

令和 3 年 4 月 30 日現在

機関番号：14301

研究種目：基盤研究(B)（特設分野研究）

研究期間：2018～2020

課題番号：18KT0015

研究課題名（和文）妊娠期間中の時系列多層情報解析による疾病前兆や病態進行の予測技術開発と適用

研究課題名（英文）Time-series multi-layered omics analysis during gestation period towards precision medicine

研究代表者

長崎 正朗（Nagasaki, Masao）

京都大学・スーパーグローバルコース医学生命系ユニット・特定教授

研究者番号：90396862

交付決定額（研究期間全体）：（直接経費） 14,100,000円

研究成果の概要（和文）：妊娠期間中の日本人の妊婦の方から得られた、時系列多層オミクスプロファイリングの情報に対し、経時的な変化を仮定した数理モデルを適用することで、妊娠糖尿病、妊娠高血圧症、高血圧（妊娠高血圧除く）、児の性別、経産婦かどうかに関連するマーカー候補の探索を進めた。さらに、これらのマーカーに関連する生体内の機能群についても同定をおこない、オミクスプロファイリングから形質の予測を行うこともできた。

研究成果の学術的意義や社会的意義

本研究において、妊娠期間中の児の性別、経産婦かどうかなどの形質、妊娠糖尿病、妊娠高血圧症、高血圧（妊娠高血圧除く）などの疾患に関連する遺伝子群および、関連する機能グループを同定することができた。さらに、一部非侵襲の情報からの予測においても一定の成果が得られている。国内外での独立した調査研究などにより、本研究で得られた候補が検証されることで、原因の解明に近づくとともに、新しい早期診断法・治療法の開発が進んでいくと考えている。

研究成果の概要（英文）：To our time-series multilayered omics profiling information (metabolome and transcriptome) from Japanese pregnant women during pregnancy, we have developed a mathematical model that can explicitly model the time-course expression data. Using this model, we detected many candidate markers associated with gestational diabetes, gestational hypertension, hypertension (excluding hypertension of pregnancy), gender of the child, and whether multipara or not. We have also identified enrichment biological groups among these markers. Additionally, part of the phenotype could be estimated from the omics profiling information.

研究分野：バイオインフォマティクス

キーワード：ライフログ オミクス メタボローム 妊娠期間 出産 統合解析 ロジスティック回帰

## 様式 C-19、F-19-1、Z-19（共通）

### 1. 研究開始当初の背景

妊娠期間中は、生体内の環境がめまぐるしく変化する。そのため、生体内が非妊娠期間中と異なる状態となることでさまざまな疾患が発症してくる。例えば、妊娠高血圧症候群はいまだ原因不明の多因子疾患であり、母児への短期・長期的なリスクから、次世代の育成にかかる重要な健康問題となっている。年間約 100 万人の児が日本では産まれるがそのうち約 5 %以上の妊婦が罹患する。近年、次世代シーケンサを中心とした計測技術の向上、情報科学の進展により、本疾患に対して遺伝因子・環境因子を統合的に解析することで、原因の解明と共に新しい早期診断法・治療法の開発が望まれている。

### 2. 研究の目的

全ゲノム情報と時系列多層情報（トランスクリプトーム、メタボローム、臨床情報などの情報）に対して多変量解析などの統計手法を用いて妊娠高血圧症候群などへのリスク因子の同定を行う。さまざまな層における個別の予測マーカーだけではなく、いままで一見関係性がないと思われる複雑な事象の関連性をつかむことで、妊娠期間中における複合予測マーカーを同定することを目的とする。さらに、侵襲的な層（血液、血漿などから得られる情報）の予測マーカーをライフログや尿などの非侵襲的な層（ライフログ、唾液などから得られる情報）のマーカーに、機械学習手法を適用することで置き換えていくことで、実社会で役に立つ簡便で利用価値の高い複合マーカーを探索することを目指す。これにより、将来的に妊娠期間中の疾病発症及び病態進行の予測を個別に行うことでリスクの軽減を行う。そのため以下の研究を目的として開発を進めた。

【問い 1】妊娠期間中の妊婦の方にライフログを含む時系列多層オミクスプロファイリングの情報を用いることで妊娠高血圧症候群などの予測がどの程度可能であるのか？そのためには、どの情報解析手法がより有効であるか？

【問い 2】どの程度侵襲的な情報を非侵襲的な情報で代替をすることが可能か？また、そのためには、どの情報解析手法がより有効であるか？

### 3. 研究の方法

国立大学法人東北大学（研究代表者：長崎 正朗）と株式会社NTTドコモとの間で平成 27 年 9 月 14 日より「マタニティログ調査」として本試験を開始し、平成 28 年 11 月 15 日をもって研究参加される妊婦の募集を完了した情報を用いて解析を行った。妊婦被験者同意獲得総数は 314 名である。平成 29 年度末に、同意撤回対象者を除くすべての全ゲノム情報、臨床情報（病歴、臨床検査情報など）、DNA 由来のトランスクリプトーム情報、尿由来の NMR およびメタボローム情報がそろっておりそれらの情報を解析対象として研究を行った（Sugawara et al PMJ Open 2019）。

妊娠期間中 2 回、および出産後に 1 回（合計 3 回）に渡って繰り返し行われた各妊婦の採血の影響をモデルに組み込むために、R の nlme パッケージに含まれる lme 関数を用いて、線形混合（linear mixed-effects）モデルによる回帰計算を実施した。出産前後で遺伝子発現量の変化に差異があると考えられるため、遺伝子発現量の経過日数における変化を出産前後の二つの回帰係数に分けてモデル化した（図 1）。二つの単回帰係数を持つ回帰式の計算を結合するために、定数項を出産時の値とし出産前後で共通とした。本回帰モデルにより、出産前後の発現量データについて定数項を通して関連付けることができ、共通の回帰式で表すことができるため、出産前後のデータを別々に解析するよりも統計的に有利となり検出力が向上した。疾患等の情報はダミー変数として回帰式に組み込み、回帰係数の統計的有意性によって関連遺伝子の候補を抽出した（妊娠経過日数（ProgressDay）に対してのみ、出産前変化  $\beta_1$  を利用）。着目している疾患等以外の共変量を取り除き解析の感度を高めるために、疾患等と相関係数の低い PEER factor（O Stegle et al PLoS Comput. Bio. 2010）の影響を発現量から除去した後に回帰分析を行った。相関係数の閾値はパラメータとし、0.01 刻みで変動させたときの全ての計算結果を集計対象にした。得られた結果から、遺伝子毎に P-value の一番小さな結果を選択し、遺伝子がユニークに登録されているファイルを作成した。

この処理に伴い、FDR の補正を行った。

図1 線形混合 (linear mixed-effects) モデルの概要

全体の回帰式 (係数AとBで時点を切り替え):

$$y_i = \beta_0 + \beta_d[\text{disease status}] + A [\beta_1 (t_i - 280)] + B [\beta_2 (t_i - 280)] + r_i + r_j$$

時点1st:

$$y_i = \beta_0 + \beta_d[\text{disease status}] + \beta_1 (t_i - 280) + r_i + r_j$$

時点2nd:

$$y_i = \beta_0 + \beta_d[\text{disease status}] + \beta_1 (t_i - 280) + r_i + r_j$$

時点after:

$$y_i = \beta_0 + \beta_d[\text{disease status}] + \beta_2 (t_i - 280) + r_i + r_j$$

- ✓  $\beta_0, \beta_d, \text{disease status}, r_i, r_j$  は共通
- ✓  $\beta_1$  は時点1stと2ndで共通

- i: 検体
- j: 妊婦 (1...301)
- $y_i$ : 遺伝子発現量
- $t_i$ : 日数
- $\beta_0$ :  $t_i = 280$ 日の切片
- $\beta_d$ : 疾患等のダミーパラメータ
- disease status: 0 or 1
- $\beta_1$ : 出産前の変化
- $\beta_2$ : 出産後の変化
- $r_i$ : 残差
- $r_j$ : 変量効果 (random effects)

時点間のデータを関連付けることにより統計量が増加

さらに、発現量に基づく各形質および疾患の分類を RFECV (交差検証) と logisticRegression を用いて学習するとともに、10-fold cross validation により同性能の評価を行った。

#### 4. 研究成果

経産婦かどうか、児の性別、妊娠高血圧症などについて同様の解析を行った。そのうち、特に以下では、経産婦かどうかにおいて得られる結果を具体例として示す。表1に関連するとして同定された遺伝子群。また、図2に妊娠期間中の発現のプロファイルと PEER 補正による効果を示す。

表1 経産婦かどうかに対して有意に検出された上位の遺伝子

gene_name	correlation	BirthTimes _Category	pval_BirthTimes _Category	fdr_BirthTimes _Category	b1
IFNLR1	0.07	-0.42188	5.84E-14	2.66E-08	-0.00083
IGHG3	0.05	-0.50494	1.34E-13	2.66E-08	0.00054
SEMA4B	0.08	-0.17133	6.01E-13	4.61E-08	-0.00063
SIGLEC6	0.07	-0.41045	1.26E-12	6.71E-08	-0.00153
ZBTB32	0.05	-0.3667	4.23E-11	8.80E-07	-0.00043
KIAA0232	0.07	0.095324	5.19E-11	1.03E-06	0.000404
PLAUR	0.08	-0.15724	1.59E-10	2.63E-06	-0.00068
DCHS1	0.09	0.320208	5.75E-10	7.05E-06	0.000916
TNFRSF13B	0.08	-0.37282	8.24E-10	9.38E-06	-0.00076
TIGIT	0.09	-0.21932	8.88E-10	9.88E-06	-0.00077
CBLB	0.1	-0.1134	1.30E-09	1.42E-05	-0.00043
BHLHE41	0.07	-0.38825	1.51E-09	1.60E-05	-0.00051
MBD6	0.07	0.093094	2.60E-09	2.27E-05	0.000408
JAK1	0.08	0.162689	2.88E-09	2.46E-05	0.000612
RP1-148H17.1	0.03	-0.43529	4.53E-09	3.44E-05	0.000678
MT-ND4	0.08	0.161009	4.97E-09	3.60E-05	0.000706
MFG8	0.1	0.175089	5.15E-09	3.68E-05	0.00064

SOWAHD	0.06	-0.2019	7.26E-09	4.82E-05	-0.00068
CADM4	0.08	-0.21855	9.19E-09	5.71E-05	-0.00095
PPCDC	0.06	-0.19234	1.23E-08	7.03E-05	-0.00065

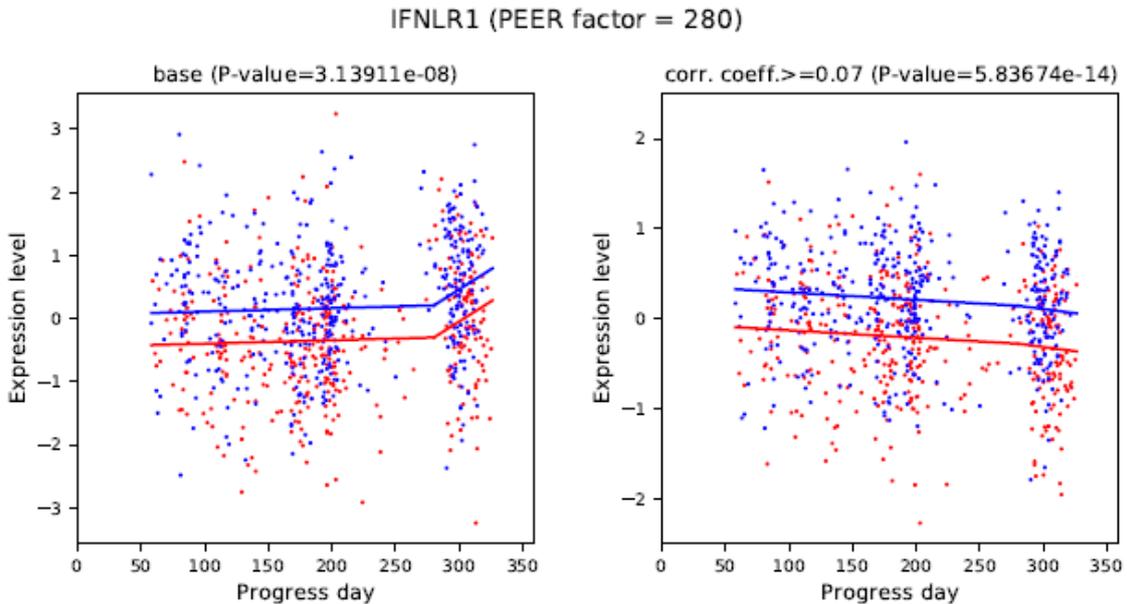


図2 経産婦かどうかに対して有意に検出された上位の遺伝子における発現量の PEER factor による補正の効果。ここでは、IFNLR1 を表示している。(左図：補正前、右図：補正後)

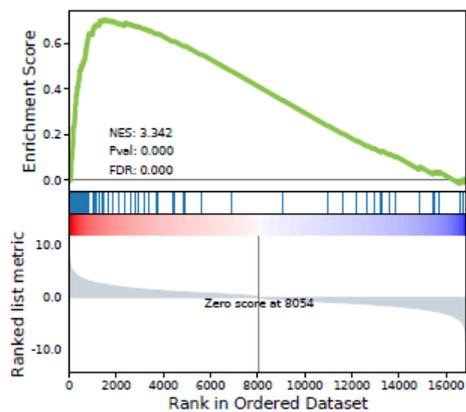
同様に、妊娠糖尿病、妊娠高血圧症、高血圧（妊娠高血圧除く）、児の性別についても同様の解析を行うことで、関連する遺伝子の同定をすることができた。さらに、得られた遺伝マーカーを使うことでオントロジー解析を行い関連する遺伝子集団の特性について解析をおこなうことで、各疾患に関連する候補パスウェイを抽出することができた。以下に、経産婦かどうかに関連するオントロジー解析の結果を表2および図3に例として示す。他の形質についても整理を行い論文として準備を進めている（Nagasaki et al in preparation）。他に、尿中の時系列のメタボローム解析による妊娠期間の予測などについても進め成果を得た（Yamauchi et al in preparation）。

表2 経産婦かどうかに対して有意に検出された上位の GO

Term	es	nes	pval	fdr
COTRANSLATIONAL_PROTEIN_TARGETING_T O_MEMBRANE	0.70689	3.342496	0	0
CYTOPLASMIC_TRANSLATION	0.569632	2.699029	0	0
CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT	0.782138	3.295097	0	0
CYTOSOLIC_RIBOSOME	0.732507	3.44945	0	0
CYTOSOLIC_SMALL_RIBOSOMAL_SUBUNIT	0.71819	2.90053	0	0
DNA_DEPENDENT_DNA_REPLICATION	-0.49244	-2.43556	0	0
ESTABLISHMENT_OF_PROTEIN_LOCALIZATIO	0.639562	3.059727	0	0

N_TO_ENDOPLASMIC_RETICULUM				
NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS	0.524989	2.792817	0	0
NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY	0.682948	3.321222	0	0
POLYSOMAL_RIBOSOME	0.707348	2.606263	0	0
POLYSOME	0.582645	2.553634	0	0
PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	0.568433	2.828479	0	0
PROTEIN_TARGETING_TO_MEMBRANE	0.541066	2.782145	0	0
RIBOSOMAL_SUBUNIT	0.484517	2.51455	0	0
STRUCTURAL_CONSTITUENT_OF_RIBOSOME	0.520529	2.673789	0	0
TRANSLATIONAL_INITIATION	0.596774	3.120694	0	0
VIRAL_GENE_EXPRESSION	0.529714	2.783519	0	0
CELL_SUBSTRATE_JUNCTION	0.421594	2.407761	0	7.30E-05
ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_MEMBRANE	0.431338	2.416558	0	7.73E-05
LARGE_RIBOSOMAL_SUBUNIT	0.49368	2.33802	0	0.000484

**COTRANSLATIONAL PROTEIN TARGETING TO MEMBRANE**



**CYTOPLASMIC TRANSLATION**

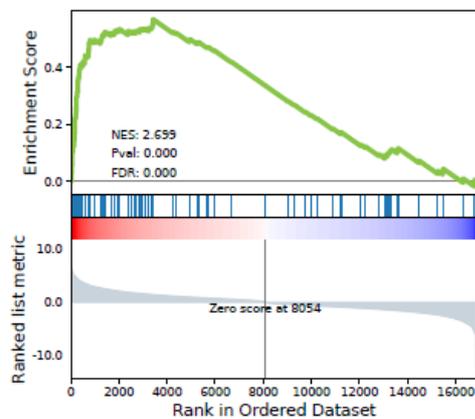


図3 経産婦かどうかに対して有意に検出された上位の GO における遺伝子の分布（横軸）と集積スコア（enrichment scores：縦軸）

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Sugawara J, Ochi D, Yamashita R, Yamauchi T, Saigusa D, Wagata M, Obara T, Ishikuro M, Tsunemoto Y, 18名略, Hiyama S, Nagasaki M	4. 巻 9(2)
2. 論文標題 Maternity Log study: a longitudinal lifelog monitoring and multiomics analysis for the early prediction of complicated pregnancy	5. 発行年 2019年
3. 雑誌名 BMJ Open	6. 最初と最後の頁 e025939
掲載論文のDOI（デジタルオブジェクト識別子） 10.1136/bmjopen-2018-025939	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計2件（うち招待講演 2件/うち国際学会 2件）

1. 発表者名 Masao Nagasaki
2. 発表標題 Genome cohort projects in Japan and biomedical analyses toward the discoveries of disease-causing variants
3. 学会等名 Genome Expo 2019（招待講演）（国際学会）
4. 発表年 2019年

1. 発表者名 Masao Nagasaki
2. 発表標題 Development and application of biomedical methods to facilitate the detection of disease-causing variants based on whole genome data from Japanese
3. 学会等名 The 3rd JCS Council Forum on Basic CardioVascular Research(BCVR 2019)（招待講演）（国際学会）
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------