

配列は、実際に切断点をまたがって異なるセグメント間に及ぶものが複数確認できるとともに、ペアエンド同士のリードが隣接する異なるセグメント間にアラインメントされていた事実からも、予想された部位と順序でリアレンジメントが起こっていることが確認された。最終的に確認された切断点の配列を Table 2 に示す。

次に、転座部位に関する正確な情報の得られていない TK6 細胞株を用いて WGS 解析により転座切断点の同定を試みた。TK6 細胞においてはこれまでの染色体解析より Fig. 13 に示すように、正常に近い核型を持つが、3 番と 21 番、および 14 番と 20 番染色体間に転座を有することが、multicolor-FISH 解析によりわかっている。この転座点の配列を WGS データより決定するために、次のようなアルゴリズムを構築した。イルミナシーケンサーからのリードは、インサート配列を両端から読んでいるため、同一インサートフラグメント約数百 bp に対して、両端から読んだ約 100bp のリード情報がペアで存在し、これらをメイトペアと読んでいる。通常メイトペアは染色体上の近傍にマップされるが、切断点をまたぐフラグメントからのシーケンス情報は、それぞれ別の領域にマッピングされることになる。この点を利用して、メイトペアがそれぞれ、3 番と 21 番、および 14 番と 20 番染色体上にマップされたリードの情報を抽出した。切断点をまたぐリードは残念ながら通常のマッピング解析ではアラインメントされず捨てられてしまうが、その近傍のリードについてはそれぞれ異なる部位に完全にマッチするためマップされたリードとして存在するので、このアルゴリズムで検出可能となる。

このようにしてマッピングデータから該当するメイトペアを抽出したところ、膨大な数の切断点の候補が得られた。ランダムなアラインメントエラーも考慮し、抽出されたリードが集

中している領域に関して、そのゲノム配列を UCSC ゲノムブラウザにて確認したところ、そのほとんどが、LINE-1, LTR などのゲノム上に散在する繰り返し配列の位置と一致した。即ち、こうした繰り返し配列はその相同性のため本来とは別の染色体上の位置にミスアラインメントされる可能性が高く、こうした解析をする場合に偽陽性結果を与えることがわかった。ただし、一部には各染色体に固有の領域にマッピングされるペアが存在し、これらは転座切断点の候補となる。今後得られた位置情報よりプライマーを設計し、切断点の増幅によるクローニングを行い、シーケンスの確認を行う予定である。

一方、別のアプローチとして、新たなアルゴリズムによるシーケンスデータ解析からの検討も行っている。本来マッピング操作時に、リファレンスシーケンスにはアラインメントされずに捨てられているデータの中に切断点を含むリードが含まれているはずであるので、これかの中から、部分的に完全にマッチして異なる部位にマッピングされる配列を拾い上げるためのアルゴリズムの構築を行っている。

C-3-2 1 分子シーケンサーを用いた高感度変異検出法の開発

TK6 細胞を代表的な変異原物質である ENU, MMS, γ 線にて処理し、tk 遺伝子を用いた遺伝子突然変異試験により、変異の誘発を確認した。前回の検討から得られた DNA の品質に問題があったため、再び新しい細胞を用いて実験を繰り返した。その結果、概ね前回と同様の結果が得られた (Table 4)。変異原物質処理による細胞の生存率は、2-8 割と処理により異なったが、tk 遺伝子の変異頻度はコントロールと比較して 40-100 倍と有意に増加した。これらの細胞から、今回は mtDNA エクストラクター CT

キット (WAKO) を用いて、ミトコンドリア DNA の抽出を行った。得られた DNA の電気泳動像を Fig. 14 に示すが、ゲノム DNA 由来と思われる全体にスミアなバンドの中に、16kb くらいのミトコンドリアサイズに相当するバンドも確認された。

これら DNA サンプルを元にして、PacBio シークエンサー用のライブラリーを調整し、1 SMART cell 分のシークエンス解析をおこなった。

その時のパフォーマンスを Fig. 15 に示すが、1 SMART cell あたり、トータルで 7 ~ 500Mb に相当するシークエンス情報が得られた。組み込まれたインサートサイズに相当する Subread length の平均長は 1.5kb ぐらいであり、リードクオリティーによるフィルター後のポリメラーゼ読み取り長が 8 kb 程度であることから、インサートは 4-5 回程度繰り返して読まれていることになる。サブリードごとの配列情報をそのままヒトミトコンドリアリファレンスシークエンス (hg38) にマッピングをすると、概ね 10% 近くの変異があり、かなりエラーレートが高いことがわかった。

そこで、Pacific Bioscience 社の解析パイプライン smartanalysis version2.3.0 に含まれる BLASR (PacBio long read aligner) プログラムを用いて、重複リードを考慮したマッピングをし、samtools (version0.1.19) を用いて変異コールをした結果、Table 5 に示した数の変異箇所が同定された。マッピングデータを可視化可能な Tablet ソフトウェアを用いて詳細に検討をしたところ、いずれの部位においても raw data 上は変異の存在が確認できた。すなわち共通して変化している部位に関しては、TK6 に元来存在している変異として検出された。

次に、問題となる新たな低頻度の変異の検出に関しては、TK6_cont-1 および ENU のサンプルにおいて、1000、および 2000bp 周辺の 2560

bese call に対してリファレンスと異なる bese call の数を計算したところ、cont-1 ではそれぞれ $48+61=109$ 、ENU では $36+27=63$ と、コントロールの方が高かった。変異 Call の頻度は、 $109/2560=0.043$ と 5% 弱であり、これはシークエンスエラーと考えられるため、誘発変異の検出のためには、よりエラー率を落とす必要があることがわかった。

C-3-3 BLM ノックアウト細胞を用いた細胞の遺伝的不安定性の評価

細胞の遺伝的不安定性を検出するためのモデル細胞として、国立医薬品食品衛生研究所変異遺伝部において開発された BLM 遺伝子ノックアウト細胞を使用した。BLM 遺伝子は染色体不安定性を示すブルーム症候群の原因遺伝子であり、DNA 二本鎖切断の修復酵素である DNA ヘリカーゼをコードしている。この遺伝子を破壊した TK6 細胞株は、親株に比べて高い染色体異常および突然変異誘発性を持つことが確かめられている。

この TK6 細胞 BLM 欠損株の遺伝子配列を、親株の TK6 細胞と比較して、突然変異およびコピー数変化を検出するため、ホールゲノムシークエンス解析を行った。また同時にミトコンドリアのシークエンス解析も行った。

WGS 解析により得られたリードをリファレンス配列 hg19 にマッピングすることにより TK6 および TK6/BLM 細胞の SNP 部位の抽出を行った。いずれも約 233Gb のデータから平均冗長度 73、カバー率 99.8% でマッピングされ、得られた SNP の数は 370 万箇所 に及んだ。このうち、二つの細胞で異なる SNP コールがされた箇所を抽出し、内容を吟味したところ、ほとんどがリファレンスに対して hetero SNP となる箇所のコールの選択の差によるものであり、新たに生じた変異であると考えられる箇所は僅かであった。

今回の検討においては、TK6/BLM 細胞を分離、培養後にクローニングを行わなかったことより、遺伝的不安定性により変異の誘発率が上がっている、NGS による検出が難しかったと考えられる。わずかに得られた真の変異はおそらく TK6/BLM 細胞の樹立過程でシングルコロニーアイソレーションが行われたことによる選択の影響であると考えられる。

TK6/BLM 細胞との比較とは別に今回得られた TK6 細胞の WGS データの解析から、CGH データの取得によるゲノム異常の検出、およびコピー数変化領域と SNP の高頻度領域が一致するという知見が得られた。17 番染色体ではゲノムコピー数が約 3 倍に増加している部分があり、この領域におけるマップデータを詳細に検討した結果、部分的に最大 6 種類の配列バリエーションがあることが判明した。このことは、3 倍に増加した計 6 本のアレルがすべて異なる配列を有していることを示している。この領域においては、SNP の発生頻度が高く、増幅を伴うゲノム異常との関連性が注目された。残念ながら、イルミナシークエンサーから得られる各リードは約 100bp と短いため、詳細な SNP の連鎖解析は難しかったが、今後 PacBio などロングリードのシークエンサーと組み合わせることにより、より詳細なアレル情報を取得したい。

TK6 細胞と BLM 欠損株の比較に関しては、プロテオーム解析からのアプローチも行っており、解析の結果得られた発現変化を示したタンパク質のリストを Table 7 に示す。

LC-MS/MS を用いたショットガンプロテオミクス解析で得られたペプチドピークの数はいくつかの総数 85845 であり、同時に取得した MS/MS データよりデータベース検索ソフトウェア MASCOT により同定されたタンパク質の総数は 1,985 個であった。このうち、タンパク質レベルの解析において、MASCOT による同定結

果の信頼性スコア 10 以上のペプチドに関して、タンパクレベルで親株に対して 2 倍以上の変化を示しかつ ANOVA 解析の p 値 0.05 未満で有意となるタンパク質の総数は 12 個であった。今後これらのタンパク質の機能と BLM 遺伝子破壊によるゲノム安定性との関連について検討を行いたい。

C-3-4 ProteoMap ソフトウェアによる Web 上でのタンパク質プロファイル情報提供

我々はこれまでに、LC-MS/MS を用いたショットガンプロテオーム解析により得られた各種細胞のタンパク発現プロファイルに関するリファレンス情報の提供と細胞間のデータ比較を可能とするためのソフトウェア「ProteoMap」を開発した。

通常質量分析装置から得られるデータは、膨大な数の数値データであり、このままではその全貌および詳細をつかみにくいことから、リテンションタイムと質量数(m/z)を各軸に取った 2 次元マップ上にイメージデータとして変換して可視化を行うことにした。この際、各ペプチドピークに対してタンデムマス (MS/MS) 測定が行われていた場合には、そのスペクトル情報が付随してくるが、これらも合わせて情報提供できるよう、クリックابلマップとして、対応するペプチドピークをマップ上でクリックした際に、ピーク情報がグラフとして表示される機能を加えた。また、MS/MS 測定がされたピークに対しては、MASCOT によるデータベース検索でのタンパク質同定結果の取り込みを行い、MASCOT 検索結果を表示させる機能も開発した。今回はこのソフトウェアの機能を利用して可視化したプロテオームデータをリファレンス情報として Web 上にて提供できるシステムを開発した。

概要を Fig. 17 に示す。2 次元マップ上 MS/MS データを持つペプチドピークは青または赤の

印がマークされるが、前者は MASCOT 検索にて同定結果が得られたピーク、後者は未同定のピークをあらわす。画面はズームイン機能を有し、それぞれのピークをクリックすることにより、MS/MS のスペクトルデータを表示させることができる。

本ソフトウェアは複数のサンプルのデータを取り込み、相互に比較することが可能であり、各ピークの濃度からおよその定量比較が可能であるが、今後より定量的な比較が可能となるよう改良を加えてゆきたい。

現在は外部サーバーにて試験的に稼働を行っているが、近日中に国立医薬品食品衛生研究所、遺伝子医薬部の HP 上にて公開を開始する予定である。

C-4 遺伝的安定性評価ツールとしての次世代シーケンサーの性能評価

C-4-1 HiSeq システム (イルミナ社) を用いたシーケンス解析

C-4-1-1 標準ゲノム DNA の品質評価

標準ゲノム DNA の品質を評価するために、二本鎖 DNA の量を測定した結果、DNA 総量は 20 μ g 以上であり、複数の読み深度 (depth) の条件でシーケンス (3 μ g 以上/1 解析) が行える量を取得することができた (Table 9)。また、核酸定量・アガロースゲル電気泳動を行い、濃度及び純度の品質検定を行った結果、すべてのサンプルにおいて問題が無いと判断された (Table 10, Fig. 18)。

C-4-1-2 標準ゲノム DNA のアレル頻度について

ゲノム上の 35 箇所において、デジタル PCR によって正確に測定されたアレル頻度が Table 8 に示されている。ただし、欠失による変異に関しては、どの位置の塩基が欠失されたかを特定することが困難であると判断し、一塩基置換

の箇所 (20 箇所) (Table 12) のアレル頻度のみに指標にして、次世代シーケンサーの精度について評価することとした。

C-4-1-3 ライブラリーの品質評価

品質確認を行った標準ゲノム DNA について、SureSelect XT Human All Exon v5 を用いてライブラリー作製を行い、それらライブラリーの品質を Agilent 2100 Bioanalyzer を用いて測定した結果、すべての検体においてクローニングサイズが約 200 bp 長からなるライブラリーを作製することができた (Fig. 19)。さらに、これらライブラリーを用いてシーケンスを行った結果、一検体あたりのペアエンドリード数が約 5 億個、総塩基数に換算すると約 50Gb 相当の配列がシーケンスされていることが確認できた (Table 11)。また、塩基配列をシーケンスするときには発生するエラー率を、以下の数式を用いて Phred クオリティスコアの値を算出した。

$$Q = -10\log_{10}p$$

その結果、Q30 (シーケンスエラーが生じる確率が 0.1%) のクオリティスコアが、いずれの検体についても 90%以上であることが確認できた (Table 11)。

C-4-2 シーケンサーの精度評価

C-4-2-1 シーケンスライブラリーごとのリード数に応じたエラー率評価

4 つの独立した標準ゲノム DNA の各ライブラリーについて、40Gbase, 30Gbase, 15Gbase 相当のシーケンスを行い、参照配列の既知変異箇所におけるシーケンスされた塩基種類の頻度を測定し、各変異箇所における 4 回の解析結果についてのバラツキの程度を CV 値 (変動係数) によって評価した。その結果、各シーケンスデータ量 (40Gbase, 30Gbase, 15Gbase) ごとに CV 値の平均を算出したところ、シーケンスデータ量が多いほど、バラツキの程度が低い

ことが確認された (Fig. 20). つまり, リード数を多く読むことで, 読み間違いを減らすことが可能であると考えられた. さらに, 変異箇所ごとについて, 同様に塩基種類の頻度を測定し, 各変異箇所における 4 回の解析結果についての CV 値を求めた. その結果, すべてのシーケンス量 (40Gbase, 30Gbase, 15Gbase) で CV 値が 20%以下であった塩基箇所は, 測定された全塩基箇所 20 個のうち 10 箇所であった. 一方, CV 値が 20%よりも極端に外れている測定箇所においては, 4 回のシーケンス解析のうち 1~3 回の外れ値が測定されているためであると考えられた. さらに, 実際の変異頻度が低い場合 (例えば, 塩基箇所#12 における変異頻度は 1%) においても, バラツキが大きくなることが確認された (Fig. 21).

C-4-2-2 すべてのライブラリーを統合して得られたシーケンスデータのエラー率評価

前述の実験で取得された 4 回のシーケンスデータをすべて統合し, 160Gbase, 120Gbase, 80Gbase, 40Gbase 相当のシーケンス量になるようにダウンサンプリングした. これらサンプリングサイズごとにマッピングを行い, 前述と同様に, 参照配列の既知変異箇所におけるシーケンスされた塩基種類の頻度を測定し, 各変異箇所におけるそれぞれのリード数 (160Gbase, 120Gbase, 80Gbase, 40Gbase) ごとのバラツキを CV 値によって評価した. Fig. 22 に示すように, 4 回のライブラリーを統合した場合の CV 値は, ほとんどの測定箇所において 20%以下であった. このことは, 複数回のシーケンスを行うことで, シーケンスデータ量に関係なく測定のバラツキを低減させることができることを示唆している. 一方で, 塩基箇所#3 においては, 40Gbase における変異頻度の測定値が他と極端に違うために, Fig. 22 に示しているような CV 値が 20%から大きくは外れた結果と

なってしまった. 次に, 実測値と公表値とのバラツキの程度についても解析を行った. Fig. 23 に示しているように, 測定箇所#7, 10, 13, 16, 18, 19 において, 実測値と公表値のバラツキが著しく高い結果となった. しかしながら, Fig. 22 の結果からも解るように, これら測定箇所におけるリード間のバラツキが低かったことから, これらの箇所における実測値と公表値のバラツキについては, シーケンサーでは正確に読み取りにくいゲノム配列 (構造) であることが推測される. または, 変異頻度の公表値を見直す必要も考えられるため, 今後は, 他のゲノム標準品を用いた解析も実施する必要があると思われる.

C-5 遺伝的安定性評価リファレンスとしての日本人ゲノムの *de novo* 配列決定

2014 年 12 月末日の段階で, 6 名の健常日本人男性 (JOM001, JOM002, JOM003, JOM004, JOM005, JOM006) から提供の申し出があり, これまで 2 回あるいは 3 回の射精分の検体提供を受けた. 本人が把握できる範囲内で先祖の国籍は全て日本であり, 出身地は本州または九州本土であることが確認された. 健康状態を正確に知ることはできないが, 6 名全てに実子がいることが確認された. それぞれの精液 250 μ L からゲノム DNA を抽出したところ, それぞれ 3.2 μ g, 1.1 μ g, 2.5 μ g, 7.7 μ g, 3.1 μ g, 4.0 μ g の収量であった (Fig. 24). 吸光定量を行い, ライブラリー作製のための品質に問題ないことを確認した. 1 回の射精分の精液量はおよそ 1 mL から 2 mL であった.

これら 6 サンプルから無作為に 5 サンプルを選び, 市販のアレイを用いた 2,294,794 サイトに対する SNP タイピングを行った (Table 13). その一方で 1000 人ゲノムプロジェクト (<http://www.1000genomes.org/>) のデータを利用し, 性染色体を除き, refSNP 番号が付けられ

ている SNP から、アセンブリ対象とするサンプルを決定するための参考になる SNP を選出した。現在利用可能なデータは 26 人種、合計 2504 人となっており、104 人の日本人が含まれていた。全データと比較して日本人に稀な SNP、ヨーロッパ人と比較して日本人に稀な SNP、東アジア人として比較して日本人に稀な SNP、また東アジア人には稀だが日本人に多い SNP を、それぞれ 22,172, 50,609, 5879, 1022 選び、今回調べた 5 サンプルがこれらをどのように持つかをまとめた (Table 14)。また、第 18 番染色体から無作為に 4000 の SNP を選び主成分分析を行った。比較サンプルとして、1000 人ゲノムデータからヨーロッパ人 (CEU) 4 名、アフリカ人 (YRI) 4 名、中国人 (CHB) 4 名、日本人 (JPT) 4 名を選び図に表した (Fig. 25)。以上の結果から、特に、他の東アジア人と区別されやすい JOM005 を選び、ペアエンド・ライブラリおよびメイトペア・ライブラリを作成し、*de novo* アセンブリを行うサンプルとした。このサンプルからはさらに 1750 μ L の精液を用い、73.0 μ g のゲノム DNA を得た。

ペアエンド・ライブラリはアダプタの配列を除いたインサート長がそれぞれ 260 bp, 360 bp, 660 bp の 3 種類を作成した。Illumina HiSeq を用いてシーケンシングを行い、それぞれ 2691 億塩基、1379 億塩基、1616 億塩基から成るリードデータを FASTQ 形式で取得した。メイトペア・ライブラリはインサート長がそれぞれ 2 kb, 5 kb, 9 kb の 3 種類を作成し、やはり Illumina HiSeq を用いてシーケンシングを行い、それぞれ 1022 億塩基、452 億塩基、461 億塩基から成るリードデータを FASTQ 形式で取得した (Table 15)。

260 bp のペアエンド・ライブラリのデータについて、トリミング処理を行い、ミトコンドリア DNA (mtDNA) の全ゲノム配列をアセンブリにより決定した。決定には、ヒトとマウスで保

存されている 5'-CCG TGC AAA GGT AGC ATA ATC ACT TGT TCC T-3' から両鎖に伸張させる、本研究グループが独自開発した GrepWalk を用いた (Hayashi et al. 2014)。その結果、16,570 bp から成る JOM005 mtDNA の完全長配列を得ることができた。リファレンス配列 NC_012920 との比較では、39 の SNP が見つかり、その全てがプリン間またはピリミジン間の transition で、transversion は一つもなかった。indel は 3 ヶ所に見られ、そのうちの 2 つは C/G ホモポリマーの長さの違いであった。

ペアエンド・ライブラリのデータのトリミングは、Cutadapt を用いてアダプタの除去、GrepWalk を用いて、両端の低品質塩基の除去、独自スクリプトでトリミングにより短くなったリードの除去を行った。メイトペア・ライブラリのデータのトリミングは、Cutadapt を用いて 3' 末端にアダプタ配列が現れるリードは除去後、Cutadapt によりジャンクション・アダプタの処理を行い、両端の低品質塩基の除去、独自スクリプトでトリミングにより短くなったリードの除去を行った。シーケンサから得られた塩基数に対し、アセンブリに用いることができる塩基数の割合は 2 kb, 5 kb, 9 kb のライブラリそれぞれについて 58.2%, 59.5%, 59.8% であった (Table 16)。今回新規に決定した mtDNA の完全配列をリファレンスとし、トリミング処理後のライブラリデータをマッピングすることでより正確なインサート長およびその偏り (標準偏差) を見積もり、*de novo* アセンブリにはこれらの値を用いた。

全てのライブラリデータを用いた *de novo* アセンブリには数ヶ月かかると予測されたため、現段階では一部のデータに限定し、ALLPATHS-LG を実行した結果、一倍体のゲノムサイズは 2,742 Mb、反復配列の存在比は 27.0% であると推定された (Table 17)。合計 2,255 Mb の contig が得られたので、ユニーク

な配列を持つ領域のほぼ全てをカバーすることができたと考えられる(Table 18). gap を含む scaffold の N50 は 45 kb に達した. 最長 scaffold は第 8 番染色体の一部と考えられる 1,300,683 bp であった.

第 3 世代シーケンサのためのライブラリは, 同サンプル JOM005 から得られたゲノム DNA を約 20 kb を目安に断片化を行い, 二本鎖両端にペアピンループ構造を持つ一本鎖アダプタを結合させることにより行った. パルスフィールド電気泳動によってサイズ分布を確認の上サイズ選別, 2 分して S1 および S2 の 2 ライブラリを作製した. それぞれのピークサイズは約 35 kb および 24 kb であった. ともに P6-C4 試薬を用い, PacBio RS II でシーケンシングを行った. S1 および S2 について, それぞれ 10 セルおよび 46 セル分のデータを取り, 合計 44,840 Mb におよぶリードデータを取得した(Table 19). それぞれの平均リード長は 12,066 bp および 9716 bp であり, 各リード長の分布をライブラリごとにヒストグラムに示した(Fig. 26). 先に完全長を決めた mtDNA の塩基配列と比較することにより, PacBio RS II から得られるロングリードは確かに mtDNA の全長を超える長さであることが確認された. これまでの報告では 10%程度のランダムなエラーが入るとされており, このことを確かめるため, mtDNA に由来すると考えられるいくつかのリードを選んで BLASTN でアラインメントしてみたところ, mtDNA 全体で 92%の identity が得られており, 最新試薬 P6-C4 により, 良好な配列データが得られていることが確認できた(Fig. 27). 各ロングリードの由来すると思われる染色体は, PacificBio Science 社が提供する BLASR を用いて調べた.

HiSeq によって得られたショートリードのアセンブル結果は, これらロングリードを用いて scaffold あるいは contig 間の gap を埋める作

業を進めている. 最終的な結果は論文発表時に一般公開予定で, 現在は次のサイトより限定公開を行っている(Fig. 28).

<http://epigenetics.nrichd.ncchd.go.jp/refjpn/>

本研究グループは, 正常出産を経験した 411 人の日本人女性のゲノムを高解像度 SNP アレイで調べ, 多くの日本人が共有するコピー数多型を報告している(Migita et al. 2014). その中でも第 2 染色体短腕側 52.77 Mb 辺り(2p16.2)にあると考えられる約 26 kb の欠失は 132 人の日本人女性で認められている. この欠失の有無を日本人男性 JOM005 で調べたところ, ヘテロで保持しており, かつ HiSeq によって得られたリードのアセンブルにより, 欠失部位は GRCh37/hg19 の chr2:52,749,687-52,785,272 で欠失長は 35,586 bp であることが判明した(Fig. 29).

C-6 分化プロペンシティを指標とした細胞特性解析法の開発

C-6-1 IPA を用いた iPS 細胞分化プロペンシティ予測マーカーの絞り込み

昨年度の研究報告書で示した分化プロペンシティ予測マーカー候補の中から, 細胞分化に機能的に関与している mRNA および miRNA を絞り込む目的で, 外胚葉, 中胚葉, および内胚葉で相関がみられた mRNA および miRNA の probe set について, IPA (Ingenuity) を用いて遺伝子ネットワーク・パスウェイ解析を行った. IPA では複数のデータベースを基に, miRNA のターゲットとなることが配列上予想される(またはターゲットであることが検証された) mRNA を, 複数の予測データベース (Target Scan Human) をもとにして調べることができ (miRNA - mRNA ペアリング解析). この解析によって選別された, ペアをつくる miRNA - mRNA は, 機能的に細胞分化に影響を及ぼしている可能性が高いと考えた. そこで, 各胚葉

において相関のあった mRNA と miRNA の間で、miRNA - mRNA ペアリング解析を行った (Tables 22-24). その結果、外胚葉マーカーの第一主成分得点と相関のあった miRNA - mRNA ペアとして 3 種類の miRNA (miR-373, miR-371-5p, miR-371-3p), 中胚葉マーカーの第一主成分得点と相関のあった miRNA-mRNA ペアとして 1 種類の miRNA (miR-524-5p), 内胚葉マーカーの第一主成分得点と相関のあった miRNA - mRNA ペアとして 6 種類の miRNA (miR-4739, miR-4505, miR-4521, miR-520g, miR-3714, miR-367) が選別された。

C-6-2 外胚葉分化における分化関連 miRNA の影響

ヒト iPS 細胞の外胚葉分化を負に制御すると予想される miRNA として、miR-373, miR-371-5p, miR-371-3p の 3 種類が同定された (Table 22, Fig. 22). そこで、ヒト iPS 細胞株 253G1 を用いてこれらの miRNA を過剰発現させた細胞を作製し、外胚葉由来細胞の一つである神経に分化誘導させたときに、miRNA の発現が神経分化にどのような影響を及ぼすかを調べた。miR-373, miR-371-5p, および miR-371-3p をヒト iPS 細胞に過剰発現させるため、Biosettia 社の miRNA 発現レンチウイルス粒子を使用した。本研究で用いたレンチウイルス粒子は、miR-373 と miR-371 (miR-371-3p, および miR-371-5p の両 miRNA 前駆体) の 2 種類の miRNA 前駆体を含む Human EF1 α promoter と puromycin 耐性遺伝子を持つものである (Fig. 23). このレンチウイルス粒子を MOI \approx 50 の濃度で細胞培地に添加し、253G1 株に感染させたのちに、1 μ g/ml の puromycin を添加することによって、①miR-373, ②miR-371-3p と miR-371-5p の miRNA を過剰発現させた 2 種類の 253G1 を作製した。また、control とし

て miRNA 前駆体を含まないレンチウイルス粒子を用いて 253G1 を感染させて 253G1 の miR-control 細胞もあわせて作製した。作製した 3 つの細胞の miRNA 発現量について、miRNA 定量 PCR を用いて定量した結果、control 株に比べ、miR-373 は約 3 倍、miR-371-3p は約 7 倍程度発現量が増加していることを確認した (Fig. 24). miR-371-5p については、miRNA の配列特異性が低く、TaqMan アッセイでのプライマー合成が不可能であったため、本研究では発現量を定量することができなかった。このようにして作製した miRNA 過剰発現 iPS 細胞を用いて、神経細胞への分化誘導実験を行った。miRNA 過剰発現 iPS 細胞をフィーダーレス条件下で培養後、Matrigel コートした 24 well プレート上に 3.6×10^4 cells/well で播種した。細胞がコンフルエントになるまで培養した後、Chambers らの報告を参考にして分化誘導を開始した。分化誘導開始から 10 日後に細胞を回収し、total RNA を抽出した。神経細胞への分化を確認するため、神経マーカー遺伝子である *PAX6*, *SOX1*, *NCAM1*, *TH*, *TUBB3*, および *NES* の発現を定量 PCR で測定した。その結果、miR-373 および miR-371-3p を過剰発現させても神経マーカーの発現に有意な差は認められなかった (Fig. 25).

C-6-3 中胚葉分化における分化関連 miRNA の影響

ヒト iPS 細胞の中胚葉分化を正に制御すると予想される miRNA として同定した miR-524-5p (Fig. 26) についても、レンチウイルスとヒト iPS 細胞株 253G1 を用いて miRNA 過剰発現細胞を作製した。作製した miRNA の発現量について、miRNA 定量 PCR を用いて定量した結果、control 株での発現量と比較して約 5 倍に増加したことを確認した (Fig. 27). miR-524-5p 過剰発現細胞を用いて、中胚葉由

来細胞の一つである心筋細胞に分化誘導させるときに、miRNA がどのような影響を及ぼすかを調べた。miRNA 過剰発現 iPS 細胞をフィーダーレス条件下で培養後、Matrigel コートした 6 well プレートに 2.0×10^5 cells/well となるよう播種した。本節でも、control として miRNA 前駆体を含まないレンチウイルス粒子を用いて作製した 253G1 の miR-control 細胞をあわせて使用した。細胞がコンフルエントになるまで培養した後、Lian らの報告を参考にして分化誘導を開始した。分化誘導開始から 15 日後に細胞を回収し、total RNA を抽出した。心筋細胞への分化を確認するため、*TNNT2*, *GATA4*, *NKX2.5*, および *MYH6* の 4 種類の心筋マーカー遺伝子の発現量を定量 PCR で測定した。その結果、miR-524-5p を過剰発現させても心筋マーカーの発現に有意な差は認められなかった (Fig. 28)。

C-6-4 内胚葉分化における分化関連 miRNA の影響

ヒト iPS 細胞の内胚葉分化を正に制御すると予想される miRNA として同定した miR-4739, miR-4505, miR-4521, miR-520g, miR-3714, miR-367 (Fig. 29) についても、レンチウイルスとヒト iPS 細胞株 253G1 を用いて miRNA 過剰発現細胞の作製を試みた。その結果、miR-4505, miR-520g の過剰発現 iPS 細胞は作製することができた。作製できた 2 つの miRNA の発現量について、miRNA 定量 PCR を用いて定量した結果、control 株に比べ、miR-4505 は約 2 倍、miR-371-3p は約 9 倍程度発現量が増加していることを確認した。しかし、miR-4739, miR-4521, miR-3714, miR-367 の 4 種類の miRNA 過剰発現 iPS 細胞を作製することができなかった (Fig. 30)。これら miR-4505 および miR-371-3p の過剰発現細胞を用いて、胚葉体を形成させることにより、分化を促した。

分化開始後 16 日目に胚葉体より total RNA を抽出し、内胚葉マーカーである AFP, SOX7, SOX17 の発現量を定量 PCR で測定した。その結果、miR-4505 および miR-371-3p を過剰発現させても内胚葉マーカーの発現に統計的に有意な差は認められなかった (Fig. 31)。

D. 考察

D-1 新規免疫不全動物を用いた造腫瘍性試験法の開発

BRGマウスにおけるHeLa細胞単体移植時のTPD50はNOGマウスと同等であり、BRG-nuマウスにおいてはNOGマウスを凌ぐ成績であった。BRG-hrマウスにおけるHeLa細胞単体移植時のTPD50はNOG-hrマウスに近似であり、異種細胞生着感度の点からヌード化(nu遺伝子導入)した系統が優位であることが考えられた。

なお、マトリゲル増強効果はNOGおよびNOG-hrマウスよりも低く、背景遺伝子の違い(NOD/ShiとBALB/c)による影響が示唆された。

D-2 幹細胞の *in vitro* 培養工程における遺伝子発現の動態解析による品質評価技術の開発

LINE-1s の転移を抑える細胞内因子として知られている A3B には、欠失多型が存在し、日本人にその割合が多いと報告されている。A3B を発現しない日本人由来 hMSCs は LINE-1s の転移によってゲノムの安定性が損なわれる可能性が考えられたので、A3B 遺伝子型と LINE-1s の発現量について解析を行った。次世代シーケンサーを用いて A3B 野生型ホモ (Ins/Ins) と A3B 欠失型ホモ (Del/Del) の RNA 配列を網羅的に解析し比較したところ、Ins/Ins と Del/Del で転移活性の残った LINE-1s の発現量には大きな違いが見られず、日本人に多いとされる A3B 欠失により hMSC における LINE-1s の転移によるゲノムの安定

性を損なう危険性は示されなかった。

さらに、iPS細胞やhMSCなど幹細胞におけるLINE-1sの発現は確認できたが、分化した正常組織ではその発現がほとんど認められないことから、hMSCにおけるLINE-1sの発現に及ぼす分化の影響について検討したところ、hMSCを脂肪分化させることにより、LINE-1s mRNAの発現量が低下することを確認した。このことから、LINE-1sの発現と分化能には関連が見られ、LINE-1sの発現がhMSCsの分化能を示すマーカーの一つとになり得る可能性が示唆された。

D-3 次世代シーケンサーを用いた細胞の遺伝的安定性評価指標の開発

今年度は、染色体転座などのゲノムリアレンジメントの検出および低頻度の突然変異の検出という観点において、次世代シーケンサーの応用に関して検討を行ってきた。既知の遺伝子転座を持つ細胞株の切断点の解析より、WGSデータ中に切断点を決定できるシーケンスリードの情報が存在することが確認できたが、通常のマッピング解析においては、こうした転座点をカバーするフラグメントはリファレンス配列への適合率が低く、アラインメントされない情報として利用できない。そこで、転座に特化したアルゴリズム、例えば全体としての合致率は高くないが、部分的には異なるゲノム上の配列に完全にマッチしている配列を拾い上げる方法により、転座のあるフラグメントのみを検出できる可能性がある。現在その有効性を検証している。

また、今回の結果から、別の方法として取りうるアプローチは、CGHデータより切断点を探す方法である。今回の検討からも、ホールゲノムシーケンスは、そのゲノムカバー率(冗長度)の利用により、より詳細なコピー数変化の解析が可能となることから、コピー数が変化

する点を可能な切断点として、可能性のあるリアレンジメント配列を人為的に作成し、この配列に対してマッピングすることにより実際の切断点を含むフラグメントを検出できる可能性が示された。今後これまでに得られたモデル細胞でのデータを用いて、この手法の有効性を検証していきたい。

さらに、もうひとつのアプローチとしては、かなり長いフラグメントを解析できるPacBioシーケンサーの利用が考えられる。残念ながら現時点ではPacBioシーケンサーのスループットはそれほど高くないため、全ゲノムをカバーできるデータを取得するのは費用的にも現実的でないが、既存のスループットの高い方法との組み合わせや、染色体分離を含めたターゲットエンリッチメントの手法を組み合わせることにより、検出は可能であると考えられる。

がん遺伝子の活性化のように融合遺伝子の特定が細胞の機能解析に重要な情報を与えることが期待されるが、細胞の品質評価という観点からは、定性的にリアレンジメントの有無を解析できれば十分であるという考え方もできる。こうした観点から、現時点で取りうる最も効率的なアプローチは、ホールゲノムシーケンス解析データを使ったゲノムワイドCGHによるコピー数変化を指標とした解析である。リアレンジメントは染色体の部分的増減を伴うことが多く、WGSデータの利用によりかなり小さい領域の変化までが検出できることがわかっている。また、コピー数変化を伴わない相互転座のような場合にも、マッピングデータの部分的減少が観察されうることから、カバー率を高めれば、ある程度の信頼性を持って検出が可能であると考えられる。

ただし、これらの検討を行う際に注意が必要なのは、ゲノム中に散在する繰り返し配列の取り扱いであり、こうした領域におけるコピー数変化の情報は、マッピング解析上のアーティファ

アクトである可能性も高い。リピート配列をマスクした解析法や、その位置情報による検証が必要であると考えられる。(ただし、相同な繰り返し配列間でのリアレンジメントも当然起こりやすいと考えられるため、その検出は今後と課題となる)

低頻度突然変異の高感度検出に関する検討では、理論的には一分子シーケンサーを用いた amplification free の方法が最適であると考えこのアプローチをとったが、シーケンスエラー率が他のシーケンサーに比べて高かったため、十分な感度が得られなかった。インサートサイズを下げることにより、シーケンスリードの冗長度を上げ、さらにコンセンサス配列に関する基準を厳しくすることによりこのエラー率を落とす必要があり、現在この観点から PacBio シーケンサーを用いた再検討を行っている。

また別のアプローチとして、スループットの高い Illumina シーケンサーを用いた PCR free 解析に、ペアエンドリードを用いたコンセンサス配列に利用するシーケンスエラーの除去法を組み合わせた方法を試みている。

これまでの検討から、自然状態での DNA ポリメラーゼによる変異誘発率は 10^{-9} から 10^{-10} であると考えられており、1細胞分裂あたり 3×10^{10} ゲノム上数個であると予想されている。このことから数 Kb 程度の遺伝子レベルでの突然変異率は 10^{-6} から 10^{-7} 程度であると予想され、今回の tk 遺伝子を用いた自然突然変異頻度とも一致する。よって、遺伝子あたりこの頻度の突然変異を検出するためには、通常の方法では少なくとも 10^7 個の細胞が必要となるが、ホールゲノムのシーケンス情報が正確に得られれば、1ホールゲノムすなわち1細胞でも 3×10^{10} bp のカバー率により、検出が可能となる。よって、クローニングした均一な細胞集団を用いることができれば、シングルコロニーアイソ

レーションをして、処理または培養前後での変異の誘発率を検討可能であるが、iPS 等のクローニングが難しい細胞に関しては、現在我々が用いているアプローチが必要となる。

通常クローニングを行わない NGS シーケンス解析においては、得られる結果は hetero なる細胞集団の平均値(majority)を反映しており、新たな生じた変異は検出できないが、リード数を増やして、エラー率を落とすことによりクローニングなしでどこまでこの変異の検出が可能とできるかが、今後の課題となる。

一方、NGS 解析データでは均一に見える細胞集団も、ホールゲノムで考えれば、分裂が起こるごとに細胞間のバリエーションが起きることとなる。通常新たな変異は、ニュートラルか劣性となるため、細胞集団全体に拡大することはないが、過去に我々の経験した hMSC 細胞の変異株の例のように、増殖性を獲得して変異を持った細胞が細胞集団全体を置き換えることも起こりうる。特に iPS 細胞のように、培養が難しい細胞においては、培養環境からの選択圧により増殖性の変異細胞が選択されやすい状況が想定される。今回 iPS 細胞をシングルコロニーに近い状態でシーケンス解析した結果から、今後細胞間の heterogeneity と変異率に関して有益な情報が得られると期待される。

(残念ながら本報告書作成までに解析が間に合わなかったため、データについては後ほど公表する予定である。)

以上の NGS 解析に関する検討に加え、プロテオームの観点から細胞の品質評価、標準化を可能とするために、得られた質量分析データを可視化して Web 上にて提供できるシステムを構築した。ユーザーがローカルに ProteoMap ソフトウェアを動作できれば、自分で取得したデータを本システムを使って Web 上で公開してリファレンスデータとすることも可能であり、今後フリーソフトウェアとしての提供とオン

ラインデータベースの構築をめざしたい。

D-4 遺伝的安定性評価ツールとしての次世代シーケンサーの性能評価

次世代シーケンサーは、従来のサンガー法とは異なり処理能力やコストの面で格段の進化を遂げ、たった一回のランで数億塩基以上も得ることが可能である。今現在も、次世代シーケンサーの性能は、すさまじい勢いで向上を続けており、今後数年の間で、個人のゲノム解読は1時間以内で終了し、数万円でできるようになると言われている。このような次世代シーケンサーの登場によって、ゲノム科学は加速度的に進展を遂げている。その一方で、リード（シーケンスの最小単位）の長さが短く、またベースコールのクオリティに問題があるなどの欠点も指摘されている。そのため、次世代シーケンサーによって得られたデータの質を評価するシステムを構築する必要がある。今回我々は、上述のような実態を把握するため、ショートリード配列の解析時に起こりうる読み間違いエラーに着目し（イルミナ社 HiSeq2500）、次世代シーケンサーの性能について検証することにした。まずエラー率の評価を厳密に行うために、ゲノム DNA のアレル頻度が正確に測定されている標準ゲノム DNA を準備し、量、質ともに次世代シーケンサー解析における条件を満たしていることを確認した。このような標準ゲノム DNA におけるライブラリーを、独立して4種類作製し、各ライブラリーについて、40Gbase, 30Gbase, 15Gbase 相当のシーケンスを行ったところ、シーケンスデータ量が多いほど、測定値のバラツキは低減されることが判明した。さらに、解析の精度を評価する目的で、すべてのライブラリーから得られたシーケンスデータを統合し、それらシーケンスデータのサンプリングサイズの幅を 160Gbase, 120Gbase, 80Gbase, 40Gbase に設定し、解析精度の閾値

を検討したところ、複数回のシーケンスデータを取得することで、マッピング量とは無関係にバラツキは低減されることが判明した。つまり、Fig. 21 で見られた各塩基間の CV 値のバラツキの程度は、Fig. 22 では低減していることから、ライブラリー間のバラツキが大きく影響していることが考えられた。即ち、実際の検体（診断用組織や細胞加工製品など）について得られたシーケンスデータを解析する場合には、複数回の独立した解析を行うことも重要であると考えられる。以上、次世代シーケンサーのエラー率評価のために標準ゲノム DNA を用いることは、標準ゲノム DNA をポジティブコントロールとして精度管理をすることが可能となるだけでなく、検出限界（Limit of Detection (LOD)) の検証、複数のプラットフォーム間での比較評価に役立つものと思われる。

D-5 遺伝的安定性評価リファレンスとしての日本人ゲノムの *de novo* 配列決定

3種類のペアエンド・ライブラリーともう3種類のメイトペア・ライブラリーを作製し、HiSeqを用いて片側 150 nt のペアエンドデータを取得した。これらに対し、最新の大規模メモリ搭載型計算機で *de novo* アセンブリを行い、N50 が 45 kb のデータを取得することができた。総 contig 長は 2,255 Mb に達し、反復配列を除いたゲノム領域をおおよそカバーすることができた。本研究ではまず mtDNA の完全長を決定し、*de novo* アセンブリを行う前に6種類の各ライブラリーのインサート長を推定するという工夫を行った。実際に見積もったインサート長は、どれもライブラリー作製時に狙ったサイズよりも小さく、この正確さによりアセンブリの精度も上がったと考えている。その一方でまだ多くの gap が残っており、PacBio から得られたリードを、塩基配列そのものではなく、scaffold の並び換え目的で使用することにより全ゲノ

ム配列決定に向けて作業を進めて行く予定である。

本研究グループが高解像度 SNP アレイを用いて日本人固有のコピー数多型を多数報告したが、欠失はともかく、コピー数獲得の場合、追加されたアレルが染色体のどこにあるのかを調べることは容易ではなかった。今回1人のデータではあるものの、そのいくつかの場所を特定するばかりでなく、ブレイクポイント配列を塩基レベルで正確に決定することができた。これまでアジア人に欠失している遺伝子は多く知られているが、リファレンス配列に欠失している未知の配列も、今回得られた scaffold の中から探し出すことも可能であると考えている。これら構造多型は、可逆的な SNP に対し、不可逆な変異であるため、これまで SNP から描かれていた曖昧な系統樹を修正することができ、人類学への貢献も期待される成果である。今回は1サンプルのアセンブリ、5サンプルのマッピングであったが、引き続きサンプル収集を行い、アノテーションも充実させ、日本人ゲノム標準リファレンス配列を、アレル頻度情報とともに完成させる予定である。

D-6 分化プロペンシティを指標とした細胞特性解析法の開発

本研究は、再生医療等製品の原材料としての iPS 細胞の適格性の評価のために、iPS 細胞株の分化プロペンシティ予測マーカーを同定することを目的に研究を行っている。昨年度の報告において、ヒト iPS 細胞株 10 種類について、外胚葉、中胚葉、および内胚葉への分化のしやすさの順位と、未分化状態での mRNA および miRNA 発現量の順位を用いて、スピアマンの順位相関解析を行い、相関のある mRNA および miRNA を分化予測マーカー候補として同定した。本年度は、これらの分化予測マーカー候補である分化プロペンシティと相関のある

mRNA と miRNA を、さらに遺伝子ネットワーク・パスウェイ解析により絞り込んだ。

IPA を用いた miRNA-mRNA ペアリング解析によって、Tables 22-24 に示す 10 種類の miRNA と、43 種類の mRNA が選別された。このうち、*BCLAF1*, *STXBP5L*, *APCDD1L*, *XPNPEP3*, *EFCAB2*, *RAB3B* はそれぞれ二つの miRNA のターゲットとして検出されており、外胚葉、および内胚葉分化との相関関係について信頼性が特に高いと考えられた。さらに本研究では、miRNA に注目し各胚葉への分化への影響を評価することを試みた。mRNA ではなく miRNA に注目した理由は、ペアリング解析で抽出された miRNA の中には複数の mRNA の発現に関わっていることが予想されたため、一つの因子で分化に対してより大きな影響が見られるのではないかと推測したためである。ペアリング解析で抽出された miRNA について、これまでに得られている知見を以下に概説する。① miR-371~3 は ES/iPS 細胞において神経への分化傾向を予測するという報告がある

(Kim et al. *Cell Stem Cell*. 2011;8: 695-706.). ②

miR-371-5p は細胞周期の G1/S 移行を加速させることによって細胞増殖に関与するという報告がある (Liu et al. *Cancer Lett*. 2013;335:351-360.). ③ miR-367 は細胞のリプログラミングに関与するという報告がある (Kuo et al. *Biochem Biophys Res Commun*. 2012;417:11-16.). その他全ての miRNA については機能未知である。

今回ヒト iPS 細胞である 253G1 にレンチウイルスを感染させた結果、miR-373, miR-371-3p, miR-524-5p, miR-4505, および miR-520g の 5 種類の miRNA 過剰発現細胞を作製することができた (Figs. 24, 27 and 30). miRNA の増加量

は、control での発現量を 1 とすると、少ない miRNA (miR-4505) で約 2 倍、多い miRNA (miR-520g) で約 9 倍と差がみられた。しかし、それぞれの miRNA を iPS 細胞株 10 種類のシグナル値 (Figs. 22, 26 and 29) でみると、miR-373 で 10 細胞株中 3 番目、miR-371-3p で 10 細胞株中 4 番目、miR-524-5p と miR-4505 で 10 細胞株中 1 番目に発現が多い細胞株となる計算になった。一方、miR-4739, miR-4521, miR-3714, および miR-367 の過剰発現細胞は作製できなかった。その原因として、これら 4 つの miRNA は細胞内においてプロセシングされにくい miRNA である可能性が考えられた。成熟型 miRNA の発現レベルは細胞内での miRNA のプロセシングに依存することが知られている。本研究で用いた BiOSETTIA 社の miRNA 発現レンチウイルス粒子には、miRBase に登録されているネイティブな前駆体配列がクローニングされているため、細胞内で Drosha と Dicer によってプロセシングされやすい miRNA もあれば、プロセシングされにくい miRNA もあり、実際に miRNA 前駆体から成熟型への変換率は 10% から 90% とバラツキがあることが報告されている。

miR-373 および miR-371-3p の過剰発現細胞における神経細胞への分化誘導では、神経マーカー遺伝子の発現で比較したときに control との差は見られなかった (Fig. 25)。また、miR-524-5p の過剰発現細胞における心筋細胞への分化誘導においても、control と比較してマーカー遺伝子の発現量に差は見られなかった (Fig. 28)。また miR-4505 および miR-371-3p の過剰発現細胞における胚葉体形成による分化においても、内胚葉マーカーの発現量に差は見られなかった (Fig. 31)。

これらの結果に関して、いくつかの理由が考えられた。一つ目の理由として考えられたのは、miRNA の発現量不足である。今回作製した

miRNA 過剰発現細胞における miRNA 発現量は、control に対し miR-373 で約 3 倍、miR-371-3p で約 7 倍、miR-524-5p で約 5 倍であったが、実際に分化に影響を及ぼすには、さらに多くの miRNA 量が必要である可能性が考えられた。また、一種類の miRNA のみが細胞の分化状態に大きな変化を与えているとは限らない。Kimらの論文においても、miR-371~3 が神経分化に影響する唯一の因子ではなく、他にこの反応に寄与している因子の存在を示唆している。そのため、単独種の miRNA のほか、複数種の miRNA を過剰発現させた細胞での検討を行うことも必要であると考えられた。二つ目の理由として、分化誘導後にマーカー遺伝子発現量を比較した日数が遅すぎたために、発現量に差が見られないことが考えられた。心筋細胞への分化誘導の予備実験において、細胞に発現するマーカー遺伝子を継時的に定量したところ、*TNNT2*, *NKX2.5*, および *MYH6* の発現量は分化誘導後 5 日目から 10 日目にかけて大きく増加していることが確認された。この結果から、miRNA が分化誘導の初期の段階に作用していた可能性も考えられる。これを確認するためには、タイムコースをとって分化誘導を行う必要があると思われた。三つ目の理由として考えられたのは、miRNA の発現が原因で目的細胞に分化しにくいのではなく、細胞へ分化しにくい細胞株では結果的に miRNA が多く発現しているという可能性である。細胞の分化プロペンシティと発現量との間で相関のある miRNA であるものの、実際の分化に機能的にはたらいっていない可能性も考えられる。

今回の実験では、残念ながら同定した miRNA が分化に機能的に関与していることを証明することができなかった。しかしながら、本研究課題で報告した目的細胞の製造に適切な iPS 細胞マーカーをスクリーニングする手法は、未分化状態のヒト iPS 細胞株における三

胚葉系細胞への分化プロペンシティ予測マーカーの探索と、これらの予測マーカーのサロゲートマーカーとしての信頼性を確認する上で有用であると思われる。目的細胞も様々で、その分化方法も様々である。iPS 細胞加工製品において、我々が報告した方法は、最終製品に適切な株の評価に貢献することが期待される。

E. 結論

E-1 新規免疫不全動物を用いた造腫瘍性試験法の開発

異種細胞生着能および被毛がないことによる実験操作性から判断した場合、BRG-nu マウスが造腫瘍性試験用動物として NOG マウスや BRG マウスよりも優位であった。

E-2 幹細胞の *in vitro* 培養工程における遺伝子発現の動態解析による品質評価技術の開発

LINE-1s の転移を抑える細胞内因子として知られている A3B には、欠失多型が存在し、日本人にその割合が多いと報告されている。A3B を発現しない日本人由来 hMSCs は LINE-1s の転移によってゲノムの安定性が損なわれる可能性が考えられたので、A3B 遺伝子型と LINE-1s の発現量について解析を行った。次世代シーケンサーを用いて A3B 野生型ホモ (Ins/Ins) と A3B 欠失型ホモ (Del/Del) の RNA 配列を網羅的に解析し比較したところ、Ins/Ins と Del/Del で転移活性の残った LINE-1s の発現量には大きな違いが見られず、日本人に多いとされる A3B 欠失により hMSC における LINE-1s の転移によるゲノムの安定性を損なう危険性は示されなかった。

さらに、iPS 細胞や hMSC など幹細胞における LINE-1s の発現は確認できたが、分化した正常組織ではその発現がほとんど認められないことから、hMSC における LINE-1s の発現に及ぼす分化の影響について検討したところ、

hMSC を脂肪分化させることにより、LINE-1s mRNA の発現量が低下することを確認した。

このことから、LINE-1s の発現と分化能には関連が見られ、LINE-1s の発現が hMSCs の分化能を示すマーカーの一つとになり得る可能性が示唆された。

E-3 次世代シーケンサーを用いた細胞の遺伝的安定性評価指標の開発

以上の検討を基に、現段階での NGS によるシーケンスデータの利用に関しての提言を以下にまとめる。

1. CGH データを利用したゲノムワイドな品質評価
2. 細胞起源同一性の確認
3. がん遺伝子を含めた全遺伝子の変異確認
4. 培養過程における増殖性変異獲得のチェック

1. に関しては WGS データの取得が必要であるが、2-4 に関してはエクソーム解析のみでも検討が可能であり、今後目的に応じて品質評価への利用が期待される。4. に関してはあまりこれまでに意識されていなかったが、細胞の heterogeneity の維持という観点からも今後さらに検討が必要な課題である。

細胞のプロテオームデータに関して、PeroteomeMap ソフトウェアを利用して Web 上にて情報提供できるシステムを構築した。今後は、ユーザーからのプロテオームデータを受け入れることにより、細胞のプロテオームデータに関するリファレンスデータベースの構築が可能となった。

E-4 遺伝的安定性評価ツールとしての次世代シーケンサーの性能評価

本研究では、次世代シーケンサーの精度を理解するため、ショートリード配列の解析時に起こりうる読み間違いエラーに着目し (イルミナ

社HiSeq2500) , 次世代シーケンサーの性能について検証した。まず, 我々は, 参照配列に対する変異塩基の頻度が既に測定されているゲノム標品を準備し, この標準ゲノムDNAを用いてシーケンスを行った。実際にシーケンスされた塩基種類の頻度を測定したところ, シーケンスデータ量が多いほど, 測定値のバラツキは低減されることが判明した。

一般的に, エキソーム解析における読み取り総塩基数は, 通常 5Gbase 程度であるが, 今回の実験のように, 15~40Gbase のシーケンスデータ量を取得した場合でも, 各ライブラリー間でのバラツキが散見された。その一方で, 複数回の独立した解析を行うことで, シーケンスデータのバラツキは抑えられることが確認された。つまり, シーケンスの精度を高めるためには, カバレッジを深くすることの他に, 独立した複数の解析を実施することも重要であることが示唆された。また, 今回の解析のように, 通常の 8 倍程度のカバレッジでシーケンスを行ったとしても, 測定する塩基の位置においては, バラツキに差が現れることも観察された。このことは, 現在のシーケンサー自体の性能の限界であると考えられるため, 今後は, コスト面や解析速度など, 目的に応じてシーケンサーの精度を理解し, 解析方法を使い分けることが重要であると思われる。

E-5 遺伝的安定性評価リファレンスとしての日本人ゲノムの *de novo* 配列決定

現在のリファレンス配列は, アジア人の SNP 情報が反映されているものの, その骨格は主としてヨーロッパおよびアフリカ由来のヒトゲノムを用いて組み立てられた配列である。日本人の由来については諸説あるが, 日本列島への最後の大規模な人類の移動は約 2000 年前であると考えられており, それ以前は多くのアジア系民族の血が混じり合い, それ以後は

比較的孤立した状態を保ってきたというユニークな遺伝的特徴を持つ。それゆえに日本人ゲノムには挿入, 欠失, 重複, 反復配列伸張や短縮, トランスポゾンの挿入や欠失, 転座, 遺伝子変換など共通し, かつ他の人種から区別される多くの構造多型があると考えても不思議ではない。エクソーム解析をはじめとする次世代シーケンサーを用いたリシーケンシングにおいては, リファレンスと大きく異なるこれらの配列データは扱うことができないため無視されている。日本人を対象とした医療を考える場合, 治療のための細胞品質評価も含め, *de novo* アセンブリと第 3 世代シーケンサーを利用し, 日本人ゲノムのリファレンス配列を用意することが有効であると考えられ, 本研究においてその最初のデータを産出することができた。このデータが細胞を用いた再生医療, パーソナル医療を始めとする将来の医療に貢献することは間違いない。

E-4 分化プロペンシティを指標とした細胞特性解析法の開発

多能性幹細胞株の目的細胞への分化を適切に予測することを目標とし, 分化プロペンシティを指標とした細胞特性解析法の開発を行った。分化プロペンシティ予測マーカー候補として絞り込まれた miRNA のヒト iPS 細胞安定発現株を樹立し, 分化実験を行い, これらの miRNA がヒト iPS 細胞の分化において機能的な役割を果たしているのかを検討した。今回報告した手法を利用することにより, 最終製品に適したヒト iPS 細胞株の選択が可能になると思われる。

「幹細胞の *in vitro* 培養工程における遺伝子発現の動態解析による品質評価技術の開発」に関する参考文献

1) Riggi N, Suvà ML, Suvà D, Cironi L, Provero P.

Tercier S, Joseph JM, Stehle JC, Baumer K, Kindler V, Stamenkovic I: EWS-FLI-1 expression triggers a Ewing's sarcoma initiation program in primary human mesenchymal stem cells. *Cancer Res* 2008, 68:2176-2185.

- 2) A. Macia, M. Munoz-Lopez, J.L. Cortes, R.K. Hastings, S. Morell, G. Lucena-Aguilar, J.A. Marchal, R.M. Badge, J.L. Garcia-Perez, Epigenetic control of retrotransposon expression in human embryonic stem cells, *Mol Cell Biol* 31 (2011) 300-316.
- 3) S. Wissing, M. Munoz-Lopez, A. Macia, Z. Yang, M. Montano, W. Collins, J.L. Garcia-Perez, J.V. Moran, W.C. Greene, Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility, *Hum Mol Genet* 21 (2012) 208-218.
- 4) S. Wissing, M. Montano, J.L. Garcia-Perez, J.V. Moran, W.C. Greene, Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells, *J Biol Chem* 286 (2011) 36427-36437.
- 5) Y.L. Chiu, W.C. Greene, The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements, *Annu Rev Immunol* 26 (2008) 317-353.
- 6) K. Sato, T. Izumi, N. Misawa, T. Kobayashi, Y. Yamashita, M. Ohmichi, M. Ito, A. Takaori-Kondo, Y. Koyanagi, Remarkable lethal G-to-A mutations in vif-proficient HIV-1 provirus by individual APOBEC3 proteins in humanized mice, *J Virol* 84 (2010) 9546-9556.
- 7) J.M. Kidd, T.L. Newman, E. Tuzun, R. Kaul, E.E. Eichler, Population stratification of a common APOBEC gene deletion polymorphism, *PLoS Genet* 3 (2007) e63.

F. 健康危険情報

なし

G. 研究発表

G-1 論文発表

1. Kono K., Takada N., Yasuda S., Sawada R., Niimi S., Matsuyama A., Sato Y. : Characterization of the cell growth analysis for detection of immortal cellular impurities in human mesenchymal stem cells. *Biologicals*, 43, 146-149 (2015)
2. Kusakawa S., Machida K., Yasuda S., Takada N., Kuroda T., Sawada R., Okura H., Tsutsumi H., Kawamata S., Sato Y. : Characterization of in vivo tumorigenicity tests using severe immunodeficient NOD/Shi-scid IL2R α ^{null} mice for detection of tumorigenic cellular impurities in human cell-processed therapeutic products. *Regenerative Therapy*, 1, 30-37 (2015)
3. 澤田留美 「再生医療等製品開発における動物実験—指針及び評価指標について—」 *オベリスク*, 20(1), 25-31 (2015)
4. 澤田留美 「再生医療等製品とバイオマテリアル, そして評価指標」 *バイオマテリアル—生体材料—*, 33(1), 7-8 (2015)
5. Sasaki H., Takeuchi I., Okada M., Sawada R., Kanie K., Kiyota Y., Honda H., and Kato R.: Label-free morphology-based prediction of multiple differentiation potentials of human mesenchymal stem cells for early evaluation of intact cells. *PLOS ONE*, 9(4), e93952 (2014).
6. Kono K., Niimi S., and Sawada R. : Cyclin

- D2 promotes the proliferation of human mesenchymal stem cells. *J. Bone Marrow Res.*, 2: 136. 1000136 (2013).
7. 鈴木孝昌 コンパニオン診断薬の現状と課題 「最先端バイオマーカーを用いた診断薬/診断装置開発と薬事対応」 p271-275 (技術情報協会) 2015
 8. Fukuda A, Tomikawa J, Miura T, Hata K, Nakabayashi K, Egan K, Akutsu H, Umezawa A. The role of maternal-specific H3K9me3 modification in establishing imprinted X-chromosome inactivation and embryogenesis in mice. *Nat Commun.* 2014, 5: 5464.
 9. Fukawatase Y, Toyoda M, Okamura K, Nakamura K, Nakabayashi K, Takada S, Yamazaki-Inoue M, Masuda A, Nasu M, Hata K, Hanaoka K, Higuchi A, Takubo K, Umezawa A. Ataxia telangiectasia derived iPS cells show preserved x-ray sensitivity and decreased chromosomal instability. *Sci. Rep.* 4, 5421 (2014)
 10. Migita O, Maehara K, Kamura H, Miyakoshi K, Tanaka M, Morokuma S, Fukushima K, Shimamoto T, Saito S, Sago H, Nishihama K, Abe K, Nakabayashi K, Umezawa A, Okamura K, Hata K. Compilation of copy number variants identified in phenotypically normal and parous Japanese women. *J. Hum. Genet.* 59, 326-331 (2014)
 11. Tano K, Yasuda S, Kuroda T, Saito H, Umezawa A, Sato Y. A novel *in vitro* method for detecting undifferentiated human pluripotent stem cells as impurities in cell therapy products using a highly efficient culture system. *PLoS One.* 2014;9:e110496.
 12. 佐藤陽治 再生医療と薬学 ファルマシア 2014;50:1213-5.
 13. 佐藤陽治 ヒト多能性幹細胞加工製品に残存する未分化多能性幹細胞の高感度検出法の開発 *再生医療* 2014;13:432-5.
 14. 村岡ひとみ, 佐藤陽治 再生医療・細胞治療の規制動向とレギュラトリーサイエンス DDS. 2014;29:207-16.
 15. 中島啓行, 佐藤陽治 薬事法改正と再生医療等安全性確保法を踏まえた再生医療／細胞治療の開発 ファームステージ 2014;10:1-5.
 16. 三浦巧, 佐藤陽治 再生医療・細胞治療に使用する細胞加工物の品質・安全性評価の原則と造腫瘍性の考え方 谷本学校毒性質問箱 2014;16:1-10.
 17. 佐藤陽治 再生医療／細胞治療における細胞培養に関する規制 『再生医療の細胞培養技術開発と応用展開』（監修：紀ノ岡正博）株式会社シーエムシー出版，東京（2014），pp. 27-36.
 18. 草川森士, 佐藤陽治 再生医療製品の造腫瘍性評価 *最新医学* 2014;69(3)増刊号:745-765.
 19. 佐藤大作, 佐藤陽治 規制関連 『再生医療用語集』（印刷中）

20. 村岡ひとみ, 佐藤陽治 再生医療・細胞治療の臨床研究から実用化までの道のり *Geriatric Medicine (老年医学)* 2014;52(3):237-239.
 21. 佐藤陽治 ヒト iPS 細胞由来移植細胞中に混入する造腫瘍性細胞/未分化細胞の in vitro 検出法 *Cytometry Research* 2014;24(1): 7-11.
 22. 安田智, 佐藤陽治 再生医療製品の品質関連規制と対応の留意点 『動物細胞培養の手法と細胞死・増殖不良・細胞変異を防止する技術』(編集:技術情報協会) 技術情報協会, 東京 (2014), pp. 517-22.
5. 加藤玲子, 齋島由二, 福井千恵, 比留間瞳, 澤田留美, 宮島敦子, 新見伸吾「ヒト単球系細胞の蛋白質発現挙動に基づく医用材料の血液適合性評価マーカの探索」第 36 回日本バイオマテリアル学会大会 (2014.11)
 6. 河野 健, 新見伸吾, 澤田留美「間葉系幹細胞における細胞分化と LINE-1 の発現について」第 14 回日本再生医療学会総会 (2015.3)
 7. 高田のぞみ, 河野健, 安田智, 澤田留美, 新見伸吾, 松山晃文, 佐藤陽治「細胞増殖特性を利用した不死化細胞検出試験法の性能評価」第 14 回日本再生医療学会総会 (2015.3)
 8. 佐々木寛人, 高橋厚妃, 蟹江慧, 澤田留美, 本多裕之, 清田泰次郎, 加藤竜司「骨髄由来および脂肪組織由来間葉系幹細胞の継代培養における品質劣化モニタリング」第 14 回日本再生医療学会総会 (2015.3)
 9. Suresh T., Maekawa K., Saito Y., Sato Y., Suzuki T. Individual variations in the human urinary proteome in relation to rat. The 3rd International Conference on Personalized Medicine (2014.6) (Prague)
 10. スレッシュ ティルパッティ, 斎藤嘉朗, 本間正充, 佐藤陽治, 鈴木孝昌 変異原暴露モニタリング手法としてのタンパクアダクトミクス日本環境変異原学会第 43 回大会 (2014. 12) (東京)

G-2 学会発表

1. 町田一彦, 草川森士, 澤田留美, 安田智, 佐藤陽治, 堤秀樹 NOGヘアレスマウスにおける免疫不全能の定量的解析 日本実験動物科学技術さっぽろ2014 札幌(平成26年5月15-17日)
2. Kono K., Niimi S., Sawada R.; Analysis of Line1 expression in human mesenchymal stem cells, 12th Annual Meeting of International Society for Stem Cell Research (2014.6)
3. Sasaki H., Okada N., Kanie K., Kiyota Y., Honda H., Sawada R., Kato R.; Image-based profiling of mesenchymal stem cells using non-label images, TERMIS-EU 2014 (2014.6)
4. 澤田留美, 河野 健, 比留間瞳, 加藤玲子, 新見伸吾「生体親和性高分子材料によるヒト単球細胞の機能の制御について—遺伝子発現の網羅的解析による検討」第 36 回日本バイオマテリアル学会大会 (2014.11)

11. Suzuki T., Suresh T. Protein adductome analysis for the human exposure monitoring to mutagens. The 4th Asian Conference on Environmental Mutagens (2014.12) (Kolkata)
12. 鈴木孝昌 医薬品開発においてヒト内在性物質を測定する際の定量分析法に関する留意点(案)の概要:規制の重要性と今後の課題 第6回JBFシンポジウム (2015. 2) (東京)
13. 阿久津 英憲, 菅原 亨, 三浦 巧, 梅澤 明弘. mir-302 マイクロRNAファミリーによるヒト多能性幹細胞の中・内胚葉初期分化制御. 第14回再生医療学会総会, 横浜 (2015年3月19~21日)
14. Takumi Miura, Tohru Sugawara, Atsushi Fukuda, Ryo Tamoto, Akihiro Umezawa, Hidenori Akutsu. Generation of Committed Neural Progenitors from Human Fibroblasts by Defined Factors. The 12th Annual Meeting International Society for Stem Cell Research, Vancouver, Canada (2014年6月18-21日)
15. 草川森士, 安田智, 黒田拓也, 川真田伸, 佐藤陽治 軟寒天コロニー形成試験を応用した再生医療製品に混在する悪性形質転換細胞の高感度検出法 第14回再生医療学会総会, 横浜 (2015年3月19-21日)
16. 田埜慶子, 安田智, 黒田拓也, 梅澤明弘, 佐藤陽治 ヒト多能性幹細胞由来再生医療製品中に残存する未分化細胞をダイレクトに検出する方法の開発 第14回再生医療学会総会, 横浜 (2015年3月19-21日)
17. Sato Y. Tumorigenicity Tests for the Quality and Safety of Cell-Based Therapeutic Products. IABS Workshop, kyoto (2015年2月18-19日)
18. Yasuda S. The New Japanese Regulatory Framework for Regenerative Medicine & Cell Therapy. World Stem Cell Summit 14, San Antonio (2014年12月3-5日)
19. 佐藤陽治 ヒト/動物細胞加工製品の品質確保に関する基本的考え方 レギュラトリーサイエンス学会シンポジウム~再生医療等製品の承認審査と再生医療新法~, 東京 (2014年11月25日)
20. 佐藤陽治 細胞技術の許認可の実情-再生医療に関する日本の新しい規制の枠組み- 第36回日本バイオマテリアル学会大会, 東京 (2014年11月18日)
21. Kusakawa S, Yasuda S., Kuroda T, Kawamata S, Sato Y. A new soft agar colony formation assay based on high-content imaging for sensitive detection of tumorigenic cellular impurities in human cell-processed therapeutic products. Global Controls in Stem Cells, Singapore (2014年11月5-7日)
22. 佐藤陽治 ヒト由来移植細胞に混入する多能性幹細胞・造腫瘍性細胞の検出法の性能評価 第87回日本生化学大会, 京都 (2014年10月15-18日)
23. Kuroda T, Tachi S, Yasuda S., Kusakawa S, Sato Y. Profiling of Human Induced Pluripotent Stem Cell Lines for Predicting the