

## FHSD の病態解明及び新規遺伝子診断法に関する研究

研究代表者 田中 裕二郎 東京医科歯科大学難治疾患研究所遺伝生化学分野 准教授

### 研究要旨

顔面肩甲上腕型筋ジストロフィー（Faciocapulothoracic muscular dystrophy, FSHD）は、4番染色体テロメア近傍のD4Z4反復配列にコードされるホメオボックス転写因子DUX4の異常発現が原因と考えられている。D4Z4反復配列の短縮、SMCHD1遺伝子変異、ASH1ヒストンメチル基転移酵素がD4Z4の転写活性化に関係することから、FSHDはエピジェネティックな疾患である。その複雑な病態のために、本邦ではFSHDの基礎的な分子機構の解析が遅れていた。本研究では、我々がこれまでに蓄積してきたクロマチン制御因子に関する解析手法を駆使し、FSHDの病態解明と新しい分子標的治療薬及び遺伝子診断法の開発を目指している。

これまで、DUX4プロモーター・レポーターの構築、ASH1及び関連クロマチン制御因子遺伝子のノックアウト、一分子DNAシーケンサーによるD4Z4反復配列のde novoシーケンシングを行って来た。本年度は、ASH1ノックアウトマウスの表現型解析、non-coding RNAノックダウンオリゴヌクレチドを用いたエピジェネティック治療薬の開発、PaBio RSを用いたD4Z4 BACクローン解析データを用いたde novoアセンブリのパラメータ解析をお行い、さらにゲノムから4番染色体のD4Z4領域を精製する方法を開発することによって簡便な遺伝子診断法を確立した。

### A. 研究目的

FSHDは、4番染色体テロメア近傍にあるD4Z4反復配列のクロマチン構造の変化を基盤とするエピジェネティックな疾患であると考えられている。健常人では、D4Z4領域はヘテロクロマチン構造を取り、YY1, EZH2, HMGB2などのヘテロクロマチン関連因子が結合するとともに、DNAがメチル化されている。一方、FSHD患者ではD4Z4領域のヘテロクロマチンの指標となる因子の結合やDNAメチル化が低下している。その結果、FSHD患者ではD4Z4領域の転写活性が高まり、反復配列自体にコードされているDUX4蛋白質が産生され、それによって筋萎縮が引き起こされると考えられている。

D4Z4領域のクロマチン構造が変化する原因は2つ知られている。一つは、D4Z4の反復数が短縮すること（1型）。もう一つは、ヘテロクロマチン制御因子SMCHD1遺伝子の変異である（2型）。本邦では、国立精神神経疾患研究センターがFSHDの遺伝子診断を行っており、2002年から2010

年までの10年間に915症例のFSHDを診断している。この中、D4Z4が7~9コピーと比較的長い患者についてSMCHD1遺伝子のエクソンシーケンシングを行った所、実に2/3でSMCHD1の変異が見つかった（三橋・西野ら）。これは、従来考えられていたように1型と2型が厳密に分けられるのではなく、FSHDはD4Z4の反復数とSMCHD1を始めとする疾患作用因子（モディファイア）によって決まる疾患スペクトラムであることを示している。

本研究は、このようなFSHDのエピジェネティックな病態基盤に焦点を当て、D4Z4領域の転写制御機構の解明、分子標的治療薬の開発、新しい遺伝子診断法の開発を目指している。

本年度は、昨年度に引き続きASH1を始めとするクロマチン制御因子欠損細胞の作成、ASH1ノックアウトマウスの表現型解析、D4Z4トランスジェニックマウス（FSHD疾患モデル動物）の交配、D4Z4反復配列を含むBACクローンのシーケンシング及び解析アルゴリズムの検証（森岡

との共同研究) lncRNA を標的とするノックダウン人工核酸 LNA gapmer の合成(横田との共同研究) さらに患者ゲノムから D4Z4 領域を選択的に精製するための技術開発を行った。

## B. 研究方法

### 【クロマチン制御因子ノックアウト細胞の作成】

前年度に引き続き、CRISPR/Cas9 システムを用いて C2C12 マウス筋芽細胞株で ASH1 及び Suz12、G9a をそれぞれ単独に欠損する細胞に加え、ASH1 + Suz12、ASH1 + G9a を同時に欠失する細胞株を樹立した。

### 【ASH1 ノックアウトマウスの作成】

前年度に CRISPR/Cas9 システムを用いて樹立した ASH1 変異マウスと C57BL/6 との交配を進めるとともに、新たに ASH1 の SET ドメインを標的にした遺伝子操作を行い、post-SET ドメインのシステイン残基を欠損する変異マウスを樹立した。

ASH1 遺伝子変異による初期発生への影響を検証するとともに、ASH1 ヘテロ欠損マウスの交配によってホモ欠損マウスが生後 1 日目に致死性であることを見出した。

### 【D4Z4 トランスジェニックマウスの交配】

前年度に樹立した D4Z4 トランスジェニックマウスと C57BL/6 マウスとの交配を進め、一部は ASH1 欠損マウスとも交配を行った。

### 【D4Z4 ゲノム領域の精製技術開発】

CRISPR/Cas9 システムにおいて、ヌクレアーゼ活性を欠損する Cas9 (dCas9) に FLAG タグを付けたリコンビナント蛋白質 (FLAG-dCas9) を合成した。また、4 番染色体と 10 番染色体の D4Z4 相同領域の 3' 末端にある転写集結シグナル周辺 (pLAM 配列) で、それぞれ配列が異なる部位に対する選択的ガイド RNA (sgRNA) を合成した。FLAG-dCas9/D4Z4-sgRNA 複合体をマグネットビーズに固相化し、これと 4 番染色体または 10 番染色体の pLAM 配列を含むプラスミド DNA を反応させ、特異的結合を PCR によって検証した。

(倫理面への配慮)

本研究は、本学組換え DNA 実験委員会及び動物実験委員会の承認を得て行っている。

## C. 研究結果

### 【クロマチン制御因子ノックアウト細胞の作成】

ASH1 は HOX 遺伝子の発現維持に必要とされる Trithorax グループ因子の一員で、HOX 遺伝子の発現を抑制する Polycomb グループ因子と拮抗的に作用してクロマチン構造を制御する。FSDH に於いては、ASH1 が D4Z4 の転写活性化因子として、ヘテロクロマチン制御因子 SMCHD1 が抑制因子としてそれぞれ知られているが、D4Z4 における Polycomb 因子の役割は不明である。そこで、ASH1 を中心とするクロマチン制御因子がどのように D4Z4 の転写を制御するか解析するため、昨年度に引き続きマウス筋芽細胞株 C2C12 において ASH1、Suz12、G9a (Ehmt2)、SMCHD1 をそれぞれ単独或いは複合欠損する細胞を樹立した。

Ash1、Suz12、G9a についてはそれぞれ 12~30 クローン当たり 1 個のホモ欠損細胞を樹立することに成功した(資料 3A)。さらに Ash1 欠損細胞を用いて Suz12 または G9a をノックアウトし、Ash1 + Suz12 及び Ash1 + G9a の組み合わせでホモ複合欠損する細胞株を樹立した。

今後、これらの細胞株へ D4Z4-BAC クローンを導入し、D4Z4 の発現制御機構を解析するとともに、C2C12 細胞分化系におけるクロマチン制御因子の役割を明らかにする予定である。

### 【ASH1 ノックアウトマウスの作成】

In vivo において ASH1 が D4Z4 の転写をどのように制御するか解析するために、昨年度に引き続き CRISPR/Cas9 システムによる ASH1 ノックアウトマウスの作成と、C57BL/6 バックグラウンドへの交配を進めた。これまでに、ASH1 のエクソン 2 の 2 箇所に於ける 4bp 欠損、13bp 欠損、1bp 挿入といういずれもフレームシフトを来すナンセンス変異体を樹立した(資料 3B)。

これまでに ASH1 ヘテロ欠損マウスの交

配によって、生後0日まではメンデルの法則に従ってホモ欠損マウスが誕生し、マクロ形態学的には胎生期を通じて特に異常を認めないことが分かった(資料3C)。また、ASH1がTrithoraxグループの一員であることから頸椎C1-C2のホメオティック変形が予想されたが、マイクロCT解析では異常所見は認めなかった(資料3D)。一方、生後1日を経過するとASH1ホモ欠損マウスは1匹も観察されないことから、生後致死性を示すことが明らかになった。現在、組織学的検索を行っている。

ASH1のSETドメインはヒストンH3リジン36(H3K36)を特異的にメチル化する(Tanaka et al., 2007)。また、ASH1をノックダウンするとD4Z4領域のH3K36メチル化が低下する(Cabianca et al., 2012)。しかし、H3K36メチル化がD4Z4の転写活性化に必要なかどうかは示されていない。そこで、メチル化活性を持たないASH1変異体を作るため、CRISPR/Cas9システムによる新たなターゲティングを行った。当初、活性中心にあるヒスチジン残基を1アミノ酸変異を狙ったが、代わりに補酵素であるS-adenosylmethionine (SAM)結合ポケットを形成するZnフィンガードメインのシステイン残基Cys<sup>2220</sup>を欠損する変異体を得られた(資料3E-F)。

#### 【D4Z4トランスジェニックマウスの交配】

ASH1欠損マウスの一部とD4Z4トランスジェニックマウスの交配を現在進めており、今後D4Z4転写活性化にASH1 あるいはそのヒストンメチル基転写活性が必須であるかどうかを検証する予定である。

#### 【D4Z4ゲノム領域の精製技術開発】

FSHDの遺伝子診断に向けて、現在の放射性同位元素を用いる方法からより簡便で将来的には保険適用も可能な新たな遺伝子診断法を確立するため、森岡との共同研究によってD4Z4領域のシーケンシング技術を開発するとともに、患者ゲノムからD4Z4だけを特異的に抽出する方法を検討した。これまで、PacBioシーケンシングにも使われているPhi29 DNAポリメラーゼの誘導体を用いてD4Z4領域の増幅を試みて来たが、本年度はさらにCRISPR/Cas9システムを

応用した新しい精製法を開発した。Cas9は標的DNAのPAM配列(*S. pyogenes*由来のCas9の場合NGG)を認識し、ガイドRNAと相補的な標的配列を切断する。Cas9のアミノ酸残基を2箇所置換しヌクレアーゼ活性を持たないCas9(dCas9)は、同様にPAM配列を認識しガイドRNAと相補的なDNAと複合体を形成するが、DNAを切断しないため強固に結合する性質を持っている。そこで、FLAGタグの付いたdCas9と、4番染色体のD4Z4配列 あるいは10番染色体のD4Z4相同領域の配列に特異的なガイドRNAを合成し、それぞれ標的DNAと選択的に結合するかどうかを検証した。

まず、FLAG-dCas9(大阪大学藤井穂高氏より譲渡)をpET24aベクターにサブクローニングし、BL21ホスト細胞に導入して30℃で発現誘導させ、リコンビナント蛋白(1,403アミノ酸)として精製した(資料3G)。4番と10番染色体に特異的なガイドRNAについては、それぞれ3箇所及び1箇所の標的配列をデザインし(特許申請中) in vitro転写反応によってRNAを合成した抗FLAG抗体によってマグネットビーズ(Dynabeads-Protein G)に固相化したFLAG-dCas9とガイドRNAを反応させ、4番染色体又は10番染色体の配列をサブクローニングしたpBSベクターと反応させた所、それぞれ予想された標的配列と結合することを確認した(資料3H)。

このようにして精製したD4Z4断片の大きさを測定するには、PicoGreen等でDNAを染色し蛍光顕微鏡下にDNAの長さを測定する(資料3I)か、またはDNAの5'または3'断端の濃度とDNA量の比を計算することによってDNAの平均長を概算する方法が考えられ、今後の検討課題である。

#### D. 考察

FSHDのエピジェネティック病態解明については、ASH1とヘテロクロマチン制御因子Suz12(Polycombグループ)、G9a(ユークロマチン抑制因子)のダブル欠損細胞株を樹立した。現在進めているSMCHD1とASH1のダブル欠損細胞株を含め、今後D4Z4-BACクローンを用いたD4Z4転写制御機構解析のための重要なツールとなる。

同様にD4Z4の発現制御機構をin vivoで

解析するためのASH1欠損マウスについても、エクソン2の二箇所におけるナンセンス変異を3種、SETドメインのシステイン残基を1アミノ酸インフレーム欠失する変異体を1種、それぞれ確立し、C57BL/6へのバッククロスを進めている。エクソン2のASH1ノックアウトマウスは、胎生期から生後0日まではメンデルの法則に従って正常に発生し、形態学的異常もホメオティック変異も観察されなかった。しかし、ASH1ホモ欠損マウスは生後1日以降は例外なく観察されなかったことから、ASH1欠失が生後致死性であることが明らかになった。この知見は、今後ASH1の機能を標的とするFSHDの治療薬を検証する際にはその副作用として慎重な検討が必要とされることを意味する。現在、ASH1が高発現する中枢神経、特にASH1と関係の深いMLLの標的遺伝子であるPITX2が発現している視床下部を中心に組織学的検索を行っている。

FSHDの新しい診断法については、2つの大きな進展があった。まず、PacBio RSによるシーケンシングについて、de novoアセンブリに成功したシーケンス情報を用いてシミュレーション実験を行い、PacBio RSの1セル当たりのアウトプットの数十分の1のデータ量でアセンブリが可能である（即ち遺伝子診断を大幅にコストダウン出来る）ことを示した。さらに、FLAG-dCas9リコンビナント蛋白質と4番染色体または10番染色体に特異的なガイドRNAを用いてD4Z4領域を高度に精製することが原理的に可能であることを示した。よって、外来患者の末梢血から精製したゲノムDNAを制限酵素処理し、これをマグネットビーズに固相化したdCas9/sgRNAのカラムを通してD4Z4領域のみ精製し、そのDNA断片の長さを直接測定するというワークフローが完成した。今後、本学の外来患者サンプルを用いた新規診断の検証、国立精神神経研究センターで遺伝子診断が確定している患者のサンプルを用いた信頼性の検証を進める予定である。

#### E. 結論

本研究は、病態解明、新規治療法の開発、

新規診断法の開発という三つの柱を掲げて進めて来たが、その中、病態機構については遺伝子ノックアウト細胞やASH1ノックアウトマウス、D4Z4トランスジェニックマウスといった研究材料の準備がほぼ完成し、これから実際の解析に移行する段階になっている。治療法についても、lncRNAに対するLNA gapmerの合成とアッセイ系（DBE-T発現ベクター、D4Z4トランスジェニックマウス）の確立まで進めることが出来た。診断法については、D4Z4領域のゲノムDNA精製からシーケンシングまたはDNA蛍光染色による長さの測定まで、基本的な条件をクリアすることが出来、三つの課題の中では最も進めることが出来た。本研究は三次公募であったため、実際の研究機関が3年に満たないことから一部予定通り完成していない課題も残されているが、今後早期の完成を目指して鋭意努力したい。

これまで、FSHDの診断には放射性同位元素を用いる方法しかなく、特に本邦では確定診断例が少なかった。海外では、これまで10万人に約5人と言われていたFSHD患者が12人という統計報告が昨年出ている。本研究を契機として日本筋ジストロフィー協会に設置されたFSHD分科会においても、臨床医からFSHDを疑われる症例は非常に多い印象があるとの発言もあり、今回作動原理を証明した新しい診断法について、保険適用を目指してさらに必要とされる検出感度と特異性の検証を進めて行きたい。

#### F. 健康危険情報

該当なし。

#### G. 研究発表

##### 1. 論文発表

Sequencing and *de novo* assembly of macrosatellite repeats of the FSHD locus. Morioka, MS, Tanaka, Y. (submitted)

*De novo* assembly of PacBio sequence data by homopolymer contraction method. Morioka, MS, Tanaka Y. (manuscript in preparation)

##### 2. 学会発表

第37回日本分子生物学会年会・平成

26年12月(横浜)・PacBio RSによる顔面  
肩甲上腕型筋ジストロフィー遺伝子座のシ  
ーケンシング(田中裕二郎・森岡勝樹)

de novo アセンブリの効率化)

2. 実用新案登録  
なし

H. 知的財産権の出願・登録状況

3. その他  
なし

1. 特許取得

学内審査中(ホモポリマー縮合による

## FSHD の新規分子標的治療薬に関する研究

研究分担者 横田 隆徳 東京医科歯科大学大学院脳神経病態学分野 教授

### 研究要旨

顔面肩甲上腕型筋ジストロフィー（Faciocapulo humeral muscular dystrophy, FSHD）は、4番染色体テロメア近傍にある D4Z4 反復領域の転写が活性化され、その結果 DUX4 ホメオボックス転写因子が発現されることが進行性筋萎縮の原因と考えられている。我々は、D4Z4 領域の転写活性化にはクロマチン構造制御因子の一つ ASH1 が必要であり、特に D4Z4 領域の 5' 側から転写される long non-coding RNA (lncRNA) の一種 DBE-T が ASH1 の SET ドメインに結合し ASH1 をリクルートすることを報告している (Cabianca et al., 2012)。

本研究は、FSHD の分子標的治療薬の開発を目指し、特に我々の研究グループが専門とする ASH1 と lncRNA との相互作用に着目して DBE-T に対するロックダウンヘテロ核酸を合成した。

### A. 研究目的

大部分の FSHD 患者では 4 番染色体テロメア近傍にある D4Z4 反復配列が短縮しているが、5' 側には全ての患者で残されている領域（NDE; non-deleted element）がある。我々は、NDE 領域から non-coding RNA の一種である DBE-T（D4Z4-binding element）が転写され、これが ASH1 の SET ドメインに結合することによって ASH1 を D4Z4 領域にリクルートすることを報告している (Cabianca et al., 2012)。

ASH1 は HOX 遺伝子群の発現維持に関わる Trithorax グループの一員であり、その SET ドメインによってヒストン H3 のリジン 36 を選択的にメチル化する (Tanaka et al., 2007)。これまでも他の SET ドメインが RNA と結合することは報告された例があるが、HOX 遺伝子以外で Trithorax 因子が non-coding RNA と結合することを示したのは、我々の報告が最初である。

D4Z4 領域の転写が活性化されると、D4Z4 反復配列の最も 3' 側の D4Z4 にコードされる DUX4 蛋白質が発現される。DUX4 はホメオボックス型の転写因子で、さらに下流の PITX1 という別の転写因子の発現を誘導し、その結果筋萎縮を引き起こすと考えられている。

D4Z4 領域の転写活性化から PITX1 の発

現に至る過程で、FSHD の分子標的治療薬となる可能性のある DBE-T, ASH1, DUX4, PITX4 の 4 つの標的候補の中で、我々は最も上流に位置する lncRNA が FSHD 患者の筋細胞でのみ発現されること（特異性）に注目し、DBE-T に対するロックダウンヘテロ核酸をデザインした。

様々な人工核酸（オリゴヌクレオチド）によるロックダウン技術の中で、架橋化核酸（Bridged Nucleic Acid）または LNA（Locked Nucleic Acid）として知られる新しい人工核酸は、二重鎖核酸のリボース立体構造を固定化することにより、標的核酸に対する結合親和性、ヌクレアーゼ耐性、さらに細胞膜透過性を高めるという特徴を有している。さらに、ロックダウンオリゴの構造を RNA-DNA-RNA というヘテロ核酸にすることにより、中央の DNA 部分が標的 RNA と結合して生じる DNA/RNA ヘテロ二重鎖を RNaseH が切断し、それによって標的 RNA を効率的に分解することが可能になる（資料 4.1A）。

本研究では、約 3.4Kb となる DBE-T に対して 20 種類の LNA 核酸を合成した。

### B. 研究方法

#### 【LNA gapmer の合成】

ジーンデザイン社 (Eziqon 社) のロック

ダウンオリゴのデザインアルゴリズムを用いて、DBE-T に対する 20 種類の LNA gapmer (ヘテロ核酸) を合成した。また、DBE-T の発現レベルを定量するための Taqman PCR プローブを合成した。

#### 【DBE-T 発現ベクターの作成】

LNA gapmer の活性を in vitro で評価するため、3,374bp の DBE-T を CMV プロモータ下にクローニングした発現ベクターを作成した (資料 4.1B)。

#### (倫理面への配慮)

本研究は FSHD 患者由来の細胞を使う予定であり、本学医学部倫理審査委員会 (承認番号 199) 及び難治疾患研究所倫理委員会 (承認番号 2014-012) の審査を経て承認を得ている。

### C. 研究結果

#### 【LNA gapmer の合成】

DBE-T に対する 20 種の LNA gapmer を合成した。

#### 【DBE-T 発現ベクター】

DBE-T は FSHD 患者細胞でのみ発現する。そこで、DBE-T のノックダウン効率をより簡便に定量するために、DBE-T を CMV プロモータ下にクローニングした発現ベクターを作成した。テンプレートとして、PacBio 解析にも用いた BAC クローン PR11-242C3 を用い、シーケンシングによって配列を確認した。DBE-T の一部は GC 含有率が極めて高く、PrimerSTAR GXL DNA ポリメラーゼ (タカラバイオ) によって増幅することを見出すまで時間が掛かってしまった。CMV-DBE-T 発現ベクターを HEK293 細胞に一過性発現させ、RT-PCR によって DBE-T の発現を確認した。

### D. 考察

FSHD の分子標的治療薬として、lncRNA の一種である DBE-T に対する LNA gapmer を 20 種合成した。

これらノックダウンオリゴの効果を評価するために、現在 DBE-T 発現ベクターを導入した細胞での Taqman-PCR による定量スクリーニングを行っている。

今後、D4Z4 トランスジェニックマウス由来の線維芽細胞 (MEF) または筋芽細胞でのノックダウン実験、さらに FSHD 患者由来の細胞 (末梢血リンパ球または iPS 細胞由来の分化筋芽細胞) を予定している。

### E. 結論

DBE-T に対する 20 種のヘテロ核酸 (LNA gapmer) を合成したが、アッセイ系の立ち上げが遅れたため (DBE-T 発現ベクター、D4Z4 トランスジェニックマウス、FSHD 患者細胞) 実際のスクリーニングは現在開始した段階である。

### F. 健康危険情報

該当なし。

### G. 研究発表

1. 論文発表  
なし
2. 学会発表  
なし

### H. 知的財産権の出願・登録状況

1. 特許取得  
なし
2. 実用新案登録  
なし
3. その他  
なし

## D4Z4 反復配列のシーケンシングに関する研究

研究分担者 森岡 勝樹 東京医科歯科大学難治疾患研究所声明情報学分野 助教

### 研究要旨

顔面肩甲上腕型筋ジストロフィー（Faciocapulo humeral muscular dystrophy, FSHD）の原因となる 4 番染色体テロメア近傍の D4Z4 領域は、約 3.3Kb の高度に保存された GC リッチな配列が数十コピー反復するマイクロサテライトリピートの一種である。従来の Sanger シーケンス法や次世代シーケンス法では、D4Z4 の反復単位よりも短い配列しか得られないため、それぞれの反復単位を正確に並べてアセンブルすることは不可能に近かった。実際、4 番染色体のテロメア近傍と、これと相同性の高い領域を含む 10 番染色体の領域は今日に至るまでシーケンス情報がほとんど無い。

本研究では、FSHD の新たな遺伝子診断技術の確立を目指して、D4Z4 領域を含む BAC クローンを用いて PacBio RS による de novo アセンブリを試みた。また、得られたシーケンス情報を用いて、実際に D4Z4 領域の de novo アセンブリに必要なリード数や配列条件をシミュレーション実験によって検証した。

### A. 研究目的

ヒト 4 番染色体のテロメア近傍には、3.3Kb のレトロ反復配列（D4Z4）が健常人では約 100 コピー存在する（資料 4.2A）。顔面肩甲上腕型筋ジストロフィー（FSHD）患者では、D4Z4 のコピー数が 10 以下に短縮しているか（1 型）またはヘテロクロマチン制御因子 SMCHD1 遺伝子の変異している（2 型）ことが知られている。その結果、D4Z4 のヘテロクロマチン構造が緩み、D4Z4 にコードされるホメオボックス転写因子 DUX4 が発現されることが筋萎縮の原因になると考えられている。患者の筋細胞で D4Z4 が転写されてしまうのは、ヒストンメチル基転移酵素 ASH1 が non-coding RNA によって D4Z4 領域にリクルートされ、D4Z4 の転写を活性化することが原因の一つと考えられており（Cabianca et al., Cell, 2012）、エピジェネティックな作用機序に着目した新たな治療法の開発が期待されている。

FSHD は 10 万人に 5~12 人の発症頻度と言われているが、その遺伝子診断には Southern プロット法や FISH（Molecular combing 法）による D4Z4 のコピー数評価が必要となるため、ハードルが高かった。

そこで、これらに代わる診断法として、一分子 DNA シーケンサーによるゲノムアセンブリが D4Z4 領域において原理的に可能かどうかを、D4Z4 を含む BAC クローンを用いて検証した。

4 番染色体の D4Z4 反復配列は、それぞれが高い相同性を持つばかりでなく（資料 4.2B）、ほぼ同じ反復配列が 10 番染色体上にも存在する。従って、3.3Kb のリピート配列を正しくアセンブルするためには、少なくともリピート単位よりも長いリードが必須となり、現状では PacBio RS に限られる。PacBio RS の長所は、第三世代シーケンサーに比べて得られる配列情報が圧倒的に長く（30Kb を越える）PCR を介さない一分子シーケンサーで GC 含有率によるバイアスが全く無いことである。一方 In/Del エラーが多い、得られるリード数が少ないといった短所もある。

### B. 研究方法

#### 【BAC クローンのシーケンシング】

D4Z4 領域全体を含む新たな BAC クローンは、BACPAC Resources (Children's Hospital Oakland Research Institute, USA) より入手した。このライブラリーのインサートの長さは平均 80Kb である。20 μ



g の DNA を D4Z4 反復の外側で切断する EcoRV で断片化し、濃縮 (AMPure) SMRT bell ライブラリ作成 (Pacific Bioscience) を行い、DNA/Polymerase Binding Kit: P4 または P5、Sequencing reagent: v2.0 (C2 または C3)、SMRT cell; v3 を用いて 1 セルのシーケンスを行った。

得られたシーケンスデータを用いて、PacBio による HGAP 解析パイプラインによる de novo アセンブリを行った。また、約 15 万リードのシーケンスデータの一部 (100~4 万) をランダムに選び、de novo アセンブリが可能かどうか、またどのような配列情報がアセンブリに影響するかをシミュレーション実験によって検証した。

(倫理面への配慮)

この実験は既存の BAC クローンを用いて行うため、倫理面での問題は発生しない。

### C. 研究結果

【de novo アセンブリの結果】

同じ BAC クローンを用いて、P4-C2 または P5-C3 の反応ケミストリを用いてシーケンシングを行った。一般的には P5-C3 の方がより長いリードが得られる。シーケンシングの結果、P4-C2 と P5-C3 ケミストリによって得られるリード数、マップされたサブリードの数や精度には大きな違いはなかったが (資料 4.2C)、HBAP2/3 解析パイプラインでは P4-C2 ケミストリのデータのみ de novo アセンブリによって D4Z4 全体をカバーするコンティグが得られた (資料 4.2D)。そこで、P4-C2 と P5-C3 の間でより詳細なリードの比較を行った所、リードの長さの分布には違いが無かったが、P5-C3 ケミストリで得られた配列は特に In/Del エラーが多いことが分かった (資料 4.2F)。

【シーケンスデータを用いたシミュレーション実験】

de novo アセンブリに成功した P4-C3 のデータを用い、ランダムに選んだ 100~4 万リードを用いて de novo アセンブリを行った結果、最小で 1000 リードから 4000 リードあれば、D4Z4 領域のアセンブリが可能であることが分かった (資料 4.2G)。即

ち、PacBio RS の 1 セル当たりのリード数約 15 万の 1/40~1/150 のデータでアセンブリが可能ということになる。さらに、このシミュレーションにおいてうまくアセンブリされたデータセットとアセンブリされなかったデータセットを比較検討したところ (資料 4.2H)、約 15Kb 以上の長さを持つ D4Z4 領域のリードが 1 つ以上含まれることが必須であることが明らかになった。

### D. 考察

これまで D4Z4 のような長い反復配列のシーケンシングは極めて困難であり、実際最新のヒトゲノムデータベースにおいても D4Z4 領域は大きなギャップとして残されている。今回、BAC クローンをを用いた検証実験ではあるが、PacBio RS によって 3.3Kb の D4Z4 配列を 13.5 コピー含む DNA を de novo アセンブリ出来ることが実証された意義は大きい。

実際のシーケンスデータを用いて、D4Z4 領域のアセンブリに必要な条件を検証し、うまくアセンブリするためにはリード数は必ずしも多い必要はなく、それよりも D4Z4 反復配列を 4 個以上含む長さのリードが 1 つ以上含まれることが重要であることが示された。FSHD 患者の大部分 (95%) は、D4Z4 が 10 コピー以下に短縮しているため、今回の検証実験の解析条件はそのまま患者の遺伝子診断に適用可能である。一方、健常人ではそれよりも長い反復配列を含むため、今回の BAC クローンよりも厳しい条件 (リード数と長さ) が要求される可能性が高い。

今回の解析では、EcoRV によって切断した DNA 断片をシーケンシング解析に用いている。患者の遺伝子診断では、このような EcoRV に相当するゲノム DNA を抽出する必要がある。そのために、我々は並行する研究で FLAG-dCas9 と D4Z4 特異的なガイド RNA により D4Z4 を含むゲノム DNA 断片を精製する方法を確立しており、今後実際の患者ゲノム DNA を用いた検証を進める予定である。

### E. 結論

PacBio RS により、D4Z4 を 10 コピー以上含む DNA 断片の de novo アセンブリに

成功した。また、現行の PacBio RS の 1 セル当たり得られるデータの数十分の 1 の少ないデータ量でも、比較的長いリードが含まれればアセンブリ可能であることを実証した。並行する研究で D4Z4 領域を特異的に抽出する技術も確立しており、今後患者ゲノム DNA を用いた遺伝子診断技術の検証を進める予定である。

#### F.健康危険情報

該当なし。

#### G.研究発表

##### 1.論文発表

Sequencing and *de novo* assembly of macrosatellite repeats of the FSHD locus. Morioka, MS, Tanaka, Y. (submitted)

*De novo* assembly of PacBio sequence data by homopolymer contraction method. Morioka, MS, Tanaka Y. (manuscript in preparation)

##### 2.学会発表

第 37 回日本分子生物学会年会・平成 26 年 12 月(横浜)・PacBio RS による顔面肩甲上腕型筋ジストロフィー遺伝子座のシーケンシング(田中裕二郎・森岡勝樹)

#### H.知的財産権の出願・登録状況

##### 1.特許取得

学内審査中(ホモポリマー縮合による *de novo* アセンブリの効率化)

##### 2.実用新案登録

なし

##### 3.その他

なし