

平成 28 年度 委託研究開発成果報告書

I. 基本情報

事業名：(日本語) ゲノム医療実現推進プラットフォーム事業先端ゲノム研究開発
(英語) Advanced Genome Research and Bioinformatics Study to Facilitate Medical Innovation

研究開発課題名：(日本語) ヒトゲノム De Novo 情報解析テクノロジーの創出
(英語) Informatics for analyzing de novo human genome assemblies

研究開発担当者 (日本語) 国立大学法人東京大学・大学院新領域創成科学研究科・教授・森下 真一
所属 役職 氏名：(英語) Professor, Graduate School of Frontier Sciences, University of Tokyo, Shinichi Morishita

実施期間：平成28年9月1日 ～ 平成29年3月31日

分担研究 (日本語)
開発課題名：(英語)

研究開発分担者 (日本語)
所属 役職 氏名：(英語)

II. 成果の概要 (総括研究報告)

和文

- (1) 長鎖 DNA 解読技術による日本人集団の De Novo ゲノム解読と、構造変異検出アルゴリズムの設計とデータ公開： PacBio Sequel、10X Chromium、Oxford nanopore MinION を入札し、平成 29 年 1 月に導入した。平成 28 年度末までに稼働させ解読を開始した。長鎖 DNA 解読の対象となるサンプル選択にあたっては、日本人ゲノム中の構造多型の多様性をカタログ化するため正常人コントロールサンプルの解読を優先し、10 サンプルの解読を終了した。収集したデータを用いて PacBio Sequel および Oxford nanopore の解読性能 (生データのリード長・エラーパターン、解読コスト、ゲノムアセンブリした結果の塩基精度等) を分析する作業を開始した。構造変異の検出、アレル特異的メチル化および発現量解析については米国の 2 大学との共同研究が順調に進んでいる。セントロメア配列のアセンブリについては、Hi-C データの利用が有効であることが分かりつつある。DNA サンプルは正常人コントロールに加えて、脳疾患罹患者、消

化器癌・造血器悪性腫瘍・肺がん・大腸がん等の臨床試料を対象に、導入した長鎖 DNA 解読装置を使った構造変異解析を開始した。平成 30 年度末のデータ公開に向けて対象とする検体のリストを作成し計画通りプロジェクトが進行中である。

- (2) 第 2 種過誤を克服する疾患関連変異探索アルゴリズム： これまでに蓄積している 5,000 名規模のエキソームデータを、長鎖 DNA 解読装置で得られた日本人 *de novo* ゲノムデータで再分析し、第 2 種過誤を克服する疾患関連変異探索・学習アルゴリズムの研究開発を開始した。
- (3) 時系列コホート情報と腸内細菌叢等の環境因子を分析するアルゴリズム設計： PacBio Sequel を用いて 12 名の腸内細菌叢サンプルを解読し分析した。微生物を単離することが困難な腸内細菌叢からでも、長鎖 DNA を利用すると短鎖 DNA を使う場合に比べて 2 桁以上長い contig 配列が得られた (N50 contig 長で評価)。一部の微生物については完全な環状ゲノムを決定でき、水平伝播する遺伝子の全貌を把握できた。さらにプラスミド、ファージの DNA を完全に決めることができ、12 サンプル間で大きな違いがあることが分かった。PacBio Sequel の 1 分子計測技術を使い DNA のメチル化状態 (6mA/4mC) を推定し、周辺モチーフを同定したところ、12 サンプル内で多様であることが分かった。腸内細菌叢では、プラスミド、ファージ、DNA メチル化状態は多様であるというこの成果を踏まえ、これらの多様性がどのように疾患に関与するかを今後は調べてゆく。
- (4) 低コストかつ高速な構造変異クリニカルシーケンシング法： 短鎖タンDEM 繰返し配列 (short tandem repeat expansion) を低コストで検出する方法を改善した。短鎖 DNA 解読装置で予測した繰返し領域を PCR 増幅し、そのアンプリコンを PacBio Sequel で読む方法を開発した。

英文

- (1) **De novo sequencing of Japanese personal genomes by long read sequencing technology, development of algorithms for detecting structural variants, and database construction**

We purchased PacBio Sequel, 10X Chromium, and Oxford nanopore MinION in January 2017, and started *de novo* sequencing of personal genomes. To generate a catalog of structural variants in the Japanese population, we first sequenced ten healthy control samples. Subsequently, we have started comparing the performance of PacBio Sequel and Oxford nanopore MinION in terms of a variety of measurements such as read length distribution, error patterns in base calling, sequencing cost, and genome assembly quality. On the analysis of structural variants in the Japanese population, we have been collaborating with a research group in the US. We developed an algorithm for estimating allele-specific methylation and have been working with a research group in the US to understand the involvement of allele-specific methylation with allele-specific gene expression. To sequence centromere regions, we found Hi-C data were quite useful. In addition to sequencing healthy control DNA samples, we also have sequenced DNA samples from brain diseased persons and cancer cells using the installed long-read sequencers so as to analyze structural variants. We listed candidate DNA samples to be sequenced for restricted or open access at the end of 2018 fiscal year.

(2) **Algorithms for identifying disease-associated mutations while minimizing false-negatives**

With ~5000 Japanese exome data and *de novo* human genome assemblies that will be generated in this project, we have started developing machine learning algorithms for detecting disease-associated mutations while minimizing false-negatives.

(3) **Algorithms for analyzing human gut microbiomes with cohort study data**

We developed an algorithm for investigating the metagenomics of 12 human gut microbiomes assembled from extremely long reads of ~9kb average length by single-molecule real-time (SMRT) sequencing. We successfully identified 101 near-complete microbial genomes, and 92 complete circular sequences of putative mobile genetic elements (MGEs), including plasmids and phages, which we call the “metamobilome.” MGEs were highly prevalent or varied between human gut microbiomes from different nations. Using methylation-sensitive SMRT sequencing, we directly determined the “metamethylome” in microbiomes and identified 503 methylation motifs, many of which were unknown. We found considerable inter-individual variations in metamethylome and metamobilome.

(4) **Fast and low-cost clinical sequencing of structural variants**

We have been attempting to improve our method of detecting short tandem repeat expansions (STRs), an important class of structural variants associated with brain diseases, in a given DNA samples in terms of both time efficiency and cost efficiency. We first approximate the positions of STRs using low-cost short read sequencing, and amplify the focal regions using PCR, sequence bar-coded amplicons using PacBio Sequel, and assemble long reads.

III. 成果の外部への発表

(1) 学会誌・雑誌等における論文一覧（国内誌 0 件、国際誌 16 件）

1. Suzuki Y, Korlach J, Turner SW, Tsukahara T, Taniguchi J, Qu W, Ichikawa K, Yoshimura J, Yurino H, Takahashi Y, Mitsui J, Ishiura H, Tsuji S, Takeda H, Morishita S. AgIn: measuring the landscape of CpG methylation of individual repetitive elements. *Bioinformatics*. 2016. 32(19), 2911-9.
2. Qu W, Tsukahara T, Nakamura R, Yurino H, Hashimoto S, Tsuji S, Takeda H, Morishita S. Assessing Cell-to-Cell DNA Methylation Variability on Individual Long Reads. *Scientific Reports*. 2016. 6:21317.
3. Ando M, Kawazu M, Ueno T, Koinuma D, Ando K, Koya J, Kataoka K, Yasuda T, Yamaguchi H, Fukumura K, Yamato A, Soda M, Sai E, Yamashita Y, Asakage T, Miyazaki Y, Kurokawa M, Miyazono K, Nimer SD, Yamasoba T, Mano H. Mutational Landscape and Antiproliferative Functions of ELF Transcription Factors in Human Cancer. *Cancer Res*. 2016. 76(7), 1814-24.
4. Fukumura K, Kawazu M, Kojima S, Ueno T, Sai E, Soda M, Ueda H, Yasuda T, Yamaguchi H, Lee J, Shishido-Hara Y, Sasaki A, Shirahata M, Mishima K, Ichimura K, Mukasa A,

- Narita Y, Saito N, Aburatani H, Nishikawa R, Nagane M, Mano H. Genomic characterization of primary central nervous system lymphoma. *Acta Neuropathol*. 2016. 131(6), 865-75.
5. Katsura A, Suzuki HI, Ueno T, Mihira H, Yamazaki T, Yasuda T, Watabe T, Mano H, Yamada Y, Miyazono K. MicroRNA-31 is a positive modulator of endothelial-mesenchymal transition and associated secretory phenotype induced by TGF-beta. *Genes Cells*. 2016. 21(1), 99-116.
 6. Yasuda T, Tsuzuki S, Kawazu M, Hayakawa F, Kojima S, Ueno T, Imoto N, Kohsaka S, Kunita A, Doi K, Sakura T, Yujiri T, Kondo E, Fujimaki K, Ueda Y, Aoyama Y, Ohtake S, Takita J, Sai E, Taniwaki M, Kurokawa M, Morishita S, Fukayama M, Kiyoi H, Miyazaki Y, Naoe T, Mano H. Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. *Nat Genet*. 2016. 48(5), 569-74.
 7. Nakamura T, Yamashita S, Fukumura K, Nakabayashi J, Tanaka K, Tamura K, Tateishi K, Kinoshita M, Fukushima S, Takami H, Fukuoka K, Yamazaki K, Matsushita Y, Ohno M, Miyakita Y, Shibui S, Kubo A, Shuto T, Kocialkowski S, Yamanaka S, Mukasa A, Sasayama T, Mishima K, Maehara T, Kawahara N, Nagane M, Narita Y, Mano H, Ushijima T, Ichimura K. Genome-wide DNA methylation profiling identifies primary central nervous system lymphoma as a distinct entity different from systemic diffuse large B-cell lymphoma. *Acta Neuropathol*. 2017. 133(2), 321-4.
 8. Nakamura R, Sone J, Atsuta N, Tohnai G, Watanabe H, Yokoi D, Nakatochi M, Watanabe H, Ito M, Senda J, Katsuno M, Tanaka F, Li Y, Izumi Y, Morita M, Taniguchi A, Kano O, Oda M, Kuwabara S, Abe K, Aiba I, Okamoto K, Mizoguchi K, Hasegawa K, Aoki M, Hattori N, Tsuji S, Nakashima K, Kaji R and Sobue G. Next-generation sequencing of 28 ALS-related genes in a Japanese ALS cohort. *Neurobiol Aging*. 2016. 39, 219 e1-8.
 9. Williams K, Topp S, Yang S, Smith B, Fifita J, Warraich S, Zhang K, Farrarwell N, Vance C, Hu X, Chesi A, Leblond C, Sundaramoorthy V, Dobson-Stone C, Lee A, Rayner S, Molloy M, van Blitterswijk M, Dickson D, Petersen R, Graff-Radford N, Boeve B, Murray M, Pottier C, Don E, Winnick C, McCann E, Hogan A, Daoud H, Levert A, Dion P, Mitsui J, Ishiura H, Takahashi Y, Goto J, Kost J, Gellera C, Gkazi AS, Miller J, Stockton J, Brooks W, Boundy K, Polak M, Muñoz-Blanco JL, Esteban-Pérez J, Rabano A, Hardiman O, Morrison K, Ticozzi N, Silani V, de Bellerocche J, Glass J, Kwok J, Guillemin G, Chung R, Tsuji S, Brown R, García-Redondo A, Rademakers R, Landers J, Gitler A, Rouleau G, Cole N, Yerbury J, Atkin J, Shaw C, Nicholson G, and Blair I. C9orf72 mutations in amyotrophic lateral sclerosis and frontotemporal dementia *Nature Commun*. 2016. 7, 11253
 10. Higuchi Y, Hashiguchi A, Yuan J, Yoshimura A, Mitsui J, Ishiura H, Tanaka M, Ishihara S, Tanabe H, Nozuma S, Okamoto Y, Matsuura E, Ohkubo R, Inamizu S, Shiraishi W, Yamasaki R, Ohyagi Y, Kira J, Oya Y, Yabe H, Nishikawa N, Tobisawa S, Matsuda N, Masuda M, Kugimoto C, Fukushima K, Yano S, Yoshimura J, Doi K, Nakagawa M, Morishita S, Tsuji S, and Takashima H. Mutations in MME cause an autosomal-recessive Charcot-Marie-Tooth disease type 2. *Ann Neurol* 2016. 79, 659-72, 2016.
 11. Mitsui J, Ishiura H, Yoshimura J, Doi K, Morishita S, Shoji H, Arimori Y, Matsumoto T,

- Shimizu J and Tsuji S. Novel COL6A2 mutation in a case of limb girdle muscular dystrophy phenotype with autosomal recessive inheritance. *Neurol. CLin. NeuroSci.* 2016. 4, 189–191, 2016.
12. Aihara K, Mukasa A, Nagae G, Nomura M, Yamamoto S, Ueda H, Tatsuno K, Shibahara J, Takahashi M, Momose T, Tanaka S, Takayanagi S, Yanagisawa S, Nejo T, Takahashi S, Omata M, Otani R, Saito K, Narita Y, Nagane M, Nishikawa R, Ueki K, Aburatani H, Saito N. Genetic and epigenetic stability of oligodendrogliomas at recurrence. *Acta Neuropathol Commun.* 2017. 5(1):18.
 13. Kondo A, Yamamoto S, Nakaki R, Shimamura T, Hamakubo T, Sakai J, Kodama T, Yoshida T, Aburatani H, Osawa T. Extracellular Acidic pH Activates the Sterol Regulatory Element-Binding Protein 2 to Promote Tumor Progression. *Cell Rep.* 2016. 18(9):2228-2242.
 14. duVerle DA, Yotsukura S, Nomura S, Aburatani H, Tsuda K. CellTree: an R/bioconductor package to infer the hierarchical structure of cell populations from single-cell RNA-seq data. *BMC Bioinformatics.* 2016. 17(1):363.
 15. Tsuji S, Midorikawa Y, Seki M, Takayama T, Sugiyama Y, Aburatani H. Network-based analysis for identification of candidate genes for colorectal cancer progression. *Biochem Biophys Res Commun.* 2016. 476(4):534-40.
 16. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, Tanaka H, Taniguchi H, Kawakami Y, Ueno M, Gotoh K, Ariizumi SI, Wardell CP, Hayami S, Nakamura T, Aikata H, Arihiro K, Boroevich KA, Abe T, Nakano K, Maejima K, Sasaki-Oku A, Ohsawa A, Shibuya T, Nakamura H, Hama N, Hosoda F, Arai Y, Ohashi S, Urushidate T, Nagae G, Yamamoto S, Ueda H, Tatsuno K, Ojima H, Hiraoka N, Okusaka T, Kubo M, Marubashi S, Yamada T, Hirano S, Yamamoto M, Ohdan H, Shimada K, Ishikawa O, Yamaue H, Chayama K, Miyano S, Aburatani H, Shibata T, Nakagawa H. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet.* 2016. 48(5):500-9.

(2) 学会・シンポジウム等における口頭・ポスター発表

1. Y. Suzuki, S. Nishijima, Y. Furuta, W. Suda, K. Oshima, J. Taniguchi, J. Yoshimura, M. Hattori and S. Morishita. Meta-methylome analysis with SMRT sequencing revealed a diversity of DNA methylation motifs in uncultured human gut microbiomes. The Biology of Genomes, *Cold Spring Harbor Laboratory Meeting*, New York. May 2016. (ポスター)
2. Y. Suzuki, S. Nishijima, Y. Furuta, W. Suda, K. Oshima, M. Hattori and S. Morishita. Metagenome, metamobilome and meta-methylome of human gut flora uncovered by SMRT sequencing. *Advances in Genome Biology and Technology General Meeting*, Hollywood, Florida. February 2017. (口頭)
3. W. Qu, H. Ishiura, J. Mitsui, K. Doi, J. Yoshimura, D. Hsu, M. Boitano, J. Korlach, S. Tsuji, S. Morishita. SMRT Sequencing of barcoded amplicons with large short tandem repeat expansions. *Advances in Genome Biology and Technology General Meeting*, Hollywood,

Florida. February 2017. (ポスター)

4. Y. Suzuki and S. Morishita. Personal diploid methylomes via analysis of phased heterozygous variants and single-molecule real-time sequencing. ***Advances in Genome Biology and Technology General Meeting***, Hollywood, Florida. February 2017. (ポスター)

(3) 「国民との科学・技術対話社会」に対する取り組み

シンポジウム等での口頭発表は無かったが、解説記事の執筆を通じてアウトリーチ活動を行った。

1. PacBio 1 分子実時間シーケンシング ゲノム, トランスクリプトーム, エピゲノム解析への応用、森下真一、実験医学増刊 「RNA-Seq 実験ハンドブック」、2016、236～241 頁
2. 個人ゲノム解析のためのゲノムインフォマティクス、森下真一、遺伝子医学 MOOK 別冊「最新精神・神経遺伝医学研究と遺伝カウンセリング」、51～60 頁
3. PacBio ロングリードを用いたメタゲノム解析、鈴木慶彦・森下真一、実験医学別冊 NGS アプリケーション「今すぐ始める！ メタゲノム解析 実験プロトコール」、2016、149～155 頁

(4) 特許出願

無し