

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [editorial.bmjopen@bmj.com](mailto:editorial.bmjopen@bmj.com)

# BMJ Open

## Maternity Log Study: protocol for a longitudinal lifelog monitoring and multi-omics analysis for the early prediction of complicated Pregnancy

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-025939
Article Type:	Cohort profile
Date Submitted by the Author:	11-Aug-2018
Complete List of Authors:	<p>Sugawara, Junichi; Tohoku Medical Megabank Organization, Tohoku University, Feto-Maternal Medical Science; Tohoku University Graduate School of Medicine, Obstetrics and Gynecology  Ochi, Daisuke; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC.  Yamashita, Riu; Tohoku Medical Megabank Organization, Tohoku University  Yamauchi, Takafumi; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC.  Saigusa, Daisuke; Tohoku Medical Megabank Organization, Tohoku University  Wagata, Maiko; Tohoku Medical Megabank Organization, Tohoku University, Feto-Maternal Medical Science; Tohoku University Graduate School of Medicine, Obstetrics and Gynecology  Ishikuro, Mami; Tohoku Medical Megabank Organization, Tohoku University  Tsunemoto, Yoshiaki; Research Laboratories, NTT DOCOMO, INC.  Harada, Yuki; Tohoku Medical Megabank Organization, Tohoku University  Shibata, Tomoko; Tohoku Medical Megabank Organization, Tohoku University  Kawashima, Junko; Tohoku Medical Megabank Organization, Tohoku University  Katsuoka, Fumiki; Tohoku Medical Megabank Organization, Tohoku University  Igarashi-Takai, Takako ; Tohoku Medical Megabank Organization, Tohoku University  Ogishima, Soichi; Tohoku Medical Megabank Organization, Tohoku University  Metoki, Hirohito; Tohoku Medical and Pharmaceutical University  Hashizume, Hiroaki; Tohoku Medical Megabank Organization, Tohoku University  Fuse, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine  Minegishi, Naoko; Tohoku Medical Megabank Organization, Tohoku University  Koshiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University  Tanabe, Osamu; Tohoku Medical Megabank Organization, Tohoku University  Kuriyama, Shinichi; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine  Kinoshita, Kengo; Tohoku Medical Megabank Organization, Tohoku University</p>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

	Kure, Shigeo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Yaegashi, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Yamamoto, Masayuki; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Hiyama, Satoshi; Research Laboratories, NTT DOCOMO, INC. Nagasaki, Masao; Tohoku Medical Megabank Organization, Tohoku University
Keywords:	lifelog, multi-omics analysis, prediction, complicated pregnancy

For peer review only

SCHOLARONE™  
Manuscripts

1 **Maternity Log Study: protocol for a longitudinal lifelog**  
2 **monitoring and multi-omics analysis for the early**  
3 **prediction of complicated pregnancy**

4  
5 Junichi Sugawara<sup>1,2,3\*</sup>, Daisuke Ochi<sup>1,4</sup>, Riu Yamashita<sup>1</sup>, Takafumi Yamauchi<sup>1,4</sup>,  
6 Daisuke Saigusa<sup>1</sup>, Maiko Wagata<sup>1,2</sup>, Mami Ishikuro<sup>1</sup>, Yoshiki Tsunemoto<sup>4</sup>, Yuki  
7 Harada<sup>1</sup>, Tomoko F. Shibata<sup>1</sup>, Junko Kawashima<sup>1</sup>, Fumiki Katsuoka<sup>1</sup>, Takako  
8 Igarashi-Takai<sup>1</sup>, Soichi Ogishima<sup>1</sup>, Hirohito Metoki<sup>5</sup>, Hiroaki Hashizume<sup>1</sup>, Nobuo  
9 Fuse<sup>1,2</sup>, Naoko Minegishi<sup>1</sup>, Seizo Koshiba<sup>1</sup>, Osamu Tanabe<sup>1</sup>, Shinichi  
10 Kuriyama<sup>1,2</sup>, Kengo Kinoshita<sup>1</sup>, Shigeo Kure<sup>1,2</sup>, Nobuo Yaegashi<sup>1,2,3</sup>, Masayuki  
11 Yamamoto<sup>1,2</sup>, Satoshi Hiyama<sup>4</sup>, and Masao Nagasaki<sup>1\*</sup>.

12  
13 1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryō-  
14 machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

15 2 Tohoku University Graduate School of Medicine, 1-1, Seiryō-machi, Aoba-ku,  
16 Sendai, Miyagi, 980-8574, Japan.

17 3 Tohoku University Hospital, 1-1, Seiryō-machi, Aoba-ku, Sendai, Miyagi, 980-  
18 8574, Japan.

19 4 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,  
20 Kanagawa, Japan 239-8536.

21 5 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima,  
22 Aobaku, Sendai, Miyagi 981-8558, Japan.

23

1  
2  
3 24 Email addresses:  
4

5 25 \*Junichi Sugawara: [jsugawara@med.tohoku.ac.jp](mailto:jsugawara@med.tohoku.ac.jp)  
6

7 26 Daisuke Ochi: [ochi@megabank.tohoku.ac.jp](mailto:ochi@megabank.tohoku.ac.jp)  
8

9 27 Riu Yamashita: [ryamasi@megabank.tohoku.ac.jp](mailto:ryamasi@megabank.tohoku.ac.jp)  
10

11 28 Takafumi Yamauchi: [t.yamauchi@megabank.tohoku.ac.jp](mailto:t.yamauchi@megabank.tohoku.ac.jp)  
12

13 29 Daisuke Saigusa: [saigusa@m.tohoku.ac.jp](mailto:saigusa@m.tohoku.ac.jp)  
14

15 30 Maiko Wagata: [wagata@med.tohoku.ac.jp](mailto:wagata@med.tohoku.ac.jp)  
16

17 31 Mami Ishikuro: [m\\_ishikuro@med.tohoku.ac.jp](mailto:m_ishikuro@med.tohoku.ac.jp)  
18

19 32 Yoshiki Tsunemoto: [yoshiki.tsunemoto@megabank.tohoku.ac.jp](mailto:yoshiki.tsunemoto@megabank.tohoku.ac.jp)  
20

21 33 Yuki Harada: [harada@megabank.tohoku.ac.jp](mailto:harada@megabank.tohoku.ac.jp)  
22

23 34 Tomoko F. Shibata: [tshibata@megabank.tohoku.ac.jp](mailto:tshibata@megabank.tohoku.ac.jp)  
24

25 35 Junko Kawashima: [kawashima@dent.tohoku.ac.jp](mailto:kawashima@dent.tohoku.ac.jp)  
26

27 36 Fumiki Katsuoka: [kfumiki@med.tohoku.ac.jp](mailto:kfumiki@med.tohoku.ac.jp)  
28

29 37 Takako Igarashi-Takai: [takai@megabank.tohoku.ac.jp](mailto:takai@megabank.tohoku.ac.jp)  
30

31 38 Soichi Ogishima: [ogishima@megabank.tohoku.ac.jp](mailto:ogishima@megabank.tohoku.ac.jp)  
32

33 39 Hirohito Metoki: [hmetoki@tohoku-mpu.ac.jp](mailto:hmetoki@tohoku-mpu.ac.jp)  
34

35 40 Hiroaki Hashizume: [hashizume@megabank.tohoku.ac.jp](mailto:hashizume@megabank.tohoku.ac.jp)  
36

37 41 Nobuo Fuse: [fusen@megabank.tohoku.ac.jp](mailto:fusen@megabank.tohoku.ac.jp)  
38

39 42 Naoko Minegishi: [nmine@med.tohoku.ac.jp](mailto:nmine@med.tohoku.ac.jp)  
40

41 43 Seizo Koshiba: [koshiba@megabank.tohoku.ac.jp](mailto:koshiba@megabank.tohoku.ac.jp)  
42

43 44 Osamu Tanabe: [otanabe@megabank.tohoku.ac.jp](mailto:otanabe@megabank.tohoku.ac.jp)  
44

45 45 Shinichi Kuriyama: [kuriyama@med.tohoku.ac.jp](mailto:kuriyama@med.tohoku.ac.jp)  
46

47 46 Kengo Kinoshita: [kengo@ecei.tohoku.ac.jp](mailto:kengo@ecei.tohoku.ac.jp)  
48

49 47 Shigeo Kure: [kure@med.tohoku.ac.jp](mailto:kure@med.tohoku.ac.jp)  
50

51 48 Nobuo Yaegashi: [yaegashi@med.tohoku.ac.jp](mailto:yaegashi@med.tohoku.ac.jp)  
52

1  
2  
3 49 Masayuki Yamamoto: masiyamamoto@med.tohoku.ac.jp  
4

5 50 Satoshi Hiyama: hiyamas@nttdocomo.com  
6

7 51 \*Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp  
8  
9

10 52  
11

12 53 \*Corresponding authors:  
13

14 54 Junichi Sugawara: jsugawara@med.tohoku.ac.jp  
15

16 55 and Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp  
17

18 56 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryomachi,  
19

20 57 Aoba-ku, 9808573, Sendai, Japan. Phone: +81-22-273-6283  
21  
22

23 58  
24

25 59 Word count: 3931 words  
26  
27

28 60  
29

30 61  
31

32 62  
33

34 63  
35

36 64  
37

38 65  
39

40 66  
41

42 67  
43

44 68  
45

46 69  
47

48 70  
49

50 71  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **72 Abstract**

4  
5 **73 Introduction:** Multifactorial diseases, including various complications of  
6 pregnancy, are caused by a complex interaction of genetic and environmental  
7 factors such as lifestyle and living environment. The evaluation of continuous  
8 lifestyle monitoring using healthcare devices provides information on latent  
9 physiologic changes prior to the onset of disease. We expect that monitoring  
10 these factors directly is more accurate than using conventional methods such  
11 as questionnaires. A prospective cohort study for pregnant women, the  
12 Maternity Log study (MLOG), was designed to construct a time-course high-  
13 resolution reference catalog of bioinformatic data in pregnancy and explore the  
14 associations between genomic and environmental factors and the onset of  
15 pregnancy complications using continuous lifestyle monitoring combined with  
16 multi-omics data on the genome, transcriptome, proteome, metabolome, and  
17 microbiome.  
18  
19

20  
21 **22 Methods and analysis:** Pregnant women were recruited at the timing of first  
23 routine antenatal visits. Study participants uploaded daily general health  
24 information including quality of sleep, condition of bowel movements, and the  
25 presence of nausea, pain, and uterine contractions. Participants also collected  
26 physiologic data, such as body weight, blood pressure, heart rate, and body  
27 temperature, using multiple home healthcare devices. Biospecimens, including  
28 maternal plasma, serum, urine, saliva, dental plaque, and cord blood, were  
29 collected for multi-omics analysis. This study is expected to elucidate the causal  
30 relationship between complicated pregnancy and maternal lifestyle and  
31 physiologic changes. Lifelog and multi-omics data will be used to construct a  
32 time-course high-resolution reference catalog of pregnancy. The reference  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 97 catalog will allow us to discover relationships among multi-dimensional  
4  
5 98 phenotypes and novel risk markers in pregnancy for the future personalized  
6  
7 99 early prediction of pregnancy complications.  
8  
9

10 **Ethics and dissemination:** This study was approved by the Tohoku Medical  
11  
12 Megabank Organization, Tohoku University (2014-1-704 and 2017-1-085).  
13  
14 Written informed consent was obtained from all participants.  
15

16 103

17  
18 **Strengths and limitations of this study:**  
19

20  
21 105 This is the first study designed to collect longitudinal lifelog information through  
22  
23 106 healthcare devices, self-administered questionnaires using smartphones, and  
24  
25 107 varieties of biospecimens throughout pregnancy.  
26

27 108 Longitudinal, continuous, individual lifelog data with a high acquisition rate will  
28  
29 109 enable us to assess dynamic physiological changes throughout pregnancy.  
30

31 110 Mutli-omics data will make it possible to understand the complex mechanisms  
32  
33 111 of multifactorial pregnancy-related diseases.  
34

35  
36 112 A time-course high-resolution reference catalog of wellness and multi-omics  
37  
38 113 data will be informative to develop a personalized predictive model for  
39  
40 114 pregnancy complications.  
41

42 115 Further study with larger sample size is needed to validate a reference catalog  
43  
44 116 of normal pregnancy and a prediction model of pregnancy complications.  
45

46 117

47  
48 118

49  
50 119

51  
52 120

53  
54 121  
55  
56  
57  
58  
59  
60



## 122 INTRODUCTION

123 The incidence of pregnancy-related disorders, including hypertensive disorders  
124 of pregnancy (HDP), gestational diabetes mellitus (GDM), and preterm delivery  
125 has been increasing worldwide [1-4]. These multifactorial conditions are caused  
126 by an interaction of genetic factors and environmental factors [5,6]. Recent  
127 reports suggest that continuous lifestyle monitoring using wearable biosensors  
128 provides important information on latent physiologic changes that are exhibited  
129 prior to the onset of disease [7]. Using these monitors, environmental factors  
130 may be estimated more accurately than by using conventional questionnaires.

131 For these reasons, we have designed a prospective cohort study for pregnant  
132 women, the Maternity Log study (MLOG). In this study, pregnant women upload  
133 daily information and physiologic data using multiple home healthcare devices.  
134 In addition, variety of biospecimens are collected for multi-omics analysis.

135 To the best of our knowledge, this study will be the first to integrate multi-  
136 omics data with objective data on environmental factors, including daily lifelog  
137 data, in pregnant women. Integrated information from the study will be utilized  
138 to discover the relationship among multi-dimensional phenotypes and novel risk  
139 markers for the future personalized early prediction of pregnancy complications.

140

## 141 METHODS AND ANALYSIS

### 142 Study setting

143 The aim of the MLOG study is to construct a time-course high-resolution  
144 reference catalog of bioinformatic data in pregnancy and thereby develop  
145 methods by which early prediction of obstetric complications, through integrated  
146 analysis of daily lifelogs and multi-omics data, *i.e.*, maternal genomes,

1  
2  
3 147 transcriptomes, metabolomes, and oral microbiomes.  
4

5 148 The MLOG study is a prospective, add-on cohort study, built on a birth- and 3-  
6  
7 149 generation cohort study established by the Tohoku Medical Megabank  
8  
9  
10 150 Organization (TMM BirThree Cohort Study) [8] in order to elucidate the  
11  
12 151 mechanisms of complicated multifactorial diseases in mothers and children in  
13  
14 152 the wake of the Great East Japan Earthquake in 2011. Epidemiological data  
15  
16 153 from extensive questionnaire surveys and accurate clinical records, including  
17  
18 154 birth outcomes, can be abstracted from the integrated biobank of Tohoku  
19  
20 155 Medical Megabank Organization (ToMMo) [8].

21  
22  
23 156 Written informed consent was obtained from all participants by the genome  
24  
25 157 medical research coordinators (GMRCs). This study was conducted in  
26  
27 158 accordance with the Declaration of Helsinki and approved by the ethics  
28  
29 159 committee of ToMMo, Tohoku University (2014-1-704, 2017-1-085). This study  
30  
31 160 was conducted under a collaborative research agreement with ToMMo, Tohoku  
32  
33 161 University and NTT DoCoMo, Inc. (Tokyo, Japan).  
34

35  
36 162

### 37 38 163 **Patient and public involvement**

39  
40 164 Patients and public were not directly involved in the development of the  
41  
42 165 research question or the design of the study. The main results will be made  
43  
44 166 available in the public domain.  
45

46  
47 167

### 48 49 168 **Participants**

50  
51 169 Participants were recruited at a first routine antenatal visit at Tohoku University  
52  
53 170 Hospital, Sendai, Japan between September 2015 and September 2016. A  
54  
55 171 flowchart of the recruitment process is shown in Figure 1. Patients who already

1  
2  
3 172 agreed to participate in the TMM BirThree Cohort Study were recruited to  
4  
5 173 provide an additional informed consent for the MLOG study. A total of 302  
6  
7 174 women were enrolled. The inclusion criteria were age  $\geq 20$  years and the ability  
8  
9 175 to access the internet using a smartphone in the Japanese language.  
10  
11  
12 176 Participants were excluded after enrollment if termination of pregnancy,  
13  
14 177 abortion, or transfer to another institution for emergency care occurred before  
15  
16 178 delivery, or if they withdrew consent for any reason.  
17  
18  
19

179

### 180 **Outline of study protocol**

181 The study protocol consisted of blood and urine sampling, saliva and dental  
182 plaque sampling, self-administered daily lifelog data collection, and data upload  
183 from multiple wearable devices through a smartphone. An overview of the  
184 protocol is provided in Figure 2. In Japan, routine antenatal visits, including  
185 ultrasounds, are scheduled every 4 weeks from early pregnancy ( $< 12$  weeks)  
186 to 23 weeks of gestation, every 2 weeks from 24 to 35 weeks, and every week  
187 from 36 weeks to delivery [9]. Lifelog data collection was continued throughout  
188 pregnancy and until 1 month after delivery. Optional data collection could be  
189 continued up to 180 days after delivery.

190

### 191 **Blood and urine sampling**

192 Blood samples were collected 3 times from each participant; the first sample  
193 was collected between 12 and 24 weeks of gestation, the second between 24  
194 and 36 weeks, and the third at 1 month after delivery. A maximum of 13 mL of  
195 blood was collected each time, from which serum and plasma were separated  
196 to be stored at  $-80^{\circ}\text{C}$  until the time of analysis. An aliquot of blood (2.5 mL) was

1  
2  
3 197 stored in a PAXgene® tube (Becton, Dickinson and Company, Franklin Lakes,  
4  
5 198 NJ, USA) at -80°C until the time of RNA extraction for transcriptome analysis.  
6  
7 199 Genomic DNA was extracted from mononuclear cells using an Autopure®  
8  
9 200 extractor (Qiagen, Venlo, The Netherlands). Approximately 10 mL of blood was  
10  
11 201 collected from the umbilical vein in a PAXgene® tube for storage at -80°C, and  
12  
13 202 in an EDTA 2K tube (Becton, Dickinson and Company, Franklin Lakes, NJ,  
14  
15 203 USA) for separation of plasma to be stored at -80°C. Urine samples (10 mL)  
16  
17 204 were collected at each antenatal visit; when participants were admitted to the  
18  
19 205 hospital ward, urine was collected once weekly. Urine samples were  
20  
21 206 immediately transferred and stored at -80°C until the time of analysis.  
22  
23  
24  
25  
26

### 27 208 **Saliva and dental plaque sampling**

28  
29 209 Samples of saliva and dental plaque were collected 3 times from each  
30  
31 210 participant, at the same time points as blood collection. Approximately 3 mL of  
32  
33 211 saliva was collected using a 50-mL conical centrifuge tube (Corning, Inc.,  
34  
35 212 Corning, NY, USA) and stored at -80°C until analysis. Dental plaque was  
36  
37 213 sampled by brushing, suspended in 0.5 mL of Tris-EDTA (10 mM Tris, 1 mM  
38  
39 214 EDTA; pH, 8.0), and immediately stored at -80°C until the time of sample  
40  
41 215 processing.  
42  
43  
44  
45

### 46 217 **Lifelog data collection**

47  
48 218 Based on previous publications on the utility for risk assessment of pregnancy-  
49  
50 219 related diseases, we selected several lifelog parameters to employ in this study,  
51  
52 220 *i.e.*, body temperature [10], home blood pressure [11], body weight [12],  
53  
54 221 physical activity (calorie expenditure) [13], as well as self-administered  
55  
56  
57  
58  
59  
60

1  
2  
3 222 information such as sleep quality [14], condition of stool [15], severity of nausea  
4 [16], fetal movement [17], severity of pain [18], uterine contractions [19], and  
5 223  
6 [16], fetal movement [17], severity of pain [18], uterine contractions [19], and  
7 224  
8 palpitations [20]. Body temperature, home blood pressure, body weight, and  
9  
10 225 physical activity were uploaded from multiple healthcare devices through a  
11  
12 226 smartphone. The self-administered information described above was input  
13  
14 227 manually on mobile applications created for this study.

15  
16 228 Data collection was started after obtaining informed consent and after giving  
17  
18 229 detailed instructions for the use of the healthcare devices. These applications  
19  
20 230 tracked quality of sleep; condition of stool using the Bristol Scale [21-23];  
21  
22 231 severity of nausea using the Pregnancy-Unique Quantification of Emesis and  
23  
24 232 nausea (PUQE) score [24,25]; headache, toothache, lumbago, and upper and  
25  
26 233 lower abdominal pain using a numerical rating scale (NRS) score; the number  
27  
28 234 of perceived uterine contractions; palpitations; and fetal movement using a  
29  
30 235 modified count-to-10 fetal movement chart [26,27].

31  
32  
33 236 Sleep quality was evaluated by the wakeup time, bedtime, sleep satisfaction  
34  
35 237 (ranked from satisfied to poor using a numeric scale of 0-4), and the number of  
36  
37 238 nocturnal awakenings (0-6).

38  
39  
40 239 The Bristol stool form scale was originally developed to assess constipation  
41  
42 240 and diarrhea [21, 22], and its use has been spread widely to evaluate functional  
43  
44 241 bowel disorders [22]. Using the Bristol scale, stool is classified into 7 types  
45  
46 242 according to cohesion and surface cracking [21, 22].

47  
48 243 The PUQE score [24, 25] was developed to estimate the severity of nausea  
49  
50 244 and vomiting in pregnancy and quantifies the number of daily vomiting and  
51  
52 245 retching episodes and the length of nausea in hours (over the preceding 12 h).

53  
54  
55 246 The total score ranges from 3 (no symptoms) to 15, and higher scores are

1  
2  
3 247 correlated with increasing severity of nausea and vomiting [24, 25].  
4

5 248 In the NRS score for headache, toothache, lumbago, and upper and lower  
6  
7 249 abdominal pain, the total score ranges from 0 (no pain) to 10 (maximum ever  
8  
9 250 experienced).

10  
11 251 Uterine contractions and palpitations were evaluated using definitions  
12  
13 252 determined for the current study. Uterine contractions were assessed using the  
14  
15 253 number of perceived contractions per day, ranging from 0 to more than 5. The  
16  
17 254 count-to-10 method was originally developed to assess fetal well-being by  
18  
19 255 recording the time, in minutes, required to count 10 fetal movements [26]. More  
20  
21 256 recently, a modified count-to-10 method has been proposed: pregnant women  
22  
23 257 are advised to start counting when they feel the first movement, then record the  
24  
25 258 time required to perceive an additional 9 movements [27]. Pregnant women are  
26  
27 259 encouraged to select a 2-hour period when they feel active fetal movements  
28  
29 260 and are instructed to count kicking and rolling movements in a favorable  
30  
31 261 maternal position after 24 weeks of gestation.  
32  
33  
34  
35

36 262 The applications also collected dietary logs and the medications taken on the  
37  
38 263 day before and the day of the antenatal visit, on which blood or urine samples  
39  
40 264 were collected.

41  
42 265 Daily home blood pressure, body weight, body temperature, and physical  
43  
44 266 activity were measured as described below with home healthcare devices, and  
45  
46 267 uploaded through wireless communications using mobile applications on a  
47  
48 268 smartphone. Daily home blood pressure was measured twice daily using an  
49  
50 269 HEM-7510 monitor (OMRON Healthcare Co., Ltd., Kyoto, Japan): within 1 hour  
51  
52 270 of awakening in the morning and just before going to bed at night. Body weight  
53  
54 271 was measured using an HBF-254C meter (OMRON Healthcare Co., Ltd.) once  
55  
56

1  
2  
3 272 daily within 1 hour of awakening in the morning. Daily body temperature was  
4  
5 273 evaluated using an MC-652LC digital thermometer (MC-652LC; OMRON  
6  
7 274 Healthcare Co., Ltd.) just after awakening. Physical activity was assessed using  
8  
9 275 an HJA-403C pedometer (HJA-403C; OMRON Healthcare Co., Ltd.) to count  
10  
11 276 steps and calculate calorie expenditure.  
12  
13

14 277

### 16 278 **Clinical and epidemiological information**

18 279 Baseline clinical information and maternal and neonatal outcomes (*e.g.*,  
19  
20 280 maternal age, clinical data and findings from each antenatal visit, gestational  
21  
22 281 age at delivery, type of delivery, birth weight, maternal and fetal complications)  
23  
24 282 were obtained from the medical records of the Tohoku University Hospital.  
25  
26 283 Epidemiological data, including extensive questionnaire surveys from the TMM  
27  
28 284 BirThree Cohort can be obtained from the ToMMo integrated biobank [8].  
29  
30

31 285

### 33 286 **Database**

35 287 A customized laboratory information management system (LIMS) was  
36  
37 288 established to track all biospecimens. All data were transferred to the TMM  
38  
39 289 integrated database after 2-step anonymization in a linkable fashion.  
40  
41 290 Data handling was strictly regulated under HIPAA (Health Insurance Portability  
42  
43 291 and Accountability Act of 1996, United States Security and Privacy Rules) [28,  
44  
45 292 29] and the Act on the Protection of Personal Information [30]. Security control  
46  
47 293 at our facility has been described previously [31].  
48  
49

50 294

### 53 295 **Omics analysis**

54  
55 296 Whole-genome sequencing  
56  
57



1  
2  
3 297 To minimize amplification bias, we adopted a PCR-free library preparation  
4  
5 298 method. After performing library quality control using the quantitative MiSeq  
6  
7 299 method [32], libraries were sequenced on HiSeq 2500 Sequencing System  
8  
9 300 (Illumina, Inc., San Diego, CA, USA) to generate 259-bp, paired-end reads. We  
10  
11 301 generated the sequencing data at over 12.5x coverage on average, and we  
12  
13 302 identified variants using the alignment tool BWA-MEM (ver. 0.7.5a-r405) with  
14  
15 303 the default option. Single nucleotide variants (SNVs) and indels were jointly  
16  
17 304 called across all samples using Genome Analysis Tool Kit's HaplotypeCaller  
18  
19 305 (ver. 3.8). Default filters were applied to SNV and indel calls using the GATK's  
20  
21 306 Variant Quality Score Recalibration (VQSR) approach. The human reference  
22  
23 307 genome was GRCh37/hg19 with the decoy sequence (hs37d5) and NC\_007605  
24  
25 308 (Human Gamma Herpesvirus 4). The complete fasta file named  
26  
27 309 hg19\_tommo\_v2.fa is available from iJGVD website  
28  
29 310 (<http://ijgvd.megabank.tohoku.ac.jp>).

31  
32  
33 311

### 312 Transcriptome

313 Whole blood were collected using the PAXgene® RNA tube, which is widely  
34  
35 314 used for transcriptome analysis. After storage at -80°C, total RNA was purified  
36  
37 315 with PAXgene Blood RNA Kit® (Qiagen, Venlo, The Netherlands) using  
38  
39 316 QiaSymphony® (Qiagen ). The amount and quality of the total RNA was  
40  
41 317 assessed with Bio Analyzer® or Tape Station® (both from Agilent  
42  
43 318 Technologies, Santa Clara, CA, USA), and we only used RNA samples with an  
44  
45 319 RNA integrity number (RIN) (or an RIN equivalent) higher than 7.0. Total RNA  
46  
47 320 was reverse-transcribed using an oligo-dT primer. We used TruSeq DNA PCR-  
48  
49 321 Free Library Preparation Kit (Illumina, Inc.) for library preparation for



1  
2  
3 322 sequencing with HiSeq 2500 Sequencing System.  
4

5 323  
6

7 324 Plasma and urine metabolome  
8

9 325 *Nuclear magnetic resonance (NMR) spectroscopy*  
10

11 326 All NMR measurements for metabolome analysis were conducted at 298 K on a  
12

13 327 Bruker Avance 600 MHz spectrometer equipped with a SampleJet sample  
14

15 328 changer (Bruker Corp., Billerica, MA, USA) [35]. Standard 1-dimensional  
16

17 329 nuclear Overhauser enhancement spectroscopy (1D NOESY) and Carr-Purcell-  
18

19 330 Meiboom-Gill (CPMG) spectra were obtained for each plasma or urine sample.  
20

21 331 All spectra for plasma or urine samples were acquired using 16 scans and 32 k  
22

23 332 of complex data points. All data were analyzed using the TopSpin 3.5 (Bruker  
24

25 333 Corp.) and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta,  
26

27 334 Canada) programs.  
28

29 335  
30

31 336 *Gas chromatography-tandem mass spectrometry (GC-MS/MS)*  
32

33 337 Sample preparation for plasma and urine (50  $\mu$ L each) was performed using a  
34

35 338 Microlab STARlet robot system (Hamilton, Reno, NV, USA) followed by the  
36

37 339 methods previously reported by Nishiumi [36, 37]. The resulting deproteinized  
38

39 340 and derivatized supernatant (1  $\mu$ L) was subjected to GC-MS/MS, performed on  
40

41 341 a GC-MS TQ-8040 system (Shimadzu Corp., Kyoto, Japan). The compound  
42

43 342 separation was performed using a fused silica capillary column (BPX-5; 30 m  $\times$   
44

45 343 0.25 mm inner diameter; film thickness, 0.25  $\mu$ m; Shimadzu Corp, Kyoto,  
46

47 344 Japan). Metabolite detection was performed using Smart Metabolites Database  
48

49 345 (Shimadzu Corp.) that contained the relevant multiple reaction monitoring  
50

51 346 (MRM) method file and data regarding the GC analytical conditions, MRM  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 347 parameters, and retention index employed for the metabolite measurement. The  
4  
5 348 database used in this study included data on 475 peaks from 334 metabolites.  
6  
7 349 All peaks of metabolites detected from each sample was annotated and  
8  
9 350 analyzed using Traverse MS® (Reifycs Inc., Tokyo, Japan).  
10  
11  
12 351

## 13 14 352 Oral Microbiome

15  
16 353 Analysis of oral microbiome was conducted by previously reported protocols  
17  
18 354 [36]. In brief, saliva was collected in a 25-mL tube. Dental plaque was sampled  
19  
20 355 by participants themselves by brushing teeth with a sterilized toothbrush, and  
21  
22 356 then suspended in saline for collection. Both samples were stored at -80°C until  
23  
24 357 the time of processing. DNA was extracted from saliva and dental plaque by  
25  
26 358 standard glass bead-based homogenization and subsequent purification with a  
27  
28 359 silica-membrane spin-column using PowerSoil DNA Isolation Kit (Mo Bio  
29  
30 360 Laboratories, Carlsbad, CA, USA). DNA was eluted from the spin column with  
31  
32 361 30-µL RNase-free water (Takara Bio, Inc., Shiga, Japan), and stored at -20°C  
33  
34 362 after determining the amount and quality of DNA with a Nanodrop  
35  
36 363 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Using  
37  
38 364 DNA extracted from saliva or dental plaque as a template, a part of the V4  
39  
40 365 variable region of the bacterial 16S rRNA gene was amplified by 2-step PCR.  
41  
42 366 Tag-indexed PCR products thus obtained were subjected to multiplex amplicon  
43  
44 367 sequencing using MiSeq System and MiSeq Sequencing Reagent Kit, v3  
45  
46 368 (Illumina, Inc.) according to the manufacturer's instructions.  
47  
48  
49  
50

51 369

## 52 53 370 **Outcomes**

54  
55 371 The following obstetric complications represented the primary outcomes. HDP  
56  
57  
58  
59  
60

1  
2  
3 372 was classified as gestational hypertension, preeclampsia, superimposed  
4  
5 373 preeclampsia, and chronic hypertension [37]. Spontaneous preterm birth was  
6  
7 374 defined as spontaneous preterm labor or preterm premature rupture of  
8  
9  
10 375 membranes resulting in preterm birth at less than 37 weeks of gestation. GDM  
11  
12 376 was diagnosed according to the International Association of the Diabetes and  
13  
14 377 Pregnancy Study Groups (IADPSG) criteria [38]. The secondary outcomes were  
15  
16 378 maternal body weight, blood pressure, physical activity, lifestyle changes,  
17  
18 379 perinatal mental disorders, fetal growth, fetal movement, and birth weight.  
19  
20  
21 380

### 22 381 **Statistical analysis**

23  
24 382 The association of outcomes with each factor will be analyzed using a statistical  
25  
26 383 hypothesis test such as Welch's t-test, Fisher's exact test, the Chi-square test,  
27  
28 384 and others as appropriate. Multiple logistic regression modelling will be used to  
29  
30 385 adjust for confounders and to assess whether each factor or combination of  
31  
32 386 factors can be used to predict outcomes. Stepwise selection algorithms or  
33  
34 387 regularized algorithms (e.g., LASSO, ridge regression, or elastic net) will be  
35  
36 388 used to select the optimal number of contributing factors that maximize the  
37  
38 389 predictive power using the leave-1-out cross validation or K-fold cross validation  
39  
40 390 methods.  
41  
42  
43

44 391 Individual genetic factors may have an effect on outcomes; therefore, some  
45  
46 392 aggregated genetic risk score should be included in the prediction model. For  
47  
48 393 example, SNVs, including rare variants in or around a chromosome region of a  
49  
50 394 known or estimated risk gene, could be aggregated by considering their impacts  
51  
52 395 on biological function of the gene or their minor allele frequencies in the  
53  
54 396 population. However, we are limited in the number of study participants, and the

397 aggregated risk score might therefore contribute only slightly to the predictive  
 398 power. To create a more reliable risk score, the estimates from other large-  
 399 scale cohort data using polygenic score tools, *e.g.*, PRSice [39], could be used  
 400 for this study.

401

## 402 INTERIM RESULTS

### 403 Clinical background

404 A total of 302 women were enrolled, and the mean gestational weeks of  
 405 recruitment was  $16.4 \pm 4.9$  weeks (mean  $\pm$  SD). A total of 285 participants have  
 406 delivered; their baseline clinical characteristics are described in Table 1. The  
 407 mean maternal age at delivery was  $33.3 \pm 4.9$  years. Approximately 42% of the  
 408 participants were over 35 years of age, 51% were parous, and 22% were  
 409 overweight or obese by their prepregnancy body mass indices ( $BMI \geq 25$   
 410  $kg/m^2$ ). Overall, 8.4% of participants had HDP, and 5.6% underwent  
 411 spontaneous preterm birth. On average, infants were delivered at  $38.0 \pm 2.3$   
 412 weeks of gestation with a mean birth weight of  $2907 \pm 572$  g. The rate of low  
 413 birth weight was 18%. Mean gestational weeks of the first and second blood  
 414 sampling were  $17.0 \pm 5.0$  and  $27.5 \pm 2.5$ , respectively. The third blood sampling  
 415 was performed at  $31.1 \pm 3.0$  days after delivery on average. The length of  
 416 enrollment ranged from 90 to 396 days with a mean of  $216 \pm 61$  days.

417  
418

**Table 1.** Participant characteristics

Characteristic	Value
Maternal (n = 285)	
• Age at delivery, y, mean (SD)	33.3 ( $\pm$ 4.9)
• Age at delivery, y, n (%)	
20-24	12 (4.2)
25-29	45 (15.8)
30-34	107 (37.5)

17

1		
2		
3	35-39	90 (31.6)
4	40-44	30 (10.5)
5	45-49	1 (0.4)
6		
7	• Parity, n (%)	
8	0	140 (49.1)
9	1	93 (32.6)
10	≥ 2	52 (18.2)
11		
12	• Prepregnancy BMI*, kg/m <sup>2</sup> , mean (SD)	22.7 (± 5.1)
13	• Prepregnancy BMI, kg/m <sup>2</sup> , n (%)	
14	< 18.5	36 (12.6)
15	18.5-24.9	186 (65.3)
16	25.0-29.9	34 (11.9)
17	≥ 30.0	29 (10.2)
18		
19	• Gestational weeks at delivery, mean (SD)	38.0 (± 2.3)
20		
21	• Mode of delivery, n (%)	
22	Noncesarean	179 (62.8)
23	Cesarean	106 (37.2)
24		
25	• Pregnancy complication, n (%)	
26	Hypertensive disorder of pregnancy	24 (8.4)
27	Spontaneous preterm birth	16 (5.6)
28		
29	Neonatal (n = 300)	
30	• Birth weight, mean (SD)	2907 (± 572)
31	• Sex, n (%)	
32	Male	168 (56)
33	Female	132 (44)
34	• Low-birth weight (< 2500 g), n (%)	54 (18)
35		
36	419	*BMI, body mass index

420

421 **Data acquisition**

422 The percentage of data uploads as of June 2017 was calculated for the 285  
423 final study participants. For each lifelog item, the upload rate for each  
424 participant was calculated from the total days of actual upload divided by the  
425 number of days from enrollment to delivery. The mean upload rate for each  
426 lifelog item was 85.3% (steps and calorie), 82.1% (body weight), 80.4% (body  
427 temperature), 78.0% (morning home blood pressure), 71.6% (evening home  
428 blood pressure), 83.5% (sleep quality), 82.1% (condition of stool, severity of  
429 pain, severity of nausea, uterine contractions, palpitations), and 67.4% (fetal

1  
2  
3 430 movement) (Figure 3).  
4  
5

6 431

### 7 432 **Number of data points**

8  
9 433 The total number of collected data points as of June 2017 was calculated for the  
10  
11 434 285 final study participants. The approximate number of registered data points  
12  
13 435 was 86 000 for body weight, 324 000 points for home diastolic and systolic  
14  
15 436 blood pressure, 86 000 for physical activity, and 74 000 for body temperature.  
16  
17 437 When physical conditions such as stool condition, severity of pain, and fetal  
18  
19 438 movement were combined, the total number of data points was over 6 million.  
20  
21  
22

23 439

### 24 440 **DISCUSSION**

25  
26 441 Herein, we have described the rationale, design, objective, data collection  
27  
28 442 methods, and interim results of the MLOG study. The study was launched in  
29  
30 443 September 2016, and baseline data collection ended in June 2017. A total of  
31  
32 444 285 participants uploaded lifelog data throughout pregnancy with a high data  
33  
34 445 acquisition rate and over 6 million total data points. Biospecimens for multi-  
35  
36 446 omics analysis were satisfactorily collected and all tracked by LIMS.

37  
38 447 There are three noteworthy features in the MLOG study. First, it is a  
39  
40 448 prospective add-on cohort study based on the ToMMo BirThree cohort study,  
41  
42 449 with a full series of epidemiological data and a highly structured follow-up  
43  
44 450 system for mothers, newborns, and families [8]. Second, we have successfully  
45  
46 451 collected longitudinal, continuous, individual lifelog data with a high acquisition  
47  
48 452 rate, which will enable us to assess dynamic changes in physiologic conditions  
49  
50 453 throughout pregnancy. Third, multi-omics data will make it possible to fully  
51  
52 454 understand the complex mechanisms of multifactorial pregnancy-related  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 455 diseases and to overcome the unpredictability of these complications.

4 456 Prediction models using clinical and epidemiological information and  
5  
6 457 circulating factors for pregnancy-related diseases have been extensively  
7  
8 458 developed [40], and risk-assessment approaches using clinical information have  
9  
10 459 also been developed [41, 42]. However, there is a lack of evidence for the  
11  
12 460 benefits of these predictive models for routine clinical use [43].  
13  
14

15  
16 461 Once the likelihood of a pregnancy-related disorder is estimated with high  
17  
18 462 sensitivity and specificity, evidence-based clinical interventions could reduce the  
19  
20 463 rate of maternal and neonatal morbidity and mortality [44]. Therefore, an early-  
21  
22 464 prediction algorithm that can be used with a high level of confidence is needed  
23  
24 465 to obtain better outcomes for patients with pregnancy complications.  
25  
26

27 466 Recently, several studies of sample sizes comparable with ours, exploiting  
28  
29 467 lifelog or multi-omics data were reported. One of the studies analyzed lifelog  
30  
31 468 and multi-omics data, collected from 108 individuals at three time points during  
32  
33 469 a nine-month period [45]. In their study, several remarkable relationships were  
34  
35 470 identified among physiological and multi-omics data through integrated  
36  
37 471 analyses. Another study investigated genome-wide associations between  
38  
39 472 genetic variants and gene expression levels across 44 human tissues from a  
40  
41 473 few hundreds of postmortem donors [46]. They studied both cis-eQTL (within 1  
42  
43 474 Mb of target-gene transcription start sites) and trans-eQTLs (more distant from  
44  
45 475 target genes or on other chromosomes) with 350 whole blood samples, and  
46  
47 476 thereby identified 5,862 cis-eQTL and one trans-eQTL associations. These  
48  
49 477 previous studies indicate that our time-course high-resolution reference catalog  
50  
51 478 with 285 pregnant women would be well applicable to high-dimensional data  
52  
53 479 analyses such as searches for quantitative trait loci and molecular risk markers.  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 480 Hopefully, our study will result in the development of a novel stratification  
4  
5 481 model for pregnancy-related diseases employing multi-omics and lifelog data.  
6  
7 482 The MLOG study will enable us to construct a time-course high-resolution  
8  
9 483 reference catalog of wellness and multi-omics data from pregnant women and  
10  
11 484 thereby develop a personalized predictive model for pregnancy complications.  
12  
13 485 Progressive data sharing and collaborative studies would make it possible to  
14  
15 486 establish a standardized early-prediction method through large clinical trials.  
16  
17  
18  
19

487

### 488 **Author affiliations**

489 1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryomachi,  
490 Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

491 2 Tohoku University Graduate School of Medicine, 1-1, Seiryomachi, Aoba-ku,  
492 Sendai, Miyagi, 980-8574, Japan.

493 3 Tohoku University Hospital, 1-1, Seiryomachi, Aoba-ku, Sendai, Miyagi, 980-  
494 8574, Japan.

495 4 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,  
496 Kanagawa, Japan 239-8536.

497 5 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima,  
498 Aobaku, Sendai, Miyagi 981-8558, Japan.

499

### 500 **Acknowledgements**

501 The authors would like to thank all the MLOG study participants, the staff of the  
502 Tohoku Medical Megabank Organization, Tohoku University (a full list of  
503 members is available at: <http://www.megabank.tohoku.ac.jp/english/a161201/>),  
504 and the department of Obstetrics and Gynecology, Tohoku University Hospital,



1  
2  
3 505 for their efforts and contributions. The MLOG study group also included Chika  
4  
5 506 Igarashi, Motoko Ishida, Yumiko Ishii, Hiroko Yamamoto, Akiko Akama, Kaori  
6  
7 507 Noro, Miyuki Ozawa, Yuka Narita, Junko Yusa, Miwa Meguro, Michiyo Sato,  
8  
9 508 Miyuki Watanabe, Mai Tomizuka, Mika Hotta, Naomi Matsukawa, Makiko Sumii,  
10  
11 509 Ayako Okumoto, Yukie Oguma, Ryoko Otokozawa, Toshiya Hatanaka, Sho  
12  
13 510 Furuhashi, Emi Shoji, Tomoe Kano, Riho Mishina, and Daisuke Inoue.  
14  
15

16 511

### 17 512 **Contributors**

18  
19 513 JS, DO, RY, TY, HM, OT, SKu, NY, SH, MN were involved in initial stages of  
20  
21 514 the strategy and design of study conception. JS, DO, RY, TY, OT, DS, SKo, SH,  
22  
23 515 MN: responsible for the draft of the manuscript. JS, DO, RY, TY, MW, MI, HM,  
24  
25 516 OY, SKu: recruitment and sample collection. DO, RY, TY, DS, YT, YH, TFS, JK,  
26  
27 517 FK, TIT, SO, NM, SKo, OT, MN: sample analysis, data processing, and  
28  
29 518 statistical analysis. JS, HH, NF, NM, SKo, OT, SKu, KK, SK, NY, MY, SH, MN:  
30  
31 519 advice and supervision of sample analysis. All authors have contributed to  
32  
33 520 revision and have approved the final manuscript.  
34  
35

36 521

### 37 522 **Funding**

38  
39 523 The present study was supported by NTT DoCoMo, Inc., with a collaborative  
40  
41 524 research agreement between NTT DoCoMo and ToMMo. This work was  
42  
43 525 supported in part by the Tohoku Medical Megabank Project from the Japan  
44  
45 526 Agency for Medical Research and Development (AMED) and the Ministry of  
46  
47 527 Education, Culture, Sports, Science and Technology (MEXT).  
48  
49

50 528

### 51 529 **Competing interests**

1  
2  
3 530 This study was funded by NTT DoCoMo, Inc.  
4

5 531 Daisuke Ochi, Takafumi Yamauchi, and Satoshi Hiyama are employees of NTT  
6

7 532 DoCoMo, Inc. All other authors declare that they have no competing interests.  
8

9  
10 533  
11

### 12 13 534 **Ethics approval and consent to participate**

14  
15 535 The TMM BirThree cohort study was approved by the ethics committee of the  
16

17 536 Tohoku University (authorization numbers, 2013-4-103 and 2017-4-010). The  
18

19 537 MLOG study was approved by the Tohoku Medical Megabank Organization,  
20

21 538 Tohoku University (2014-1-704 and 2017-1-085). Written informed consent was  
22

23 539 obtained from all participants.  
24

25  
26 540  
27

### 28 541 **Availability of data and materials**

29 542 The datasets used during the current study are available from the  
30

31 543 corresponding author on reasonable request.  
32

33  
34 544  
35

36 545  
37

38 546  
39

40 547  
41

42 548  
43

44 549  
45

46 550  
47

48 551  
49

50 552  
51

## 52 553 **REFERENCES**

- 1  
2  
3 554 1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public  
4 health perspective. *Diabetes Care*. 2007;30 Suppl 2: S141-6.  
5  
6 555  
7  
8 556  
9  
10 557 2. Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, Requejo JH, et al. The  
11 worldwide incidence of preterm birth: a systematic review of maternal mortality  
12 and morbidity. *Bull World Health Organ*. 2010; 88:31-8.  
13  
14 559  
15  
16 560  
17  
18 561 3. Duley L. The global impact of pre-eclampsia and eclampsia. *Semin Perinatol*.  
19 2009; 33:130-7.  
20  
21 562  
22  
23 563  
24  
25 564 4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United  
26 States, 1980-2010: age-period-cohort analysis. *BMJ*. 2013;347: f 6564.  
27  
28 565  
29  
30 566  
31  
32 567 5. Waken RJ, de Las Fuentes L, Rao DC. A Review of the Genetics of  
33 Hypertension with a Focus on Gene-Environment Interactions. *Curr*  
34 *Hypertens Rep*. 2017; 19:23.  
35  
36 568  
37  
38 569  
39  
40 570  
41  
42 571 6. Ward K, Lindheimer MD. Genetic factors in the etiology of preeclampsia /  
43 eclampsia. In: Chesley's Hypertensive Disorders in pregnancy. London:  
44 Elsevier; 2990: 51-72.  
45  
46 572  
47  
48 573  
49  
50 574  
51  
52 575 7. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schu "ssler-Fiorenza Rose SM, et  
53 al. Digital Health: Tracking Physiomes and Activity Using Wearable  
54 Biosensors Reveals Useful Health Related Information. *PLoS Biol*. 2017; 15:  
55 e2001402.  
56  
57  
58  
59  
60

- 1  
2  
3 579  
4  
5 580 8. Kuriyama S, Yaegashi N, Nagami F, Arai T, Kawaguchi Y, Osumi N, et al.  
6  
7 581 The Tohoku Medical Megabank Project: Design and Mission.  
8  
9 582 J Epidemiol. 2016; 26:493-511.  
10  
11 583  
12  
13 584 9. Japan Society of Obstetrics and Gynecology, Guideline for Obstetrical  
14  
15 585 Practice in Japan, Japan Society of Obstetrics and Gynecology, Tokyo,  
16  
17 586 Japan, pp. 1–4, 2017 [in Japanese].  
18  
19 587  
20  
21 588 10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the  
22  
23 589 thermoneutral zone: a longitudinal study of normal human pregnancy. Acta  
24  
25 590 Physiol (Oxf). 2011; 201: 467-74.  
26  
27 591  
28  
29 592 11. Metoki H, Ohkubo T, Watanabe Y, Nishimura M, Sato Y, Kawaguchi M,  
30  
31 593 Hara A, Hirose T, Obara T, Asayama K, Kikuya M, Yagihashi K, Matsubara Y,  
32  
33 594 Okamura K, Mori S, Suzuki M, Imai Y; BOSHI Study Group. Seasonal trends of  
34  
35 595 blood pressure during pregnancy in Japan: the babies and their parents'  
36  
37 596 longitudinal observation in Suzuki Memorial Hospital in Intrauterine Period  
38  
39 597 study. J Hypertens. 2008; 26: 2406-13.  
40  
41 598  
42  
43 599 12. Haugen M, Brantsæter AL, Winkvist A, Lissner L, Alexander J, Oftedal B,  
44  
45 600 Magnus P, Meltzer HM. Associations of pre-pregnancy body mass index and  
46  
47 601 gestational weight gain with pregnancy outcome and postpartum weight  
48  
49 602 retention: a prospective observational cohort study. BMC Pregnancy Childbirth.  
50  
51 603 2014 Jun 11; 14: 201.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 604

4  
5 605 13. Sorensen TK, Williams MA, Lee IM, Dashow EE, Thompson ML, Luthy DA.

6  
7 606 Recreational physical activity during pregnancy and risk of preeclampsia.

8  
9 607 Hypertension. 2003 Jun; 41:1273-80.

10  
11 608

12  
13 609 14. Reutrakul S, Zaidi N, Wroblewski K, Kay HH, Ismail M, Ehrmann DA, Van

14  
15 610 Cauter E. Sleep disturbances and their relationship to glucose tolerance in

16  
17 611 pregnancy. Diabetes Care. 2011; 34: 2454-7.

18  
19 612

20  
21 613 15. Cornish J, Tan E, Teare J, Teoh TG, Rai R, Clark SK, Tekkis PP. A meta-

22  
23 614 analysis on the influence of inflammatory bowel disease on pregnancy. Gut.

24  
25 615 2007; 56: 830-7.

26  
27 616

28  
29 617 16. Huxley RR. Nausea and vomiting in early pregnancy: its role in placental

30  
31 618 development. Obstet Gynecol. 2000; 95:779-82.

32  
33 619

34  
35 620 17. Holm Tveit JV, Saastad E, Stray-Pedersen B, Børdahl PE, Frøen JF.

36  
37 621 Maternal characteristics and pregnancy outcomes in women presenting with

38  
39 622 decreased fetal movements in late pregnancy. Acta Obstet Gynecol Scand.

40  
41 623 2009; 88: 1345-51.

42  
43 624

44  
45 625 18. Facchinetti F, Allais G, D'Amico R, Benedetto C, Volpe A. The relationship

46  
47 626 between headache and preeclampsia: a case-control study. Eur J Obstet

48  
49 627 Gynecol Reprod Biol. 2005; 121: 143-8.

50  
51 628

- 1  
2  
3 629 19. Iams JD, Newman RB, Thom EA, Goldenberg RL, Mueller-Heubach E,  
4  
5 630 Moawad A, Sibai BM, Caritis SN, Miodovnik M, Paul RH, Dombrowski MP,  
6  
7 631 Thurnau G, McNellis D; National Institute of Child Health and Human  
8  
9 632 Development Network of Maternal-Fetal Medicine Units. Frequency of uterine  
10  
11 633 contractions and the risk of spontaneous preterm delivery. *N Engl J Med.* 2002 ;  
12  
13 634 346: 250-5.  
14  
15  
16 635  
17  
18 636 20. Abbas AE, Lester SJ, Connolly H. Pregnancy and the cardiovascular  
19  
20 637 system. *Int J Cardiol.* 2005; 98: 179-89.  
21  
22 638  
23  
24 639 21. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit  
25  
26 640 time. *Scand J Gastroenterol.* 1997; 32: 920-4.  
27  
28 641  
29  
30 642 22. Riegler G, Esposito I. Bristol scale stool form. A still valid help in medical  
31  
32 643 practice and clinical research. *Tech Coloproctol* 2001; 5: 163-4  
33  
34 644  
35  
36 645 23. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller  
37  
38 646 RC. Functional Bowel Disorders. *Gastroenterology* 2006; 130: 1480-91.  
39  
40 647  
41  
42 648 24. Koren G, Boskovic R, Hard M, Maltepe C, Navioz Y, Einarson A.  
43  
44 649 Motherisk-PUQE (pregnancy-unique quantification of emesis and nausea)  
45  
46 650 scoring system for nausea and vomiting of pregnancy. *Am J Obstet Gynecol.*  
47  
48 651 2002;186: S228-31.  
49  
50 652  
51  
52 653 25. Koren G, Piwko C, Ahn E, Boskovic R, Maltepe C, Einarson A, et al.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 654 Validation studies of the Pregnancy Unique-Quantification of Emesis (PUQE)  
4  
5 655 scores. J Obstet Gynaecol. 2005; 25: 241-4.  
6  
7 656  
8  
9  
10 657 26. Pearson JF, Weaver JB. Fetal activity and fetal wellbeing: an evaluation. Br  
11  
12 658 Med J 1976; 1:1305–7.  
13  
14 659  
15  
16 660 27. Winje BA, Saastad E, Gunnes N, Tveit JV, Stray-Pedersen B, Flenady V, et  
17  
18 661 al. Analysis of 'count-to-ten' fetal movement charts: a prospective cohort  
19  
20 662 study. BJOG. 2011;118: 1229-38.  
21  
22 663  
23  
24 664 28. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach  
25  
26 665 Notification rules under the Health Information Technology for Economic and  
27  
28 666 Clinical Health Act and the Genetic Information Nondiscrimination Act; other  
29  
30 667 modifications to the HIPAA rules. Fed Regist. 2013; 78: 5565-702.  
31  
32 668  
33  
34 669 29. Health Insurance Portability and Accountability Act of 1996. Public Law 104-  
35  
36 670 191. US Statut Large. 1996; 110:1936-2103.  
37  
38 671  
39  
40 672 30. Amended Act on the Protection of Personal Information.  
41  
42 673 [https://www.ppc.go.jp/files/pdf/Act\\_on\\_the\\_Protection\\_of\\_Personal\\_Informatio](https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Informatio)  
43  
44 674 [n.pdf](https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Informatio).  
45  
46 675  
47  
48  
49 676 31. Takai-Igarashi T, Kinoshita K, Nagasaki M, Ogishima S, Nakamura N,  
50  
51 677 Nagase S, et al. Security controls in an integrated Biobank to protect privacy in  
52  
53 678 data sharing: rationale and study design. BMC Med Inform Decis Mak. 2017;  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 679 17:100.  
4  
5 680  
6  
7 681 32. Katsuoka F, Yokozawa J, Tsuda K, Ito S, Pan X, Nagasaki M, et al. An  
8  
9 682 efficient quantitation method of next-generation sequencing libraries by using  
10  
11 683 MiSeq sequencer. *Anal Biochem.* 2014; 466: 27-9.  
12  
13 684  
14  
15 685 33. Koshihara S, Motoike I, Kojima K, Hasegawa T, Shirota M, Saito T, et al.  
16  
17 686 The structural origin of metabolic quantitative diversity. *Sci Rep.* 2016; 6:  
18  
19 687 31463.  
20  
21 688  
22  
23 689 34. Nishiumi S, Kobayashi T, Ikeda A, Yoshie T, Kibi M, Izumi Y, et al. A novel  
24  
25 690 serum metabolomics-based diagnostic approach for colorectal cancer. *PLoS*  
26  
27 691 *One.* 2012; 7: e40459.  
28  
29 692  
30  
31 693 35. Nishiumi S, Kobayashi T, Kawana S, Unno Y, Sakai T, Okamoto K, et al.  
32  
33 694 Investigations in the possibility of early detection of colorectal cancer by gas  
34  
35 695 chromatography/triple-quadrupole mass spectrometry. *Oncotarget.* 2017; 8,  
36  
37 696 17115-17126.  
38  
39 697  
40  
41 698 36. Sato Y, Yamagishi J, Yamashita R, Shinozaki N, Ye B, Yamada T, et al.  
42  
43 699 Inter-Individual Differences in the Oral Bacteriome Are Greater than Intra-Day  
44  
45 700 Fluctuations in Individuals. *PLoS One.* 2015;10: e0131607.  
46  
47 701  
48  
49 702 37. Watanabe K, Naruse K, Tanaka K, Metoki H, Suzuki Y. Outline of definition  
50  
51 703 and classification of pregnancy induced hypertension (PIH). *Hypertens Res*



- 1  
2  
3 704 Pregnancy 2013; 1: 3–4.  
4  
5 705  
6  
7 706 38. IADPSG Consensus Panel: International association of diabetes and  
8 pregnancy study groups recommendations on the diagnosis and classification  
9 of hyperglycemia in pregnancy. Diabetes Care 2010; 33: 676-682.  
10  
11 707  
12 708  
13  
14 709  
15  
16 710 39. #PRSice PRSice: Polygenic Risk Score software. Euesden J, Lewis CM,  
17 O'Reilly PF, Bioinformatics, 2015; 31:1466-8.  
18  
19 711  
20 712  
21  
22 713 40. Wax JR, Cartin A, Pinette MG. Biophysical and Biochemical Screening for  
23 the Risk of Preterm Labor: An Update. Clin Lab Med. 2016; 36: 369-83.  
24  
25 714  
26  
27 715  
28  
29 716 41. Al-Rubaie Z, Askie LM, Ray JG, Hudson HM, Lord SJ. The performance of  
30 risk prediction models for pre-eclampsia using routinely collected maternal  
31 characteristics and comparison with models that include specialised tests and  
32 with clinical guideline decision rules: a systematic review. BJOG. 2016;  
33 123:1441-1452.  
34  
35 718  
36 719  
37 720  
38 721  
39  
40 722 42. Koullali B, Oudijk MA, Nijman TA, Mol BW, Pajkrt E. Risk assessment and  
41 management to prevent preterm birth. Semin Fetal Neonatal Med. 2016 ;21:  
42 80-8.  
43  
44 723  
45 724  
46  
47 725  
48  
49 726 43. Henderson JT, Thompson JH, Burda BU, Cantor A. Preeclampsia  
50 Screening: Evidence Report and Systematic Review for the US Preventive  
51 Services Task Force. JAMA. 2017; 317: 1668-1683.  
52  
53 727  
54 728  
55  
56  
57  
58  
59  
60

- 1  
2  
3 729  
4  
5 730 44. Broekhuijsen K, van Baaren GJ, van Pampus MG, Ganzevoort W, Sikkema  
6  
7 731 JM, Woiski MD, et al; HYPITAT-II Study Group. Immediate delivery versus  
8  
9 732 expectant monitoring for hypertensive disorders of pregnancy between 34  
10  
11 733 and 37 weeks of gestation (HYPITAT-II): an open-label, randomised  
12  
13 734 controlled trial. *Lancet*. 2015; 385: 2492-2501.  
14  
15  
16 735  
17  
18 736 45. Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, McDonald  
19  
20 737 DT, Kusebauch U, Moss CL, Zhou Y, Qin S, Moritz RL, Brogaard K, Omenn  
21  
22 738 GS, Lovejoy JC, Hood L. A wellness study of 108 individuals using personal,  
23  
24 739 dense, dynamic data clouds. *Nat Biotechnol*. 2017; 35: 747-756.  
25  
26  
27 740  
28  
29 741 46. GTEx Consortium; Laboratory, Data Analysis & Coordinating Center  
30  
31 742 (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis  
32  
33 743 Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund;  
34  
35 744 NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source  
36  
37 745 Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core  
38  
39 746 Resource—VARI; Brain Bank Repository—University of Miami Brain  
40  
41 747 Endowment Bank; Leidos Biomedical—Project Management; ELSI Study;  
42  
43 748 Genome Browser Data Integration & Visualization—EBI; Genome Browser Data  
44  
45 749 Integration & Visualization—UCSC Genomics Institute, University of California  
46  
47 750 Santa Cruz; Lead analysts;; Laboratory, Data Analysis & Coordinating Center  
48  
49 751 (LDACC);; NIH program management;; Biospecimen collection;; Pathology;;  
50  
51 752 eQTL manuscript working group.; Battle A, Brown CD, Engelhardt BE,  
52  
53 753 Montgomery SB. Genetic effects on gene expression across human tissues.  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 754 Nature. 2017; 550: 204-213.

4  
5 755

6  
7 756

8  
9  
10 757 **Figure titles and legends**

11 758

12  
13  
14 759 **Figure 1. Flowchart of Maternity Log Study (MLOG) participants**

15 760

16  
17  
18 761 **Figure 2. Overview of the MLOG study protocol**

19 762 **A:** Participant timeline for the MLOG study.

20  
21  
22 763 **B:** Physiologic information collected using healthcare devices. Specific  
23  
24 764 measures were uploaded each day from the time of enrollment (solid horizontal  
25  
26 765 lines). Participants had the option to continue uploading data until 180 days  
27  
28 766 after delivery (dashed horizontal lines).

29  
30  
31 767 **C:** Daily lifelogs of self-reported information using a smartphone application.  
32  
33 768 Basic lifelog information was input manually from the time of enrollment (solid  
34  
35 769 horizontal lines). Participants had the option to continue uploading data until  
36  
37 770 180 days after delivery (dashed horizontal lines). Fetal movement and uterine  
38  
39 771 contractions were recorded from 24 and 20 weeks of gestation, respectively.

40  
41  
42 772

43  
44 773 **Figure 3. Data acquisition rate**

45  
46 774 The mean data upload rate of specific measures was calculated from the total  
47  
48 775 days of actual uploads divided by the number of days from enrollment to  
49  
50 776 delivery in each participant.

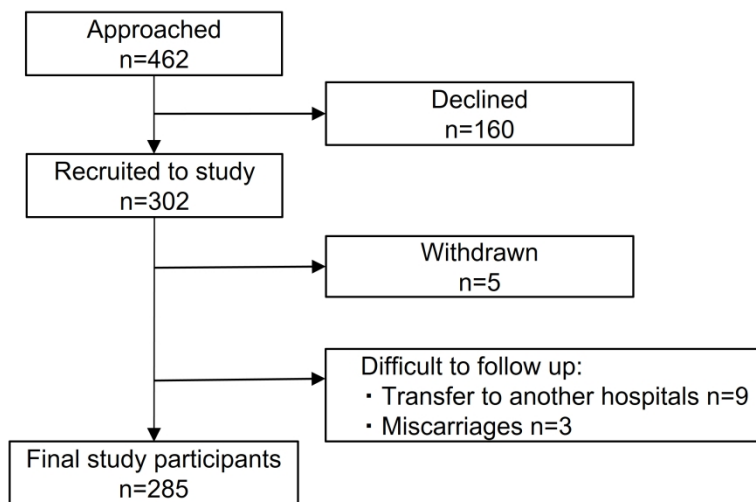


Figure 1.

254x190mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

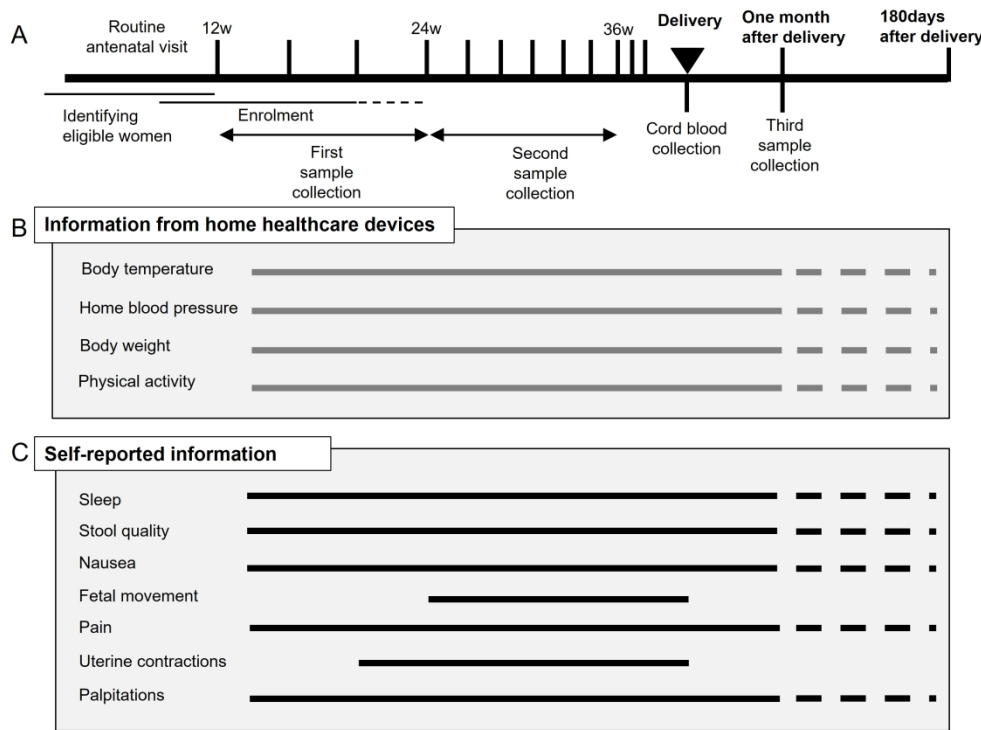


Figure 2.

254x190mm (300 x 300 DPI)

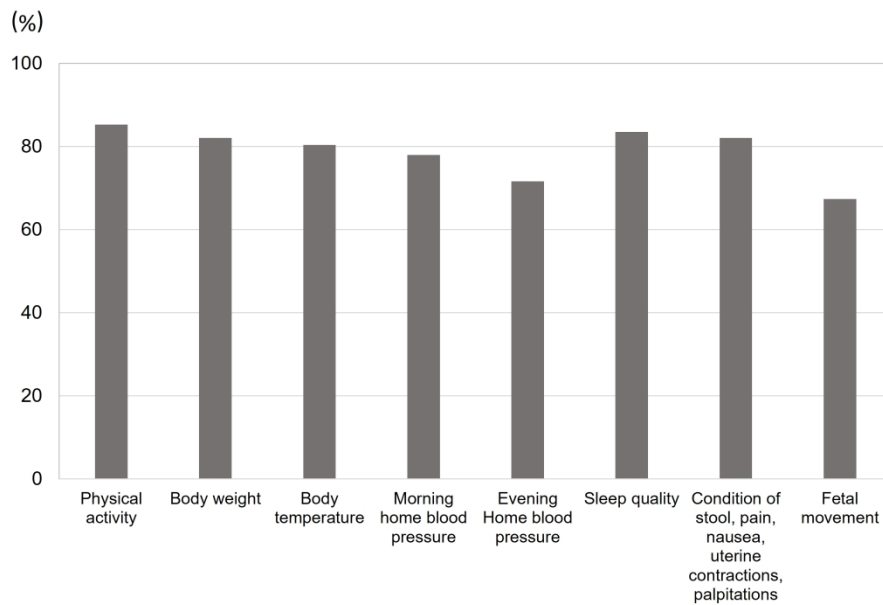


Figure 3.

254x190mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# BMJ Open

## Cohort Profile: Maternity Log Study: protocol for a longitudinal lifelog monitoring and multi-omics analysis for the early prediction of complicated Pregnancy

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-025939.R1
Article Type:	Cohort profile
Date Submitted by the Author:	09-Nov-2018
Complete List of Authors:	<p>Sugawara, Junichi; Tohoku Medical Megabank Organization, Tohoku University, Feto-Maternal Medical Science; Tohoku University Graduate School of Medicine, Obstetrics and Gynecology  Ochi, Daisuke; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC.  Yamashita, Riu; Tohoku Medical Megabank Organization, Tohoku University  Yamauchi, Takafumi ; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC.  Saigusa, Daisuke; Tohoku Medical Megabank Organization, Tohoku University  Wagata, Maiko; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine  Obara, Taku; Tohoku Medical Megabank Organization, Tohoku University  Ishikuro, Mami; Tohoku Medical Megabank Organization, Tohoku University  Tsunemoto, Yoshiki; Research Laboratories, NTT DOCOMO, INC.  Harada, Yuki; Tohoku Medical Megabank Organization, Tohoku University  Shibata, Tomoko; Tohoku Medical Megabank Organization, Tohoku University  Mimori, Takahiro ; Tohoku Medical Megabank Organization, Tohoku University  Kawashima, Junko; Tohoku Medical Megabank Organization, Tohoku University  Katsuoka, Fumiki; Tohoku Medical Megabank Organization, Tohoku University  Igarashi-Takai, Takako ; Tohoku Medical Megabank Organization, Tohoku University  Ogishima, Soichi; Tohoku Medical Megabank Organization, Tohoku University  Metoki, Hirohito; Tohoku Medical and Pharmaceutical University  Hashizume, Hiroaki; Tohoku Medical Megabank Organization, Tohoku University  Fuse, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine  Minegishi, Naoko; Tohoku Medical Megabank Organization, Tohoku University  Koshiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University</p>

	Tanabe, Osamu; Tohoku Medical Megabank Organization, Tohoku University; Radiation Effects Research Foundation Kuriyama, Shinichi; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Kinoshita, Kengo; Tohoku Medical Megabank Organization, Tohoku University Kure, Shigeo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Yaegashi, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; TUH Yamamoto, Masayuki; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Hiyama, Satoshi; Research Laboratories, NTT DOCOMO, INC. Nagasaki, Masao; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine
<b>Primary Subject Heading</b>:	Obstetrics and gynaecology
Secondary Subject Heading:	Health informatics
Keywords:	lifelog, multi-omics analysis, prediction, complicated pregnancy

SCHOLARONE™  
Manuscripts



1 **Cohort profile: Maternity Log Study: protocol for a**  
2 **longitudinal lifelog monitoring and multi-omics analysis**  
3 **for the early prediction of complicated pregnancy**

4  
5 Junichi Sugawara<sup>1,2\*</sup>, Daisuke Ochi<sup>1,3</sup>, Riu Yamashita<sup>1</sup>, Takafumi Yamauchi<sup>1,3</sup>,  
6 Daisuke Saigusa<sup>1</sup>, Maiko Wagata<sup>1,2</sup>, Taku Obara<sup>1</sup>, Mami Ishikuro<sup>1</sup>, Yoshiki  
7 Tsunemoto<sup>3</sup>, Yuki Harada<sup>1</sup>, Tomoko F. Shibata<sup>1</sup>, Takahiro Mimori<sup>1</sup>, Junko  
8 Kawashima<sup>1</sup>, Fumiki Katsuoka<sup>1</sup>, Takako Igarashi-Takai<sup>1</sup>, Soichi Ogishima<sup>1</sup>,  
9 Hirohito Metoki<sup>4</sup>, Hiroaki Hashizume<sup>1</sup>, Nobuo Fuse<sup>1,2</sup>, Naoko Minegishi<sup>1</sup>, Seizo  
10 Koshiba<sup>1</sup>, Osamu Tanabe<sup>1,5</sup>, Shinichi Kuriyama<sup>1,2</sup>, Kengo Kinoshita<sup>1</sup>, Shigeo  
11 Kure<sup>1,2</sup>, Nobuo Yaegashi<sup>1,6</sup>, Masayuki Yamamoto<sup>1,2</sup>, Satoshi Hiyama<sup>3</sup>, and  
12 Masao Nagasaki<sup>1,2\*</sup>.

13  
14 1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryomachi,  
15 Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

16 2 Tohoku University Graduate School of Medicine, 1-1, Seiryomachi, Aoba-ku,  
17 Sendai, Miyagi, 980-8574, Japan.

18 3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,  
19 Kanagawa, Japan 239-8536.

20 4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-  
21 ku, Sendai, Miyagi 981-8558, Japan.

22 5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku,  
23 Hiroshima 732-0815, Japan.

24 6 Tohoku University Hospital, 1-1, Seiryomachi, Aoba-ku, Sendai, Miyagi, 980-

1  
2  
3  
4 25 8574, Japan.  
5  
6 26  
7  
8 27  
9

10 28 Email addresses:

11  
12  
13 29 \*Junichi Sugawara: jsugawara@med.tohoku.ac.jp  
14

15 30 Daisuke Ochi: ochi@megabank.tohoku.ac.jp  
16

17 31 Riu Yamashita: ryamasi@megabank.tohoku.ac.jp  
18

19 32 Takafumi Yamauchi: t.yamauchi@megabank.tohoku.ac.jp  
20

21 33 Daisuke Saigusa: saigusa@m.tohoku.ac.jp  
22

23 34 Maiko Wagata: wagata@med.tohoku.ac.jp  
24

25 35 Taku Obara: obara-t@hosp.tohoku.ac.jp  
26

27 36 Mami Ishikuro: m\_ishikuro@med.tohoku.ac.jp  
28

29 37 Yoshiki Tsunemoto: yoshiki.tsunemoto@megabank.tohoku.ac.jp  
30

31 38 Yuki Harada: harada@megabank.tohoku.ac.jp  
32

33 39 Tomoko F. Shibata: tshibata@megabank.tohoku.ac.jp  
34

35 40 Takahiro Mimori: mimori@megabank.tohoku.ac.jp  
36

37 41 Junko Kawashima: kawashima@dent.tohoku.ac.jp  
38

39 42 Fumiki Katsuoka: kfumiki@med.tohoku.ac.jp  
40

41 43 Takako Igarashi-Takai: takai@megabank.tohoku.ac.jp  
42

43 44 Soichi Ogishima: ogishima@megabank.tohoku.ac.jp  
44

45 45 Hirohito Metoki: hmetoki@tohoku-mpu.ac.jp  
46

47 46 Hiroaki Hashizume: hashizume@megabank.tohoku.ac.jp  
48

49 47 Nobuo Fuse: fusen@megabank.tohoku.ac.jp  
50

51 48 Naoko Minegishi: nmine@med.tohoku.ac.jp  
52

53 49 Seizo Koshiba: koshiba@megabank.tohoku.ac.jp  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 50 Osamu Tanabe: otanabe@rerf.or.jp

5  
6 51 Shinichi Kuriyama: kuriyama@med.tohoku.ac.jp

7  
8 52 Kengo Kinoshita: kengo@ecei.tohoku.ac.jp

9  
10 53 Shigeo Kure: kure@med.tohoku.ac.jp

11  
12 54 Nobuo Yaegashi: yaegashi@med.tohoku.ac.jp

13  
14 55 Masayuki Yamamoto: masiyamamoto@med.tohoku.ac.jp

15  
16 56 Satoshi Hiyama: hiyamas@nttdocomo.com

17  
18 57 \*Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp

19  
20  
21  
22 58

23  
24 59 \*Corresponding authors:

25  
26 60 Junichi Sugawara: jsugawara@med.tohoku.ac.jp

27  
28 61 and Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp

29  
30 62 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryomachi,

31  
32 63 Aoba-ku, 980-8573, Sendai, Japan. Phone: +81-22-273-6283

33  
34  
35  
36 64

37  
38 65 Word count: 5314 words

39  
40 66 Key words: lifelog, multi-omics analysis, prediction, complicated pregnancy

41  
42  
43 67

44  
45 68

46  
47 69

48  
49 70

50  
51 71

52  
53 72

54  
55 73

56  
57 74

58  
59  
60

1  
2  
3  
4 75 **Abstract**

5  
6 76 **Purpose:** A prospective cohort study for pregnant women, the Maternity Log  
7  
8 77 study (MLOG), was designed to construct a time-course high-resolution  
9  
10 78 reference catalog of bioinformatic data in pregnancy and explore the  
11  
12 79 associations between genomic and environmental factors and the onset of  
13  
14 80 pregnancy complications, such as hypertensive disorders of pregnancy,  
15  
16 81 gestational diabetes mellitus, and preterm labor, using continuous lifestyle  
17  
18 82 monitoring combined with multi-omics data on the genome, transcriptome,  
19  
20 83 proteome, metabolome, and microbiome.

21  
22 84 **Participants:** Pregnant women were recruited at the timing of first routine  
23  
24 85 antenatal visits at Tohoku University Hospital, Sendai, Japan between  
25  
26 86 September 2015 and November 2016. Of the eligible women who were invited,  
27  
28 87 65.4% agreed to participate, and a total of 302 women were enrolled. The  
29  
30 88 inclusion criteria were age  $\geq$  20 years and the ability to access the internet using  
31  
32 89 a smartphone in the Japanese language.

33  
34 90 **Findings to date:**

35  
36 91 Study participants uploaded daily general health information including quality of  
37  
38 92 sleep, condition of bowel movements, and the presence of nausea, pain, and  
39  
40 93 uterine contractions. Participants also collected physiologic data, such as body  
41  
42 94 weight, blood pressure, heart rate, and body temperature, using multiple home  
43  
44 95 healthcare devices. The mean upload rate for each lifelog item was ranging  
45  
46 96 from 67.4 % (fetal movement) to 85.3% (physical activity) and the total number  
47  
48 97 of data points was over 6 million. Biospecimens, including maternal plasma,  
49  
50 98 serum, urine, saliva, dental plaque, and cord blood, were collected for multi-  
51  
52 99 omics analysis.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 100 **Future plans:**

5  
6 101 Lifelog and multi-omics data will be used to construct a time-course high-  
7  
8 102 resolution reference catalog of pregnancy. The reference catalog will allow us to  
9  
10 103 discover relationships among multi-dimensional phenotypes and novel risk  
11  
12 104 markers in pregnancy for the future personalized early prediction of pregnancy  
13  
14 105 complications.  
15  
16

17 106

18  
19  
20 107 **Strengths and limitations of this study:**

- 21  
22 108 ■ This is the first study designed to collect longitudinal lifelog information  
23  
24 109 through healthcare devices, self-administered questionnaires using  
25  
26 110 smartphones, and varieties of biospecimens throughout pregnancy.  
27  
28 111 ■ Longitudinal, continuous, individual lifelog data with a high acquisition rate  
29  
30 112 will enable us to assess dynamic physiological changes throughout  
31  
32 113 pregnancy. Multi-omics data will make it possible to understand the complex  
33  
34 114 mechanisms of multifactorial pregnancy-related diseases.  
35  
36 115 ■ Potential limitations of the present study are as follows: 1) the limited  
37  
38 116 sample size, and 2) participant recruitment only at a tertiary hospital for  
39  
40 117 high-risk populations. Therefore, the results might not be applicable to the  
41  
42 118 general populations.  
43  
44 119 ■ Inclusion criteria of the present study limited the eligibility to pregnant  
45  
46 120 women with age >20 years and the ability to access the internet using a  
47  
48 121 smartphone. Therefore, results of the present study might not be applicable  
49  
50 122 to pregnancies with lower coverage of smartphone use.  
51  
52 123 ■ Further study with a larger sample size of general populations is needed to  
53  
54 124 validate a reference catalog of normal pregnancy and a prediction model of  
55  
56  
57  
58  
59  
60

1		
2		
3		
4	125	pregnancy complications.
5		
6	126	
7		
8	127	
9		
10	128	
11		
12		
13	129	
14		
15	130	
16		
17	131	
18		
19	132	
20		
21	133	
22		
23	134	
24		
25	135	
26		
27	136	
28		
29	137	
30		
31	138	
32		
33	139	
34		
35	140	
36		
37	141	
38		
39	142	
40		
41	143	
42		
43	144	
44		
45	145	
46		
47	146	
48		
49	147	
50		
51	148	
52		
53	149	
54		
55		
56		
57		
58		
59		
60		

## 150 INTRODUCTION

151 The incidence of pregnancy-related disorders, including hypertensive disorders  
152 of pregnancy (HDP), gestational diabetes mellitus (GDM), and preterm delivery  
153 has been increasing worldwide [1-4]. These multifactorial conditions are caused  
154 by an interaction of genetic factors and environmental factors [5,6]. Recent  
155 reports suggest that continuous lifestyle monitoring using wearable biosensors  
156 provides important information on latent physiologic changes that are exhibited  
157 prior to the onset of disease [7]. Using these monitors, environmental factors  
158 may be estimated more accurately than by using conventional questionnaires.

159 For these reasons, we have designed a prospective cohort study for pregnant  
160 women, the Maternity Log study (MLOG). In this study, pregnant women upload  
161 daily information and physiologic data using multiple home healthcare devices.  
162 In addition, a variety of biospecimens are collected for multi-omics analysis.

163 To the best of our knowledge, this study will be the first to integrate multi-  
164 omics analyses and objective data on environmental factors, including daily  
165 lifelog data, in pregnant women. This study may demonstrate correlations  
166 between specific lifelog patterns and pregnancy related physiological changes,  
167 such as blood pressure, gestational weight gain, and onset of obstetric diseases.  
168 Furthermore, studies on associations among lifelog patterns, plasma and urine  
169 metabolomes, transcriptomes, and genomic variations may reveal relationships  
170 among multi-dimensional phenotypes, and lead to identification of novel risk  
171 markers in pregnancy for the future personalized early prediction of pregnancy  
172 complications, e.g. hypertensive disorders of pregnancy, gestational diabetes,  
173 and preterm labor.

174

## 175 **COHORT DESCRIPTION**

### 176 **Study setting**

177 The aim of the MLOG study is to construct a time-course high-resolution  
178 reference catalog of bioinformatic data in pregnancy and thereby develop  
179 methods for early prediction of obstetric complications, through integrated  
180 analysis of daily lifelogs and multi-omics data, *i.e.*, maternal genomes,  
181 transcriptomes, metabolomes, and oral microbiomes.

182 The MLOG study is a prospective, add-on cohort study, built on a birth- and 3-  
183 generation cohort study established by the Tohoku Medical Megabank  
184 Organization (TMM BirThree Cohort Study) [8] in order to elucidate the  
185 mechanisms of complicated multifactorial diseases in mothers and children in  
186 the wake of the Great East Japan Earthquake in 2011. Epidemiological data  
187 from extensive questionnaire surveys and accurate clinical records, including  
188 birth outcomes, can be abstracted from the integrated biobank of the Tohoku  
189 Medical Megabank Organization (ToMMo) [8]. TMM BirThree Cohort Study was  
190 started in July 2013 in one obstetric clinic and expanded throughout Miyagi  
191 Prefecture, and approximately 50 obstetric clinics and hospitals (including  
192 Tohoku University Hospital) participated in the recruiting process. We planned  
193 to recruit 20,000 pregnant women as probands, and her family members from  
194 three generations, a total of over 70,000 participants [8]. Written informed  
195 consent was obtained from all participants by the genome medical research  
196 coordinators (GMRCs). The MLOG study was conducted in accordance with the  
197 Declaration of Helsinki and approved by the ethics committees of Graduate  
198 School of Medicine (2014-1-704) and ToMMo (22017-1-085), Tohoku University  
199 under a collaborative research agreement among ToMMo, Tohoku University



1  
2  
3  
4 200 and NTT DoCoMo, Inc. (Tokyo, Japan).  
5  
6

7 201

## 8 202 **Patient and public involvement**

9  
10 203 Patients or the public were not directly involved in the development of the  
11  
12 204 research question or the design of the study. The main results will be made  
13  
14 205 available in the public domain.  
15  
16

17 206

## 18 207 **Participants**

19  
20 208 Participants were recruited at a first routine antenatal visit at Tohoku University  
21  
22 209 Hospital, Sendai, Japan between September 2015 and November 2016. A  
23  
24 210 flowchart of the recruitment process is shown in Figure 1. GMRCs at Tohoku  
25  
26 211 University Hospital approached eligible pregnant women for TMM BirThree  
27  
28 212 Cohort Study (n= 631), and patients who already agreed to participate in TMM  
29  
30 213 BirThree Cohort Study (n=513) were assessed for eligibility for the MLOG study.  
31  
32 214 Finally, 462 pregnant women were asked to provide informed consent for the  
33  
34 215 MLOG study. A total of 302 women were enrolled. The inclusion criteria were  
35  
36 216 the age  $\geq 20$  years and the ability to access the internet using a smartphone in  
37  
38 217 the Japanese language. Participants were excluded after enrollment if  
39  
40 218 termination of pregnancy, abortion, or transfer to another institution for  
41  
42 219 emergency care occurred before delivery, or if they withdrew consent for any  
43  
44 220 reason.  
45  
46

47 221

## 48 222 **Outline of study protocol**

49  
50 223 The study protocol consisted of blood and urine sampling, saliva and dental  
51  
52 224 plaque sampling, self-administered daily lifelog data collection, and data upload  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 225 from multiple healthcare devices through a smartphone. An overview of the  
5  
6 226 protocol is provided in Figure 2. In Japan, routine antenatal visits, including  
7  
8 227 ultrasounds, are scheduled every 4 weeks from early pregnancy (< 12 weeks)  
9  
10 228 to 23 weeks of gestation, every 2 weeks from 24 to 35 weeks, and every week  
11  
12 229 from 36 weeks to delivery [9]. Lifelog data collection was continued throughout  
13  
14 230 pregnancy and until 1 month after delivery. Optional data collection could be  
15  
16 231 continued up to 180 days after delivery.  
17  
18  
19  
20  
21

232

### 233 **Blood and urine sampling**

234 Blood samples were collected 3 times from each participant; the first sample  
25  
26 235 was collected between 12 and 24 weeks of gestation, the second between 24  
27  
28 236 and 36 weeks, and the third at 1 month after delivery. A maximum of 13 mL of  
29  
30 237 blood was collected each time, from which serum and plasma were separated  
31  
32 238 to be stored at -80°C until the time of analysis. An aliquot of blood (2.5 mL) was  
33  
34 239 stored in a PAXgene® tube (Becton, Dickinson and Company, Franklin Lakes,  
35  
36 240 NJ, USA) at -80°C until the time of RNA extraction for transcriptome analysis.  
37  
38 241 Genomic DNA was extracted from mononuclear cells using an Autopure®  
39  
40 242 extractor (Qiagen, Venlo, The Netherlands). Approximately 10 mL of cord blood  
41  
42 243 was collected from the umbilical vein in a PAXgene® tube for storage at -80°C,  
43  
44 244 and in an EDTA 2K tube (Becton, Dickinson and Company, Franklin Lakes, NJ,  
45  
46 245 USA) for separation of plasma to be stored at -80°C. Urine samples (10 mL)  
47  
48 246 were collected at each antenatal visit; when participants were admitted to the  
49  
50 247 hospital ward, urine was collected once weekly. Urine samples were  
51  
52 248 immediately transferred and stored at -80°C until the time of analysis.  
53  
54  
55  
56  
57  
58  
59 249

## 250 **Saliva and dental plaque sampling**

251 Samples of saliva and dental plaque were collected 3 times from each  
252 participant, at the same time points as blood collection. Approximately 3 mL of  
253 saliva was collected using a 50-mL conical centrifuge tube (Corning, Inc.,  
254 Corning, NY, USA) and stored at -80°C until analysis. Dental plaque was  
255 sampled by brushing, suspended in 0.5 mL of Tris-EDTA (10 mM Tris, 1 mM  
256 EDTA; pH, 8.0), and immediately stored at -80°C until the time of sample  
257 processing.

258

## 259 **Lifelog data collection**

260 Based on previous publications on the utility for risk assessment of pregnancy-  
261 related diseases, we selected several lifelog parameters to employ in this study,  
262 *i.e.*, body temperature [10], home blood pressure [11], body weight [12], and  
263 physical activity (calorie expenditure) [13], as well as self-administered  
264 information such as sleep quality [14], condition of stool [15], severity of nausea  
265 [16], fetal movement [17], severity of pain [18], uterine contractions [19], and  
266 palpitations [20]. Body temperature, home blood pressure, body weight, and  
267 physical activity were uploaded from multiple healthcare devices through a  
268 smartphone. The self-administered information described above was input  
269 manually on mobile applications created for this study.

270 Data collection was started after obtaining informed consent and after giving  
271 detailed instructions for the use of the healthcare devices. These applications  
272 tracked quality of sleep; condition of stool using the Bristol Scale [21-23];  
273 severity of nausea using the Pregnancy-Unique Quantification of Emesis and  
274 nausea (PUQE) score [24,25]; headache, toothache, lumbago, and upper and

1  
2  
3  
4 275 lower abdominal pain using a numerical rating scale (NRS) score; the number  
5  
6 276 of perceived uterine contractions; palpitations; and fetal movement using a  
7  
8 277 modified count-to-10 fetal movement chart [26,27].  
9

10 278 Sleep quality was evaluated by the wakeup time, bedtime, sleep satisfaction  
11  
12 279 (ranked from satisfied to poor using a numeric scale of 0-4), and the number of  
13  
14 280 nocturnal awakenings (0-6).

15  
16  
17 281 The Bristol stool form scale was originally developed to assess constipation  
18  
19 282 and diarrhea [21, 22], and its use has been spread widely to evaluate functional  
20  
21 283 bowel disorders [22]. Using the Bristol scale, stool is classified into 7 types  
22  
23 284 according to cohesion and surface cracking [21, 22].  
24  
25

26 285 The PUQE score [24, 25] was developed to estimate the severity of nausea  
27  
28 286 and vomiting in pregnancy and quantifies the number of daily vomiting and  
29  
30 287 retching episodes and the length of nausea in hours (over the preceding 12 h).  
31  
32 288 The total score ranges from 3 (no symptoms) to 15, and higher scores are  
33  
34 289 correlated with increasing severity of nausea and vomiting [24, 25].  
35  
36  
37

38 290 In the NRS score for headache, toothache, lumbago, and upper and lower  
39  
40 291 abdominal pain, the total score ranges from 0 (no pain) to 10 (maximum ever  
41  
42 292 experienced).  
43  
44

45 293 Uterine contractions and palpitations were evaluated using definitions  
46  
47 294 determined for the current study. Uterine contractions were assessed using the  
48  
49 295 number of perceived contractions per day, ranging from 0 to more than 5. The  
50  
51 296 count-to-10 method was originally developed to assess fetal well-being by  
52  
53 297 recording the time, in minutes, required to count 10 fetal movements [26]. More  
54  
55 298 recently, a modified count-to-10 method has been proposed: pregnant women  
56  
57 299 are advised to start counting when they feel the first movement, then record the  
58  
59  
60

1  
2  
3  
4 300 time required to perceive an additional 9 movements [27]. Pregnant women are  
5  
6 301 encouraged to select a 2-hour period when they feel active fetal movements  
7  
8 302 and are instructed to count kicking and rolling movements in a favorable  
9  
10 303 maternal position after 24 weeks of gestation.

11  
12  
13 304 The applications also collected dietary logs and the medications taken on the  
14  
15 305 day before and the day of the antenatal visit, on which blood or urine samples  
16  
17 306 were collected.

18  
19  
20 307 Daily home blood pressure, body weight, body temperature, and physical  
21  
22 308 activity were measured as described below with home healthcare devices, and  
23  
24 309 uploaded through wireless communications using mobile applications on a  
25  
26 310 smartphone. Daily home blood pressure was measured twice daily using an  
27  
28 311 HEM-7510 monitor (OMRON Healthcare Co., Ltd., Kyoto, Japan): within 1 hour  
29  
30 312 of awakening in the morning and just before going to bed at night. Body weight  
31  
32 313 was measured using an HBF-254C meter (OMRON Healthcare Co., Ltd.) once  
33  
34 314 daily within 1 hour of awakening in the morning. Daily body temperature was  
35  
36 315 evaluated using an MC-652LC digital thermometer (MC-652LC; OMRON  
37  
38 316 Healthcare Co., Ltd.) just after awakening. Physical activity was assessed using  
39  
40 317 an HJA-403C pedometer (HJA-403C; OMRON Healthcare Co., Ltd.) to count  
41  
42 318 steps and calculate calorie expenditure.

43  
44  
45  
46  
47 319

## 48 49 320 **Clinical and epidemiological information**

50  
51  
52 321 Baseline clinical information and maternal and neonatal outcomes (*e.g.*,  
53  
54 322 maternal age, clinical data and findings from each antenatal visit, gestational  
55  
56 323 age at delivery, type of delivery, birth weight, and maternal and fetal  
57  
58 324 complications) were obtained from the medical records of Tohoku University  
59  
60

1  
2  
3  
4 325 Hospital. Epidemiological data, including extensive questionnaire surveys by  
5  
6 326 TMM BirThree Cohort Study can be obtained from the ToMMo integrated  
7  
8 327 biobank [8].  
9

10 328

### 11 329 **Database**

12  
13  
14  
15 330 A customized laboratory information management system (LIMS) was  
16  
17 331 established to track all biospecimens. All data were transferred to the TMM  
18  
19 332 integrated database after 2-step anonymisation in a linkable fashion.  
20  
21  
22 333 Data handling was strictly regulated under HIPAA (Health Insurance Portability  
23  
24 334 and Accountability Act of 1996, United States Security and Privacy Rules) [28,  
25  
26 335 29] and the Act on the Protection of Personal Information [30]. Security control  
27  
28 336 at our facility has been described previously [31].  
29  
30

31 337

### 32 338 **Omics analysis**

33  
34  
35  
36 339 Whole-genome sequencing  
37  
38 340 To minimize amplification bias, we adopted a PCR-free library preparation  
39  
40 341 method. After performing library quality control using the quantitative MiSeq  
41  
42 342 method [32], libraries were sequenced on HiSeq 2500 Sequencing System  
43  
44 343 (Illumina, Inc., San Diego, CA, USA) to generate 259-bp, paired-end reads. We  
45  
46 344 generated the sequencing data at over 12.5x coverage on average, and we  
47  
48 345 identified variants using the alignment tool BWA-MEM (ver. 0.7.5a-r405) with  
49  
50 346 the default option. Single nucleotide variants (SNVs) and indels were jointly  
51  
52 347 called across all samples using Genome Analysis Tool Kit's HaplotypeCaller  
53  
54 348 (ver. 3.8). Default filters were applied to SNV and indel calls using the GATK's  
55  
56 349 Variant Quality Score Recalibration (VQSR) approach. The human reference  
57  
58  
59  
60

1  
2  
3  
4 350 genome was GRCh37/hg19 with the decoy sequence (hs37d5) and NC\_007605  
5  
6 351 (Human Gamma Herpesvirus 4). The complete fasta file named  
7  
8 352 hg19\_tommo\_v2.fa is available from iJGVD website  
9  
10 353 (<http://ijgvd.megabank.tohoku.ac.jp>) [33]. For the quality assurance, we have  
11  
12 354 checked the ratio of the bases with the phred quality score over 30, the total  
13  
14 355 variant numbers in each chromosome, and the ratio of transitions to  
15  
16 356 transversions for a pair of sequences.  
17  
18  
19  
20  
21

## 22 358 Transcriptome

23  
24 359 Whole blood was collected using the PAXgene® RNA tube, which is widely  
25  
26 360 used for transcriptome analysis. After storage at -80°C, total RNA was purified  
27  
28 361 with PAXgene Blood RNA Kit® (Qiagen, Venlo, The Netherlands) using  
29  
30 362 QiaSymphony® (Qiagen). Total RNA was reverse-transcribed using an oligo-dT  
31  
32 363 primer. We used TruSeq DNA PCR-Free Library Preparation Kit (Illumina, Inc.)  
33  
34 364 for library preparation for sequencing with HiSeq 2500 Sequencing System. For  
35  
36 365 the quality assurance, we randomly selected 11 samples in one batch (usually  
37  
38 366 48 samples) and checked an RNA integrity number (RIN) (or an RIN equivalent)  
39  
40 367 using BioAnalyzer® or Tape Station® (both from Agilent Technologies, Santa  
41  
42 368 Clara, CA, USA). The batch with RIN (or an RIN equivalent) higher than 7.0 for  
43  
44 369 all tested samples was used for the downstream analysis. The minimum  
45  
46 370 threshold for the total sequence reads for each sample was set to thirty millions.  
47  
48 371 For computing a series of quality control metrics for RNA-seq data, RNA-SeQC  
49  
50 372 was used to check the quality of sequence reads [34].  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 374 Plasma and urine metabolome



1  
2  
3  
4 375 *Nuclear magnetic resonance (NMR) spectroscopy*

5  
6 376 All NMR measurements for metabolome analysis were conducted at 298 K on a  
7  
8 377 Bruker Avance 600 MHz spectrometer equipped with a SampleJet sample  
9  
10 378 changer (Bruker Corp., Billerica, MA, USA) [35]. Standard 1-dimensional  
11  
12 379 nuclear Overhauser enhancement spectroscopy (1D NOESY) and Carr-Purcell-  
13  
14 380 Meiboom-Gill (CPMG) spectra were obtained for each plasma or urine sample.  
15  
16 381 All spectra for plasma or urine samples were acquired using 16 scans and 32 k  
17  
18 382 of complex data points. All data were analyzed using the TopSpin 3.5 (Bruker  
19  
20 383 Corp.) and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta,  
21  
22 384 Canada) programs. All spectra were referenced to an internal standard (DSS-  
23  
24 385 d6). As necessary, those spectra were aligned using hierarchical cluster-based  
25  
26 386 peak alignment method, which is implemented as an R package called "speaq"  
27  
28 387 [36].  
29  
30  
31  
32

33 388

34  
35  
36 389 *Gas chromatography-tandem mass spectrometry (GC-MS/MS)*

37  
38 390 Sample preparation for plasma and urine (50 µL each) was performed using a  
39  
40 391 Microlab STARlet robot system (Hamilton, Reno, NV, USA) followed by the  
41  
42 392 methods previously reported by Nishiumi [37, 38]. The resulting deproteinized  
43  
44 393 and derivatized supernatant (1 µL) was subjected to GC-MS/MS, performed on  
45  
46 394 a GC-MS TQ-8040 system (Shimadzu Corp., Kyoto, Japan). The compound  
47  
48 395 separation was performed using a fused silica capillary column (BPX-5; 30 m ×  
49  
50 396 0.25 mm inner diameter; film thickness, 0.25 µm; Shimadzu Corp, Kyoto,  
51  
52 397 Japan). Metabolite detection was performed using Smart Metabolites Database  
53  
54 398 (Shimadzu Corp.) that contained the relevant multiple reaction monitoring  
55  
56 399 (MRM) method file and data regarding the GC analytical conditions, MRM  
57  
58  
59  
60



1  
2  
3  
4 400 parameters, and retention index employed for the metabolite measurement. The  
5  
6 401 database used in this study included data on 475 peaks from 334 metabolites.  
7  
8 402 All peaks of metabolites detected from each sample was annotated and  
9  
10 403 analyzed using Traverse MS® (Reifycs Inc., Tokyo, Japan). Then, two types of  
11  
12 404 normalization were performed to these annotated metabolites. The first  
13  
14 405 normalization was performed using the peak of 2-isopropylmalic acid as an  
15  
16 406 internal standard which was added to each sample before analysis with GC-  
17  
18 407 MS/MS. Then the second normalization was performed using quality control  
19  
20 408 (QC) samples which were injected after every 12 study samples according to  
21  
22 409 the RQC normalization methods [39]. Normalized values of each metabolite in  
23  
24 410 the QC samples were assessed by calculating coefficients of variation (CVs),  
25  
26 411 and metabolites with CVs over 20% were eliminated.  
27  
28  
29  
30  
31  
32

### 33 413 Oral Microbiome

34  
35 414 Analysis of oral microbiome was conducted by previously reported protocols  
36  
37 415 [40]. In brief, saliva was collected in a 50-mL tube. Dental plaque was sampled  
38  
39 416 by participants by brushing teeth with a sterilized toothbrush, and then  
40  
41 417 suspending it in 0.5 mL Tris-EDTA for collection. Both samples were stored at -  
42  
43 418 80°C until the time of processing. DNA was extracted from saliva and dental  
44  
45 419 plaque by standard glass bead-based homogenization and subsequent  
46  
47 420 purification with a silica-membrane spin-column using PowerSoil DNA Isolation  
48  
49 421 Kit (Mo Bio Laboratories, Carlsbad, CA, USA). DNA was eluted from the spin  
50  
51 422 column with 30- $\mu$ L RNase-free water (Takara Bio, Inc., Shiga, Japan), and  
52  
53 423 stored at -20°C after determining the amount and purity of DNA with a  
54  
55 424 Nanodrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).  
56  
57  
58  
59  
60

1  
2  
3  
4 425 Using DNA extracted from saliva or dental plaque as a template, a part of the  
5  
6 426 V4 variable region of the bacterial 16S rRNA gene was amplified by 2-step PCR.  
7  
8 427 Tag-indexed PCR products thus obtained were subjected to multiplex amplicon  
9  
10 428 sequencing using MiSeq System with MiSeq Sequencing Reagent Kit, v3  
11  
12 429 (Illumina, Inc.) according to the manufacturer's instructions. For the quality  
13  
14 430 assurance, the minimum threshold of the total sequence reads for each sample  
15  
16 431 was set to ten thousands, and the principal component analysis was used to  
17  
18 432 eliminate outliers.  
19  
20  
21  
22  
23

433

#### 434 **Outcomes**

435 The following obstetric complications represented the primary outcomes.  
436 Gestational age was confirmed by measuring fetal crown rump length from 9 to  
437 13 weeks of gestation using transvaginal ultrasound. HDP was defined as  
438 gestational hypertension, preeclampsia, superimposed preeclampsia, or chronic  
439 hypertension [41,42]. Preterm birth was defined as spontaneous preterm labor,  
440 medically induced preterm labor, or preterm premature rupture of membranes  
441 resulting in preterm birth at less than 37 weeks of gestation. GDM was  
442 diagnosed according to the International Association of the Diabetes and  
443 Pregnancy Study Groups (IADPSG) criteria [43]. The secondary outcomes were  
444 maternal body weight, blood pressure, physical activity, lifestyle changes,  
445 perinatal mental disorders, fetal growth, fetal movement, and birth weight.

446

#### 447 **Sample size calculation**

448 At this time, there is little reliable evidence to demonstrate how time-dependent  
449 trends of longitudinal dense data would differ by pregnancy outcomes.

1  
2  
3  
4 450 Therefore, a priori sample size calculation is not provided in the present study.  
5  
6 451 However, considering that one of the main purposes of the MLOG study is to  
7  
8 452 explore the relationship between patterns of longitudinal home blood pressure  
9  
10 453 and the onset of HDP, we estimated a required sample size as follows. Based  
11  
12 454 on the HDP incidence of approximately 10% at Tohoku University Hospital, with  
13  
14 455 a statistical power of 90% and a significance level of 5%, a sample of 250  
15  
16 456 participants is required to detect a 5-mmHg difference in average home blood  
17  
18 457 pressure (with a 7-mmHg standard deviation) in the HDP group. To allow for  
19  
20 458 15% attrition and withdrawals during pregnancy, a minimum of 300 participants  
21  
22 459 at baseline was required.  
23  
24  
25  
26  
27  
28

### 29 461 **Statistical analysis of longitudinal lifelog data**

30  
31 462 One of the major advantages of the MLOG study is the dense information for  
32  
33 463 each participant. Especially, time points for lifelog data collection are highly  
34  
35 464 dense for each participant. For these datasets, per-person analysis of dynamic  
36  
37 465 relationships between variables can be applied [44]. Vector autoregressive  
38  
39 466 (VAR) modeling is a promising solution to find the predicates for each outcome.  
40  
41 467 In addition, the Granger causality test can elucidate the temporal ordering of  
42  
43 468 dynamic relationship between two or more variables and indicate putative  
44  
45 469 causal associations [45]. Some types of lifelog data were generated  
46  
47 470 automatically; the others were manually input. We will first detect outlier data  
48  
49 471 points, depending on the type of each lifelog, and eliminate them. The missing  
50  
51 472 time-series lifelog data, ranging in 15-33% of the total data points, would be  
52  
53 473 imputed using the EM-imputation algorithm - e.g. Amelia library [46], after  
54  
55 474 normalising the data by data transformation if required. For downstream  
56  
57  
58  
59  
60

1  
2  
3  
4 475 analysis, the data might be collapsed with time scale, e.g. taking trimmed mean  
5  
6 476 or median for each week, month, or trimester.  
7

8 477  
9

## 10 478 **Statistical analysis of multi-omics data**

11  
12  
13 479 The present study allows combination of longitudinal lifelog data with multi-  
14  
15 480 omics data. In contrast to single omics analysis, the multi-omics analysis would  
16  
17 481 reveal the complicated interactions between one and another. However, the  
18  
19 482 sample size for multi-omics analysis is usually relatively small. Dimension  
20  
21 483 reduction via unsupervised or supervised learning for each omics data would be  
22  
23 484 key ingredients to derive meaningful patterns from high dimensional data sets.  
24  
25 485 Also, obtaining low dimensional representations provides a mean to deal with  
26  
27 486 the multiple testing problem by decreasing number of statistical tests. For gene  
28  
29 487 expression data, surrogate variable analysis [47] and sparse factor analysis [48]  
30  
31 488 are frequently used to capture unknown batch effects in advance to expression  
32  
33 489 quantitative trait locus (eQTL) analysis. The extracted factors can be removed  
34  
35 490 from raw expression data to increase power for detecting associated genes [49].  
36  
37 491 Several unsupervised clustering methods [50,51,52] would be also applicable to  
38  
39 492 obtain hidden patterns from dense time-course lifelog measurements, which  
40  
41 493 might be related to pregnancy complications. Recently developed multi-view  
42  
43 494 factor analysis approaches [53,54] have been used to integrate heterogeneous  
44  
45 495 omics data to identify essential components to distinguish disease subtypes  
46  
47 496 from few hundreds of samples. This line of approach would be a promising way  
48  
49 497 to characterize biological status such as gestational age, and to predict clinical  
50  
51 498 outcomes such as spontaneous preterm birth.  
52  
53  
54  
55  
56  
57

58  
59 499 Standard analyses would be also applicable for the selected variables and  
60

1  
2  
3  
4 500 extracted factors (features). The association of outcomes with each feature will  
5  
6 501 be analyzed using statistical hypothesis tests such as Welch's t-test, Fisher's  
7  
8 502 exact test, the Chi-square test, and others as appropriate. Multiple logistic  
9  
10 503 regression modeling will be used to adjust for confounders and to assess  
11  
12 504 whether each feature or combination of features can be used to predict  
13  
14 505 outcomes. Stepwise selection algorithms or regularized algorithms (e.g.,  
15  
16 506 LASSO, ridge regression, or elastic net) will be used to select the optimal  
17  
18 507 number of contributing features that maximize the predictive power using the  
19  
20 508 leave-1-out cross validation or K-fold cross validation methods.

21  
22  
23  
24 509 Individual genetic features may have an effect on outcomes; therefore, some  
25  
26 510 aggregated genetic risk score should be included in the prediction model. For  
27  
28 511 example, SNVs, including rare variants in or around a chromosome region of a  
29  
30 512 known or estimated risk gene, could be aggregated by considering their impacts  
31  
32 513 on biological function of the gene or their minor allele frequencies in the  
33  
34 514 population. However, this study is limited in the number of study participants,  
35  
36 515 and the aggregated risk score might therefore contribute only slightly to the  
37  
38 516 predictive power. To create a more reliable risk score, the estimates from other  
39  
40 517 large-scale cohort data using polygenic score tools, e.g., PRSice [55], could be  
41  
42 518 used for this study.

43  
44  
45  
46  
47 519

## 520 **FINDINGS TO DATE**

### 521 **Clinical background**

522 A total of 302 women were enrolled, and the mean gestational weeks of  
53  
54 523 recruitment was  $16.4 \pm 4.9$  weeks (mean  $\pm$  SD). A total of 285 participants have  
55  
56 524 been followed up to delivery; their baseline clinical characteristics are described  
57  
58  
59  
60

in Table 1. The mean maternal age at delivery was  $33.3 \pm 4.9$  years. As for educational levels, 62% of the participants were high school graduates with or without vocational college education, and 21% had a college degree. The majority were employed (65%) in early pregnancy, and about 40% had a high household income (over 6 million yen per year). Approximately 42% of the participants were over 35 years of age, 51% were parous, and 22% were overweight or obese by their prepregnancy body mass indices ( $\text{BMI} \geq 25 \text{ kg/m}^2$ ). Overall, 8.4% of the participants had HDP, and 5.6% underwent spontaneous preterm birth. On average, infants were delivered at  $38.0 \pm 2.3$  weeks of gestation with a mean birth weight of  $2907 \pm 572$  g. The rate of low birth weight was 18%. Mean gestational weeks of the first and second blood sampling were  $17.0 \pm 5.0$  and  $27.5 \pm 2.5$ , respectively. The third blood sampling was performed at  $31.1 \pm 3.0$  days after delivery on average. The length of enrollment ranged from 90 to 396 days with a mean of  $216 \pm 61$  days.

**Table 1.** Participant characteristics

Characteristics	Value
<b>Maternal (n = 285)</b>	
• Age at delivery, y, mean (SD)	33.3 ( $\pm 4.9$ )
• Age at delivery, y, n (%)	
20-24	12 (4.2)
25-29	45 (15.8)
30-34	107 (37.5)
35-39	90 (31.6)
40-44	30 (10.5)
45-49	1 (0.4)
• Education (n=81) n (%)	
Elementary school / Junior high school	5 (6.2)
High school	35 (43.2)
Vocational college	23 (28.4)

1		
2		
3		
4	College degree and above	17 (21.0)
5		
6	Others	1 (1.2)
7		
8	Data not available	204
9		
10		
11		
12	• Occupation (n=270) n (%)	
13		
14	Housewife or unemployed	93 (34.4)
15		
16	Employed	175 (64.8)
17		
18	Student	2 (0.7)
19		
20		
21	• Annual household income, yen (n=248) n (%)	
22		
23	< 2 million	17 (6.9)
24		
25	2-4 million	59 (23.8)
26		
27	4-6 million	73 (29.4)
28		
29	6-8 million	51 (20.6)
30		
31	8-10 million	22 (8.9)
32		
33	> 10 million	26 (10.5)
34		
35		
36	• Parity, n (%)	
37	0	140 (49.1)
38	1	93 (32.6)
39		
40	≥ 2	52 (18.2)
41		
42	• Prepregnancy BMI*, kg/m <sup>2</sup> , mean (SD)	22.7 (± 5.1)
43	• Prepregnancy BMI, kg/m <sup>2</sup> , n (%)	
44	< 18.5	36 (12.6)
45	18.5-24.9	186 (65.3)
46	25.0-29.9	34 (11.9)
47	≥ 30.0	29 (10.2)
48		
49	• Gestational weeks at delivery, mean (SD)	38.0 (± 2.3)
50		
51		
52	• Mode of delivery, n (%)	
53	Noncesarean	179 (62.8)
54	Cesarean	106 (37.2)
55		
56	• Pregnancy complication, n (%)	
57	Hypertensive disorder of pregnancy	24 (8.4)
58	Spontaneous preterm birth	16 (5.6)
59		
60		

**Neonatal (n = 300)**

• Birth weight, g, mean (SD)	2907 (± 572)
• Sex, n (%)	
Male	168 (56)
Female	132 (44)
• Low-birth weight (< 2500 g), n (%)	54 (18)

541 \*BMI, body mass index

542

**543 Data acquisition**

544 The percentage of data uploads as of June 2017 was calculated for the 285  
 545 final study participants. For each lifelog item, the upload rate for each  
 546 participant was calculated from the total number of days of actual uploads  
 547 divided by the number of days from enrollment to delivery. The mean upload  
 548 rate for each lifelog item was 85.3% (physical activity), 82.1% (body weight),  
 549 80.4% (body temperature), 78.0% (morning home blood pressure), 71.6%  
 550 (evening home blood pressure), 83.5% (sleep quality), 82.1% (condition of stool,  
 551 severity of pain, severity of nausea, uterine contractions, and palpitations), and  
 552 67.4% (fetal movement) (Figure 3).

553

**554 Number of data points**

555 The total number of collected data points as of June 2017 was calculated for the  
 556 285 final study participants. The approximate number of registered data points  
 557 was 86 000 for body weight, 324 000 points for home diastolic and systolic  
 558 blood pressure, 86 000 for physical activity, and 74 000 for body temperature.  
 559 When physical conditions such as stool condition, severity of pain, and fetal  
 560 movement were combined, the total number of data points was over 6 million.

561

**562 STRENGTHS AND LIMITATIONS**



1  
2  
3  
4 563 Herein, we have described the rationale, design, objective, data collection  
5  
6 564 methods, and interim results of the MLOG study. The study was launched in  
7  
8 565 September 2016, and baseline data collection ended in June 2017. A total of  
9  
10 566 285 participants uploaded lifelog data throughout pregnancy with a high data  
11  
12 567 acquisition rate and over 6 million total data points. Biospecimens for multi-  
13  
14 568 omics analysis were satisfactorily collected and all tracked by LIMS.

15  
16  
17 569 There are three noteworthy features in the MLOG study. First, it is a  
18  
19  
20 570 prospective add-on cohort study based on TMM BirThree Cohort Study, with a  
21  
22 571 full series of epidemiological data and a highly structured follow-up system for  
23  
24 572 mothers, newborns, and families [8]. Second, we have successfully collected  
25  
26 573 longitudinal, continuous, individual lifelog data with a high acquisition rate,  
27  
28 574 which will enable us to assess dynamic changes in physiologic conditions  
29  
30 575 throughout pregnancy. Third, multi-omics data will make it possible to fully  
31  
32 576 understand the complex mechanisms of multifactorial pregnancy-related  
33  
34 577 diseases and to overcome the unpredictability of these complications.

35  
36  
37  
38 578 Prediction models using clinical and epidemiological information and  
39  
40 579 circulating factors for pregnancy-related diseases have been developed  
41  
42 580 extensively [56], and risk-assessment approaches using clinical information  
43  
44 581 have also been developed [57, 58]. However, there is a lack of evidence for the  
45  
46 582 benefits of these predictive models for routine clinical use [59]. Once the  
47  
48 583 likelihood of a pregnancy-related disorder is estimated with high sensitivity and  
49  
50 584 specificity, evidence-based clinical interventions could reduce the rate of  
51  
52 585 maternal and neonatal morbidity and mortality [60]. Therefore, an early-  
53  
54 586 prediction algorithm that can be used with a high level of confidence is needed  
55  
56 587 to obtain better outcomes for patients with pregnancy complications.  
57  
58  
59  
60

1  
2  
3  
4 588 Recently, several studies of sample sizes comparable with ours, exploiting  
5  
6 589 lifelog or multi-omics data were reported. One of the studies analyzed lifelog  
7  
8 590 and multi-omics data, collected from 108 individuals at three time points during  
9  
10 591 a nine-month period [61]. In their study, several remarkable relationships were  
11  
12 592 identified among physiological and multi-omics data through integrated  
13  
14 593 analyses. Another study investigated genome-wide associations between  
15  
16 594 genetic variants and gene expression levels across 44 human tissues from a  
17  
18 595 few hundreds of postmortem donors [49]. They studied both cis-eQTL (within 1  
19  
20 596 Mb of target-gene transcription start sites) and trans-eQTLs (more distant from  
21  
22 597 target genes or on other chromosomes) with 350 whole blood samples, and  
23  
24 598 thereby identified 5,862 cis-eQTL and one trans-eQTL associations. These  
25  
26 599 previous studies indicate that our time-course high-resolution reference catalog  
27  
28 600 with 285 pregnant women would be well applicable to high-dimensional data  
29  
30 601 analyses such as searches for quantitative trait loci and molecular risk markers.

31  
32  
33  
34  
35  
36 602 Potential limitation of the present study is participant recruitment only at  
37  
38 603 Tohoku University Hospital that is one of the tertiary hospitals in Miyagi  
39  
40 604 Prefecture for high-risk populations. Therefore, the sample size is limited, and  
41  
42 605 the results might not be applicable to the general populations. Inclusion criteria  
43  
44 606 of the present study limited the eligibility to pregnant women with age >20 years  
45  
46 607 and the ability to access the internet using a smartphone. Therefore, results of  
47  
48 608 the present study might not be applicable to pregnancies with lower coverage of  
49  
50 609 smartphone use.

51  
52  
53  
54 610 Hopefully, our study will result in the development of a novel stratification  
55  
56 611 model for pregnancy-related diseases employing multi-omics and lifelog data.  
57  
58 612 The MLOG study will enable us to construct a time-course high-resolution

1  
2  
3  
4 613 reference catalog of wellness and multi-omics data from pregnant women and  
5  
6 614 thereby develop a personalized predictive model for pregnancy complications.  
7  
8 615 Progressive data sharing and collaborative studies would make it possible to  
9  
10 616 establish a standardized early-prediction method through large clinical trials.  
11  
12  
13 617

### 618 **Author affiliations**

17 619 1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryomachi,  
18  
19 620 machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

22 621 2 Tohoku University Graduate School of Medicine, 1-1, Seiryomachi, Aoba-ku,  
23  
24 622 Sendai, Miyagi, 980-8574, Japan.

27 623 3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,  
28  
29 624 Kanagawa, Japan 239-8536.

31 625 4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-  
32  
33 626 ku, Sendai, Miyagi 981-8558, Japan.

36 627 5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku,  
37  
38 628 Hiroshima 732-0815, Japan.

40 629 6 Tohoku University Hospital, 1-1, Seiryomachi, Aoba-ku, Sendai, Miyagi, 980-  
41  
42 630 8574, Japan.

45 631

48 632

### 633 **Acknowledgements**

51 634 The authors would like to thank all the MLOG study participants, the staff of the  
52  
53 635 Tohoku Medical Megabank Organization, Tohoku University (a full list of  
54  
55 636 members is available at: <http://www.megabank.tohoku.ac.jp/english/a161201/>),  
56  
57 637 and the Department of Obstetrics and Gynecology, Tohoku University Hospital,  
58  
59  
60

1  
2  
3  
4 638 for their efforts and contributions. The MLOG study group also included Chika  
5  
6 639 Igarashi, Motoko Ishida, Yumiko Ishii, Hiroko Yamamoto, Akiko Akama, Kaori  
7  
8 640 Noro, Miyuki Ozawa, Yuka Narita, Junko Yusa, Miwa Meguro, Michiyo Sato,  
9  
10 641 Miyuki Watanabe, Mai Tomizuka, Mika Hotta, Naomi Matsukawa, Makiko Sumii,  
11  
12 642 Ayako Okumoto, Yukie Oguma, Ryoko Otokozawa, Toshiya Hatanaka, Sho  
13  
14 643 Furuhashi, Emi Shoji, Tomoe Kano, Riho Mishina, and Daisuke Inoue.  
15  
16  
17  
18  
19

644

### 645 **Contributors**

20  
21  
22 646 JS, DO, RY, TY, HM, OT, SKu, NY, SH, MN were involved in initial stages of  
23  
24 647 the strategy and design of study conception. JS, DO, RY, TY, OT, DS, SKo, SH,  
25  
26 648 MN: responsible for the draft of the manuscript. JS, DO, RY, TY, MW, MI, HM,  
27  
28 649 OY, SKu: recruitment and sample collection. DO, RY, TY, DS, TO, YT, YH, TFS,  
29  
30 650 TM, JK, FK, TIT, SO, NM, SKo, OT, MN: sample analysis, data processing, and  
31  
32 651 statistical analysis. JS, HH, NF, NM, SKo, OT, SKu, KK, SK, NY, MY, SH, MN:  
33  
34 652 advice and supervision of sample analysis. All authors have contributed to  
35  
36 653 revision and have approved the final manuscript, and agreed to be accountable  
37  
38 654 for all aspects of the work in ensuring that questions related to the accuracy or  
39  
40 655 integrity of any part of the work are appropriately investigated and resolved.  
41  
42  
43  
44

656

### 657 **Funding**

45  
46  
47  
48  
49 658 The present study was supported by NTT DoCoMo, Inc., with a collaborative  
50  
51 659 research agreement between NTT DoCoMo and ToMMo. This work was  
52  
53 660 supported in part by the Tohoku Medical Megabank Project from the Japan  
54  
55 661 Agency for Medical Research and Development (AMED) and the Ministry of  
56  
57 662 Education, Culture, Sports, Science and Technology (MEXT).  
58  
59  
60

1  
2  
3  
4 663

5  
6 664 **Competing interests**

7  
8 665 This study was funded by NTT DoCoMo, Inc.

9  
10 666 Daisuke Ochi, Takafumi Yamauchi, and Satoshi Hiyama are employees of NTT

11  
12 667 DoCoMo, Inc. All other authors declare that they have no competing interests.

13  
14  
15 668

16  
17  
18  
19 669 **Ethics approval and consent to participate**

20  
21 670 TMM BirThree Cohort Study was approved by the ethics committees of the

22  
23 671 Tohoku University (authorization numbers, 2013-4-103 and 2017-4-010). The

24  
25 672 MLOG study was approved by the ethics committees of the Graduate School of

26  
27 673 Medicine (2014-1-704) and the Tohoku Medical Megabank Organization (2017-

28  
29 674 1-085), Tohoku University. Written informed consent was obtained from all

30  
31 675 participants.

32  
33  
34  
35 676

36  
37 677 **Provenance and peer review**

38  
39 678 Not commissioned; externally peer reviewed.

40  
41  
42 679

43  
44 680 **Data sharing statement**

45  
46 681 The datasets used during the current study are available from the

47  
48 682 corresponding authors on reasonable request.

49  
50  
51 683

52  
53 684

54  
55 685

56  
57  
58 686

59  
60

1  
2  
3  
4 6875  
6 6887  
8 6899  
10 690 **REFERENCES**11  
12 691 1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public13  
14 692 health perspective. *Diabetes Care*. 2007;30 Suppl 2: S141-6. doi:15  
16 693 10.2337/dc07-s206.17  
18 69419  
20 695 2. Beck S, Wojdyla D, Say L, et al. The worldwide incidence of preterm birth: a21  
22 696 systematic review of maternal mortality and morbidity. *Bull World Health Organ*.23  
24 697 2010; 88:31-8. doi: 10.2471/BLT.08.062554.25  
26 69827  
28 699 3. Duley L. The global impact of pre-eclampsia and eclampsia. *Semin Perinatol*.29  
30 700 2009; 33:130-7. doi: 10.1053/j.semperi.2009.02.010.31  
32 70133  
34 702 4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United States,35  
36 703 1980-2010: age-period-cohort analysis. *BMJ*. 2013;347: f 6564. doi:37  
38 704 10.1136/bmj.f6564.39  
40 70541  
42 706 5. Waken RJ, de Las Fuentes L, Rao DC. A Review of the Genetics of43  
44 707 Hypertension with a Focus on Gene-Environment Interactions. *Curr Hypertens*45  
46 708 *Rep*. 2017; 19:23. doi: 10.1007/s11906-017-0718-1.47  
48 70949  
50 710 6. Ward K, Lindheimer MD. Genetic factors in the etiology of preeclampsia /51  
52 711 eclampsia. In: Chesley's Hypertensive Disorders in pregnancy. London:

1  
2  
3  
4 712 Elsevier; 2990: 51-72.  
5

6 713  
7

8 714 7. Li X, Dunn J, Salins D, Zhou G, et al. Digital Health: Tracking Physiomes and  
9  
10 715 Activity Using Wearable Biosensors Reveals Useful Health Related Information.  
11  
12 716 PLoS Biol. 2017; 15: e2001402. doi: 10.1371/journal.pbio.2001402.  
13  
14

15 717  
16

17 718 8. Kuriyama S, Yaegashi N, Nagami F, et al. The Tohoku Medical Megabank  
18  
19 719 Project: Design and Mission. J Epidemiol. 2016; 26:493-511. doi:  
20  
21 720 10.2188/jea.JE20150268.  
22  
23

24 721  
25

26 722 9. Japan Society of Obstetrics and Gynecology, Guideline for Obstetrical  
27  
28 723 Practice in Japan, Japan Society of Obstetrics and Gynecology, Tokyo, Japan,  
29  
30 724 pp. 1–4, 2017 [in Japanese].  
31  
32

33 725  
34

35 726 10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the  
36  
37 727 thermoneutral zone: a longitudinal study of normal human pregnancy. Acta  
38  
39 728 Physiol (Oxf). 2011; 201: 467-74. doi: 10.1111/j.1748-1716.2010.02228.x.  
40  
41

42 729  
43

44 730 11. Metoki H, Ohkubo T, Watanabe Y, et al. Seasonal trends of blood pressure  
45  
46 731 during pregnancy in Japan: the babies and their parents' longitudinal  
47  
48 732 observation in Suzuki Memorial Hospital in Intrauterine Period study. J  
49  
50 733 Hypertens. 2008; 26: 2406-13. doi: 10.1097/HJH.0b013e32831364a7.  
51  
52

53 734  
54

55 735 12. Haugen M, Brantsæter AL, Winkvist A, et al. Associations of pre-pregnancy  
56  
57 736 body mass index and gestational weight gain with pregnancy outcome and  
58  
59  
60

- 1  
2  
3  
4 737 postpartum weight retention: a prospective observational cohort study. BMC  
5  
6 738 Pregnancy Childbirth. 2014 Jun 11; 14: 201. doi: 10.1186/1471-2393-14-201.  
7  
8 739  
9  
10 740 13. Sorensen TK, Williams MA, Lee IM, et al. Recreational physical activity  
11  
12 741 during pregnancy and risk of preeclampsia. Hypertension. 2003 Jun; 41:1273-  
13  
14 742 80. doi: 10.1161/01.HYP.0000072270.82815.91  
15  
16 743  
17  
18 744 14. Reutrakul S, Zaidi N, Wroblewski K, et al. Sleep disturbances and their  
19  
20 745 relationship to glucose tolerance in pregnancy. Diabetes Care. 2011; 34: 2454-7.  
21  
22 746 doi: 10.2337/dc11-0780.  
23  
24 747  
25  
26 748 15. Cornish J, Tan E, Teare J, et al. A meta-analysis on the influence of  
27  
28 749 inflammatory bowel disease on pregnancy. Gut. 2007; 56: 830-7. doi:  
29  
30 750 10.1136/gut.2006.108324.  
31  
32 751  
33  
34 752 16. Huxley RR. Nausea and vomiting in early pregnancy: its role in placental  
35  
36 753 development. Obstet Gynecol. 2000; 95:779-82.  
37  
38 754  
39  
40 755 17. Holm Tveit JV, Saastad E, Stray-Pedersen B, et al. Maternal characteristics  
41  
42 756 and pregnancy outcomes in women presenting with decreased fetal movements  
43  
44 757 in late pregnancy. Acta Obstet Gynecol Scand. 2009; 88: 1345-51. doi:  
45  
46 758 10.3109/00016340903348375.  
47  
48 759  
49  
50 760 18. Facchinetti F, Allais G, D'Amico R, et al. The relationship between  
51  
52 761 headache and preeclampsia: a case-control study. Eur J Obstet Gynecol  
53  
54  
55  
56  
57  
58  
59  
60



- 1  
2  
3  
4 762 Reprod Biol. 2005; 121: 143-8. doi: 10.1016/j.ejogrb.2004.12.020.  
5  
6 763  
7  
8 764 19. Iams JD, Newman RB, Thom EA, et al; National Institute of Child Health  
9  
10 765 and Human Development Network of Maternal-Fetal Medicine Units. Frequency  
11  
12 766 of uterine contractions and the risk of spontaneous preterm delivery. N Engl J  
13  
14 767 Med. 2002; 346: 250-5. doi: 10.1056/NEJMoa002868  
15  
16 768  
17  
18 769 20. Abbas AE, Lester SJ, Connolly H. Pregnancy and the cardiovascular  
19  
20 770 system. Int J Cardiol. 2005; 98: 179-89. doi: 10.1016/j.ijcard.2003.10.028.  
21  
22 771  
23  
24 772 21. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit  
25  
26 773 time. Scand J Gastroenterol. 1997; 32: 920-4.  
27  
28 774 doi: 10.3109/00365529709011203  
29  
30 775  
31  
32 776 22. Riegler G, Esposito I. Bristol scale stool form. A still valid help in medical  
33  
34 777 practice and clinical research. Tech Coloproctol 2001; 5: 163-4. doi:  
35  
36 778 10.1007/s101510100019  
37  
38 779  
39  
40 780 23. Longstreth GF, Thompson WG, Chey WD, et al. Functional Bowel Disorders.  
41  
42 781 Gastroenterology 2006; 130: 1480-91. doi: 10.1053/j.gastro.2005.11.061  
43  
44 782  
45  
46 783 24. Koren G, Boskovic R, Hard M, et al. Motherisk-PUQE (pregnancy-unique  
47  
48 784 quantification of emesis and nausea) scoring system for nausea and vomiting of  
49  
50 785 pregnancy. Am J Obstet Gynecol. 2002;186: S228-31.  
51  
52 786  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 787 25. Koren G, Piwko C, Ahn E, et al. Validation studies of the Pregnancy Unique-  
5  
6 788 Quantification of Emesis (PUQE) scores. *J Obstet Gynaecol*. 2005; 25: 241-4.  
7  
8 789 doi: 10.1080/01443610500060651  
9  
10 790  
11  
12  
13 791 26. Pearson JF, Weaver JB. Fetal activity and fetal wellbeing: an evaluation. *Br*  
14  
15 792 *Med J* 1976; 1:1305–7.  
16  
17 793  
18  
19  
20 794 27. Winje BA, Saastad E, Gunnes N, et al. Analysis of 'count-to-ten' fetal  
21  
22 795 movement charts: a prospective cohort study. *BJOG*. 2011;118: 1229-38. doi:  
23  
24 796 10.1111/j.1471-0528.2011.02993.x  
25  
26 797  
27  
28  
29 798 28. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach  
30  
31 799 Notification rules under the Health Information Technology for Economic and  
32  
33 800 Clinical Health Act and the Genetic Information Nondiscrimination Act; other  
34  
35 801 modifications to the HIPAA rules. *Fed Regist*. 2013; 78: 5565-702.  
36  
37 802  
38  
39  
40 803 29. Health Insurance Portability and Accountability Act of 1996. Public Law 104-  
41  
42 804 191. *US Statut Large*. 1996; 110:1936-2103.  
43  
44 805  
45  
46  
47 806 30. Amended Act on the Protection of Personal Information.  
48  
49 807 [https://www.ppc.go.jp/files/pdf/Act\\_on\\_the\\_Protection\\_of\\_Personal\\_Information.](https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Information.pdf)  
50  
51 808 pdf.  
52  
53 809  
54  
55  
56 810 31. Takai-Igarashi T, Kinoshita K, Nagasaki M, et al. Security controls in an  
57  
58 811 integrated Biobank to protect privacy in data sharing: rationale and study design.

- 1  
2  
3  
4 812 BMC Med Inform Decis Mak. 2017; 17:100. doi: 10.1186/s12911-017-0494-5  
5  
6 813  
7  
8 814 32. Katsuoka F, Yokozawa J, Tsuda K, et al. An efficient quantitation method of  
9  
10 815 next-generation sequencing libraries by using MiSeq sequencer. Anal Biochem.  
11  
12 816 2014; 466: 27-9. doi: 10.1016/j.ab.2014.08.015  
13  
14  
15 817  
16  
17 818 33. Yamaguchi-Kabata Y, et al, Nariai N, Kawai Y, et al. iJGVD: an  
18  
19 819 integrative Japanese genome variation database based on whole-genome  
20  
21 820 sequencing. Hum Genome Var. 2015; 2:15050. doi: 10.1038/hgv.2015.50.  
22  
23 821 <https://www.ncbi.nlm.nih.gov/pubmed/27081555>  
24  
25 822  
26  
27 823 34. DeLuca DS, Levin JZ, Sivachenko A, et al. RNA-SeQC: RNA-seq metrics  
28  
29 824 for quality control and process optimization. Bioinformatics. 2012; 28: 1530-  
30  
31 825 1532. doi: 10.1093/bioinformatics/bts196  
32  
33 826  
34  
35 827 35. Koshiha S, Motoike I, Kojima K, et al. The structural origin of metabolic  
36  
37 828 quantitative diversity. Sci Rep. 2016; 6: 31463. doi: 10.1038/srep31463  
38  
39 829  
40  
41 830 36. Vu TN, Valkenborg D, Smets K, et al. An integrated workflow for robust  
42  
43 831 alignment and simplified quantitative analysis of NMR spectrometry data. BMC  
44  
45 832 Bioinformatics. 2011; 12: 405. doi: 10.1186/1471-2105-12-405.  
46  
47 833  
48  
49 834 37. Nishiumi S, Kobayashi T, Ikeda A, et al. A novel serum metabolomics-based  
50  
51 835 diagnostic approach for colorectal cancer. PLoS One. 2012; 7: e40459. doi:  
52  
53 836 10.1371/journal.pone.0040459.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 837  
5  
6 838 38. Nishiumi S, Kobayashi T, Kawana S, et al. Investigations in the possibility of  
7  
8 839 early detection of colorectal cancer by gas chromatography/triple-quadrupole  
9  
10 840 mass spectrometry. *Oncotarget*. 2017; 8, 17115-17126. doi:  
11  
12 841 10.18632/oncotarget.15081.  
13  
14  
15 842  
16  
17 843 39. Saigusa D, Okamura Y, Motoike IN, et al. Establishment of Protocols for  
18  
19 844 Global Metabolomics by LC-MS for Biomarker Discovery. *PLoS One*.  
20  
21 845 2016;11(8): e0160555. doi: 10.1371/journal.pone.0160555.  
22  
23  
24 846  
25  
26 847 40. Sato Y, Yamagishi J, Yamashita R, et al. Inter-Individual Differences in the  
27  
28 848 Oral Bacteriome Are Greater than Intra-Day Fluctuations in Individuals. *PLoS*  
29  
30 849 *One*. 2015;10: e0131607. doi: 10.1371/journal.pone.0131607.  
31  
32  
33 850  
34  
35 851 41. Brown MA, Magee LA, Kenny LC, et al. Hypertensive Disorders of  
36  
37 852 Pregnancy: ISSHP Classification, Diagnosis, and Management  
38  
39 853 Recommendations for International Practice. *Hypertension*. 2018; 72: 24–43.  
40  
41 854 doi: 10.1161/HYPERTENSIONAHA.117.10803.  
42  
43 855  
44  
45 856 42. Watanabe K, Naruse K, Tanaka K, et al. Outline of definition and  
46  
47 857 classification of pregnancy induced hypertension (PIH). *Hypertens Res*  
48  
49 858 *Pregnancy* 2013; 1: 3–4.  
50  
51 859  
52  
53 860 43. IADPSG Consensus Panel: International association of diabetes and  
54  
55 861 pregnancy study groups recommendations on the diagnosis and classification  
56  
57  
58  
59  
60

- 1  
2  
3  
4 862 of hyperglycemia in pregnancy. *Diabetes Care* 2010; 33: 676-682. doi:  
5  
6 863 10.2337/dc09-1848.  
7  
8 864  
9  
10 865 44. Box GEP, Jenkins GM, Reinsel GC. *Time series Analysis: Forecasting and*  
11  
12 *Control*. 5th ed. New Jersey: Wiley; 2015.  
13 866  
14  
15 867  
16  
17 868 45. Brandt PT, Williams JT. *Multiple Time Series Models*, Thousand Oaks, CA:  
18  
19 Sage Publications, 2007.  
20 869  
21  
22 870  
23  
24 871 46. Honaker J, King G, Blackwell M. *Amelia II: A Program for Missing Data*,  
25  
26 *Journal of Statistical Software*, 45 (7) 2011.  
27 872  
28  
29 873  
30  
31 874 47. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by  
32  
33 surrogate variable analysis. *PLoS Genet*. 2007; 3: 1724-35. doi:  
34 875  
35 10.1371/journal.pgen.0030161  
36 876  
37  
38 877  
39  
40 878 48. Stegle O, Parts L, Piipari M, et al. Using probabilistic estimation of  
41  
42 expression residuals (PEER) to obtain increased power and interpretability of  
43 879  
44 gene expression analyses. *Nat Protoc*. 2012; 7: 500-7. doi:  
45 880  
46 10.1038/nprot.2011.457.  
47 881  
48  
49 882  
50  
51 883 49. GTEx Consortium, Battle A, Brown CD, et al. Genetic effects on gene  
52  
53 expression across human tissues. *Nature*. 2017; 550: 204-213. doi:  
54 884  
55 10.1038/nature24277.  
56 885  
57  
58 886  
59  
60

- 1  
2  
3  
4 887 50. Polgreen PM, Yang M, Kuntz JL, et al. Using oral vancomycin prescriptions  
5  
6 888 as a proxy measure for *Clostridium difficile* infections: a spatial and time series  
7  
8 889 analysis. *Infect Control Hosp Epidemiol*. 2011; 32: 723-6. doi: 10.1086/660858.  
9  
10 890  
11  
12 891 51. McDowell IC, Manandhar D, Vockley CM, et al. Clustering gene expression  
13  
14 892 time series data using an infinite Gaussian process mixture model. *PLoS*  
15  
16 893 *Comput Biol*. 2018; 14: e1005896. doi: 10.1371/journal.pcbi.1005896.  
17  
18 894  
19  
20 895 52. Hensman J, Rattray M, Lawrence ND. Fast Nonparametric Clustering of  
21  
22 896 Structured Time-Series. *IEEE Trans Pattern Anal Mach Intell*. 2015; 37: 383-93.  
23  
24 897 doi: 10.1109/TPAMI.2014.2318711.  
25  
26 898  
27  
28 899 53. Rohart F, Gautier B, Singh A et al. mixOmics: An R package for 'omics  
29  
30 900 feature selection and multiple data integration. *PLoS Comput Biol*. 2017;13:  
31  
32 901 e1005752. doi: 10.1371/journal.pcbi.1005752.  
33  
34 902  
35  
36 903 54. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis-a  
37  
38 904 framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*.  
39  
40 905 2018; 14: e8124. doi: 10.15252/msb.20178124.  
41  
42 906  
43  
44 907 55. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software.  
45  
46 908 *Bioinformatics*, 2015; 31:1466-8. doi: 10.1093/bioinformatics/btu848.  
47  
48 909  
49  
50 910 56. Wax JR, Cartin A, Pinette MG. Biophysical and Biochemical Screening for  
51  
52 911 the Risk of Preterm Labor: An Update. *Clin Lab Med*. 2016; 36: 369-83. doi:  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 912 10.1016/j.cll.2016.01.019.  
5  
6 913  
7  
8 914 57. Al-Rubaie Z, Askie LM, Ray JG, et al. The performance of risk prediction  
9  
10 915 models for pre-eclampsia using routinely collected maternal characteristics and  
11  
12 916 comparison with models that include specialised tests and with clinical guideline  
13  
14 917 decision rules: a systematic review. BJOG. 2016; 123:1441-1452. doi:  
15  
16 918 10.1111/1471-0528.  
17  
18  
19  
20 919  
21  
22 920 58. Koullali B, Oudijk MA, Nijman TA, et al. Risk assessment and management  
23  
24 921 to prevent preterm birth. Semin Fetal Neonatal Med. 2016 ;21: 80-8. doi:  
25  
26 922 10.1016/j.siny.2016.01.005.  
27  
28  
29 923  
30  
31 924 59. Henderson JT, Thompson JH, Burda BU, et al. Preeclampsia Screening:  
32  
33 925 Evidence Report and Systematic Review for the US Preventive Services Task  
34  
35 926 Force. JAMA. 2017; 317: 1668-1683. doi: 10.1001/jama.2016.18315.  
36  
37  
38 927  
39  
40 928 60. Broekhuijsen K, van Baaren GJ, van Pampus MG, et al; HYPITAT-II Study  
41  
42 929 Group. Immediate delivery versus expectant monitoring for hypertensive  
43  
44 930 disorders of pregnancy between 34 and 37 weeks of gestation (HYPITAT-II): an  
45  
46 931 open-label, randomised controlled trial. Lancet. 2015; 385: 2492-2501. doi:  
47  
48 932 10.1016/S0140-6736(14)61998-X.  
49  
50  
51 933  
52  
53 934 61. Price ND, Magis AT, Earls JC, et al. A wellness study of 108 individuals  
54  
55 935 using personal, dense, dynamic data clouds. Nat Biotechnol. 2017; 35: 747-756.  
56  
57  
58 936 doi: 10.1038/nbt.3870.  
59  
60

1  
2  
3  
4 9375  
6 9387  
8 9399  
10 940 **FIGURE TITLES AND LEGENDS**11  
12  
13 94114  
15 942 **Figure 1. Flowchart of Maternity Log Study (MLOG) participants**16  
17 94318  
19 944 **Figure 2. Overview of the MLOG study protocol**20  
21 945 **A:** Participant timeline for the MLOG study.22  
23 946 **B:** Physiologic information collected using healthcare devices. Specific  
24  
25 947 measures were uploaded each day from the time of enrollment (solid horizontal  
26  
27 948 lines). Participants had the option to continue uploading data until 180 days  
28  
29 949 after delivery (dashed horizontal lines).30  
31 950 **C:** Daily lifelogs of self-reported information using a smartphone application.  
32  
33 951 Basic lifelog information was input manually from the time of enrollment (solid  
34  
35 952 horizontal lines). Participants had the option to continue uploading data until  
36  
37 953 180 days after delivery (dashed horizontal lines). Fetal movement and uterine  
38  
39 954 contractions were recorded from 24 and 20 weeks of gestation, respectively.40  
41 95542  
43 956 **Figure 3. Data acquisition rate**44  
45 957 The mean data upload rate of specific measures was calculated from the total  
46  
47 958 number of days of actual uploads divided by the number of days from  
48  
49 959 enrollment to delivery for each participant.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



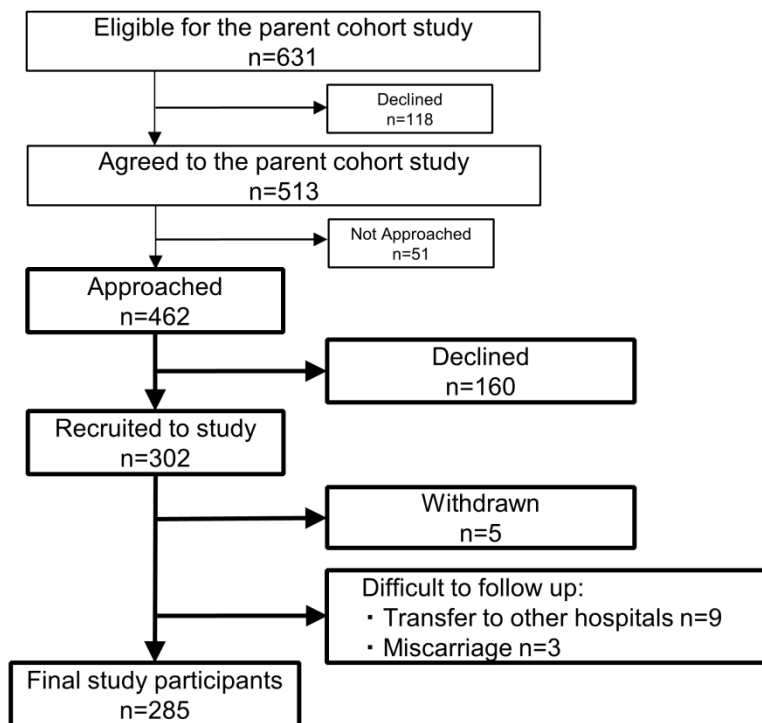


Figure 1.

254x190mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

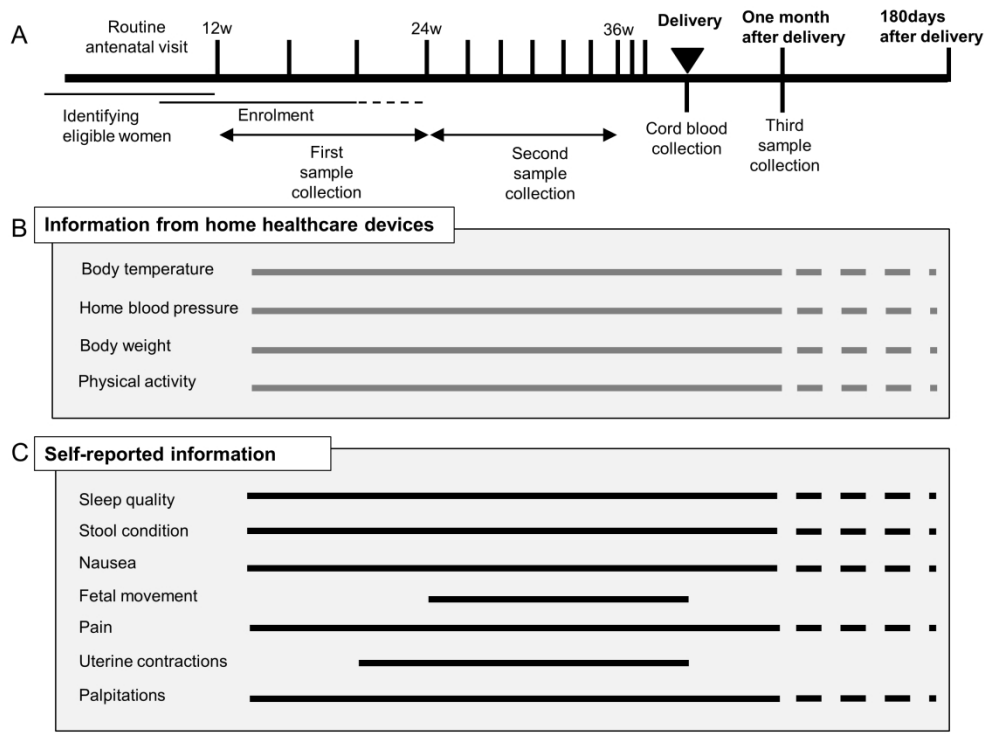


Figure 2.

254x190mm (300 x 300 DPI)

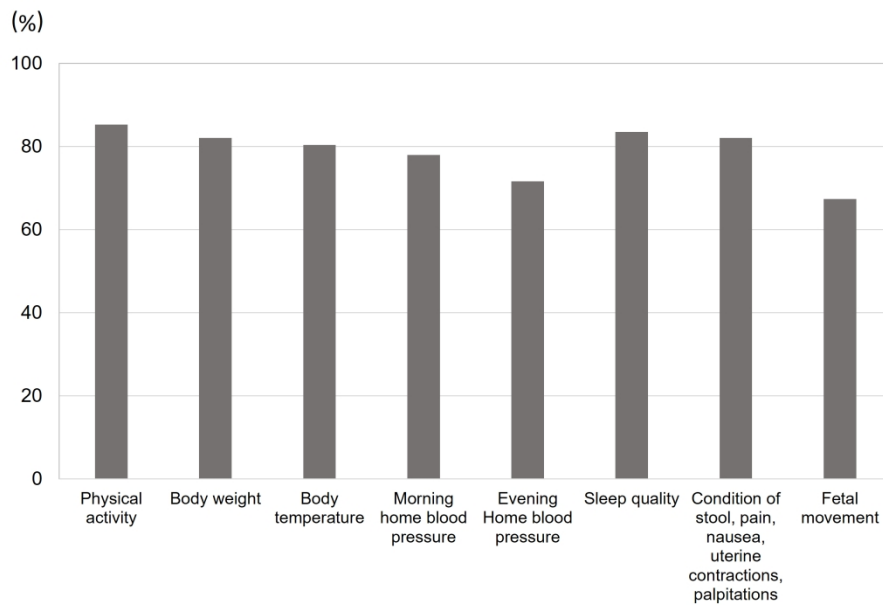


Figure 3.

254x190mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# BMJ Open

## Cohort Profile: Maternity Log Study: a longitudinal lifelog monitoring and multi-omics analysis for the early prediction of complicated Pregnancy

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-025939.R2
Article Type:	Cohort profile
Date Submitted by the Author:	20-Dec-2018
Complete List of Authors:	<p>Sugawara, Junichi; Tohoku Medical Megabank Organization, Tohoku University, Feto-Maternal Medical Science; Tohoku University Graduate School of Medicine, Obstetrics and Gynecology  Ochi, Daisuke; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC.  Yamashita, Riu; Tohoku Medical Megabank Organization, Tohoku University  Yamauchi, Takafumi ; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC.  Saigusa, Daisuke; Tohoku Medical Megabank Organization, Tohoku University  Wagata, Maiko; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine  Obara, Taku; Tohoku Medical Megabank Organization, Tohoku University  Ishikuro, Mami; Tohoku Medical Megabank Organization, Tohoku University  Tsunemoto, Yoshiki; Research Laboratories, NTT DOCOMO, INC.  Harada, Yuki; Tohoku Medical Megabank Organization, Tohoku University  Shibata, Tomoko; Tohoku Medical Megabank Organization, Tohoku University  Mimori, Takahiro ; Tohoku Medical Megabank Organization, Tohoku University  Kawashima, Junko; Tohoku Medical Megabank Organization, Tohoku University  Katsuoka, Fumiki; Tohoku Medical Megabank Organization, Tohoku University  Igarashi-Takai, Takako ; Tohoku Medical Megabank Organization, Tohoku University  Ogishima, Soichi; Tohoku Medical Megabank Organization, Tohoku University  Metoki, Hirohito; Tohoku Medical and Pharmaceutical University  Hashizume, Hiroaki; Tohoku Medical Megabank Organization, Tohoku University  Fuse, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine  Minegishi, Naoko; Tohoku Medical Megabank Organization, Tohoku University  Koshiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University</p>

	Tanabe, Osamu; Tohoku Medical Megabank Organization, Tohoku University; Radiation Effects Research Foundation Kuriyama, Shinichi; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Kinoshita, Kengo; Tohoku Medical Megabank Organization, Tohoku University Kure, Shigeo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Yaegashi, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; TUH Yamamoto, Masayuki; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Hiyama, Satoshi; Research Laboratories, NTT DOCOMO, INC. Nagasaki, Masao; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine
<b>Primary Subject Heading</b>:	Obstetrics and gynaecology
Secondary Subject Heading:	Health informatics
Keywords:	lifelog, multi-omics analysis, prediction, complicated pregnancy

SCHOLARONE™  
Manuscripts

1 **Cohort profile: Maternity Log Study: a longitudinal**  
2 **lifelog monitoring and multi-omics analysis for the**  
3 **early prediction of complicated pregnancy**

4  
5 Junichi Sugawara<sup>1,2\*</sup>, Daisuke Ochi<sup>1,3</sup>, Riu Yamashita<sup>1</sup>, Takafumi Yamauchi<sup>1,3</sup>,  
6 Daisuke Saigusa<sup>1</sup>, Maiko Wagata<sup>1,2</sup>, Taku Obara<sup>1</sup>, Mami Ishikuro<sup>1</sup>, Yoshiki  
7 Tsunemoto<sup>3</sup>, Yuki Harada<sup>1</sup>, Tomoko F. Shibata<sup>1</sup>, Takahiro Mimori<sup>1</sup>, Junko  
8 Kawashima<sup>1</sup>, Fumiki Katsuoka<sup>1</sup>, Takako Igarashi-Takai<sup>1</sup>, Soichi Ogishima<sup>1</sup>,  
9 Hirohito Metoki<sup>4</sup>, Hiroaki Hashizume<sup>1</sup>, Nobuo Fuse<sup>1,2</sup>, Naoko Minegishi<sup>1</sup>, Seizo  
10 Koshiba<sup>1</sup>, Osamu Tanabe<sup>1,5</sup>, Shinichi Kuriyama<sup>1,2</sup>, Kengo Kinoshita<sup>1</sup>, Shigeo  
11 Kure<sup>1,2</sup>, Nobuo Yaegashi<sup>1,6</sup>, Masayuki Yamamoto<sup>1,2</sup>, Satoshi Hiyama<sup>3</sup>, and  
12 Masao Nagasaki<sup>1,2\*</sup>.

13  
14 1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryomachi,  
15 Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

16 2 Tohoku University Graduate School of Medicine, 1-1, Seiryomachi, Aoba-ku,  
17 Sendai, Miyagi, 980-8574, Japan.

18 3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,  
19 Kanagawa, Japan 239-8536.

20 4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-  
21 ku, Sendai, Miyagi 981-8558, Japan.

22 5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku,  
23 Hiroshima 732-0815, Japan.

24 6 Tohoku University Hospital, 1-1, Seiryomachi, Aoba-ku, Sendai, Miyagi, 980-

1  
2  
3  
4 25 8574, Japan.  
5  
6 26  
7  
8 27  
9

10 28 Email addresses:

11  
12  
13 29 \*Junichi Sugawara: jsugawara@med.tohoku.ac.jp  
14

15 30 Daisuke Ochi: ochi@megabank.tohoku.ac.jp  
16

17 31 Riu Yamashita: ryamasi@megabank.tohoku.ac.jp  
18

19 32 Takafumi Yamauchi: t.yamauchi@megabank.tohoku.ac.jp  
20

21 33 Daisuke Saigusa: saigusa@m.tohoku.ac.jp  
22

23 34 Maiko Wagata: wagata@med.tohoku.ac.jp  
24

25 35 Taku Obara: obara-t@hosp.tohoku.ac.jp  
26

27 36 Mami Ishikuro: m\_ishikuro@med.tohoku.ac.jp  
28

29 37 Yoshiki Tsunemoto: yoshiki.tsunemoto@megabank.tohoku.ac.jp  
30

31 38 Yuki Harada: harada@megabank.tohoku.ac.jp  
32

33 39 Tomoko F. Shibata: tshibata@megabank.tohoku.ac.jp  
34

35 40 Takahiro Mimori: mimori@megabank.tohoku.ac.jp  
36

37 41 Junko Kawashima: kawashima@dent.tohoku.ac.jp  
38

39 42 Fumiki Katsuoka: kfumiki@med.tohoku.ac.jp  
40

41 43 Takako Igarashi-Takai: takai@megabank.tohoku.ac.jp  
42

43 44 Soichi Ogishima: ogishima@megabank.tohoku.ac.jp  
44

45 45 Hirohito Metoki: hmetoki@tohoku-mpu.ac.jp  
46

47 46 Hiroaki Hashizume: hashizume@megabank.tohoku.ac.jp  
48

49 47 Nobuo Fuse: fusen@megabank.tohoku.ac.jp  
50

51 48 Naoko Minegishi: nmine@med.tohoku.ac.jp  
52

53 49 Seizo Koshiba: koshiba@megabank.tohoku.ac.jp  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 50 Osamu Tanabe: otanabe@rerf.or.jp  
5  
6 51 Shinichi Kuriyama: kuriyama@med.tohoku.ac.jp  
7  
8 52 Kengo Kinoshita: kengo@ecei.tohoku.ac.jp  
9  
10 53 Shigeo Kure: kure@med.tohoku.ac.jp  
11  
12 54 Nobuo Yaegashi: yaegashi@med.tohoku.ac.jp  
13  
14 55 Masayuki Yamamoto: masiyamamoto@med.tohoku.ac.jp  
15  
16 56 Satoshi Hiyama: hiyamas@nttdocomo.com  
17  
18 57 \*Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp  
19  
20  
21  
22 58  
23  
24 59 \*Corresponding authors:  
25  
26 60 Junichi Sugawara: jsugawara@med.tohoku.ac.jp  
27  
28 61 and Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp  
29  
30 62 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryomachi,  
31  
32 Aoba-ku, 980-8573, Sendai, Japan. Phone: +81-22-273-6283  
33  
34  
35  
36 64  
37  
38 65 Word count: 5305 words  
39  
40 66 Key words: lifelog, multi-omics analysis, prediction, complicated pregnancy  
41  
42  
43 67  
44  
45 68  
46  
47 69  
48  
49 70  
50  
51 71  
52  
53 72  
54  
55 73  
56  
57 74  
58  
59  
60



1  
2  
3  
4 75 **Abstract**

5  
6 76 **Purpose:** A prospective cohort study for pregnant women, the Maternity Log  
7  
8 77 study (MLOG), was designed to construct a time-course high-resolution  
9  
10 78 reference catalog of bioinformatic data in pregnancy and explore the  
11  
12 79 associations between genomic and environmental factors and the onset of  
13  
14 80 pregnancy complications, such as hypertensive disorders of pregnancy,  
15  
16 81 gestational diabetes mellitus, and preterm labor, using continuous lifestyle  
17  
18 82 monitoring combined with multi-omics data on the genome, transcriptome,  
19  
20 83 proteome, metabolome, and microbiome.

21  
22 84 **Participants:** Pregnant women were recruited at the timing of first routine  
23  
24 85 antenatal visits at Tohoku University Hospital, Sendai, Japan between  
25  
26 86 September 2015 and November 2016. Of the eligible women who were invited,  
27  
28 87 65.4% agreed to participate, and a total of 302 women were enrolled. The  
29  
30 88 inclusion criteria were age  $\geq$  20 years and the ability to access the internet using  
31  
32 89 a smartphone in the Japanese language.

33  
34 90 **Findings to date:**

35  
36 91 Study participants uploaded daily general health information including quality of  
37  
38 92 sleep, condition of bowel movements, and the presence of nausea, pain, and  
39  
40 93 uterine contractions. Participants also collected physiologic data, such as body  
41  
42 94 weight, blood pressure, heart rate, and body temperature, using multiple home  
43  
44 95 healthcare devices. The mean upload rate for each lifelog item was ranging  
45  
46 96 from 67.4 % (fetal movement) to 85.3% (physical activity) and the total number  
47  
48 97 of data points was over 6 million. Biospecimens, including maternal plasma,  
49  
50 98 serum, urine, saliva, dental plaque, and cord blood, were collected for multi-  
51  
52 99 omics analysis.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 100 **Future plans:**

5  
6 101 Lifelog and multi-omics data will be used to construct a time-course high-  
7  
8 102 resolution reference catalog of pregnancy. The reference catalog will allow us to  
9  
10 103 discover relationships among multi-dimensional phenotypes and novel risk  
11  
12 104 markers in pregnancy for the future personalized early prediction of pregnancy  
13  
14 105 complications.  
15

16  
17 106

18  
19  
20 107 **Strengths and limitations of this study:**

21  
22 108 ■ This is the first study designed to collect longitudinal lifelog information  
23  
24 109 through healthcare devices, self-administered questionnaires using  
25  
26 110 smartphones, and varieties of biospecimens throughout pregnancy.  
27

28  
29 111 ■ Longitudinal, continuous, individual lifelog data with a high acquisition rate  
30  
31 112 will enable us to assess dynamic physiological changes throughout  
32  
33 113 pregnancy.  
34

35  
36 114 ■ Multi-omics data will make it possible to understand the complex  
37  
38 115 mechanisms of multifactorial pregnancy-related diseases.  
39

40  
41 116 ■ Potential limitations are the limited sample size and participant recruitment  
42  
43 117 only at a tertiary hospital for high-risk populations.  
44

45 118 ■ Inclusion criteria of the present study limited the eligibility to pregnant  
46  
47 119 women with age >20 years and the ability to access the internet using a  
48  
49 120 smartphone.  
50

51 121

52 122

53 123

54 124  
55  
56  
57  
58  
59  
60

## 125 INTRODUCTION

126 The incidence of pregnancy-related disorders, including hypertensive disorders  
127 of pregnancy (HDP), gestational diabetes mellitus (GDM), and preterm delivery  
128 has been increasing worldwide [1-4]. These multifactorial conditions are caused  
129 by an interaction of genetic factors and environmental factors [5,6]. Recent  
130 reports suggest that continuous lifestyle monitoring using wearable biosensors  
131 provides important information on latent physiologic changes that are exhibited  
132 prior to the onset of disease [7]. Using these monitors, environmental factors  
133 may be estimated more accurately than by using conventional questionnaires.

134 For these reasons, we have designed a prospective cohort study for pregnant  
135 women, the Maternity Log study (MLOG). In this study, pregnant women upload  
136 daily information and physiologic data using multiple home healthcare devices.  
137 In addition, a variety of biospecimens are collected for multi-omics analysis.

138 To the best of our knowledge, this study will be the first to integrate multi-  
139 omics analyses and objective data on environmental factors, including daily  
140 lifelog data, in pregnant women. This study may demonstrate correlations  
141 between specific lifelog patterns and pregnancy related physiological changes,  
142 such as blood pressure, gestational weight gain, and onset of obstetric  
143 diseases. Furthermore, studies on associations among lifelog patterns, plasma  
144 and urine metabolomes, transcriptomes, and genomic variations may reveal  
145 relationships among multi-dimensional phenotypes, and lead to identification of  
146 novel risk markers in pregnancy for the future personalized early prediction of  
147 pregnancy complications, e.g. hypertensive disorders of pregnancy, gestational  
148 diabetes, and preterm labor.

149

## 150 COHORT DESCRIPTION

### 151 Study setting

152 The aim of the MLOG study is to construct a time-course high-resolution  
153 reference catalog of bioinformatic data in pregnancy and thereby develop  
154 methods for early prediction of obstetric complications, through integrated  
155 analysis of daily lifelogs and multi-omics data, *i.e.*, maternal genomes,  
156 transcriptomes, metabolomes, and oral microbiomes.

157 The MLOG study is a prospective, add-on cohort study, built on a birth- and 3-  
158 generation cohort study established by the Tohoku Medical Megabank  
159 Organization (TMM BirThree Cohort Study) [8] in order to elucidate the  
160 mechanisms of complicated multifactorial diseases in mothers and children in  
161 the wake of the Great East Japan Earthquake in 2011. Epidemiological data  
162 from extensive questionnaire surveys and accurate clinical records, including  
163 birth outcomes, can be abstracted from the integrated biobank of the Tohoku  
164 Medical Megabank Organization (ToMMo) [8]. TMM BirThree Cohort Study was  
165 started in July 2013 in one obstetric clinic and expanded throughout Miyagi  
166 Prefecture, and approximately 50 obstetric clinics and hospitals (including  
167 Tohoku University Hospital) participated in the recruiting process. We planned  
168 to recruit 20,000 pregnant women as probands, and her family members from  
169 three generations, a total of over 70,000 participants [8]. Written informed  
170 consent was obtained from all participants by the genome medical research  
171 coordinators (GMRCs). The MLOG study was conducted in accordance with the  
172 Declaration of Helsinki and approved by the ethics committees of Graduate  
173 School of Medicine (2014-1-704) and ToMMo (22017-1-085), Tohoku University  
174 under a collaborative research agreement among ToMMo, Tohoku University

1  
2  
3  
4 175 and NTT DoCoMo, Inc. (Tokyo, Japan).

5  
6 176

## 7 8 177 **Patient and public involvement**

9  
10 178 Patients or the public were not directly involved in the development of the  
11  
12 179 research question or the design of the study. The main results will be made  
13  
14 180 available in the public domain.

15  
16  
17 181

## 18 19 20 182 **Participants**

21  
22 183 Participants were recruited at a first routine antenatal visit at Tohoku University  
23  
24 184 Hospital, Sendai, Japan between September 2015 and November 2016. A  
25  
26 185 flowchart of the recruitment process is shown in Figure 1. GMRCs at Tohoku  
27  
28 186 University Hospital approached eligible pregnant women for TMM BirThree  
29  
30 187 Cohort Study (n= 631), and patients who already agreed to participate in TMM  
31  
32 188 BirThree Cohort Study (n=513) were assessed for eligibility for the MLOG study.  
33  
34 189 Finally, 462 pregnant women were asked to provide informed consent for the  
35  
36 190 MLOG study. A total of 302 women were enrolled. The inclusion criteria were  
37  
38 191 the age  $\geq$  20 years and the ability to access the internet using a smartphone in  
39  
40 192 the Japanese language. Participants were excluded after enrollment if  
41  
42 193 termination of pregnancy, abortion, or transfer to another institution for  
43  
44 194 emergency care occurred before delivery, or if they withdrew consent for any  
45  
46 195 reason.

47  
48  
49  
50  
51 196

## 52 53 197 **Outline of study protocol**

54  
55 198 The study protocol consisted of blood and urine sampling, saliva and dental  
56  
57 199 plaque sampling, self-administered daily lifelog data collection, and data upload  
58  
59  
60

1  
2  
3  
4 200 from multiple healthcare devices through a smartphone. An overview of the  
5  
6 201 protocol is provided in Figure 2. In Japan, routine antenatal visits, including  
7  
8 202 ultrasounds, are scheduled every 4 weeks from early pregnancy (< 12 weeks)  
9  
10 203 to 23 weeks of gestation, every 2 weeks from 24 to 35 weeks, and every week  
11  
12 204 from 36 weeks to delivery [9]. Lifelog data collection was continued throughout  
13  
14 205 pregnancy and until 1 month after delivery. Optional data collection could be  
15  
16 206 continued up to 180 days after delivery.  
17  
18  
19  
20  
21

207

### 208 **Blood and urine sampling**

209 Blood samples were collected 3 times from each participant; the first sample  
210 was collected between 12 and 24 weeks of gestation, the second between 24  
211 and 36 weeks, and the third at 1 month after delivery. A maximum of 13 mL of  
212 blood was collected each time, from which serum and plasma were separated  
213 to be stored at -80°C until the time of analysis. An aliquot of blood (2.5 mL) was  
214 stored in a PAXgene® tube (Becton, Dickinson and Company, Franklin Lakes,  
215 NJ, USA) at -80°C until the time of RNA extraction for transcriptome analysis.  
216 Genomic DNA was extracted from mononuclear cells using an Autopure®  
217 extractor (Qiagen, Venlo, The Netherlands). Approximately 10 mL of cord blood  
218 was collected from the umbilical vein in a PAXgene® tube for storage at -80°C,  
219 and in an EDTA 2K tube (Becton, Dickinson and Company, Franklin Lakes, NJ,  
220 USA) for separation of plasma to be stored at -80°C. Urine samples (10 mL)  
221 were collected at each antenatal visit; when participants were admitted to the  
222 hospital ward, urine was collected once weekly. Urine samples were  
223 immediately transferred and stored at -80°C until the time of analysis.  
224

## 225 **Saliva and dental plaque sampling**

226 Samples of saliva and dental plaque were collected 3 times from each  
227 participant, at the same time points as blood collection. Approximately 3 mL of  
228 saliva was collected using a 50-mL conical centrifuge tube (Corning, Inc.,  
229 Corning, NY, USA) and stored at -80°C until analysis. Dental plaque was  
230 sampled by brushing, suspended in 0.5 mL of Tris-EDTA (10 mM Tris, 1 mM  
231 EDTA; pH, 8.0), and immediately stored at -80°C until the time of sample  
232 processing.

233

## 234 **Lifelog data collection**

235 Based on previous publications on the utility for risk assessment of pregnancy-  
236 related diseases, we selected several lifelog parameters to employ in this study,  
237 *i.e.*, body temperature [10], home blood pressure [11], body weight [12], and  
238 physical activity (calorie expenditure) [13], as well as self-administered  
239 information such as sleep quality [14], condition of stool [15], severity of nausea  
240 [16], fetal movement [17], severity of pain [18], uterine contractions [19], and  
241 palpitations [20]. Body temperature, home blood pressure, body weight, and  
242 physical activity were uploaded from multiple healthcare devices through a  
243 smartphone. The self-administered information described above was input  
244 manually on mobile applications created for this study.

245 Data collection was started after obtaining informed consent and after giving  
246 detailed instructions for the use of the healthcare devices. These applications  
247 tracked quality of sleep; condition of stool using the Bristol Scale [21-23];  
248 severity of nausea using the Pregnancy-Unique Quantification of Emesis and  
249 nausea (PUQE) score [24,25]; headache, toothache, lumbago, and upper and



1  
2  
3  
4 250 lower abdominal pain using a numerical rating scale (NRS) score; the number  
5  
6 251 of perceived uterine contractions; palpitations; and fetal movement using a  
7  
8 252 modified count-to-10 fetal movement chart [26,27].  
9

10 253 Sleep quality was evaluated by the wakeup time, bedtime, sleep satisfaction  
11  
12 254 (ranked from satisfied to poor using a numeric scale of 0-4), and the number of  
13  
14 255 nocturnal awakenings (0-6).  
15

16  
17 256 The Bristol stool form scale was originally developed to assess constipation  
18  
19 257 and diarrhea [21, 22], and its use has been spread widely to evaluate functional  
20  
21 258 bowel disorders [22]. Using the Bristol scale, stool is classified into 7 types  
22  
23 259 according to cohesion and surface cracking [21, 22].  
24

25  
26 260 The PUQE score [24, 25] was developed to estimate the severity of nausea  
27  
28 261 and vomiting in pregnancy and quantifies the number of daily vomiting and  
29  
30 262 retching episodes and the length of nausea in hours (over the preceding 12 h).  
31  
32 263 The total score ranges from 3 (no symptoms) to 15, and higher scores are  
33  
34 264 correlated with increasing severity of nausea and vomiting [24, 25].  
35

36  
37 265 In the NRS score for headache, toothache, lumbago, and upper and lower  
38  
39 266 abdominal pain, the total score ranges from 0 (no pain) to 10 (maximum ever  
40  
41 267 experienced).  
42

43  
44 268 Uterine contractions and palpitations were evaluated using definitions  
45  
46 269 determined for the current study. Uterine contractions were assessed using the  
47  
48 270 number of perceived contractions per day, ranging from 0 to more than 5. The  
49  
50 271 count-to-10 method was originally developed to assess fetal well-being by  
51  
52 272 recording the time, in minutes, required to count 10 fetal movements [26]. More  
53  
54 273 recently, a modified count-to-10 method has been proposed: pregnant women  
55  
56 274 are advised to start counting when they feel the first movement, then record the  
57  
58  
59  
60



1  
2  
3  
4 275 time required to perceive an additional 9 movements [27]. Pregnant women are  
5  
6 276 encouraged to select a 2-hour period when they feel active fetal movements  
7  
8 277 and are instructed to count kicking and rolling movements in a favorable  
9  
10 278 maternal position after 24 weeks of gestation.

11  
12  
13 279 The applications also collected dietary logs and the medications taken on the  
14  
15 280 day before and the day of the antenatal visit, on which blood or urine samples  
16  
17 281 were collected.

18  
19  
20 282 Daily home blood pressure, body weight, body temperature, and physical  
21  
22 283 activity were measured as described below with home healthcare devices, and  
23  
24 284 uploaded through wireless communications using mobile applications on a  
25  
26 285 smartphone. Daily home blood pressure was measured twice daily using an  
27  
28 286 HEM-7510 monitor (OMRON Healthcare Co., Ltd., Kyoto, Japan): within 1 hour  
29  
30 287 of awakening in the morning and just before going to bed at night. Body weight  
31  
32 288 was measured using an HBF-254C meter (OMRON Healthcare Co., Ltd.) once  
33  
34 289 daily within 1 hour of awakening in the morning. Daily body temperature was  
35  
36 290 evaluated using an MC-652LC digital thermometer (MC-652LC; OMRON  
37  
38 291 Healthcare Co., Ltd.) just after awakening. Physical activity was assessed using  
39  
40 292 an HJA-403C pedometer (HJA-403C; OMRON Healthcare Co., Ltd.) to count  
41  
42 293 steps and calculate calorie expenditure.

43  
44  
45  
46  
47 294

## 48 49 295 **Clinical and epidemiological information**

50  
51  
52 296 Baseline clinical information and maternal and neonatal outcomes (*e.g.*,  
53  
54 297 maternal age, clinical data and findings from each antenatal visit, gestational  
55  
56 298 age at delivery, type of delivery, birth weight, and maternal and fetal  
57  
58 299 complications) were obtained from the medical records of Tohoku University  
59  
60

1  
2  
3  
4 300 Hospital. Epidemiological data, including extensive questionnaire surveys by  
5  
6 301 TMM BirThree Cohort Study can be obtained from the ToMMo integrated  
7  
8 302 biobank [8].  
9

10 303

### 13 304 **Database**

15 305 A customized laboratory information management system (LIMS) was  
16  
17 306 established to track all biospecimens. All data were transferred to the TMM  
18  
19 307 integrated database after 2-step anonymisation in a linkable fashion.  
20  
21  
22 308 Data handling was strictly regulated under HIPAA (Health Insurance Portability  
23  
24 309 and Accountability Act of 1996, United States Security and Privacy Rules) [28,  
25  
26 310 29] and the Act on the Protection of Personal Information [30]. Security control  
27  
28 311 at our facility has been described previously [31].  
29  
30

31 312

### 33 313 **Omics analysis**

35 314 Whole-genome sequencing

36  
37  
38 315 To minimize amplification bias, we adopted a PCR-free library preparation  
39  
40 316 method. After performing library quality control using the quantitative MiSeq  
41  
42 317 method [32], libraries were sequenced on HiSeq 2500 Sequencing System  
43  
44 318 (Illumina, Inc., San Diego, CA, USA) to generate 259-bp, paired-end reads. We  
45  
46 319 generated the sequencing data at over 12.5x coverage on average, and we  
47  
48 320 identified variants using the alignment tool BWA-MEM (ver. 0.7.5a-r405) with  
49  
50 321 the default option. Single nucleotide variants (SNVs) and indels were jointly  
51  
52 322 called across all samples using Genome Analysis Tool Kit's HaplotypeCaller  
53  
54 323 (ver. 3.8). Default filters were applied to SNV and indel calls using the GATK's  
55  
56 324 Variant Quality Score Recalibration (VQSR) approach. The human reference  
57  
58  
59  
60

1  
2  
3  
4 325 genome was GRCh37/hg19 with the decoy sequence (hs37d5) and NC\_007605  
5  
6 326 (Human Gamma Herpesvirus 4). The complete fasta file named  
7  
8 327 hg19\_tommo\_v2.fa is available from iJGVD website  
9  
10 328 (<http://ijgvd.megabank.tohoku.ac.jp>) [33]. For the quality assurance, we have  
11  
12 329 checked the ratio of the bases with the phred quality score over 30, the total  
13  
14 330 variant numbers in each chromosome, and the ratio of transitions to  
15  
16 331 transversions for a pair of sequences.  
17  
18  
19  
20  
21

### 22 332 Transcriptome

23  
24 334 Whole blood was collected using the PAXgene® RNA tube, which is widely  
25  
26 335 used for transcriptome analysis. After storage at -80°C, total RNA was purified  
27  
28 336 with PAXgene Blood RNA Kit® (Qiagen, Venlo, The Netherlands) using  
29  
30 337 QiaSymphony® (Qiagen). Total RNA was reverse-transcribed using an oligo-dT  
31  
32 338 primer. We used TruSeq DNA PCR-Free Library Preparation Kit (Illumina, Inc.)  
33  
34 339 for library preparation for sequencing with HiSeq 2500 Sequencing System. For  
35  
36 340 the quality assurance, we randomly selected 11 samples in one batch (usually  
37  
38 341 48 samples) and checked an RNA integrity number (RIN) (or an RIN equivalent)  
39  
40 342 using BioAnalyzer® or Tape Station® (both from Agilent Technologies, Santa  
41  
42 343 Clara, CA, USA). The batch with RIN (or an RIN equivalent) higher than 7.0 for  
43  
44 344 all tested samples was used for the downstream analysis. The minimum  
45  
46 345 threshold for the total sequence reads for each sample was set to thirty millions.  
47  
48 346 For computing a series of quality control metrics for RNA-seq data, RNA-SeQC  
49  
50 347 was used to check the quality of sequence reads [34].  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

349 Plasma and urine metabolome

1  
2  
3  
4 350 *Nuclear magnetic resonance (NMR) spectroscopy*

5  
6 351 All NMR measurements for metabolome analysis were conducted at 298 K on a  
7  
8 352 Bruker Avance 600 MHz spectrometer equipped with a SampleJet sample  
9  
10 353 changer (Bruker Corp., Billerica, MA, USA) [35]. Standard 1-dimensional  
11  
12 354 nuclear Overhauser enhancement spectroscopy (1D NOESY) and Carr-Purcell-  
13  
14 355 Meiboom-Gill (CPMG) spectra were obtained for each plasma or urine sample.  
15  
16 356 All spectra for plasma or urine samples were acquired using 16 scans and 32 k  
17  
18 357 of complex data points. All data were analyzed using the TopSpin 3.5 (Bruker  
19  
20 358 Corp.) and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta,  
21  
22 359 Canada) programs. All spectra were referenced to an internal standard (DSS-  
23  
24 360 d6). As necessary, those spectra were aligned using hierarchical cluster-based  
25  
26 361 peak alignment method, which is implemented as an R package called "speaq"  
27  
28 362 [36].  
29  
30  
31  
32  
33

34 363  
35  
36 364 *Gas chromatography-tandem mass spectrometry (GC-MS/MS)*

37  
38 365 Sample preparation for plasma and urine (50  $\mu$ L each) was performed using a  
39  
40 366 Microlab STARlet robot system (Hamilton, Reno, NV, USA) followed by the  
41  
42 367 methods previously reported by Nishiumi [37, 38]. The resulting deproteinized  
43  
44 368 and derivatized supernatant (1  $\mu$ L) was subjected to GC-MS/MS, performed on  
45  
46 369 a GC-MS TQ-8040 system (Shimadzu Corp., Kyoto, Japan). The compound  
47  
48 370 separation was performed using a fused silica capillary column (BPX-5; 30 m  $\times$   
49  
50 371 0.25 mm inner diameter; film thickness, 0.25  $\mu$ m; Shimadzu Corp, Kyoto,  
51  
52 372 Japan). Metabolite detection was performed using Smart Metabolites Database  
53  
54 373 (Shimadzu Corp.) that contained the relevant multiple reaction monitoring  
55  
56 374 (MRM) method file and data regarding the GC analytical conditions, MRM  
57  
58  
59  
60

1  
2  
3  
4 375 parameters, and retention index employed for the metabolite measurement. The  
5  
6 376 database used in this study included data on 475 peaks from 334 metabolites.  
7  
8 377 All peaks of metabolites detected from each sample was annotated and  
9  
10 378 analyzed using Traverse MS® (Reifycs Inc., Tokyo, Japan). Then, two types of  
11  
12 379 normalization were performed to these annotated metabolites. The first  
13  
14 380 normalization was performed using the peak of 2-isopropylmalic acid as an  
15  
16 381 internal standard which was added to each sample before analysis with GC-  
17  
18 382 MS/MS. Then the second normalization was performed using quality control  
19  
20 383 (QC) samples which were injected after every 12 study samples according to  
21  
22 384 the RQC normalization methods [39]. Normalized values of each metabolite in  
23  
24 385 the QC samples were assessed by calculating coefficients of variation (CVs),  
25  
26 386 and metabolites with CVs over 20% were eliminated.

31 387

### 33 388 Oral Microbiome

35 389 Analysis of oral microbiome was conducted by previously reported protocols  
36  
37 [40]. In brief, saliva was collected in a 50-mL tube. Dental plaque was sampled  
38  
39 by participants by brushing teeth with a sterilized toothbrush, and then  
40  
41 391 by suspending it in 0.5 mL Tris-EDTA for collection. Both samples were stored at -  
42  
43 392 80°C until the time of processing. DNA was extracted from saliva and dental  
44  
45 393 plaque by standard glass bead-based homogenization and subsequent  
46  
47 394 purification with a silica-membrane spin-column using PowerSoil DNA Isolation  
48  
49 395 Kit (Mo Bio Laboratories, Carlsbad, CA, USA). DNA was eluted from the spin  
50  
51 396 column with 30- $\mu$ L RNase-free water (Takara Bio, Inc., Shiga, Japan), and  
52  
53 397 stored at -20°C after determining the amount and purity of DNA with a  
54  
55 398 Nanodrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).  
56  
57 399  
58  
59  
60

1  
2  
3  
4 400 Using DNA extracted from saliva or dental plaque as a template, a part of the  
5  
6 401 V4 variable region of the bacterial 16S rRNA gene was amplified by 2-step  
7  
8 402 PCR. Tag-indexed PCR products thus obtained were subjected to multiplex  
9  
10 403 amplicon sequencing using MiSeq System with MiSeq Sequencing Reagent Kit,  
11  
12 404 v3 (Illumina, Inc.) according to the manufacturer's instructions. For the quality  
13  
14 405 assurance, the minimum threshold of the total sequence reads for each sample  
15  
16 406 was set to ten thousands, and the principal component analysis was used to  
17  
18 407 eliminate outliers.  
19  
20  
21

22 408

### 23 24 409 **Outcomes**

25  
26 410 The following obstetric complications represented the primary outcomes.  
27  
28 411 Gestational age was confirmed by measuring fetal crown rump length from 9 to  
29  
30 412 13 weeks of gestation using transvaginal ultrasound. HDP was defined as  
31  
32 413 gestational hypertension, preeclampsia, superimposed preeclampsia, or chronic  
33  
34 414 hypertension [41,42]. Preterm birth was defined as spontaneous preterm labor,  
35  
36 415 medically induced preterm labor, or preterm premature rupture of membranes  
37  
38 416 resulting in preterm birth at less than 37 weeks of gestation. GDM was  
39  
40 417 diagnosed according to the International Association of the Diabetes and  
41  
42 418 Pregnancy Study Groups (IADPSG) criteria [43]. The secondary outcomes were  
43  
44 419 maternal body weight, blood pressure, physical activity, lifestyle changes,  
45  
46 420 perinatal mental disorders, fetal growth, fetal movement, and birth weight.  
47  
48  
49

50 421

### 51 52 422 **Sample size calculation**

53  
54 423 At this time, there is little reliable evidence to demonstrate how time-dependent  
55  
56 424 trends of longitudinal dense data would differ by pregnancy outcomes.  
57  
58  
59  
60

1  
2  
3  
4 425 Therefore, a priori sample size calculation is not provided in the present study.  
5  
6 426 However, considering that one of the main purposes of the MLOG study is to  
7  
8 427 explore the relationship between patterns of longitudinal home blood pressure  
9  
10 428 and the onset of HDP, we estimated a required sample size as follows. Based  
11  
12 429 on the HDP incidence of approximately 10% at Tohoku University Hospital, with  
13  
14 430 a statistical power of 90% and a significance level of 5%, a sample of 250  
15  
16 431 participants is required to detect a 5-mmHg difference in average home blood  
17  
18 432 pressure (with a 7-mmHg standard deviation) in the HDP group. To allow for  
19  
20 433 15% attrition and withdrawals during pregnancy, a minimum of 300 participants  
21  
22 434 at baseline was required.  
23  
24  
25  
26  
27  
28

### 29 436 **Statistical analysis of longitudinal lifelog data**

30  
31 437 One of the major advantages of the MLOG study is the dense information for  
32  
33 438 each participant. Especially, time points for lifelog data collection are highly  
34  
35 439 dense for each participant. For these datasets, per-person analysis of dynamic  
36  
37 440 relationships between variables can be applied [44]. Vector autoregressive  
38  
39 441 (VAR) modeling is a promising solution to find the predicates for each outcome.  
40  
41 442 In addition, the Granger causality test can elucidate the temporal ordering of  
42  
43 443 dynamic relationship between two or more variables and indicate putative  
44  
45 444 causal associations [45]. Some types of lifelog data were generated  
46  
47 445 automatically; the others were manually input. We will first detect outlier data  
48  
49 446 points, depending on the type of each lifelog, and eliminate them. The missing  
50  
51 447 time-series lifelog data, ranging in 15-33% of the total data points, would be  
52  
53 448 imputed using the EM-imputation algorithm - e.g. Amelia library [46], after  
54  
55 449 normalising the data by data transformation if required. For downstream  
56  
57  
58  
59  
60



1  
2  
3  
4 450 analysis, the data might be collapsed with time scale, e.g. taking trimmed mean  
5  
6 451 or median for each week, month, or trimester.  
7

8 452  
9

### 10 453 **Statistical analysis of multi-omics data**

11  
12 454 The present study allows combination of longitudinal lifelog data with multi-  
13  
14 455 omics data. In contrast to single omics analysis, the multi-omics analysis would  
15  
16 456 reveal the complicated interactions between one and another. However, the  
17  
18 457 sample size for multi-omics analysis is usually relatively small. Dimension  
19  
20 458 reduction via unsupervised or supervised learning for each omics data would be  
21  
22 459 key ingredients to derive meaningful patterns from high dimensional data sets.  
23  
24 460 Also, obtaining low dimensional representations provides a mean to deal with  
25  
26 461 the multiple testing problem by decreasing number of statistical tests. For gene  
27  
28 462 expression data, surrogate variable analysis [47] and sparse factor analysis [48]  
29  
30 463 are frequently used to capture unknown batch effects in advance to expression  
31  
32 464 quantitative trait locus (eQTL) analysis. The extracted factors can be removed  
33  
34 465 from raw expression data to increase power for detecting associated genes  
35  
36 466 [49]. Several unsupervised clustering methods [50,51,52] would be also  
37  
38 467 applicable to obtain hidden patterns from dense time-course lifelog  
39  
40 468 measurements, which might be related to pregnancy complications. Recently  
41  
42 469 developed multi-view factor analysis approaches [53,54] have been used to  
43  
44 470 integrate heterogeneous omics data to identify essential components to  
45  
46 471 distinguish disease subtypes from few hundreds of samples. This line of  
47  
48 472 approach would be a promising way to characterize biological status such as  
49  
50 473 gestational age, and to predict clinical outcomes such as spontaneous preterm  
51  
52 474 birth.  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3  
4 475 Standard analyses would be also applicable for the selected variables and  
5  
6 476 extracted factors (features). The association of outcomes with each feature will  
7  
8 477 be analyzed using statistical hypothesis tests such as Welch's t-test, Fisher's  
9  
10 478 exact test, the Chi-square test, and others as appropriate. Multiple logistic  
11  
12 479 regression modeling will be used to adjust for confounders and to assess  
13  
14 480 whether each feature or combination of features can be used to predict  
15  
16 481 outcomes. Stepwise selection algorithms or regularized algorithms (*e.g.*,  
17  
18 482 LASSO, ridge regression, or elastic net) will be used to select the optimal  
19  
20 483 number of contributing features that maximize the predictive power using the  
21  
22 484 leave-1-out cross validation or K-fold cross validation methods.

26 485 Individual genetic features may have an effect on outcomes; therefore, some  
27  
28 486 aggregated genetic risk score should be included in the prediction model. For  
29  
30 487 example, SNVs, including rare variants in or around a chromosome region of a  
31  
32 488 known or estimated risk gene, could be aggregated by considering their impacts  
33  
34 489 on biological function of the gene or their minor allele frequencies in the  
35  
36 490 population. However, this study is limited in the number of study participants,  
37  
38 491 and the aggregated risk score might therefore contribute only slightly to the  
39  
40 492 predictive power. To create a more reliable risk score, the estimates from other  
41  
42 493 large-scale cohort data using polygenic score tools, *e.g.*, PRSice [55], could be  
43  
44 494 used for this study.

49 495

## 52 496 **FINDINGS TO DATE**

### 54 497 **Clinical background**

56 498 A total of 302 women were enrolled, and the mean gestational weeks of  
57  
58 499 recruitment was  $16.4 \pm 4.9$  weeks (mean  $\pm$  SD). A total of 285 participants have

500 been followed up to delivery; their baseline clinical characteristics are described  
 501 in Table 1. The mean maternal age at delivery was  $33.3 \pm 4.9$  years. As for  
 502 educational levels, 62% of the participants were high school graduates with or  
 503 without vocational college education, and 21% had a college degree. The  
 504 majority were employed (65%) in early pregnancy, and about 40% had a high  
 505 household income (over 6 million yen per year). Approximately 42% of the  
 506 participants were over 35 years of age, 51% were parous, and 22% were  
 507 overweight or obese by their prepregnancy body mass indices ( $\text{BMI} \geq 25$   
 508  $\text{kg/m}^2$ ). Overall, 8.4% of the participants had HDP, and 5.6% underwent  
 509 spontaneous preterm birth. On average, infants were delivered at  $38.0 \pm 2.3$   
 510 weeks of gestation with a mean birth weight of  $2907 \pm 572$  g. The rate of low  
 511 birth weight was 18%. Mean gestational weeks of the first and second blood  
 512 sampling were  $17.0 \pm 5.0$  and  $27.5 \pm 2.5$ , respectively. The third blood sampling  
 513 was performed at  $31.1 \pm 3.0$  days after delivery on average. The length of  
 514 enrollment ranged from 90 to 396 days with a mean of  $216 \pm 61$  days.

515  
 516

**Table 1.** Participant characteristics

Characteristics	Value
<b>Maternal (n = 285)</b>	
• Age at delivery, y, mean (SD)	33.3 ( $\pm 4.9$ )
• Age at delivery, y, n (%)	
20-24	12 (4.2)
25-29	45 (15.8)
30-34	107 (37.5)
35-39	90 (31.6)
40-44	30 (10.5)
45-49	1 (0.4)
• Education (n=81) n (%)	
Elementary school / Junior high school	5 (6.2)
High school	35 (43.2)

1		
2		
3		
4	Vocational college	23 (28.4)
5		
6	College degree and above	17 (21.0)
7		
8	Others	1 (1.2)
9		
10	Data not available	204
11		
12		
13		
14	• Occupation (n=270) n (%)	
15		
16	Housewife or unemployed	93 (34.4)
17		
18	Employed	175 (64.8)
19		
20	Student	2 (0.7)
21		
22		
23	• Annual household income, yen (n=248) n (%)	
24		
25	< 2 million	17 (6.9)
26		
27	2-4 million	59 (23.8)
28		
29	4-6 million	73 (29.4)
30		
31	6-8 million	51 (20.6)
32		
33	8-10 million	22 (8.9)
34		
35	> 10 million	26 (10.5)
36		
37		
38	• Parity, n (%)	
39	0	140 (49.1)
40	1	93 (32.6)
41	≥ 2	52 (18.2)
42		
43	• Prepregnancy BMI*, kg/m <sup>2</sup> , mean (SD)	22.7 (± 5.1)
44		
45	• Prepregnancy BMI, kg/m <sup>2</sup> , n (%)	
46	< 18.5	36 (12.6)
47	18.5-24.9	186 (65.3)
48	25.0-29.9	34 (11.9)
49	≥ 30.0	29 (10.2)
50		
51	• Gestational weeks at delivery, mean (SD)	38.0 (± 2.3)
52		
53		
54	• Mode of delivery, n (%)	
55	Noncesarean	179 (62.8)
56	Cesarean	106 (37.2)
57		
58	• Pregnancy complication, n (%)	
59	Hypertensive disorder of pregnancy	24 (8.4)
60		

	Spontaneous preterm birth	16 (5.6)
	<b>Neonatal (n = 300)</b>	
	• Birth weight, g, mean (SD)	2907 (± 572)
	• Sex, n (%)	
	Male	168 (56)
	Female	132 (44)
	• Low-birth weight (< 2500 g), n (%)	54 (18)
517	<hr/> *BMI, body mass index	

518

519 **Data acquisition**

520 The percentage of data uploads as of June 2017 was calculated for the 285  
521 final study participants. For each lifelog item, the upload rate for each  
522 participant was calculated from the total number of days of actual uploads  
523 divided by the number of days from enrollment to delivery. The mean upload  
524 rate for each lifelog item was 85.3% (physical activity), 82.1% (body weight),  
525 80.4% (body temperature), 78.0% (morning home blood pressure), 71.6%  
526 (evening home blood pressure), 83.5% (sleep quality), 82.1% (condition of  
527 stool, severity of pain, severity of nausea, uterine contractions, and  
528 palpitations), and 67.4% (fetal movement) (Figure 3).

529

530 **Number of data points**

531 The total number of collected data points as of June 2017 was calculated for the  
532 285 final study participants. The approximate number of registered data points  
533 was 86 000 for body weight, 324 000 points for home diastolic and systolic  
534 blood pressure, 86 000 for physical activity, and 74 000 for body temperature.  
535 When physical conditions such as stool condition, severity of pain, and fetal  
536 movement were combined, the total number of data points was over 6 million.

537

## 538 **STRENGTHS AND LIMITATIONS**

539 Herein, we have described the rationale, design, objective, data collection  
540 methods, and interim results of the MLOG study. The study was launched in  
541 September 2016, and baseline data collection ended in June 2017. A total of  
542 285 participants uploaded lifelog data throughout pregnancy with a high data  
543 acquisition rate and over 6 million total data points. Biospecimens for multi-  
544 omics analysis were satisfactorily collected and all tracked by LIMS.

545 There are three noteworthy features in the MLOG study. First, it is a  
546 prospective add-on cohort study based on TMM BirThree Cohort Study, with a  
547 full series of epidemiological data and a highly structured follow-up system for  
548 mothers, newborns, and families [8]. Second, we have successfully collected  
549 longitudinal, continuous, individual lifelog data with a high acquisition rate,  
550 which will enable us to assess dynamic changes in physiologic conditions  
551 throughout pregnancy. Third, multi-omics data will make it possible to fully  
552 understand the complex mechanisms of multifactorial pregnancy-related  
553 diseases and to overcome the unpredictability of these complications.

554 Prediction models using clinical and epidemiological information and  
555 circulating factors for pregnancy-related diseases have been developed  
556 extensively [56], and risk-assessment approaches using clinical information  
557 have also been developed [57, 58]. However, there is a lack of evidence for the  
558 benefits of these predictive models for routine clinical use [59]. Once the  
559 likelihood of a pregnancy-related disorder is estimated with high sensitivity and  
560 specificity, evidence-based clinical interventions could reduce the rate of  
561 maternal and neonatal morbidity and mortality [60]. Therefore, an early-  
562 prediction algorithm that can be used with a high level of confidence is needed

1  
2  
3  
4 563 to obtain better outcomes for patients with pregnancy complications.

5  
6 564 Recently, several studies of sample sizes comparable with ours, exploiting  
7  
8 565 lifelog or multi-omics data were reported. One of the studies analyzed lifelog  
9  
10 566 and multi-omics data, collected from 108 individuals at three time points during  
11  
12 567 a nine-month period [61]. In their study, several remarkable relationships were  
13  
14 568 identified among physiological and multi-omics data through integrated  
15  
16 569 analyses. Another study investigated genome-wide associations between  
17  
18 570 genetic variants and gene expression levels across 44 human tissues from a  
19  
20 571 few hundreds of postmortem donors [49]. They studied both cis-eQTL (within 1  
21  
22 572 Mb of target-gene transcription start sites) and trans-eQTLs (more distant from  
23  
24 573 target genes or on other chromosomes) with 350 whole blood samples, and  
25  
26 574 thereby identified 5,862 cis-eQTL and one trans-eQTL associations. These  
27  
28 575 previous studies indicate that our time-course high-resolution reference catalog  
29  
30 576 with 285 pregnant women would be well applicable to high-dimensional data  
31  
32 577 analyses such as searches for quantitative trait loci and molecular risk markers.

33  
34 578 Potential limitation of the present study is participant recruitment only at  
35  
36 579 Tohoku University Hospital that is one of the tertiary hospitals in Miyagi  
37  
38 580 Prefecture for high-risk populations. Therefore, the sample size is limited, and  
39  
40 581 the results might not be applicable to the general populations. Inclusion criteria  
41  
42 582 of the present study limited the eligibility to pregnant women with age >20 years  
43  
44 583 and the ability to access the internet using a smartphone. Therefore, results of  
45  
46 584 the present study might not be applicable to pregnancies with lower coverage of  
47  
48 585 smartphone use.

49  
50 586 Hopefully, our study will result in the development of a novel stratification  
51  
52 587 model for pregnancy-related diseases employing multi-omics and lifelog data.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 588 The MLOG study will enable us to construct a time-course high-resolution  
5  
6 589 reference catalog of wellness and multi-omics data from pregnant women and  
7  
8 590 thereby develop a personalized predictive model for pregnancy complications.  
9  
10 591 Progressive data sharing and collaborative studies would make it possible to  
11  
12 592 establish a standardized early-prediction method through large clinical trials.  
13  
14

15 593

## 17 594 **COLLABORATION**

19 595 We are very much interested in collaborating with other research groups and  
20  
21  
22 596 are open for specific and detailed proposals approved by the institutional ethical  
23  
24 597 review committee. We are planning to share the full data of the MLOG study in  
25  
26 598 the TMM biobank [8] by the end of 2022, and a portion of the data have been  
27  
28 599 distributed to researchers approved by the Sample and Data Access Committee  
29  
30  
31 600 of the biobank.  
32

33 601

## 36 602 **Author affiliations**

37  
38 603 1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryomachi,  
39  
40 604 Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

41  
42  
43 605 2 Tohoku University Graduate School of Medicine, 1-1, Seiryomachi, Aoba-ku,  
44  
45 606 Sendai, Miyagi, 980-8574, Japan.

46  
47 607 3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,  
48  
49 608 Kanagawa, Japan 239-8536.

50  
51  
52 609 4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-  
53  
54 610 ku, Sendai, Miyagi 981-8558, Japan.

55  
56 611 5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku,  
57  
58 612 Hiroshima 732-0815, Japan.  
59  
60

1  
2  
3  
4 613 6 Tohoku University Hospital, 1-1, Seiryomachi, Aoba-ku, Sendai, Miyagi, 980-  
5  
6 614 8574, Japan.

7  
8 615

9  
10 616

## 11 617 **Acknowledgements**

12  
13 618 The authors would like to thank all the MLOG study participants, the staff of the  
14  
15 619 Tohoku Medical Megabank Organization, Tohoku University (a full list of  
16  
17 620 members is available at: <http://www.megabank.tohoku.ac.jp/english/a161201/>),  
18  
19 621 and the Department of Obstetrics and Gynecology, Tohoku University Hospital,  
20  
21 622 for their efforts and contributions. The MLOG study group also included Chika  
22  
23 623 Igarashi, Motoko Ishida, Yumiko Ishii, Hiroko Yamamoto, Akiko Akama, Kaori  
24  
25 624 Noro, Miyuki Ozawa, Yuka Narita, Junko Yusa, Miwa Meguro, Michiyo Sato,  
26  
27 625 Miyuki Watanabe, Mai Tomizuka, Mika Hotta, Naomi Matsukawa, Makiko Sumii,  
28  
29 626 Ayako Okumoto, Yukie Oguma, Ryoko Otokozawa, Toshiya Hatanaka, Sho  
30  
31 627 Furuhashi, Emi Shoji, Tomoe Kano, Riho Mishina, and Daisuke Inoue.

32  
33 628

## 34 629 **Contributors**

35  
36 630 JS, DO, RY, TY, HM, OT, SKu, NY, SH, MN were involved in initial stages of  
37  
38 631 the strategy and design of study conception. JS, DO, RY, TY, OT, DS, SKo, SH,  
39  
40 632 MN: responsible for the draft of the manuscript. JS, DO, RY, TY, MW, MI, HM,  
41  
42 633 OT, SKu: recruitment and sample collection. DO, RY, TY, DS, TO, YT, YH,  
43  
44 634 TFS, TM, JK, FK, TIT, SO, NM, SKo, OT, MN: sample analysis, data  
45  
46 635 processing, and statistical analysis. JS, HH, NF, NM, SKo, OT, SKu, KK, SK,  
47  
48 636 NY, MY, SH, MN: advice and supervision of sample analysis. All authors have  
49  
50 637 contributed to revision and have approved the final manuscript, and agreed to



1  
2  
3  
4 638 be accountable for all aspects of the work in ensuring that questions related to  
5  
6 639 the accuracy or integrity of any part of the work are appropriately investigated  
7  
8 640 and resolved.  
9

10 641

### 11 642 **Funding**

12  
13 643 The present study was supported by NTT DoCoMo, Inc., with a collaborative  
14  
15 644 research agreement between NTT DoCoMo and ToMMo. This work was  
16  
17 645 supported in part by the Tohoku Medical Megabank Project from the Japan  
18  
19 646 Agency for Medical Research and Development (AMED) and the Ministry of  
20  
21 647 Education, Culture, Sports, Science and Technology (MEXT).  
22  
23  
24  
25

26 648

### 27 649 **Competing interests**

28  
29 650 This study was funded by NTT DoCoMo, Inc.  
30  
31 651 Daisuke Ochi, Takafumi Yamauchi, and Satoshi Hiyama are employees of NTT  
32  
33 652 DoCoMo, Inc. All other authors declare that they have no competing interests.  
34  
35  
36  
37

38 653

### 39 654 **Ethics approval and consent to participate**

40  
41  
42 655 TMM BirThree Cohort Study was approved by the ethics committees of the  
43  
44 656 Tohoku University (authorization numbers, 2013-4-103 and 2017-4-010). The  
45  
46 657 MLOG study was approved by the ethics committees of the Graduate School of  
47  
48 658 Medicine (2014-1-704) and the Tohoku Medical Megabank Organization (2017-  
49  
50 659 1-085), Tohoku University. Written informed consent was obtained from all  
51  
52  
53 660 participants.  
54  
55  
56  
57

58 661  
59  
60

1  
2  
3  
4 662 **Provenance and peer review**

5  
6 663 Not commissioned; externally peer reviewed.  
7

8 664  
9

10 665 **Data sharing statement**

11  
12 666 We are planning to share the full deidentified data of the MLOG study in the  
13  
14 667 TMM biobank. Investigators interested in the MLOG study are encouraged to  
15  
16 668 contact the corresponding authors, Dr. Junichi Sugawara at  
17  
18 669 [jsugawara@med.tohoku.ac.jp](mailto:jsugawara@med.tohoku.ac.jp) or Dr. Masao Nagasaki at  
19  
20  
21 670 [nagasaki@megabank.tohoku.ac.jp](mailto:nagasaki@megabank.tohoku.ac.jp). Currently, no additional data are available.  
22  
23

24 671  
25

26 672  
27

28 673  
29

30 674  
31

32 675  
33

34 676  
35

36 677  
37

38 678  
39

40 679  
41

42 680  
43

44 681  
45

46 682  
47

48 683  
49

50 684  
51

52 685  
53

54 686  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 6875  
6 6887  
8 6899  
10 690 **REFERENCES**11  
12 691 1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public13  
14 692 health perspective. *Diabetes Care*. 2007;30 Suppl 2: S141-6. doi:15  
16 693 10.2337/dc07-s206.17  
18 69419  
20 695 2. Beck S, Wojdyla D, Say L, et al. The worldwide incidence of preterm birth: a21  
22 696 systematic review of maternal mortality and morbidity. *Bull World Health Organ*.23  
24 697 2010; 88:31-8. doi: 10.2471/BLT.08.062554.25  
26 69827  
28 699 3. Duley L. The global impact of pre-eclampsia and eclampsia. *Semin Perinatol*.29  
30 700 2009; 33:130-7. doi: 10.1053/j.semperi.2009.02.010.31  
32 70133  
34 702 4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United35  
36 703 States, 1980-2010: age-period-cohort analysis. *BMJ*. 2013;347: f 6564. doi:37  
38 704 10.1136/bmj.f6564.39  
40 70541  
42 706 5. Waken RJ, de Las Fuentes L, Rao DC. A Review of the Genetics of43  
44 707 Hypertension with a Focus on Gene-Environment Interactions. *Curr Hypertens*45  
46 708 *Rep*. 2017; 19:23. doi: 10.1007/s11906-017-0718-1.47  
48 70949  
50 710 6. Ward K, Lindheimer MD. Genetic factors in the etiology of preeclampsia /51  
52 711 eclampsia. In: Chesley's Hypertensive Disorders in pregnancy. London:

1  
2  
3  
4 712 Elsevier; 2990: 51-72.  
5

6 713  
7

8 714 7. Li X, Dunn J, Salins D, Zhou G, et al. Digital Health: Tracking Physiomes and  
9  
10 715 Activity Using Wearable Biosensors Reveals Useful Health Related Information.  
11  
12 716 PLoS Biol. 2017; 15: e2001402. doi: 10.1371/journal.pbio.2001402.  
13  
14

15 717  
16

17 718 8. Kuriyama S, Yaegashi N, Nagami F, et al. The Tohoku Medical Megabank  
18  
19 719 Project: Design and Mission. J Epidemiol. 2016; 26:493-511. doi:  
20  
21 720 10.2188/jea.JE20150268.  
22  
23

24 721  
25

26 722 9. Japan Society of Obstetrics and Gynecology, Guideline for Obstetrical  
27  
28 723 Practice in Japan, Japan Society of Obstetrics and Gynecology, Tokyo, Japan,  
29  
30 724 pp. 1–4, 2017 [in Japanese].  
31  
32

33 725  
34

35 726 10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the  
36  
37 727 thermoneutral zone: a longitudinal study of normal human pregnancy. Acta  
38  
39 728 Physiol (Oxf). 2011; 201: 467-74. doi: 10.1111/j.1748-1716.2010.02228.x.  
40  
41

42 729  
43

44 730 11. Metoki H, Ohkubo T, Watanabe Y, et al. Seasonal trends of blood pressure  
45  
46 731 during pregnancy in Japan: the babies and their parents' longitudinal  
47  
48 732 observation in Suzuki Memorial Hospital in Intrauterine Period study. J  
49  
50 733 Hypertens. 2008; 26: 2406-13. doi: 10.1097/HJH.0b013e32831364a7.  
51  
52

53 734  
54

55 735 12. Haugen M, Brantsæter AL, Winkvist A, et al. Associations of pre-pregnancy  
56  
57 736 body mass index and gestational weight gain with pregnancy outcome and  
58  
59  
60

- 1  
2  
3  
4 737 postpartum weight retention: a prospective observational cohort study. BMC  
5  
6 738 Pregnancy Childbirth. 2014 Jun 11; 14: 201. doi: 10.1186/1471-2393-14-201.  
7  
8 739  
9  
10 740 13. Sorensen TK, Williams MA, Lee IM, et al. Recreational physical activity  
11  
12 741 during pregnancy and risk of preeclampsia. Hypertension. 2003 Jun; 41:1273-  
13  
14 742 80. doi: 10.1161/01.HYP.0000072270.82815.91  
15  
16 743  
17  
18 744 14. Reutrakul S, Zaidi N, Wroblewski K, et al. Sleep disturbances and their  
19  
20 745 relationship to glucose tolerance in pregnancy. Diabetes Care. 2011; 34: 2454-  
21  
22 746 7. doi: 10.2337/dc11-0780.  
23  
24 747  
25  
26 748 15. Cornish J, Tan E, Teare J, et al. A meta-analysis on the influence of  
27  
28 749 inflammatory bowel disease on pregnancy. Gut. 2007; 56: 830-7. doi:  
29  
30 750 10.1136/gut.2006.108324.  
31  
32 751  
33  
34 752 16. Huxley RR. Nausea and vomiting in early pregnancy: its role in placental  
35  
36 753 development. Obstet Gynecol. 2000; 95:779-82.  
37  
38 754  
39  
40 755 17. Holm Tveit JV, Saastad E, Stray-Pedersen B, et al. Maternal characteristics  
41  
42 756 and pregnancy outcomes in women presenting with decreased fetal movements  
43  
44 757 in late pregnancy. Acta Obstet Gynecol Scand. 2009; 88: 1345-51. doi:  
45  
46 758 10.3109/00016340903348375.  
47  
48 759  
49  
50 760 18. Facchinetti F, Allais G, D'Amico R, et al. The relationship between  
51  
52 761 headache and preeclampsia: a case-control study. Eur J Obstet Gynecol  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 762 Reprod Biol. 2005; 121: 143-8. doi: 10.1016/j.ejogrb.2004.12.020.  
5  
6 763  
7  
8 764 19. Iams JD, Newman RB, Thom EA, et al; National Institute of Child Health  
9  
10 765 and Human Development Network of Maternal-Fetal Medicine Units. Frequency  
11  
12 766 of uterine contractions and the risk of spontaneous preterm delivery. N Engl J  
13  
14 767 Med. 2002; 346: 250-5. doi: 10.1056/NEJMoa002868  
15  
16 768  
17  
18 769 20. Abbas AE, Lester SJ, Connolly H. Pregnancy and the cardiovascular  
19  
20 770 system. Int J Cardiol. 2005; 98: 179-89. doi: 10.1016/j.ijcard.2003.10.028.  
21  
22 771  
23  
24 772 21. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit  
25  
26 773 time. Scand J Gastroenterol. 1997; 32: 920-4.  
27  
28 774 doi: 10.3109/00365529709011203  
29  
30 775  
31  
32 776 22. Riegler G, Esposito I. Bristol scale stool form. A still valid help in medical  
33  
34 777 practice and clinical research. Tech Coloproctol 2001; 5: 163-4. doi:  
35  
36 778 10.1007/s101510100019  
37  
38 779  
39  
40 780 23. Longstreth GF, Thompson WG, Chey WD, et al. Functional Bowel  
41  
42 781 Disorders. Gastroenterology 2006; 130: 1480-91. doi:  
43  
44 782 10.1053/j.gastro.2005.11.061  
45  
46 783  
47  
48 784 24. Koren G, Boskovic R, Hard M, et al. Motherisk-PUQE (pregnancy-unique  
49  
50 785 quantification of emesis and nausea) scoring system for nausea and vomiting of  
51  
52 786 pregnancy. Am J Obstet Gynecol. 2002;186: S228-31.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 787  
5  
6 788 25. Koren G, Piwko C, Ahn E, et al. Validation studies of the Pregnancy Unique-  
7  
8 789 Quantification of Emesis (PUQE) scores. J Obstet Gynaecol. 2005; 25: 241-4.  
9  
10 790 doi: 10.1080/01443610500060651  
11  
12  
13 791  
14  
15 792 26. Pearson JF, Weaver JB. Fetal activity and fetal wellbeing: an evaluation. Br  
16  
17 793 Med J 1976; 1:1305–7.  
18  
19  
20 794  
21  
22 795 27. Winje BA, Saastad E, Gunnes N, et al. Analysis of 'count-to-ten' fetal  
23  
24 796 movement charts: a prospective cohort study. BJOG. 2011;118: 1229-38. doi:  
25  
26 797 10.1111/j.1471-0528.2011.02993.x  
27  
28  
29 798  
30  
31 799 28. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach  
32  
33 800 Notification rules under the Health Information Technology for Economic and  
34  
35 801 Clinical Health Act and the Genetic Information Nondiscrimination Act; other  
36  
37 802 modifications to the HIPAA rules. Fed Regist. 2013; 78: 5565-702.  
38  
39  
40 803  
41  
42 804 29. Health Insurance Portability and Accountability Act of 1996. Public Law 104-  
43  
44 805 191. US Statut Large. 1996; 110:1936-2103.  
45  
46  
47 806  
48  
49 807 30. Amended Act on the Protection of Personal Information.  
50  
51 808 [https://www.ppc.go.jp/files/pdf/Act\\_on\\_the\\_Protection\\_of\\_Personal\\_Information.](https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Information.pdf)  
52  
53 809 pdf.  
54  
55 810  
56  
57  
58 811 31. Takai-Igarashi T, Kinoshita K, Nagasaki M, et al. Security controls in an

- 1  
2  
3  
4 812 integrated Biobank to protect privacy in data sharing: rationale and study  
5  
6 813 design. *BMC Med Inform Decis Mak.* 2017; 17:100. doi: 10.1186/s12911-017-  
7  
8 814 0494-5  
9  
10 815  
11  
12  
13 816 32. Katsuoka F, Yokozawa J, Tsuda K, et al. An efficient quantitation method of  
14  
15 817 next-generation sequencing libraries by using MiSeq sequencer. *Anal Biochem.*  
16  
17 818 2014; 466: 27-9. doi: 10.1016/j.ab.2014.08.015  
18  
19  
20 819  
21  
22 820 33. Yamaguchi-Kabata Y, et al, Nariai N, Kawai Y, et al. iJGVD: an  
23  
24 821 integrative Japanese genome variation database based on whole-genome  
25  
26 822 sequencing. *Hum Genome Var.* 2015; 2:15050. doi: 10.1038/hgv.2015.50.  
27  
28 823 <https://www.ncbi.nlm.nih.gov/pubmed/27081555>.  
29  
30 824  
31  
32  
33 825 34. DeLuca DS, Levin JZ, Sivachenko A, et al. RNA-SeQC: RNA-seq metrics  
34  
35 826 for quality control and process optimization. *Bioinformatics.* 2012; 28: 1530-  
36  
37 827 1532. doi: 10.1093/bioinformatics/bts196.  
38  
39 828  
40  
41  
42 829 35. Koshihara S, Motoike I, Kojima K, et al. The structural origin of metabolic  
43  
44 830 quantitative diversity. *Sci Rep.* 2016; 6: 31463. doi: 10.1038/srep31463  
45  
46 831  
47  
48  
49 832 36. Vu TN, Valkenburg D, Smets K, et al. An integrated workflow for robust  
50  
51 833 alignment and simplified quantitative analysis of NMR spectrometry data. *BMC*  
52  
53 834 *Bioinformatics.* 2011; 12: 405. doi: 10.1186/1471-2105-12-405.  
54  
55 835  
56  
57  
58 836 37. Nishiumi S, Kobayashi T, Ikeda A, et al. A novel serum metabolomics-based



- 1  
2  
3  
4 837 diagnostic approach for colorectal cancer. PLoS One. 2012; 7: e40459. doi:  
5  
6 838 10.1371/journal.pone.0040459.  
7  
8 839  
9  
10 840 38. Nishiumi S, Kobayashi T, Kawana S, et al. Investigations in the possibility of  
11  
12 841 early detection of colorectal cancer by gas chromatography/triple-quadrupole  
13  
14 842 mass spectrometry. Oncotarget. 2017; 8, 17115-17126. doi:  
15  
16 843 10.18632/oncotarget.15081.  
17  
18 844  
19  
20 845 39. Saigusa D, Okamura Y, Motoike IN, et al. Establishment of Protocols for  
21  
22 846 Global Metabolomics by LC-MS for Biomarker Discovery. PLoS One.  
23  
24 847 2016;11(8): e0160555. doi: 10.1371/journal.pone.0160555.  
25  
26 848  
27  
28 849 40. Sato Y, Yamagishi J, Yamashita R, et al. Inter-Individual Differences in the  
29  
30 850 Oral Bacteriome Are Greater than Intra-Day Fluctuations in Individuals. PLoS  
31  
32 851 One. 2015;10: e0131607. doi: 10.1371/journal.pone.0131607.  
33  
34 852  
35  
36 853 41. Brown MA, Magee LA, Kenny LC, et al. Hypertensive Disorders of  
37  
38 854 Pregnancy: ISSHP Classification, Diagnosis, and Management  
39  
40 855 Recommendations for International Practice. Hypertension. 2018; 72: 24–43.  
41  
42 856 doi: 10.1161/HYPERTENSIONAHA.117.10803.  
43  
44 857  
45  
46 858 42. Watanabe K, Naruse K, Tanaka K, et al. Outline of definition and  
47  
48 859 classification of pregnancy induced hypertension (PIH). Hypertens Res  
49  
50 860 Pregnancy 2013; 1: 3–4.  
51  
52 861  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 862 43. IADPSG Consensus Panel: International association of diabetes and  
5  
6 863 pregnancy study groups recommendations on the diagnosis and classification  
7  
8 864 of hyperglycemia in pregnancy. *Diabetes Care* 2010; 33: 676-682. doi:  
9  
10 865 10.2337/dc09-1848.  
11  
12  
13 866  
14  
15 867 44. Box GEP, Jenkins GM, Reinsel GC. *Time series Analysis: Forecasting and*  
16  
17 868 *Control*. 5th ed. New Jersey: Wiley; 2015.  
18  
19  
20 869  
21  
22 870 45. Brandt PT, Williams JT. *Multiple Time Series Models*, Thousand Oaks, CA:  
23  
24 871 Sage Publications, 2007.  
25  
26  
27 872  
28  
29 873 46. Honaker J, King G, Blackwell M. *Amelia II: A Program for Missing Data*,  
30  
31 874 *Journal of Statistical Software*, 45 (7) 2011.  
32  
33 875  
34  
35  
36 876 47. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by  
37  
38 877 surrogate variable analysis. *PLoS Genet*. 2007; 3: 1724-35. doi:  
39  
40 878 10.1371/journal.pgen.0030161  
41  
42  
43 879  
44  
45 880 48. Stegle O, Parts L, Piipari M, et al. Using probabilistic estimation of  
46  
47 881 expression residuals (PEER) to obtain increased power and interpretability of  
48  
49 882 gene expression analyses. *Nat Protoc*. 2012; 7: 500-7. doi:  
50  
51 883 10.1038/nprot.2011.457.  
52  
53  
54 884  
55  
56 885 49. GTEx Consortium, Battle A, Brown CD, et al. Genetic effects on gene  
57  
58 886 expression across human tissues. *Nature*. 2017; 550: 204-213. doi:

- 1  
2  
3  
4 887 10.1038/nature24277.  
5  
6 888  
7  
8 889 50. Polgreen PM, Yang M, Kuntz JL, et al. Using oral vancomycin prescriptions  
9  
10 890 as a proxy measure for Clostridium difficile infections: a spatial and time series  
11  
12 891 analysis. Infect Control Hosp Epidemiol. 2011; 32: 723-6. doi: 10.1086/660858.  
13  
14  
15 892  
16  
17 893 51. McDowell IC, Manandhar D, Vockley CM, et al. Clustering gene expression  
18  
19 894 time series data using an infinite Gaussian process mixture model. PLoS  
20  
21 895 Comput Biol. 2018; 14: e1005896. doi: 10.1371/journal.pcbi.1005896.  
22  
23  
24 896  
25  
26 897 52. Hensman J, Rattray M, Lawrence ND. Fast Nonparametric Clustering of  
27  
28 898 Structured Time-Series. IEEE Trans Pattern Anal Mach Intell. 2015; 37: 383-93.  
29  
30 899 doi: 10.1109/TPAMI.2014.2318711.  
31  
32  
33 900  
34  
35 901 53. Rohart F, Gautier B, Singh A et al. mixOmics: An R package for 'omics  
36  
37 902 feature selection and multiple data integration. PLoS Comput Biol. 2017;13:  
38  
39 903 e1005752. doi: 10.1371/journal.pcbi.1005752.  
40  
41  
42 904  
43  
44 905 54. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis-a  
45  
46 906 framework for unsupervised integration of multi-omics data sets. Mol Syst Biol.  
47  
48 907 2018; 14: e8124. doi: 10.15252/msb.20178124.  
49  
50  
51 908  
52  
53 909 55. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software.  
54  
55 910 Bioinformatics, 2015; 31:1466-8. doi: 10.1093/bioinformatics/btu848.  
56  
57  
58 911  
59  
60

- 1  
2  
3  
4 912 56. Wax JR, Cartin A, Pinette MG. Biophysical and Biochemical Screening for  
5  
6 913 the Risk of Preterm Labor: An Update. *Clin Lab Med.* 2016; 36: 369-83. doi:  
7  
8 914 10.1016/j.cll.2016.01.019.  
9  
10 915  
11  
12 916 57. Al-Rubaie Z, Askie LM, Ray JG, et al. The performance of risk prediction  
13  
14 917 models for pre-eclampsia using routinely collected maternal characteristics and  
15  
16 918 comparison with models that include specialised tests and with clinical guideline  
17  
18 919 decision rules: a systematic review. *BJOG.* 2016; 123:1441-1452. doi:  
19  
20 920 10.1111/1471-0528.  
21  
22 921  
23  
24 922 58. Koullali B, Oudijk MA, Nijman TA, et al. Risk assessment and management  
25  
26 923 to prevent preterm birth. *Semin Fetal Neonatal Med.* 2016 ;21: 80-8. doi:  
27  
28 924 10.1016/j.siny.2016.01.005.  
29  
30 925  
31  
32 926 59. Henderson JT, Thompson JH, Burda BU, et al. Preeclampsia Screening:  
33  
34 927 Evidence Report and Systematic Review for the US Preventive Services Task  
35  
36 928 Force. *JAMA.* 2017; 317: 1668-1683. doi: 10.1001/jama.2016.18315.  
37  
38 929  
39  
40 930 60. Broekhuijsen K, van Baaren GJ, van Pampus MG, et al; HYPITAT-II Study  
41  
42 931 Group. Immediate delivery versus expectant monitoring for hypertensive  
43  
44 932 disorders of pregnancy between 34 and 37 weeks of gestation (HYPITAT-II): an  
45  
46 933 open-label, randomised controlled trial. *Lancet.* 2015; 385: 2492-2501. doi:  
47  
48 934 10.1016/S0140-6736(14)61998-X.  
49  
50 935  
51  
52 936 61. Price ND, Magis AT, Earls JC, et al. A wellness study of 108 individuals  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 937 using personal, dense, dynamic data clouds. Nat Biotechnol. 2017; 35: 747-  
5  
6 938 756. doi: 10.1038/nbt.3870.  
7

8 939

9  
10 940 **FIGURE TITLES AND LEGENDS**

11 941

12  
13  
14  
15 942 **Figure 1. Flowchart of Maternity Log Study (MLOG) participants**

16 943

17  
18  
19 944 **Figure 2. Overview of the MLOG study protocol**

20 945 **A:** Participant timeline for the MLOG study.

21  
22 946 **B:** Physiologic information collected using healthcare devices. Specific  
23  
24 947 measures were uploaded each day from the time of enrollment (solid horizontal  
25  
26 948 lines). Participants had the option to continue uploading data until 180 days  
27  
28 949 after delivery (dashed horizontal lines).

29 950 **C:** Daily lifelogs of self-reported information using a smartphone application.

30  
31 951 Basic lifelog information was input manually from the time of enrollment (solid  
32  
33 952 horizontal lines). Participants had the option to continue uploading data until  
34  
35 953 180 days after delivery (dashed horizontal lines). Fetal movement and uterine  
36  
37 954 contractions were recorded from 24 and 20 weeks of gestation, respectively.  
38  
39 955

40 956

41 957 **Figure 3. Data acquisition rate**

42  
43 958 The mean data upload rate of specific measures was calculated from the total  
44  
45 959 number of days of actual uploads divided by the number of days from  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

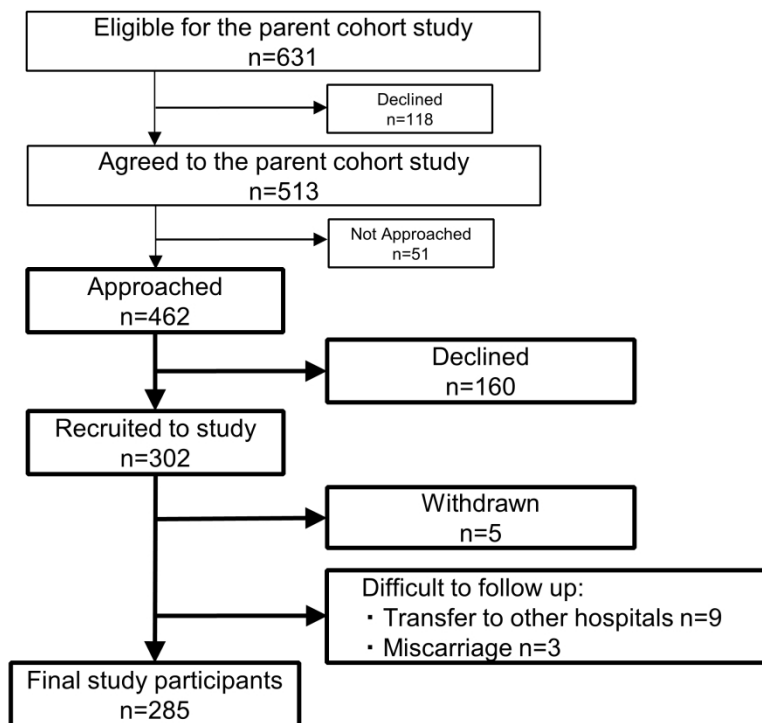


Figure 1.

254x190mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

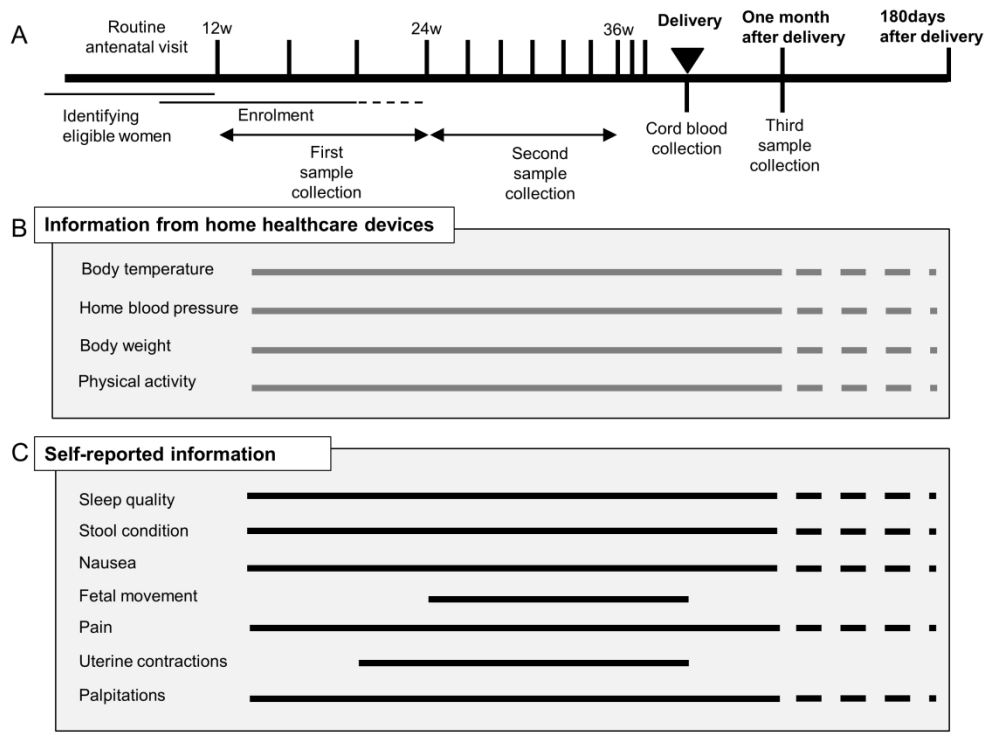


Figure 2.

254x190mm (300 x 300 DPI)

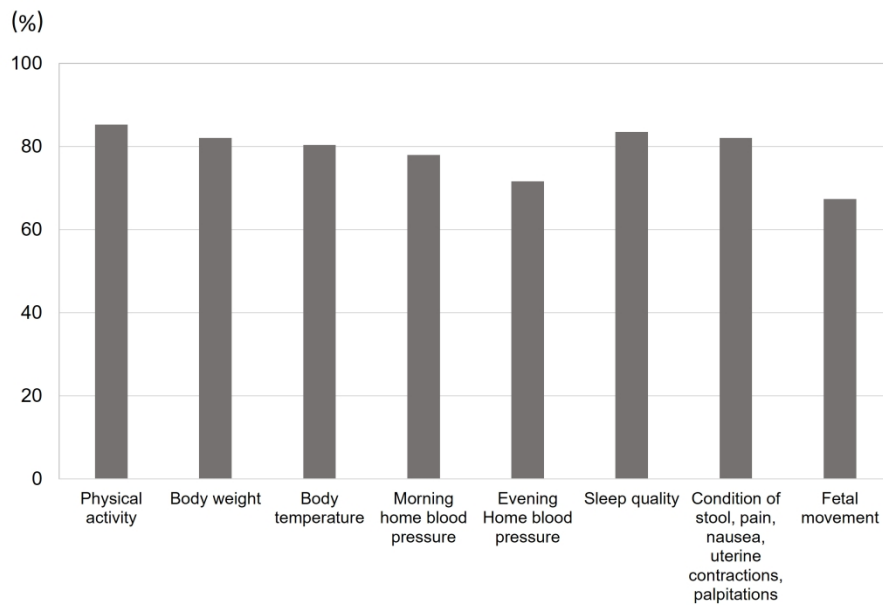


Figure 3.

254x190mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60