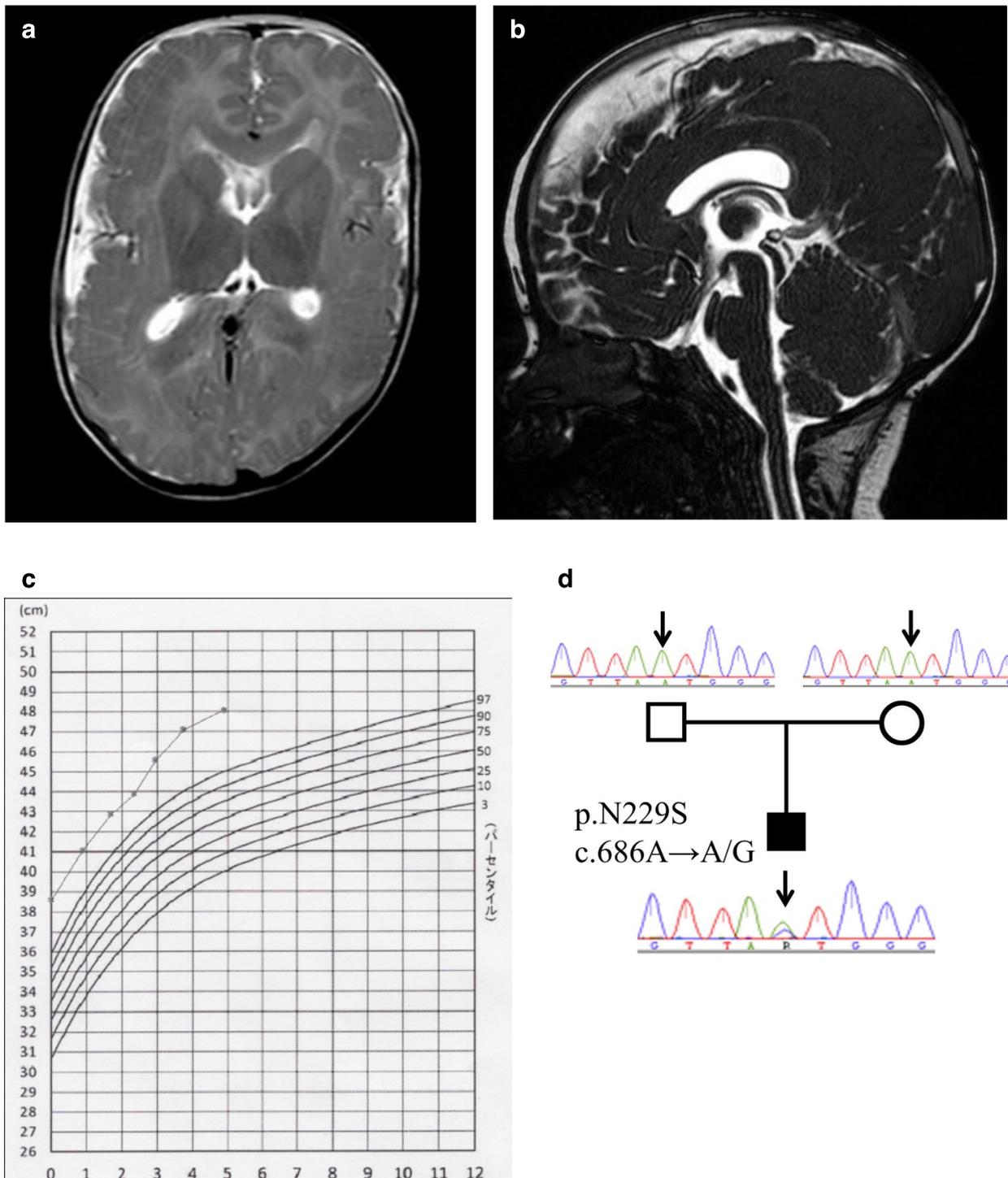


図 12. ATK3 遺伝子に変異が認められた巨脳症 (MCAP) 患者

(a) 生後 5 日目の MRI 画像。両側シルビウス裂近傍に多小脳回を認めた。(b) 生後 4 ヶ月目の MRI 画像。小脳の肥大による後頭蓋窩の狭小化、脳梁の肥大を認めた。(c) 頭囲の推移 (横軸は月齢で 12 月齢まで表示している値は平均値に基づく)。出生児から明らかに頭囲が大きく、生後 3 ヶ月の時点で 4 SD を超えていた。(d) AKT3 遺伝子で変異が認められた箇所の患者と両親の Sanger シーケンスの確認結果。

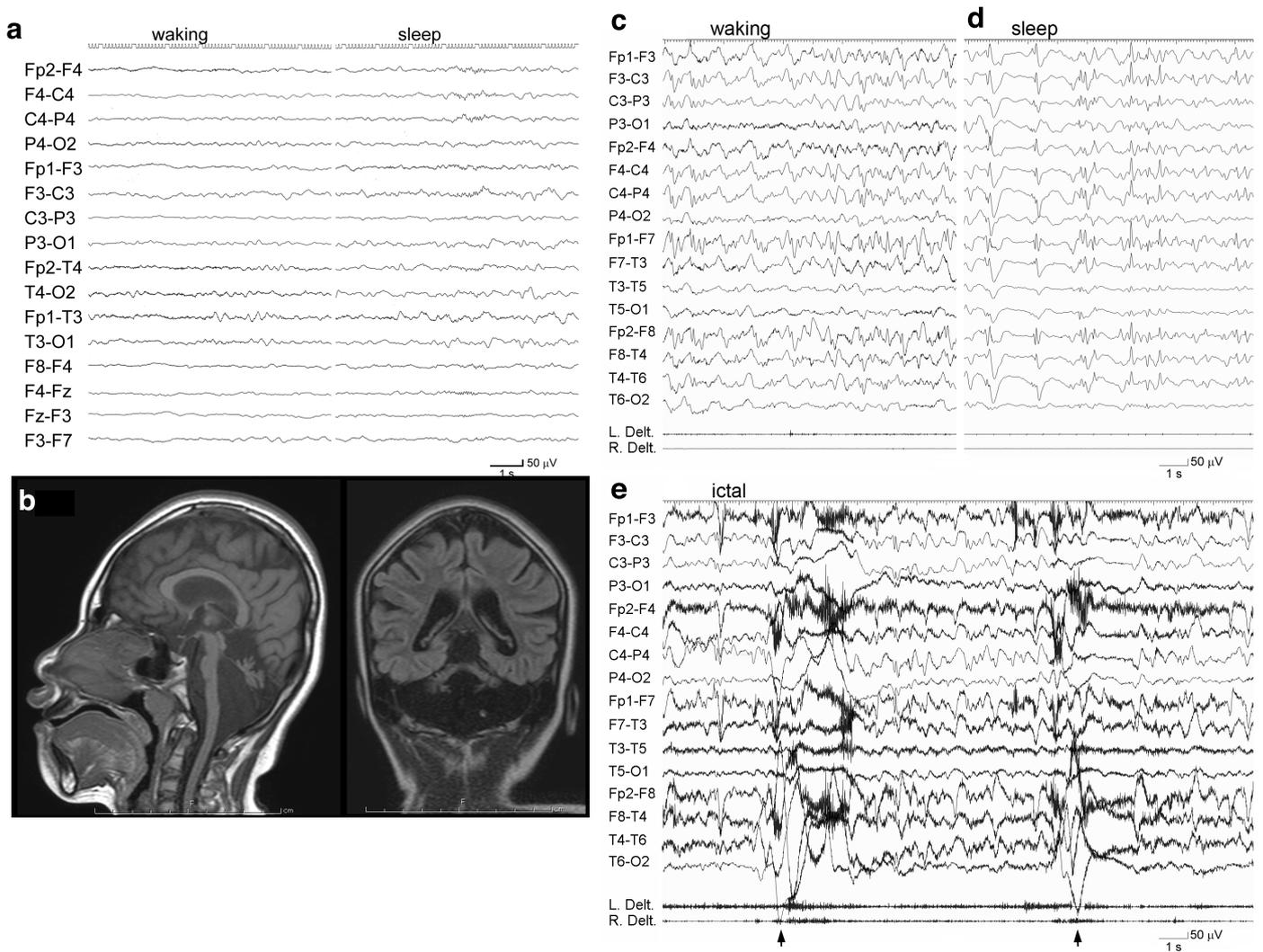


図 13. CASK 遺伝子に変異が認められた MICPCH 患者

(a) 幼児期の EEG。起床時、睡眠時共にてんかん性の異常は認められなかった。(b) 13 歳時の MRI 画像。重篤な橋および小脳の形成不全が確認された一方で脳梁は保持されていた。(c) 11 歳時の覚醒時の発作間欠期 EEG。(d) 11 歳時の睡眠時の発作間欠期 EEG。(e) 11 歳時のてんかん性攣縮発作時の EEG。(Delt は三角筋を示す)

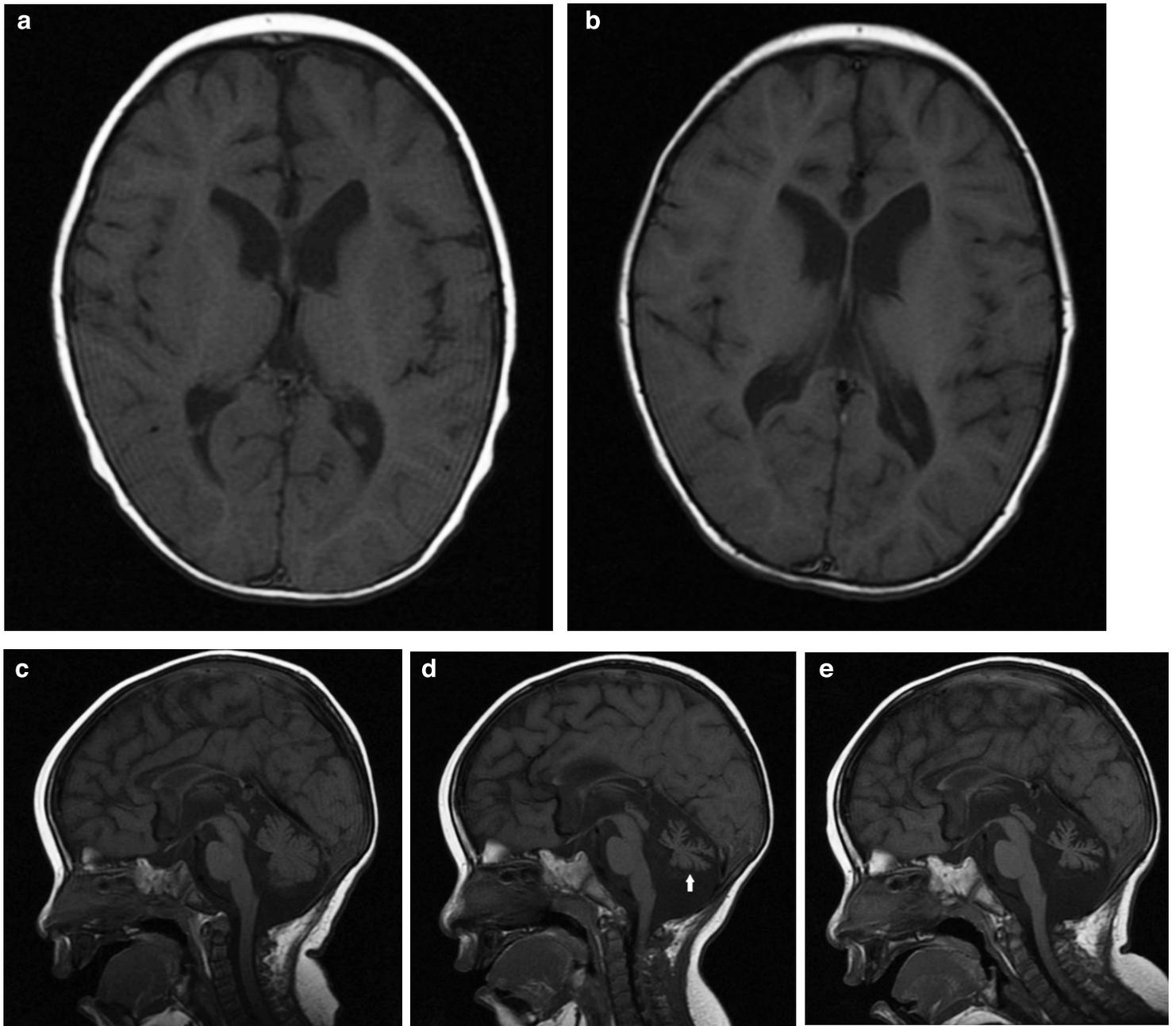


図 14. KIF1A 変異が見られた進行性神経変性症患者

(a) 1 歳時、(b) 4 歳時の MRI。脳室拡大、白質の減少が確認できる。軽微の側脳室拡大も見られる。

(c) 1 歳時、(d) 3 歳時、(e) 4 歳時の MRI。脳梁低形成と小脳萎縮が確認できる。小脳虫部は萎縮性であった (d の矢印部)。



図 15. NFIA 変異が見られた脳梁低形成と尿路異常症の患者

(a) 4 歳時の MRI により、嚢胞と脳室拡大、多小脳回 (矢印部) が確認された。(b) 4 歳時の MRI で小脳低形成 (アスタリスク印) が確認された。(c) 排尿時膀胱尿道造影 (VCUG) で grade IV の膀胱尿管逆流現象 (VUR) が確認された。

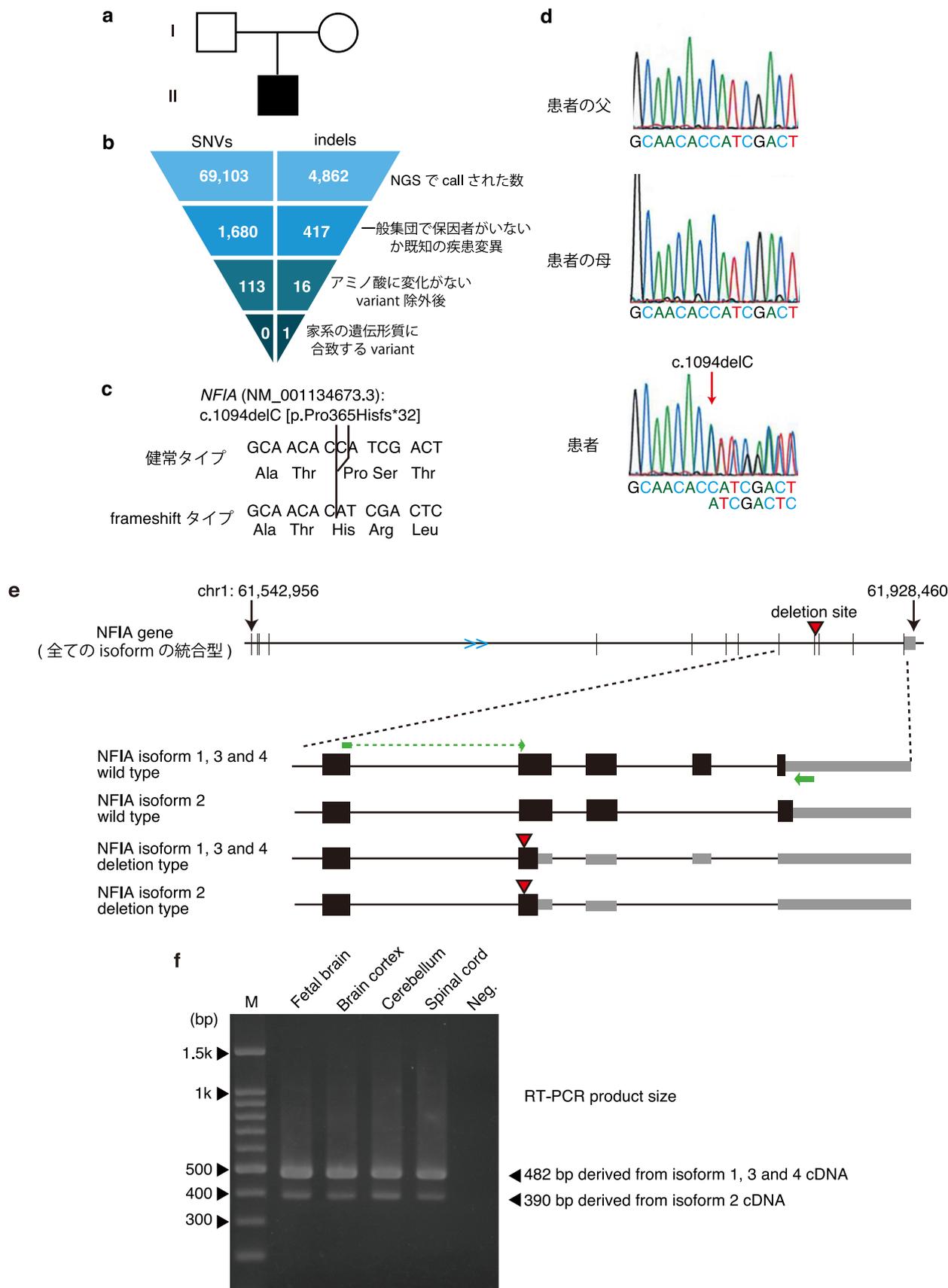


図 16. NFIA 変異が見られた脳梁低形成と尿路異常症の患者の遺伝子解析結果

(a) 患者の家系図。(b) NGS の call された variant の絞り込み結果数。最終的に NFIA の欠失 1 個だけが残った。(c) 見つかった欠失によるタンパク質の frameshift。(d) 患者とその両親の Sanger シーケンスによる確認結果。患者のみに欠失が見られた。(e) NFIA 遺伝子の各 isoform の遺伝子構造と、欠失部位、および PCR での確認用の primer の位置。三角が欠失箇所を示す。緑色の矢印は primer の部位を示し、L 側の primer は intron を挟むように設計した。primer の配列は L 側が 5' -CATTACAGGACCCAGAGCAAG-3'、R 側が 5' -GCTGCAACTTTTATCCCAGGT-3' とし、isoform 2 の推定増幅産物長は 390 bp、それ以外の isoform の推定増幅産物長は 482 bp にて設定した。(f) 各種神経系組織での RT-PCR の結果。isoform 2 とそれ以外の両方の予想位置にバンドが確認された。M はラダーマーカー、Neg. は DNA を入れていないネガティブコントロールを示す。

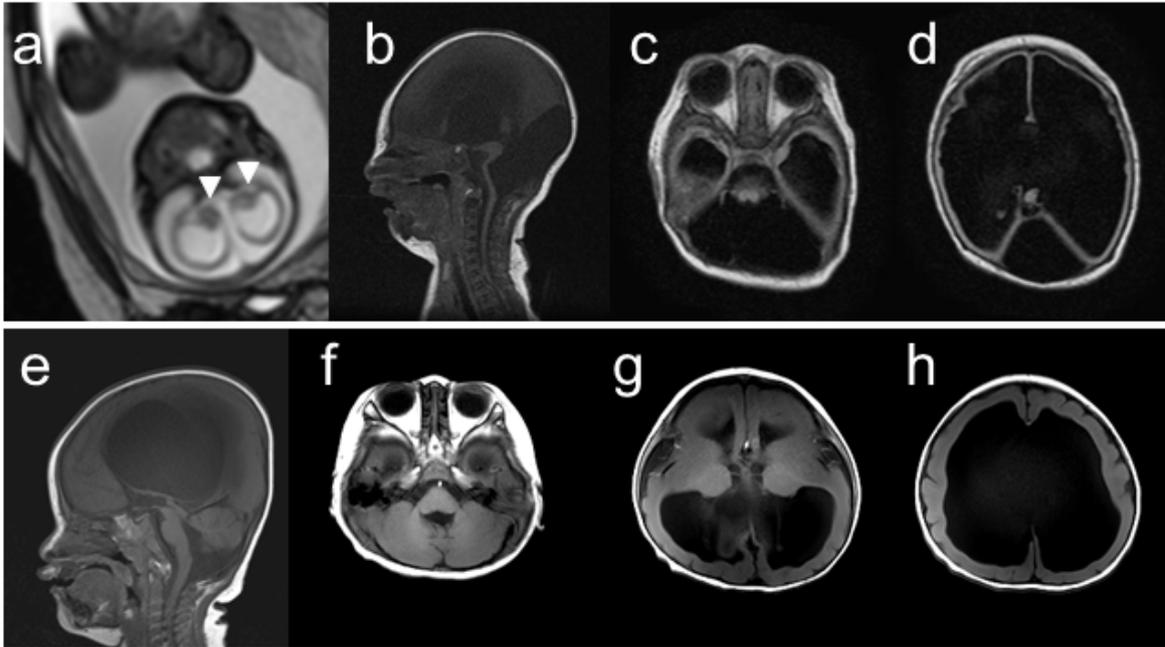
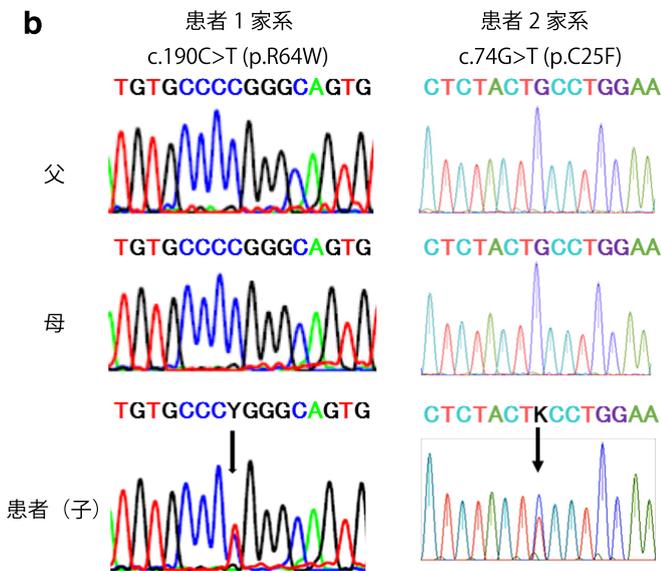
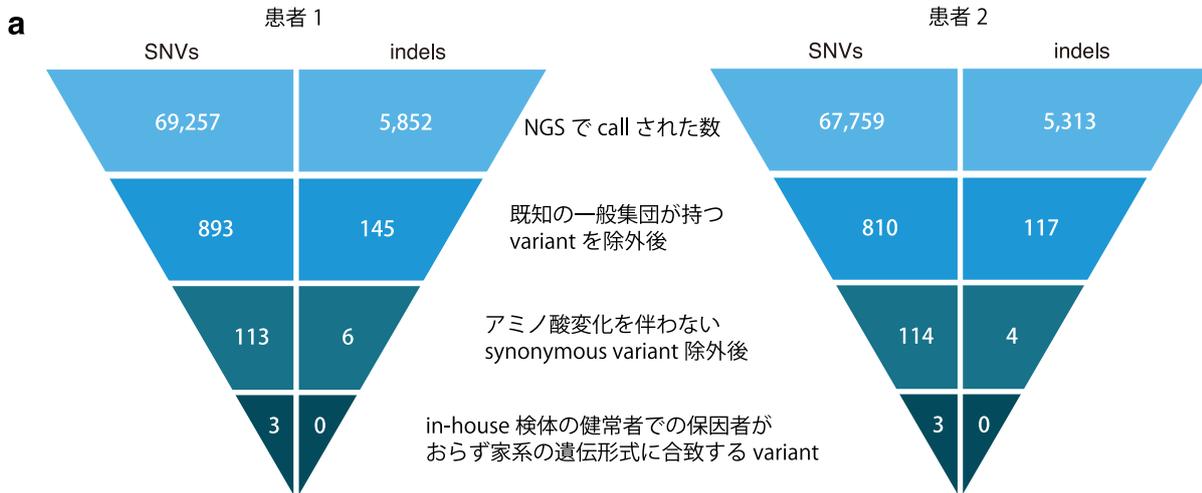


図17. TUBA1A変異が見られた2例の患者のMRI

(a) 1例目の患者の妊娠28週目。三角印は大腦基底核を示す。(b - d) 1例目の患者の生後6日目MRI。薄脳実質、脳幹・小脳・脳梁形成不全が確認され、大部分の頭蓋内は脳脊髄液で満たされている。(e - h) 2例目の患者の生後1年時のMRI。薄脳実質・脳室拡大・脳回欠損・厚脳回・大腦基底核の形成不全・小脳虫部の低形成が確認された。脳幹は正常であった。



c

	R64	C25
<i>TUBA1A</i> mutant	...GAGKHVP W AVFVDLE...	...NACWELY F LEHGIQP...
<i>H.sapiens TUBA1A</i>	...GAGKHVP P RAVFVDLE...	...NACWELY C LEHGIQP...
<i>P.troglodytes</i>	...GAGKHVP P RAVFVDLE...	...NACWELY C LEHGIQP...
<i>M.mulatta</i>	...GAGKHVP P RAVFVDLE...
<i>C.lupus</i>	...GAGKHVP P RAVFVDLE...	...NACWELY C LEHGIQP...
<i>B.taurus</i>	...GAGKHVP P RAVFVDLE...	...NACWELY C LEHGIQP...
<i>M.musculus</i>	...GAGKHVP P RAVFVDLE...	...NACWELY C LEHGIQP...
<i>R.norvegicus</i>	...GAGKHVP P RAVFVDLE...	...NACWELY C LEHGIQP...
<i>G.gallus</i>	...GAGKHVP P RAVFVDLE...
<i>D.rerio</i>	...GAGKHVP P RAVFVDLE...	...NACWELY C LEHGIQP...
<i>C.elegans</i>	...PSGKHVP P RAIFVDLE...	...NACWELY C LEHGITP...
<i>X.tropicalis</i>	...GAGKHVP P RAVFVDLE...	...NACWELY C LEHGIQP...

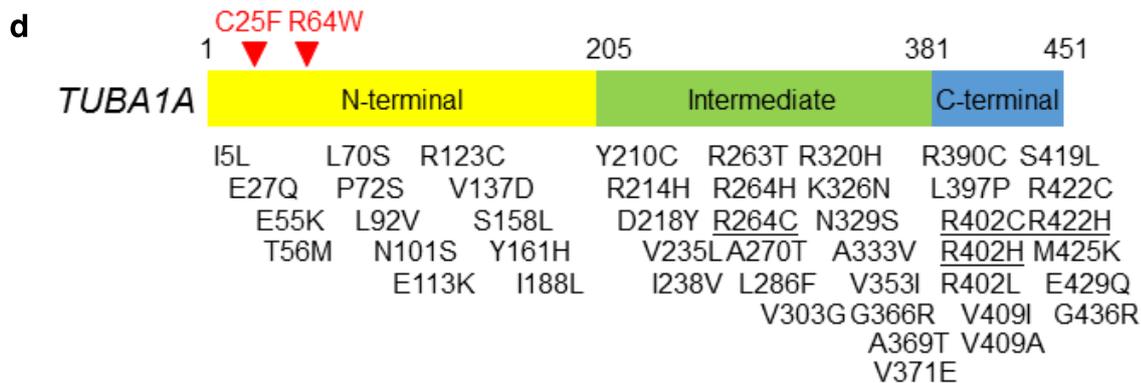


図 18. TUBA1A 変異が見られた 2 例の患者家系の遺伝子解析

(a) WES 解析による候補変異の絞り込み。両家系共に最終的に 3 個の SNV が残ったが、2 家系に共通して TUBA1A 遺伝子変異が確認された。
 (b) 見つかった TUBA1A 変異の Sanger シーケンスによる確認結果。両家系とも患者のみが変異を保因し、両親には変異が見られなかった。
 (c) 見つかった TUBA1A 変異の種間保存性。変異が見つかった箇所の保存性は高い。(d) TUBA1A タンパク構造内における見つかった変異の位置。両変異共に N 末端側に存在する。下部に表記したのは神経変性に関わる既報変異。下線を引いた変異は重複箇所に報告がある変異。

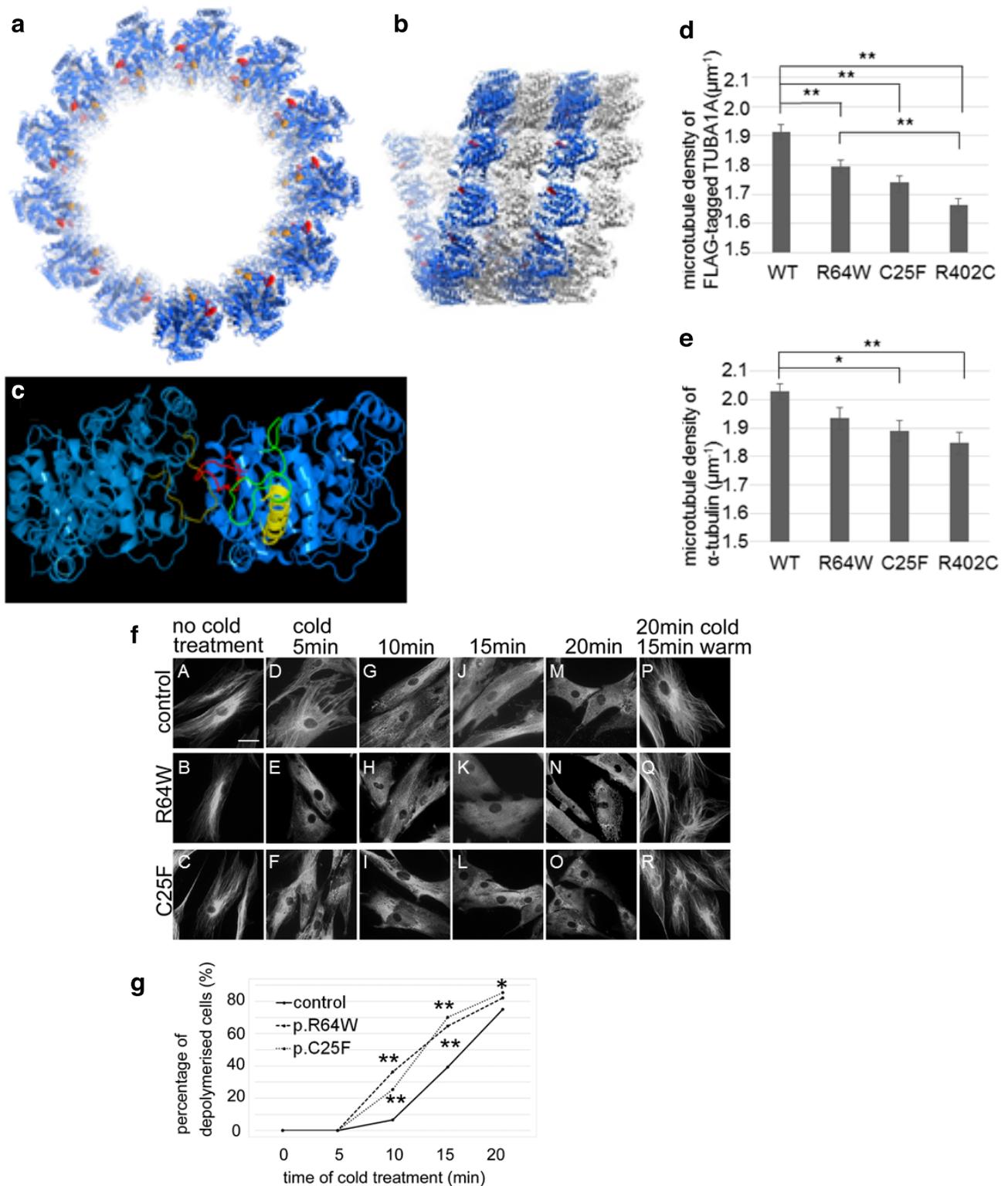


図 19. TUBA1A 変異の機能解析

(a) tubulin のタンパク質の立体構造を上部から見た模式図。13 個のヘテロダイマーが連なった構造をしている。赤色が R64、橙色が C25 の α -tubulin のアミノ酸残基の位置を示す。(b) tubulin のタンパク質の立体構造を横側から見た模式図。 α -tubulin は青色で、 β -tubulin は灰色で示す。赤色が R64 の α -tubulin の位置を示し、C25 は内側にあり横方向からは見えない。(c) α -tubulin の高解像度模式図。左右共に α -tubulin を示す。茶色は M ループ、赤は H1' -S2 ループ、緑は H1-H1' ループ、黄色は helix H1 構造をそれぞれ示す。(d) COS7 で各 TUBA1A を強制発現させ、その細胞内での TUBA1A の密度を測定した。WT は野生型、R64W と C25F は今回発見された変異、R402C は以前に脳回欠損の原因変異であることが知られている変異型の TUBA1A。** $P < 0.01$ 。One-Way ANOVA と Tukey の post-hoc test による。(e) (d) と同様に TUBA1A を強制発現させ、その細胞を α -tubulin 抗体で染色し α -tubulin の密度を測定した結果。エラーバーは標準誤差を示す。TUBA1A の定量的場合 (d) と同様の傾向が得られた。 $*P < 0.05$ 、** $P < 0.01$ 。One-Way ANOVA と Tukey の post-hoc test による。(f) WT 型の control と各変異型の TUBA1A を持つ繊維芽細胞で水冷による解重合を調査した。A-C は水冷未処理状態の細胞。水冷 15 分後、変異型の TUBA1A を持つ tubulin は大部分の細胞に解重合が見られたが、control ではまだ解重合が起こっていない細胞が散見された (J-L)。(g) (f) の実験について細胞をカウントし、時間経過による解重合の割合を算出した。 $*P < 0.05$ 、** $P < 0.01$ 。control に対する Welch's t-test で Bonferroni 補正後の P 値による。

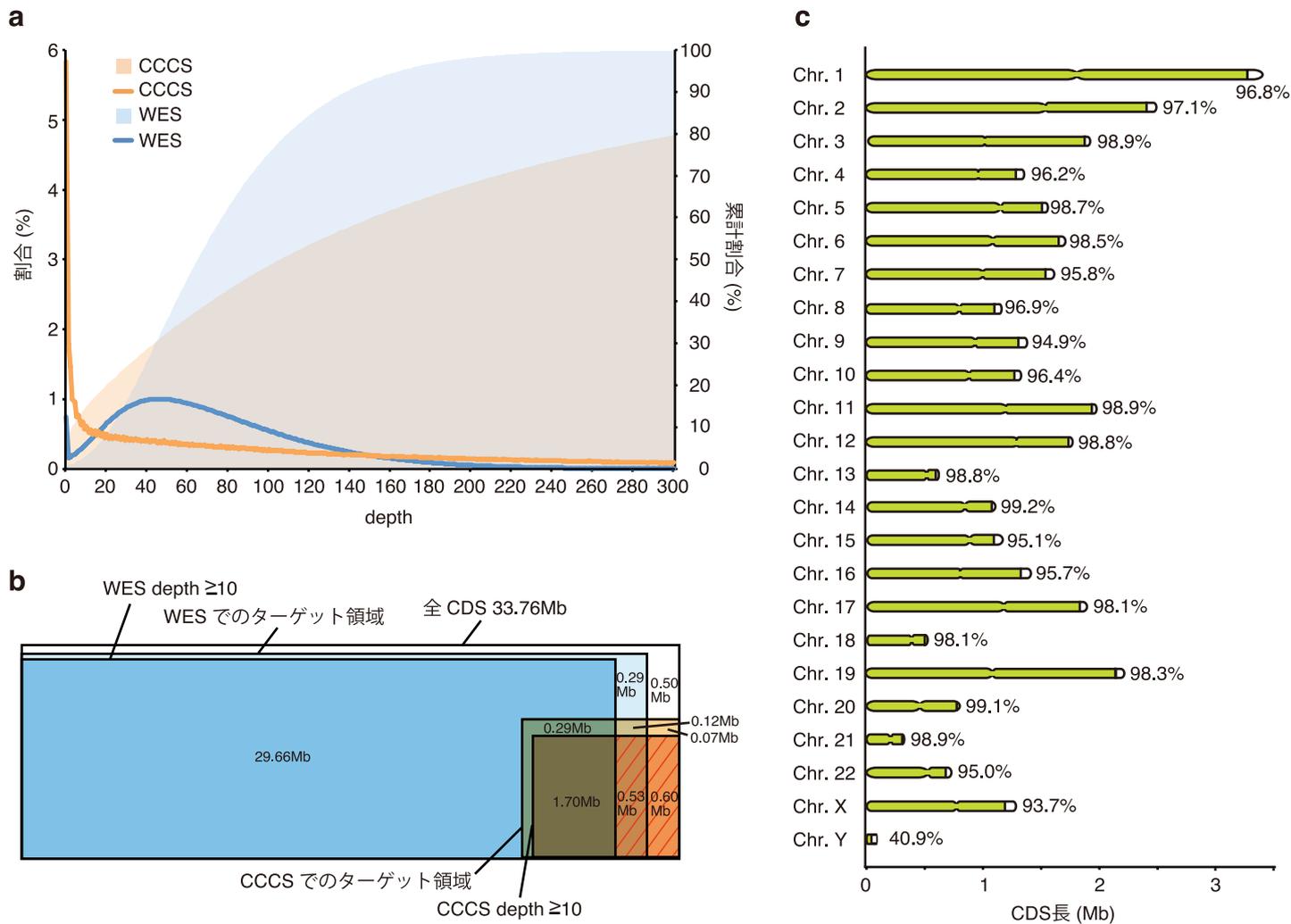
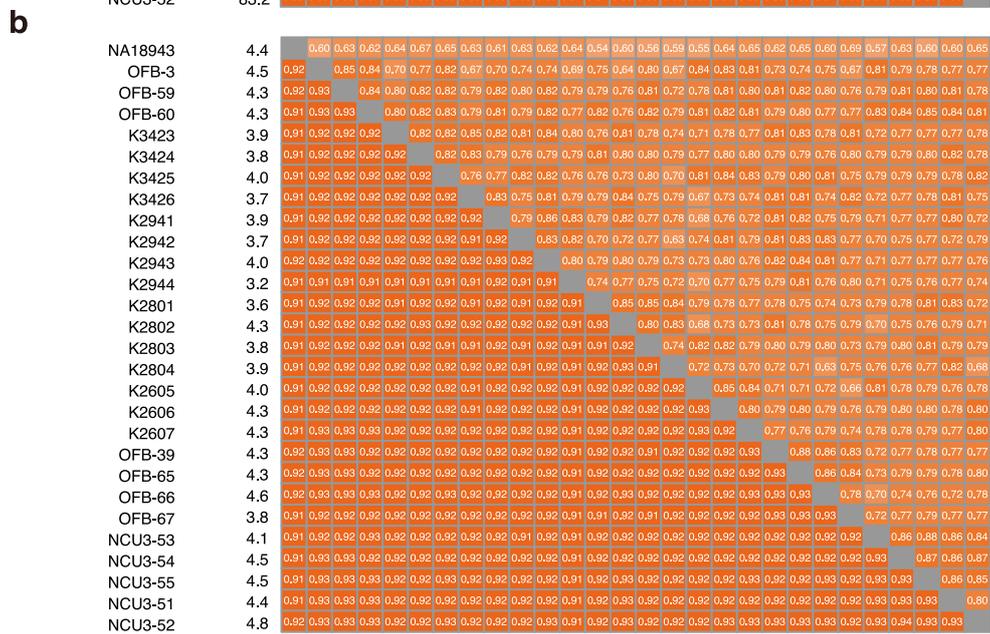
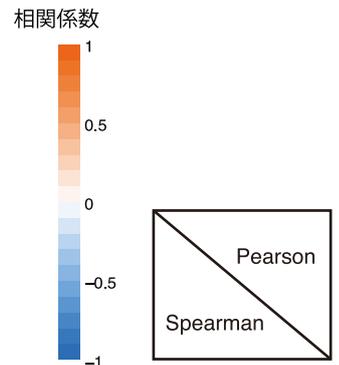
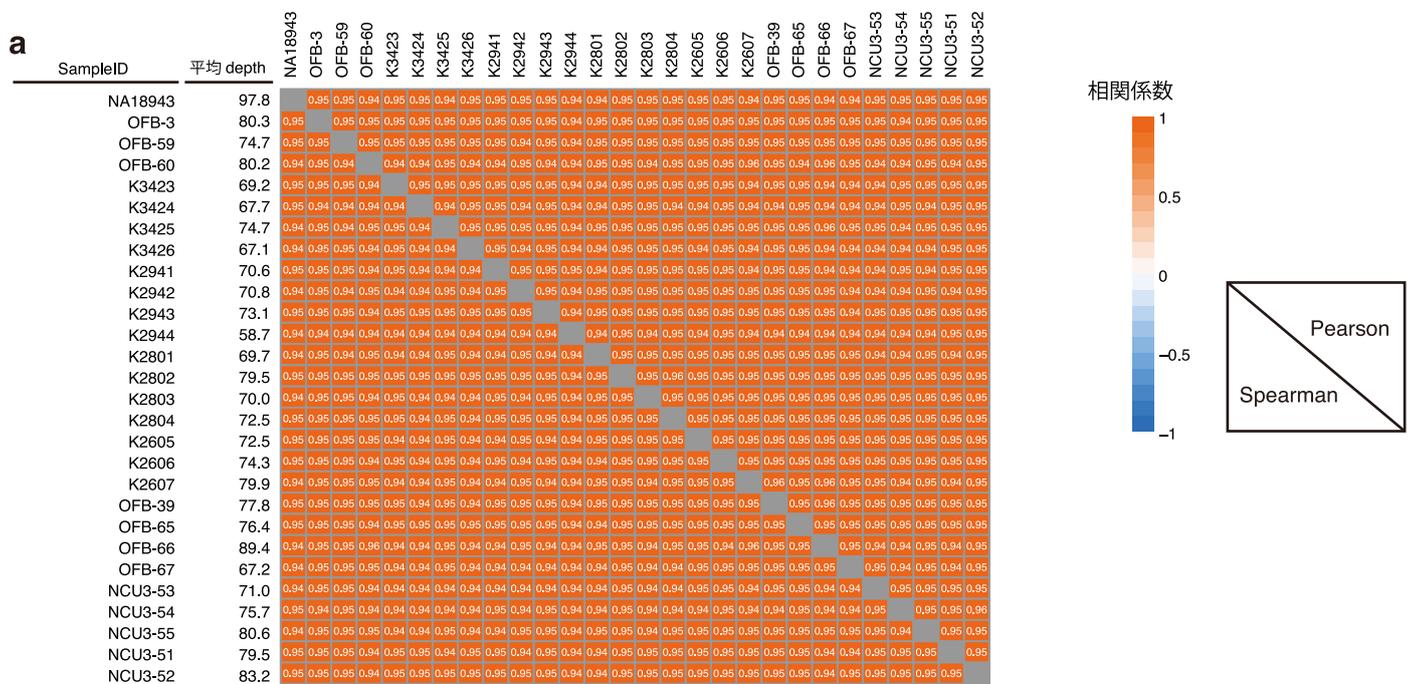


図 20. CDS 領域の NGS シーケンス

(a) WES と CCCS の depth の分布とその累計分布。線で示したのが WES と CCCS の各 depth での割合（左側の数値）で、面で示したのがその累計割合（右側の数値）。(b) WES と CCCS による CDS のカバー率。赤い斜線で示した部分が WES では call が得られなかったが、新たに CCCS によって call が得られた CDS 領域。(c) WES と CCCS の統合結果として、各染色体で call が得られた CDS の割合。各染色体の棒の長さは実際の染色体自身の長さではなく、各染色体に含まれる CDS 領域の長さを示す。Y 染色体のみは男性検体のみについて調査した。全ての図は CCCS を実施した 28 検体の平均値を示す。



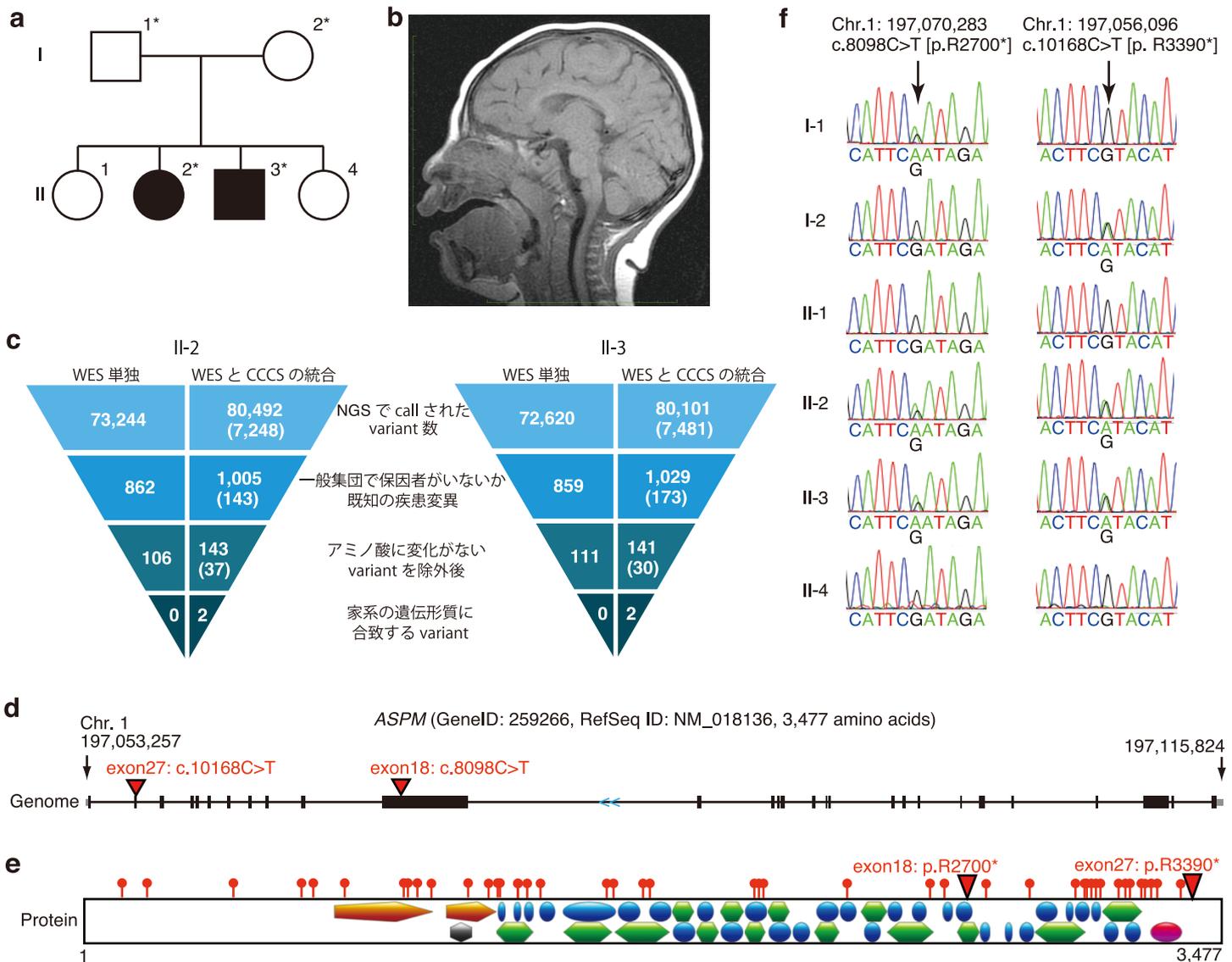


図 22. WES と CCCS の統合解析により見つかった小頭症家系の ASPM 変異

(a) 小頭症家系の家系図。アスタリスクマークは WES および CCCS を実施した個人を示す。(b) II-2 患者の 4 歳時の MRI。脳容積の減少が見られ、特に大脳前頭葉の容積縮小が目立つ。(c) II-2 患者と II-3 患者での NGS 解析での疾患変異絞り込み推移。逆三角形の左側は CCCS を行う前の WES 単独の絞り込み結果を、逆三角形の右側は WES と CCCS を統合解析した場合の絞り込み結果を示す。最終的に ASPM 遺伝子上の 2 個の compound heterozygous な変異のみが残った。(d) ASPM のゲノム上の構造と見つかった変異箇所。薄いグレーは非翻訳領域 (UTR) を示し、黒い箱は CDS 領域を示す。水色の矢印 (<<) は転写の方向を示す。赤三角は今回見つかった変異箇所を示す。(e) ASPM タンパク質の構造を示す。3,477 アミノ酸から成り、上部の赤いピンはこれまでに報告されている小頭症原因変異箇所を示す。赤三角は今回見つかった変異箇所を示す。枠の色の付いたマークは各タンパクドメインを示し、オレンジの五角形は calponin homology (IPR001717: InterPro ID)、黒の六角形は calmodulin-regulated spectrin-associated protein CH (IPR022613)、緑色の六角形は P-loop containing nuclease triphosphate hydrolase (IPR027417)、赤色の円形は armadillo-type hold domain (IPR016024) を示す。青色の円形は IQ-motif EF-hand binding site (IPR0000048) を示す。(f) 今回見つかった変異について、健常者の子を含め、家系図内の全員について Sanger シーケンスで確認した結果を示す。primer は c.8093C>T の確認の際が、exon18F: TGCCTCTAAAAGCAGCCTGAA および exon18R: CAGTGCGTACCCAAGCAGTTA の組み合わせにて、c.10168C>T の確認の際が、exon 27F: TGGTCCTTACAGGTGTTTCTGG および exon27R: GGAGGCAGAGATTGCATTGAG の合成 oligo を用いた。

表 1. 神経系遺伝子カスタムリシーケンスにてターゲットにした 284 遺伝子リスト

遺伝子シンボル	遺伝子名	染色体上の位置 *	EntrezGeneID	RefSeqID
ACTB	actin, beta	7: 5,566,779 - 5,570,232	60	NM_001101
AHI1	Abelson helper integration site 1	6: 135,605,110 - 135,818,903	54806	NM_017651
AKT3	v-akt murine thymoma viral oncogene homolog 3	1: 243,651,535 - 244,006,886	10000	NM_005465
ALDH1A2	aldehyde dehydrogenase 1 family, member A2	15: 58,245,622 - 58,358,121	8854	NM_003888
ALDH1L1	aldehyde dehydrogenase 1 family, member L1	3: 125,822,404 - 125,900,029	10840	NM_012190
ALG6	ALG6, alpha-1,3-glucosyltransferase	1: 63,833,261 - 63,904,233	29929	NM_013339
ANK3	ankyrin 3, node of Ranvier (ankyrin G)	10: 61,786,056 - 62,493,284	288	NM_020987
AP1S2	adaptor-related protein complex 1, sigma 2 subunit	X: 15,843,929 - 15,873,100	8905	NM_003916
APEX1	APEX nuclease (multifunctional DNA repair enzyme) 1	14: 20,923,290 - 20,925,931	328	NM_001641
ARC	activity-regulated cytoskeleton-associated protein	8: 143,692,405 - 143,695,833	23237	NM_015193
ARFGEF2	ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin A-inhibited)	20: 47,538,275 - 47,653,230	10564	NM_006420
ARHGEF15	Rho guanine nucleotide exchange factor (GEF) 15	17: 8,213,556 - 8,225,834	22899	NM_025014
ARL13B	ADP-ribosylation factor-like 13B	3: 93,698,983 - 93,774,522	200894	NM_144996
ARX	aristales related homeobox	X: 25,021,811 - 25,034,065	170302	NM_139058
ASPM	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)	1: 197,053,257 - 197,115,824	259266	NM_018136
ATP2C1	ATPase, Ca ⁺⁺ transporting, type 2C, member 1	3: 130,569,369 - 130,735,556	27032	NM_014382
ATP6V0A2	ATPase, H ⁺ transporting, lysosomal V0 subunit a2	12: 124,196,865 - 124,246,302	23545	NM_012463
ATXN10	ataxin 10	22: 46,067,678 - 46,241,187	25814	NM_013236
BHMT	betaine--homocysteine S-methyltransferase	5: 78,407,604 - 78,428,113	635	NM_001713
BHMT2	betaine--homocysteine S-methyltransferase 2	5: 78,365,547 - 78,385,897	23743	NM_017614
BRCA1	breast cancer 1, early onset	17: 41,196,312 - 41,277,500	672	NM_007294
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	15: 40,453,210 - 40,513,337	701	NM_001211
CACNA1A	calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	19: 13,317,256 - 13,617,274	773	NM_000068
CACNA1H	calcium channel, voltage-dependent, T type, alpha 1H subunit	16: 1,203,241 - 1,271,772	8912	NM_021098
CACNB4	calcium channel, voltage-dependent, beta 4 subunit	2: 152,689,285 - 152,955,593	785	NM_000726
CAMK2A	calcium/calmodulin-dependent protein kinase II alpha	5: 149,599,054 - 149,669,403	815	NM_015981
CASK	calcium/calmodulin-dependent serine protein kinase (MAGUK family)	X: 41,374,187 - 41,782,287	8573	NM_003688
CBS	cystathionine-beta-synthase	21: 44,473,301 - 44,496,472	875	NM_000071
CC2D2A	coiled-coil and C2 domain containing 2A	4: 15,471,489 - 15,603,180	57545	NM_020785
CDK5RAP2	CDK5 regulatory subunit associated protein 2	9: 123,151,147 - 123,342,437	55755	NM_018249
CDKL5	cyclin-dependent kinase-like 5	X: 18,443,725 - 18,671,749	6792	NM_003159
CELSR1	cadherin, EGF LAG seven-pass G-type receptor 1	22: 46,756,731 - 46,933,067	9620	NM_014246
CENPJ	centromere protein J	13: 25,456,412 - 25,497,027	55835	NM_018451
CEP152	centrosomal protein 152kDa	15: 49,030,135 - 49,103,343	22995	NM_014985
CEP290	centrosomal protein 290kDa	12: 88,442,790 - 88,535,993	80184	NM_025114
CEP57	centrosomal protein 57kDa	11: 95,523,625 - 95,565,857	9702	NM_014679
CFL1	cofilin 1 (non-muscle)	11: 65,622,282 - 65,625,804	1072	NM_005507
CHKA	choline kinase alpha	11: 67,820,326 - 67,888,858	1119	NM_001277
CHRNA2	cholinergic receptor, nicotinic, alpha 2 (neuronal)	8: 27,317,278 - 27,336,813	1135	NM_000742
CHRNA4	cholinergic receptor, nicotinic, alpha 4 (neuronal)	20: 61,974,662 - 61,992,748	1137	NM_000744
CHRN2	cholinergic receptor, nicotinic, beta 2 (neuronal)	1: 154,540,257 - 154,552,354	1141	NM_000748
CLCN2	chloride channel, voltage-sensitive 2	3: 184,063,973 - 184,079,439	1181	NM_004366
COL4A1	collagen, type IV, alpha 1	13: 110,801,310 - 110,959,496	1282	NM_001845
CTH	cystathionase (cystathionine gamma-lyase)	1: 70,876,901 - 70,905,534	1491	NM_001902
CUBN	cubilin (intrinsic factor-cobalamin receptor)	10: 16,865,965 - 17,171,816	8029	NM_001081
CYCS	cytochrome c, somatic	7: 25,158,270 - 25,164,980	54205	NM_018947
DCX	doublecortin	X: 110,537,007 - 110,655,460	1641	NM_000555
DHCR7	7-dehydrocholesterol reductase	11: 71,145,457 - 71,159,477	1717	NM_001360
DHFR	dihydrofolate reductase	5: 79,922,045 - 79,950,800	1719	NM_000791
DISC1	disrupted in schizophrenia 1	1: 231,762,561 - 232,177,018	27185	NM_018662
DISP1	dispatched homolog 1 (Drosophila)	1: 222,988,431 - 223,179,337	84976	NM_032890
DKK1	dickkopf WNT signaling pathway inhibitor 1	10: 54,074,041 - 54,077,417	22943	NM_012242
DLX5	distal-less homeobox 5	7: 96,649,702 - 96,654,143	1749	NM_005221
DNMT1	DNA (cytosine-5-)-methyltransferase 1	19: 10,244,022 - 10,305,755	1786	NM_001379
DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	2: 25,455,830 - 25,565,459	1788	NM_153759
DNMT3L	DNA (cytosine-5-)-methyltransferase 3-like	21: 45,666,222 - 45,682,099	29947	NM_013369
DVL1P1	dishevelled segment polarity protein 1 pseudogene 1	22: 19,240,644 - 19,240,992	8215	-
DVL2	dishevelled segment polarity protein 2	17: 7,128,661 - 7,137,863	1856	NM_004422
DVL3	dishevelled segment polarity protein 3	3: 183,873,284 - 183,891,314	1857	NM_004423
DYNC2H1	dynein, cytoplasmic 2, heavy chain 1	11: 102,980,160 - 103,350,591	79659	NM_001377
DYNC2LI1	dynein, cytoplasmic 2, light intermediate chain 1	2: 44,001,178 - 44,037,149	51626	NM_015522
DYRK1A	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A	21: 38,739,859 - 38,887,679	1859	NM_001396
ECT2	epithelial cell transforming sequence 2 oncogene	3: 172,468,475 - 172,539,264	1894	NM_018098
EFHC1	EF-hand domain (C-terminal) containing 1	6: 52,284,994 - 52,360,583	114327	NM_018100
EFNB1	ephrin-B1	X: 68,048,840 - 68,062,007	1947	NM_004429
EN1	engrailed homeobox 1	2: 119,599,747 - 119,605,759	2019	NM_001426
EN2	engrailed homeobox 2	7: 155,250,824 - 155,257,526	2020	NM_001427
EOMES	eomesodermin	3: 27,757,868 - 27,763,785	8320	NM_005442
EP300	E1A binding protein p300	22: 41,488,614 - 41,576,081	2033	NM_001429
EYA4	eyes absent homolog 4 (Drosophila)	6: 133,562,495 - 133,853,258	2070	NM_004100
FGF17	fibroblast growth factor 17	8: 21,900,428 - 21,906,320	8822	NM_003867

FGFR1	fibroblast growth factor receptor 1	8: 38,268,656 - 38,326,352	2260	NM_015850
FGFR2	fibroblast growth factor receptor 2	10: 123,237,844 - 123,357,972	2263	NM_000141
FGFR3	fibroblast growth factor receptor 3	4: 1,795,039 - 1,810,599	2261	NM_000142
FKRP	fukutin related protein	19: 47,249,303 - 47,261,832	79147	NM_024301
FKTN	fukutin	9: 108,320,411 - 108,403,399	2218	NM_006731
FLNA	filamin A, alpha	X: 153,576,900 - 153,603,006	2316	NM_001456
FOLH1	folate hydrolase (prostate-specific membrane antigen) 1	11: 49,168,187 - 49,230,222	2346	NM_004476
FOLR1	folate receptor 1 (adult)	11: 71,900,602 - 71,907,367	2348	NM_000802
FOLR2	folate receptor 2 (fetal)	11: 71,927,819 - 71,932,994	2350	NM_000803
FOLR3	folate receptor 3 (gamma)	11: 71,846,771 - 71,850,934	2352	NM_000804
FOXB1	forkhead box B1	15: 60,296,421 - 60,298,142	27023	NM_012182
FOXC1	forkhead box C1	6: 1,610,681 - 1,614,132	2296	NM_001453
FOXC2	forkhead box C2 (MFH-1, mesenchyme forkhead 1)	16: 86,600,857 - 86,602,539	2303	NM_005251
FOXD1	forkhead box D1	5: 72,742,083 - 72,744,352	2297	NM_004472
FOXH1	forkhead box H1	8: 145,699,115 - 145,701,718	8928	NM_003923
FPGS	folylpolyglutamate synthase	9: 130,565,154 - 130,576,556	2356	NM_004957
FREM1	FRAS1 related extracellular matrix 1	9: 14,734,664 - 14,910,993	158326	NM_144966
FUZ	fuzzy planar cell polarity protein	19: 50,310,123 - 50,316,567	80199	NM_025129
FZD3	frizzled family receptor 3	8: 28,351,722 - 28,431,785	7976	NM_017412
FZD6	frizzled family receptor 6	8: 104,310,661 - 104,345,094	8323	NM_003506
GABRA1	gamma-aminobutyric acid (GABA) A receptor, alpha 1	5: 161,274,197 - 161,326,965	2554	NM_000806
GABRB3	gamma-aminobutyric acid (GABA) A receptor, beta 3	15: 26,788,693 - 27,018,935	2562	NM_000814
GABRD	gamma-aminobutyric acid (GABA) A receptor, delta	1: 1,950,768 - 1,962,192	2563	NM_000815
GABRG2	gamma-aminobutyric acid (GABA) A receptor, gamma 2	5: 161,494,648 - 161,582,545	2566	NM_000816
GART	phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase	21: 34,876,238 - 34,915,198	2618	NM_000819
GAS1	growth arrest-specific 1	9: 89,559,277 - 89,562,104	2619	NM_002048
GJA1	gap junction protein, alpha 1, 43kDa	6: 121,756,745 - 121,770,873	2697	NM_000165
GJA5	gap junction protein, alpha 5, 40kDa	1: 147,228,332 - 147,245,484	2702	NM_005266
GLI2	GLI family zinc finger 2	2: 121,554,867 - 121,750,229	2736	NM_005270
GLI3	GLI family zinc finger 3	7: 42,000,547 - 42,276,618	2737	NM_000168
GPR56	G protein-coupled receptor 56	16: 57,653,910 - 57,698,944	9289	NM_005682
GPR98	G protein-coupled receptor 98	5: 89,854,617 - 90,460,033	84059	NM_032119
GRHL3	grainyhead-like 3 (Drosophila)	1: 24,645,812 - 24,690,972	57822	NM_021180
HESX1	HESX homeobox 1	3: 57,231,944 - 57,234,280	8820	NM_003865
HK1	hexokinase 1	10: 71,029,756 - 71,161,638	3098	NM_000188
HK2	hexokinase 2	2: 75,059,782 - 75,120,481	3099	NM_000189
HMX1	H6 family homeobox 1	4: 8,868,773 - 8,873,543	3166	NM_018942
HSPG2	heparan sulfate proteoglycan 2	1: 22,148,737 - 22,263,750	3339	NM_005529
HYLS1	hydrolethalus syndrome 1	11: 125,753,509 - 125,770,541	219844	NM_145014
IER3IP1	immediate early response 3 interacting protein 1	18: 44,681,390 - 44,702,745	51124	NM_016097
INPP5E	inositol polyphosphate-5-phosphatase, 72 kDa	9: 139,323,067 - 139,334,256	56623	NM_019892
KCNA1	potassium voltage-gated channel, shaker-related subfamily, member 1 (episodic ataxia with myokymia)	12: 5,019,073 - 5,027,422	3736	NM_000217
KCNAB1	potassium voltage-gated channel, shaker-related subfamily, beta member 1	3: 155,838,337 - 156,256,927	7881	NM_003471
KCNJ11	potassium inwardly-rectifying channel, subfamily J, member 11	11: 17,406,795 - 17,410,878	3767	NM_000525
KCNJ6	potassium inwardly-rectifying channel, subfamily J, member 6	21: 38,996,778 - 39,288,741	3763	NM_002240
KCNMA1	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	10: 78,629,359 - 79,397,577	3778	NM_002247
KCNQ2	potassium voltage-gated channel, KQT-like subfamily, member 2	20: 62,037,542 - 62,103,993	3785	NM_004518
KCNQ3	potassium voltage-gated channel, KQT-like subfamily, member 3	8: 133,133,105 - 133,493,004	3786	NM_004519
KIF7	kinesin family member 7	15: 90,171,201 - 90,198,682	374654	NM_198525
L1CAM	L1 cell adhesion molecule	X: 153,126,971 - 153,141,399	3897	NM_000425
LARGE	like-glycosyltransferase	22: 33,669,062 - 34,316,416	9215	NM_004737
LEP	leptin	7: 127,881,331 - 127,897,682	3952	NM_000230
LEPR	leptin receptor	1: 65,886,335 - 66,103,176	3953	NM_002303
LGI1	leucine-rich, glioma inactivated 1	10: 95,517,566 - 95,557,916	9211	NM_005097
LGI2	leucine-rich repeat LGI family, member 2	4: 25,000,471 - 25,032,414	55203	NM_018176
LRP2	low density lipoprotein receptor-related protein 2	2: 169,983,619 - 170,219,122	4036	NM_004525
MAT1A	methionine adenosyltransferase I, alpha	10: 82,031,576 - 82,049,434	4143	NM_000429
MAT2A	methionine adenosyltransferase II, alpha	2: 85,766,101 - 85,772,403	4144	NM_005911
MAT2B	methionine adenosyltransferase II, beta	5: 162,930,231 - 162,946,328	27430	NM_013283
MBD5	methyl-CpG binding domain protein 5	2: 148,778,580 - 149,271,046	55777	NM_018328
MCPH1	microcephalin 1	8: 6,264,113 - 6,501,140	79648	NM_024596
MECP2	methyl CpG binding protein 2 (Rett syndrome)	X: 153,287,264 - 153,363,188	4204	NM_004992
MED12	mediator complex subunit 12	X: 70,338,406 - 70,362,304	9968	NM_005120
MED17	mediator complex subunit 17	11: 93,517,405 - 93,546,496	9440	NM_004268
MEF2C	myocyte enhancer factor 2C	5: 88,014,058 - 88,199,922	4208	NM_002397
MKS1	Meckel syndrome, type 1	17: 56,282,797 - 56,296,966	54903	NM_017777
MSI1	musashi RNA-binding protein 1	12: 120,779,133 - 120,806,983	4440	NM_002442
MSX1	msh homeobox 1	4: 4,861,392 - 4,865,663	4487	NM_002448
MSX2	msh homeobox 2	5: 174,151,575 - 174,157,902	4488	NM_002449
MTHFD1	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methylenetetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase	14: 64,854,759 - 64,926,725	4522	NM_005956

MTHFD1L	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like	6: 151,186,815 - 151,423,023	25902	NM_015440
MTHFD2	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase	2: 74,425,690 - 74,442,425	10797	NM_006636
MTHFR	methylenetetrahydrofolate reductase (NAD(P)H)	1: 11,845,787 - 11,866,160	4524	NM_005957
MTHFS	5,10-methenyltetrahydrofolate synthetase (5-formyltetrahydrofolate cyclo-ligase)	15: 80,135,889 - 80,189,627	10588	NM_006441
MTR	5-methyltetrahydrofolate-homocysteine methyltransferase	1: 236,958,581 - 237,067,281	4548	NM_000254
MTRR	5-methyltetrahydrofolate-homocysteine methyltransferase reductase	5: 7,869,217 - 7,901,237	4552	NM_002454
MTTP	microsomal triglyceride transfer protein	4: 100,485,240 - 100,545,154	4547	NM_000253
NAT1	N-acetyltransferase 1 (arylamine N-acetyltransferase)	8: 18,027,971 - 18,081,198	9	NM_000662
NAT2	N-acetyltransferase 2 (arylamine N-acetyltransferase)	8: 18,248,755 - 18,258,723	10	NM_000015
NCAM1	neural cell adhesion molecule 1	11: 112,831,969 - 113,149,158	4684	NM_000615
NDE1	nudE neurodevelopment protein 1	16: 15,737,124 - 15,820,210	54820	NM_017668
NDEL1	nudE neurodevelopment protein 1-like 1	17: 8,339,170 - 8,371,495	81565	NM_030808
NDST1	N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 1	5: 149,887,674 - 149,937,773	3340	NM_001543
NEUROG2	neurogenin 2	4: 113,434,672 - 113,437,328	63973	NM_024019
NF1	neurofibromin 1	17: 29,421,945 - 29,704,695	4763	NM_000267
NHEJ1	nonhomologous end-joining factor 1	2: 219,940,046 - 220,025,587	79840	NM_024782
NLGN3	neuroligin 3	X: 70,364,681 - 70,391,051	54413	NM_018977
NLGN4X	neuroligin 4, X-linked	X: 5,808,083 - 6,146,706	57502	NM_020742
NOG	noggin	17: 54,671,060 - 54,672,951	9241	NM_005450
NOS1	nitric oxide synthase 1 (neuronal)	12: 117,645,947 - 117,799,607	4842	NM_000620
NOS2	nitric oxide synthase 2, inducible	17: 26,083,792 - 26,127,555	4843	NM_000625
NOS3	nitric oxide synthase 3 (endothelial cell)	7: 150,688,144 - 150,711,687	4846	NM_000603
NOTCH3	notch 3	19: 15,270,444 - 15,311,792	4854	NM_000435
NPHP1	nephronophthisis 1 (juvenile)	2: 110,880,913 - 110,962,639	4867	NM_000272
NRG2	neuregulin 2	5: 139,226,364 - 139,422,884	9542	NM_004883
NRXN1	neurexin 1	2: 50,145,643 - 51,259,674	9378	NM_004801
NUMB	numb homolog (Drosophila)	14: 73,741,918 - 73,925,286	8650	NM_003744
OCN	occludin	5: 68,788,119 - 68,853,931	100506658	NM_002538
OFD1	oral-facial-digital syndrome 1	X: 13,752,832 - 13,787,480	8481	NM_003611
OPHN1	oligophrenin 1	X: 67,262,186 - 67,653,299	4983	NM_002547
PAFAH1B1	platelet-activating factor acetylhydrolase 1b, regulatory subunit 1 (45kDa)	17: 2,496,923 - 2,588,909	5048	NM_000430
PAX1	paired box 1	20: 21,686,297 - 21,699,124	5075	NM_006192
PAX2	paired box 2	10: 102,505,468 - 102,589,698	5076	NM_000278
PAX3	paired box 3	2: 223,064,606 - 223,163,715	5077	NM_000438
PAX6	paired box 6	11: 31,806,340 - 31,839,509	5080	NM_000280
PAX7	paired box 7	1: 18,957,500 - 19,075,360	5081	NM_002584
PCMT1	protein-L-isoaspartate (D-aspartate) O-methyltransferase	6: 150,070,831 - 150,132,557	5110	NM_005389
PCNT	pericentrin	21: 47,744,036 - 47,865,682	5116	NM_006031
PFN1	profilin 1	17: 4,848,945 - 4,852,381	5216	NM_005022
PIPSK1C	phosphatidylinositol-4-phosphate 5-kinase, type I, gamma	19: 3,630,179 - 3,700,477	23396	NM_012398
PKD1	polycystic kidney disease 1 (autosomal dominant)	16: 2,138,711 - 2,185,899	5310	NM_000296
PMM2	phosphomannomutase 2	16: 8,891,670 - 8,943,194	5373	NM_000303
PNKP	polynucleotide kinase 3'-phosphatase	19: 50,364,460 - 50,370,822	11284	NM_007254
POLL	polymerase (DNA directed), lambda	10: 103,338,639 - 103,348,027	27343	NM_013274
POMGNT1	protein O-linked mannose N-acetylglucosaminyltransferase 1 (beta 1,2-)	1: 46,654,353 - 46,685,977	55624	NM_017739
POMT1	protein-O-mannosyltransferase 1	9: 134,378,289 - 134,399,193	10585	NM_007171
POMT2	protein-O-mannosyltransferase 2	14: 77,741,299 - 77,787,225	29954	NM_013382
POR	P450 (cytochrome) oxidoreductase	7: 75,544,420 - 75,616,173	5447	NM_000941
PRMT1	protein arginine methyltransferase 1	19: 50,180,409 - 50,191,707	3276	NM_001536
PRMT2	protein arginine methyltransferase 2	21: 48,055,507 - 48,085,155	3275	NM_001535
PTCH1	patched 1	9: 98,205,264 - 98,279,247	5727	NM_000264
PTK7	protein tyrosine kinase 7	6: 43,044,006 - 43,129,458	5754	NM_002821
PTPN11	protein tyrosine phosphatase, non-receptor type 11	12: 112,856,536 - 112,947,717	5781	NM_002834
PURA	purine-rich element binding protein A	5: 139,493,708 - 139,499,001	5813	NM_005859
RAB23	RAB23, member RAS oncogene family	6: 57,053,581 - 57,087,078	51715	NM_016277
RAB3GAP1	RAB3 GTPase activating protein subunit 1 (catalytic)	2: 135,809,835 - 135,928,280	22930	NM_012233
RARA	retinoic acid receptor, alpha	17: 38,465,423 - 38,513,895	5914	NM_000964
RARS2	arginyl-tRNA synthetase 2, mitochondrial	6: 88,224,096 - 88,299,735	57038	NM_020320
RELN	reelin	7: 103,112,231 - 103,629,963	5649	NM_005045
RNF2	ring finger protein 2	1: 185,014,551 - 185,071,740	6045	NM_007212
RNMT	RNA (guanine-7-) methyltransferase	18: 13,726,704 - 13,764,554	8731	NM_003799
RPGRIP1L	RPGRIP1-like	16: 53,633,818 - 53,737,771	23322	NM_015272
RPS6KA3	ribosomal protein S6 kinase, 90kDa, polypeptide 3	X: 20,168,029 - 20,284,750	6197	NM_004586
SALL1	sal-like 1 (Drosophila)	16: 51,169,886 - 51,185,183	6299	NM_002968
SALL4	sal-like 4 (Drosophila)	20: 50,400,581 - 50,419,048	57167	NM_020436
SARDH	sarcosine dehydrogenase	9: 136,528,682 - 136,605,077	1757	NM_007101
SCN1A	sodium channel, voltage-gated, type I, alpha subunit	2: 166,845,670 - 167,005,642	6323	NM_006920
SCN1B	sodium channel, voltage-gated, type I, beta subunit	19: 35,521,592 - 35,531,353	6324	NM_001037
SCN2A	sodium channel, voltage-gated, type II, alpha subunit	2: 166,095,912 - 166,248,820	6326	NM_021007
SCRIB	scribbled planar cell polarity protein	8: 144,873,090 - 144,897,549	23513	NM_015356
SEPSECS	Sep (O-phosphoserine) tRNA:Sec (selenocysteine) tRNA synthase	4: 25,121,627 - 25,162,204	51091	NM_016955
SHANK3	SH3 and multiple ankyrin repeat domains 3	22: 51,113,070 - 51,171,640	85358	NM_033517

SHH	sonic hedgehog	7: 155,595,558 - 155,604,967	6469	NM_000193
SHMT1	serine hydroxymethyltransferase 1 (soluble)	17: 18,231,187 - 18,266,856	6470	NM_004169
SIX3	SIX homeobox 3	2: 45,169,037 - 45,173,216	6496	NM_005413
SLC12A6	solute carrier family 12 (potassium/chloride transporter), member 6	15: 34,522,197 - 34,630,265	9990	NM_005135
SLC16A2	solute carrier family 16, member 2 (thyroid hormone transporter)	X: 73,641,328 - 73,753,764	6567	NM_006517
SLC19A1	solute carrier family 19 (folate transporter), member 1	21: 46,934,629 - 46,962,385	6573	NM_194255
SLC25A19	solute carrier family 25 (mitochondrial thiamine pyrophosphate carrier), member 19	17: 73,269,061 - 73,285,530	60386	NM_021734
SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1	1: 43,391,046 - 43,424,847	6513	NM_006516
SLC40A1	solute carrier family 40 (iron-regulated transporter), member 1	2: 190,425,316 - 190,445,537	30061	NM_014585
SLC9A6	solute carrier family 9, subfamily A (NHE6, cation proton antiporter 6), member 6	X: 135,067,583 - 135,129,428	10479	NM_006359
SNAP29	synaptosomal-associated protein, 29kDa	22: 21,213,292 - 21,245,502	9342	NM_004782
SOCS7	suppressor of cytokine signaling 7	17: 36,508,007 - 36,561,846	30837	NM_014598
SOX2	SRY (sex determining region Y)-box 2	3: 181,429,712 - 181,432,224	6657	NM_003106
SPINT2	serine peptidase inhibitor, Kunitz type, 2	19: 38,755,098 - 38,783,254	10653	NM_021102
SPTBN4	spectrin, beta, non-erythrocytic 4	19: 40,973,126 - 41,082,365	57731	NM_020971
SRGAP2	SLIT-ROBO Rho GTPase activating protein 2	1: 206,516,197 - 206,637,783	23380	NM_015326
SRPX2	sushi-repeat containing protein, X-linked 2	X: 99,899,163 - 99,926,296	27286	NM_014467
STIL	SCL/TAL1 interrupting locus	1: 47,715,811 - 47,779,819	6491	NM_003035
T	T, brachyury homolog (mouse)	6: 166,571,144 - 166,582,157	6862	NM_003181
TBCE	tubulin folding cofactor E	1: 235,530,728 - 235,612,280	6905	NM_003193
TBR1	T-box, brain, 1	2: 162,272,620 - 162,281,573	10716	NM_006593
TCF4	transcription factor 4	18: 52,889,562 - 53,303,188	6925	NM_003199
TCTN2	tectonic family member 2	12: 124,155,660 - 124,192,950	79867	NM_024809
TDGF1	teratocarcinoma-derived growth factor 1	3: 46,616,045 - 46,623,953	6997	NM_003212
TGIF1	TGFB-induced factor homeobox 1	18: 3,412,072 - 3,458,409	7050	NM_003244
TMEM216	transmembrane protein 216	11: 61,159,832 - 61,166,335	51259	NM_016499
TMEM67	transmembrane protein 67	8: 94,767,072 - 94,831,462	91147	NM_153704
TP53	tumor protein p53	17: 7,571,720 - 7,590,868	7157	NM_000546
TP73	tumor protein p73	1: 3,569,129 - 3,652,765	7161	NM_005427
TRPM6	transient receptor potential cation channel, subfamily M, member 6	9: 77,337,411 - 77,503,010	140803	NM_017662
TSC1	tuberous sclerosis 1	9: 135,766,735 - 135,820,020	7248	NM_000368
TSC2	tuberous sclerosis 2	16: 2,097,990 - 2,138,713	7249	NM_000548
TSEN2	TSEN2 tRNA splicing endonuclease subunit	3: 12,525,931 - 12,574,820	80746	NM_025265
TSEN34	TSEN34 tRNA splicing endonuclease subunit	19: 54,694,119 - 54,698,394	79042	NM_024075
TSEN54	TSEN54 tRNA splicing endonuclease subunit	17: 73,512,609 - 73,520,820	283989	NM_207346
TUBA1A	tubulin, alpha 1a	12: 49,578,578 - 49,583,107	7846	NM_006009
TUBA1B	tubulin, alpha 1b	12: 49,521,565 - 49,525,304	10376	NM_006082
TUBA1C	tubulin, alpha 1c	12: 49,658,865 - 49,667,113	84790	NM_032704
TUBA3C/TUBA3D	tubulin, alpha 3c	13: 19,747,910 - 19,755,992	7278	NM_006001
TUBA3C/TUBA3D	tubulin, alpha 3c	2: 132,233,580 - 132,240,507	113457	NM_080386
TUBA3E	tubulin, alpha 3e	2: 130,949,318 - 130,956,034	112714	NM_207312
TUBA4A	tubulin, alpha 4a	2: 220,115,001 - 220,118,638	7277	NM_006000
TUBA8	tubulin, alpha 8	22: 18,593,453 - 18,614,498	51807	NM_018943
TUBB	tubulin, beta class I	6: 30,688,157 - 30,693,199	203068	NM_178014
TUBB1	tubulin, beta 1 class VI	20: 57,594,309 - 57,601,709	81027	NM_030773
TUBB2A	tubulin, beta 2A class IIa	6: 3,153,900 - 3,157,783	7280	NM_001069
TUBB2B	tubulin, beta 2B class IIb	6: 3,224,495 - 3,227,968	347733	NM_178012
TUBB4B	tubulin, beta 4B class IVb	9: 140,135,711 - 140,138,159	10383	NM_006088
TUBB3	tubulin, beta 3 class III	16: 89,988,417 - 90,002,505	10381	NM_006086
TUBB4A	tubulin, beta 4A class IVa	19: 6,494,330 - 6,502,595	10382	NM_006087
TUBB6	tubulin, beta 6 class V	18: 12,308,257 - 12,326,568	84617	NM_032525
TUBB8	tubulin, beta 8 class VIII	10: 92,828 - 95,178	347688	NM_177987
TWIST1	twist basic helix-loop-helix transcription factor 1	7: 19,155,091 - 19,157,295	7291	NM_000474
TWIST2	twist basic helix-loop-helix transcription factor 2	2: 239,756,673 - 239,795,893	117581	NM_057179
UBE3A	ubiquitin protein ligase E3A	15: 25,582,396 - 25,684,175	7337	NM_000462
VANGL1	VANGL planar cell polarity protein 1	1: 116,184,574 - 116,240,845	81839	NM_138959
VANGL2	VANGL planar cell polarity protein 2	1: 160,370,364 - 160,398,468	57216	NM_020335
VAX1	ventral anterior homeobox 1	10: 118,888,032 - 118,897,812	11023	NM_199131
VLDLR	very low density lipoprotein receptor	9: 2,621,793 - 2,654,485	7436	NM_003383
VRK1	vaccinia related kinase 1	14: 97,263,684 - 97,347,951	7443	NM_003384
WDR62	WD repeat domain 62	19: 36,545,783 - 36,596,012	284403	NM_173636
WNT1	wingless-type MMTV integration site family, member 1	12: 49,372,236 - 49,376,396	7471	NM_005430
WNT11	wingless-type MMTV integration site family, member 11	11: 75,897,370 - 75,917,574	7481	NM_004626
WNT3A	wingless-type MMTV integration site family, member 3A	1: 228,194,723 - 228,248,972	89780	NM_033131
WNT5A	wingless-type MMTV integration site family, member 5A	3: 55,499,743 - 55,521,670	7474	NM_003392
YWHAE	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide	17: 1,247,834 - 1,303,556	7531	NM_006761
ZIC1	Zic family member 1	3: 147,127,181 - 147,134,506	7545	NM_003412
ZIC2	Zic family member 2	13: 100,634,026 - 100,639,019	7546	NM_007129
ZIC3	Zic family member 3	X: 136,648,346 - 136,654,259	7547	NM_003413
ZIC5	Zic family member 5	13: 100,615,275 - 100,624,178	85416	NM_033132
ZNF525	zinc finger protein 525	19: 53,868,968 - 53,889,843	170958	NR_003699

* 染色体上の位置は GRCh37/hg19 の build バージョンに基づく。

表 2. HapMap データと比較した際の我々の神経系カスタムリシーケンス手法での NGS call の精度検証

	HapMap のcall	我々のNGS call	Sangerシーケンス前	Sangerシーケンス後
我々の NGS と HapMap の双方でジェノタイプが 得られた箇所	Ref-Homo	All	2,172	-
	Hetero	All	463	-
	No-Ref-Homo	All	494	-
	Total		3,129	-
False positive (偽陽性)	Ref-Homo	No-Ref-Homo	0	0
	Ref-Homo	Hetero	2	0
	No-Ref-Homo	Hetero	1	0
	No-Ref-Homo	No-Ref-Homo (塩基の異なる Homo)	0	0
	Hetero	Hetero (塩基の異なる Hetero)	0	0
	Hetero	No-Ref-Homo	2	0
	Total		5	0
False negative (偽陰性)	Hetero	Ref-Homo	2	0
	No-Ref-Homo	Ref-Homo	9	0
	Total		11	0

No-Ref-Homo, リファレンス配列とは異なるhomozygote; Ref-Homo, リファレンス配列と同一のhomozygote.

表 3. 神経系カスタムリシーケンスの NGS call で HapMap データと不一致だった 16 箇所の Sanger シーケンスによる検証

ID	HapMap	NGS call	Chr	Chr coordinate	SNP rsID	Reference	HapMap	Our NGS	Sanger	NGS	L primer	R primer	PCR product size	SNV position in the PCR product
							genotype	call	sequence	read depth				
MisMatch_1	Ref-Homo	Hetero	9	139328369	rs10781543	T	T	Y	Y	23	TTCTACTCTCGCACGCCAAC	TGCCCAGGACCTGTATGTCA	408	241
MisMatch_2	Ref-Homo	Hetero	17	7579472	rs1042522	G	G	S	S	47	GACTTGGCTGTCCCAGAATG	TCCCAAGCAATGGATGATTT	254	148
MisMatch_3	No-Ref-Homo	Hetero	8	144887734	rs11783725	G	A	R	R	13	CAGAGCCACACCATTACCT	CAGTGTGTGTCTGGCCCTA	258	147
MisMatch_4	Hetero	No-Ref-Homo	6	133827354	rs2277083	A	R	G	G	369	TCTTCCACCACAGCAATATGT	TGCTTCTCTTGAGCTGAATTTTT	297	201
MisMatch_5	Hetero	No-Ref-Homo	9	123152241	rs2297456	A	M	C	C	31	GAGAGCTGTGTGAAAGCCAAG	GTGCACAGCTGTAGGGTGTTC	251	129
MisMatch_6	Hetero	Ref-Homo	19	41034912	rs1461008	G	R	G	G	211	GCTTTGCCCACTAATTCCT	CACCTGGAGTTCCTACCTCGT	253	111
MisMatch_7	Hetero	Ref-Homo	20	62038757	rs3746364	C	Y	C	C	17	CATACAGCCAGACAGCTCCA	TACCTGCTACCACCCTCTC	342	284
MisMatch_8	No-Ref-Homo	Ref-Homo	1	22176959	rs2290499	C	T	C	C	39	GATCCTGCCGTTGCAAGAGT	GGTTATAGGAACAGGGGTCTCC	267	177
MisMatch_9	No-Ref-Homo	Ref-Homo	1	206574659	rs1754497	C	G	C	C	19	GCCCTCTAGGGTGTGTTAACT	CAGAGAGGAAGGTGCGTAGA	266	85
MisMatch_10	No-Ref-Homo	Ref-Homo	3	12573336	rs299641	A	G	A	A	51	GGAACATGTGGTCCCAAGAA	CCTCTGGTCTCTGGGAATGA	263	119
MisMatch_11	No-Ref-Homo	Ref-Homo	4	4861780	rs10213286	C	T	C	C	11	ACATGACTTCTTTGCCACTCG	CTTCCTGTGGTGGCCATGA	200	138
MisMatch_12	No-Ref-Homo	Ref-Homo	6	151413828	rs2295734	C	A	C	C	59	GTGGACAAGGCTCTCACAGG	CGCAGGACACTAAGGGTTGT	255	136
MisMatch_13	No-Ref-Homo	Ref-Homo	7	103206013	rs2711839	T	A	T	T	143	CTGGAGCTGTACTACTGTGGGAGT	CAAAGCACATAGGGGGCACT	250	134
MisMatch_14	No-Ref-Homo	Ref-Homo	7	103244981	rs362681	A	G	A	A	74	ACTGCTGCCAATGTAATG	GCATTGGACAATAATTGAGTGG	250	140
MisMatch_15	No-Ref-Homo	Ref-Homo	16	2143849	rs2854581	C	G	C	C	45	AAGTACAGGGCTCCAGCAA	GTCTCTGCTTCCCAGGACT	314	133
MisMatch_16	No-Ref-Homo	Ref-Homo	16	2162979	rs2855349	G	A	G	G	19	GCTCCTGTGCACCCAGTTAC	GAGGGAGGGACGCCAATC	390	335

Chr, 染色体番号 ; No-Ref-Homo, リファレンス配列と不一致の homozygous; Ref-Homo, リファレンス配列の homozygous.

* 染色体の位置はリファレンスゲノムのバージョン GRCh37/hg19 を用い、塩基の表記は正鎖方向の塩基情報に基づく

表 4. 神経系遺伝子ターゲットリシーケンスでの疾患原因変異同定率

	検体数 (家系数)	割合 (%)
既知あるいは遺伝子機能から確定的な原因変異が同定できた	14	11.8%
確定的ではないが既知の遺伝子機能から原因変異である可能性が高いと推測される変異が見つかった	3	2.5%
疾患原因かどうかは不明だが何かしらの候補変異が最終的に残った	47	39.5%
一つも疾患原因候補変異が見つからなかった	55	46.2%
総数	119	100.0%

表 5. これまでに報告されているてんかん性攣縮があり CASK 遺伝子に変異を持つ患者

患者 No.	性別	表現型 *	てんかん性攣縮発作年齢	CASK 変異	報告論文
1	男	ID、MICPCH	生後 4 ヶ月	c.1061T>C (p.L348P)	引用文献 60、患者 No.16
2	男	ID、MICPCH、多発奇形	生後 2 ヶ月	exon 2 を含む 111 kb 欠失	引用文献 59、患者 No.1
3	男	ID、MICPCH、多発奇形	生後 3 ヶ月	c.1A>G	引用文献 59、患者 No.2
4	男	ID、MICPCH	NA	c.278+1G>A	引用文献 58、患者 No.13
5	女	ID、MICPCH	生後 17 ヶ月	c.316C>G (p.R28*)	引用文献 60、患者 No.3
6	女	ID、MICPCH	生後 34 ヶ月	c.243_244delTA (p.Y81*)	引用文献 60、患者 No.10
7	女	ID、大脳・小脳萎縮、 舞踏アテトーゼ	生後 3 年	c.82C>T (p.R28*)	引用文献 61、患者 No.14
8	女	ID、MICPCH	乳児期 (詳細不明)	c.2074C>T (p.Q692*)	引用文献 58、患者 No.7
9	女	ID、MICPCH	生後 3 年 8 ヶ月	c.1896dupC (p.C633fs*2)	引用文献 62 (本論文の報告患者)

*ID, 知能障害; MICPCH, 小頭症と橋および小脳低形成

表 6. WES データと HumanOmniExpress Exome SNP chip データとの一致率

サンプル ID*	NGS でリファレンスと異なる homozygote の call が得られた箇所	NGS で heterozygote の call が 得られた箇所	NGS でリファレンスと一致する homozygote の call が得られた箇所	合計
NA18943 (V4)	99.92% (7 mismatch / 8,911)	99.44% (64 mismatch / 11,500)	99.97% (69 mismatch / 223,162)	99.94% (140 mismatch / 243,573)
NA18943 (V5)	99.91% (8 mismatch / 9,263)	99.85% (18 mismatch / 11,710)	99.97% (73 mismatch / 229,963)	99.96% (99 mismatch / 250,936)
OFB35 (V4)	99.98% (2 mismatch / 9,133)	99.54% (53 mismatch / 11,474)	99.97% (72 mismatch / 224,092)	99.95% (127 mismatch / 244,699)
NCU2_6 (V4)	99.97% (3 mismatch / 8,840)	99.52% (52 mismatch / 10,852)	99.97% (73 mismatch / 222,167)	99.95% (128 mismatch / 241,859)
NCU2_77 (V4)	99.99% (1 mismatch / 8,906)	99.94% (7 mismatch / 11,498)	99.97% (71 mismatch / 225,507)	99.97% (79 mismatch / 245,911)
K3423 (V5)	99.99% (1 mismatch / 8,697)	99.77% (26 mismatch / 11,228)	99.97% (72 mismatch / 226,575)	99.96% (99 mismatch / 246,500)
I-042-2 (V4)	99.97% (3 mismatch / 8,850)	99.46% (60 mismatch / 11,109)	99.97% (68 mismatch / 222,755)	99.95% (131 mismatch / 242,714)
I-057 (V4)	99.93% (6 mismatch / 8,752)	99.54% (51 mismatch / 11,174)	99.96% (82 mismatch / 221,468)	99.94% (139 mismatch / 241,394)
平均	99.96%	99.63%	99.97%	99.95%

* サンプル ID 名の括弧表記の V4 または V5 は WES での capture probe のプラットフォームである SureSelect V4 または V5 を示す

表 7. WES と SNP chip で不一致だった 140 箇所の検証データ

Sanger 検証用 ID	Chr	Coordinate	Ref.	SNP chip call	我々の NGS call	Sanger シー ケンス結果	Sanger 用の PCR L 側 primer 配列	Sanger 用の PCR R 側 primer 配列
SangerVali_1	1	12776344	A	A	W	W	CACTCTGTGGGTCATTTGCAG	AACACCCAAAGAGGCTGAGAA
SangerVali_2	1	15850613	G	R	G	-	-	-
SangerVali_3	1	19447752	G	G	K	G	TTTGGGCTTTAAGTCCCCTCT	CTTGACGCCGTCTGTACTTC
SangerVali_4	1	22186113	T	T	K	T	ATGCACCATCTGCCTGAATCT	GGGCTGGATGTGAGAAAGAGT
SangerVali_5	1	38262417	T	T	K	T	TGAACTCAGCCTCTCCTCCAG	CTGCCCCCAAACCTTCTAT
SangerVali_6	1	53158524	A	A	M	A	AGCTTATCATGGGGCTCAGGT	AGCCCCAGACATGAAGAACAA
SangerVali_7	1	110233147	C	Y	T	T	GGAGGAGTGATATGGGGAATGA	GTTCTGTGGCATGAAACCAGTA
SangerVali_8	1	116226631	C	C	M	C	AGGGGGTGGGTAGACAACACT	AAGAGGGTACTGGCATCAGGA
SangerVali_9	1	144874815	T	C	Y	Y	ACCCAGAAAGTGGAGGCTACC	TGACCTTGGTTTTACCCGTTG
SangerVali_10	1	152128689	C	T	C	C	CTTTTCTGTCCATTTGACCA	TCGGTGTGAAAAACCAATTCAGG
SangerVali_11	1	155209756	G	C	G	G	GTATCAGTACCCAGCGGGAAA	TACAGCTCGTGGTGTGTGTC
SangerVali_12	1	155209795	G	C	G	G	GTTACCTGTGCCCGTGTGATT	GGGGTGTAGGAATTTTAAAACC
SangerVali_13	1	202399880	C	Y	C	C	GGTCCCCACTACCTCTTTGG	CAGAGCAAAGCTCATCCCATT
SangerVali_14	1	208252715	A	A	M	A	TGCAGGACTTACCTCGGTGAT	ATCCATGGTGGGGAAGAGTTT
SangerVali_15	1	208272313	A	A	M	A	TACTTACCAACCGGCTGTGCT	AGGCCCTCCGGAGTACAGTAA
SangerVali_16	1	218480973	A	G	R	R	TGTTTTGAGAGGGATGGATGG	GCAAATGGCAGGATTTTCAGA
SangerVali_17	1	227920447	C	Y	T	Y	AGGACCAGGAGCCTCTTCTCA	CTTTTAGCTGCATGGCTCAG
SangerVali_18	2	27352795	G	R	K	K	TCATGGTCCCTGTCCAATTC	CAAGCACTAGGAGAGCCTGGA
SangerVali_19	2	28800894	C	Y	C	Y	AGAATCACCCCTCCTACGTG	CCACCGTGTATTCTTCATCA
SangerVali_20	2	46707886	G	R	G	R	CGAGATGGAGCTGGAGAAGGT	AGTGGATGCTCAGCCCCTTT
SangerVali_21	2	95537544	A	G	A	A	CTGCCCTGCAAAGAGTACGAC	GATGGAGAAGGGCACCTCTGT
SangerVali_22	2	121747688	G	T	G	G	TACAGCAGGCTTTGGCCTAGT	ACATGTGGATCTGGCCGTAGT
SangerVali_23	2	128051291	T	C	T	T	GCATTTGCAGCCTGTAGTCCT	GGGATGCAGGTGGCTCTTAG
SangerVali_24	2	179634421	T	T	K	T	CCCACCAATCAAGGATTTTT	GGCTGAAGAATGGTGTGGAAA
SangerVali_25	2	206305364	C	C	M	C	CCCCAATTATGCCAGAGTGAA	GGTCAAAGGATGGCACTTGTG
SangerVali_26	3	41278119	C	C	M	C	TTCCTCAAGGGCCTTTTTCTC	TCAAGCCAGGAAACATCAAT
SangerVali_27	3	43389767	G	G	K	G	GGCAACCCACCAATACATC	AAAAACTTATGCCGCCCACT
SangerVali_28	3	97868931	G	R	G	G	GCAGGTTCAATTCAAGTTTTACC	TCTCAGGCTGTAGATCATGGGATT
SangerVali_29	3	129303386	G	A	G	G	TTGCAGTAGGCAATCTGGTGA	GGGTCACATGTAGGGCATAGC
SangerVali_30	3	195925183	T	T	W	W	AACGCCCTCCAGTGTAGAAC	GGTACATGGCTGAAGCTGTCC
SangerVali_31	4	3230343	T	T	K	T	GTAACCTGGGGTGTCTGAACG	ACCTGGAGTTGACTGGAGACG
SangerVali_32	4	48559517	G	G	K	G	TCACAGGATTATGGAACCTTTCA	AATCCAGACAGCTCACCTGCT
SangerVali_33	4	187004217	C	A	T	T	TGGTTGGCCACCTAGAAGTA	TGTTGGCTATGTTGTTGTTGCT
SangerVali_34	5	33937593	C	T	Y	Y	GGTGGCCTCGGCTCTGAAG	ACCAGCAGCAGGTAGCACAAAGA
SangerVali_35	5	38950153	G	G	K	G	TGTAGCATAGCCAAAAGCATCTC	TTCATGAGCTCAAGTCCTTCACG
SangerVali_36	5	112824045	C	T	C	C	AGGGACTGATCTCGCTCCTGT	GCGTACTCCTCCTCCAAGTA
SangerVali_37	6	10756728	C	T	C	C	CCCCAGGGATGGAATTTTA	GTCAGATGAAAAATGTCTGCAC
SangerVali_38	6	31383071	A	R	A	A	AATTGAATCCCTGCCTGGAT	CCTGGCCTGAGACTCTGTCTT
SangerVali_39	6	32713693	T	K	T	T	GCTGTTTCTCACACAGAGG	TTTCAGAAGAGGCTCGTCCAG
SangerVali_40	6	41774685	C	C	S	C	AACACGTAGAGATCCCGGACTT	ACTCTGCTGCCAGGTTGAG
SangerVali_41	6	44221316	G	R	G	G	TGAAGATGAAGTGGCAGCAGA	ATGGGGCTTGACACCCTTAGT
SangerVali_42	6	48036327	A	A	M	A	GCACGGTGGGAAAAAGACC	AAAGACCAGGCGGAGTGGTT
SangerVali_43	6	76376551	C	Y	C	C	ATCTCCTCAGCTGCTGATTC	TGGCACTACCAAACCTGCCTTC
SangerVali_44	6	137814707	C	A	C	C	CCATGTTGGCTGTGGAGAGAG	GGAGGAGATGAAGAGGCTGGT
SangerVali_45	6	148835416	C	Y	C	Y	GAGGGCAAGGATGTGTGTTTC	TTCTTTCTGACTCTGCGAGTGG
SangerVali_46	7	20198700	G	G	K	G	GGAGGCTCCACTTGAACACAA	CACAAAACCCTCAATTGGAA
SangerVali_47	7	47408712	T	T	K	T	GACACTGAGGGCTTTCTCAGC	AATGGAGACGCTGCTCTGAAG
SangerVali_48	7	55004546	C	Y	T	T	CCAACACCTAGTGACCATTCCA	TGTATCAATGGGTTCCGCATA
SangerVali_49	7	73771749	C	C	M	C	GTGCTTACAGTCCCTGAGACC	GGGGACAGGTAAGCTGTGT
SangerVali_50	7	82784501	G	A	G	G	GCTGCAGAGTTTTCCAGATCC	TAAGGCACCTGGGCCTACAAAG

(表 7 の続き)

Sanger 検証用 ID	Chr	Coordinate	Ref.	SNP chip call	我々の NGS call	Sanger シー ケンス結果	Sanger 用の PCR L 側 primer 配列	Sanger 用の PCR R 側 primer 配列
SangerVali_51	7	96810439	A	M	A	A	TTTAGGTGTATGCAACAGCGTTAT	GGGGTTCTAAATTTTGCATGTC
SangerVali_52	7	129680877	A	A	M	A	GCTGGTTGTAAGTGGCACAGAGA	CTGGGAGACAGAGTGAGAAGTTT
SangerVali_53	7	131128350	G	G	K	G	TCAATCTGTTGCCTTGTCTGA	CGCAATGTACGGAATGTTCTT
SangerVali_54	8	37691330	T	G	K	K	AAGGAACAGGCGAGATGATCC	GAGAGGCTTGAGCAAATGAGG
SangerVali_55	9	463655	C	C	M	C	GGCAGACCAGAGGGAATATCA	CTCCACCCTTTCGAGGTTCT
SangerVali_56	9	35811508	A	R	A	-	-	-
SangerVali_57	9	77403574	C	C	M	C	TCATCACCAGAGCCAAGGAGT	GTTTCTGCTCCATCTGCGTCT
SangerVali_58	9	86474262	T	Y	T	T	TCCACATTGTGTCTGCGTTT	AATGAATACCCCTCCCTCAA
SangerVali_59	9	96428334	G	A	R	R	GGGACTTGTCTGTGAGTTGG	TGCCATCTCATGCAGAAACAC
SangerVali_60	9	133325472	T	Y	T	Y	TGTTGGCACATTACCCATCTG	TCCAGAGGCCCTAAACCTTA
SangerVali_61	9	138836946	C	M	C	C	CACCTCAAGGGAAGACACAGC	CGTTAGTCTGGGGATGTTGGA
SangerVali_62	9	139413211	T	T	K	T	ATGGGGACACTCGCAGTAGAA	TTAGGGGACAGGGAGCTCAG
SangerVali_63	10	102057362	A	A	M	A	GAAGTCGATGAACACCACTCG	CCTGTAAAGCCCAGACAATGA
SangerVali_64	10	117075090	T	T	K	T	GTTCTGTACGCAGCTCAAAA	AAGCAGAAACTGAAATTGTTCCA
SangerVali_65	10	135203135	C	T	C	C	GAGCCTCTCCTGACACACTCTGC	CTGAGCCAGCAGGTCCAGGT
SangerVali_66	11	373404	G	R	G	-	-	-
SangerVali_67	11	6806827	G	K	G	G	ATGTGGCCATTTGTCATCCTC	GGAAGGTCACACCCATCACAT
SangerVali_68	11	12225829	G	G	K	C	ATTGGGTGGATGAATGACAGG	GATGGAGCCAGCACAGAAGTT
SangerVali_69	11	18267005	G	A	G	G	CCTCATAGCCAGGTCTCCTGA	ACTCCTCCTTCTGGCCCTT
SangerVali_70	11	26718732	C	C	M	C	TGTTGTGTACCAATTAATGAGCAA	ATTGCTAAACAGCTATGTTCCACA
SangerVali_71	11	56467945	G	K	G	G	CACGAAACACTGCAAATGTGA	GATCAACACGATGAGGGTGTCT
SangerVali_72	11	56468440	G	K	G	G	AAGCTGTGTGCATTGCTGGTA	AAGGCTTTGAGGTAGCCCTTG
SangerVali_73	11	56468694	C	T	C	C	GATCTCCTCCTCAAGGGCTA	ACGTGTCAAGAAGGGGACTGA
SangerVali_74	11	65413817	A	C	A	A	TTTGGGGCACAGAAGACCTCAT	GGGGAGGGAGGCGTTAGTGA
SangerVali_75	11	82984713	T	Y	T	T	GTTCCCAAAAGTGCTGGTTA	AAGCATGAAACCAGTCCAATCA
SangerVali_76	11	82991264	C	Y	C	C	TTTCCAAGTCTGAACCAATGC	AAAAGAGAGGCTCAGAGGAGTCAA
SangerVali_77	11	93540717	G	A	G	G	GGGCAACAGGCTTTGAGACTT	CTATGCGAAGAAGGGGGCTTA
SangerVali_78	11	126146290	T	T	K	-	-	-
SangerVali_79	12	9243846	G	R	G	G	AAGAGATCCTTACCCGGATGC	ACACTCCTCTCCAGCTCAGCA
SangerVali_80	12	14613917	C	C	M	C	TTCCTTCTGTGCCAGTCCTA	TTGGGGTGGTATTGCTGTTCT
SangerVali_81	12	30848675	C	T	Y	Y	CTCACCTGGTTGAGCTCGTTC	GGTGAATGGAGAGCGAGTTGT
SangerVali_82	12	42512844	A	M	A	A	GTTTCTGAGGCCTCTCCAGCA	TGCAGGCATGAGTCACAACATC
SangerVali_83	12	52885316	A	R	A	-	-	-
SangerVali_84	12	55420426	A	C	M	M	AATCCAAGGAGATGGGAGAGC	TGACTCTTCGAGCCCTGAATC
SangerVali_85	12	58018979	C	C	Y	Y	GGTGCATGTTTATGCTGTGTT	CCTCACCTGCACCACTTCTCT
SangerVali_86	12	83251120	C	C	S	C	AGGGTTGAATCCCTGGAGCTA	TGCGAGAACAGTCACTCCTTG
SangerVali_87	12	119866496	C	C	M	C	CCAAAACCTTACCATGTCCA	TTGTAGGCAACCCCATAGC
SangerVali_88	13	25367325	C	C	Y	Y	CGATACATACCACCGCTAGAA	CAAGATCTTCCCTCTGCTTT
SangerVali_89	13	101258623	T	C	T	T	AGCCAGGCTAGCTTTGAAAAA	CCGAGTGTCTGAGGAATG
SangerVali_90	13	114537583	T	T	K	T	GGAGATCCCAACACCTTAGC	TGTAATTCCTTGAGCGCACCT
SangerVali_91	14	39818145	G	A	R	R	CCTGCTGAAAATGAAGCCACT	CCCAATTCGAAAAGTAATGAGA
SangerVali_92	14	50750699	A	A	M	A	TGGTTAAAACCAAGATGCAAAA	AGCAAGAAAAGCTTGGCAAAA
SangerVali_93	14	75416199	A	A	M	A	CTACCTGCATGGTGACATTGG	AGAGTTGAGGGAGCTCGGATG
SangerVali_94	15	42104243	G	G	K	G	GGCTGTATGCCTGTGGAGAAG	AATGGTTCAGTGGGGTGAAT
SangerVali_95	15	50784950	C	T	C	C	GATGCCTGTGGTACAGGTGCT	TGGAGCGTTACATAGGCACTG
SangerVali_96	15	58729144	C	Y	C	Y	TCCGGACAGAAGAGAATGAGG	ATTCAAAACCATGCCTGTTTCG
SangerVali_97	15	73023725	G	G	K	G	CAGCCTTTTGGGAATGACTGA	GCACTCAATGAGAAGGGCAAG
SangerVali_98	16	67913783	G	A	G	G	GAGACCAAGCCAAGTTGATG	CTGCCATCCTCCATCTCTGAC
SangerVali_99	16	70896033	C	T	C	-	-	-
SangerVali_100	16	72143357	C	C	S	C	TCCCATGTGGCTCCTAATTG	AATACCAAGCGGAGGGAACAC

(表 7 の続き)

Sanger 検証用 ID	Chr	Coordinate	Ref.	SNP chip call	我々の NGS call	Sanger シーケンス結果	Sanger 用の PCR L 側 primer 配列	Sanger 用の PCR R 側 primer 配列
SangerVali_101	16	85942643	A	A	M	A	ATTCATGGAAGGGAAGGATG	CTGGAAGGTTTCATGCCATA
SangerVali_102	17	5037195	G	A	G	G	ACAGACAACTCGGCTGTGGT	GTGGCTCCCGTAGGCTTACTT
SangerVali_103	17	42552212	G	G	K	G	TGGCATGCAAGAAAACAACCTC	TGAACGATTGCAGTGTCTTT
SangerVali_104	17	45232068	A	C	A	A	CCTCTACCTGCTGGGTTCAAA	CTTCTCAGTTTTGGGATTTTGC
SangerVali_105	17	45234597	A	G	A	A	CACTGAGGAATCTGTATTCAAGGA	CTGCCAACTCTTGACAACAC
SangerVali_106	17	45234645	G	C	G	G	GAAAATTTCTACCAAGTCATTACAAGGCCTATTTCTGTTTCCACTG	
SangerVali_107	17	45820022	A	A	M	A	GCTACAGCACACCACTGATGC	AATTAGGGGTAGGGGCTTGGT
SangerVali_108	17	61512597	A	A	M	A	GTTGAACAGCAGTGCCTCCTT	TAGTGCTCCTCCCACTGT
SangerVali_109	17	62854928	G	R	G	G	AACAAAGCATACACCTCAGCCTTA	CACAAAAGAACTTCCAGGATATGG
SangerVali_110	17	73500739	C	T	C	C	CAGGAGCCAACTCAGCATTTG	CAGACAGCCCAGGTACATCCT
SangerVali_111	17	75186962	C	T	C	C	GCTGTCCAGGGAGAACACAG	CGTCTTTCCGACCTTTACCT
SangerVali_112	19	434848	T	C	T	T	CCCGGATGCTGTTGTAGTAA	AAGTAGTGGGCTGGTCACTG
SangerVali_113	19	6008188	A	M	A	A	CTCCAGCTCCTCAGCGTCAC	TCCAGTTGTCCACGGTGTCTTA
SangerVali_114	19	9922242	A	G	A	A	AGCACGATGCATTCAAGCAG	CTCTGTACCCCCAATCTCTGG
SangerVali_115	19	12060947	A	M	A	A	TGGGAATGGATTACATCTGC	AAGTTTTGCCACATTGCTTGC
SangerVali_116	19	17837598	C	C	S	C	ACTTGAGGTTCTGCGAGAGC	AGACACTCGGTTTGGGGTCTT
SangerVali_117	19	37677689	C	Y	C	C	GAATTCTTCGGTGTGCAATCA	CGTCTTTTTAGGCATCACA
SangerVali_118	19	38795554	A	R	A	A	TCCAGCCCAGAACCATCTCTTC	CGTGGACAGCCTCTCCTTGG
SangerVali_119	19	38851455	A	A	M	A	CCCTGCAGCTGGTGTACCTT	AAGTGCTGAACTCCTGCCATC
SangerVali_120	19	40540697	G	R	G	G	GCAAGCTGTGTGAGAAGACCA	TCACACCCAGCTTAATCACCA
SangerVali_121	19	48282172	T	Y	T	-	-	-
SangerVali_122	19	52383649	T	T	Y	Y	GCCATAGAAAACATCCCATGC	TATGGATCAGGGGTTTGTGA
SangerVali_123	19	55316329	A	R	A	R	TGGGGAGAATCTTCTGAGCAC	GGGTGAAAACCTCGACATCTG
SangerVali_124	20	42204913	A	A	M	A	ATTGCACAAGACAGGGGTCAT	ATGTGCTTCTGAGCTGCCTCT
SangerVali_125	20	45014787	C	M	C	C	AATTCTGGCAGGAGTGTGGAA	TGGACGTGTCAATCCTTCAGA
SangerVali_126	20	57597971	G	G	S	S	TTTTCGGGGTCGGTTTACTCT	GCCTCCCTGAGCATAGACATC
SangerVali_127	20	58330382	C	C	M	C	GATGAAGGAAGGGAGGTGCAG	GCCATTCTAGGAGCAGCTCAA
SangerVali_128	20	62200738	G	A	G	G	GTGCATCCTCTGCCGATAGTT	TTTGAGCAATGGGTGGTCTTC
SangerVali_129	21	10907041	C	M	C	-	-	-
SangerVali_130	21	45211298	C	C	M	C	TTGGTCTCATTGAGGAGTCA	AAAAGTGGGTTGACCAGATGC
SangerVali_131	22	37260160	A	A	M	A	CGCACAGTGACTTTGAACAGC	TCCCAGGGTAATGTCAAATGG
SangerVali_132	X	3533962	T	Y	C	C	GCAGTCCCTGCTAATGGAAGA	CAGAATGACTGGATTTGTTCTGC
SangerVali_133	X	6069116	T	Y	T	T	CCGTGGGCACGTAGATGTTTA	CTTGGGTCCAGTGGAGCAGTA
SangerVali_134	X	9677599	G	R	A	A	TACTGACCACCTCCCTTCT	CCTGAAGATCATGGATGCACA
SangerVali_135	X	31089628	T	T	A	A	GAACGTTCTGAAAGGGCAACA	GCCTTCCACCTGGAGAAGAAC
SangerVali_136	X	39911528	A	A	M	A	CGTTCGTGAATCCACCAGAT	TAGTTGGGCGCACTTTTCATT
SangerVali_137	X	48049632	T	Y	T	T	TAGCGAGAAAAGCCAGAGTGG	TCTCCTAGCATTGTTGAGAACAT
SangerVali_138	X	153761203	G	C	G	G	TCTCTCAGGGTGTGGACCAGT	TGGCTTCTCTCAGGTCAAGG
SangerVali_139	X	153762708	G	C	G	G	GTCGATGCGGTAGATCTGGTC	CAGAGGGTCACTGGGAACAC
SangerVali_140	X	153764245	A	T	A	A	CTGGTAGAGAGGGCAGAACCA	CCCAGCCACTTCTAACACAC

染色体の位置およびリファレンス配列 (Ref.) はヒトリファレンスゲノムの build GRCh37/hg19 に基づく。

heterozygote のアレルは M が A/C、R が A/G、W が A/T、S が C/G、Y が C/T、K が G/T を示す。

Sanger の結果および primer 配列の部分が「-」(ハイフン) のものは Sanger シーケンスがうまく実施できなかった箇所を示す。

表 8. SNP chip データと比較した際の我々の WES での NGS call の精度検証

	SNP chip のcall	我々のNGS call	Sangerシーケンス前	Sangerシーケンス後
我々の NGS と SNP chip の双方でジェノタイプが 得られた箇所	Ref-Homo	All	223,149	-
	Hetero	All	11,479	-
	No-Ref-Homo	All	8,945	-
	Total		243,573	-
False positive (偽陽性)	Ref-Homo	No-Ref-Homo	1	0
	Ref-Homo	Hetero	55	49 (1)
	No-Ref-Homo	Hetero	8	0
	No-Ref-Homo	No-Ref-Homo (塩基の異なる Homo)	1	0
	Hetero	Hetero (塩基の異なる Hetero)	1	0
	Hetero	No-Ref-Homo	5	1
	Total		71	50 (1)
False negative (偽陰性)	Hetero	Ref-Homo	37	11 (5)
	No-Ref-Homo	Ref-Homo	32	2 (2)
	Total		69	13 (7)

No-Ref-Homo, リファレンス配列とは異なるhomozygote; Ref-Homo, リファレンス配列と同一のhomozygote。
括弧内の数字はSangerシーケンスがうまく実施できなかった数を示す。

表 9. 神経系遺伝子ターゲットリシーケンスでの疾患原因変異同定率

	家系数		合計割合 (%)
	SureSelect V4*	SureSelect V5*	
既知あるいは遺伝子機能から確定的な原因変異が同定できた	4	9	30.2%
確定的ではないが既知の遺伝子機能から原因変異である可能性が高いと推測される変異が見つかった	2	5	16.3%
疾患原因かどうかは不明だが何かしらの候補変異が最終的に残った	3	11	32.6%
一つも疾患原因候補変異が見つからなかった	2	7	20.9%
総数	11	32	100.0%

*SureSelect V4 および V5 は WES の capture probe の platform の種類

表 10. WES の各プラットフォームでの CDS 領域のカバー率

WES プラットフォーム	ターゲット領域長 *	デザイン段階での CDS 領域カバー率	NGS シーケンス データ量	CDS 領域の 平均 depth	CDS 領域の depth 中央値	CDS 領域内で depth が 10 以上の割合	CDS 領域内で depth が 15 以上の割合
Agilent SureSelect V4 [†]	87.49 Mb	93.88%	6.84 Gb	79.15	61.78	90.57%	87.56%
Agilent SureSelect V5 [†]	89.48 Mb	96.53%	6.38 Gb	73.10	65.02	93.85%	91.90%
NimbleGen SeqCap ez v2 [‡]	81.58 Mb	96.08%	18.68 Gb	199.07	172.00	93.54%	92.98%
Illumina TruSeq [‡]	100.33 Mb	94.28%	11.40 Gb	61.72	60.00	90.23%	88.63%

* キャプチャープローブ領域の前後 100 bp を含む長さ

† SureSelect V4 と V5 のデータは我々の検体のそれぞれ 38 検体と 104 検体の平均値を示す

‡ オリジナル生データは Clark らの論文データを用い、それを我々の解析パイプラインで再解析した結果を示す

表 11. CCCS の結果統計データ

ターゲット領域長 (CDS 領域以外も含み、 プローブデザイン領域の前後 100 bp を含む)	7,444,566 bp
ターゲット領域内の CDS 領域	3,303,926 bp
得られたシーケンス read 平均	17,209,225 reads (1,736,171,493 bases)
リファレンスゲノムにマッピングされた read 数平均	16,932,011 reads (97.92%)
単一箇所にもマッピングできた read 数平均	13,183,344 reads
単一箇所にもマッピングされ、paired-end の距離が 1,000 bp 以内に存在した read 数平均	11,412,668 reads
BLAT 前のターゲット領域上の read の base 数平均	965,239,576 bases
BLAT 前のターゲット CDS 領域の平均 depth	169.94
BLAT 前のターゲット CDS 領域の depth が 10 以上の base 数と割合の平均	2,740,520 bases (ターゲット CDS 領域の 82.95%)
BLAT を実施した read 数平均	6,998,677 reads
BLAT でマッピングされた read 数平均	3,164,628 reads
BLAT により追加された base 数平均	106,256,016 bases
BLAT データを追加後のターゲット領域上の read の base 数平均	1,071,495,592 bases
BLAT 後のターゲット CDS 領域の平均 depth	194.55
BLAT 後のターゲット CDS 領域の depth が 10 以上の base 数と割合の平均	2,827,100 bases (ターゲット CDS 領域の 85.57%)

各データは CCCS を実施した 28 検体の平均値を示す。

表 12. CCCS データと HumanOmniExpress Exome SNP chip データとの一致率

サンプル ID*	NGS でリファレンスと異なる homozygote の call が得られた箇所	NGS で heterozygote の call が 得られた箇所	NGS でリファレンスと一致する homozygote の call が得られた箇所	合計
NA18943	99.22% (4 mismatch / 516)	99.30% (4 mismatch / 569)	99.91% (11 mismatch / 11,814)	99.85% (19 mismatch / 12,899)
MCPH_II-2	100.00% (0 mismatch / 501)	99.35% (4 mismatch / 619)	99.84% (17 mismatch / 12,149)	99.84% (21 mismatch / 13,269)
平均	99.61%	99.33%	99.89%	99.85%

表 13. CCCS と SNP chip で不一致だった 40 箇所の検証データ

Sanger 検証用 ID	Chr	Coordinate	Ref.	SNP chip call	我々の NGS call	Sanger シー ケンス結果	Sanger 用の PCR L 側 primer 配列	Sanger 用の PCR R 側 primer 配列
SangerVali_2_1	1	1458567	A	R	A	-	-	-
SangerVali_2_2	1	12837153	G	R	G	G	TCTGCCCTGAATGAGCAGAGT	AACTGAGAGGGGCTCAGGACT
SangerVali_2_3	1	12837669	G	K	G	G	ATGATGCCCAGGGTGCTCTC	TCCCTTCTGCATCCCCTTT
SangerVali_2_4	1	12837720	G	R	G	G	ATGATGCCCAGGGTGCTCTC	TCCCTTCTGCATCCCCTTT
SangerVali_2_5	1	12837728	G	R	G	G	ATGATGCCCAGGGTGCTCTC	TCCCTTCTGCATCCCCTTT
SangerVali_2_6	1	12921311	G	R	G	G	GTGTCTCTCCAGTTCCAAGC	GCTGCCTGAATTCCTCAGTGT
SangerVali_2_7	1	225702574	T	T	W	T	GATGCCAAAAAGAATCCAGATG	TTTAGCATGGCATTCTGCGTA
SangerVali_2_8	2	133875794	G	G	R	R	TGCCCCCTTCTCAGTTCTCTATT	AAAAATTGGTTTGGCTCTATGCT
SangerVali_2_9	2	172411273	G	R	G	-	-	-
SangerVali_2_10	3	75790513	T	Y	T	T	AATTTGCTGGGCATTACTCCTC	CCAGCTGTCGTCATTCTGAAA
SangerVali_2_11	3	97868931	G	R	G	G	TGGTTTTTATTTTGCAGTTCA	TTACTTGCTTGTTTCTCAGGCTGT
SangerVali_2_12	3	97868931	G	R	G	G	TGGTTTTTATTTTGCAGTTCA	TTACTTGCTTGTTTCTCAGGCTGT
SangerVali_2_13	4	56502178	G	R	A	R	GAGCAGGGCCTGGGTATTTTC	GAGATGCTGCGAACAGAGAGC
SangerVali_2_14	4	69434042	A	R	A	G	AGCAGATGATTTACTGGCATTGA	AACAACAACCTGAAAAGAAGCACTG
SangerVali_2_15	7	82784501	G	A	G	G	AAGGAGGCTGTTTTGCAGATGG	GGGACACCAAAACCCCTAGCTC
SangerVali_2_16	7	82784501	G	A	G	G	AAGGAGGCTGTTTTGCAGATGG	GGGACACCAAAACCCCTAGCTC
SangerVali_2_17	7	149499412	G	R	G	A	CGAAGCCAGTCCTGCTTTGT	CACAGCTGGCTGACACATGA
SangerVali_2_18	8	102213962	C	C	G	C	CCCACGGGAGAGCTGTATCTT	GTTTTCCACAGGGTTGTGTCA
SangerVali_2_19	11	20101690	A	A	M	A	CCTGCCTTGGTCTCCAGGAA	TCCTGGGATCTACGCCCAA
SangerVali_2_20	11	73760510	A	A	M	A	TGATTTTGAGGTCTCCACTCCA	CCTCCATCCCCTGAATCAGCTT
SangerVali_2_21	11	82984713	T	Y	T	T	GTTCCCCAAAAGTGCTGGTTA	AAGCATGAAACCACTCGAATCA
SangerVali_2_22	11	82984713	T	Y	T	T	GTTCCCCAAAAGTGCTGGTTA	AAGCATGAAACCACTCGAATCA
SangerVali_2_23	11	82991264	C	Y	C	C	TTCAGGAGGTATCCATGTTTCTCA	AAAGAGAGGCTCAGAGGAGTCAA
SangerVali_2_24	11	82991264	C	Y	C	C	TTCAGGAGGTATCCATGTTTCTCA	AAAGAGAGGCTCAGAGGAGTCAA
SangerVali_2_25	12	133682655	A	R	A	R	AAACCCCATGAGTGAAGGACTG	TACGGGGTTTTGGTTGTATGG
SangerVali_2_26	14	36140611	G	G	A	G	AATGTAGGAGAAGCAATCAATTCCA	TACAATCATCAAACACTGCTCACC
SangerVali_2_27	16	70883694	G	R	G	G	AGGAGCTTCTGGACACATGGA	TGGTTTGCATTACTGCTCTG
SangerVali_2_28	16	70993553	A	W	A	A	AATTTTCCAAAAGGCACACA	AGGTGGAAAACAACCCAGTGA
SangerVali_2_29	17	45232068	A	C	A	A	CCTCTACCTGCTGGGTTCAAA	GACTGATTTGCCATTCTCTTCC
SangerVali_2_30	17	45232068	A	C	A	A	CCTCTACCTGCTGGGTTCAAA	GACTGATTTGCCATTCTCTTCC
SangerVali_2_31	17	45234597	A	G	A	A	CACTGAGGAATCTGTATTCAAGGA	CTGCCCAACTCTTGACAACAC
SangerVali_2_32	17	45234597	A	G	A	A	CACTGAGGAATCTGTATTCAAGGA	CTGCCCAACTCTTGACAACAC
SangerVali_2_33	17	45234645	G	C	G	G	CAATTGTGCTCTGGGGTGTTC	GGCCTATTTCTGTTTCCACTGA
SangerVali_2_34	17	45234645	G	C	G	G	CAATTGTGCTCTGGGGTGTTC	GGCCTATTTCTGTTTCCACTGA
SangerVali_2_35	19	622193	C	C	M	C	TACTGACGGCTGCGCTGAGA	CTGACCGCCCTGCTCTACA
SangerVali_2_36	19	36168914	T	Y	C	Y	TATCTCGTGGTTTGGGTTTGC	GGCTCAGGAACCTGGAAAAAG
SangerVali_2_37	19	43371294	C	C	M	M	CGTACAAGGGTTTTCCCATGA	CCACATTTGAAGAACTCCACCT
SangerVali_2_38	19	51919263	G	G	K	G	GTATCCCTCTGCCCTCCAAT	CCAGGGAAAATGTCCCATACCT
SangerVali_2_39	22	17264565	G	K	G	K	GCACGTTACAGGAGAACATAAA	ATGGCAGTCAGGCAAAGTGT
SangerVali_2_40	X	48814258	T	T	K	-	-	-

染色体の位置およびリファレンス配列 (Ref.) はヒトリファレンスゲノムの build GRCh37/hg19 に基づく。

heterozygote のアレルは M が A/C、R が A/G、W が A/T、S が C/G、Y が C/T、K が G/T を示す。

Sanger の結果および primer 配列の部分が「-」(ハイフン) のものは Sanger シーケンスがうまく実施できなかった箇所を示す。

表 14. SNP chip データと比較した際の我々の CCCS での NGS call の精度検証

	SNP chip のcall	我々のNGS call	Sangerシーケンス前	Sangerシーケンス後
我々の NGS と SNP chip の双方でジェノタイプが 得られた箇所	Ref-Homo	All	23,945	-
	Hetero	All	1,200	-
	No-Ref-Homo	All	1,023	-
	Total		26,168	-
False positive (偽陽性)	Ref-Homo	No-Ref-Homo	2	2
	Ref-Homo	Hetero	8	6 (1)
	No-Ref-Homo	Hetero	0	0
	No-Ref-Homo	No-Ref-Homo (塩基の異なる Homo)	0	0
	Hetero	Hetero (塩基の異なる Hetero)	0	0
	Hetero	No-Ref-Homo	2	2
	Total		12	10 (1)
False negative (偽陰性)	Hetero	Ref-Homo	20	6 (2)
	No-Ref-Homo	Ref-Homo	8	0
	Total		28	6 (2)

No-Ref-Homo, リファレンス配列とは異なるhomozygote; Ref-Homo, リファレンス配列と同一のhomozygote。
括弧内の数字はSangerシーケンスがうまく実施できなかった数を示す。

表 15. CCCS を実施した 28 検体の検体情報とその WES および CCCS のデータ量と CDS カバー率

Sample ID	性別	疾患名	WES のデータ量 (base)	WES で CDS 領域 の depth が 10 以上の塩基数	CCCS のデータ量 (base)	CCCS で CDS 領域 の depth が 10 以上の塩基数	WES と CCCS の統合 データで CDS 領域の depth が 10 以上の塩基数	WES と CCCS の統合 データで CDS 領域の depth が 10 以上の割合
NA18943	男	健常 control	7,723,799,866	31,970,268	1,343,806,200	2,732,114	32,864,913	97.36%
OFB-3	男	小頭症	6,455,890,600	31,743,212	1,449,126,600	2,786,082	32,793,358	97.15%
OFB-59	男	健常 (OFB-3 の父親)	6,220,816,400	31,697,499	1,133,761,600	2,700,654	32,752,365	97.03%
OFB-60	女	健常 (OFB-3 の母親)	6,782,985,000	31,695,811	1,561,996,200	2,769,126	32,746,589	97.01%
K3423	女	小頭症	5,621,947,800	31,588,057	1,717,952,228	2,843,819	32,740,175	96.99%
K3424	男	小頭症	5,373,814,000	31,552,304	1,895,463,364	2,877,576	32,769,424	97.08%
K3425	男	健常 (K3243 と K3244 の父親)	6,058,804,200	31,613,101	1,958,575,436	2,875,799	32,793,804	97.15%
K3426	女	健常 (K3243 と K3244 の母親)	5,486,626,400	31,548,655	1,808,916,464	2,858,552	32,736,556	96.98%
K2941	女	両側シルビウス裂周囲多小脳回	5,753,303,000	31,596,174	1,860,680,378	2,863,327	32,747,746	97.01%
K2942	女	両側シルビウス裂周囲多小脳回	5,855,006,600	31,532,350	2,345,938,312	2,904,453	32,735,803	96.98%
K2943	男	健常 (K2941 と K2942 の父親)	5,905,630,000	31,608,348	1,770,633,222	2,836,051	32,779,628	97.11%
K2944	女	健常 (K2941 と K2942 の母親)	4,944,724,600	31,300,424	1,668,052,370	2,823,832	32,600,095	96.58%
K2801	男	小頭症と脳回欠損	5,738,220,200	31,438,462	1,896,317,218	2,864,281	32,715,312	96.92%
K2802	男	小頭症と脳回欠損	6,656,280,600	31,690,354	1,805,333,792	2,858,745	32,811,085	97.20%
K2803	男	健常 (K2801 と K2802 の父親)	5,738,374,400	31,509,347	2,267,141,344	2,909,018	32,760,173	97.05%
K2804	女	健常 (K2801 と K2802 の母親)	6,113,201,600	31,579,058	1,696,266,922	2,838,740	32,735,800	96.98%
K2605	男	小頭症	6,141,457,400	31,583,198	1,447,285,358	2,718,717	32,736,228	96.98%
K2606	男	健常 (K2605 の父親)	6,358,131,800	31,692,874	1,812,551,858	2,843,725	32,808,895	97.19%
K2607	女	健常 (K2605 の母親)	6,932,848,600	31,698,413	2,100,156,024	2,865,803	32,778,919	97.11%
OFB-39	男	精神遅滞と小頭症	6,434,297,800	31,682,539	1,681,480,724	2,832,617	32,810,510	97.20%
OFB-65	男	精神遅滞と小頭症	6,142,420,200	31,720,085	1,659,908,538	2,819,290	32,825,268	97.24%
OFB-66	男	健常 (OFB-39 と OFB-65 の父親)	7,502,036,000	31,761,975	1,659,938,030	2,806,906	32,833,930	97.27%
OFB-67	女	健常 (OFB-39 と OFB-65 の母親)	5,760,032,000	31,561,128	1,620,079,390	2,778,348	32,711,620	96.91%
NCU3-53	女	軟骨低形成疑い、透明中隔欠損、脳室拡大	5,914,430,400	31,641,240	1,657,435,654	2,821,600	32,764,345	97.06%
NCU3-54	女	健常 (NCU3-53 と NCU3-55 の姉妹)	6,360,736,200	31,729,864	1,906,358,234	2,852,411	32,801,792	97.17%
NCU3-55	女	軟骨低形成疑い、透明中隔欠損、脳室拡大	6,868,378,600	31,722,725	1,701,860,908	2,831,085	32,792,549	97.15%
NCU3-51	男	健常 (NCU3-53 と NCU3-55 の父親)	6,537,301,800	31,716,825	1,591,500,632	2,819,852	32,815,320	97.21%
NCU3-52	女	健常 (NCU3-53 と NCU3-55 の母親)	6,815,622,200	31,804,071	1,594,284,798	2,826,288	32,819,764	97.23%
平均値			6,221,325,652	31,642,084	1,736,171,493	2,827,100	32,770,785	97.08%

謝辞

学生時代にご指導ご鞭撻を賜り、博士論文の審査委員への推薦を行っていただきました筑波大学大学院人間総合化学研究科医学系専攻疾患制御医学専攻神経病態医学分野の玉岡晃教授に深謝いたします。

ご多忙の中、本学位論文の審査委員として審査いただきました主査の須磨崎亮教授、副査の野口恵美子教授、増本幸二教授、山本哲也准教授、およびご協力いただきました筑波大学大学院人間総合科学研究科の先生方や職員の方々に心より感謝いたします。

本研究で共同研究者としてご助力、ご指導、ご助言をいただきました慶應義塾大学医学部臨床遺伝学センターの小崎健次郎教授、独立行政法人国立病院機構大阪医療センター再生医療研究室の金村米博室長、独立行政法人大阪府立病院機構大阪府立母子保健総合医療センター遺伝診療科の岡本伸彦主任部長、名古屋市立大学大学院医学研究科新生児・小児医学分野の齋藤伸治教授、昭和大学医学部小児科学講座の加藤光広講師、社会医療法人愛仁会高槻病院神経外科の山崎麻美副院長、千葉大学大学院医学研究院小児病態学の塩浜直博士、独立行政法人国立病院機構東京医療センター感覚器センター聴覚障害研究室の松永達雄室長、同務台英樹博士、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の清水厚志特命教授に心より感謝の意を表します。

理化学研究所での研究において、上司としてご指導ご鞭撻をいただきました理化学研究所統合生命医科学研究センターおよび東京医科歯科大学難治疾患研究所医科学数理分野の角田達彦教授に深甚の謝意を表します。

最後に、本研究において疾患の解明や治療を目指した研究のために快く検体を提供いただきました患者さん及びそのご家族に深く感謝いたします。