

総説 (2018年度横浜市立大学医学会賞受賞論文)

遺伝性疾患の原因遺伝子単離と ゲノム異常発生分子機構に関する研究

水 口 剛

横浜市立大学医学部 遺伝学

要 旨: 希少遺伝性疾患の原因は、現在でも全体の40-60%程度の疾患について未解明のままである。疾患を引き起こす遺伝子変異の種類やサイズは多様で全てを網羅的にカバーするゲノム解析技術は存在しない。従って未解明の疾患については既存の解析技術の穴をうめるような新規解析技術の適用が有用である。実際、染色体核型分析、FISH法、キャピラリーシーケンサー、マイクロアレイ、次世代シークエンサーに代表される染色体・ゲノム解析法はそれぞれ異なる解像度を有し、新規解析技術の登場が新たな種類・サイズの病的変化を明らかにしてきた。本稿ではこれらの解析技術を駆使して筆者が行ってきた遺伝性疾患の原因・病態解明を目的とした多角的取り組みについて紹介する。

Key words: ポジショナルクローニング (positional cloning),
次世代シーケンサー (next generation sequencer),
全エクソーム解析 (whole exome sequencing), Hi-C (Hi-C),
ロングリードシーケンサー (long-read sequencer)

はじめに

希少遺伝疾患は発症頻度こそ少ないものの、7,000疾患以上存在すると推定されている。ゲノム解析技術の進歩に伴いこのうち約60%程度(ヒトの遺伝性疾患データベースであるOMIMには4,099個の表現型を規定する遺伝子が登録されている)において原因遺伝子が解明されてきた^{1, 2)}。特に次世代シーケンサーの出現は希少遺伝疾患研究を一変させ、これまでの遺伝学的解析法ではアプローチが困難であった小家系や孤発例においても疾患原因遺伝子が同定可能な時代となった。一方で網羅的遺伝子解析によって検出される膨大な数の遺伝的変化について病原性の評価・解釈が課題となっている。また遺伝的変化のうち疾患の原因となるゲノム構造異常発生のメカニズムについて、ヒトではDNA塩基配列決定以上の解析が難しく未解明な点が多い。筆者はこれまでにゲノム解析技術の進歩の恩恵を受けながらこれらの新規解析技術を積

極的に取り入れることで遺伝性疾患の原因解明研究を行ってきた。筆者がこれまで携わった(1)染色体構造異常症例を端緒とする疾患遺伝子単離(ポジショナルクローニング)、(2)全エクソーム解析と酵母機能解析による遺伝子型-表現型関連の研究、(3)ゲノム異常発生分子機構の研究(4)ロングリードシーケンス解析法の確立にむけた取り組みについて紹介する。

I. 染色体構造異常症例を端緒とする疾患遺伝子単離(ポジショナルクローニング)

疾患座の位置情報に基づく原因遺伝子単離(ポジショナルクローニング)において染色体転座や微細欠失を合併する患者は絶好の解析対象であり、患者が唯一でも原因遺伝子の単離が可能ながある。2004年、マルファン症候群様症状を呈する患者の染色体転座切断点解析を端緒として新規責任遺伝子 *TGFBR2* を単離した³⁾。マルファン症候群(MFS)は5,000-10,000出生に1例の罹患

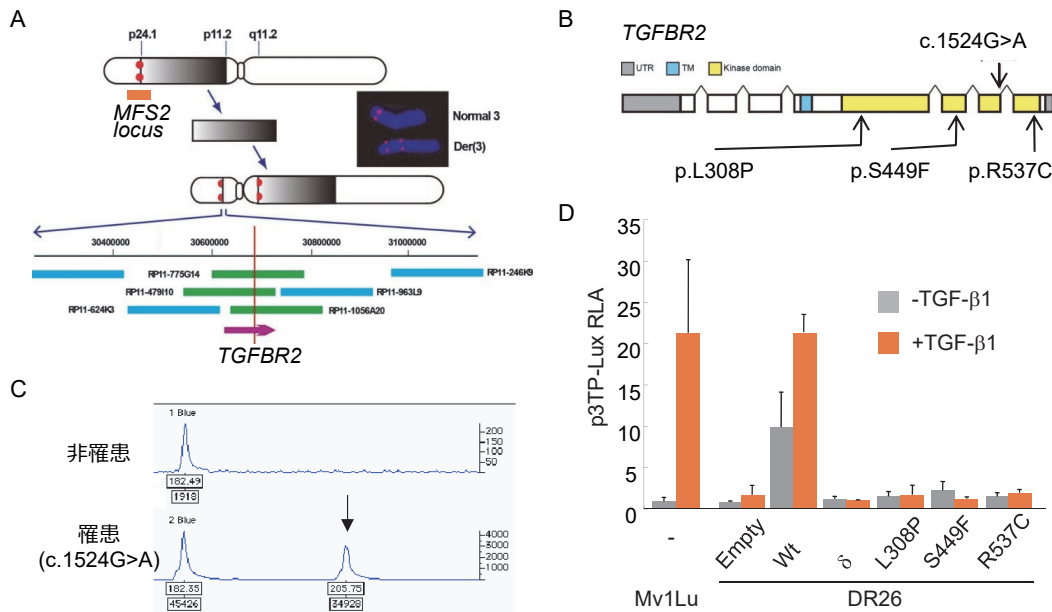


図1 マルファン症候群2型（現在ロイス・デーツ症候群と呼称）の責任遺伝子 *TGFBR2* の単離

A: マルファン症候群様の臨床症状を呈し染色体構造異常を合併した患者の染色体 FISH 解析. 3 番染色体 3 p24.1 転座切断点が *MFS2* 疾患遺伝子座 (orange rectangle) と合致した. 転座切断点をまたぐ BAC クローン (green rectangle) を用いた FISH 解析において *TGFBR2* 遺伝子の断裂を示唆するスプリットシグナル所見が派生 3 番染色体で観察された. B: *TGFBR2* の変異解析により同定された 4 個の病的バリエント. スプライシング変異 1 例 (c.1524G>A), ミスセンス変異 3 例 (p.L308P, p.S449F, p.R537C). C: c.1524G>A によるスプライシング異常の解析. RT-PCR により罹患者にのみ異常転写産物が検出された (矢印). D: ルシフェラーゼレポーターアッセイによる TGF- β シグナル伝達活性の検討. 内在性の *TGFBR2* 発現が欠損した DR26 細胞に野生型および変異型 *TGFBR2* を発現させ TGF- β 刺激後のシグナル活性を定量した. 野生型がシグナル活性を回復 (rescue) させたのに対し 3 個のミスセンス変異 (p.L308P, p.S449F and p.R537C) の強制発現でシグナル活性は回復しなかった. δ は細胞内キナーゼドメインを欠損させた短縮型 *TGFBR2*. (Mizuguchi T, Colod-Beroud G, et al.³⁾ より改変).

率で発症する常染色体優性遺伝性の結合織疾患である. 心血管系, 骨格系, 眼の異常を主徴とし, 細胞外マトリックスのマイクロフィブリルを構成するフィブリリン 1 をコードする *FBN1* 遺伝子変異によって引き起こされる⁴⁾. しかし *FBN1* 変異を認めない症例が多数存在すること, また連鎖解析により第 2 の疾患遺伝子座 (*MFS2*: 染色体 3 p25-p24.2 領域) が報告されたことから遺伝的異質性 (異なる原因遺伝子異常により同一の臨床表現型を呈すること) が指摘されていた⁵⁾. 筆者らは染色体 3 p24.1 に染色体切断点を持つ染色体異常とマルファン症候群を合併する症例を経験した. 転座切断点が *MFS2* 座と合致する事から 3 p24.1 染色体切断点においてマルファン症候群 2 型 (現在ロイス・デーツ症候群と呼称される) の原因遺伝子が断裂し, それが発症の原因となっていると想定し染色体切断点付近の染色体・ゲノム解析を行った. 染色体蛍光 in situ ハイブリダイゼーション (Fluorescence in situ hybridization, FISH) 法により転座切断点で TGF- β II 型受容体をコードする *TGFBR2* 遺伝子が断裂している事を突き止めた (図 1 A). さらに *FBN1* 遺伝子に異常を認めないマルファン症候群 20 家系について *TGFBR2* のシーケンス変異解析を行い 4 種類の点変異を 5 症例で同定した (スプライシング変異 1 個 (c.1524G>A), ミスセンス変異 3 個 (p.L308P, p.S449F, p.R537C)). (図 1 B). c.

1524G>A は同義的置換であったがエクソン 6 の最終塩基でありスプライシング異常を引き起こすこと, フレームシフト (コドンの読み枠が変わる変異) により未成熟終始コドンが出現することを cDNA 解析により確認した (図 1 C). この結果から未成熟終始コドンによりタンパク質の早期翻訳終結または異常な RNA 分解などを引き起こす病的変異であると結論した. ミスセンス変異についてはルシフェラーゼレポーターアッセイを用いて TGF- β シグナル活性への影響を検討した. 内在性 TGF- β II 型受容体を欠損した DR26 細胞に野生型および変異型 *TGFBR2* を発現させ TGF- β 1 刺激後のシグナル活性を定量した. 野生型 *TGFBR2* の強制発現が TGF- β シグナルを回復 (rescue) させたのに対し 3 個のミスセンス変異型 *TGFBR2* (p.L308P, p.S449F, p.R537C) ではシグナル活性が回復しなかった事から機能喪失性変異であることが示唆された (図 1 D). 以上の結果より *TGFBR2* がマルファン症候群 2 型の疾患原因遺伝子であるとの結論に至った. マルファン症候群においては *FBN1* 遺伝子異常が結合織の物理的脆弱性を引き起こすと考えられている. この構造的破綻に加えて, マルファン症候群モデルマウスの研究からフィブリリン 1 が TGF- β の活性調節を担い, TGF- β シグナル伝達異常が病態に関与することが示唆されていた⁶⁾. この仮説を支持し, ヒトにおいて TGF- β シグナル伝達異常

が結合織疾患を惹起することを明らかにした。その後、マルファン症候群類縁疾患においてTGF- β シグナルパスウェイに属する原因遺伝子が相次いで報告されロイス・ディーツ症候群を始めとするTGF- β シグナル伝達異常による先天性大動脈瘤症候群の疾患概念が確立される契機となった⁷⁾。更にマルファン症候群およびマルファン症候群が疑われる49名の患者集団を用いた遺伝子変異解析を行い、55% (27/49症例) に*FBNI* 遺伝子変異、6% (3/49症例) にTGF- β 受容体遺伝子変異 (*TGFBR1* または *TGFBR2*) を同定した。臨床症状 (表現型) と遺伝子・遺伝子変異の種類 (遺伝子型) の検討 (遺伝子型-表現型関連) からマルファン症候群およびロイス・ディーツ症候群の主要な原因遺伝子がそれぞれ*FBNI*、TGF- β 受容体遺伝子 (*TGFBR1* または *TGFBR2*) である事を報告した⁸⁾。

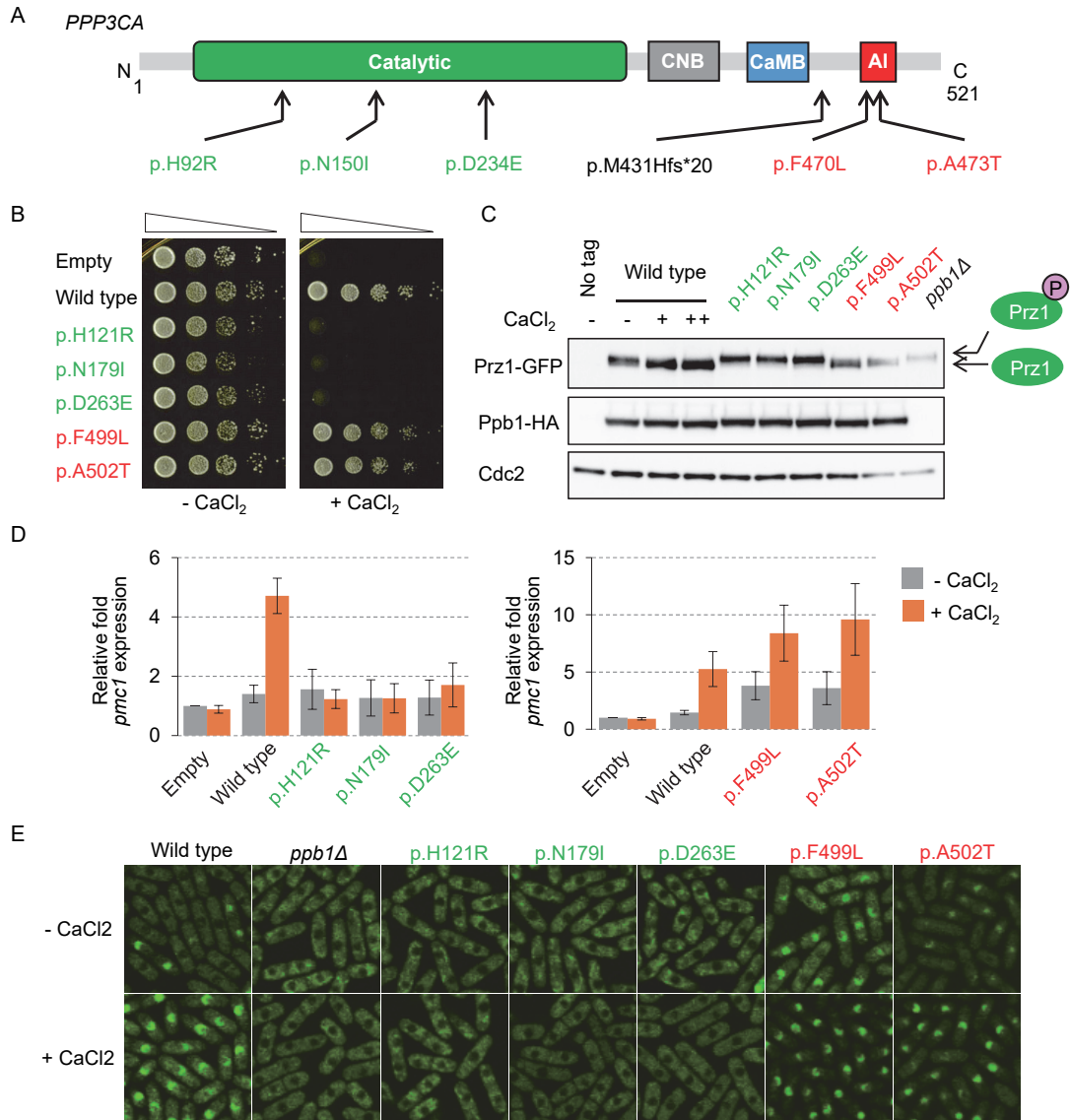
上述のFISH解析はゲノムの特定領域を高解像度 (100kb程度) で解析することはできるがゲノム全体を俯瞰的に解析することはできない。そこでポジショナルクローニングによる疾患遺伝子同定の端緒となる染色体構造異常をゲノムワイドに探索するための自主開発BAC (Bacterial Artificial Chromosome) マイクロアレイの作製に関わり様々な疾患のアレイ比較ゲノムハイブリダイゼーション解析 (array-based comparative genomic hybridization, array CGH) を実施した⁹⁾。4,234個のBACクローンを全ゲノムに均等に配置したマイクロアレイを作製することで、平均0.75Mbの解像度でゲノムコピー数を解析することが可能となった (染色体分析の検出限界が5-10Mb程度であることから、従来の1/10以下のコピー数変化をゲノムワイドに検出できる)。この自主開発BACマイクロアレイを用いたarray CGH解析により、難治性てんかん患者 (生後2か月から大田原症候群を発症しその後West症候群に移行した1例の女兒) に2.0Mbの染色体9q33.3-q34.11微細欠失を検出した。この微細欠失領域に存在する候補遺伝子の変異解析 (13例の大田原症候群患者で*STXBPI*変異解析を行い4例で変異が同定された) により大田原症候群の新規責任遺伝子*STXBPI*が単離された¹⁰⁾。

II. 全エクソーム解析と酵母機能解析による遺伝子型-表現型関連の研究

現在、ショートリード次世代シーケンサーを用いてゲノムの約2%にあたるタンパク質翻訳領域を効率よくシーケンスする全エクソーム解析が遺伝学解析の主流となっている。筆者も所属研究室のリソースを活用した全エクソーム解析を行うとともに^{11, 12)}、次世代シーケンサーで検出される膨大な数のバリエーションの病原性評価にシンプルかつスピーディーな酵母機能解析を応用することで

同一遺伝子の機能喪失性変異と獲得性変異が異なる臨床表現型と関連する事 (遺伝子型-表現型関連)¹³⁾ や病的意義不明のバリエーションの病原性を明らかにした¹⁴⁾。同一遺伝子の機能的に異なる変異が原因となって異なる表現型を発現した*PPP3CA* 遺伝子異常の例を以下に紹介する。

てんかんは、人口100人のうち0.5~1人にみられる最も頻度が高い神経疾患の一つである (発症率0.5~1%)。てんかんの原因は様々 (脳血管障害、脳腫瘍、脳外傷、感染症など) であるが小児の難治性てんかんについては遺伝的要因の関与が強く示唆されている¹⁵⁾。実際、難治性てんかんコホートを対象とした全エクソーム解析では患者のおよそ30-50%程度に遺伝子異常が同定される¹⁶⁾。これまでに頻度の高いメジャーなてんかん関連遺伝子が単離され遺伝的背景が明らかになる一方で、頻度の低いレアな原因遺伝子については検出頻度が低い故に同定されたバリエーションと患者表現型が偶然の合併であるのか、疾患原因であるのか見極めが難しい。従ってバリエーションの病的意義について遺伝子機能変化を含めた評価が重要となる。筆者らは難治性てんかん患者コホートの全エクソーム解析 (n=1,456) を端緒として6個のヘテロ接合性*PPP3CA* 変異を同定した (5個のミスセンス変異、1個のフレームシフト変異)¹³⁾。*PPP3CA* はタンパク質脱リン酸化酵素カルシニューリンの触媒サブユニットをコードする遺伝子で、ミスセンス変異のうち3個は触媒ドメインに、2個は自己抑制ドメインに位置した (図2A)。興味深いことに、たんぱく質ドメインにおけるバリエーションの位置と臨床表現型に対応関係を認め、触媒ドメイン変異は早期発症型てんかん性脳症、抑制ドメイン変異が多発先天奇形/知的障害、骨系統異常と関連していた (表1)。そこで「触媒ドメイン変異と自己抑制ドメイン変異による異なる機能変化が異なる疾患を引き起こす」という仮説を設定し機能解析を実施した。カルシニューリンは脳神経系に豊富に発現し、細胞内Ca²⁺濃度の上昇により活性化、多様な細胞内基質を脱リン酸化することで免疫系、神経系、筋骨格系を含む様々な器官の発生と機能に関わることが知られている。触媒ドメインが基質の脱リン酸化を担うのに対し、自己抑制ドメインは触媒ドメインの活性中心に近接し脱リン酸化活性を抑制している。細胞内Ca²⁺濃度が上昇すると触媒サブユニットのたんぱく質立体構造が変化し自己抑制ドメインによる抑制が解除されることで脱リン酸化活性が上昇すると考えられている。カルシニューリンシグナル系の代表的な脱リン酸化基質として転写因子NFAT (Nuclear Factor of Activated T cells) が知られている。NFATはカルシニューリンによる脱リン酸化依存性に細胞質から核内に移行し多数の遺伝子の発現を制御している¹⁷⁾。この一連のシグナルカスケードは進化的に酵母から哺乳類に至るまで高度に保存されていることから酵母機能解析系を用いてバ

図2 *PPP3CA*の機能喪失性変異と獲得性変異は臨床的に異なる疾患と関連する

A: 全エクソーム解析により同定された6個の病的バリエント。3個の触媒ドメイン変異、2個の自己抑制ドメイン変異、1個のフレームシフト変異を同定した。CNB, カルシニューリン B 結合ドメイン; CaMB, カルモジュリン結合ドメイン; AI, 自己抑制ドメイン。B: *ppb1* (酵母 *PPP3CA*) 欠損株は Ca^{2+} 感受性を示すが、野生型の *ppb1* を過剰発現させると正常な Ca^{2+} ホメオスタシスを回復し Ca^{2+} 添加培地で生育した (Wild type)。触媒ドメイン変異型 *ppb1* では回復を認めなかった (p.H121R, p.N179I, p.D263E)。自己抑制ドメイン変異型 *ppb1* は野生型と同様に Ca^{2+} 添加培地で生育した (p.F499L, p.A502T)。C: *ppb1* (酵母 *PPP3CA*) に変異を導入した knock-in 細胞株における転写因子 Prz1 のリン酸化状態をウェスタンブロット調べた。野生型では Ca^{2+} 添加後、脱リン酸化により Prz1 の移動度が早くなっている。触媒ドメイン変異 knock-in 細胞株ではリン酸化型 Prz1, 自己抑制ドメイン変異 knock-in 細胞株では脱リン酸化型 Prz1 が検出された。Cdc2, ローディングコントロール。D: *ppb1* (酵母 *PPP3CA*) 欠損株では培地への Ca^{2+} 添加に対する転写活性応答が起こらない (Empty)。野生型の *ppb1* を過剰発現させると転写活性は回復するが (Wild type)、触媒ドメイン変異型 *ppb1* (p.H121R, p.N179I, p.D263E) の過剰発現では不応性のままであった。自己抑制ドメイン変異型 *ppb1* (p.F499L, p.A502T) を過剰発現させると Ca^{2+} 刺激のない環境下で標的遺伝子の過剰発現を認め恒常的活性化が示唆された。E: 野生型では Ca^{2+} 刺激に応答した転写因子 Prz1 の核内移行が観察される (Wild type)。触媒ドメイン変異 knock-in 細胞株では Ca^{2+} 刺激後も基質である転写因子 Prz1 の核内移行は観察されなかった (p.H121R, p.N179I, p.D263E)。自己抑制ドメイン変異 knock-in 細胞株では Ca^{2+} 刺激のない環境下で核内移行を示す細胞の増加が認められた (p.F499L, p.A502T)。酵母 *ppb1* p.H121R, p.N179I, p.D263E, p.F499L, p.A502T はヒト *PPP3CA* p.H92R, p.N150I, p.D234E, p.F470L, p.A473TThr に対応する。(Mizuguchi T, et al.¹³) より改変)。

リアントによって引き起こされるカルシニューリンシグナル系の変化を調べた。具体的には1) 基質である転写因子 Prz1 (NFAT 類似の酵母転写因子) のリン酸化状態、2) Prz1 の核内移行、3) Prz1 により制御される遺伝子の転写活性、4) Ca^{2+} ストレスに対する細胞レベルの感受性を解析した。*ppb1* (*PPP3CA* の酵母ホモログ) 欠損

株は Ca^{2+} ホメオスタシスの破綻により Ca^{2+} に対して感受性を示す。そこでこの *ppb1* 欠損株に野生型、変異型 *ppb1* を過剰発現させる事でたんに質機能を評価した。野生型導入細胞が Ca^{2+} 添加培地で生育し、野生型の表現型を回復 (rescue) するのに対し、触媒ドメイン変異導入細胞では Ca^{2+} 感受性を示し表現型の回復がおこらないこと

表1 PPP3CA変異をもつ患者の臨床表現型

Individuals		1	2	3	4	5	6
Variants		p.D234E	p.N150I	p.H92R	p.M431Hfs*20	p.F470L	p.A473T
Clinical diagnosis		WS	WS, LGS, RS	WS	WS	MCA/ID	DD with seizures
Birth weight		N.D.	N.D.	-0.2 SD	-0.56 SD	+1.5 SD	-1.7 SD
Birth length		+1.9 SD	-0.4 SD	+0.1 SD	-0.07 SD	-3.4 SD	-6.4 SD
Seizure	Onset	7 m	8 m	22 m	6 m	-	9 m
	Type	ES and M	ES	ES	ES	-	G
	Prognosis	Intractable	Intractable	Intractable	Intractable	-	Tractable
Developmental delay		+	+	+	+	+	+
Intellectual disability		Profound	Profound	Severe	Severe	Severe	Moderate
Dysmorphism	Craniosynostosis	-	-	-	-	+	+
	Micrognathia	-	-	-	-	+	+
	Cleft palate	-	-	-	-	+	+
	Brachydactyly	-	-	-	-	+	+
Delayed calvarial ossification		-	-	-	-	+	N.D.
Skeletal	Short stature	-	-	-	-	+	+
	Arthrogyriposis	-	-	-	-	+	+
	Gracile bones	-	-	-	-	+	N.D.
	Fractures	-	-	-	-	+	-

Abbreviations: +, present; -, absent; y, years; m, months; N.D.; not determined; WS, West syndrome; LGS; Lennox-Gastaut syndrome; RS, Rett syndrome; MCA, multiple congenital anomalies; ES, epileptic spasm; M, myoclonus; G, generalized seizure.

から機能喪失性変異であることが示唆された (図2B)。実際、触媒ドメイン変異は転写因子Prz1の脱リン酸化、核内移行、標的遺伝子の転写活性の障害を引き起こし、遺伝子機能を失ったヌル変異であることが分かった (図2C-E)。これに対して自己抑制ドメイン変異はCa²⁺刺激のない環境下でPrz1の過剰な脱リン酸化、核内移行、転写活性化を引き起こし、抑制制御の破綻が示唆された (恒常的活性化) (図2C-E)。このことから触媒ドメイン変異、自己抑制ドメイン変異がそれぞれカルシニューリンシグナルの減弱と恒常的活性化を引き起こし、臨床的に異なる疾患と関連するとの結論に至った¹³⁾。

III. ゲノム異常発生分子機構の研究

上述の自作BACマイクロアレイ作製とアレイCGH解析を行っていた当時、世界的にも全ゲノムマイクロアレイによってゲノムコピー数多型・異常に関する知見が集積し始めた時期であった。このようなゲノム構造異常検出法の目覚ましい発展に比し、疾患の原因となる染色体転座、欠失、重複などのゲノム構造異常発生のメカニズムについては生物学の重要課題であるにも関わらず未解明な点が多かった。原因のひとつは、ヒトを対象とした

研究ではDNA塩基配列決定以上の解析が難しく実験的にアプローチすることが困難である点が挙げられる。生命現象の基本は生物種を超えて類似点が多い。そこで筆者は染色体・ゲノム研究モデルとしての酵母の有用性に着目し、ゲノムの3次元局所構造と空間配置という観点からゲノム異常発生分子機構の研究を行った。遺伝情報の担い手であるゲノムDNAはヌクレオソームを最小単位としたクロマチンを形成し核内に収納されている。このゲノムの物理的折り畳みと空間配置の制御は転写、DNA複製、DNA損傷修復、DNA組み換えなどあらゆる細胞機能とリンクし、ゲノムが安定に継承されるための重要な要素と考えられている。しかし従来の顕微鏡を用いた解析では生体内で刻々と変化するゲノム構造・核内配置をゲノムワイドかつ高解像度に解析することが困難である。そこでHi-C解析とよばれる分子生物学的手法によるゲノムの3次元構造解析に取り組んだ (図3A)。Hi-C解析は空間的に近い位置に存在するDNAをホルムアルデヒド固定後に制限酵素で切断、連結し、連結されたDNA断片 (空間的近接を反映したキメラDNA) を次世代シーケンサーでシーケンス解析する。連結が起こった頻度 (シーケンスによって読まれた回数) から空間的な近接関係を

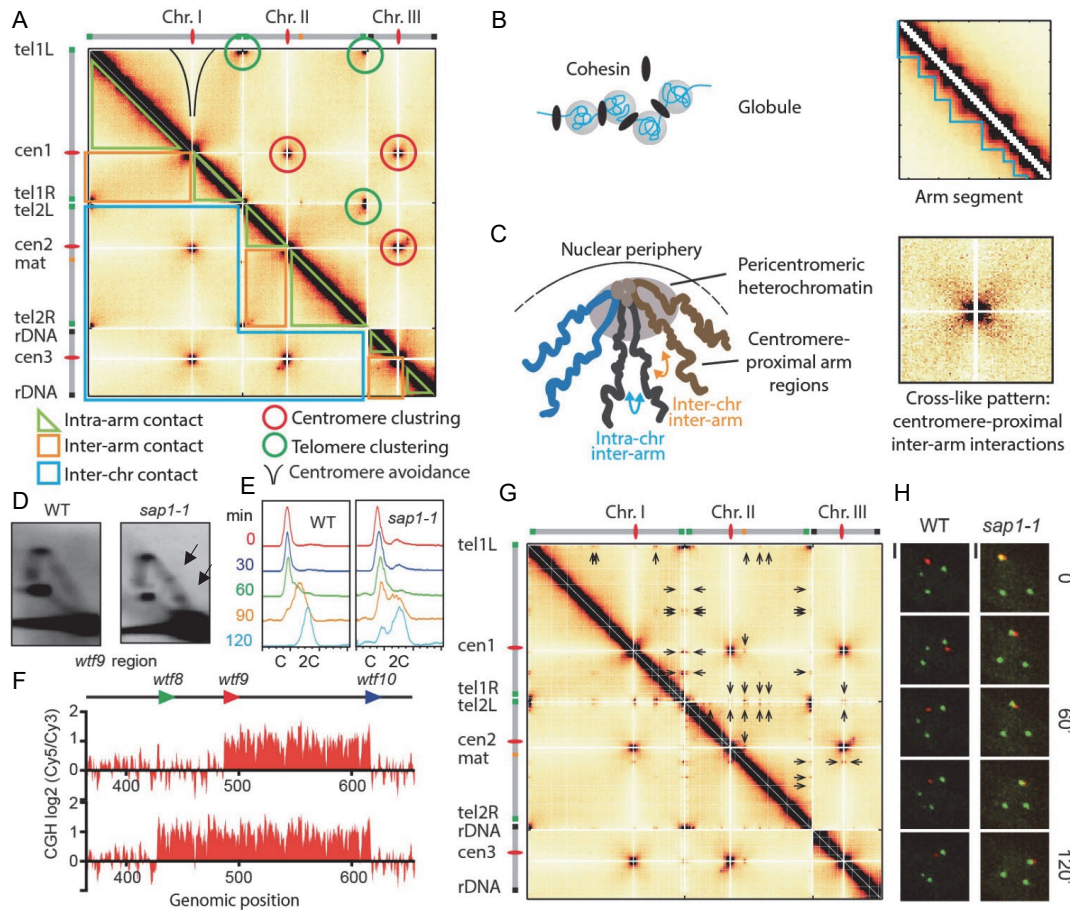


図3 Hi-C解析によるゲノムワイドコンタクトマップ

A: 野生型のゲノムワイドコンタクトマップ。コンタクト頻度は intra-arm, inter-arm, inter-chromosomal contact の順に減少し間期核において染色体テリトリーの傾向があることが示唆された。セントロメア、テロメア周辺においてそれぞれコンタクト頻度の増加を認めセントロメアおよびテロメアのクラスター形成が示唆された。B: 染色体腕において、コヒーシンによって境界されたドメインが存在した。C: 同一染色体および異なる染色体間のセントロメア周辺領域にコンタクト頻度の増加を認め (Cross-like pattern) セントロメアが中心体の近傍にクラスターしたラブル配向と呼ばれる規則性のある配置をとることが示唆された。D: 2次元ゲル電気泳動による DNA の複製中間体の解析。sap1-1変異体ではリピード配列である wtf9 領域に DNA 複製フォークの停止を示唆するシグナルが検出された (矢印)。E: フローサイトメトリーによる細胞周期の解析。G1 停止させた細胞 (c) の DNA 複製を再開させると野生型が120分で DNA 複製を完了 (2c) させるのに対し sap1-1 変異体では複製の遅延、複製障害が観察された。F: array CGH によるコピー数解析。DNA 複製が障害された sap1-1 変異体の培養を継続するとリピード配列を介したゲノム構造異常が誘導される。ここでは重複によるゲノムコピー数の増加が検出されている。G: DNA 複製が障害された sap1-1 変異体のゲノムワイドコンタクトマップ。野生型には認められない特異的なコンタクトが観察された (矢印)。H: 生細胞タイムラプスイメージング。ゲノム上の2領域を異なる色素 (赤と緑のシグナル) で可視化し変異体でのみ2つのシグナルの共局在を観察した。(Mizuguchi T, et al.¹⁸⁻²⁰ より改変)。

明らかにし、この情報を基にゲノムの3次元構造を解釈する手法である。この解析により、間期核において染色体テリトリーの傾向がある事、セントロメアおよびテロメアがクラスターを形成している事、異なる染色体のセントロメアがクラスターシラブル配向と呼ばれる規則性のある配置をとる事、染色体腕においてゲノム配列上離れていても空間的には近接しているゲノム領域・ドメイン (ヒトにおける TAD, Topologically Associating Domain に類似のユニット) が存在する事が示唆された。更にコヒーシン変異体、ヘテロクロマチン変異体の Hi-C 解析を行い、野生型のデータと比較することで、ゲノムの高次構造形成に重要な2つの要素とその形成にかかわる分子を明らかにした^{18, 19)}。染色体腕において、コヒーシンに

よって境界されたドメインが存在し、その形成はコヒーシン機能に依存した (図3B)。一方、ヘテロクロマチンはセントロメアヘテロクロマチンの凝縮を介してセントロメア辺縁領域の空間的近接と染色体テリトリーに影響を与えることを明らかにした (図3C)。この予備実験を踏まえ、DNA複製が障害される酵母変異体 sap1-1 (高温培養下で、複製フォークの停止 (図3D)、DNA複製障害 (図3E) とゲノム再構成 (図3F) が誘導される変異体) を用いた Hi-C 解析を行い、DNA複製ストレスがゲノムの空間配置に与える影響を解析した。ゲノム構造異常について DNA複製障害が誘因となる事はよく知られた事実であるが、DNA複製障害 (原因) によりゲノム構造異常 (結果) が引き起こされる過程で、ゲノムが核内でどのよ

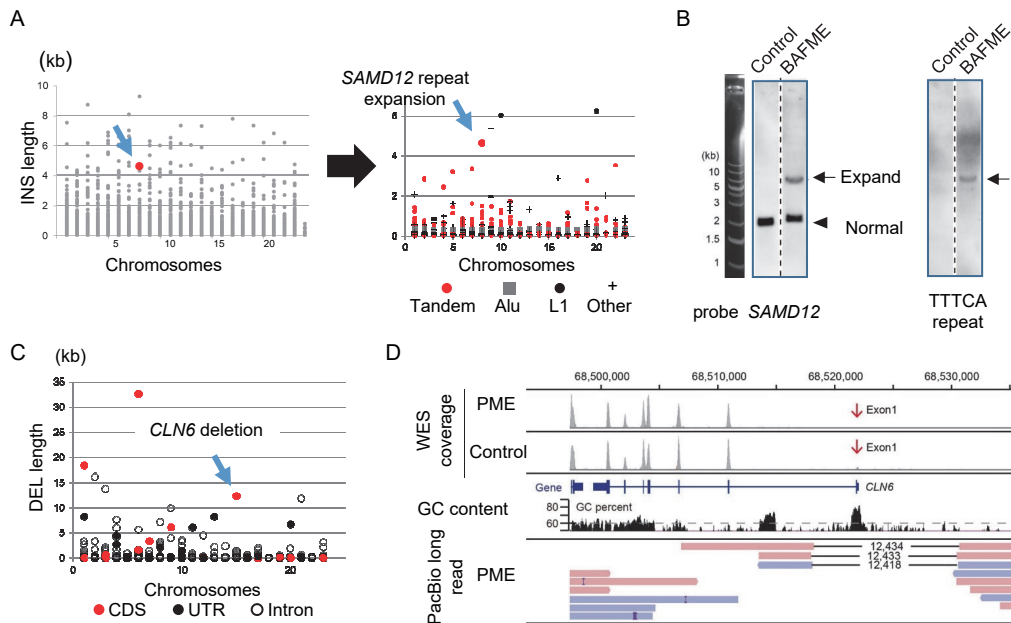


図4 ロングリードシーケンス解析法の確立

A: BAFME 患者のロングリード全ゲノム解析により検出された9,138個の insertions (INS). 3名の健常者コントロールデータに存在するバリエーションを除外することで候補を1,086個に絞り込んだ。更にBAFME 1 (8 q24) 座に位置する事、サイズが大きい事、シンプルリピート配列である事に着目することで *SAMD12* リピート伸長が有力な候補として抽出された。X 軸, 22本の常染色体と2本の性染色体; Y 軸, insertion の長さ (kb)。B: BAFME 患者のサザンブロット解析によりヘテロ接合性のリピート伸長 (矢印) が確認された。TGAAA リピートプローブで伸長アレル (矢印) にシグナルが検出され TTTCA リピートを含む伸長であることが確認された。C: PME 患者のロングリード全ゲノム解析により検出された7,216個の deletions (DEL) についてバリエーションの絞り込みを行い、*CLN6* 遺伝子を巻き込んだ12kbの病的欠失を検出した。CDS, たんぱく質翻訳領域; UTR, 非翻訳領域。D: 全エクソーム解析でのカバレッジ (上段)。GC 含量 (中段)。PacBio ロングリードシーケンスのリード (下段)。全エクソーム解析では欠失の存在するエクソン1領域に十分なシーケンスカバレッジが得られていないが (赤矢印)、ロングリードシーケンスではバイアスのない均等なカバレッジにより12kb欠失を検出した。(Mizuguchi T, et al.^{22, 23} より改変)

うな応答, 変化をするのかゲノムワイドに検出することはこれまで不可能であった。Hi-C解析の結果, *sap1-1* 変異体特異的なクロマチン間相互作用を検出し, DNA複製ストレス下における空間配置の変化を捉えることに成功した²⁰⁾ (図3G)。Hi-C解析の結果をもとに, 生細胞イメージングでゲノム上の2領域を異なる色素 (赤と緑のシグナル) で可視化したところ変異体でのみ2つのシグナルの共局在を観察し変異体特異的なコンタクトを確認した (図3H)。更にテロメアタンパク質複合体 Shelterin 構成因子を欠失させた細胞 (*sap1-1* とテロメアタンパク質複合体構成因子の二重変異体) ではこの変異体特異的なクロマチン間相互作用が消失したことから, この変化がテロメアタンパク質複合体 Shelterin 構成因子に依存的であることを突き止めた²⁰⁾。DNA複製の適切な制御はゲノム安定性に必須であり, その破綻は染色体の構造異常につながる。Hi-C解析がゲノム構造と機能との関わりを研究するのに有用な解析手法となりうる事を示した。

IV. ロングリードシーケンス解析法の確立

現在, 希少遺伝性疾患の40-60%程度は原因が未解明である。この事実は現状の遺伝解析法では技術的に検出できない遺伝的变化が潜在することを示唆する。ロング

リードシーケンサーはショートリード解析でシーケンスが難しいゲノム難読領域 (リピート, GCリッチな配列など) の解析が可能で, これまでの変異解析を補完する新規解析技術として最適である。これまでは微生物研究など適応が限定的であったが, シーケンサー性能が向上し技術的にヒトのロングリード全ゲノム解析が可能状況となってきた。実際ロングリード解析を用いる事でこれまでのショートリード解析で見逃されてきた構造多型 (SV, Structural Variant) が多数検出され疾患との関連が注目されている。現在この膨大な数のSV (ロングリード全ゲノム解析では1人あたり約15,000個のSVが検出される) から病的変異を抽出するための解析フローの構築が課題となっている。シーケンスリード長150bpのショートリード解析に比し, リード長が20kbを超えるロングリード解析ではリピート配列全長を一本のリードでカバーできる事からこのロングリードの特性を最大限に活用できるリピート病をモデルとしてこの課題に取り組んだ。良性成人型家族性ミオクローヌスてんかん (BAFME, Benign Adult Familial Myoclonic Epilepsy) は常染色体優性遺伝で成人発症の皮質振戦と稀発全般てんかん発作を主徴とする疾患である。連鎖解析によって疾患遺伝子座が染色体8q24領域にマップされていたが, 2018年に

SAMD12 遺伝子のTTTCAの5塩基リピート伸長が原因である事が明らかとなった²¹⁾。そこで既知原因遺伝子*SAMD12* 遺伝子のリピート伸長を抽出する解析系の確立を目的としてBAFME患者のロングリード全ゲノム解析を行った。15,636個のSVがコールされたが(9,138個の挿入, 6,498個の欠失), 1) コントロールに存在しない, 2) *BAFME1*座(染色体8q24)に位置する, 3) サイズが大きい構造多型, 4) シンプルリピート配列であることに着目する事で*SAMD12* 遺伝子内のリピート伸長を有力な候補として抽出する事ができた(図4A)²²⁾。

さらにこのSV解析系を用いて全エクソーム解析で変異未同定の進行性ミオクローヌスてんかん(PME, Progressive Myoclonus Epilepsy)の兄妹発症例を解析した。兄妹はともに4歳頃より運動失調, ミオクローヌス, 全般てんかん発作を発症し, 神経学的退行により寝たきりである。候補遺伝子変異スクリーニング, 酵素診断, 皮膚生検を用いた電子顕微鏡所見で異常を認めず, 全エクソーム解析を実施したが原因の同定には至らなかった。同様の臨床経過をたどる重度の神経変性疾患の兄妹例であり遺伝要因が示唆された。そこで全エクソーム解析でカバーされない領域(イントロンや遺伝子間領域)やゲノム難読領域に病的変化がある可能性を想定しロングリード全ゲノム解析を行った。17,165個のSVコールのなかから(9,949個の挿入, 7,216個の欠失), 病的意義が示唆されるバリエーションの選択(フィルタリング)を行った。フィルタリング後に残った1,451個の欠失のうち, 機能的に影響が大きいと考えられる遺伝子たんぱく質翻訳領域を含むこと, 影響を受ける遺伝子の機能に着目することで, 染色体15q23領域に12kbの病的欠失を同定した。この12kb欠失は神経変性疾患セロイドリボフスチン症の原因遺伝子*CLN6*のエクソン1を巻き込んだ欠失であった。エクソンを含む欠失であるにも関わらず全エクソーム解析でなぜ検出できなかったかについてゲノム配列の特徴を検討したところ, 該当領域は高GC含量, 多数のリピートエレメントに特徴づけられるゲノム難読領域で, 解析に必要なシーケンスカバレッジ(データ量)が十分に得られていないことが明らかになった。全エクソーム解析で未解決の症例の一部は, このようなシーケンスカバレッジの偏重が原因となって病的構造異常が見逃されているケースが潜在している可能性が示唆された(図4B)²³⁾。今後, ロングリード全ゲノム解析によってこれまでショートリードで読めなかった領域のゲノム変化が明らかになることで繰り返し配列の異常や複雑な構造異常に起因する疾患の原因究明が進展することが期待される。

おわりに

現在ゲノム解析の主流である全エクソーム解析の変異検出率は対象疾患や解析コホートにもよるが, おおむね30-40%と言われている。診断は適切な支援を受けるための出発点であり今後も新旧解析技術を駆使して新たな疾患原因遺伝子が同定されこれらの遺伝情報が, ますます医療に活かされることが期待される。

謝 辞

本研究は, 横浜市立大学倫理委員会の承認を得て行った。研究にご協力いただいた患者様とご家族の皆様, 共同研究していただいた諸先生方, およびご指導いただきました横浜市立大学医学研究科遺伝学教室の松本直通教授と研究室の先生方に深謝申し上げます。

文 献

- 1) Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE: Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*, **14**: 681-691, 2013.
- 2) Ferlini A, Scotton C, Novelli G: Biomarkers in rare diseases. *Public Health Genomics*, **16**: 313-321, 2013.
- 3) Mizuguchi T, Collod-Beroud G, Akiyama T, et al: Heterozygous *TGFBR2* mutations in Marfan syndrome. *Nat Genet*, **36**: 855-860, 2004.
- 4) Dietz HC, Cutting GR, Pyeritz RE, et al: Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature*, **352**: 337-339, 1991.
- 5) Collod G, Babron MC, Jondeau G, et al: A second locus for Marfan syndrome maps to chromosome 3 p24.2-p25. *Nat Genet*, **8**: 264-268, 1994.
- 6) Neptune ER, Frischmeyer PA, Arking DE, et al: Dysregulation of TGF- β activation contributes to pathogenesis in Marfan syndrome. *Nat Genet*, **33**: 407-411, 2003.
- 7) MacFarlane EG, Haupt J, Dietz HC, Shore EM: TGF- β Family Signaling in Connective Tissue and Skeletal Diseases. *Cold Spring Harb Perspect Biol*, **9**: 2017.
- 8) Sakai H, Visser R, Ikegawa S, et al: Comprehensive genetic analysis of relevant four genes in 49 patients with Marfan syndrome or Marfan-related phenotypes. *Am J Med Genet A*, **140**: 1719-1725, 2006.
- 9) Mizuguchi T, Hashimoto R, Itokawa M, et al: Microarray comparative genomic hybridization analysis of 59 patients with schizophrenia. *J Hum Genet*, **53**: 914-919,

- 2008.
- 10) Saitsu H, Kato M, Mizuguchi T, et al: De novo mutations in the gene encoding *STXBPI* (*MUNC18-1*) cause early infantile epileptic encephalopathy. *Nat Genet*, **40**: 782 – 788, 2008.
 - 11) Mizuguchi T, Nakashima M, Kato M, et al: *PARS2* and *NARS2* mutations in infantile-onset neurodegenerative disorder. *J Hum Genet*, **62**: 525 – 529, 2017.
 - 12) Mizuguchi T, Nakashima M, Moey LH, et al: A novel homozygous truncating variant of *NECAP1* in early infantile epileptic encephalopathy: the second case report of EIEE21. *J Hum Genet*, **64**: 347 – 350, 2019.
 - 13) Mizuguchi T, Nakashima M, Kato M, et al: Loss-of-function and gain-of-function mutations in *PPP3CA* cause two distinct disorders. *Hum Mol Genet*, **27**: 1421 – 1433, 2018.
 - 14) Uchiyama Y, Yanagisawa K, Kunishima S, et al: A novel *CYCS* mutation in the alpha-helix of the *CYCS* C-terminal domain causes non-syndromic thrombocytopenia. *Clin Genet*, **94**: 548 – 553, 2018.
 - 15) El Achkar CM, Olson HE, Poduri A, Pearl PL: The genetics of the epilepsies. *Current neurology and neuroscience reports*, **15**: 39, 2015.
 - 16) Tsuchida N, Nakashima M, Kato M, et al: Detection of copy number variations in epilepsy using exome data. *Clin Genet*, **93**: 577 – 587, 2018.
 - 17) Rusnak F, Mertz P: Calcineurin: Form and function. *Physiol Rev*, **80**: 1483 – 1521, 2000.
 - 18) Mizuguchi T, Fudenberg G, Mehta S, et al: Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature*, **516**: 432 – 435, 2014.
 - 19) Mizuguchi T, Barrowman J, Grewal SI: Chromosome domain architecture and dynamic organization of the fission yeast genome. *FEBS Lett*, **589**: 2975 – 2986, 2015.
 - 20) Mizuguchi T, Taneja N, Matsuda E, et al: Shelterin components mediate genome reorganization in response to replication stress. *Proc Natl Acad Sci U S A*, **114**: 5479 – 5484, 2017.
 - 21) Ishiura H, Doi K, Mitsui J, et al: Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat Genet*, **50**: 581 – 590, 2018.
 - 22) Mizuguchi T, Toyota T, Adachi H, Miyake N, Matsumoto N, Miyatake S: Detecting a long insertion variant in *SAMD12* by SMRT sequencing: implications of long-read whole-genome sequencing for repeat expansion diseases. *J Hum Genet*, **64**: 191 – 197, 2019.
 - 23) Mizuguchi T, Suzuki T, Abe C, et al: A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing. *J Hum Genet*, w : 359 – 368, 2019.

Abstract

IDENTIFICATION OF DISEASE-CAUSING GENES AND ANALYSIS OF MOLECULAR MECHANISMS FOR GENOMIC ALTERATIONS

Takeshi MIZUGUCHI

Department of Human Genetics, Yokohama City University Graduate School of Medicine

Genetic studies have identified the genes responsible for approximately 60% of rare monogenic disorders (>4,000 rare diseases). However, 40-60% still remain without genetic answers. Types and sizes of genetic alterations are highly variable, and no single genetic analysis method can detect the entire size range of these variants. Thus, application of new technologies with different resolutions is useful to find a genetic answer in unresolved cases. In fact, systematic discovery of new types of variants has been facilitated by technological advances that have accelerated disease gene discovery. In this review, Mendelian disease gene discovery and study of the pathomechanisms using a variety of methods will be introduced.

