

博士論文

論文題目 エキソーム解析データを用いた構造変異検出
パイプラインの構築と神経疾患の診断への応用

氏名 田中 真生

目次

要旨	3
第 1 章 序論	4
第 2 章 Materials and Methods	12
第 3 章 構造変異解析パイプラインの構築	19
第 4 章 網羅的な構造変異の解析	40
第 5 章 考察	74
謝辞	82
引用文献	83

要旨

次世代シーケンサーを用いたエキソーム解析は、メンデル遺伝形式の遺伝性疾患診断における強力な解析手法であるが、構造変異の検出は通常の解析方法では不可能である。今回、構造変異検出プログラムである **Conifer** を中心として、エキソーム解析後のデータから構造変異を検出する解析パイプラインを構築し、過去当科にてエキソーム解析を行った 1230 例について解析を行った。結果として 26 例に原因遺伝子変異の可能性がある構造変異を検出し、また 4 例に性染色体のコピー数異常（トリソミー）を検出した。これらの変異はこれまでのエキソーム解析では見逃されていた変異であり、本解析法がエキソーム解析による遺伝子診断の補助として有用であることを示した。

第 1 章 序論

1-1 次世代シーケンサーを用いたエキソーム解析について

次世代シーケンサーを用いた遺伝子解析法は，近年の解析技術の進歩およびコストの削減に伴い，メンデル遺伝形式の遺伝性疾患の診断ツールとして，急速に使用されるようになってきている．エキソーム解析は蛋白質をコードする DNA 配列のみをハイブリダイゼーション法を用いて抽出し，次世代シーケンサーで解析を行う手法である[1-3] (図 1-1)．全ゲノム解析と比較して低コストでありながら，原因変異の存在する可能性の高いエキソン領域に全体に渡って変異を検出することが可能であり，神経疾患領域を含めてクリニカルシーケンシングとして応用されるようになってきている[4-5]．ただし，エキソーム解析では，エキソン領域内の塩基置換変異および短い挿入・欠失変異の検出は可能であるが，エキソン領域外の変異の検出はできず，また，構造変異についても通常の解析方法では検出ができない．

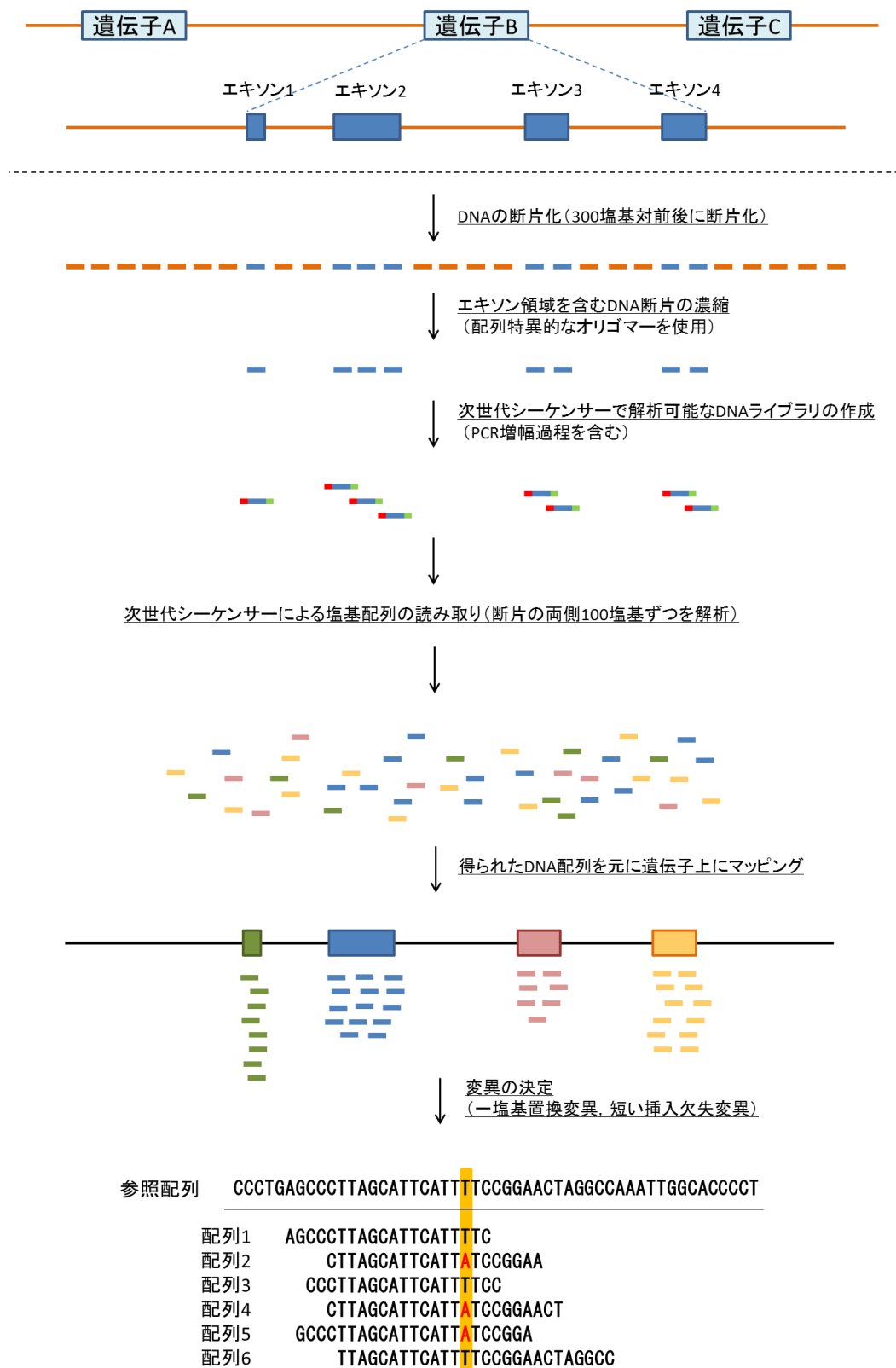


図 1-1. エキソーム解析における解析手順.

1-2 構造変異について

構造変異は1 kb以上の塩基配列が挿入、欠失あるいは重複することにより生じる遺伝子変異を指す[6]。構造変異は健常者においても高い頻度で認められることが報告されており[7]、ゲノム配列の個人差を生み出す一つの要素として注目されている。また、疾患との関連では、自閉症[8-9]、統合失調症[10-11]、てんかん[12-13]などの疾患で、低頻度の構造変異の集積が見られることが報告されている。

神経内科領域においても構造変異が原因で生じる疾患は少なくなく、*PMP22* 遺伝子の重複によるシャルコー・マリー・トゥース病 (CMT1A) [14]、*PMP22* 遺伝子の欠失による遺伝性圧脆弱性ニューロパチー (HNPP) [15]、*SNCA* 遺伝子の重複による家族性パーキンソン病 (PARK4) [16-17]、*APP* 遺伝子の重複による家族性アルツハイマー病[18]、*SMN1* 遺伝子の欠失による脊髄性筋萎縮症 (SMA) [19]などが代表的疾患として挙げられ、感度の高い検出方法が求められる。

1-3 エキソーム解析データを用いた構造変異の検出

通常のエキソーム解析におけるデータ解析パイプラインは構造変異を念頭に置いて作成されておらず、構造変異の検出は不可能である。しかし、実際にはエキソーム解析で構造変異に関する情報が得られないわけではなく、通常着目しない

解析データを利用することによって情報を得ることが可能である。次世代シーケンサーにより得られた配列データを用いて構造変異を検出する解析プログラムは、現在までに複数報告されているが[20]、多くは全ゲノム解析データを利用するものであり、エキソーム解析データを利用するプログラムはそれに比べると少ない。

次世代シーケンサーデータを用いた構造変異検出手法としては、(1) ペアエンドによるマッピングを利用する方法（異常に長いインサート長のペアリードを抽出し構造変異を検出する方法）[21]、(2) 切断点を含むリードを検出し構造変異を検出する方法（リードの一部のみしかマップされないリードを分割して再度マッピングし、変異を検出する方法）[22]、(3) マップされたリードの数（read depth）を利用する方法[23]の3種類が主に用いられる。全ゲノム解析データを用いた解析ではこれらいずれの手法も使用でき、複数の手法を組み合わせた解析方法も報告されている。しかし、エキソーム解析ではゲノム上にまばらに散らばる、全ゲノムの1~2%を占めるエキソン領域のみの情報しか得られないことから、(1)、(2)の手法を用いて解析することは難しく、(3)のマップされたリード数を元にした解析プログラムがほとんどである。

また、エキソーム解析においては、ハイブリダイゼーション法を用いて一部の

DNA 断片を抽出する過程が含まれているため、全ゲノム解析のデータを利用する場合と比較して解析が難しい。エキソーム解析において、マップされたリード数は構造変異以外にも様々な要因で変化しうるため、それらの要因の影響を少なくし、ノイズを減らす工夫が必要である[20]。マップされたリード数に影響を与える因子として考えうるものについて、表 1-1 に示す。

<DNA 自体>

遺伝子のコピー数（構造変異）

DNA の質（抽出・精製方法の違い，DNA の劣化）

<Wet の解析過程>

エキソンの濃縮用プローブによるキャプチャー効率

次世代シーケンサー内でのクラスター形成効率（GC 率等による影響）

塩基コールの正確性（クオリティ）

得られるリードデータの総数

<Dry の解析過程>

配列構造自体によるマッピング効率の変化（偽遺伝子や反復配列の存在）

マッピングに影響する変異の有無（特に挿入・欠失変異の存在）

表 1-1. エキソーム解析データにおいて，リード数に影響を与えうる因子についてまとめたもの．遺伝子コピー数の変動（構造変異）以外にも，様々な要素がリード数に影響を及ぼしうる．

1-4 今回解析に使用した Conifer プログラムの特徴

今回解析に使用した Conifer は[24], 多数例のエキソーム解析データを利用して, 領域毎のリード数を解析全リード数で補正した RPKM (Reads Per Kilobase per Million mapped reads) 値から Z-score (Z-RPKM 値) を算出し, その値をもとに変異の検出を行うプログラムである. また, 解析データのノイズ削除のため, 特異値分解法 (Singular Value Decomposition 法) を使用している点が特徴である. 欠点としては, 3 領域以上にわたって連続して有意な変動が認められた領域を変異として検出するため, 長さの短い構造変異が検出されないこと, および Z-score を元にして判断を行うことから, 解析集団中に特定の変異を有する頻度が高い場合は検出ができないことなどが挙げられる.

1-5 今回の解析の目的

当科ではこれまで, エキソーム解析で変異が認められないが臨床像から遺伝子変異の存在が強く疑われる症例については, (1) 一塩基置換変異や短い挿入欠失変異を見逃している可能性, (2) 構造変異が存在する可能性を考え, 他の手法による解析をその都度追加することで対応してきたが, エキソーム解析を行った全例に対して他の手法を用いて構造変異の有無について確認することは, コスト面等の問題から難しく, 行えていなかった. もし, エキソーム解析データを用いて全例

について構造変異の有無についての情報が得られるのであれば、遺伝子診断の正確性を高める上で有用であると考えられる。また文献的にも、多数例の神経疾患症例に対して、エキソーム解析データを用いて構造変異の有無について網羅的な評価を行った報告はこれまでにない。

本研究では、既存の構造変異検出プログラムである **Conifer** を中心に据えて、エキソーム解析後のデータから構造変異を検出する解析パイプラインを構築し、過去当院にてエキソーム解析を行った症例を対象に上記パイプラインを用いて解析を行い、神経疾患診断における本解析法の有用性を明らかにすることを目的とする。

第 2 章 Materials & Methods

2-1 解析対象とした症例について

これまでに、当院当科にて DNA を収集しエキソーム解析を行った症例、および、鹿児島大学大学院医歯学総合研究科神経内科・老年病学講座 高嶋 博 教授、山梨大学医学部神経内科学講座 瀧山 嘉久 教授より DNA を提供され、当科にてエキソーム解析を行った 1230 例のデータを用いて解析を行った。なお、1230 例のうち 1 例については、解析前の時点で構造変異の存在が他の手法により確認されていた (*PMP22* 遺伝子の重複)。解析を行った症例の臨床診断名毎の症例数について表 2-1 に示す。

診断名	症例数
筋萎縮性側索硬化症	371
シャルコー・マリー・トゥース病	289
痙性対麻痺	172
多系統萎縮症	114
副腎白質ジストロフィー	54
脊髄小脳変性症	51
周期性四肢麻痺	48
筋疾患	23
白質脳症	16
神経核内封入体病	13
パーキンソン病	11
てんかん	11
片頭痛	8
その他の疾患	37
正常対照	12
合計	1230

表 2-1. 疾患毎の症例数を上記に示す. なお, その他の疾患の内訳は以下の通り. ファール病 4 例, ジストニア 3 例, 前頭側頭葉変性症 3 例, ミトコンドリア病 3 例, Alternating Hemiplegia 3 例, アルツハイマー病 2 例, 正常圧水頭症 2 例, 進行性核上性麻痺 2 例, 脊髄性筋萎縮症 2 例, Pelizaeus Merzbacher 病 2 例, 舞踏運動 1 例, Episodic ataxia 1 例, 糖原病 1 例, Hallervorden-Spatz 病 1 例, 髄鞘低形成 1 例, Inclusion body myopathy associated with Paget disease of bone and/or frontotemporal dementia (IBMPFTD) 1 例, ニューロフェリチノパチー 1 例, 神経線維腫症 1 型 1 例, 眼咽頭遠位型ミオパチー 1 例, Startle disease 1 例, 結節性硬化症 1 例.

2-2 エキソーム解析

エキソン領域の濃縮およびシーケンサー用のライブラリ作成については、SureSelect Human All Exon V4+UTR kit (以下 V4 キットと記載) もしくは SureSelect Human All Exon V5+UTR kit (以下 V5 キットと記載, いずれも Agilent 社) を使用し, Agilent 社のプロトコルに従って行った. 原理的には, いずれのキットにおいても, 標的領域毎に作成された 120 塩基のビオチン化された RNA プローブを使用し, 断片化された DNA と RNA プローブをハイブリダイズさせ, ストレプトアビジンビーズを用いて RNA プローブに結合した DNA 断片のみを回収することによって, 標的領域の DNA 断片の濃縮を行っている.

作成されたライブラリは HiSeq2000 シーケンサー (illumina 社) を用いて解析を行い, 100 塩基×2 のペアエンド法により塩基配列の読み取りを行った. 解析により得られたリード情報 (fastq 形式) のマッピングには BWA [25] を使用し, 参照配列として GRCh37/hg19 を使用した. なお, ここで得られたマッピング後の配列データ (bam 形式) を元に, RPM 値 (Reads Per Million mapped reads) を算出した. また, 変異の検出プログラムとして Samtools [26] を使用した.

2-3 Conifer による構造変異解析

Conifer へのデータ入力は, 症例毎の全解析領域の RPM 値を RPKM 値 (Reads Per Kilobase per Million mapped reads) に変換したのち, 使用したキット毎 (V4 キットもしくは V5 キット) に分けて行った. また, 解析中の各種パラメータについては, SVD 値 (Singular Value Decomposition value) は 5 に設定し, その他についてはデフォルトの値を使用して行った[24].

2-4 アレイ CGH 法による構造変異解析

既知の疾患原因遺伝子を含む構造変異が検出され, 診断名などの臨床情報と一致する症例については, アレイ CGH 法を用いて変異が実際に存在するかどうか確認を行った. SurePrint G3 CGH Microarray 8x60 K アレイ (Agilent 社) を使用し, 検出された構造変異領域およびその周辺に高密度にプローブ (45~60 bp) をデザインして解析を行った. 検討を行った変異毎に作成したアレイプローブ情報の詳細は表 2-2 に示した通りである. なお, 性染色体のアリル数の判定については, SurePrint G3 CGH Microarray 8x60 K にあらかじめ搭載されているバックボーンプローブのデータを用いて判定を行った.

解析手順は Agilent 社のプロトコルに従い、制限酵素 (AluI, NsaI) を用いた DNA の断片化、断片化 DNA の蛍光色素 (Cyanine3 および Cyanine5) によるラベル化、マイクロアレイ上へのハイブリダイゼーションを行い、G2505B (Agilent 社) を用いて画像の読み取りを行った。なお、ラベル化反応については、コントロール DNA は Cy-3、患者 DNA は Cy-5 を用いてそれぞれ標識した。得られた画像データを、Feature Extraction Ver 11.0 および Agilent Genomic Workbench Ver 6.5 を用いて解析した後、最終的な色素強度比率のデータを算出して、関心領域のデータの描画を行った。

Target gene	Interval	Size [bp]	Total Probes	Probes per 1 kb
<i>PARK2</i>	chr6:161753591-163163834	1410244	9839	6.97
<i>PMP22</i>	chr17:13957720-15537826	1580107	10484	6.63
<i>CYP2U1</i>	chr4:108837789-108946210	108422	748	6.89
<i>FA2H</i>	chr16:74731854-74938859	207006	1366	6.59
<i>FXN</i>	chr9:71534669-71730094	195426	1297	6.63
<i>KIF5A</i>	chr12:57922456-57993554	71099	536	7.53
<i>LITAF</i>	chr16:11395426-11786506	391081	2373	6.06
<i>OPTN</i>	chr10:12862250-13195276	333027	2225	6.68
<i>SACS</i>	chr13:23393591-25031252	1637662	11273	6.88
<i>SPAST</i>	chr2:32173296-32397706	224411	1136	5.06
<i>ZFYVE26</i>	chr14:68198238-68298306	100069	768	7.67

表 2-2. Conifer 解析で検出した構造変異毎にデザインした、アレイ CGH プローブの情報を示す。切断点を決定するため、解析領域は標的遺伝子のみではなく、構造変異の検出された範囲全体を完全に含むようにデザインした。

2-5 SMN1 遺伝子欠失の制限酵素断片長多型解析法による確認

SMN1 遺伝子の欠失はアレイ CGH 法では原理的に行えないため、制限酵素断片長多型解析法によって確認を行った。まず、SMN1 遺伝子および SMN2 遺伝子の Exon 7 を既報のプライマーペア (forward: 5'-AGACTATCAACTTAATTTCTGATCA-3, reverse: 5'-CCTTCCTTCTTTTTGATTTTGTTT-3') を用いて PCR 法によって増幅した[27]。PCR 産物を精製後、SMN2 遺伝子由来の PCR 産物のみを切断する制限酵素 DraI により処理し、得られた DNA 断片をアガロースゲル電気泳動によって分離して、SMN1 および SMN2 遺伝子由来のバンドが見られるかどうか確認し判定を行った。

2-6 倫理面に関する配慮

本研究は東京大学ヒトゲノム・遺伝子解析研究倫理審査委員会および参加各機関の研究倫理委員会の承認に基づいて施行された (承認番号 1396)。

第3章 構造変異解析パイプラインの構築

3-1 構造変異解析パイプラインの全体像

以下、順に構造変異の解析方法について記述するが、解析全体の流れを図 3-1 に示す。なお、Conifer プログラムを用いた構造解析および、アレイ CGH 用の解析プログラム (Feature Extraction および Agilent Genomic Workbench) を用いたデータ解析以外の情報処理については、全て自作の解析プログラム (Perl script) を作成して行った。具体的には、Conifer へ入力するためのデータ加工、Conifer 出力データへの遺伝子情報などの追加、性染色体アレル数の判定、解析除外症例の選択などについて、複数 (10 以上) のプログラムを作成しパイプラインを構築した。

3-2 解析領域の設定について

V4 キットおよび V5 キットで使用されている全ての RNA プローブについて、プローブの両側 100 塩基を含めた領域を、それぞれのライブラリ調整キットにおける解析対象領域として設定した。また、周囲 100 塩基を含めたプローブ領域が重なっているものについては、結合させて一つの領域として扱うこととした。使用プローブ数は V4 キットが 789140 個、V5 キットが 881870 個であり、結合後の領域数は V4 キットが 177394 領域、V5 キットが 184751 領域であった。

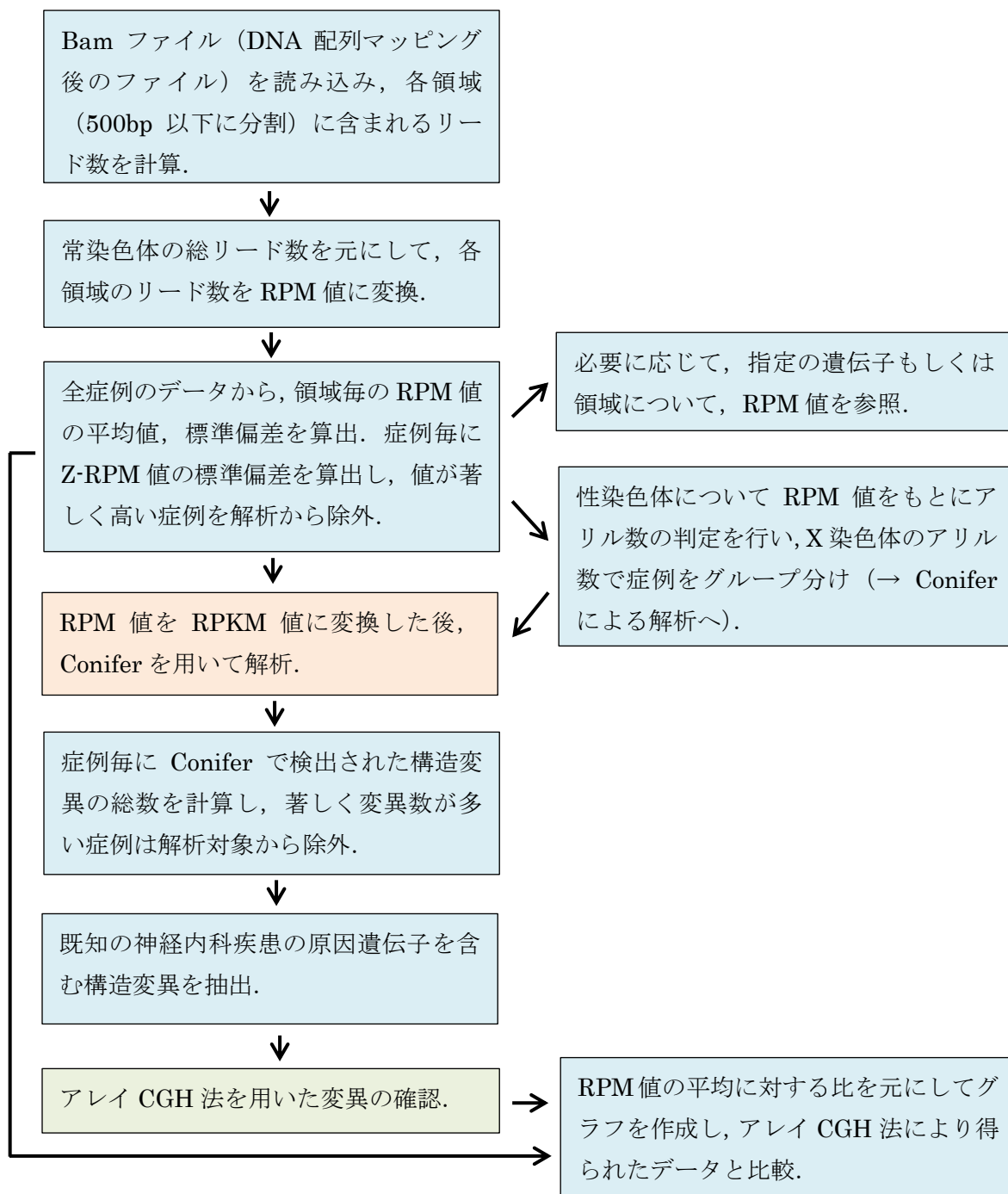


図 3-1. 解析全体の流れ. 青色の部分は自分で作成したプログラム (Perl script) を用いてデータ処理を行った.

その後、領域の大きさを全体的に揃え、解析時のポイント数を増やす目的で、領域の大きさが 500 塩基以下の場合はそのまま解析に使用し、領域の大きさが 500 塩基を超える場合には、500 塩基未満となるように領域を分割した（図 3-2）。最終的に、V4 キットでは 269834 領域、V5 キットでは 294419 領域となった。

なお、V4 キットと V5 キットで解析対象領域が異なるため、以後の構造変異に関する解析は使用したキット毎に分けて解析を行った。キット毎の症例数は、V4 キットが 361 例、V5 キットが 869 例であった。

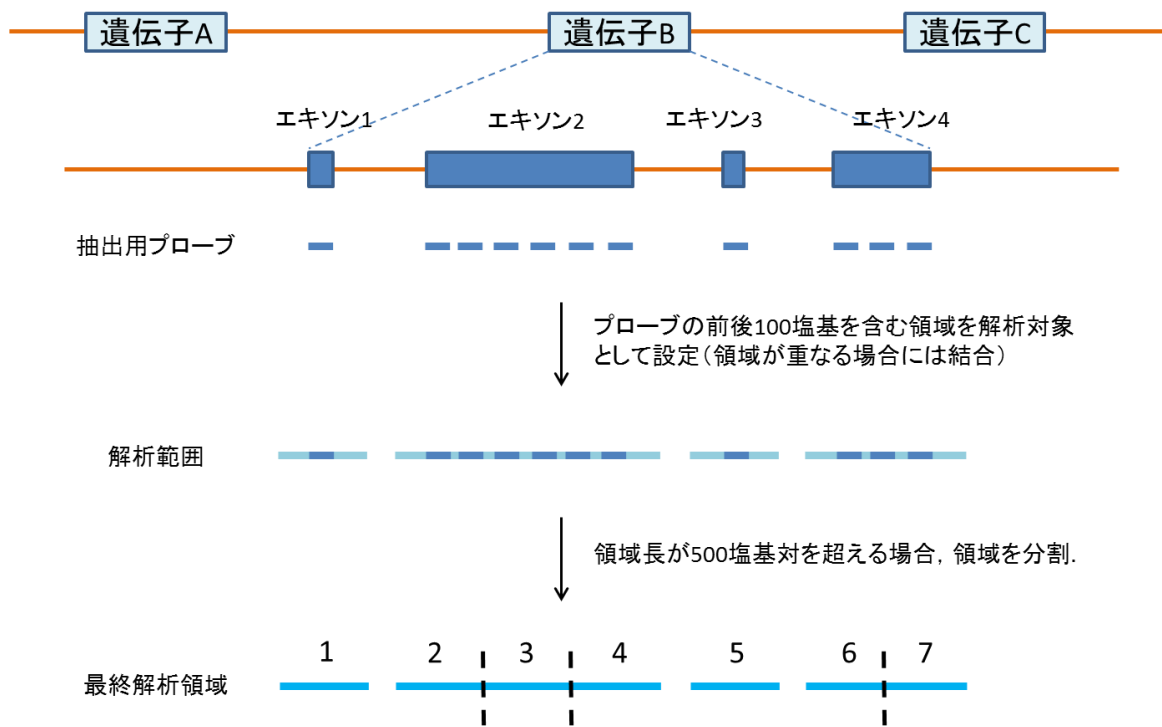


図 3-2. リード数を集計する解析領域の設定方法

3-3 領域毎のリード数および RPM 値の計算方法について

各症例について、マッピング後（BWA 処理後）の bam ファイルから個々の配列データを抽出し、ペアリードの開始位置と終端位置の情報からリードの中心位置を求め（小数点以下切り捨て）、中心位置が存在する領域を、該当リードの由来する領域として判定した（図 3-3）。なお、ペアリードのうちいずれか一方でも正しくマッピングされていないもの、およびリード全体の長さが著しく大きい（500bp 以上の）リードは、解析対象から除外した。

続いて、領域毎のリード数を症例間で比較する際に、解析総リード数のバラつきによる影響をなくすため、以下の計算式により各領域の RPM 値（Reads Per Million mapped reads）を算出した。なお、性別の違いによる影響を避けるため、常染色体の領域内に含まれるリード数の合計を分母として用いた。

$$\text{RPM 値} = (\text{領域内のリード数}) / (\text{常染色体の総リード数}) \times 10^6$$

また、常染色体の領域内に含まれるリード数の合計を元に算出した、解析症例毎の平均カバレッジの分布を図 3-4 に示す。

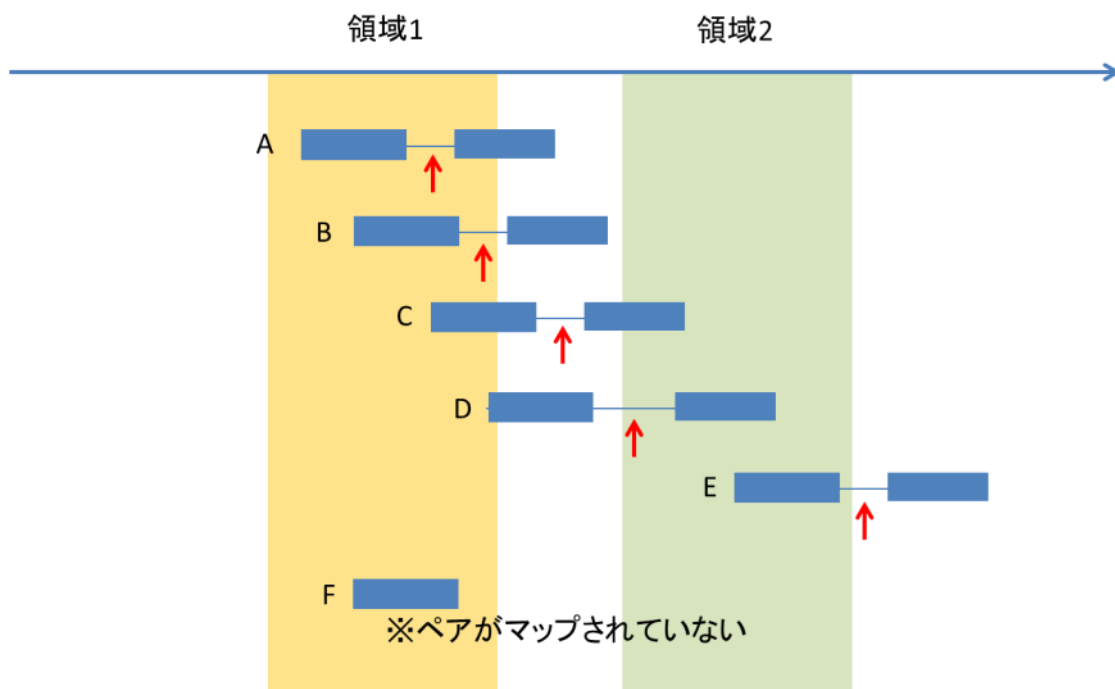


図 3-3. マッピングされたリードがどの領域由来のリードか判定する方法について図示したもの。青い長方形の部分は次世代シーケンサーによって読み取られた配列領域（100 塩基長）を示し、ペアとなっている配列領域同士を実線をつないで図示している。赤矢印はリード全体の中央位置を示す。

上記のうち、リード A、B については領域 1 由来のリードと判定される。リード C、D についてはいずれも配列の一部が領域 1 および領域 2 に含まれているが、リードの中央位置で判断するため C はいずれの領域にもカウントされず、D は領域 2 由来としてカウントされる。E はペアリードの一方が領域 2 に完全に含まれるが、中央位置が領域 2 の外にあるためカウントされない。F はリードの一方が領域 1 に完全に含まれるが、ペアリードがマップされていないため中央位置が判定できず、カウントされない。

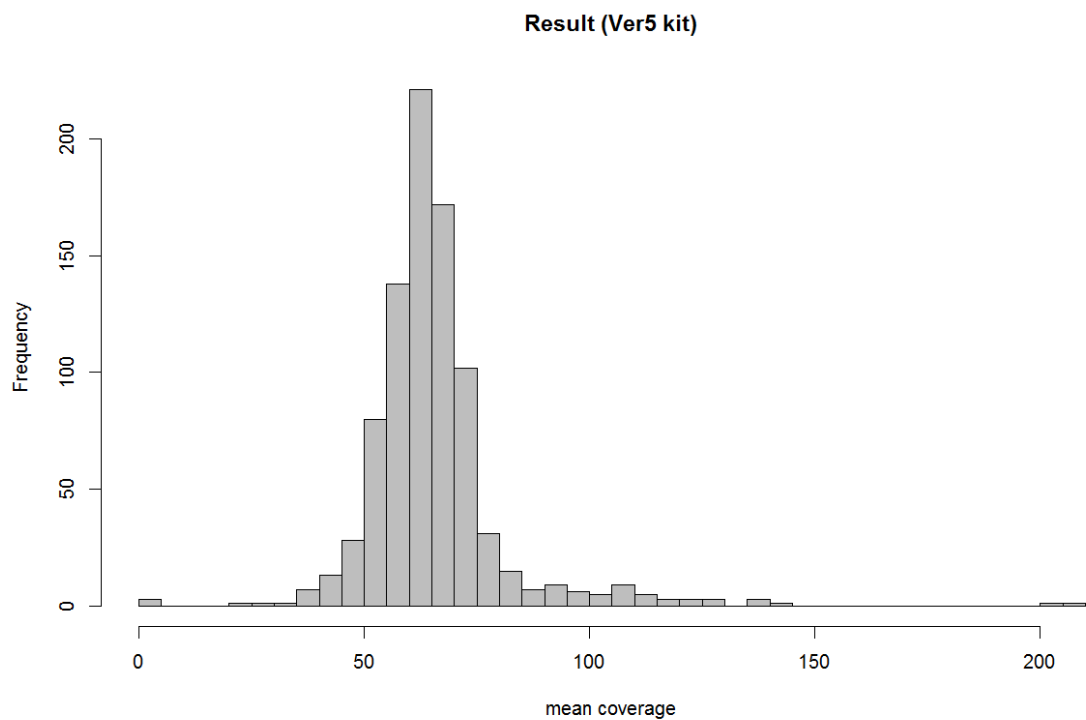
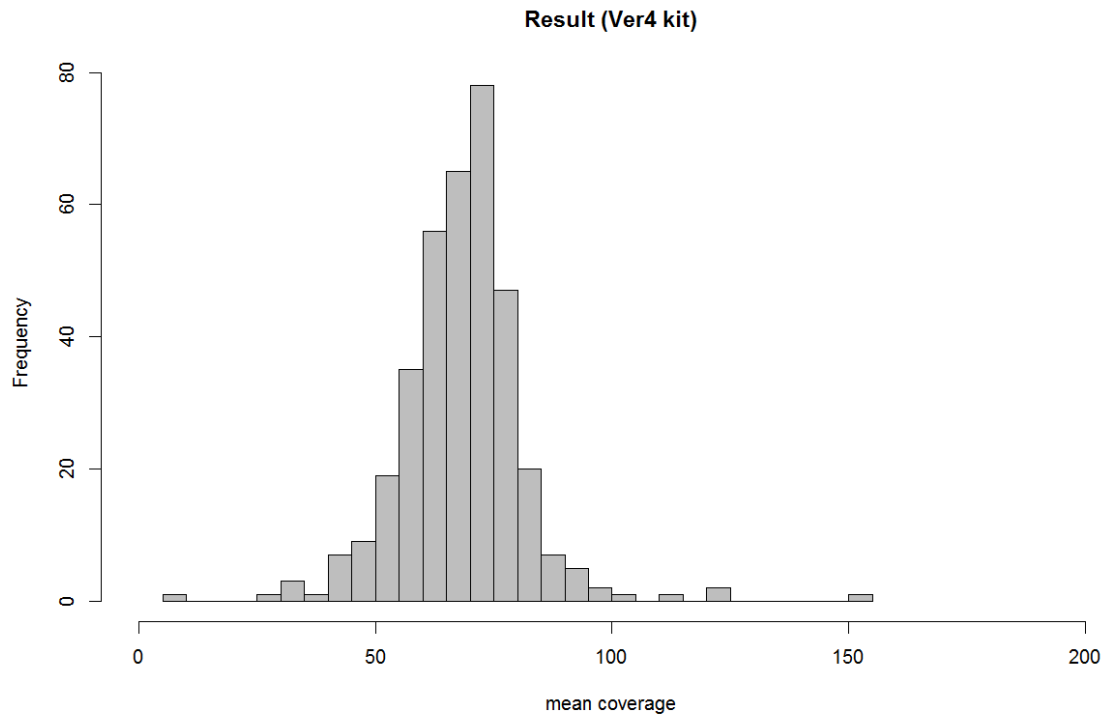


図 3-4. V4 キット（上）および V5 キット（下）における，解析症例毎の平均カバレッジの分布．なお平均カバレッジは，（常染色体の領域内に含まれるリード数の合計）／（常染色体の領域の総塩基長）× 200 により算出した．

3-4 リード数の分布が著しく異なる症例の解析対象からの除外

構造変異の検出を行う前に、平均値からのバラつきが著しい症例の削除を行った。まず、領域毎に RPM 値の平均値と標準偏差を求め、症例毎に以下の計算式を用いて RPM 値の Z-score (Z-RPM 値) を算出した。

$$\text{Z-RPM 値} = (\text{症例の RPM 値} - \text{平均 RPM 値}) / (\text{RPM 値の標準偏差})$$

Z-RPM 値の常染色体の全領域における標準偏差を、各症例の領域全体に渡るバラつきの大きさの指標として使用することとした。使用したライブラリ調整キットごとに、全症例についてデータを計算し、中央値の 2 倍をカットオフ値として、リードの分布が著しく異なる症例の削除を行った (図 3-5)。結果、V4 キットでは 361 例中 9 例 (2.5%)、V5 キットでは 869 例中 13 例 (1.5%) が解析除外となった。なお、削除された症例の多くは、全リード数が著しく少ない症例であった。

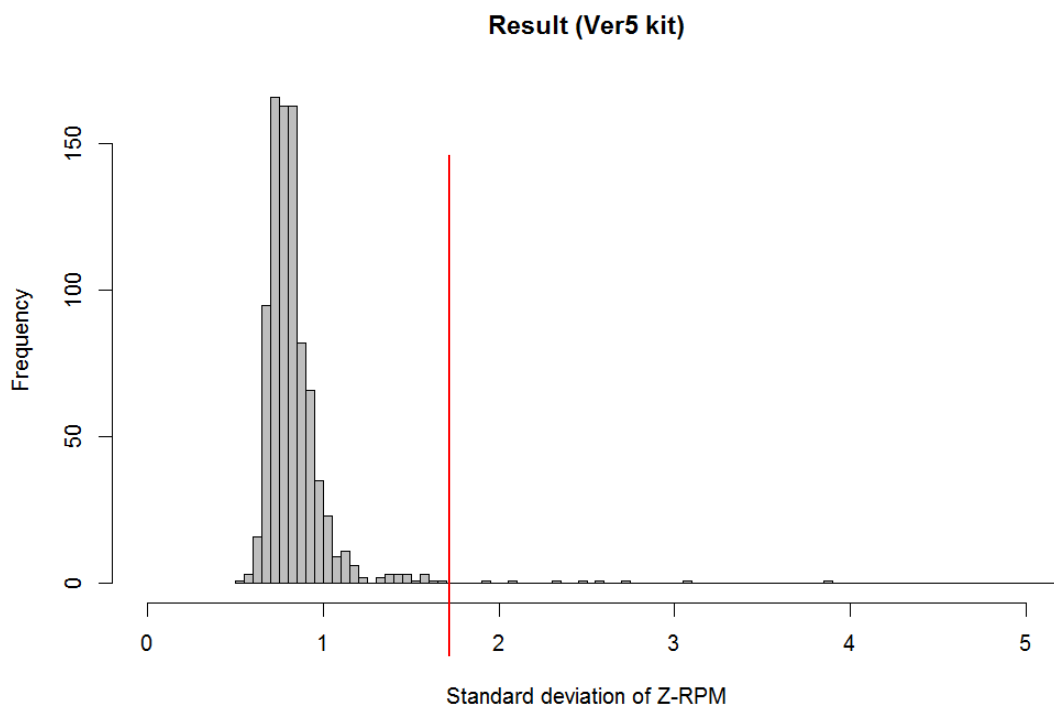
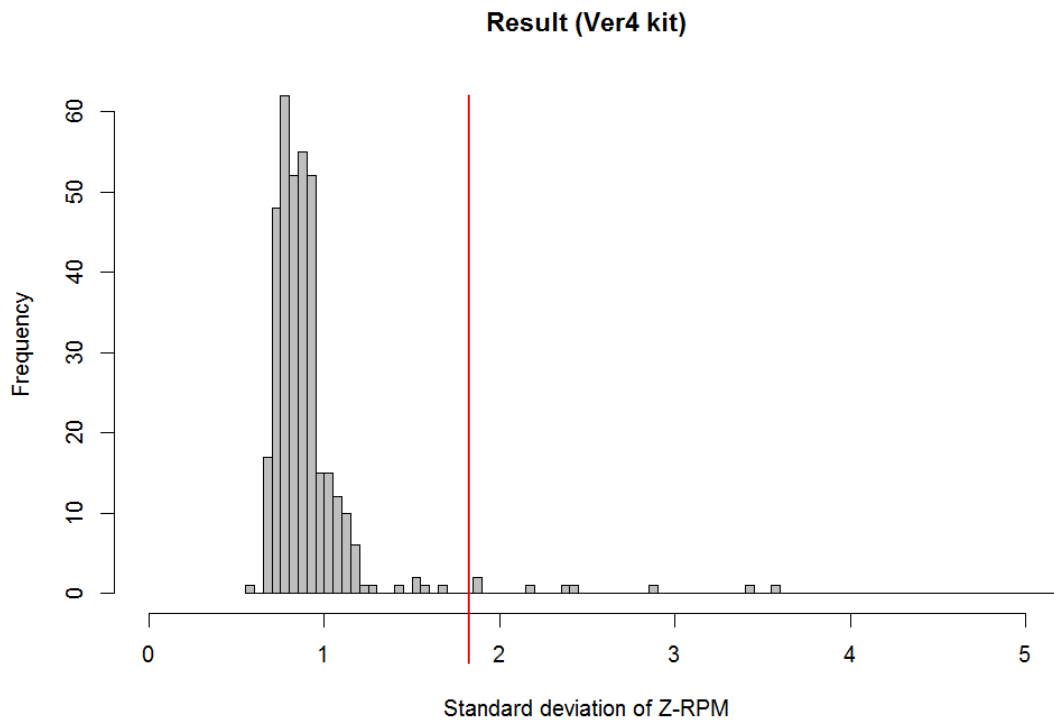


図 3-5. V4 キット（上）および V5 キット（下）における、Z-RPM 値の標準偏差の分布。中央値 \pm nMAD (normalized median absolute deviation) は、V4 キットで 0.85 ± 0.12 、V5 キットで 0.80 ± 0.10 。赤線はカットオフ値（中央値の 2 倍）。

3-5 性染色体の RPM 値に基づくコピー数判定

X 染色体および Y 染色体のコピー数の判定は、X 染色体にも Y 染色体にも共通して存在する領域である、Pseudoautosomal region (PAR) [28]を除いた領域について RPM 値を合計し、症例間で値を比較することにより行った。

以下に、Pseudoautosomal region に存在するデータの具体的な削除方法を述べる。短腕側に存在する PAR1 に関しては、X 染色体に関しては XG 遺伝子が PAR1 と X 染色体固有の領域にまたがって存在すると報告されており [29]、XG 遺伝子の Exon3 までのデータを除外し、XG 遺伝子の Exon4 以降のデータを X 染色体固有の領域として評価対象に含めた。Y 染色体では CD99 遺伝子までのデータを除外し、SRY 遺伝子以降のデータを Y 染色体固有の領域として対象に含めた。また、長腕側に存在する PAR2 に関しては、X 染色体、Y 染色体ともに SPRY3 遺伝子以降のデータを PAR2 に含まれる領域として除外した[28]。

各症例について Pseudo autosomal region (PAR) を除いた X 染色体および Y 染色体全体の RPM 値を算出し、グラフとして表示したものを図 3-6、図 3-7 に示す。

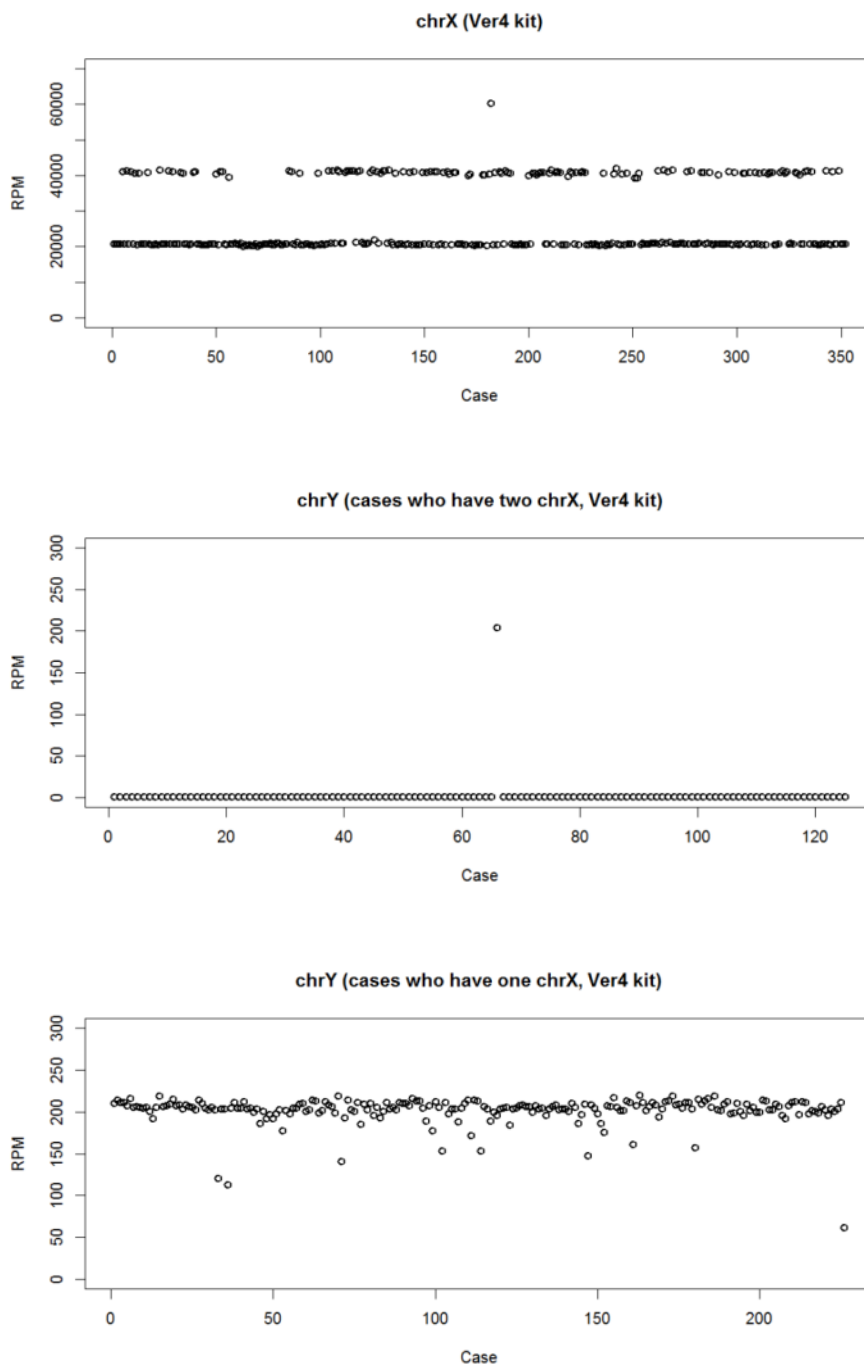


図 3-6. V4 キットを使用した症例における，chrX および chrY 全体（pseudo autosomal region を除く）の RPM 値を示している．上：chrX の RPM 値を図示したもの．アレル数 1，2 と考えられる症例以外に，アレル数 3 の症例が 1 例認められる（XXX）．中：chrX のアレル数が 2 と判定された症例の chrY の RPM 値．chrY のアレル数が 1 と考えられる症例が 1 例ある（XXY）．下：chrX のアレル数が 1 と推定された症例の chrY の RPM 値．Y 染色体のモザイククロスによると考えられる RPM 値のバラつきを認める．

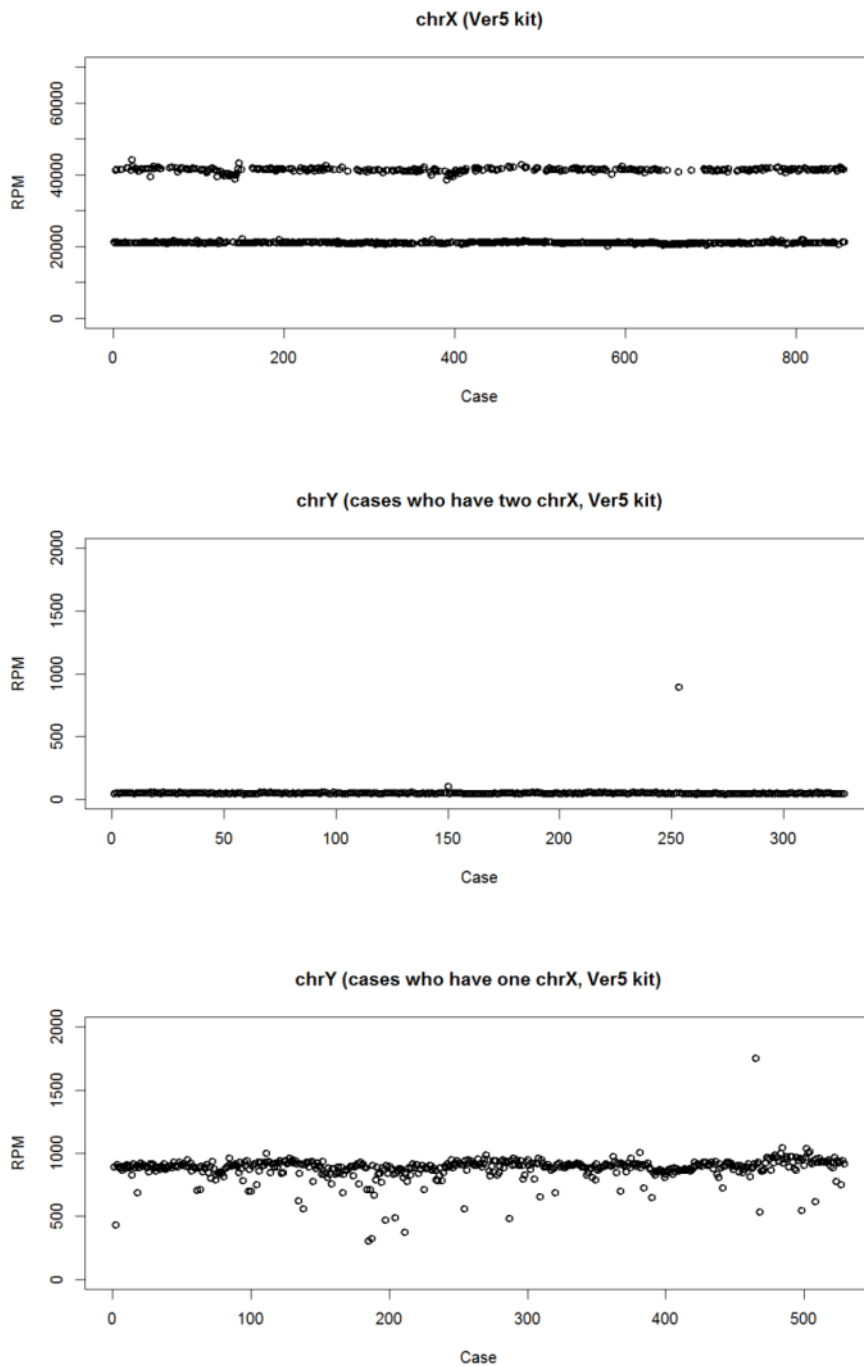


図 3-7. V5 キットを使用した症例における, chrX および chrY 全体 (pseudo autosomal region を除く) の RPM 値を示している. 上: chrX の RPM 値を図示したもの. 中: chrX のアレル数が 2 と判定された症例の chrY の RPM 値. 1 例で chrY のアレル数が 1 と考えられる (XXY). 下: chrX のアレル数が 1 と推定された症例の chrY の RPM 値. 1 例で chrY のアレル数が 2 と考えられる (XYY) .

X 染色体に関するアレル数の判定は，図 3-6，図 3-7 に示す通り，明確に判定することが可能であった．また，Y 染色体に関しては，男性例（X 染色体のアレル数が 1 であり，核型 XY が推定される症例）において RPM 値のバラつきが目立ったが，これは血球細胞においては Y 染色体のモザイクロスが生じることが報告されており[30]，Exome 解析で血球由来の DNA を使用しているため，それを反映した結果と考えられた．

3-6 網羅的な構造変異解析パイプライン

Z-RPM 値の標準偏差に基づき、平均値からのバラつきが著しい症例を除外した後の 1208 例 (V4 キット 352 例, V5 キット 856 例) について、構造変異検出用のプログラムである **Conifer** を用いて網羅的な解析を行った。解析は使用したキット毎に分けて行い、また常染色体と性染色体は分けて解析を行った。性染色体のうち X 染色体については、性別によるアレル数の違いが影響することを避けるため、前述の方法に基づいてアレル数毎 (アレル数 1, アレル数 2) に分けて解析を行った。アレル数が 3 の症例 (1 例) については解析から除外した。また、Y 染色体については **Conifer** を用いた解析は施行しなかったが、これは血球細胞における Y 染色体のモザイクスの影響で、症例ごとに全染色体数に占める Y 染色体の割合が変化するため、データ比較が困難であるためである[30]。

データ解析に関しては、**Conifer** で解析可能な形式とするため、RPM 値を以下のようにして RPKM 値 (Reads Per Kilobase per Million mapped reads) に変換し、**Conifer** プログラムでデータの読み込みを行った。

$$\text{RPKM 値} = \text{RPM 値} / (\text{領域の塩基長 [bp]} \times 1000)$$

Conifer による解析は、まず常染色体について行った。常染色体の解析において、検出された構造変異数の中央値（ \pm normalized median absolute deviation）は、V4 キットで 8（ \pm 3.0）、V5 キットで 19（ \pm 5.9）であった。検出された領域数の分布を図 3-8 に示す。図に示す通り、一部の症例は検出された領域数が他の症例と比べて著しく多く、解析上のエラーの可能性が高いと考えられたため、カットオフ値を中央値の 5 倍として以降の解析から除外することとした。結果、V4 キットでは 13 例、V5 キットでは 25 例が解析から除外され、Conifer 解析前に除外された症例と合わせると、V4 キットでは 361 例中 22 例（6.1%）、V5 キットでは 869 例中 38 例（4.4%）が解析から除外された。それ以外の 1170 例（V4 キット 339 例、V5 キット 831 例）について、X 染色体についての解析を行った後、症例毎に構造変異の検出された領域に含まれる遺伝子を確認し、診断病名などの臨床情報と一致する、既知の疾患原因遺伝子が含まれるかどうかについて確認した（図 3-9）。

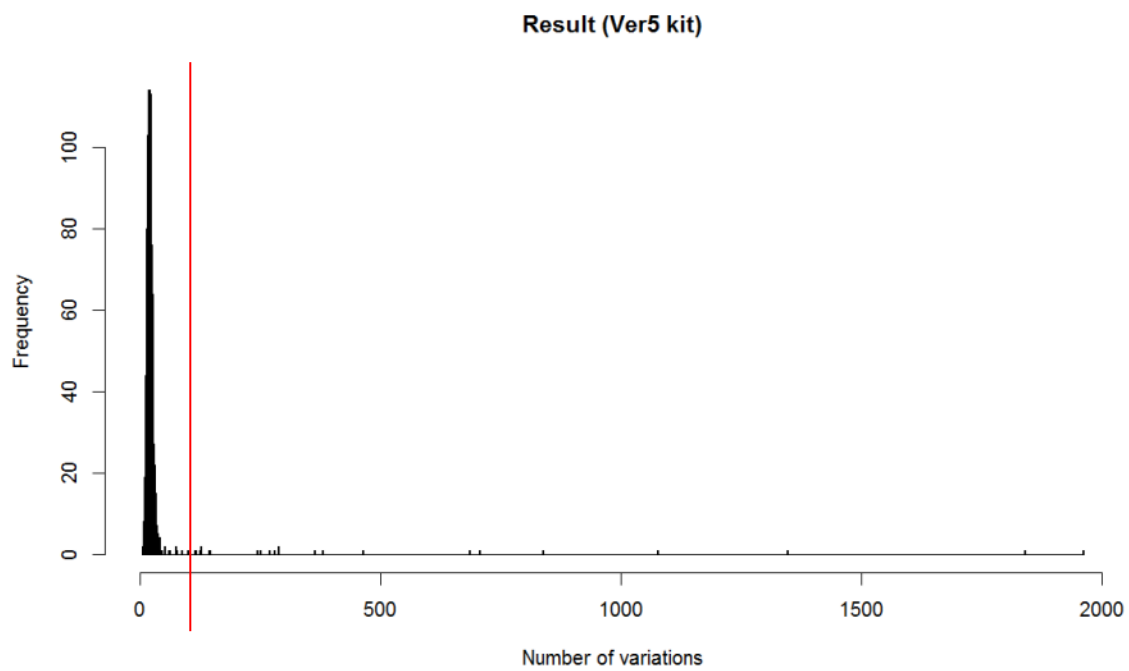
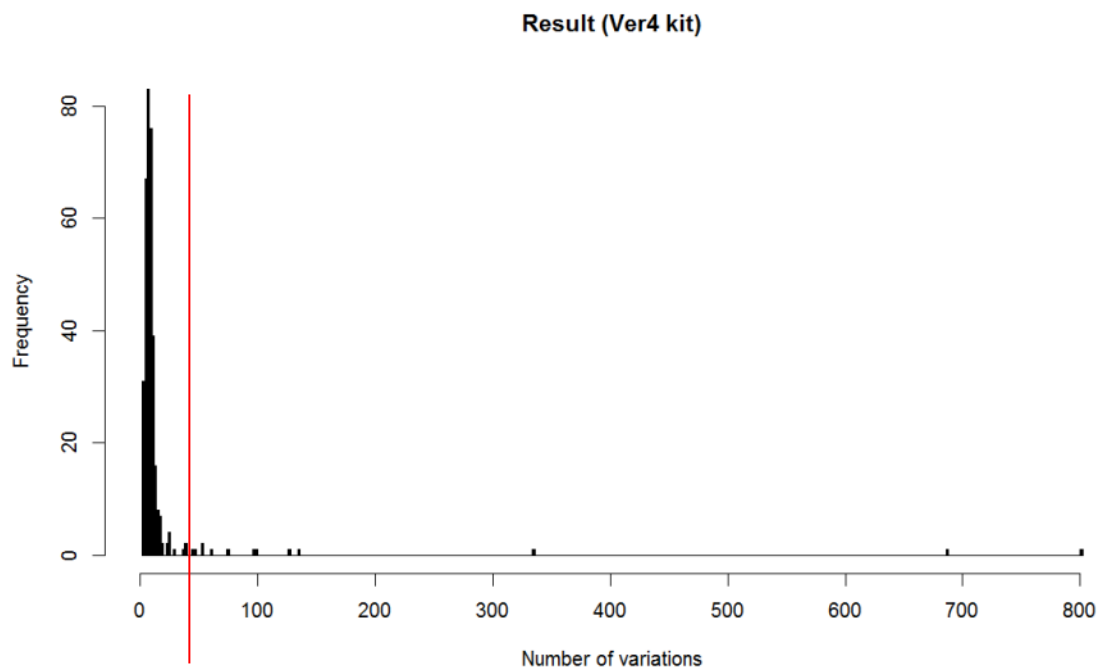


図 3-8. V4キット (上), V5キット (下) を用いて解析した症例における, Conifer 解析後の構造変異数の分布. 赤線はカットオフ値 (中央値の 5 倍).

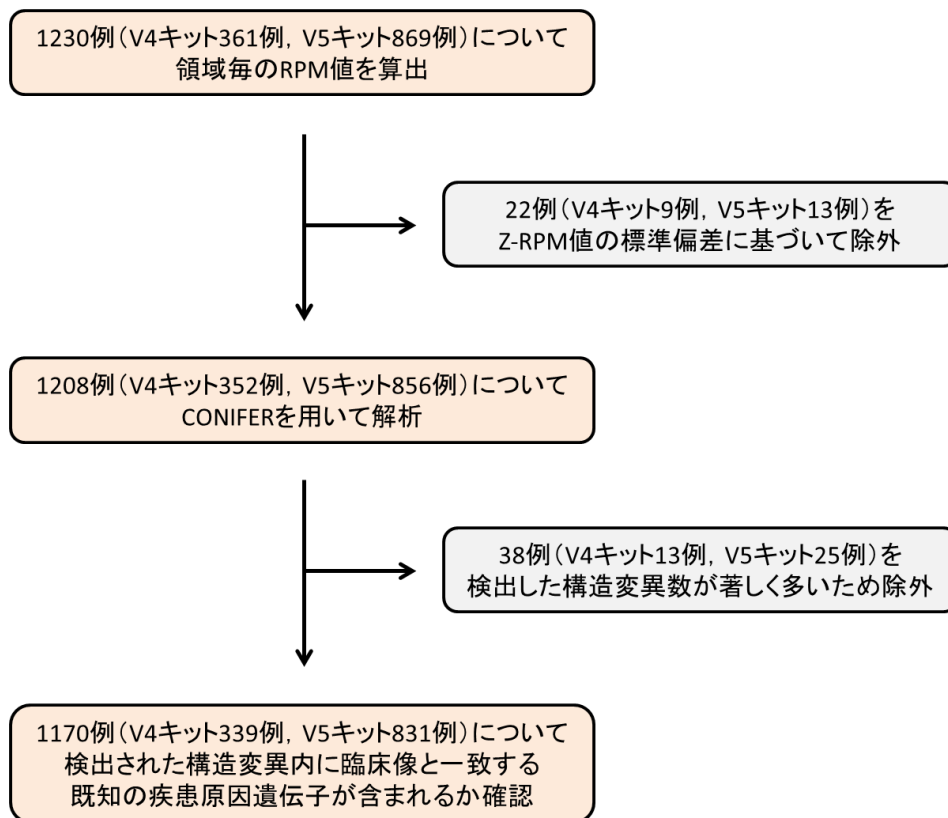


図 3-9. 本解析における除外症例数についてのまとめ.

3-7 アレイ CGH 法により得られた解析結果と RPM 値に基づくデータの比較

アレイ CGH 法で結果として出力されるデータは、Cyanine 3 および Cyanine 5 の色素強度の比率を、2 を底とする対数値に変換したものである。

出力値： \log_2 （患者由来 Cy-5 色素強度／コントロール由来 Cy-3 色素強度）

上記のデータとエクソーム解析データをもとにした解析結果を比較できるようにするため、各領域の平均 RPM 値をコントロールデータとして使用し、以下の変換を行ってアレイ CGH 法により得られたデータと比較できるようにした。なお、RPM 値が低い領域はデータのバラつきが大きくなるため、平均 RPM 値が 0.5 未満の領域データは描画に用いなかった。

変換後の値： \log_2 （患者の RPM 値／平均 RPM 値）

3-8 SMN1 遺伝子の RPM 値に基づく構造変異検出

SMN1 遺伝子については、配列一致率の高い偽遺伝子である SMN2 遺伝子が存在すること、および健常人での SMN1、SMN2 遺伝子のアレル数変異の頻度が高いことから、Conifer による変異の検出はできず、異なる解析手法を考案した。

これまでの報告で、SMN1 遺伝子と SMN2 遺伝子の配列は、IVS 6 から Exon 8 に存在する 5 ヶ所の塩基配列のみに相違があると報告されている[19] (図 3-10)。うち 2 ヶ所が Exon 領域 (Exon 7 および Exon 8) に存在しており、これを用いて Exome 解析により検出された配列情報を、SMN1 由来のリードと SMN2 由来のリードに分け、Exon 7 および Exon 8 の RPM 値を算出し解析することとした。具体的には、マッピング後 (BWA 処理後) の bam ファイルから、Exon7 もしくは Exon8 の塩基相違部位の配列を含むリードを抽出し、その塩基情報をもとにして、SMN1 遺伝子由来か SMN2 遺伝子由来かを判定し、それぞれの RPM 値を算出し、グラフ化してアレル数の推定を行った。

V4 キットには SMN1 および SMN2 遺伝子が解析対象として含まれておらず、V5 キットを使用した症例のみについて検討した。なお V5 キットでは、Exon 7 の塩基相違部位を含むプローブは 4 つあり、SMN1 遺伝子に一致するプローブが 2 つ、

SMN2 遺伝子に一致するプローブが2つであった。また、Exon 8 の塩基相違部位を含むプローブは2つで、2つとも *SMN1* 遺伝子に一致するプローブであったが、DNA キャプチャー用のプローブ長が十分長いため、一塩基置換変異の存在によるキャプチャー効率への影響は小さいとされている。

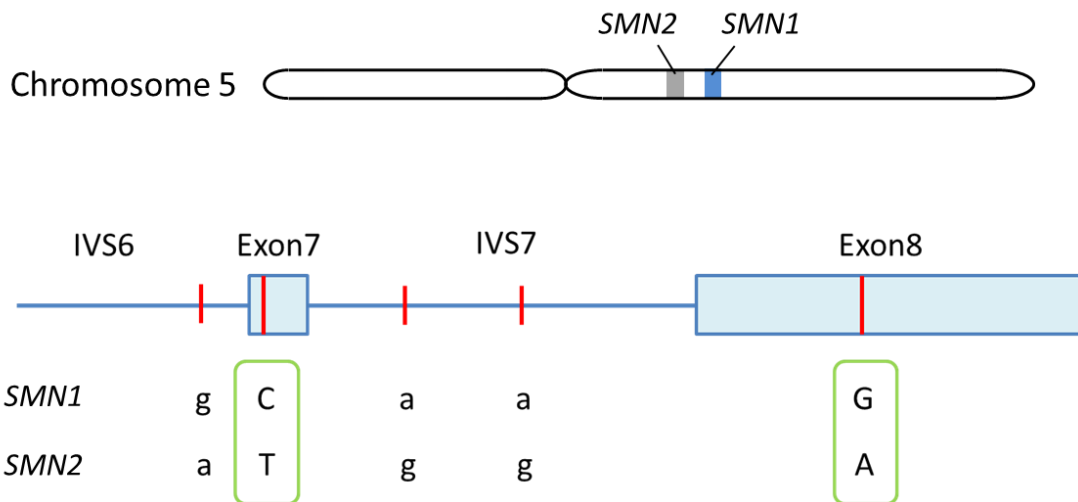


図 3-10. *SMN1* 遺伝子および *SMN2* 遺伝子において、2 者の弁別に使用可能な塩基相違部位を示したもの。うち 2 ヶ所が Exon 内に存在する。

Exome 解析での参照配列である GRCh37/hg19 において、*SMN1* 遺伝子は chr5:70,220,767-70,248,838、*SMN2* 遺伝子は chr5:69,345,349-69,373,421 にマップされており、Exon 7 の塩基相違部位 (C/T) は *SMN1* では chr5:70,247,773、*SMN2* では chr5:69,372,353 にあたり、Exon 8 の塩基相違部位 (G/A) は *SMN1* では chr5:70,248,501、*SMN2* では chr5:69,373,081 にあたる。

第4章 網羅的な構造変異の解析

4-1 検出された構造変異数および構造変異内に含まれる遺伝子数

不適切症例削除後の 1170 例 (V4 キット 339 例, V5 キット 831 例) について, 常染色体および X 染色体において検出された構造変異数 (中央値 \pm normalized median absolute deviation) は, V4 キットで 8 ± 3.0 , V5 キットで 20 ± 5.9 であり, 構造変異内に含まれる遺伝子の総数 (中央値 \pm normalized median absolute deviation) は, V4 キットで 19 ± 7.4 , V5 キットで 35 ± 11.8 であった (図 4-1, 図 4-2).

また, これまでに神経疾患との関連の既報告がある遺伝子 (約 800 遺伝子, なお *SMN1* 遺伝子など全集団に対する構造変異の頻度が著しく高いものを除く) に限定すると, 検出した遺伝子の数 (平均値) は V4 キットで 0.42, V5 キットで 0.57 であり, 一つ以上の遺伝子が検出された症例は, V4 キットで 339 例中 100 例 (29%), V5 キットで 831 例中 431 例 (40%) であった (図 4-3).

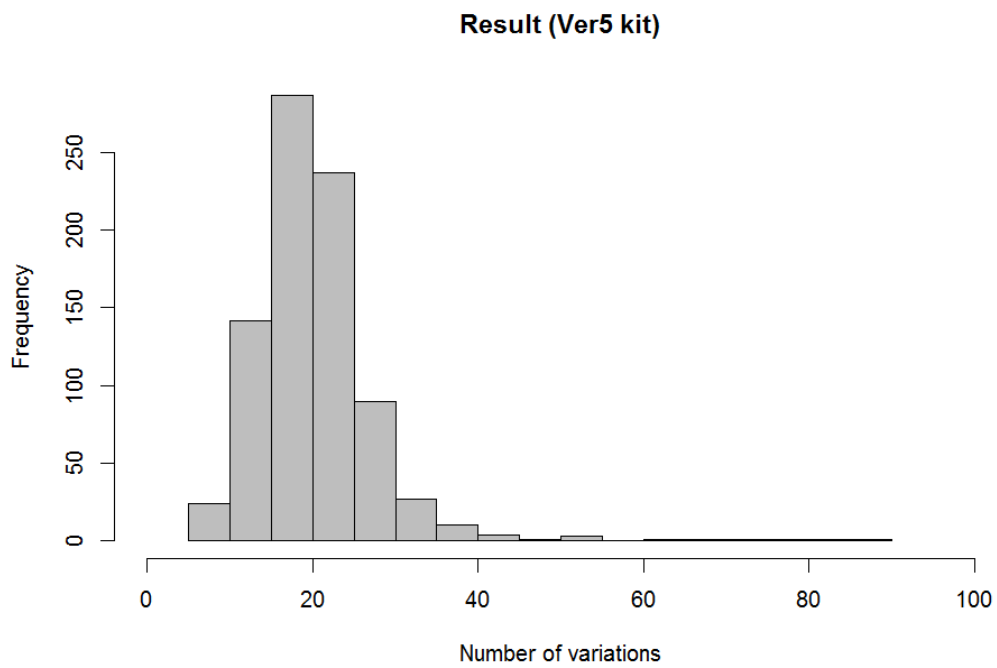
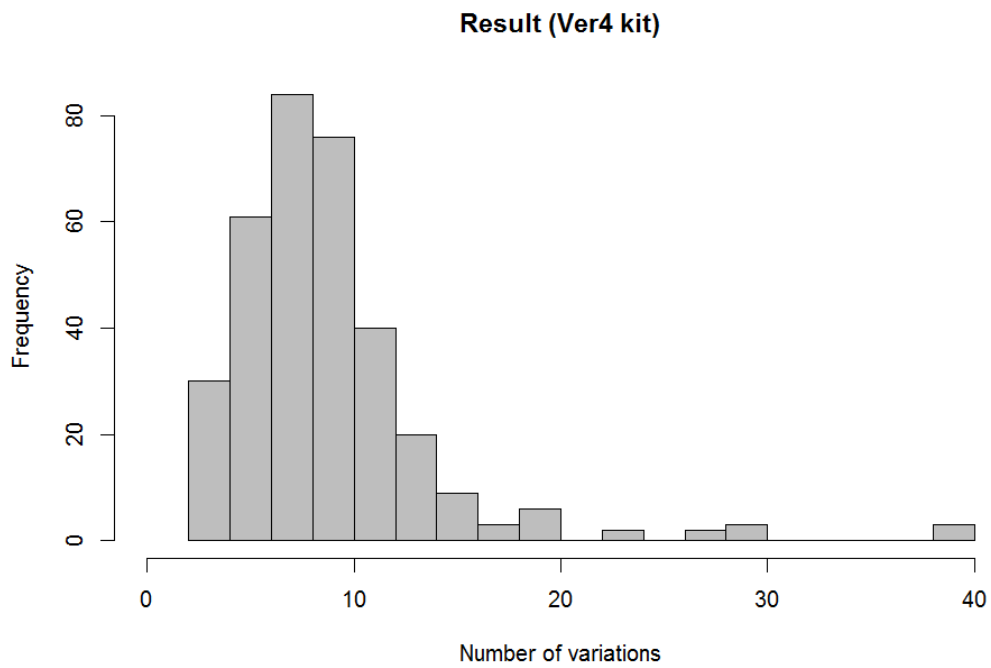


図 4-1. Conifer により検出された構造変異数の分布.

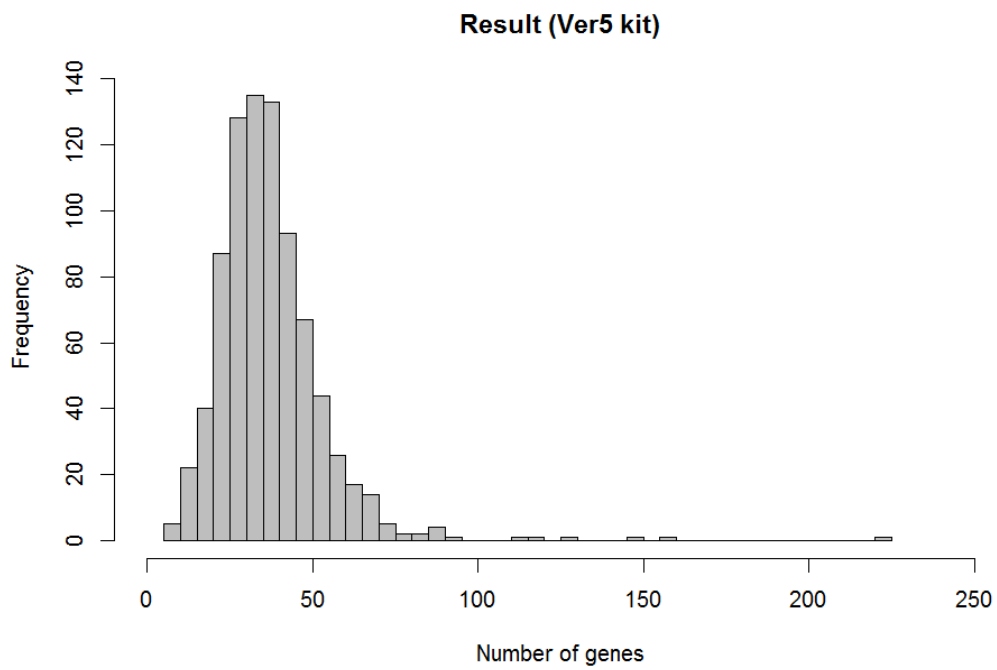
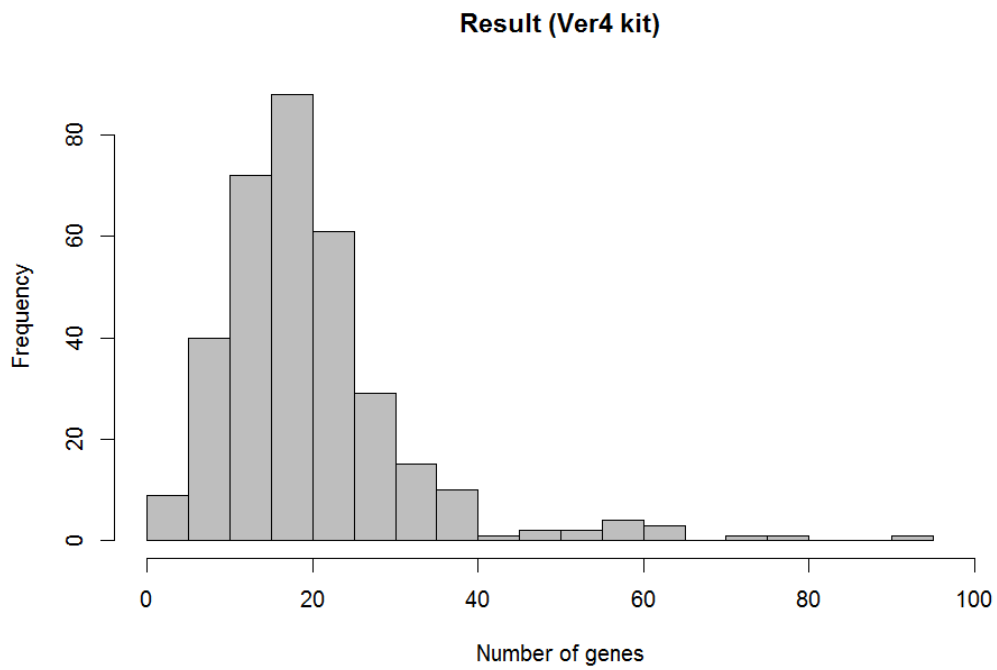


図 4-2. Conifer により検出された構造変異内に含まれる総遺伝子数の分布.

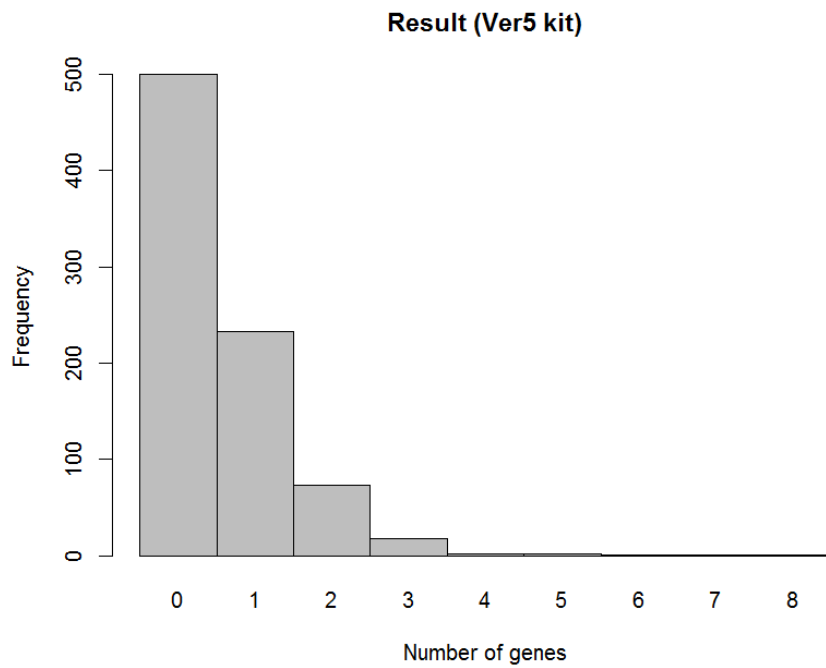
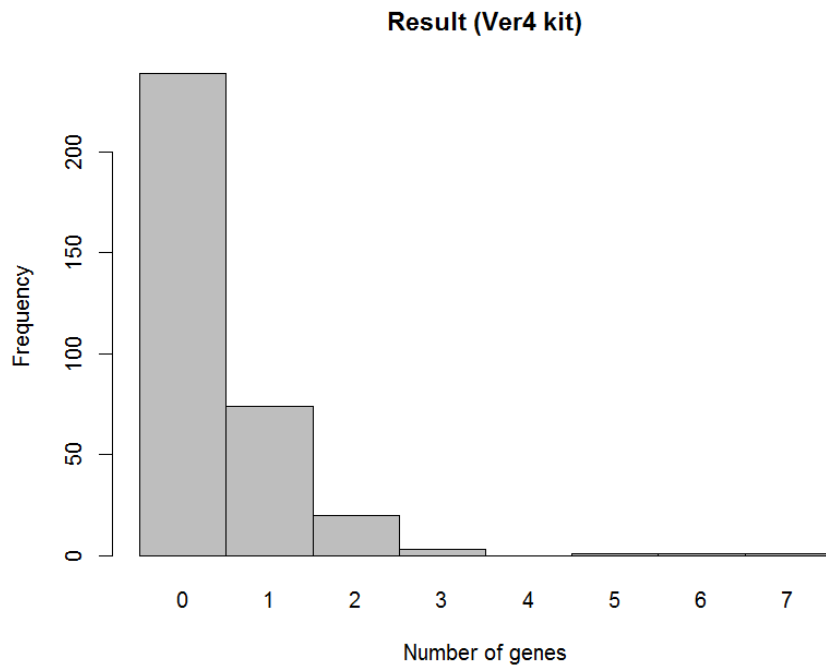


図 4-3. Conifer により検出された構造変異内に含まれる，神経疾患との関連の既報告のある遺伝子の数の分布.

4-2-1 既知の疾患原因遺伝子を含む症例の抽出

Conifer による解析結果を元に、神経疾患との関連の既報告がある遺伝子の一部もしくは全体の RPM 値変動が認められ、遺伝子変異により生じる疾患と臨床情報が一致するものに絞って検索を行った。結果、合計 24 症例において原因遺伝子の可能性のある変異が検出された（表 4-1 から表 4-4）。なお、本解析施行前に変異の存在が判明していた 1 例については、該当の変異が検出された（#2, *PMP22* 重複）。

Case	Kit	Diagnosis	Mutated gene	Disease	Inheritance pattern	Mutation type	Included region	Other rare variant in the gene	Validation by array CGH
#1	V5	CMT	<i>PMP22</i> (NM_000304)	CMT	AD	duplication	entire gene	-	yes
#2	V5	CMT				duplication	entire gene	-	yes
#3	V5	CMT				duplication	entire gene	-	yes
#4	V5	CMT				duplication	entire gene	-	yes
#5	V5	CMT				duplication	entire gene	-	yes
#6	V5	CMT				duplication	entire gene	-	yes
#7	V4	CMT				duplication	entire gene	-	yes
#8	V5	CMT				deletion (heterozygous)	entire gene	-	yes
#9	V5	CMT				deletion (heterozygous)	entire gene	-	yes
#10	V4	CMT				deletion (heterozygous)	entire gene	-	yes
#11	V5	HSP	<i>SPAST</i> (NM_014946)	HSP	AD	deletion (heterozygous)	Exon 1-3	-	yes
#12	V5	HSP				deletion (heterozygous)	Exon 1-4	-	yes
#13	V4	ALS	<i>OPTN</i> (NM_1008211)	ALS	AR/AD	deletion (homozygous)	Exon 6-8	-	yes
#14	V5	ALS				deletion (heterozygous)	Exon 1-7	-	yes

表 4-1. 同様の構造変異の既報告がある症例一覧. なお”other rare variant in the gene”欄に記載した変異の有無の情報は、通常の Exome 解析結果により得られたものである. CMT: Charcot-Marie-Tooth disease, HSP: hereditary spastic paraplegia, ALS: amyotrophic lateral sclerosis, AD: autosomal dominant, AR: autosomal recessive.

Case	Kit	Diagnosis	Mutated gene	Disease	Inheritance pattern	Mutation type	Included region	Other rare variant in the gene	Validation by array CGH
#15	V5	HSP	<i>ZFYVE26</i> (NM_015346)	HSP	AR	deletion (compound heterozygous)	Exon 18-26 del + Exon 22 del	-	yes (also by Sanger sequencing)

表 4-2. 同様の構造変異による疾患発症の報告はないが、遺伝形式が常染色体劣性で同一遺伝子に2つの変異を有する症例.

Case	Kit	Diagnosis	Mutated gene	Disease	Inheritance pattern	Mutation type	Included region	Other rare variant in the gene	Validation by array CGH
#16	V5	CMT+HSP	<i>KIF5A</i> (NM_004984)	CMT/HSP	AD	deletion (heterozygous)	Exon 2-6	-	yes
#17	V4	CMT	<i>LITAF</i> (NM_004862)	CMT	AD	deletion (heterozygous)	Exon 2-4	-	yes

表 4-3. 同様の構造変異による疾患発症の報告はないが、遺伝形式が常染色体優性である症例.

Case	Kit	Diagnosis	Mutated gene	Disease	Inheritance pattern	Mutation type	Included region	Other rare variant in the gene	Validation by array CGH
#18	V5	HSP	<i>SACS</i> (NM_014363)	ARSACS	AR	deletion (heterozygous)	entire gene	c.5151dupA p.S1718Ifs*20	yes
#19	V5	CMT(HSAN)				deletion (heterozygous)	entire gene	c.2402C>T p.P801L	yes
#20	V5	HSP				duplication (heterozygous)	entire gene	-	yes
#21	V5	HSP	<i>FA2H</i> (NM_024306)	HSP	AR	deletion (heterozygous)	Exon 1	c.908G>A p.G303E	yes
#22	V5	HSP	<i>CYP2U1</i> (NM_183075)	HSP	AR	deletion (heterozygous)	Exon 5	-	yes
#23	V4	MSA	<i>PARK2</i> (NM_004562)	PD	AR	duplication (heterozygous)	Exon 2-3	-	yes
#24	V4	CMT	<i>SBF1</i> (NM_002972)	CMT	AR	deletion (heterozygous)	entire gene	-	yes

表 4-4. 遺伝形式が常染色体劣性であるが、検出された変異がヘテロ接合性であった症例.

CMT: Charcot-Marie-Tooth disease, HSP: hereditary spastic paraplegia, HSAN: hereditary sensory and autonomic neuropathy, MSA: multiple system atrophy, ARSACS: autosomal recessive spastic ataxia of Charlevoix-Saguenay.

24 症例のうち、表 4-1 に示した 14 症例については、同様の構造変異により疾患を発症した既報告があり、原因遺伝子変異と考えられた。

表 4-2 に示した 1 症例は、同様の構造変異により疾患を発症した既報告はないものの、遺伝形式が常染色体劣性遺伝で、同一遺伝子に 2 種類の構造変異を有していることから、原因遺伝子変異である可能性が高いと考えられた。また、表 4-3 に示した 2 症例は、同様の構造変異により疾患を発症した既報告はないものの、遺伝形式が常染色体優性遺伝形式であり、原因遺伝子変異である可能性が考えられた。ただし、これら 3 症例について原因遺伝子変異と確定するには、既報告の遺伝子変異を有する症例と臨床的特徴が合致するかどうか、また家系内の発端者以外の遺伝子検査など、追加の確認が必要と考えられた。

表 4-4 に示した 7 症例については、遺伝形式が常染色体劣性であるが、検出された構造変異はヘテロ接合性であった。うち 3 例については健常者集団には認められない **rare variant** が通常のエキソーム解析で検出されており、原因遺伝子変異の可能性があると考えられた。以下、変異の検出された 24 症例について、遺伝子毎に分けて考察する。

4-2-2 PMP22 遺伝子の重複および欠失変異について (症例 #1 - #10)

PMP22 遺伝子の重複はシャルコー・マリー・トゥース病 (CMT1A) の原因遺伝子変異であり[14], 脱髄型の CMT (CMT1) の中で最も頻度が高く, 70%程度を占めるとされる[31]. また *PMP22* 遺伝子の欠失は遺伝性圧脆弱性ニューロパチー (HNPP) を引き起こす[15].

重複, 欠失いずれにおいても, *PMP22* 遺伝子を挟むようにして存在する, 2つの相同性の非常に高い (99%) 塩基配列を介して, 非対立遺伝子間の相同組み換え (non-allelic homologous recombination) が起こることによる変異の頻度が高く[32, 33], その場合には2つの塩基相同性の高い領域に挟まれた共通する 1.4 Mb の領域の重複あるいは欠失が認められる (図 4-4). 一方で, 上記とは異なる位置での組み換えによる重複, 欠失の症例についても少ないながら報告はあり, 機序として NHEJ (non-homologous end joining), MMEJ (microhomology-mediated end joining), FoSTeS/ MMBIR (fork stalling and template swithing/ microhomology-mediated break-induced replication) 等が推定されている[34,35].

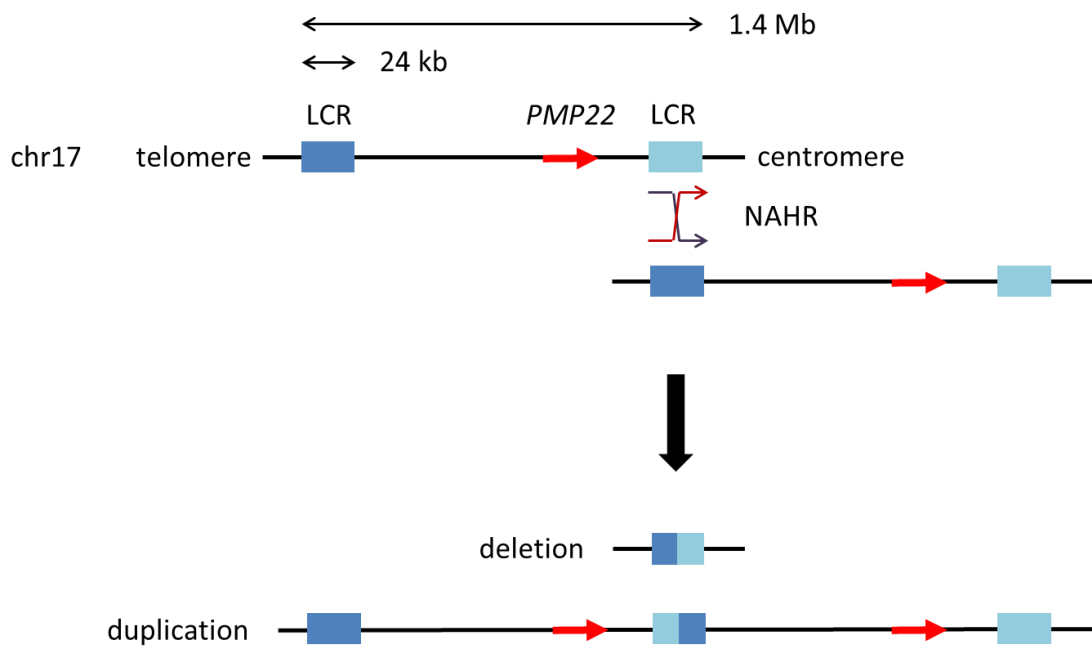


図 4-4. 相同性の高い塩基配列 (low copy repeat) を介した非対立遺伝子間相同組み換えによる, *PMP22* 遺伝子の構造変異生成の模式図. LCR: low copy repeat, NAHR: non-allelic homologous recombination.

今回変異が確認された 10 例について、アレイ CGH 法による解析データも含めて構造変異の範囲について確認した結果では、9 例が相同性の高い塩基配列を介した相同組み換えによる 1.4Mb の重複、欠失と考えられたが、1 例 (#10) は構造変異の範囲が異なっており、異なるメカニズムにより生じた変異と考えられた (図 4-5)。

4-2-3 SPAST 遺伝子の欠失変異について (症例 #11, #12)

SPAST 遺伝子は常染色体優性遺伝の痙性対麻痺 (SPG4) の原因遺伝子である [36]。塩基置換変異、欠失変異ともに多くの種類の変異が報告されており、本邦からも部分欠失変異による発症の報告がある [37]。#11 では Exon 1-3, #12 では Exon 1-4 の欠失が確認され、原因遺伝子変異と考えられる。

4-2-4 OPTN 遺伝子の欠失変異について (症例 #13, #14)

OPTN 遺伝子は ALS の原因遺伝子として報告されており、遺伝形式は常染色体劣性のことも優性のこともあるとされている [38, 39]。今回の解析では 2 例で変異が現出され、#13 はホモ接合性変異の exon 6-8 の欠失、#14 ではヘテロ接合性の exon1-7 の欠失が検出された。

4-2-5 ZFYVE26 遺伝子の欠失変異 (症例 #15)

ZFYVE26 遺伝子は常染色体劣性の痙性対麻痺 (SPG15) の原因遺伝子として報告されているが[40], 過去に欠失変異により本疾患を発症した報告は無い.

#15 では Exon 18 から Exon 26 に渡って RPM 値の低下が認められたが, 内部の Exon 22 のみ RPM 値が 0 となっており, 2 種類の構造変異の複合ヘテロ接合性変異であると考えられた (図 4-5). 同領域についてエキソーム解析のリードデータを見直すと Exon 18 と Exon 26 が結合した配列が検出され, このアリルは 9958 塩基欠失 (chr14: 68242728-68252685) +8 塩基挿入 (tgcaagct) 変異と判明した. これにより, もう一方のアリルは Exon22 のみを含む変異と推定され, Sanger 法により同領域の塩基配列を確認したところ, Exon22 を含む 459 塩基の欠失変異 (chr14:68247969-68247969) であると確認された.

4-2-6 KIF5A 遺伝子の欠失変異 (症例 #16)

KIF5A 遺伝子は常染色体優性の痙性対麻痺 (SPG10) の原因遺伝子であるが[41], 末梢神経障害の合併が特徴で, シャルコー・マリー・トゥース病と診断される症例もあると報告されている[42]. これまでに構造変異による同疾患発症の報告

は無いが、同症例はシャルコー・マリー・トゥース病および痙性対麻痺と診断されており、原因遺伝子変異の可能性が高いと考えられる。

4-2-7 LITAF 遺伝子の欠失変異 (症例 #17)

LITAF 遺伝子は常染色体優性のシャルコー・マリー・トゥース病 (CMT 1C) の原因遺伝子として報告されている[43]。これまでに構造変異による同疾患発症の報告は無い。

4-2-8 SACS 遺伝子の重複および欠失変異について (症例 #18, #19, #20)

SACS 遺伝子は常染色体劣性の遺伝性疾患である ARSACS の原因遺伝子である[44]。SACS 遺伝子全体を含む欠失変異の報告は過去にも複数あり、PMP22 と同様に SACS 遺伝子を挟むようにして 2 つの相同性の高い塩基配列があり、非対立遺伝子間の相同組み換え (NAHR) が生じる結果、約 1.5 Mb の欠失を生じるとされる[45,46]。

#18, #19, #20 における構造変異の範囲はいずれも 2 つの相同性の高い塩基配列間の約 1.5Mb の領域であった (図 4-2)。#18 については通常のエキソーム解析結果で既報告のある一塩基挿入変異 (c.5151dupA) を指摘されており[47]、同

変異をホモ接合性に有することが原因と診断されていたが、今回の解析によって欠失変異と一塩基挿入変異の複合ヘテロ接合性変異であったことが明らかとなった。#19についても、通常のエキソーム解析において健常者で認められないアミノ酸置換を伴う塩基置換変異 (c.2042C>T) が存在していたが、同変異は既報告の無い変異であった。#20 では SACS 遺伝子の重複変異を認めたが、これまで重複変異による ARSACS 発症の報告は無く、SACS 遺伝子に構造変異以外の変異を認めないこと、および同家系は近親婚を有する家系で劣性遺伝形式が疑われていることから、疾患原因変異である可能性は低いと考えられた。

4-2-9 FA2H 遺伝子の欠失変異について (症例 #21)

FA2H 遺伝子は常染色体劣性の痙性対麻痺 (SPG35) の原因遺伝子として報告されており、ジストニアと白質脳症の合併が特徴とされる[48]。また FA2H 遺伝子の Exon 3-7 の欠失と一塩基置換変異の複合ヘテロ接合性変異により、同疾患を発症した報告がある[49]。#21 は欠失変異に加えて、健常者では認められないアミノ酸置換を伴う塩基置換変異 (c.908G>A) を有していたが、これら2つの変異が疾患発症の原因となったかどうかについては、2つの変異が異なるアリルに存在すること (複合ヘテロ接合性であること) の確認、および臨床症状に矛盾が無いかどうかについての確認が必要である。

4-2-10 CYP2U1 遺伝子の欠失変異 (症例 #22)

CYP2U1 遺伝子は常染色体劣性の痙性対麻痺 (SPG 56) として報告されている [50]. これまでに構造変異による同疾患発症の報告は無い. #22 では 1 アリルの遺伝子の一部を含む欠失変異以外に rare variant は認められず, 原因遺伝子変異とは考えにくい.

4-2-11 PARK2 遺伝子の重複変異 (症例 #23)

PARK2 遺伝子は常染色体劣性のパーキンソン病 (*PARK2*) として報告されている. 構造変異の頻度が高い遺伝子であり, 最初に報告された変異も構造変異である [51]. ただし, #23 においては 1 アリルの遺伝子内の一部を含む重複変異のほかに rare variant は認められず, 原因遺伝子変異とは考えにくい.

4-2-12 SBF1 遺伝子の欠失変異 (症例 #24)

SBF1 遺伝子は常染色体劣性のシャルコー・マリー・トゥース病 (CMT4B3) の原因遺伝子として報告されている [52]. #24 では 1 アリルの遺伝子の一部を含む欠失変異以外に rare variant は認められず, 原因遺伝子変異とは考えにくい.

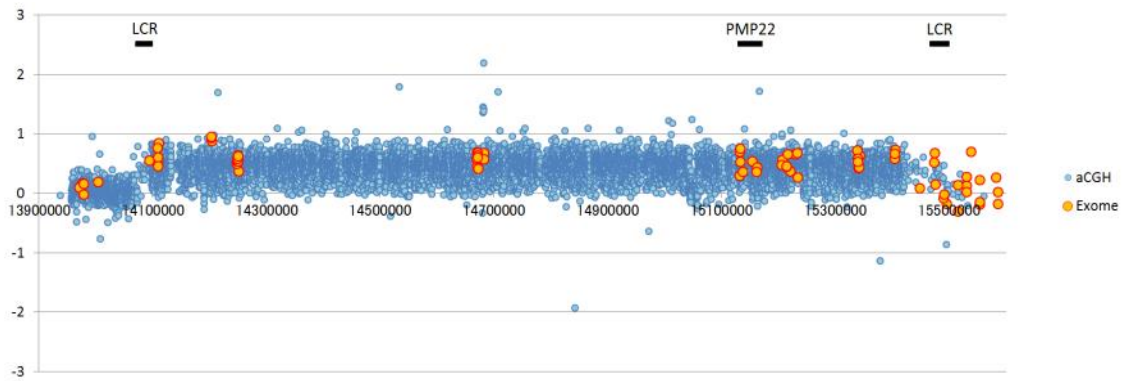
4-3 アレイ CGH 法により得られたデータと RPM 値に基づき算出されたデー

タの比較

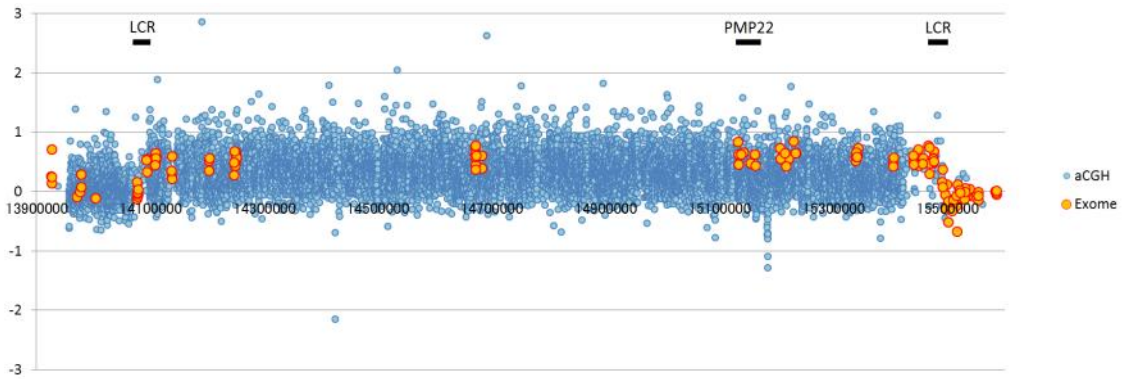
Conifer による解析で構造変異が検出され、アレイ CGH 法により確認を行った症例を用いて、アレイ CGH 法により算出されたデータ（患者由来サンプルとコントロール由来サンプルの蛍光色素強度の比により算出されたデータ）と、RPM 値に基づいて算出したデータ（全患者の平均 RPM 値をコントロールデータとし、患者の RPM 値と平均 RPM 値の比をとり算出したデータ）がどの程度一致しているのかについて確認を行った、

両者の値を同一の単位（2 を底とする log 値）に合わせ、変異ごとに一つのグラフにまとめて表示したものを図 4-5 に示す。いずれにおいても、アレイ CGH 法によるデータと、RPM 値に基づいて算出したデータはよく一致していることがわかる。

<#1 PMP22 dup>



<#2 PMP22 dup>



<#3 PMP22 dup>

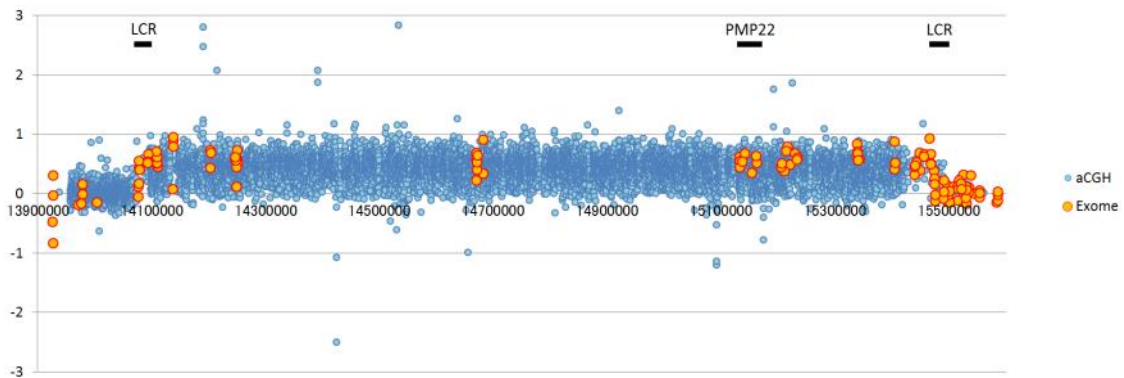
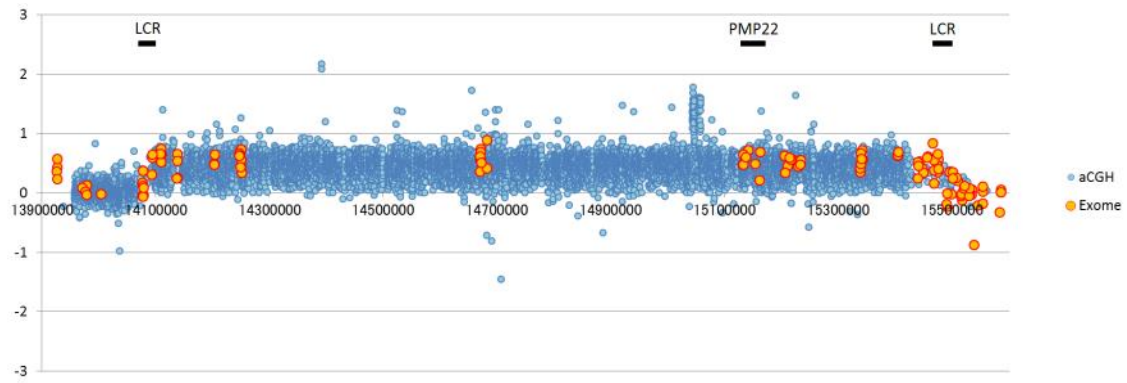
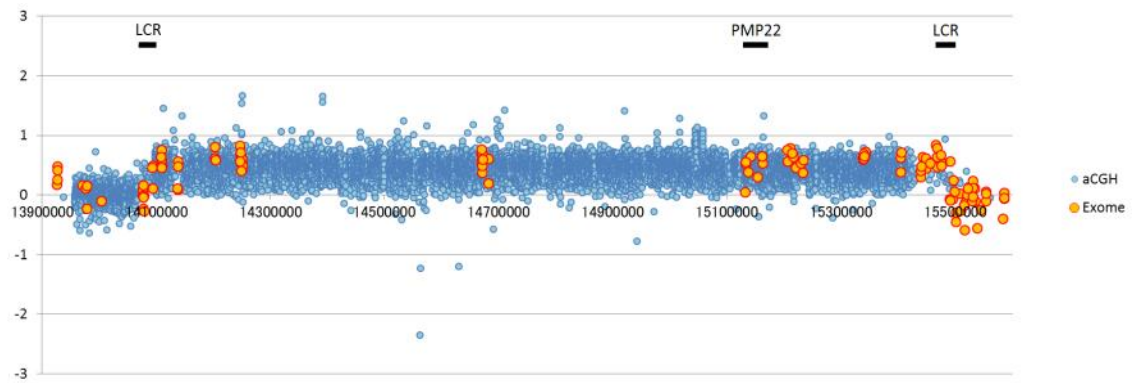


図 4-5. 各症例について、アレイ CGH 法により算出された log 比 (青) と RPM 値データより算出した log 比 (オレンジ色) を比較したもの. 横軸は GRCh37/hg19 上での位置を示す. LCR: low copy repeat.

<#4 PMP22 dup>



<#5 PMP22 dup>



<#6 PMP22 dup>

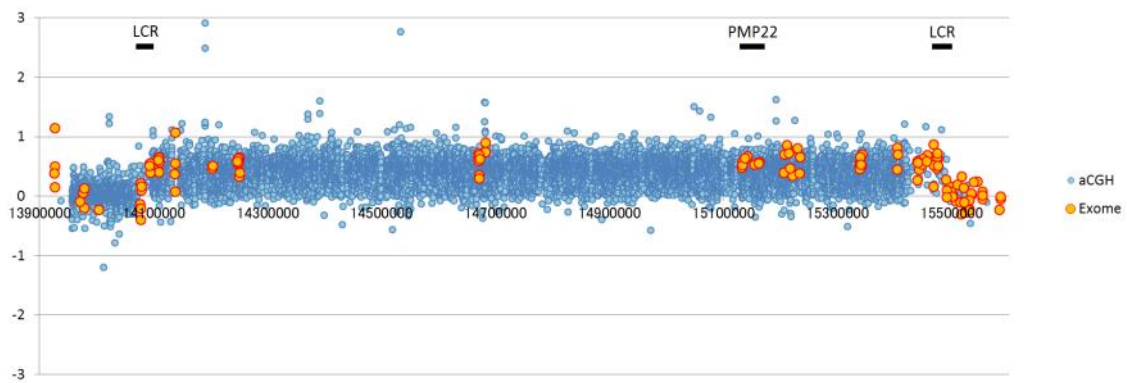
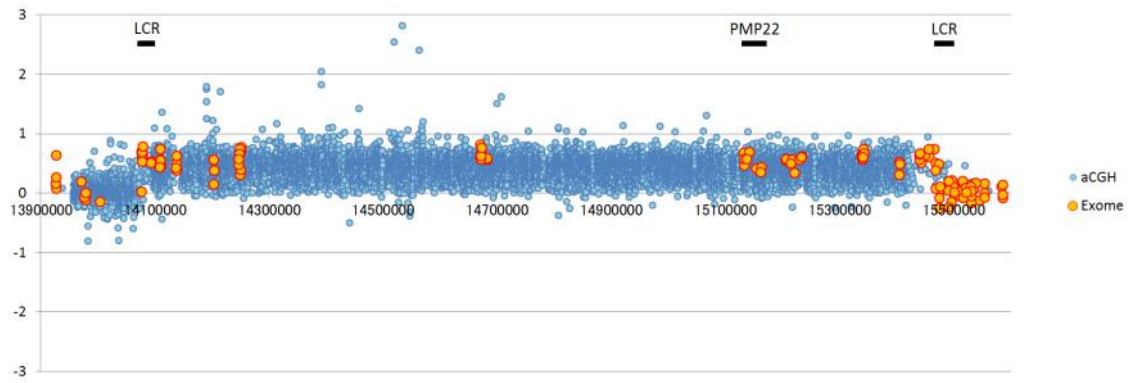
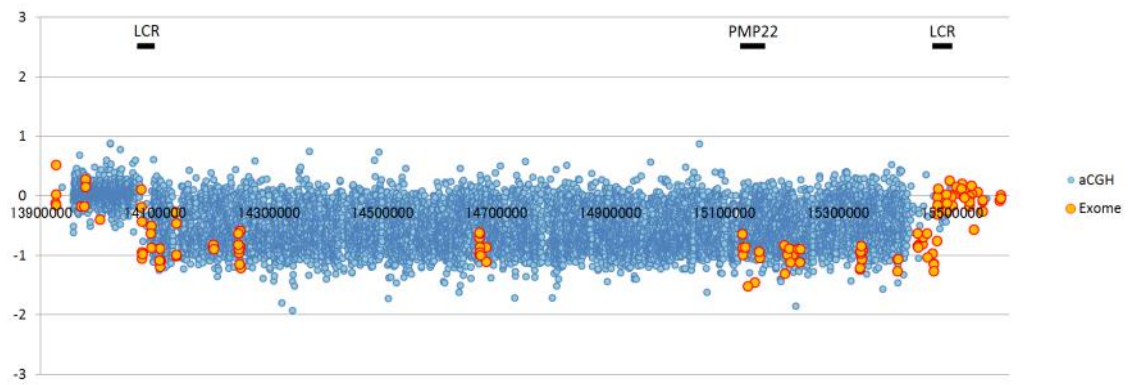


図 4-5. 続き.

<#7 PMP22 dup>



<#8 PMP22 del>



<#9 PMP22 del>

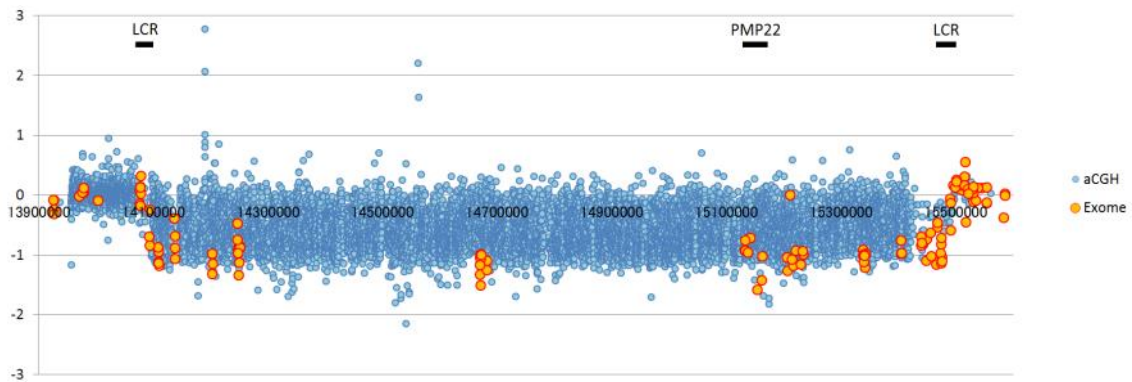
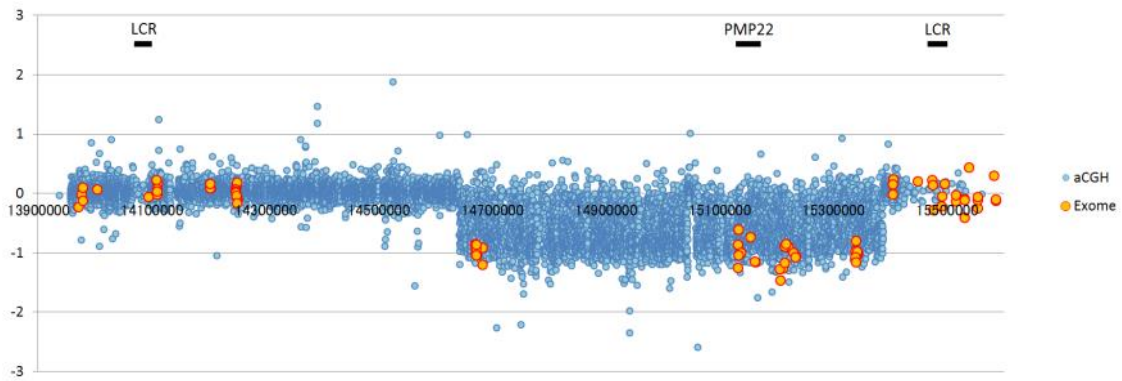
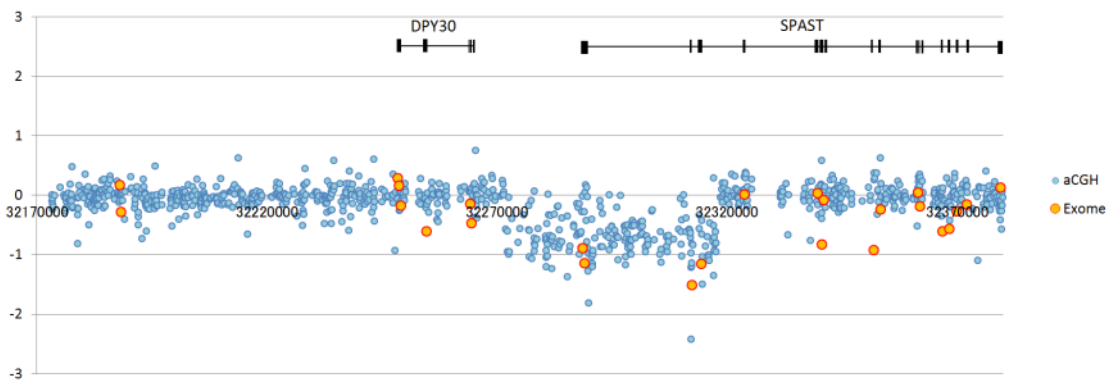


図 4-5. 続き.

<#10 *PMP22* del>



<#11 *SPAST* del>



<#12 *SPAST* del>

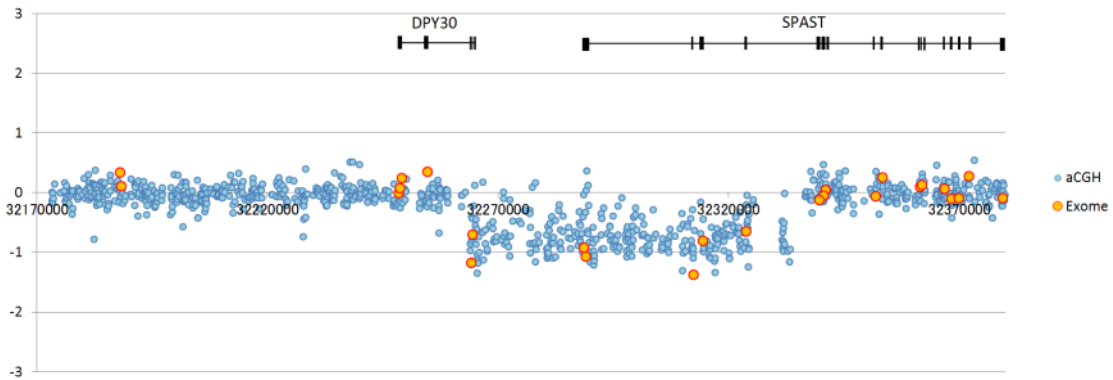
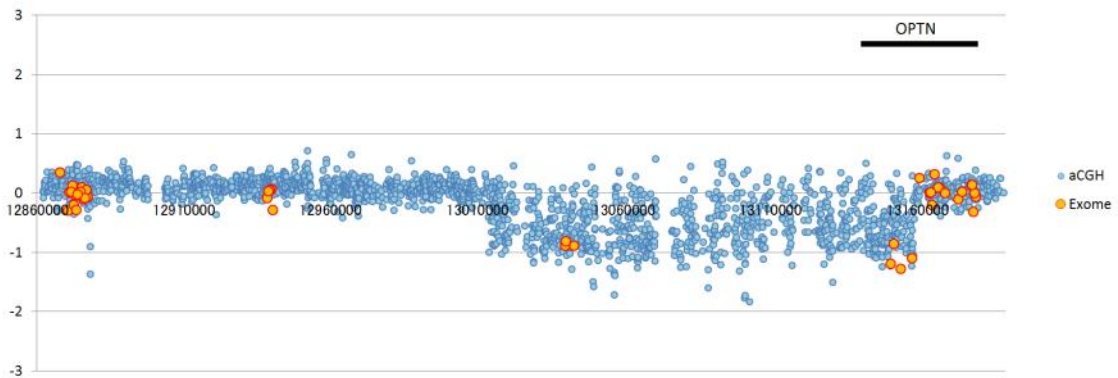
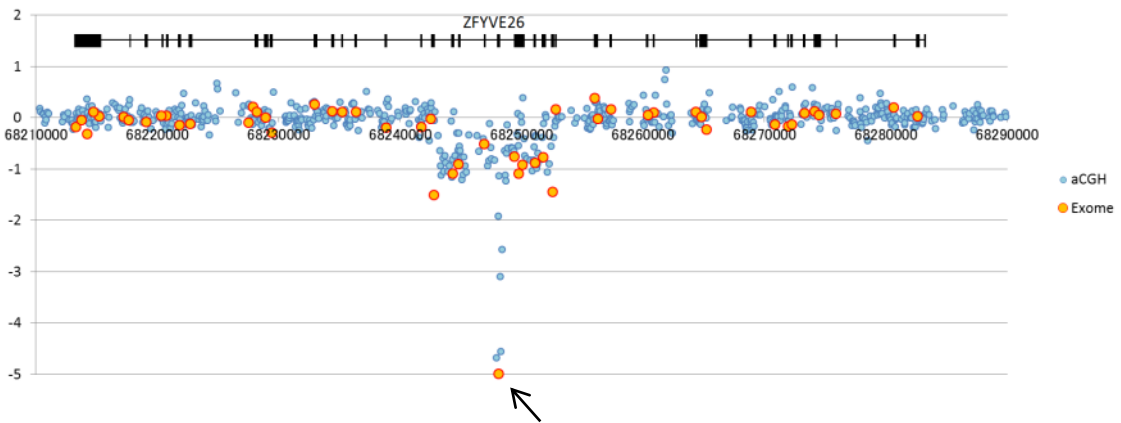


図 4-5. 続き.

<#14 *OPTN* del>



<#15 *ZFYVE26* del>



<#16 *KIF5A* del>

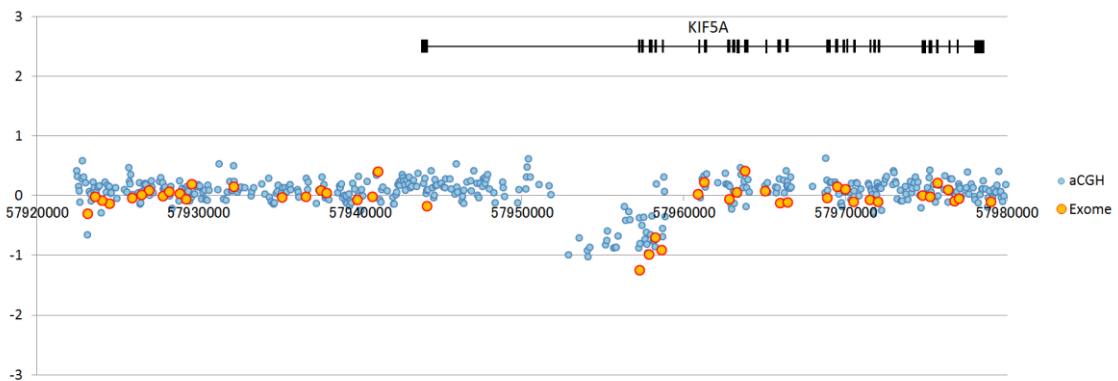
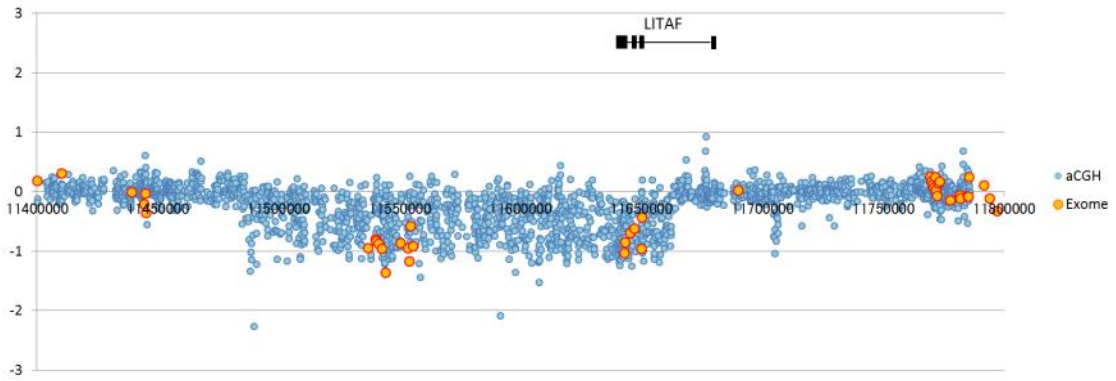
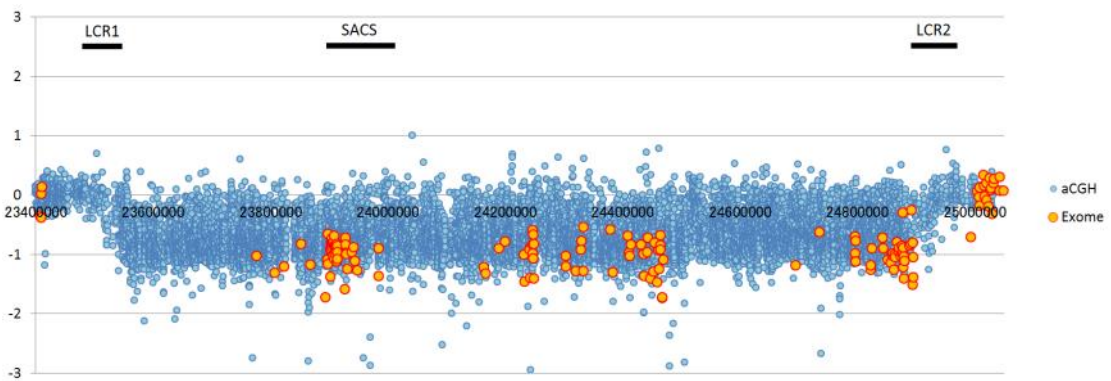


図 4-5. 続き. なお, #15 の矢印で示した *ZFYVE26* Exon22 の RPM 値に基づく log 比 (黒矢印) は実際には $-\infty$ であるが, 図内に記載するため-5 として表示した.

<#17 LITAF del>



<#18 SACS del>



<#19 SACS del>

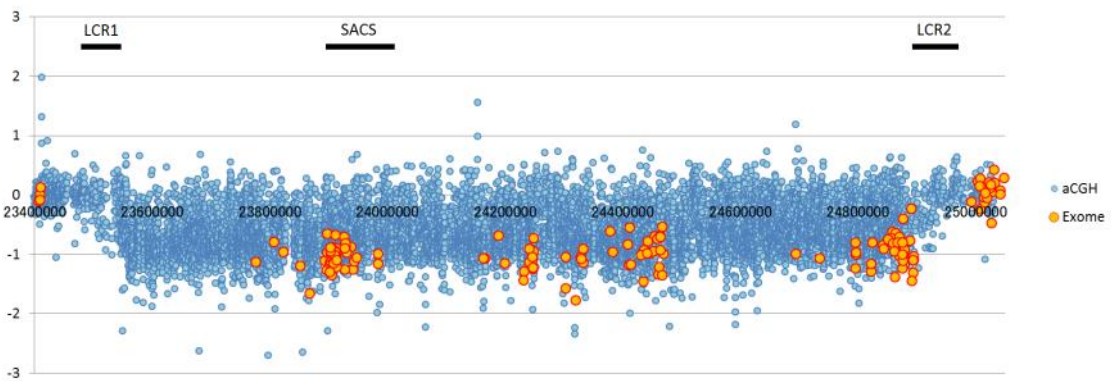
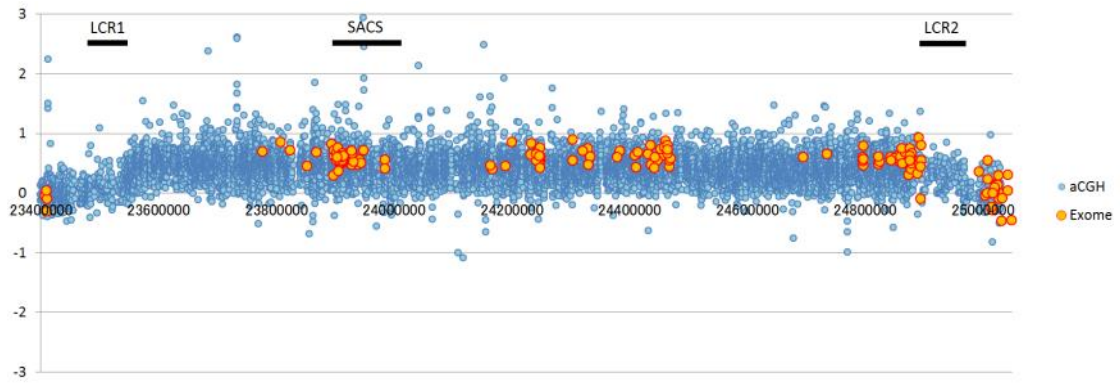
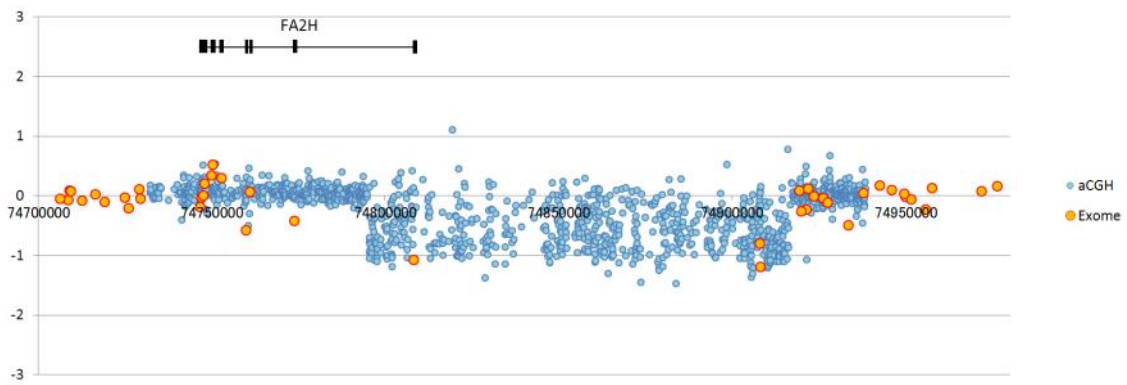


図 4-5. 続き.

<#20 SACS dup>



<#21 FA2H del>



<#22 CYP2U1 del>

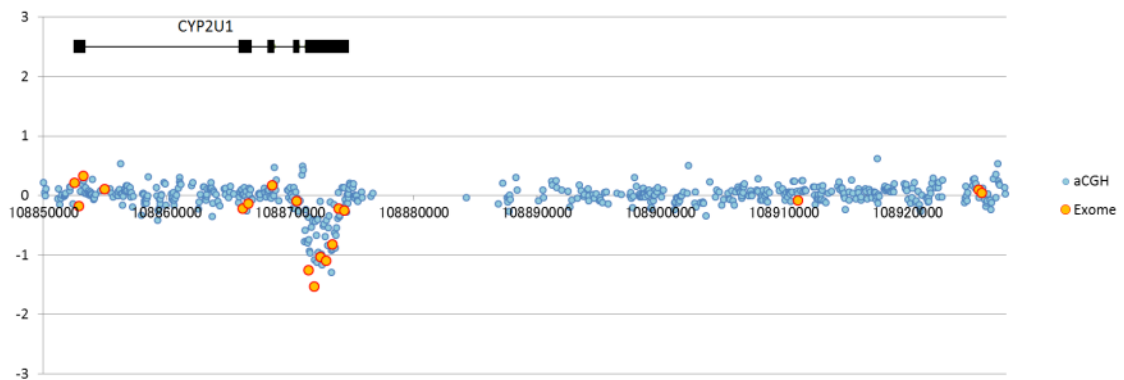


図 4-5. 続き.

<#23 *PARK2* dup>

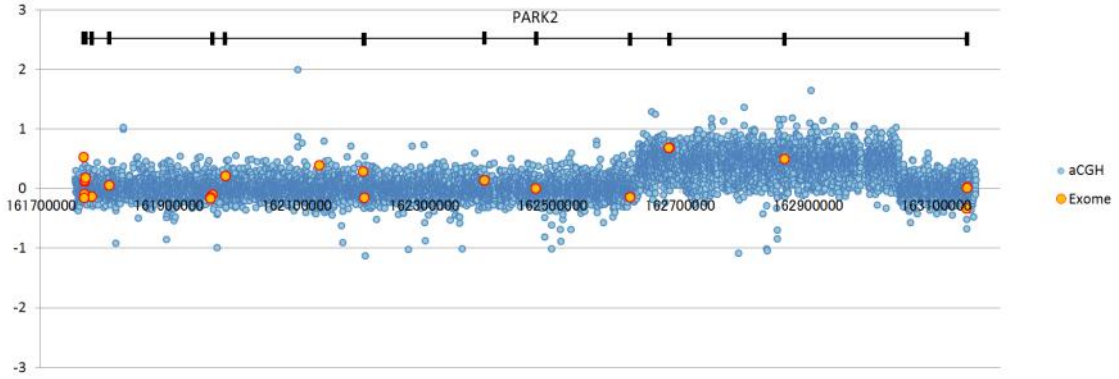


図 4-5. 続き

4-3 性染色体数異常について

性染色体の RPM 値に基づくアレル数判定により、4 例で性染色体数の異常を検出した。いずれも性染色体のトリソミーであり、XXX が 1 例、XXY が 2 例、XYY が 1 例であった（表 4-5）。

XXX, XXY, XYY の性染色体トリソミーの頻度については従来、1960 年代、1970 年代の文献に基づき、出生 1000 例あたり 1 例程度と考えられている。最近の報告によると XXX および XYY についてはおおむね同程度の頻度だが、XXY については 1000 例あたり 1.72 例とやや増加しているとの報告もある[53]。XXY の核型異常はクラインフェルター症候群として知られ、不妊症を契機に診断されることが多い。またいずれの性染色体トリソミーにおいても、精神発達過程への影響がみられることが報告されている[54, 55]。ただし、XXX 症候群、XYY 症候群についてはほとんど症状を呈しない例も多く、診断されていない症例が多く存在するとされている[55]。

今回 4 症例で性染色体トリソミーを検出したが、シャルコー・マリー・トゥース病、痙性対麻痺、周期性四肢麻痺の症例との関連を示す報告は過去になく、多数例の解析により偶発的に変異が発見されたものと考えられた。

Case	Diagnosis	Sex chromosomal genotype	Validation by array CGH
#25	HSP	XXX	yes
#26	Periodic paralysis	XXY	yes
#27	CMT	XXY	yes
#28	CMT	XYY	yes

表 4-5. 性染色体トリソミーを認めた 4 例の診断名および検出された性染色体数の異常を示す.

4-4-1 SMN1 遺伝子の構造変異解析

Exon 7 および Exon 8 に存在する *SMN1* と *SMN2* 遺伝子の塩基相違部位を用いて算出した、*SMN1* および *SMN2* 遺伝子の Exon 7, Exon 8 の各症例の RPM 値を図 4-6, 図 4-7 に示す。また、Exon 7 と Exon 8 の RPM 値の相関を図 4-8 に示す。

いずれにおいても、中央に存在する最も症例密度の高い領域が正常のコピー数

(2 アリル) の場合の RPM 値と考えられるが、データのバラつきが大きく、1 アリルと 2 アリル、2 アリルと 3 アリルの明確な区別は困難であった。ただし、0 アリルの場合には RPM 値がほぼ 0 と著しく低い値をとることから、0 アリルと 1 アリル以上は明確に区別することが可能であった。

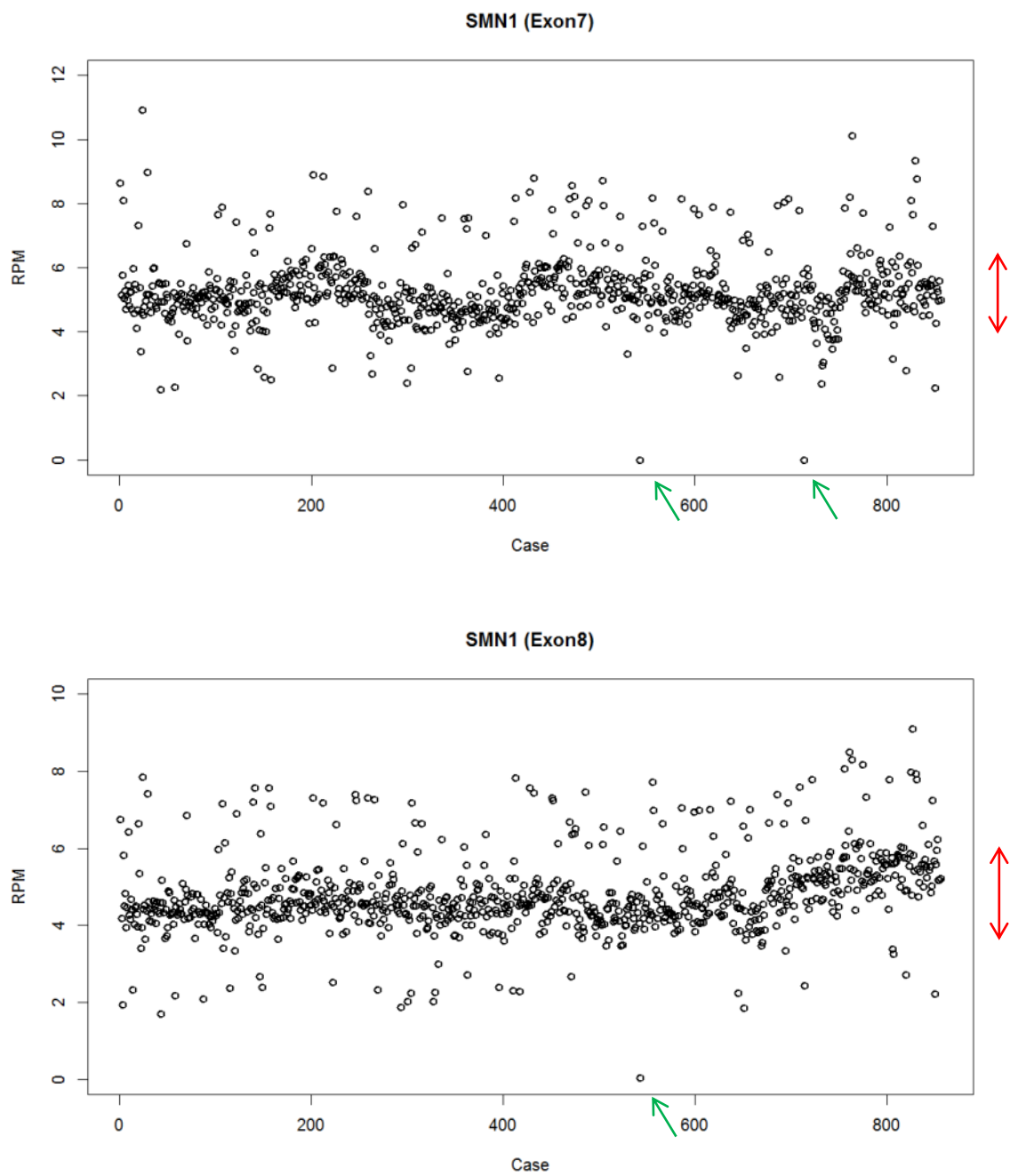


図 4-6. *SMN1* 遺伝子の Exon 7 (上図), Exon 8 (下図) の RPM 値を示したものの。個々の点が各症例の RPM 値を示す。赤矢印は正常のコピー数 (2 アリル) と考えられるゾーンを示す。緑矢印はアリル数が 0 と推定される症例を示す。

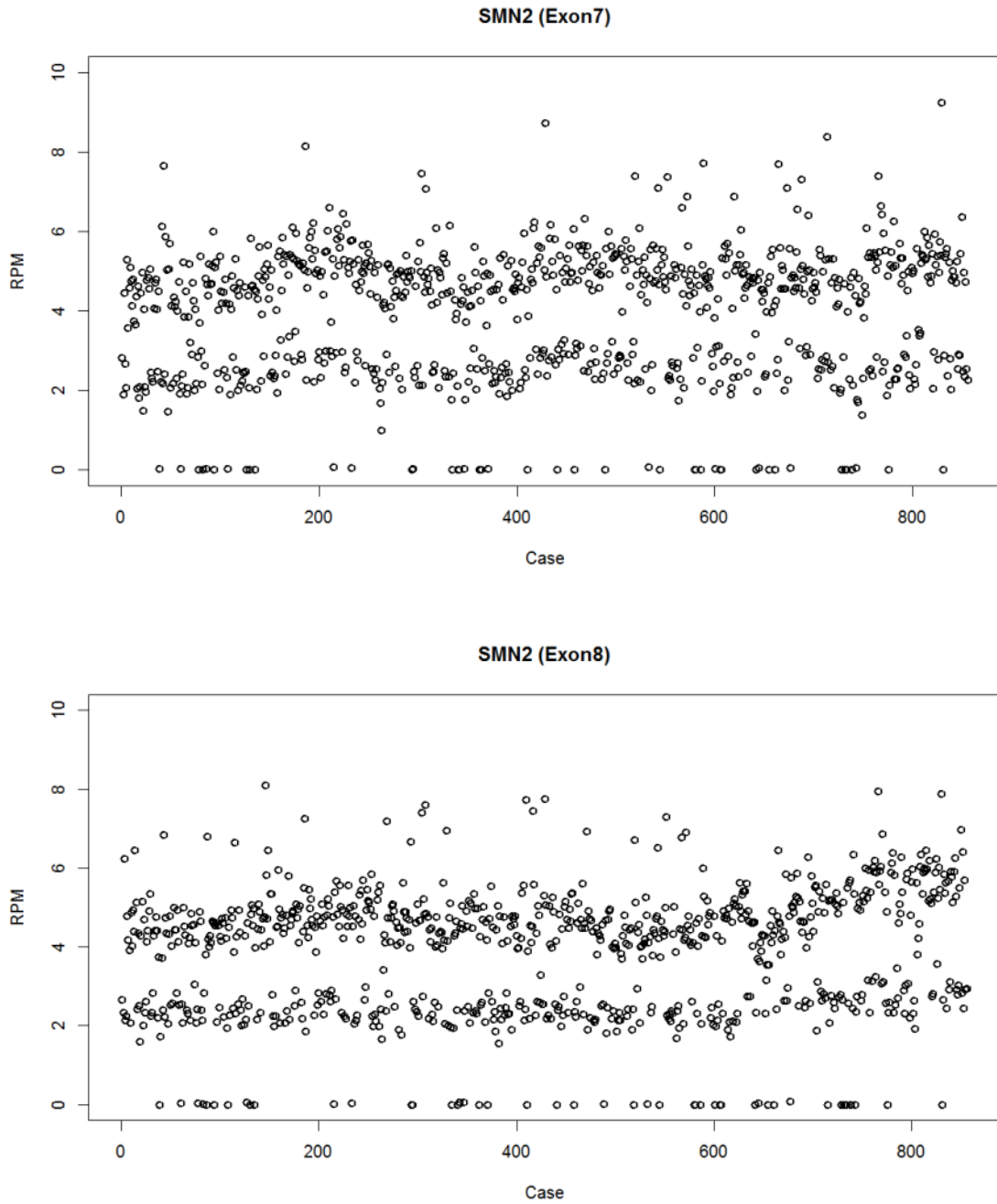


図 4-7. *SMN2* 遺伝子の Exon 7 (上図), Exon 8 (下図) の RPM 値を示したものの。個々の点が各症例の RPM 値を示す。

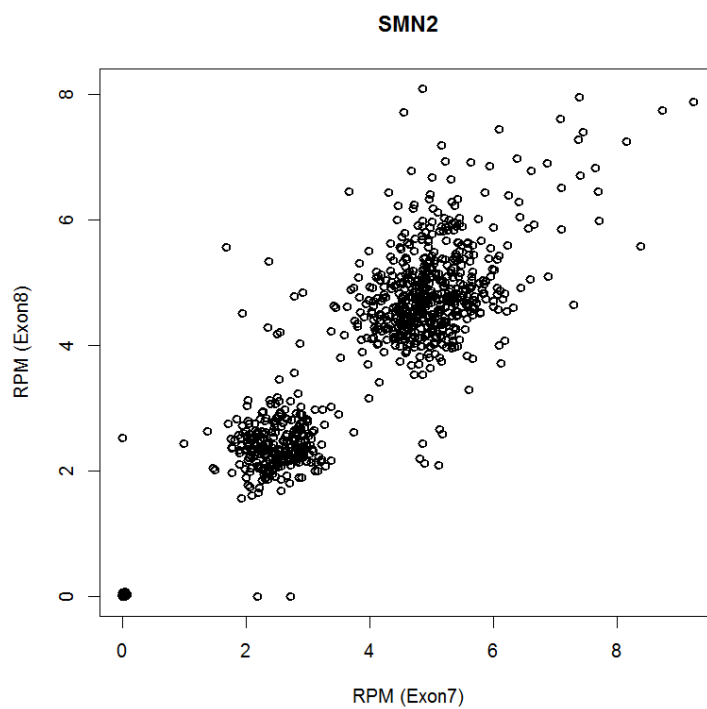
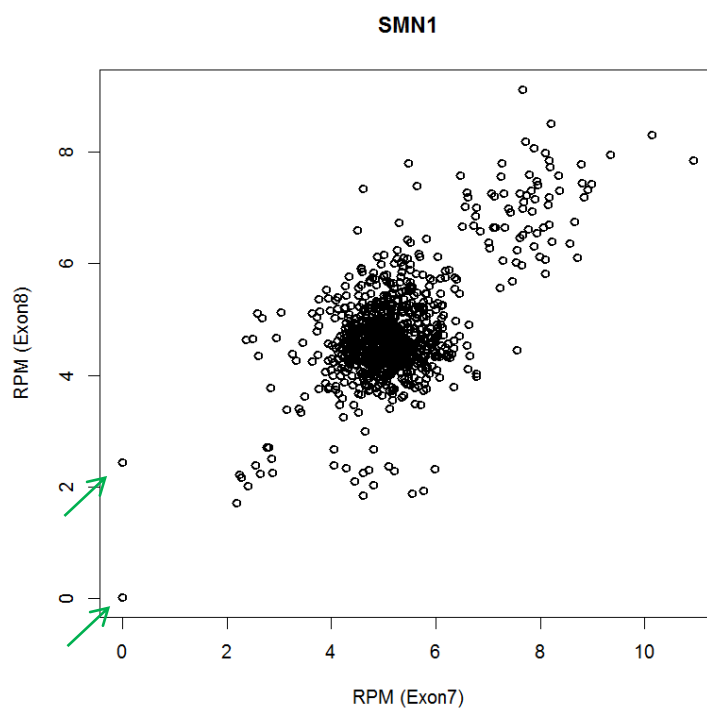


図 4-8. SMN1 (上図) および SMN2 遺伝子 (下図) の, Exon 7 と Exon 8 の RPM 値の比較. 緑矢印は SMN1 の Exon 7 から推定されるアレル数が 0 と考えられる症例を示す.

SMN1 遺伝子の変異によって引き起こされる脊髄性筋萎縮症では、95%以上の症例で *SMN1* 遺伝子の完全欠失 (0 アリル) が認められるとされ[56]、特に Exon 7 に存在する *SMN2* との塩基相違が、*SMN1* 蛋白質の発現において最も重要であるとされる[57, 58]。今回の解析で、Exon 7 のアリル数が 0 であった症例は 2 例のみであった。2 症例の RPM 値、臨床情報について表 4-6 に示す。また両症例について、Exon 7 において *SMN1* 遺伝子由来と考えられるアリルが存在しないことを、制限酵素断片長多型法[27]によって確認した結果を図 4-9 に示す。

1 例 (#30) においては Exon 7, Exon 8 とともに 0 アリルであり、もう 1 例 (#29) においては Exon 7 のアリル数が 0 であったが、Exon 8 のアリル数は 1 と推定された。後者に関しては、*SMN2* の RPM 値から、Exon 7 の塩基相違部位の塩基が *SMN2* に一致する T、Exon 8 の塩基相違部位の塩基が *SMN1* に一致する G である *SMN* 遺伝子が 1 アリル存在することが推定されるが、前述の通り Exon 7 の相違部位の塩基が C であることが *SMN1* 遺伝子の活性発現に最も重要とされており、#29 においても原因遺伝子変異である可能性が高いと考えられる。

Case	Diagnosis	<i>SMN1</i> Exon7	<i>SMN1</i> Exon8	<i>SMN2</i> Exon7	<i>SMN2</i> Exon8
#29	Motor neuron disease (Spinal muscular atrophy s/o)	0 (0)	2.44 (1)	8.38 (1以上)	5.59 (1以上)
#30	Congenital hereditary neuropathy	0 (0)	0.03 (0)	7.09 (1以上)	6.51 (1以上)

表 4-6. *SMN1* 遺伝子の完全欠失と考えられる 2 例の診断名, および *SMN* 遺伝子の各 exon の RPM 値を示す. 括弧内の数値は推定アレル数.

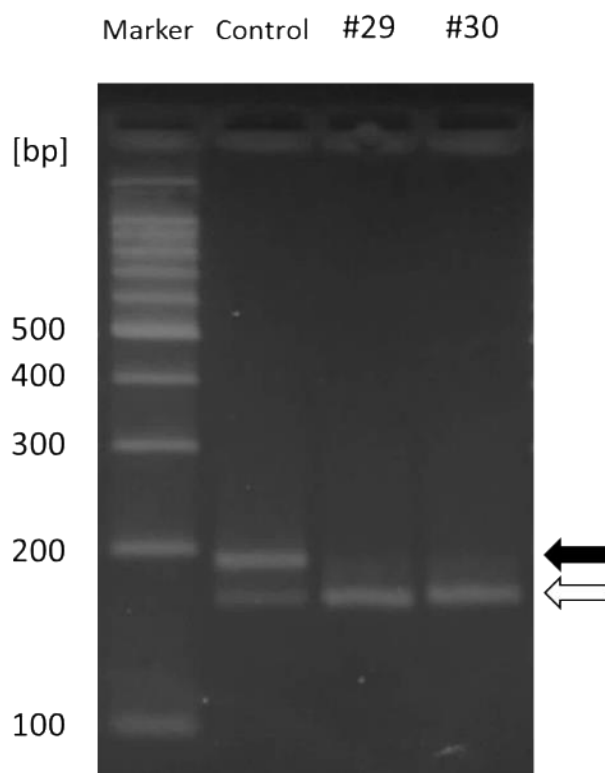


図 4-9. *SMN1* および *SMN2* 遺伝子の Exon7 を既報の primer [27]を用いた PCR 法により増幅し, 制限酵素(Dra I)により処理後, アガロースゲル電気泳動を行った. Control では Dra I により切断を受けた *SMN2* 由来のやや短いバンド(白矢印)と, *SMN1* 由来の切断を受けていないバンド (黒矢印) の 2 つのバンドが認められるが, #1, #2 では黒矢印のバンドが認められず, *SMN1* の完全欠失と判定される.

4-5 疾患群ごとの構造変異の検出数について

表 4-7 は，疾患の原因と考えられる構造変異を検出した 26 例（性染色体のア
リル数異常を認めた 4 例を除く）について，臨床診断名ごとにまとめて記載し
たものである．シャルコー・マリー・トゥース病，痙性対麻痺において多くの
構造変異を検出したが，シャルコー・マリー・トゥース病では *PMP22* 変異がそ
の中の大部分を占めていた．

臨床診断	解析症例	変異例	遺伝子名（括弧内は症例数）
CMT	289	15	<i>PMP22</i> (10), <i>SACS</i> (1), <i>SBF1</i> (1), <i>KIF5A</i> (1), <i>LITAF</i> (1), <i>SMN1</i> (1)
痙性対麻痺	172	7	<i>SACS</i> (2), <i>SPAST</i> (2), <i>FA2H</i> (1), <i>ZFYVE26</i> (1), <i>CYP2U1</i> (1)
筋萎縮性側索硬化症 (運動ニューロン疾患)	371	3	<i>OPTN</i> (2), <i>SMN1</i> (1)
多系統萎縮症	114	1	<i>PARK2</i> (1)

表 4-7. 疾患群ごとに病的構造変異を検出した症例数についてまとめたもの.

第6章 考察および今後の課題

今回、既存の構造変異検出用プログラムである Conifer を中心に据えた、エキソーム解析データから構造変異を検出するパイプラインを新たに構築し、結果として解析を行った全 1230 例中 26 例に、原因遺伝子変異の可能性のある構造変異を検出した。また、疾患との関連性は低いと考えられるが、4 例に性染色体数の異常を検出した。これらの変異はこれまでのエキソーム解析では見逃されていた変異であり、今回の解析で新たに診断が確定した症例が多くあったほか、過去にエキソーム解析でホモ接合性変異と診断されていたが、今回の解析により欠失変異との複合ヘテロ接合性変異であると診断が修正された症例もあり、本解析法がエキソーム解析による遺伝子診断の正確性を高める上で、有用であることを示すことができたと考える。また、神経内科領域の疾患においても構造変異はまれではなく、一塩基置換変異や短い挿入・欠失変異のみならず、構造変異の有無についても評価を行うことが必要であることが今回の研究により明らかとなった。

検出された変異に関して疾患毎に見ると、シャルコー・マリー・トゥース病と診断されていた症例では 6 遺伝子に構造変異が検出されたが、臨床的に最も頻度の高い原因遺伝子変異である *PMP22* 遺伝子の構造変異が大部分を占めてい

た。 *PMP22* 遺伝子の構造変異は脱髄型のシャルコー・マリー・トゥース病で最も頻度の高い変異であり、エキソーム解析を施行する前に FISH 法による解析で診断されることが普通であるが、 *PMP22* 遺伝子の構造変異はほとんどが 2 つの塩基相同性の高い領域に挟まれた複数の遺伝子を含む巨大なものであることから、本解析法でも十分に診断が可能であり、臨床的に有用と考えられる。

痙性対麻痺症例では、 *SACS*, *SPAST*, *FA2H*, *ZFYVE26*, *CYP2U1* の 5 遺伝子について構造変異を検出し、さらにシャルコー・マリー・トゥース病と診断されていた 1 例において、家族性痙性対麻痺 (SPG10) の原因遺伝子である *KIF5A* の構造変異を認めた。うち、 *ZFYVE26*, *KIF5A*, *CYP2U1* についてはこれまで構造変異の報告はない。痙性対麻痺の診断においては、 *SACS* 遺伝子や *SPAST* 遺伝子のように構造変異の報告が多い遺伝子があり、塩基置換変異や短い挿入変異、欠失変異のみならず、構造変異の評価も併せて行うことが重要と考えられる。

多系統萎縮症の 1 例では、常染色体劣性遺伝形式の家族性パーキンソン病の原因遺伝子である *PARK2* 遺伝子に、1 アリルのみの欠失変異を認めた。多系統萎縮症とパーキンソン病では臨床像が異なり、同変異については偶発的に検出された変異の可能性が高いと考えられるが、このような変異が疾患発症のリス

クとして寄与している可能性は完全には否定できない。同遺伝子について、変異を 1 アリルのみ有する場合の影響についてさらに検討するには、多数例の健常者での構造変異の検出頻度を調べ、比較する必要があると考えられる。

筋萎縮性側索硬化症と診断された症例では、2 例で *OPTN* 遺伝子の構造変異を検出した。同遺伝子については過去にホモ接合性変異を有する症例（常染色体劣性）、ヘテロ接合性変異を有する症例（常染色体優性）がいずれも報告されているが、今回変異を検出した 2 例（ホモ接合性 1 例、ヘテロ接合性 1 例）の間で、発症年齢など臨床像に相違が認められた。特にヘテロ接合性の症例に関しては、健常者あるいは筋萎縮性側索硬化症以外の患者で、同遺伝子の構造変異が存在するかどうかなど、追加の検討が必要と考えられる。

なお、今回の解析では、過去当科にてエキソーム解析を行い、今回作成したパイプラインで解析可能であった全ての症例を対象としているが、エキソーム解析施行に至った背景、臨床情報、家族内発症者の有無などに基づいた症例の分類・選択は行えていない。従って、疾患群毎の構造変異の検出頻度に関して考察することは難しいと考えられる。また、検出された変異のうち、過去に報告の無いものに関しては、実際に疾患発症に関わっているかどうかについて、

症例毎に臨床症状，家族歴の確認や，追加の遺伝子解析が必要と考えられる。

SMN1 遺伝子については自作の解析手法を用い，2例において *SMN1* 遺伝子の欠失を検出したが，次世代シーケンサーデータを利用して変異を検出した報告はこれまでにない。*SMN1* 遺伝子の変異検出には制限酵素断片長多型解析法（RFLP法）を始めとして，多くの手法が考案されているが，いずれも *SMN1* 遺伝子のみ限定した解析手法であり，本解析法は非典型例の診断において有用と考えられる。なお本解析法は，Exon7，Exon8の *SMN1* 固有の2塩基のみに着目しており，同部位のホモ接合性の欠失のみ検出可能である。脊髄性筋萎縮症では大多数（～95%）が *SMN1* 遺伝子のホモ接合性の欠失によるとされ，本法により多くの症例で診断が可能と考えられるが，それ以外の変異（複雑な構造変異や，一塩基置換と欠失の複合ヘテロ接合性変異など）は現在の解析法では検出できないため，今後の課題である。

続いて，本解析法と他の解析手法との比較についてであるが，既存の構造変異の検出方法には，一部の標的領域のみを対象とする手法（定量PCR法，MLPA法（Multiplex ligation-dependent probe amplification），FISH法など）と，ゲノム全域に渡って解析可能な手法（アレイCGH法，SNPアレイ法）がある。既存の手

法にもそれぞれ長所，短所があるが，今回のエキソーム解析データを利用した解析には以下のような短所が挙げられる．(1) 含まれるエキソン領域が少ない(1～2個) 構造変異は検出困難である．(2) リード数に影響を与える要因が構造変異以外に多数あり，データのバラつきが大きく，精度面での問題が存在する．(3) 多数例の平均値からの解離を利用して判定する方法のため，解析症例中に高頻度で存在する構造変異は検出できない．

一部の標的領域のみを対象とする手法との比較では，本解析法では標的領域以外についてもデータが得られる点で優れるが，広範囲を対象とする手法のうち，最適化された全エキソンを対象とするアレイ CGH 法は，解析対象範囲が同等であり，かつ検出精度の点で劣ると考えられる．ただしアレイ CGH 法と比較して以下の利点があると考えられる．(1) アレイ CGH 法では得られない，塩基配列情報（一塩基置換変異，短い挿入・欠失変異の有無）が同時に得られる．(2) アレイ CGH 法で解析困難な偽遺伝子の存在する遺伝子についても，条件次第で解析が可能である（*SMN1* 遺伝子など）．

本解析法には，短い構造変異や高頻度に存在する構造変異など原理的に検出できない変異があり，また検出された変異の中には偽陽性も含まれるため[59,60]，

本解析法のみで全エクソンに関して構造変異の有無について確定診断することは難しいと考えられる。従って、コストの問題を考えなければ、エクソーム解析を施行した症例全例について、アレイ CGH 法による解析を併用する方法が最も良いと思われる。ただし、本解析法でも一部の標的遺伝子に対象を限定すれば確定診断的に用いることができる可能性があり、例えば *PMP22* 遺伝子や *SACS* 遺伝子など、構造変異の範囲が一定していて多くのエクソン領域を含むものについては、データ (RPM 値) のバラつきが非常に少なく、本法のみでも診断可能と思われる。

なお、エクソーム解析データに基づく構造変異検出プログラムの中でも、**Conifer** は偽陽性の少ないプログラムと評価されている[59,60]。しかし、実際に今回の解析において偽陽性がどの程度存在するかに関しては、使用した次世代シーケンサーおよびライブラリ作成用試薬の種類および条件、RPM 値算出時の領域の設定、同時に解析した症例数、解析パラメータ値など、様々な要因が影響を与えると考えられるため、過去の文献の偽陽性率を本解析に当てはめることはできないと考えられ、検出された構造変異全てについてアレイ CGH 法による確認を行って算出する必要があると考えられる。

今回の解析における一つの問題点は、Conifer による解析で著しく多い構造変異が検出される症例が少ないながらも含まれていたことである。これらの症例に共通する因子は現時点で明らかにできていないが、サンプルの調整方法からデータ解析までの間に何らかの問題点・相違点が無かったかどうか、検討する必要があると考えられる。また、多数例のエキソーム解析データに基づいて構造変異を検出するプログラムは Conifer 以外にも、XHMM[61]、ExoCNVTest[62] など複数報告されており、他のプログラムを使用して得られた解析データとの比較を行うことも検討される。

さらに、今回は既報告のある変異についての検討のみしか行えなかったが、既報告の無い遺伝子に特定の疾患において共通する構造変異がないかなど、全遺伝子を対象とした解析方法についても、今後検討していきたいと考える。

最後に、現在はコスト面から全ゲノム解析を多くの症例に利用することは難しいが、全ゲノム解析に基づく構造変異検出はエキソーム解析と比較して、(1) エキソン領域のみを抽出する作業がないためバイアスが少ない、(2) エキソン領域以外も含むすべての領域のデータが概ね均一に得られる、という非常に大きな利点がある。今後、コストの削減により全ゲノム解析が日常的に行えるよう

になれば，次世代シーケンサーデータに基づいた構造変異の検出を，より正確に行うことが可能になると考えられる．

謝辞

本稿を終えるに当たり，本研究にご指導賜りました東京大学大学院医学系研究科脳神経医学専攻神経内科学教室 辻 省次 教授，後藤 純 准教授，三井 純 特任助教，石浦 浩之 助教，伊達 英俊 特任助教，Budrul Ahsan 特任研究員，松川 敬志 先生，東京大学大学院新領域創成科学研究科情報生命科学専攻バイオ情報科学大講座バイオデータベース分野森下研究室 森下 真一 教授，土井 晃一郎 特任講師，吉村 淳 特任助教をはじめとする教室員各員に深謝申し上げます。

また，貴重な検体をご提供下さりました，鹿児島大学大学院医歯学総合研究科神経内科・老年病学講座 高嶋 博 教授，山梨大学医学部神経内科学講座 瀧山 嘉久 教授に深謝申し上げます。

引用文献

1. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461:272-6
2. Mamanova L, Coffey AJ, Scott CE, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods*. 2010;7:111-8
3. Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet*. 2010;19:R145-51
4. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011;12:745-55.
5. Foo JN, Liu JJ, Tan EK. Whole-genome and whole-exome sequencing in neurological diseases. *Nat Rev Neurol*. 2012;8:508-17
6. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7:85-97
7. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444-54
8. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007;316:445-9
9. Kumar RA, KaraMohamed S, Sudi J, et al. Recurrent 16p11.2 microdeletions in

- autism. *Hum Mol Genet.* 2008;17:628-38
10. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature.* 2008;455:237-41
 11. Kirov G, Grozeva D, Norton N, et al. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet.* 2009;18:1497-503
 12. Olson H, Shen Y, Avallone J, et al. Copy number variation plays an important role in clinical epilepsy. *Ann Neurol.* 2014;75:943-58
 13. de Kovel CG, Trucks H, Helbig I, et al. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain.* 2010;133:23-32
 14. Lupski JR, de Oca-Luna RM, Slaugenhaupt S, et al. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell.* 1991;66:219-32
 15. Chance PF, Alderson MK, Leppig KA, et al. DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell.* 1993;72:143-51
 16. Singleton AB, Farrer M, Johnson J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science.* 2003;302:841
 17. Ibáñez P, Bonnet AM, Débarges B, et al. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet.* 2004;364:1169-71

18. Rovelet-Lecrux A, Hannequin D, Raux G, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet.* 2006;38:24-6
19. Lefebvre S, Bürglen L, Reboullet S, et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell.* 1995;80:155-65
20. Zhao M, Wang Q, Wang Q, et al. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics.* 2013;14 Suppl 11:S1
21. Korbel JO, Urban AE, Affourtit JP, et al. Paired-end mapping reveals extensive structural variation in the human genome. *Science.* 2007;318:420-6
22. Ye K, Schulz MH, Long Q, Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics.* 2009;25:2865-71
23. Yoon S, Xuan Z, Makarov V, et al. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res.* 2009;19:1586-92
24. Krumm N, Sudmant PH, Ko A, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Res.* 2012;22:1525-32
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler

- transform. *Bioinformatics*. 2009;25:1754-60
26. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078-9
27. Van der Steege G, Grootsholten PM, et al. PCR-based DNA test to confirm clinical diagnosis of autosomal recessive spinal muscular atrophy. *Lancet*. 1995;345:985-6.
28. Ross MT, Grafham DV, Coffey AJ, et al. The DNA sequence of the human X chromosome. *Nature*. 2005;434:325-37
29. Johnson NC. XG: the forgotten blood group system. *Immunohematology*. 2011;27:68-71
30. Forsberg LA, Rasi C, Malmqvist N, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet*. 2014;46:624-8
31. Szigeti K, Garcia CA, Lupski JR. Charcot-Marie-Tooth disease and related hereditary polyneuropathies: molecular diagnostics determine aspects of medical management. *Genet Med*. 2006;8:86-92
32. Lopes J, LeGuern E, Gouider R, et al. Recombination hot spot in a 3.2-kb region of the Charcot-Marie-Tooth type 1A repeat sequences: new tools for molecular diagnosis of hereditary neuropathy with liability to pressure palsies and of

- Charcot-Marie-Tooth type 1A. French CMT Collaborative Research Group. *Am J Hum Genet.* 1996;58:1223-30
33. Inoue K, Dewar K, Katsanis N, et al. The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res.* 2001;11:1018-33
34. Zhang F, Khajavi M, Connolly AM, et al. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet.* 2009;41:849-53
35. Choi BO, Kim NK, Park SW, et al. Inheritance of Charcot-Marie-Tooth disease 1A with rare nonrecurrent genomic rearrangement. *Neurogenetics.* 2011 Feb;12(1):51-8
36. Hazan J, Fonknechten N, Mavel D, et al. Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia. *Nat Genet.* 1999;23:296-303
37. Miura S, Shibata H, Kida H, et al. Partial SPAST and DPY30 deletions in a Japanese spastic paraplegia type 4 family. *Neurogenetics.* 2011;12:25-31
38. Maruyama H, Morino H, Ito H, et al. Mutations of optineurin in amyotrophic lateral sclerosis. *Nature.* 2010;465:223-6
39. Iida A, Hosono N, Sano M, et al. Novel deletion mutations of OPTN in amyotrophic

- lateral sclerosis in Japanese. *Neurobiol Aging*. 2012;33:1843.e19-24
40. Hanein S, Martin E, Boukhris A, et al. Identification of the SPG15 gene, encoding spastizin, as a frequent cause of complicated autosomal-recessive spastic paraplegia, including Kjellin syndrome. *Am J Hum Genet*. 2008;82:992-1002
41. Reid E, Kloos M, Ashley-Koch A, et al. A kinesin heavy chain (KIF5A) mutation in hereditary spastic paraplegia (SPG10). *Am J Hum Genet*. 2002;71:1189-94
42. Crimella C, Baschiroto C, Arnoldi A, et al. Mutations in the motor and stalk domains of KIF5A in spastic paraplegia type 10 and in axonal Charcot-Marie-Tooth type 2. *Clin Genet*. 2012;82:157-64
43. Street VA, Bennett CL, Goldy JD, et al. Mutation of a putative protein degradation gene LITAF/SIMPLE in Charcot-Marie-Tooth disease 1C. *Neurology*. 2003;60:22-6
44. Engert JC, Bérubé P, Mercier J, et al. ARSACS, a spastic ataxia common in northeastern Québec, is caused by mutations in a new gene encoding an 11.5-kb ORF. *Nat Genet*. 2000;24:120-5
45. Breckpot J, Takiyama Y, Thienpont B, et al. A novel genomic disorder: a deletion of the SACS gene leading to spastic ataxia of Charlevoix-Saguenay. *Eur J Hum Genet*. 2008;16:1050-4
46. Terracciano A, Casali C, Grieco GS, et al. An inherited large-scale rearrangement in

- SACS associated with spastic ataxia and hearing loss. *Neurogenetics*. 2009;10:151-5
47. Stevens JC, Murphy SM, Davagnanam I, et al. The ARSACS phenotype can include supranuclear gaze palsy and skin lipofuscin deposits. *J Neurol Neurosurg Psychiatry*. 2013;84:114-6.
48. Edvardson S, Hama H, Shaag A, et al. Mutations in the fatty acid 2-hydroxylase gene are associated with leukodystrophy with spastic paraparesis and dystonia. *Am J Hum Genet*. 2008;83:643-8
49. Pierson TM, Simeonov DR, Sincan M, et al. Exome sequencing and SNP analysis detect novel compound heterozygosity in fatty acid hydroxylase-associated neurodegeneration. *Eur J Hum Genet*. 2012;20:476-9
50. Tesson C, Nawara M, Salih MA, et al. Alteration of fatty-acid-metabolizing enzymes affects mitochondrial form and function in hereditary spastic paraplegia. *Am J Hum Genet*. 2012;91:1051-64
51. Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*. 1998;392:605-8.
52. Nakhro K, Park JM, et al. SET binding factor 1 (SBF1) mutation causes Charcot-Marie-Tooth disease type 4B3. *Neurology*. 2013;81:165-73.
53. Morris JK, Alberman E, Scott C, et al. Is the prevalence of Klinefelter syndrome

- increasing? *Eur J Hum Genet.* 2008;16:163-70
54. Tartaglia NR, Howell S, Sutherland A, et al. A review of trisomy X (47,XXX).
Orphanet J Rare Dis. 2010;5:8
55. Leggett V, Jacobs P, Nation K, et al. Neurocognitive outcomes of individuals with a sex chromosome trisomy: XXX, XYY, or XXY: a systematic review. *Dev Med Child Neurol.* 2010;52:119-29
56. Hahnen E, Forkert R, Marke C, et al. Molecular analysis of candidate genes on chromosome 5q13 in autosomal recessive spinal muscular atrophy: evidence of homozygous deletions of the SMN gene in unaffected individuals. *Hum Mol Genet.* 1995;4:1927-33
57. Lorson CL, Hahnen E, Androphy EJ, et al. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc Natl Acad Sci U S A.* 1999;96:6307-11.
58. Jodelka FM, Ebert AD, Duelli DM, et al. A feedback loop regulates splicing of the spinal muscular atrophy-modifying gene, SMN2. *Hum Mol Genet.* 2010;19:4906-17
59. de Ligt J, Boone PM, Pfundt R, et al. Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat.* 2013;34:1439-48
60. Samarakoon PS, Sorte HS, Kristiansen BE, Identification of copy number variants

from exome sequence data. *BMC Genomics*. 2014;15:661

61. Fromer M, Moran JL, Chambert K, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet*. 2012;91:597-607

62. Coin LJ, Cao D, Ren J, et al. An exome sequencing pipeline for identifying and genotyping common CNVs associated with disease with application to psoriasis. *Bioinformatics*. 2012;28:i370-i374