

課題番号 : 25指102

研究課題名 : 「第三世代シーケンサーPacBio RSを活用した糖尿病・代謝疾患の病態解析」

主任研究者名 : 安田和基

分担研究者名 : 南茂隆生

キーワード : ロングリード、ハプロタイプ決定、選択的スプライシング、DNA 修飾

研究成果 : 本研究では、第三世代シーケンサーである PacBioRS を用いて、技術開発とともに糖尿病関連疾患の病態解明を行う。このシステムは、A. 数 kb にわたる長いリード配列、B. 高い GC 含量の領域も解析可能、C. Kinetics 情報（特に波形間の inter-pulse duration : IPD）の取得により塩基修飾情報も得られる可能性、など、これまでの次世代シーケンサー（NGS）にない特徴がある。そこで、A によりハプロタイプの決定や allelic expression の解析、ゲノム構造異常の解析、B によりプロモーター領域や繰り返し配列を含めたゲノム解析、C により細胞分化や糖尿病の病態や合併症に関する DNA 修飾検出、などを試みる。本技術はまだ発展途上の部分があり、ゲノムサイズの小さい病原微生物などでは有用性が確立されつつあるものの、ゲノムサイズの大きい哺乳類ゲノムや RNA についての報告はきわめて限定的である。本研究は既存の解析方法と比較する「feasibility study」と、新たなアプリケーションの開発を行う「potential の開拓」という、2つの方向性をもつ。

既に平成 25 年度に本システムの性能を検証するために、ラムダファージ DNA を 10kb のサイズで DNA を断片化してテンプレートを作成して、シーケンスを行ったところ、「リード長」は 10～15kbp まで分布しており、両端の adaptor には含まれた「サブリード」長は 2～5kbp 程度が多く、第三世代 NGS としての PacBioRS の「ロングリード」の特性および有用性を確認することができた。

本機器の有効性と限界について、平成 25 年度に引き続いて共同研究者であるトミーデジタルバイオロジーズ（以下 TDB）社（実験部門、解析部門）と情報交換を行った。その結果、本テクノロジーの特徴的課題である①DNA 領域濃縮法、②定量性、③真核生物の DNA 修飾（特にメチル化）の IPD パターン、のいずれも進捗は大きくなかったが、①については、NimbleGen のテクノロジーを用いることで、PCR による増幅が 7 サイクル含まれ DNA メチル化の情報は失われるものの PacBio に有用な 4-5kb フラグメントの DNA 濃縮法が報告された（BMC Genomics 2015）。そこで諸々の検討の結果、この方法を採用することとし、ハプロタイプなど、解析すべき DNA 修飾候補領域を選別し、capture 用のプローブ作成を行っている。

機能性をもつ GWAS の SNP 領域のヒストン修飾の解析について、白人由来初代培養ヒト肝細胞を購入して培養した後、ゲノム DNA、Total RNA を抽出すると共に、活動性のあるプロモータ、エンハンサーのマーカーとなる H3K27ac について、第 2 世代シーケンサーであるイルミナ社 HiSeq200 を用いて ChIP-Seq を行った。その結果、5 万ヶ所の cis 調節ドメインが同定され、他疾患も含めた GWAS SNP の約 2200 が、その H3K27ac 陽性領域に存在した。今後、PacBio を用いたハプロタイプの決定および、ヘテロ接合体における allelic expression、あるいは allelic imbalance の同定を考えている。

Subject No. : 25D02

Title : The application of the third generation sequencer, PacBioRS, to the researches on diabetes mellitus and metabolic diseases.

Researchers : Yasuda K, Nammo T, and others.

Key word : the third generation sequencer, long read, haplotype phasing, alternative splicing, DNA modification

Abstract : PacBioRS, a new platform called “the third generation sequencer”, has several unique features over other next-generation sequencers(NGS), including extra-long reads(over several kb), robust base calls for GC-rich or repetitive sequences, and possible applications for the analysis of DNA modification. However, this technology is still on the progress, and there have been only a limited number of reports on its application to mammalian genomes and RNA. In this project, we are planning to utilize this unique system for the diabetes research. We perform ‘feasibility studies’ for assessing the performance of PacBioRS compared with existing technologies using known sequences, and ‘innovative studies’ trying new applications of this technology to the genomic and epigenomic studies of diabetes mellitus.

We already assessed the performance of PacBioRS by sequencing lambda phage DNAs, and found that the read length was extended up to 10~15kb, and most of the subread length was 2~5kb, thus the striking ‘long read’ feature was confirmed.

We continued discussing the advantages and disadvantages of PacBioRS, with scientists from PacBio and from Tomy Digital Biology Ltd. We focused on DNA enrichment methods, quantitative analyses, and detection of methylated DNA. We concluded that size selection of fragmented genomic DNA using Blue Pippin followed by DNA capture with NimbleGen probes is the most promising method of DNA enrichment for PacBioRS, and selected target genomic regions.

In order to obtain regulatory genomic regions, we performed a ChIP-seq analysis for H3K27ac, a hallmark of active promoters, using human primary hepatocytes derived from a Caucasian donor. We identified more than 50,000 active marks, and found that disease-associated SNPs previously reported by various GWASs were significantly enriched in these regions. We are planning to examine heplotype phasing and to detect allelic expression or imbalance of these regions in combination with analyses by PacBioRS.

第三世代NGS: PacBio RSの特性とサブテーマの吟味

(SMRT bell™ sequencing)

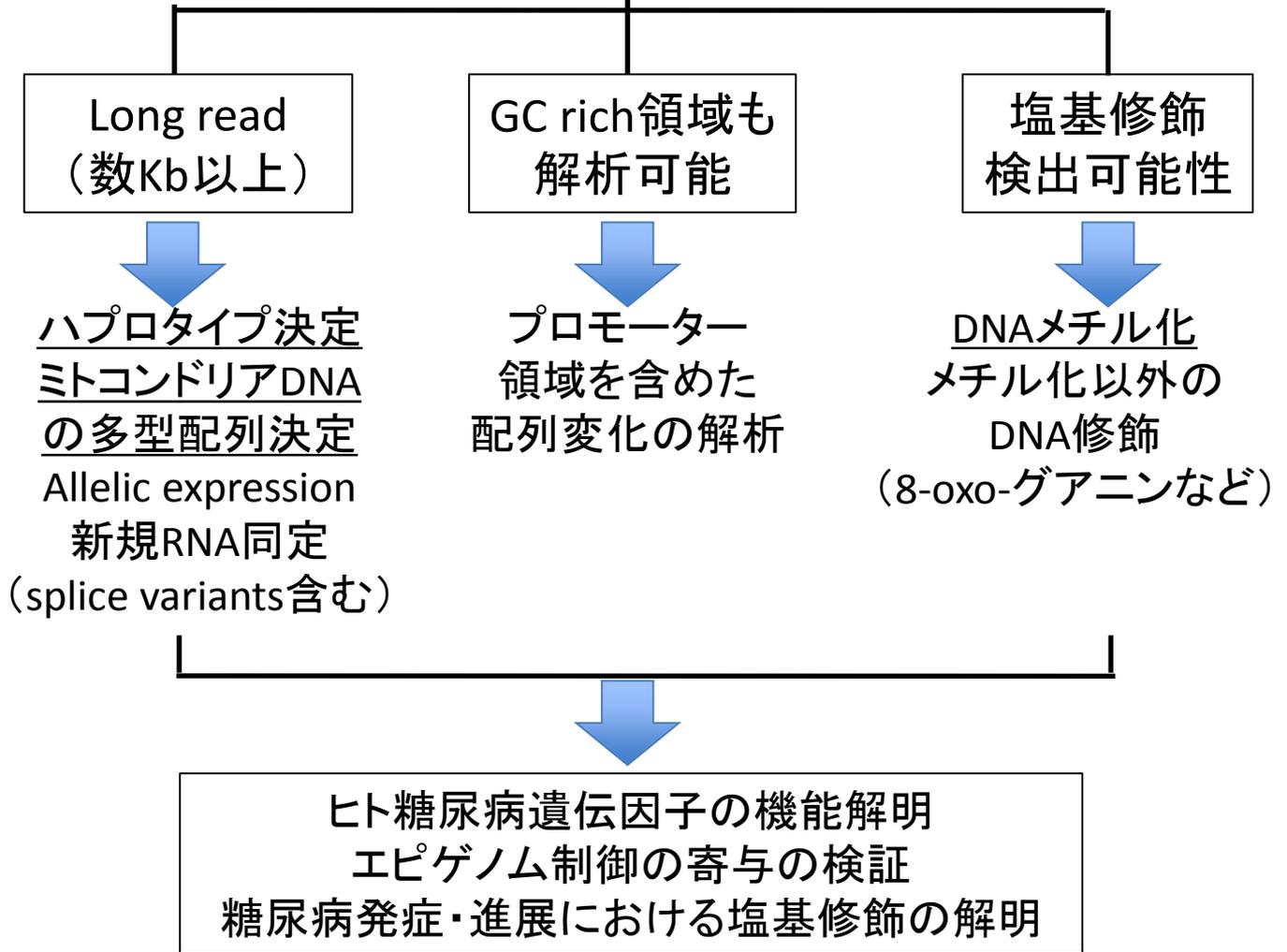
第二世代NGS
にない特徴

↓

ヒトを中心とした
高等動物ゲノム
への応用
(挑戦的分野)

↓

糖尿病関連病態
における
未解決課題への
ソリューション



下線部分: 本システムの特徴を踏まえて、改めて優先課題と位置づけたもの

PacBioRSの特性の検証

(ラムダファージDNA、インサート10kbpで作成、CLRモード、「120」分でデータ取得)

図1:リード長分布

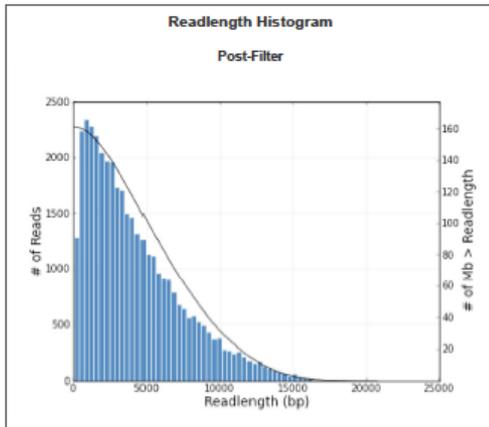


図2:サブリード長分布

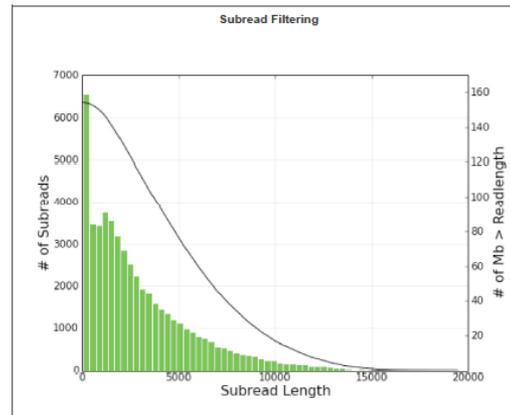
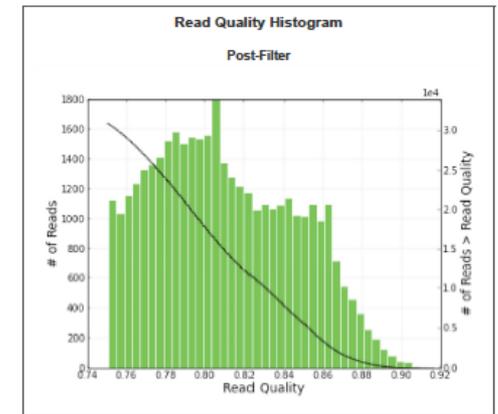


図3: Read Quality (RQ)



PacBioRSに適したDNA濃縮法の検証

- PCR増幅: 濃縮の効率が不良で、かつ多サイクルの増幅が必要。
- オリゴキャプチャー: 短いフラグメントもかなりcaptureしてしまい、PacBio用に十分サポートされていない。
- サイズ・セクション→PCR増幅→NimbleGenテクノロジーによる増幅 (BMC Genomics 2015): DNAメチル化の情報は失われるが、十分な長さのリードが得られることがわかり、ターゲット領域の吟味を施行中。

課題番号 : 25指102

研究課題名 : 「第三世代シーケンサーPacBio RS を活用した糖尿病・代謝疾患の病態解析」

分担研究課題名 : 「研究の総括、及び PacBio RS の特性を活かしたアプリケーションの探索と 2 型糖尿病研究への応用」

分担研究者名 : 安田和基 (協力研究者 ; 西村渉、宇田川陽秀、舟橋伸昭、川口美穂、ほか)

キーワード : ロングリード、DNA 濃縮、DNA 修飾

研究成果 :

我々は既にラムダファージ DNA を用いた予備実験により、「リード長」が 10~15kbp まで分布し、両端の adaptor には含まれた「サブリード」も、2~5kbp 程度が多いなど第三世代シーケンサーPacBioRS の「ロングリード」の特性を検証している。また一方、サブリードの accuracy は平均 81.6%であり、fidelity の問題から十分な depth での解読が必要と考えた。

共同研究者であるトミーデジタルバイオロジーズ (以下 TDB) 社の本機器担当者 (実験部門、解析部門) と会合を重ねた結果、平成 26 年度は、技術的課題のうち特に①DNA 領域の濃縮法、②定量的解析、③特に真核生物における DNA メチル化の検出、の 3 点について、情報交換を行った。特に①について、目的ゲノム領域の濃縮に対して、既存の試薬のプロトコルをベースに、本テクノロジーに適したカスタマイズが可能かどうかの検討を行った。しかし、汎用されている市販の試薬では、PCR、ハイブリダイゼーションのいずれをベースとした方法でも、長いフラグメントを効率的に濃縮するプロトコルはサポートされていなかった。

最近になって、ゲノム DNA を断片化したのち、Blue Pippin を用いたサイズセレクション (5-9kb) を行った上で、アダプター付加後 PCR で 7 サイクル増幅してから NimbleGen の濃縮プロトコルを用いて capture すると、PacBioRS に適した 4-5kb 平均のライブラリーが作成できる、という報告がなされた (BMC Genomics 16:214, 2015)。詳細な検討の後、濃縮すべき領域として、単一遺伝子病タイプの原因遺伝子領域、GWAS で 2 型糖尿病に関連した SNP 関連の領域などを中心にプローブ作成領域の選択を行っている。

課題番号 : 25指102

研究課題名 : 2型糖尿病感受性座位におけるハプロタイプと組織特異的遺伝子発現変化の意義の検討

主任研究者名 : 安田 和基

分担研究者名 : 南茂 隆生

キーワード : 2型糖尿病、第三世代シーケンサー、ターゲットリシーケンシング、全長転写産物解析、疾患感受性多型、ヒト組織解析

研究成果 :

長いリード配列(数 kb 以上)を特徴とする第三世代シーケンサー「PacBio RS」は、構成塩基の影響をほとんど受けることなく塩基配列決定が可能であることから、ゲノム構造異常解析、ハプロタイプ決定、全長 mRNA 配列決定など、従来のテクノロジーでは困難であった解析の実現可能性に期待が集まっている。最近になり、ヒト半数体細胞の全ゲノム解析結果が報告されるに至り(Nature 517:608-611, 2015)、レファレンスゲノムに残されたギャップ領域の50%以上が新たに判読可能となったことから、その有用性が示されつつある。

本分担研究課題では、2型糖尿病(T2D)病態解明の一環として、GWAS(ゲノムワイド相関解析)によって同定された疾患感受性領域内における cis 調節作用が、ハプロタイプ構造によってどのような変化がもたらされるのかを明らかにする。解析対象の T2D 責任臓器としては、組織特異的ヒストン修飾と T2D SNP(一塩基多型)との関連性から肝と膵島が有用であると考えられる(Nat Genet. 45:124-130, 2012)。日本人ヒト肝組織は JCRB 生物資源バンク(旧ヒューマンサイエンス研究資源バンク)から分譲を受けることが可能であるが、検体の入手・処理の簡便性や GWAS 結果の人種的背景から、市販の男性 Caucasian 由来初代培養ヒト肝細胞を用いることとした。

機能性をもつ GWAS SNP は cis エLEMENTのマーカーとなるヒストン修飾が豊富なゲノム領域と重複することが多い。このため、入手した $5 \times 10^6 \sim 1 \times 10^7$ 細胞/バイアルの一検体からは、ゲノム DNA とトータル RNA に加えて、ヒストン修飾の ChIP(クロマチン免疫沈降)サンプルを得られるよう、分割して培養を行った。ヒストン修飾の H3K27ac はアクティブなプロモーター・エンハンサーのマーカーとして有用であり(Nat Genet. 46:136-143, 2014)、検討候補の機能性 SNP を検索するために特異抗体を用いて ChIP を行い、次世代シーケンサー(HiSeq2000、イルミナ社)によるゲノム網羅的解析により同定を行った。

総計約 5 万か所におよぶ cis 調節ドメインが決定され、NHGRI Catalog データベースを参考に、様々な疾患の GWAS SNP との重複の有無を検討したところ、約 2,200 種類の GWAS SNP がヒト初代培養肝細胞の H3K27ac 領域中に見出され、機能性 SNP の可能性が高いと思われた。これら SNP は肥満関連形質、身長、血中代謝産物レベル(特に脂質)、血小板数、T2D などと関連するものが高頻度に含まれており、肝臓の生理的意義とよく合致するものと考えられた。SNP の genotype が遺伝子発現に及ぼす影響を検討するためには、ヘテロ接合体のサンプルを用いて H3K27ac エンリッチメントおよび mRNA 発現の allelic imbalance を検討する必要がある、ショートリードを用いた最近の検討から両者の方向性はエンハンサー内における塩基多型が規定するという仮説を支持している(Nature 518:350-354, 2015)。

H3K27ac 領域は数 kb 前後のサイズのもが多く、PacBio RS のリード長の点からは解析に好適と考えられた。最近になって、PacBioRS を用いたターゲットリシーケンシングのための新しい遺伝子濃縮方法が報告され(BMC Genomics 16:214, 2015)、数 kb にわたる検討が可能となってきた。ハプロタイプブロック内に存在し、かつ H3K27ac 領域と重複する複数の common SNP についてハプロタイプフェイジングを行うことができれば、アレル別の遺伝子発現との関連性についてさらに高精度な検討を行うことが可能となり、様々な疾患の遺伝素因に関する新たな知見が得られるものと期待される。

研究発表及び特許取得報告について

課題番号： 25指102

研究課題名： 「第三世代シーケンサーPacBio RSを活用した糖尿病・代謝疾患の病態解析」

主任研究者名： 安田 和基

論文発表

論文タイトル	著者	掲載誌	掲載号	年
A genome-wide association study for diabetic retinopathy in a Japanese population: potential association with a long intergenic non-coding RNA.	Awata T, Yamashita H, Kurihara S, Morita-Ohkubo T, Miyashita Y, Katayama S, Mori K, Yoneya S, Kohda M, Okazaki Y, Maruyama T, Shimada A, Yasuda K, Nishida N, Tokunaga K, Koike A.	PLoS One	9(11): e111715	2014
ATF7IP as a novel PDGFRB fusion partner in acute lymphoblastic leukemia in children. Br.	Kobayashi K, Mitsui K, Ichikawa H, Nakabayashi K, Matsuoka M, Kojima Y, Takahashi H, Iijima K, Ohtubo K, Oboki K, Okita H, Yasuda K, Sakamoto H, Hata K, Yoshida T, Matsumoto K, Kiyokawa N, Ohara A.	J Haematol	165(6): 836-841	2014
Comprehensive exploration of novel fusion genes in clear cell renal cell carcinomas using whole transcriptome analysis.	Gotoh M, Ichikawa H, Arai E, Chiku S, Sakamoto H, Fujimoto H, Hiramoto M, Nammo T, Yasuda K, Yoshida T, Kanai Y.	Genes Chromosome Canc	53(12): 1018-1032	2014
Druggable oncogene fusions in invasive mucinous lung adenocarcinoma.	Nakaoku T, Tsuta K, Ichikawa H, Shiraishi K, Sakamoto H, Enari M, Furuta K, Shimada Y, Ogiwara H, Watanabe S, Nokihara H, Yasuda K, Hiramoto M, Nammo T, Schetter AJ, Okayama H, Harris CC, Kim YH, Mishima M, Yokota J, Yoshida T, Kohno T.	Clin Cancer Res	20(12) 3087-3093	2014

研究発表及び特許取得報告について

学会発表

タイトル	発表者	学会名	場所	年月
「KCNQ1リスクアレルと生活習慣介入の耐糖能改善に及ぼす影響」	後藤温、山田晃一、宮地元彦、森田明美、後藤麻貴、出浦喜丈、竹澤純、村上晴香、佐々木敏、饗場直美、安田和基、寺内康夫、野田光彦、渡邊昌	第49回日本成人病（生活習慣病）学会学術集会	東京	平成27年1月
「糖尿病と遺伝子-今、臨床でどこまで応用できるのか-」	安田和基	湖東糖尿病セミナー	彦根	平成27年1月24日
Paternal allelic mutation at the Kcnq1 locus reduces pancreatic β -cell mass via epigenetic modification	Inoue H, Asahara S-i, Etoh H, Teruyama K, Shibutani Y, Ihara Y, Bartolome A, Hashimoto N, Matsuda T, Koyanagi-Kimura M, Kanno A, Nagashima K, Nishimura W, Yasuda K, Inagaki N, Seino S, Kasuga M, Kido T.	50th Annual Meeting of the European Association for the Study of Diabetes.	Vienna, Austria.	September 2014
「生活習慣介入とKCNQ1リスクアレルの耐糖能改善に及ぼす影響」	後藤温、山田晃一、宮地元彦、森田明美、後藤麻貴、佐々木敏、饗場直美、安田和基、寺内康夫、野田光彦、渡邊昌	第64回日本体質医学会講演会	大阪	平成26年9月
「生活習慣介入による耐機能改善に際してKCNQ1リスクアレルの及ぼす影響」	後藤温、山田晃一、宮地元彦、森田明美、後藤麻貴、出浦喜丈、竹澤純、久保田恵理、村上晴香、安田和基、野田光彦、渡邊昌	第57回日本糖尿病学会年次学術集会	大阪	平成26年5月
「膝島のゲノム網羅的解析による糖尿病発症機序の考察」	南茂隆生、宇田川陽秀、川口美穂、舟橋伸昭、上番増喬、平本正樹、西村渉、安田和基	第57回日本糖尿病学会年次学術集会	大阪	平成26年5月
「次世代シーケンサーを用いたゲノム、エピゲノム研究」	安田和基	第57回日本糖尿病学会年次学術集会	大阪	平成26年5月
「腎細胞がんが発現する新規融合遺伝子の同定」	後藤政広、市川仁、荒井恵吏、知久季倫、坂本裕美、藤元博行、安田和基、吉田輝彦、金井弥栄	第103回日本病理学会総会	広島	平成26年4月

研究発表及び特許取得報告について

その他発表(雑誌、テレビ、ラジオ等)

タイトル	発表者	発表先	場所	年月日

特許取得状況について ※出願申請中のものは()記載のこと。

発明名称	登録番号	特許権者(申請者) (共願は全記載)	登録日(申請日)	出願国

※該当がない項目の欄には「該当なし」と記載のこと。
※主任研究者が班全員分の内容を記載のこと。