

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) 革新的先端研究開発支援事業
(英語) Advanced Research and Development Programs for Medical Innovation
- 研究開発課題名： (日本語) 遺伝統計学が紐解く微生物叢・宿主・疾患・創薬のクロストーク
(英語) Crosstalk among microbiome, host, disease, and drug discovery enhanced by statistical genetics
- 研究開発担当者 (日本語) 国立大学法人大阪大学大学院医学系研究科 教授 岡田随象
所属 役職 氏名： (英語) Osaka University Graduate School of Medicine, Professor, Yukinori Okada
- 実施期間： 平成28年10月 1日 ～ 平成29年 3月31日
- 分担研究 (日本語) 遺伝統計学が紐解く微生物叢・宿主・疾患・創薬のクロストーク
開発課題名： (英語) Crosstalk among microbiome, host, disease, and drug discovery enhanced by statistical genetics
- 研究開発分担者 (日本語) 国立大学法人大阪大学大学院医学系研究科 教授 岡田随象
所属 役職 氏名： (英語) Osaka University Graduate School of Medicine, Professor, Yukinori Okada

II. 成果の概要（総括研究報告）

・ 研究開発代表者による報告の場合

平成 28 年度における研究成果を、本研究開発課題を構成する 4 テーマに沿って記述する。

1. メタゲノムデータ機械学習

疾患罹患者および対照群より得られた既存のメタゲノムデータに対する機械学習シミュレーションを実施した。PCA に代表される既存の機械学習手法は、外れ値を取るデータ単位の同定に有用である一方、比較的類似度の近いデータ単位間の峻別に対する検出力が欠ける傾向が認められたため、次元削減手法の適用を検討している。

2. メタゲノムショットガンシーケンスデータの網羅的解析

既存メタゲノムサンプルを対象に、大阪大学微生物病学研究所に新規設置された次世代シーケンサーを用いて、ショットガンシーケンスを実施した。metagenome-wide association study の実施に向けたデータ解析を準備中である。

3. LDSC による大規模疾患ゲノム解析結果の組織特異性の検討

Linkage Disequilibrium Score regression (LDSC) とは、複数の疾患における GWAS 結果を横断的に比較し、遺伝的疾患リスクの共有や組織特異性を検討する手法である。既存の LDSC 解析は欧米人集団を中心に実施されている点考慮し、日本人集団における GWAS 結果とあわせて LDSC が実施可能とするために、人種間の連鎖不平衡関係 (Linkage Disequilibrium) の違いを加味した LDSC のオプションの実装を行った。

4. HLA imputation 法および MIGWAS によるバイオマーカー同定

HLA 遺伝子多型をコンピューター上で高精度に決定し追加費用無しでの網羅的解析を可能にする「HLA imputation 法」に基づく HLA 遺伝子多型解析を実施した。免疫疾患に対する HLA imputation 解析を国際共同研究として実施し、疾患リスクを有する HLA 遺伝子および多型を同定した (論文投稿準備中)。

We conducted the following research activities.

1. Machine learning analysis using metagenome data

We applied the machine learning methods using metagenome data obtained from the disease-affected and healthy individuals. We found that some machine learning methods such as principal component analysis showed high performance to identify outliers, but not for discrimination of related clusters. We are thus considering to additionally apply the dimension reduction methods.

2. Metagenome shotgun sequence data analysis

We conducted metagenome shotgun sequencing using next generation sequencer installed at

Research Institute for Microbial Disease, Osaka University. Subsequent *in silico* analysis for metagenome-wide association study is currently prepared.

3. Cell type-specific heritability analysis of large-scale human genome data using LDSC

Linkage Disequilibrium Score regression (LDSC) is a method that evaluates shared genetic heritability and cell type specificity among the human complex traits, by utilizing the summary statistics of large-scale genome-wide association study (GWAS) data. Since the previous LDLC analysis mainly focused on the data from Europeans, we expanded the target of LDLC to additionally include other populations, and applied to the GWAS data of Japanese.

4. HLA imputation and MIGWAS analysis

The HLA imputation method computationally estimates HLA alleles from the GWAS genotype data without additional costs. By applying HLA imputation, we identified the risk HLA genotype that has a disease risk (manuscript in preparation).

- ・ 研究開発分担者による報告の場合

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 件、国際誌 件）

1. Kanai M, Tanaka T, Okada Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J Hum Genet* 2016, 61, 861-866.
2. Jing J, Pattaro C, Hoppmann A, Okada Y, CKDGen Consortium, Fox CS, Köttgen A. Combination of mouse models and genomewide association studies highlights novel genes associated with human kidney function. *Kidney Int* 2016, 90, 764-773.
3. van Rooij FJ, Qayyum R, Smith AV, Zhou Y, Trompet S, Tanaka T, Keller MF, Chang LC, Schmidt H, Yang ML, Chen MH, Hayes J, Johnson AD, Yanek LR, Mueller C, Lange L, Floyd JS, Ghanbari M, Zonderman AB, Jukema JW, Hofman A, van Duijn CM, Desch KC, Saba Y, Ozel AB, Snively BM, Wu JY, Schmidt R, Fornage M, Klein RJ, Fox CS, Matsuda K, Kamatani N, Wild PS, Stott DJ, Ford I, Slagboom PE, Yang J, Chu AY, Lambert AJ, Uitterlinden AG, Franco OH, Hofer E, Ginsburg D, Hu B, Keating B, Schick UM, Brody JA, Li JZ, Chen Z, Zeller T, Guralnik JM, Chasman DI, Peters LL, Kubo M, Becker DM, Li J, Eiriksdottir G, Rotter JI, Levy D, Grossmann V, Patel KV, Chen CH; BioBank Japan Project., Ridker PM, Tang H, Launer LJ, Rice KM, Li-Gao R, Ferrucci L, Evans MK, Choudhuri A, Trompouki E, Abraham BJ, Yang S, Takahashi A, Kamatani Y, Kooperberg C, Harris TB, Jee SH, Coresh J, Tsai FJ, Longo DL, Chen YT, Felix JF, Yang Q, Psaty BM, Boerwinkle E, Becker LC, Mook-Kanamori DO, Wilson JG, Gudnason V, O'Donnell CJ,

- Dehghan A, Cupples LA, Nalls MA, Morris AP, Okada Y, Reiner AP, Zon LI, Ganesh SK. Genome-wide Trans-ethnic Meta-analysis Identifies Seven Genetic Loci Influencing Erythrocyte Traits and a Role for RBPMS in Erythropoiesis. *Am J Hum Genet* 2017, 100, 51-63.
4. Molineros JE, Yang W, Zhou XJ, Sun C, Okada Y, Zhang H, Chua KH, Lau YL, Kochi Y, Suzuki A, Yamamoto K, Ma J, Bang SY, Lee HS, Kim K, Bae SC, Zhang H, Shen N, Looger LL, Nath SK. Confirmation of five novel susceptibility loci for Systemic Lupus Erythematosus (SLE) and integrated network analysis of 82 SLE susceptibility loci. *Hum Mol Genet* 2017, 26, 1205-1216.
 5. 岡田随象 ゲノム創薬 アレルギー 2016, 65, 1231-1232.
 6. 岡田随象 自己免疫疾患における大規模ヒトゲノム解析の現状 アレルギー 2016, 65, 921-925
 7. 坂上 沙央里、岡田随象 遺伝統計学による創薬 腎臓内科・泌尿器科 2016, 5, 82-86.
 8. 岸川 敏博、岡田随象 ゲノム医療時代のバイオインフォマティクス 医学のあゆみ 2016, 260, 332-334.
 9. 岡田随象 概論－「ゲノム情報」＋「遺伝統計学」＝「ゲノム医療」！？ 実験医学 2016, 34, 2630-2635.

(2) 学会・シンポジウム等における口頭・ポスター発表

1. Dosage Contribution of a Non-Classical HLA Gene, HLA-DOA, to the Risk of Rheumatoid Arthritis, 口頭, Yukinori Okada, 2016 ACR/ARHP Annual Meeting, 2016/11/14, アメリカ合衆国.

(3) 「国民との科学・技術対話社会」に対する取り組み

1. Statistical genetics, disease biology, and drug discovery, 口頭, 岡田随象, 2017 International Symposium on Joint Research Project of the Institute of IMS-UT, 2017/3/21, 国内.
2. Statistical genetics, disease biology, and drug discovery, 口頭, 岡田随象, Genome Evolution at Mishima 2017, 2017/3/28, 国内.

(4) 特許出願

(該当無し)