

令和 3 年 5 月 17 日現在

機関番号：14301

研究種目：基盤研究(B)（一般）

研究期間：2018～2020

課題番号：18H03326

研究課題名（和文）妊娠期間中の遺伝子発現変動と遺伝要因との関連性の探索によるデータベース構築

研究課題名（英文）Construction of eQTL database during pregnant period

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） 13,300,000円

研究成果の概要（和文）：日本人妊婦約300名の全ゲノム情報と妊娠期間、前期、中期、出産後の3点のRNAシーケンス情報を情報解析し、RNAの各時点および、時点間の変化のプロファイルを構築するとともに、同一妊婦の全ゲノムシーケンス情報を用い、これらのRNAプロファイルと相関する遺伝要因の探索を行った。また、これらの情報について整理を行い重要なRNAと個人の変異情報との関連のデータベースを構築した。

研究成果の学術的意義や社会的意義

妊娠期間中における妊娠期間毎の一般集団と健常妊婦における日本人のeQTLデータベースを世界で初めて構築した。当データベースは、今後得られてくるRNA転写物、プロテオームなどのマーカーについて個人毎の遺伝的な背景を検討するための貴重な情報資源の1つとなる。さらに、国内外のゲノムワイド関連解析によって得られたデータベースGWASCatalog等で報告されているSNPについて当該データベースに関連したSNPも得られていることから、妊娠期間毎の解析に限らず、日本人の遺伝要因と転写物の発現解析にも幅広く活用できる。

研究成果の概要（英文）：The study has conducted the association study (eQTL analysis) to about 300 Japanese pregnant women between more than ten millions variants from whole genome sequencing and RNA sequencing at three points (two data points in pregnancy and one point in postpartum). The gene-variants relationship was categorized into seven patterns (detected the common relationship in three points and specific in pregnancy or postpartum). Based on this result, the study also constructed an eQTL database of Japanese pregnant women. The database is especially useful for both for domestic and global researchers to understand the mechanisms from pregnancy to postpartum periods.

Parts of the gene-variants relationship in this database matched to the relationship in the diseases associated SNP catalog database, GWASCatalog. The result naturally suggests that the reported disease might also affect the Japanese women via these gene-variant cascades. Thus, the database can be used also for general periods.

研究分野：バイオインフォマティクス

キーワード：RNAシーケンシング eQTL 妊娠期間 遺伝要因探索 全ゲノムシーケンシング マタニティログ

## 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は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、次世代の成育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生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。妊娠期間中に母体における胎児の生育に関連して、特異的に発現する遺伝子、また遺伝子の RNA 発現量の変化が考えられる。これらの RNA の変化は、個人毎の遺伝型に大きく依存していると考えられるが、このような遺伝型と RNA の発現変化のプロファイルはどこにも存在しない。

### 2. 研究の目的

妊婦 300 名の妊娠期間中、前期、中期、出産後の 3 時点の RNA シーケンサ情報を解析することで【問い 1】に答えるとともに、同一妊婦の全ゲノムシーケンサ情報を用いることで【問い 1】から得られた各時点の状態、および時点間の変化と相関する遺伝要因の探索を行い【問い 2】【問い 3】に答えることを目的とする。さらに、これらの情報について整備を行うことで重要な関連についてはデータベースとして整備を行うことを目的とする。

【問い 1】 妊娠期間中の前期、中期、出産後で遺伝子の個人毎の発現状態は変化するのか？

【問い 2】 妊娠期間中の前期、中期、出産後で個人の SNP 情報と遺伝子の発現量の関係性がどの程度共通なのか？

【問い 3】 妊娠期間中の前期、中期、出産後で個人の SNP 情報と遺伝子の発現量との関係がある特有の期間のみにみられるものが存在するのか？

### 3. 研究の方法

全血由来の RNA 試料について、次世代シーケンサを用いて得られた情報に対して、featureCount を用いることで、各試料における各遺伝子の発現量の推定を行った。この得られた値について、シーケンサ毎に総量が異なること、また、グロビンが含まれることから、これらの情報を除去するとともに、TMM 正規化 (Trimmed mean of M values)を行うことで試料由来の情報の発現総量の正規化を行った。さらに、Rank based inverse normal transformation をすることで遺伝子毎の正規化を行った。これらの情報について PEER 解析を行うことで各試料の計測におけるさまざまな環境由来の外乱要素を除いた。これにより、56,202 個の転写物から、28,222 個の転写物に絞り込む準備を整えた。

さらに、RNA 試料の検体と同一の被験者の全血由来の DNA 試料に対して全ゲノムシーケンサにより得られた配列情報を国際参照配列にアライメントすることで得られた全ゲノム中の変異情報を整理することで約 3,000 万個の変異カタログを整備した。

これら転写物の発現量と変異カタログとの関連性を検討するため eQTL 解析を行った。特に、妊娠初期、後期、出産後と 3 時点における計測結果が各検体において取得されていることから、期間毎に eQTL 解析を行うことで共通した関係性、ある時期にのみ得られる関係性の比較検討を行った。特に、302 名のうち 187 名が妊娠期間中に疾患に関連する形質が見られないことから、一般集団と区別し、健常妊婦として比較解析ができるようにした。さらに、これらの解析結果をデータベースとして整備を進めた (Nagasaki et al in preparation)。

### 4. 研究成果

下表に得られた eQTL のデータベースの一般集団、健常妊婦における、3 時点の検出パターンによって 7 つの群に分類した比較解析の結果を統計情報として示す。

(a) eQTL 解析と得られた SNP-Gene の計測時期間の検出パターン

計測時期	対象の検体数 (健常集団)	群 1	群 2	群 3	群 4	群 5	群 6	群 7
1st (妊娠前期)	294 (183)	Yes	Yes	Yes	No	Yes	No	No
2nd (妊娠後期)	283 (181)	Yes	Yes	No	Yes	No	Yes	No
3rd (出産後)	267 (176)	Yes	No	Yes	Yes	No	No	Yes

(b) eQTL 解析の結果得られた SNP-Gene の関係の総数(日本人一般集団)

Transcript category	-log10(p-value) threshold	群 1	群 2	群 3	群 4	群 5	群 6	群 7
protein coding	12	220,839	695	28	21	21	0	79
	11	254,370	949	75	45	21	0	235
	10	293,467	1,469	335	57	36	4	569
	9	344,391	2,788	800	153	61	27	1,133
	8	409,808	5,461	1,367	531	141	112	2,113
	7	498,652	10,811	3,030	1,494	660	434	3,808
	6	614,488	21,882	7,808	4,360	2,091	1,497	8,211
	5	765,412	48,909	21,260	12,936	8,486	7,466	21,991
	4	952,772	117,237	59,358	41,703	45,671	40,491	80,290
	FDR 補正 で有意	1,107,962	465,759	290,962	244,514	2,035,497	1,907,279	2,016,913

(c) eQTL 解析の結果得られた SNP-Gene の関係の総数(日本人健常集団)

Transcript category	-log10(p-value) threshold	群 1	群 2	群 3	群 4	群 5	群 6	群 7
protein coding	12	110,377	265	13	1	5	1	45
	11	131,519	440	33	36	36	1	155
	10	158,250	718	148	70	54	1	584
	9	193,077	1,378	272	114	60	4	1,074
	8	235,728	2,765	570	257	153	54	2,016
	7	291,648	5,428	1,511	920	380	256	4,096
	6	370,988	11,636	4,329	2,767	1,224	1,238	8,479
	5	478,883	29,416	14,313	9,596	5,598	5,669	23,629
	4	620,554	80,773	44,077	29,662	32,593	34,083	76,459
	FDR 補正 で有意	761,316	373,952	244,460	198,338	1,652,655	1,672,386	1,767,692

本研究が当初目的としていた、妊娠期間中における妊娠期間毎の一般集団と健常妊婦における日本人の eQTL データベースを構築することができた。得られた関連のうち、国内外のゲノムワイド関連解析によって得られたデータベース GWASCatalog で報告されている SNP について当該データベースに関連した SNP も得られている。そのため、当該データベースを用いることで、今後得られてくる RNA 転写物、プロテオームなどのマーカーについて個人毎の遺伝的な背景の検討を検討するための貴重な情報資源の 1 つになることを期待している。なお、本研究の情報は主に、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 (Sugawara et al PMJ Open 2019)。

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. 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4件（うち招待講演 2件/うち国際学会 3件）

1. 発表者名 長崎 正朗
2. 発表標題 日本人集団のタンパク質発現量と 関連するゲノム遺伝子座との統合解析
3. 学会等名 日本人類遺伝学会第65回大会
4. 発表年 2020年

1. 発表者名 Masao Nagasaki
2. 発表標題 Genome-Wide mapping of 4,483 protein QTLs Of the 743 Japanese population
3. 学会等名 The American Society of Human Genetics (ASHG) 2020 Virtual Meeting (国際学会)
4. 発表年 2020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

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

〔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
---------	---------