

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 : 「第三世代シーケンサー」である PacBioRS 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、などを試みる計画を立案した。本技術はまだ発展途上の部分が大きく、特に哺乳類ゲノム・RNA についての報告はきわめて限定的であるため、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つの方向性を目指した。

[1] 本システムの性能を検証するために、ラムダファージ DNA からテンプレートを作成して、シーケンスを行った。10kb のサイズで DNA を断片化し、バイオアナライザーで断片化したサイズを確認したのち、テンプレートを作成し、ロングリードの特性を活かした「Continuous Long Reads (CLR)」モードで解析を行なった。データは、1次解析（ベースコール）の後、SMRT Portal 上で 2次解析を行い、「BLASR (Basic Local Alignment with Successive Refinement)」というツールを用いてレファレンス配列へマッピングを行った。その結果、「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、「ロングリード」の特性が確かめられた。マップされたサブリードについては、平均 accuracy が 81.6% と、本システムの特徴として知られている通り、第二世代 NGS よりやや低い。しかし、ケミストリーの特性に依存して特定の塩基でエラーを生じる他の NGS と異なり、いわゆるエラーはランダムに生じることから、50x 以上の depth を得られれば解決できると考えられた。実際この実験では、ほぼ例外なく 2000x 以上となっており、特定の配列でカバー率が下がることはないことが確かめられた。

以上より、第三世代 NGS としての PacBioRS の特性および有用性を確認することができた。

[2] 他の方法（第二世代 NGS など）に対する優位性を軸とした研究計画の再吟味

本機器の有効性と限界について、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、本テクノロジーに適した DNA 領域濃縮法は、まだ開発途上であること、定量性が求められるプロジェクト（定量的発現解析のための RNA-Seq など）については、現時点で第二世代 NGS に対する優位性は乏しいこと、真核生物の DNA 修飾（特にメチル化）の IPD パターンは複雑でまだレファレンスデータ収集が必要な段階であること、が判明した。以上をうけて、予定していた各サブテーマについて再吟味を行った結果、定量を目的とした RNA-Seq を中止し、目的ゲノム領域の濃縮方法についての検討を開始した。

[3] PacBioRS に適した標的ゲノム領域の濃縮法の検討：

① 既存のプロトコルの検討：

本テクノロジーでは、特徴的な「SMRTbell テンプレート」を作らせるため、既存の DNA 濃縮法はそのままでは使えないが、既存の試薬のプロトコルをベースに本テクノロジーの特性に合わせてカスタマイズ可能かどうかの検討を行った。PacBioRS の「long read」あるいは

「DNA 修飾」の利点を活かすため、まず相補的プローブのハイブリダイゼーション (hybridization) による方法のカスタマイズを試みたが、やはり各プロトコールはショートリード用フラグメントの濃縮に最適化されており、PacBio 用には問題があることがわかった。一方、PCR 増幅により大量のアンプリコン (amplicon) を得る方法は、PCR による長いフラグメントの増幅効率やエラーの問題、さらに DNA 修飾情報の消失など、そのままではやはり PacBioRS の特徴を活かしきれない可能性が高いと考えられた。

② NimbleGen プロトコールを用いた検討：

平成 27 年度になり、目的ゲノム領域について、長鎖 DNA シークエンス用のテンプレートを全ゲノムからエンリッチする方法が報告された (BMC Genomics 16: 214, 2015)。この方法をベースに SeqCap システム (ロシュ) のプロトコールを改良して、PacBio に適用できる方法が発表された。具体的には、ゲノム全体から 5~9kb のフラグメントにアダプターを付加後、PCR で増幅し、目的領域に応じて作成したプローブの mixture と液相でハイブリダイズさせ、ビーズキャプチャーを行った後に PacBio 用のアダプターを付加してシークエンスを行う、というもので、DNA 濃縮については、NimbleGen プラットフォームをベースとしている。これまでも類似のアイデアはあったが、従来プロトコールと比べ、①ゲノム断片化に、短いフラグメント (~200kb) 用の Covaris でなく、g-TUBE を用いる、②断片化後、フラグメントの回収には、AMPure Beads でなく、Blue Pippin を用いて電気泳動の原理で長いサイズのフラグメントを選択的に回収する、③末端にアダプター付加後の PCR 増幅では、酵素や反応時間、サイクルをフラグメントから変えて最適化する、の 3 点が大きな変化であり、これにより PacBioRS に適した 6~7kb 前後のフラグメントが得られるという。

プライマーは NimbleDesign を用いてヒトゲノムからプライマーデザインを行う本法では、ゲノム DNA の断片が大きいため、プローブをエクソン領域中心にデザインしてもイントロンを含んだ遺伝子領域全体を広くカバーできる可能性がある。

この方法の検証のため、標的領域の選定を行った。ターゲット領域は、MODY など単一遺伝子病タイプの原因遺伝子領域約 10、*KCNQ1* や *TCF7L2* など GWAS で 2 型糖尿病に関連した SNP 関連の領域約 15、などを中心とした。現在、プローブデザインが進行中であり、作成次第 validation を行う段階である。

[4] ハプロタイプフェイジング (haplotype phasing) に適した機能的 SNP 領域の探索：

ハプロタイプが機能的な意義を持つ領域は、プロモーター/エンハンサー活性をもつ等が考えられるが、ヒストン修飾 (H3K27ac など) ほかエピゲノム解析が、その機能及び機能変化の手がかりとなる。そこまで、まず解析用サンプルとして、エピゲノム解析に耐えうるヒト組織の入手方法を検討したところ、米国の企業経由で市販されている初代培養ヒト肝細胞 (1 バイアルあたり 500 万~1,000 万細胞) が比較的均一な細胞集団が期待され、少なくとも 12 例分入手可能であることを確認し、購入した。

GWAS により報告された疾患感受性 SNP は、遺伝子発現調節の cis エレメントのマーカール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ルとして有用であり、機能的ゲノム領域を

検索するために特異抗体を用いてChIPを行い、この目的に適したショートリードNGS (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。

その結果、総計約5万か所 (49,485) におよぶ cis 調節ドメインが決定された。NHGRI Catalog データベースを参考に、様々な疾患の GWAS SNP (全21,928) との重複の有無を検討したところ、約2,442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、そして2型糖尿病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。特に2型糖尿病関連 SNP のうち30SNP が、この肝 H3K27ac 領域に存在し、うち15SNP はリファレンス配列で転写因子モチーフの構成塩基で、さらにそのうち12SNP (80%) は variant 配列で、モチーフが消失した。これらは cis に作用する機能的エレメントとしての解析が興味深いと思われた。

今後、発現レベルの高い遺伝子について、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 についてヘテロ接合体のサンプルを用いてハプロタイプフェイジングを行うことができれば、遺伝子発現との関連をアレル特異的に高精度に検討を行うこと (allelic expression、あるいは allelic imbalance) 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。

[5] 今後の見通し：

上述した DNA 濃縮法によるターゲットリシーケンスでは、やはり PCR で増幅するため、DNA 修飾状態、特にメチル化状態に関する情報は、十分解析できない。DNA 修飾に関する情報については、やはり PCR を用いない方法により標的領域を濃縮する必要があるが、十分量の DNA からスタートして生データ (波形データ) を保存しておけば、近い将来メチル化コールができた場合、過去にさかのぼって検出・同定することができる。

また、PacBio における「エラー」は、特定の chemical (塩基) に依存せずにランダムに生じるが、ターゲットリシーケンスの場合は、全領域にわたってかなりの depth が必要であり、特に heterozygous な SNP のハプロタイプ解析には、500x 以上の depth が望ましいと思われる。そのため、上述の方法をやはり通常のショートリードに比べるとスタートゲノムテンプレート量が圧倒的に必要である。さらに、長いフラグメントの PCR 増幅により、ランダムに入るエラーの判定、などが重要となり、これらの点は今後技術革新が必要となろう。

最近 PacBioRS を用いて ヒトゲノム配列の「gap」を埋めるのに成功したという報告が目まぐるしく注目を集めている。また日本人標準ゲノム配列についても、ショートリード NGS と PacBio システムをくみあわせて完成し報告された。このように、本システムのヒトゲノム科学における重要性・有用性はますます高まりつつあり、今後も引き続き糖尿病関連病態の解明への応用を目指したい。

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 words : 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 planned to utilize this unique system for the diabetes research. We tried to perform ‘feasibility studies’ for assessing the performance of PacBioRS compared with existing technologies using known sequences, and ‘innovative studies’ towards new applications of this technology to the genomic and epigenomic studies of diabetes mellitus.

1) In order to assess the performance of PacBioRS, we sequenced lambda phage DNAs. Templates were prepared from 10kb fragments on average, and run was done by Continuous Long Reads (CLR) modes. After base calling, data were analyzed on SMRT Portal, and mapped to the reference sequence using of the lambda phage BLASR (Basic Local Alignment with Successive Refinement) algorithm.

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. The average base accuracy was 81.6%, which was lower than commonly used short-read NGS ; however, since base call errors happened randomly and were not dependent on chemical or structural features of some specific bases, we think we can expect reliable base calls if the coverage was reasonably deep.

2) We made discussion several times with scientists from Tomy Digital Biology Ltd., over the advantages and disadvantages of PacBioRS, and reached a couple of conclusions. Although DNA enrichment of targeted regions is critical for the success of this technology when we deal with large genomes such as human, there were no established methods for this step at the beginning of this project. Quantitative experiments such as tag counting of RNA-Seq would be better performed, at least for the present, by other short-read NGS due to the overwhelming read counts. DNA modification of mammalian genomes is more complicated than previously supposed and proves not to exhibit consistent patterns of inter-pulse duration(IPD) by PacBio sequencing. Therefore, we modified the whole plan, and started to develop efficient methods for DNA enrichment of targeted regions.

Researchers には、分担研究者を記載する。

There are basically two strategies for DNA enrichment, namely, hybridization-based and amplicon-based methods. In general, hybridization-based DNA enrichment is more suitable for long reads and analysis of DNA modification, but most of the enrichment kits available are optimized only for short-read NGS. A great amount of DNA is also necessary as a starting material. On the other hand, amplicon-based method has an advantage in sparing DNA, but also has problems such as PCR errors, and loss of information on DNA modification.

We first tried to modify liquid-phase hybridization method for optimization, but were not successful enough.

In 2015, a new method based on NimbleGen platform was made available. This method has several unique features. 1) genome fragmentation is performed by g-TUBE rather than Covaris, which enabled us to get longer fragments. 2) Selective recovery of long DNA fragments are done with electrophoresis-based Blue Pippin system rather than AMPure Beads. 3) DNA fragments are captured by a mixture of NimbleGen probes designed by the specific software (NimbleDesign). 4) PCR amplification is best optimized after adapter ligation. These features made enrichment of targeted DNA fragments with the size of 6-7kb possible.

We can even expect to recover long fragments by this method using exon-centric probes, if the targeted regions don't have too large introns. Since we concluded this system is the most promising method of DNA enrichment for PacBioRS, we selected target genomic regions for verification of this method. We decided to focus on monogenic type of diabetes and regions containing diabetes susceptibility SNPs. We picked up 13 genes including MODY genes (*HNF4A*, *GCK*, *HNF1A*, *HNF1B*, *PDX1* etc) and others (*INS*, *INSR*, *MC4R* etc), and about 15 SNP loci including *KCNQ1* and *TCF7L2* which confer susceptibility for type 2 diabetes, and are now designing the probes.

3) We also planned to perform genomic (haplotype phasing), epigenomic and transcriptomic (allelic expression and alternative splicing) analyses in the tissues or cells important for energy metabolism and decided to use human primary hepatocytes derived from Caucasian donors. In order to obtain regulatory genomic regions, we performed a ChIP-seq analysis for H3K27ac, a hallmark of active promoters and enhancers. We used $5 \times 10^6 \sim 1 \times 10^7$ cells for RNA extraction and ChIP-seq analysis for H3K27Ac, and identified more than 50,000 active marks for H3K27ac.

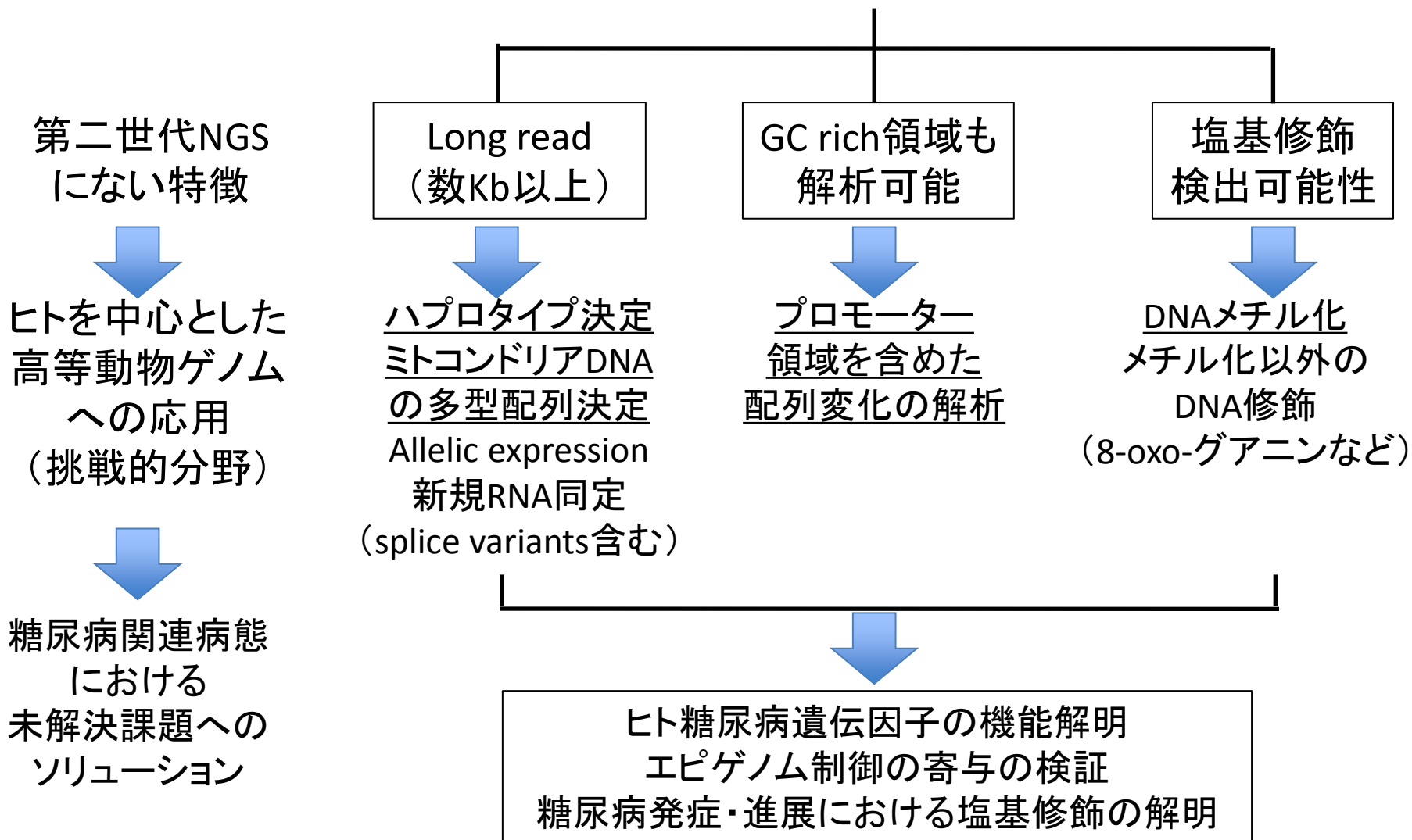
We also picked up more than 15,000 disease-associated SNPs from GWAS database for various diseases including diabetes mellitus, as potential candidates of functional genetic variations. Among them, ~2200 SNPs were localized in H3K27ac peaks in primary human hepatocytes, which means susceptibility genes previously reported by various GWASs were significantly enriched in these regions. These SNPs were related to obesity-related traits, height, blood metabolite levels (including lipids), platelet counts, and type 2 diabetes.

Some genes exhibited substantial variations in the signal peaks among different samples. We hypothesized that SNP genotypes and/or haplotypes might have impact on the epigenetic status, and are now in the process of validating the relationship between the haplotype phase revealed by target resequencing using long-read NGS and allelic imbalance of expression of the genes.

Researchers には、分担研究者を記載する。

4) Recently, the human genome re-annotation using PacBioRS was reported. The reference Japanese genome sequence was also reported, which could not be achieved without the help of PacBioRS. There have been increasing applications of the long read NGS to the genomic research of human health, and it is highly expected that several remaining issues for this unique technology would be solved in the near future.

第三世代NGS: PacBio RSの特性とサブテーマの吟味(当初計画)



下線部分: 本システムの特性を踏まえて、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

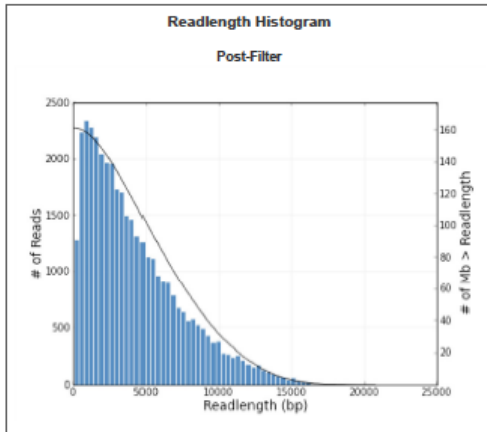


図2:サブリード長分布

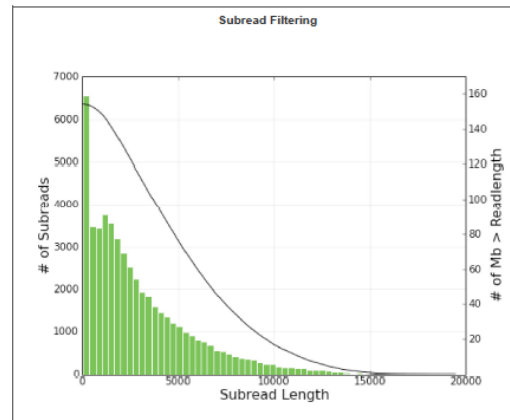


図3: Read Quality (RQ)

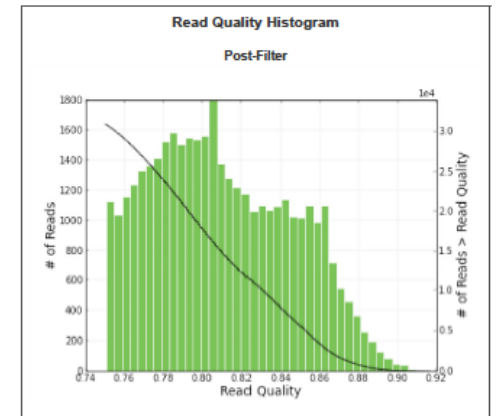


図4: Accuracy

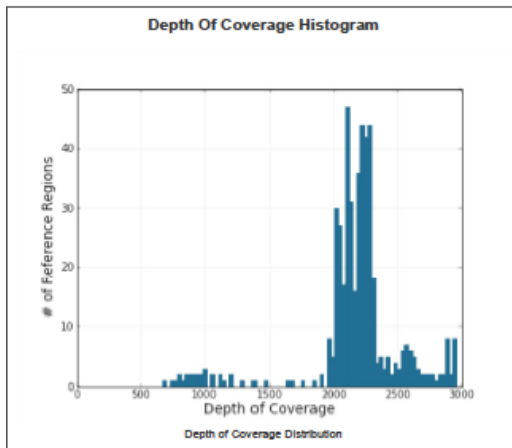
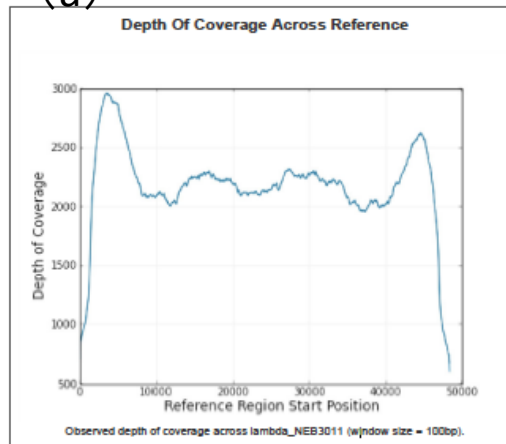
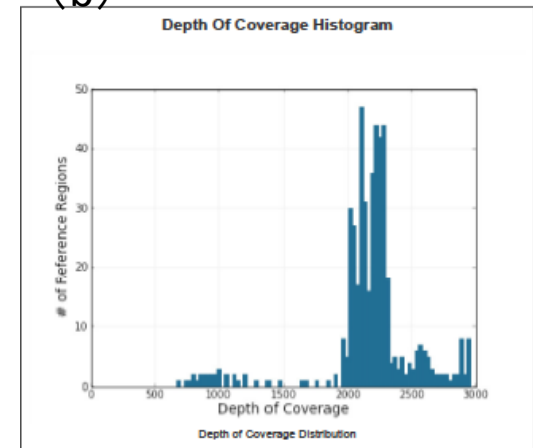


図5:カバー率分布

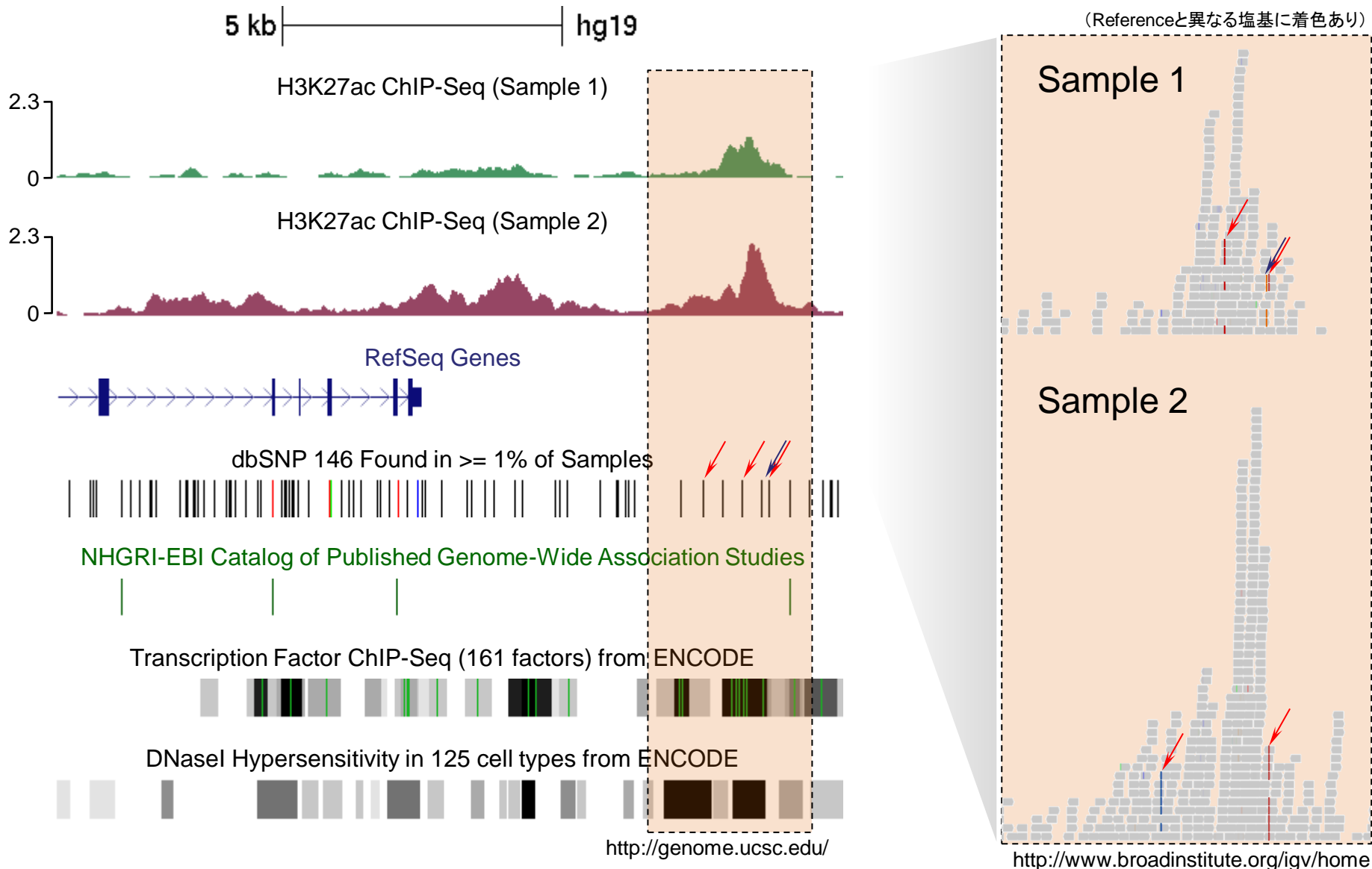
(a)



(b)



初代培養肝細胞サンプル間に認められたH3K27acと遺伝子型の相違

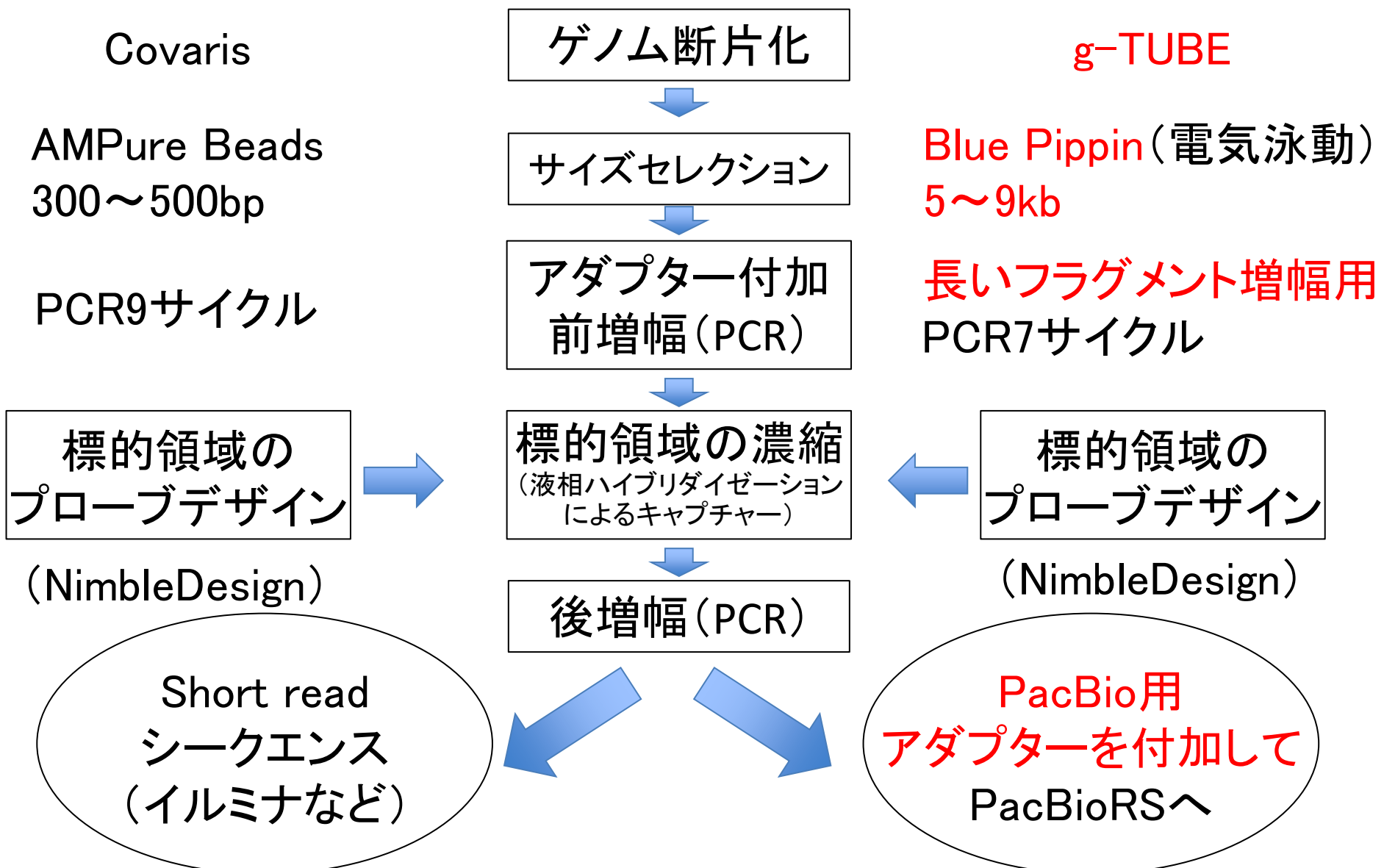


破線内のゲノム領域では少なくとも4か所に遺伝子型の相違が認められており、ゲノムアクセシビリティと転写因子の結合部位によって特徴付けられている。

PacBio用に最適化したゲノム濃縮法 (SeqCapシステムによる)

<標準プロトコール>

<PacBio用>



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 :

[1] 第三世代シーケンサーPacBioRS の性能を検証するために、ラムダファージ DNA から Template を作成して、シーケンスを行った。

PacBioRS で用いられる「SMRTbell テンプレート」については、「Continuous Long Reads (CLR)」、「Circular Consensus Sequencing (CCS)」という 2 つのモードがあるが、我々は長いインサートを用いたロングリードの特性を検証するために前者を採用し、10kb のサイズで DNA を断片化して、バイオアナライザーにより断片化したサイズを確認したのち、テンプレートを作成した。データは、1 次解析 (ベースコール) の後、SMRT Portal 上で 2 次解析を行い、レファレンス配列へマッピングを行う、「BLASR (Basic Local Alignment with Successive Refinement)」というツールを用いて解析した。

「リード長」は、10~15kbp まで分布しており、「ロングリード」という本システムの有用性が示された。リードのうち、両端の adaptor には含まれた「サブリード」も、2~5kbp 程度が多く、またリードごとの精度「Read Quality (RQ)」も良好であった。マップされたサブリードの accuracy は平均 81.6%であったが、いわゆるエラーはケミストリーに依存せず、50x 以上の depth で「コンセンサス配列」を作成すれば精度は 100%に近づくとされており、depth を十分得られれば課題である fidelity も解決可能と考えられた。今回はほぼどの領域も 2000x 以上となっており、この点で全く問題はなかった。

[2] 他の方法 (第二世代 NGS など) に対する優位性を軸とした研究計画の吟味

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、以下のことが明らかになった。

- 1) 他の NGS と比較してリード数が少ないため、単独でヒトゲノム全域を対象とするのは難しく、目的領域の十分な濃縮 (エンチッリメント) が必要だが、本テクノロジーに適した DNA 領域濃縮法は、まだ開発途上であり、プロジェクト開始時には公式にサポートされたキット・プロトコールはなかった。
- 2) リード数が多くないこととも関連して、定量性が求められるプロジェクト (RNA-Seq による定量的発現解析など) については、現時点では第二世代 NGS に対する優位性は乏しい。
- 3) 真核生物では DNA 修飾がこれまでの予想以上に多彩であり、最も頻度が高くかつ機能的に重要と思われる「シトシンメチル化」だけをとっても、波形間の inter-pulse duration (IPD) が単一のパターンを示さず、その原因も未解決で、まだ十分なデータが蓄積されていない。

[3] PacBioRS に適した標的ゲノム領域の濃縮法の検討 :

1) 既存のプロトコールの検討 :

本テクノロジーでは特徴的な「SMRTbell テンプレート」を作らせるため、既存の DNA 濃縮法はそのままでは使えず、カスタマイズ可能かどうかの検討を行った。

一般に目的のゲノム領域の濃縮法には、相補的プローブのハイブリダイゼーション (hybridization) による方法と、PCR 増幅により大量のアンプリコン (amplicon) を得る方

法とがある。PacBioRSの「ロングリード」あるいは「DNA修飾解析可能」の利点を活かすためには、前者が、汎用されているキットはショートリードNGS用に最適化されており、そのままではPacBioRSのライブラリー作成に適さないことがわかった。一方、後者のアンプリコンによる方法は、スタートDNAを節約できるなどのメリットはあるが、PCRによる増幅効率やエラーの問題、さらにDNA候補情報の消失など、PacBioRSの特徴を活かしきれない可能性が高い。そのため前者にあたる某社のプロトコルを検討して、カスタマイズを試みたが、やはり問題があることがわかった。

なお、非常に多型の多いミトコンドリアDNAの「phasing」も本システムに適した検討テーマであるが、ヒトミトコンドリアDNAを特異的に濃縮するキットは市販されていないものの、細胞あたりのコピー数が多いため、エクソーム用の市販濃縮キットでもミトコンドリアDNAの濃縮ができる可能性は残っていると考えている。

2) NimbleGenプロトコル：

平成27年度になり、目的ゲノム領域について、長鎖DNAシーケンス用のテンプレートを全ゲノムからエンリッチする方法が報告された(BMC Genomics 16: 214, 2015)。この方法をベースに、SeqCapシステム(ロシュ)のプロトコルの改良により、ロングリード用のDNA濃縮用のプロトコルとして発表された。具体的には、ゲノム全体を比較的大きく断片化してサイズセレクションを行い、5~9kbのフラグメントにシーケンス用アダプターを付加後、PCRで7サイクル増幅し、目的領域に応じて作成したプローブのmixtureと液相でハイブリダイズさせ、ビーズキャプチャーを行った後にPacBio用のアダプターを付加してシーケンスを行う、というもので、DNA濃縮については、NimbleGenプラットフォームをベースとしている。これまでも類似のアイデアはあったが、従来プロトコルと比べ、①ゲノム断片化に、短いフラグメント(~200kb)用のCovarisでなく、g-TUBEを用いる、②断片化後、AMPure Beadsでなく、Blue Pippinを用いて電気泳動の原理で長いサイズのフラグメントを選択的に回収する、③アダプター付加後のPCR増幅では、酵素や反応時間、サイクルをフラグメントから変えて最適化する、の3点が大きな変化であり、これによりPacBioRSに適した6~7kb前後のフラグメントが得られる。

この方法の検証のため、標的領域の選定を行った。プライマーはNimbleDesignを用いてヒトゲノムからプライマーデザインを行う。本法では、ゲノムDNAの断片が大きいため、イントロンが1-2kb程度であれば、プローブをエクソン領域中心にデザインしてもイントロンを含んで隣接エクソンをまたぐフラグメントも得られ、遺伝子領域全体を広くカバーできる。その結果として、より大きな構造変異の検出、ハプロタイプフェージングなどが、期待される。実際この方法にて非常に多型性の高い、かつハプロタイプ構造が重要な意味をもつHLA領域の解析に成功したとの報告がある。

ターゲット領域は、MODY 遺伝子を含む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(*KCNQ1*、*TCF7L2*ほか)、などを中心に25領域を選択した。今後実サンプルも用いて検討予定である。

課題としては、やはりPacBioにおける「エラー」のため、リシーケンスの場合は、全領域にわたってかなりのdepthが必要であり、特にheterozygousなSNPのハプロタイプ解析には、500x以上が望ましく、通常のショートリードに比べるとテンプレート量が圧倒的に多く必要である。今後この点でもさらに、技術革新が期待される。

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 :

【背景・目的】

ゲノムワイド相関解析 (GWAS) により、2型糖尿病をはじめとした common disease の感受性一塩基多型 (SNP) の報告は 2007 年頃から増加したが、主任研究者の安田は国家プロジェクトとしての「ミレニアムゲノムプロジェクト」において東アジア人の強力な 2 型糖尿病関連遺伝因子 *KCNQ1* の同定に成功した (Nat Genet. 40:1092-1097, 2008)。これら SNP のほとんどは非コード DNA 領域に認められたことから、遺伝子発現調節の変化が疾患感受性のメカニズムとして重要であると推察されたが、実際にヒト臍島を用いた検討によって、2 型糖尿病と最も関連の強い *TCF7L2* 遺伝子の SNP (rs7903146) 周辺のゲノム領域が、クロマチン構造の変化によってエンハンサー活性が変化することを私たちは明らかにした。ところで、疾患感受性座位には複数の cis 調節領域がしばしば存在し、このような領域にも SNP が認められる場合がある。よって、ハプロタイプ構造を明らかにした上で (Haplotype phasing)、疾患感受性領域内における cis 調節作用の総和を考慮すれば、疾患感受性の要素である遺伝子発現変化をより正確に理解できるかもしれない。

第三世代シーケンサー「PacBio RS」は、長いリード配列 (現在、平均~10kbp) を特徴とし、構成塩基 (GC/AT 含量) 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に威力を発揮しつつある。最近になり、ヒト半数体細胞 CHM1 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。

本分担研究課題では、2 型糖尿病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をもたらされるのかを明らかにすることを目的として開始した。

【方法】

1. 解析対象の 2 型糖尿病責任臓器としては、組織特異的ヒストン修飾様式と 2 型糖尿病 SNP との関連性から肝と臍島が有用であると考えられた (Nat Genet. 45:124-130,2012)。検体の入手・実験方法の簡便性や GWAS 結果の人種的背景から、市販されている男性 Caucasian 由来初代培養ヒト肝細胞を用いた。
2. 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。入手した $5 \times 10^6 \sim 1 \times 10^7$ 細胞/バイアルの検体からは、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。
3. 様々な疾患や形質の GWAS SNP (NHGRI-EBI Catalog) を UCSC ゲノムブラウザー (hg19) からダウンロードした (~15,000 種類)。また、H3K27ac の ChIP-Seq ピーク領域に存在する GWAS SNP を明らかにした。

4. PacBioRS を用いたターゲットリシーケンシングのためのカスタムデザインツール (NimbleDesign) 、および遺伝子濃縮方法 (Roche NimbleGen's SeqCap EZ システム) を確認した。

【結果】

初代培養肝細胞の H3K27ac ChIP-Seq 結果の再現性はサンプル間で良好であり、総計約 5 万か所の cis 調節領域が決定された。約 1 万 5 千種類の GWAS SNP のうち、約 2,200 種類の SNP が初代培養肝細胞の H3K27ac 領域内に見出され、機能性 SNP である可能性が考えられた。これら SNP が相関する形質・疾患としては頻度順に“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”、“2 型糖尿病”などが認められ、肝臓の生理的意義に矛盾しない所見と考えられた。

サンプル間で各 H3K27ac ピークのシグナルを比較すると、遺伝子によってはかなり大きな変化を示すものが観察された (パワーポイント参照)。H3K27ac エンリッチメントと mRNA 発現の個人差は H3K27ac 領域内の塩基多型が規定すると報告されているが (Nature 518:350-354,2015) 、初代培養肝細胞を用いた本検討でも、H3K27ac が大きく変化しているサンプル間では遺伝子型に関しても多くの場所で異なっていることが判明し、本研究のための詳細な検討に適していると考えられた。解析遺伝子のコーディング SNP にヘテロ接合体が認められるサンプルでは、Haplotype phasing によってアレル別に H3K27ac 領域の遺伝子型の違いも全て明らかにした上で、mRNA および各 H3K27ac の allelic imbalance の観察を行えば、疾患感受性に関与する SNP の詳細な検討が可能となるために、PacBioRS によるターゲットリシーケンシングの検討を行っている。

【考察】

第三世代シーケンサーはその登場以来着実にパフォーマンスを伸ばしつつあり、100Mbp までのゲノムサイズであれば、ほぼ染色体全域にわたるアセンブリが可能となっている。ハプロタイプ情報も加味された構造多型の正確な把握は、遺伝子発現や疾患感受性の理解に不可欠であることが明らかとなりつつあり (Nat Genet. 47:921-925, 2015) 、超ロングリードによる塩基配列決定が疾患研究を含めて生物学に与えるインパクトは重要である。

本研究で用いた初代培養肝細胞は、各サンプルの検討を同一条件下に行うことが可能であるために、特に遺伝素因研究には適していると考えられた。最近になって、2 型糖尿病感受性座位のファインマッピングが検討されるようになり (Nat Genet. 47:1415-1425, 2015) 、臓器特異的クロマチン構造との関連性から、病態に関与する臓器の同定が具体化しつつある。第三世代シーケンサーを用いた様々な臓器の詳細な検討によって得られる知見は、新たな創薬標的の同定など広範な応用も期待できるであろう。技術進歩が現状のペースを維持できれば、数年以内にヒトゲノムのアセンブリが可能となるリード長が達成されるという試算もあるが (doi: <http://dx.doi.org/10.1101/048603>) 、莫大な費用の問題からは容易な日常的使用は困難である。この点からは、SeqCap EZ システムを用いたターゲットリシーケンシング (BMC Genomics 16:214,2015) の有効利用には期待が持てる。また、液滴バーコーディングシステム (Nat Biotechnol. 34:303-311, 2016) や DNA クロスリンクとメイトペア法を組み合わせた手法 (Genome Res 26:342-350, 2016) などの工夫によって、ショートリードシーケンスを活用する方法は費用や開始 DNA 量の点からも高い有用性が期待できる。しかし、真に正確なゲノムアセンブリを行うには安定した「シーケンスの超ロングリード化」が何事にも代えがたく、今後の技術発展には大きな期待が寄せられている。

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
Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes.	2. Imamura M, Takahashi A, Yamauchi T, Hara K, Yasuda K, Grarup N, Zhao W, Wang X, Huerta-Chagoya A, Hu C, Moon S, Long J, Kwak SH, Rasheed A, Saxena R, Ma RCW, Okada Y, Iwata M, Hosoe J, Shojima N, Iwasaki M, Fujita H, Suzuki K, Danesh J, Jørgensen T, Jørgensen ME, Witte DR, Brandslund I, Christensen C, Hansen T, Mercader JM, Flannick J, Moreno-Macias H, Burtt NP, Zhang R, Kim YJ, Zheng W, Singh JR, Tam CHT, Hirose H, Maegawa H, Ito C, Kaku K, Watada H, Tanaka Y, Tobe K, Kawamori R, Kubo M, Cho YS, Chan JCN, Sanghera D, Frossard P, Park KS, Shu X-O, Kim B-J, Florez JC, Tusie-Luna T, Jia W, Tai ES, Pedersen O, Saleheen D, Maeda S, Kadowaki T.	Nature Commun	7: 10531	2016
Comparable analysis of type 2 diabetes-associated SNP alleles identifies allele-specific DNA-binding proteins for the KCNQ1 locus.	Hiramoto M, Udagawa H, Watanabe A, Miyazawa K, Ishibashi N, Kawaguchi M, Uebanso T, Nishimura W, Nammo T, Yasuda K.	Int J Mol Med	36(1): 222-230	2015
Establishment of maturity-onset diabetes of the young (MODY)-induced pluripotent stem cells from a Japanese patient.	Yabe S, Iwasaki N, Yasuda K, Hamazaki T, Konno M, Fukuda S, Takeda F, Kasuga M, Okochi H.	J Diabetes Invest	6(5): 543-547	2015
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
Genome-wide association study identifies three novel loci for type 2 diabetes.	Hara K, Fujita H, Johnson TA, Yamauchi T, Yasuda K, Horikoshi M, Peng C, Hu C, Ma RC, Imamura M, Iwata M, Tsunoda T, Morizono T, Shojima N, So WY, Leung TF, Kwan P, Zhang R, Wang J, Yu W, Maegawa H, Hirose Hi DIAGRAM Consortim, Kaku K, Ito C, Watada H, Tanaka Y, Tobe K, Kashiwagi A, Kawamori R, Jia W, Chan JC, Teo YY, Shyong TE, Kamatani N, Kubo M, Maeda S, Kadowaki T.	Hum Mol Genet	23(1): 239-246	2014
Poor responses to tyrosine kinase inhibitors in a child with precursor B-cell acute lymphoblastic leukemia with SNX2-ABL1 chimeric transcript.	Masuzawa A, Kiyotani C, Osumi T, Shiota Y, Iijima K, Tomita O, Nakabayashi K, Oboki K, Yasuda K, Sakamoto H, Ishikawa H, Hata K, Yoshida T, Matsumoto K, Kiyokawa N, Mori T.	Eur J Haematol	92(3): 263-267	2014
Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants.	Pasquali L, Gaulton KJ, Rodriguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Morán I, Gómez-Marín C, van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T, Cebola I, García-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Gómez-Skarmeta JL, Müller F, McCarthy MI, Ferrer J	Nat Genet	46(2): 163-143	2014

学会発表

タイトル	発表者	学会名	場所	年月
「2型糖尿病など代謝性疾患のゲノム解析等の現状と展望」	安田和基	第18回ヒューマンサイエンス総合研究ワークショップ:ゲノム解析技術は創薬を変えるか	東京	平成28年3月
「糖尿病にかかわる遺伝学・体質の基礎」	安田和基	第50回糖尿病学の進歩	東京	平成28年2月
「糖代謝異常の罹患体質を考える」	安田和基	第43回内科学の展望	岐阜	平成27年11月

研究発表及び特許取得報告について

「A meta-analysis of Japanese genome-wide association studies identified seven novel susceptibility loci to type 2 diabetes.」	Imamura M, Maeda S, Yamauchi T, Hara K, Iwata M, Hirose H, Yasuda K, Watada H, Maegawa H, Ito C, Tanaka Y, Tobe K, Kaku K, Kawamori R, Kadowaki T.	51th Annual Meeting of the European Association for the Study of Diabetes	Stockholm, Sweden	September 2015
「糖尿病の成因の解明にむけて—KCNQ1の同定から多層的疾患オミックス研究まで—」	安田和基	第4回糖尿病トランスレーションリサーチ研究会	弘前	平成27年7月
「網羅的cis調節領域の検討による、高脂肪食に続発する膵島代謝機序の解明」	南茂隆生、宇田川陽秀、舟橋伸昭、安田和基	第4回NGS現場の会	筑波	平成27年7月
「なぜ糖尿病になる人とならない人がいるのか—ゲノムからバイオバンクまで—」	安田和基	第3回国立国際医療研究センター (NCGM) 市民公開講座『糖尿病の明日を考える—センターが進める治す医療への第一歩—』	東京	平成27年6月
「A meta-analysis of Japanese genome-wide association studies identified seven novel susceptibility loci to type 2 diabetes.」	Imamura M, Maeda S, Yamauchi T, Hara K, Takahashi A, Kubo M, Iwata M, Hirose H, Yasuda K, Watada H, Maegawa H, Ito C, Tanaka Y, Tobe K, Kaku K, Kawamori R, Kasuga M, Kadowaki T.	75th Scientific Sessions, American Diabetes Association.	Boston, USA	June 2016
「膵島のゲノム網羅的解析による膵島代償機序/糖尿病発症機序関連因子の同定」	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58回日本糖尿病学会年次学術集会	下関	平成27年5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回日本内科学会講演会	大阪	平成26年9月
「生活習慣介入による耐糖能改善に際して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
「膵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
「教育講演：遺伝子異常と糖尿病」	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
「糖尿病の遺伝子解析の現状と課題—遺伝・環境相互作用も含めて」	安田和基	Scientific Exchange Meeting糖尿病における遺伝・環境の相互作用	宇都宮	平成26年2月
「糖尿病の遺伝子と臨床」	安田和基	第28回糖尿病研修セミナー	神戸	平成25年12月
「2型糖尿病遺伝因子はどこまでわかったか」	安田和基	第24回新潟糖尿病臨床研究会	新潟	平成25年10月
「ゲノム網羅的解析を用いた高脂肪食摂取による膵島の代償機序の解明」	南茂隆生、宇田川陽秀、川口美穂、衛藤弘城、上番増喬、平本正樹、西村渉、安田和基	第3回NGS現場の会	神戸	平成25年9月
「教育講演：糖尿病診療に遺伝情報は役立つのか」	安田和基	第13回糖尿病・情報学会	徳島	平成25年8月
「ゲノム網羅的解析を用いた、高脂肪食摂取による膵島の代償機序の解明」	南茂隆生、宇田川陽秀、川口美穂、衛藤弘城、上番増喬、平本正樹、西村渉、安田和基	第56回日本糖尿病学会年次学術集会	熊本	平成25年5月
「Type 2 diabetes-associated SNPs with in KCNQ1 gene modulate the affinity of the locus for DNA-binding factors」	Hiramoto M, Udagawa H, Watanabe A, Kawaguchi M, Nishimura W, Nammo T, Yasuda K.	Annual Scientific Meeting, American Diabetes Association	Chicago, USA	2013年6月

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

研究発表及び特許取得報告について

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。