RBMO

ELSEVIER

ARTICLE

Whole exome sequencing, a hypothesis-free approach to investigate recurrent early miscarriage

BIOGRAPHY

Lénaick Gourhant has been an engineer in molecular biology and genetics at CHRU in Brest since 2006. She works in the research team EA3878 on pregnancy and venous thrombosis.

Lenaick Gourhant[1], Ozvan Bocher[2], Luc De Saint Martin[1], Thomas E. Ludwig[2], Anne Boland[3], Jean F. Deleuze[3], Philippe Merviel[1,4], Pierre F. Dupré[2,4], Catherine A. Lemarié[1,2], Francis Couturaud[1,5], Cédric Le Maréchal[2,6], Emmanuelle Génin[2], Elisabeth Pasquier[1,*]

KEY MESSAGE

This study confirms the involvement of the vitamin B12 pathway (especially *TCN2*) and reinforces the rationale for seeking vitamin B12 deficiency concomitant with folate deficiency in women with recurrent early miscarriage. Four new genes of interest were also highlighted, two involved in embryonic development (clathrin-mediated endocytosis, ciliary pathway).

ABSTRACT

Research question: Are there genetic determinants shared by unrelated women with unexplained recurrent early miscarriage (REM)?

Design: Thirty REM cases and 30 controls were selected with extreme phenotype among women from Eastern Brittany (France), previously enrolled in an incident case–control study on thrombophilic mutations. Cases and controls were selected based on the number of early miscarriages or live births, respectively. Peripheral blood was collected for DNA extraction at initial visit. The burden of low-frequency variants in the coding part of the genes was compared using whole exome sequencing (WES).

Results: Cases had 3 to 17 early miscarriages (20 cases: ≥5 previous losses). Controls had 1 to 4 live births (20 controls: ≥3 previous live births) and no miscarriages. WES data were available for 29 cases and 30 controls. A total of 209,387 variants were found (mean variant per patient: 59,073.05) with no difference between groups ($P = 0.68$). The top five most significantly associated genes were *ABCA4*, *NFAM1*, *TCN2*, *AL078585.1* and *EPS15*. Previous studies suggest the involvement of vitamin B12 deficiency in REM. *TCN2* encodes for vitamin B12 transporter into cells. Therefore, holotranscobalamin (active vitamin B12) was measured for both cases and controls (81.2 ± 32.1 versus 92.9 ± 34.3 pmol/l, respectively, $P = 0.186$). Five cases but no controls were below 50 pmol/l ($P = 0.052$).

Conclusions: This study highlights four new genes of interest in REM, some of which belong to known networks of genes involved in embryonic development (clathrin-mediated endocytosis and ciliary pathway). The study also confirms the involvement of *TCN2* (vitamin B12 pathway) in the early first trimester of pregnancy.

[1] Université de Brest, EA 3878 (GETBO) Brest 29609, France
[2] Université de Brest, Inserm, EFS, UMR 1078, GGB, CHU de Brest, Brest 29200, France
[3] Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, Evry 91057, France
[4] Division of Gynaecology, University Hospital of Brest Brest 29609, France
[5] Université de Brest, Centre d'Investigation Clinique, INSERM CIC 1412 Brest 29609, France
[6] Université de Brest, Laboratoire de Génétique Moléculaire et d'Histocompatibilité Brest 29609, France

*Corresponding author. E-mail address: elisabeth.pasquier@chu-brest.fr (E. Pasquier). https://doi.org/10.1016/j.rbmo.2021.01.008 1472-6483/
Declaration: The authors report no financial or commercial conflicts of interest.

KEYWORDS

Extreme phenotype
Holotranscobalamin
Rare variants
Recurrent early miscarriage
TCN2 gene
Whole exome sequencing

INTRODUCTION

Recurrent early miscarriage (REM) is currently defined as three or more consecutive losses occurring in the first trimester of pregnancy (*Delabaere et al., 2014*). REM affects 1–2% of couples trying to have children. Standard investigations fail to reveal any well-recognized cause in more than 50% of couples (*Branch et al., 2010*). To date, on the basis of animal models and clinical studies, there are several hypotheses regarding possible underlying mechanisms of unexplained REM: altered ovarian reserve, progesterone deficit, prothrombotic state and/or vascular endothelial dysfunction, immunological impairment, chronic endometritis or vitamin deficiencies (e.g. B9, B12 and D). In a large incident case–control study of couples, some possible determinants of REM were investigated, such as factor V Leiden or mutation G20210A of the prothrombin gene in both parents (*Pasquier et al., 2009*), skewed X-chromosome inactivation in female cases (*Pasquier et al., 2007*) and blood anti-Müllerian hormone concentration (*Leclercq et al., 2019*). Additional studies from other groups suggest that a family history of miscarriage is a risk factor for unexplained REM (*Woolner et al., 2019*; *Zhang et al., 2010*). However, no evidence for strong genetic determinants of unexplained REM was found, particularly based on findings of hypothesis-based candidate gene studies (odds ratios ranging from 0.5 to 2.4), according to a recent systematic review (*Pereza et al., 2017*). This emphasizes the need for hypothesis-free studies. In unexplained REM, two genome-wide association studies have been reported in women (*Kolte et al., 2011*; *Wang et al., 2010*). They highlighted a few areas of interest in the genome, such as the *PDEA* and *GRIK2* genes (*Kolte et al., 2011*), but none of these genomic regions reached the genome-wide significance level. Moreover, these results were not reproduced in an independent study.

In addition, next-generation sequencing tools were used to identify a possible genetic cause in some families with pregnancy losses, especially when a specific phenotype of fetal development defect was identified, such as fetal akinesia (*Rajcan-Separovic et al., 2020*). Only a few studies involved early losses. These were conducted in related subjects and abortion products (*Qiao et al., 2016*) or among unrelated women for a subset of candidate genes (*Quintero-Ronderos et al., 2017*). Specific mutations or affected genes (e.g. *DYNC2H1*) shared between families or between different studies were rare. Nevertheless, some gene networks were often involved, such as cilia function or cell division (*Rajcan-Separovic et al., 2020*).

Thus, further studies in REM are needed. Using next-generation sequencing, this study aimed to seek mechanisms of unexplained REM through a hypothesis-free approach, among unrelated subjects. An extreme-phenotype study can be effective with modest sample sizes to detect low-frequency variants in complex traits, even more effective than studies with larger sample sizes involving subjects with less extreme individual phenotypes (*Emond et al., 2012*). Therefore, female cases and controls were selected at both ends of the phenotype distribution (i.e. with extreme phenotypes) and the burden of low-frequency variants in the coding part of the different genes was compared using whole exome sequencing (WES).

MATERIALS AND METHODS

Selection of cases and controls with extreme phenotypes

Cases were selected among women enrolled in an incident case–control study on thrombophilic mutations in unexplained pregnancy loss (*Pasquier et al., 2009*). Controls of similar age were recruited from the local electoral lists if they had given birth to at least one living child (FIGURE 1). All women were seen in consultation by one of the physician investigators, who provided information, obtained informed consent and conducted a standard questionnaire survey before venous blood sampling.

Then, all women participating in the case–control study on thrombophilic mutations were secondly enrolled in a prospective cohort follow-up study. They were contacted by phone to collect new obstetric and medical events. In addition, all documents related to medical visits and hospitalizations were retrieved.

This ancillary study on genetic determinants involved studying a subgroup of women with unexplained REM, i.e. unexplained iterative losses (≥3) in the first trimester of pregnancy. Exclusion criteria consisted of: a maternal or paternal carrier of a structural chromosomal rearrangement, maternal antiphospholipid antibodies or any anatomical abnormality likely to be responsible for REM. Thirty female cases and 30 female controls with extreme phenotypes were selected, primarily on the basis of the number of early miscarriages and live births, respectively. The main selection criterion for the cases was the number of previous early miscarriages, i.e. mostly before 35 years old, reported at the initial visit. The main selection criterion for the controls was the absence of previous miscarriages both at inclusion and follow-up. Their age at inclusion was within the same age range as the cases. In addition, the controls were selected according to the number of live births before inclusion and during follow-up.

Samples

Biological samples (DNA, plasma and serum) were collected at the initial visit, at least 2 months after any recognized obstetric event or hormonal treatment (FIGURE 1).

In particular, peripheral blood was collected in EDTA. DNA from these 30 cases and 30 controls was extracted from blood lymphocytes by a standard salting-out procedure. The concentrations of DNA were determined using Qubit fluorescence technology (*Miller et al., 1988*).

WES was performed by the Centre National de Recherche en Génomique Humaine (CNRGH, Institut de Biologie François Jacob, CEA). After quality control by the DNA and Cell Bank Laboratory, genomic DNA (3 µg) was captured using in-solution enrichment methodology (SureSelect Human All Exon V5, 50 Mb; Agilent Technologies, Santa Clara, CA, USA). Library preparation and exome enrichment protocol (~20,000 targeted genes) were performed automatically on a Bravo liquid handling robot (Agilent Technologies), according to the manufacturer's instructions. After normalization and quality control, exome-enriched libraries were sequenced on a HiSeq 2000 Sequencing System (Illumina Inc., San Diego, CA, USA) as paired-end 100 bp reads. Samples were sequenced in pools of three samples on each lane of the HiSeq 2000 flow cell, in order to reach at least 90 to 100 of good-quality samples. Image analysis and base calling

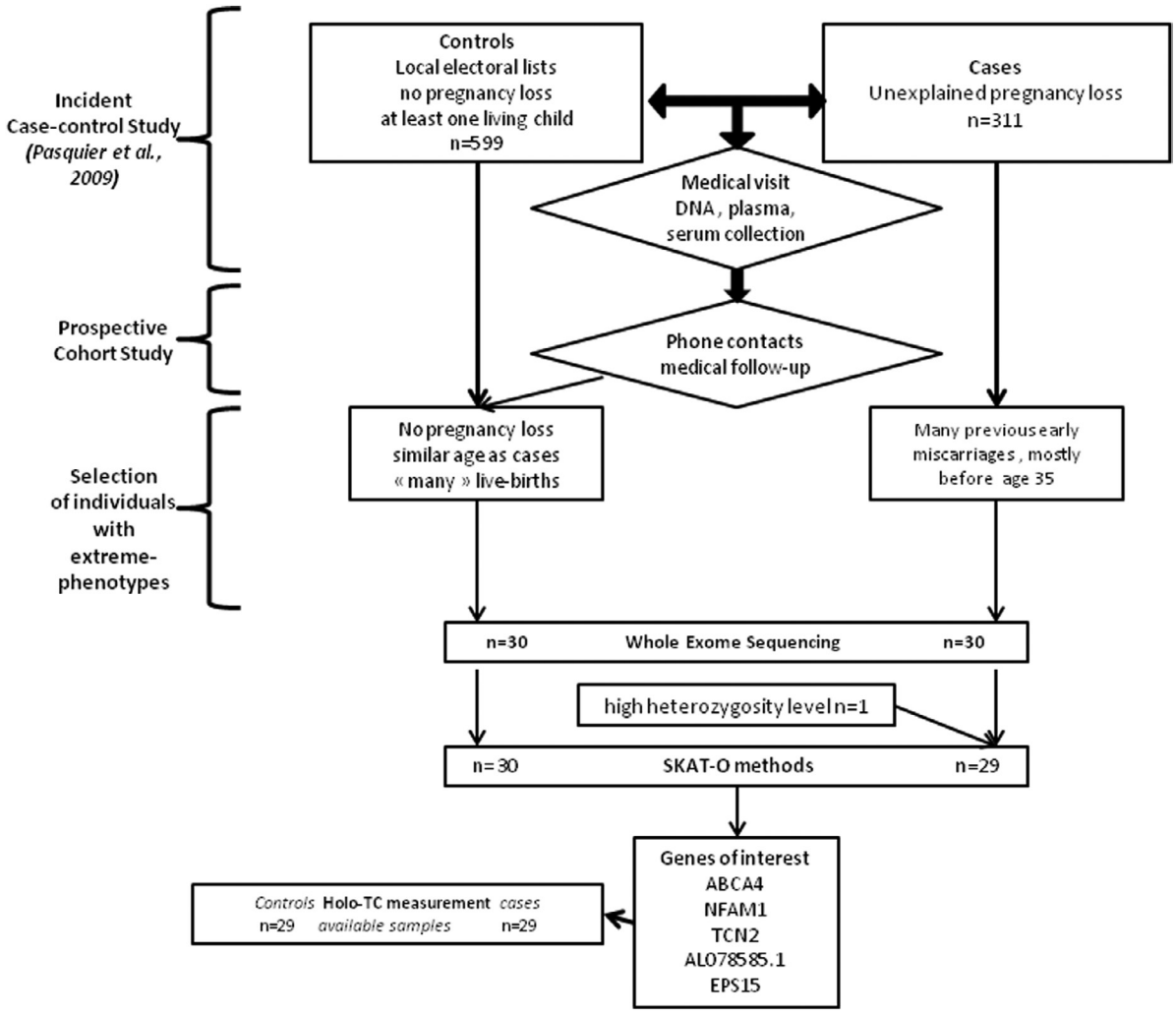

FIGURE 1 Flow diagram.

were performed using the Illumina Real-Time Analysis (RTA) Pipeline. Sequence quality parameters were assessed throughout the sequencing run. Data were processed using a dedicated next-generation sequencing (NGS) pipeline in line with the Broad Institute Genome Analysis Toolkit (GATK) 'Best Practices' (https://software.broadinstitute.org/gatk/best-practices). Briefly, reads were mapped on the reference genome (Build 37) using bwa-mem and the variants were called with Haplotype Caller in GVCF mode. The files were then merged using the GATK's CombineGVCFs, recalibrated and finally annotated with Variant Effect Predictor (VEP).

WES data analyses

Quality control (QC)

A stringent QC was performed first at the genotype and variant level and then at the individual level. A quality score was estimated for each variant with the VQSR function from the GATK and only the variants (SNV or indels) falling in the VQSR tranche PASS or 90.00 to 99.00 were kept. Only the variants with Quality by depth (QD) ≥2, FS (phred-scaled *P*-value using Fisher's exact test to detect strand bias) ≤60 for SNV or ≤200 for indels, SOR (symmetric odds ratio of 2 × 2 contingency table to detect strand bias) ≤3 for SNV or ≤10 for indels, MQ (overall mapping quality of reads supporting a variant call) ≥40 for SNV or ≥10 for indels were kept. When available, the following parameters were also used to filter variants: inbreeding coefficient (InbreedingCoeff) ≥(−0.8), MQRankSum (*Z*-score from Wilcoxon rank sum test of Alt versus Ref read mapping qualities) ≥(−12.5), ReadPosRankSum (*Z*-score from Wilcoxon rank sum test of Alt versus Ref read position bias) was either ≥(−8) for SNV or ≥(−20) for indels. Genotypes with a depth <10 or a genotype quality (gq) <20 were set to missing and comparison of missing rates was done between cases and controls. Only the variants with a genotype call rate greater than 90% and not significantly different between cases and controls (*P*-value of the Fisher's exact test comparing the number of missing genotypes between groups >0.001) were kept. A QC on the allelic balance was also performed to ensure that the mean allelic balance calculated over heterozygous genotypes was within 25–75% in each group (cases and controls). Once the set of eligible variants was determined, QC was performed at the individual level by checking the ancestry within Europe [using principal component analysis (PCA) on a pruned subset of common variants (minor allele frequency, MAF ≥0.05)], heterozygosity levels and kinship between individuals. All these tests were done with plink1.9 (*Chang et al., 2015*).

Rare variant association tests

To test the hypothesis that rare variants located in the same gene could be involved in disease susceptibility, rare variant association tests were performed using the SKAT-O method (*Lee et al., 2012*) that combines burden and variance component tests and was shown to be the most powerful under different disease susceptibility models. Only SNV that passed all QC had an expected effect on the protein (missense, stoploss, stopgain, canonical splice site variants) and with a MAF below 0.05 in either the case or the control group were included in the analysis. Data extraction and formatting were performed in R using our own scripts implemented in the Ravages package (*Bocher et al., 2019*) and association tests were performed using the SKAT R package (https://cran.r-project.org/web/packages/SKAT/) (*Lee et al., 2012*).

Active vitamin B12 measurement

Vitamin B12 (cobalamin) is bound to two proteins, transcobalamin and haptocorrin. The transcobalamin–vitamin B12 complex, known as holotranscobalamin (holo-TC), contained the biologically active cobalamin that is delivered to all cells after interaction with a specific receptor. For both cases and controls, the measurement of holo-TC blood concentration was simultaneously carried out using a chemiluminescent microparticle immunoassay (ARCHITECT Active-B12 from Abbott Laboratories, Abbott Park, IL, USA), whose results are linear from 5.0 to 128 pmol/l (*Merrigan et al., 2015*).

The main characteristics of women and their holo-TC blood concentrations were compared between the two groups using SPSS Statistics for Windows, Version 19.0 (IBM Corp., Armonk, NY, USA). Student's *t*-test and chi-squared test were used for continuous variables (ages, body mass index [BMI], tobacco use, parity ratios by pregnancy (calculated at inclusion in the case–control study and at the end of follow-up), duration of follow-up, holo-TC blood concentrations) and categorical variables (allelic distribution), respectively. For each woman, the parity ratios were calculated as follows: the number of previous live births divided by the number of previous started pregnancies. The odds ratio (OR) was calculated using logistical regression. $P < 0.05$ was considered statistically significant.

The study was approved by the local ethics committee, the CPP of Brest University Hospital (CPP Ouest-336) on 17 December 2002.

RESULTS

Case and control characteristics

Thirty cases and 30 controls with an extreme phenotype were selected from women previously recruited to a master incident case–control study on thrombophilic mutations in Eastern Brittany, France (*Pasquier et al., 2009*):

(i) Cases had a previous history of three to 17 early miscarriages at inclusion in the master incident case–control study (median: 5, interquartile range [IQR]: 4–6.25). Twenty of the cases had five or more losses. Of these 30 cases, 12 (40%) had a previous live birth. The cases often experienced a further early miscarriage (median: 1, min–max: 0–3) during the prospective follow-up of 5.7 ± 3.0 years. When considering the whole reproductive life, their parity ratio by pregnancy was 0.16 ± 0.13 (number of losses: median: 6, IQR: 5–7).

(ii) Controls had one to four previous live births at inclusion in the case–control study (median: 3, IQR: 2–3). Twenty of the controls had three or more live births; they did not suffer any miscarriage even during the prospective follow-up of 6.8 ± 1.6 years.

The main case and control characteristics are summarized in TABLE 1. Age at inclusion, duration of follow-up and BMI were normally distributed and not statistically different between case and control groups ($P = 0.170$, 0.117 and 0.807, respectively). No controls but four cases were lost to follow-up. Cases were significantly younger than controls at the end of follow-up (38.0 ± 5.7 versus 41.1 ± 3.3 years, $P = 0.019$). The number of live births (0 to 2) during follow-up was not significantly different between cases and controls ($P = 0.173$).

Whole exome sequencing

Only one individual (a case) was removed after the QC: this individual had a high heterozygosity level and was found to be an outlier on the PCA. Thus, WES data were available for 29 cases and 30 controls. A total of 209,387 variants were found with an average of 59,073.05 variants per individual and no difference between cases and controls (average numbers of variants were respectively 59,049.52 in cases and 59,095.80 in

TABLE 1 MAIN CHARACTERISTICS OF CASES COMPARED WITH THE CONTROLS

	Cases	Controls	*P*-value
n	**30 (*29*)**	**30**	
Age at inclusion, years (mean ± SD) Median [min–max]	32.9 ± 4.6 *(32.9 ± 4.7)* 33.7 [20.3–41.8]	34.3 ± 3.2 34.7 [28.1–42]	0.170 *(0.162)*
At inclusion, smoking, pack-years (mean ± SD)	3.3 ± 4.6 *(3.3 ± 4.6)*	4.5 ± 7.0	0.467 *(0.467)*
Body mass index (kg/m^2)	23.1 ± 4.1 *(22.8 ± 4.4)*	23.3 ± 3.6	0.807 *(0.603)*
Parity ratio by pregnancy at inclusion, (mean ± SD)	0.09 ± 0.13 *(0.09 ± 0.13)*	1	<0.001
Duration of follow-up, years (mean ± SD)[a]	5.7 ± 3.0 *(5.5 ± 3.0)*	6.8 ± 1.6	0.117 *(0.067)*
Age at the end of follow-up, years (mean ± SD)[a]	38.0 ± 5.7 *(38.0 ± 5.6)*	41.1 ± 3.3	0.019 *(0.022)*
Parity ratio by pregnancy (whole reproductive life), (mean ± SD)[a]	0.16 ± 0.13 *(0.16 ± 0.13)*	1	<0.001

Data reported for all 30 cases or the 29 cases with available whole exome sequencing (italic).

[a] Four missing data among cases (omitted from the analyses).

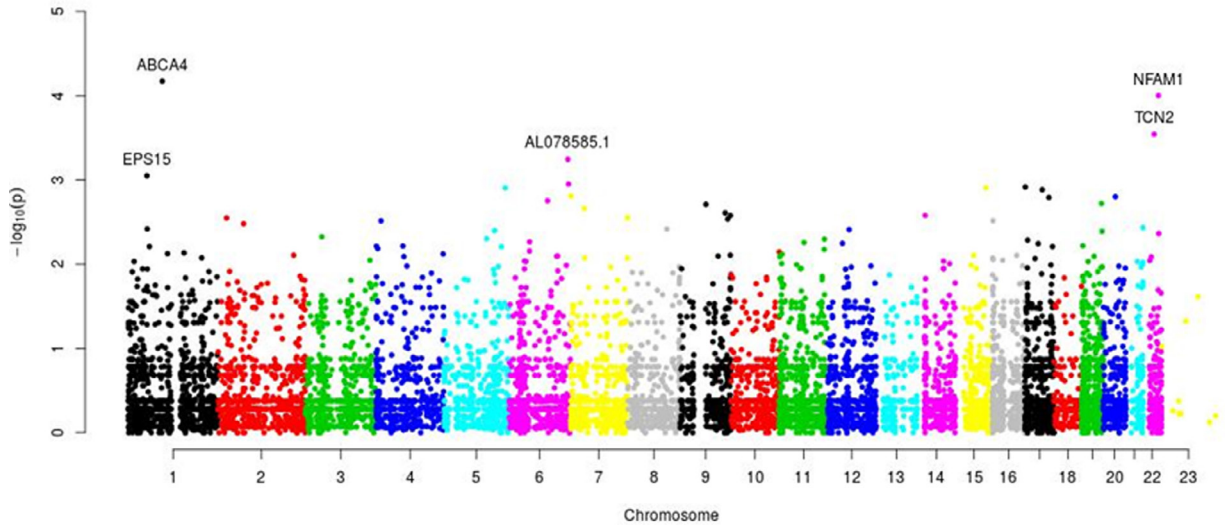

FIGURE 2 Manhattan plot of the rare variant association test SKAT-O. For each gene ordered by chromosome and position on chromosome, the –log10p-value of the SKAT-0 test is given. The top five most associated genes are named.

controls, Student's *t*-test $P = 0.668$). Among these variants, 195,090 were SNV and on average an individual carries 54,112.41 SNV (54,091.93 on average in cases and 54,132.20 in controls, Student's *t*-test $P = 0.683$).

Rare variant association tests

For the rare variant association test, only missense, stoploss, stopgain and canonical splice site SNV with a MAF below 5% in at least one of the two groups of cases or controls were considered. This represented a total of 25,219 variants located in 10,908 genes.

Results of the rare variant association test (SKAT-O) are displayed in FIGURE 2 and the top five list of the most significantly associated genes is given in TABLE 2. None of the genes reached genome-wide significance (usually defined as a *P*-value below 2.5×10^{-6} for rare variant association tests). The five most significant genes were *ABCA4*, *NFAM1*, *TCN2*, *AL078585.1* and *EPS15*. Interestingly, *TCN2* encodes for a protein involved in vitamin B12 transport into cells. As shown in TABLE 3, six variants were retained in this gene for the rare variant association test, including rs9606756 [c.67A>G,p.Ile23Val] that has already been associated with recurrent spontaneous abortion in Korean patients (*Kim et al., 2014*). This variant was also more frequent in cases than in controls ($P = 0.03$; OR 6; 95% CI 1.17–30.72). Furthermore, this variant has an increased frequency in cases and a decreased frequency in controls compared with the GnomAD reference population (*Karczewski et al., 2020*) (0.190 and 0.033, respectively, compared with 0.120 in GnomAD European populations).

Holo-TC measurements

The holo-TC blood concentrations were available for 29 cases and 29 controls (FIGURE 3). They were normally distributed from 30.1 to 188.0 pmol/l and slightly, but not significantly, lower in the cases (81.2 ± 32.1) than in the controls (92.9 ± 34.3 pmol/l, $P = 0.186$).

TABLE 2 FIVE MOST ASSOCIATED GENES IN THE SKAT-0 RARE VARIANT ASSOCIATION TEST

Gene	*P*-value of SKAT-0 test	Number of variants	cMAF	
			Cases	Controls
ABCA4	6.74e-05	7	0.259	0.017
NFAM1	9.97e-05	2	0.483	0.033
TCN2	2.86–04	6	0.534	0.083
AL078585.1	5.69e-04	1	0.224	0.017
EPS15	8.928e-04	3	0.190	0.017
Gene	*P*-value of WSS	OR	CI 95% of OR	
ABCA4	1.46e-05	135.64	[4.45–4136.10]	
NFAM1	5.88e-05	56.36	[3.54–896.05]	
TCN2	1.33e-04	18.90	[2.59–137.91]	
AL078585.1	3.7e-04	970.58	[4.60–204861.9]	
EPS15	7.66e-04	106.23	[2.12–5320.35]	

The *P*-value of the test is given as well as the number of variants and their cumulative minor allele frequency (cMAF) in cases and controls. *P*-values, OR and their 95% CI are also given using the burden test WSS.

TABLE 3 LIST OF VARIANTS FROM THE TOP FIVE GENES INCLUDED IN THE RARE VARIANT ASSOCIATION TEST

Gene	CHR	POS	RS	REF	ALT	WORST_ANNOT	FreqA-LT_Cases	FreqALT_Controls	FreqALT_GnomAD	FreqALT_GnomAD_NFE
ABCA4	1	94463617	rs1800555	C	T	missense_variant	0.017	0.000	0.011	0.012
		94476467	rs1801466	T	A	missense_variant	0.069	0.017	0.042	0.066
		94496610	rs62642573	C	T	missense_variant	0.017	0.000	0.000	0.000
		94512565	rs1801581	C	T	missense_variant	0.086	0.000	0.030	0.041
		94514466	rs61754030	T	C	missense_variant	0.017	0.000	0.002	0.003
		94568675	rs62646863	T	C	missense_variant	0.017	0.000	0.001	0.002
		94568686	rs62646862	C	T	missense_variant	0.034	0.000	0.003	0.004
NFAM1	22	42783070	rs34963472	G	A	stop_gained	0.241	0.017	0.136	0.153
		42793873	rs5996153	A	G	missense_variant	0.241	0.017	0.153	0.160
TCN2	22	31006860	rs9606756	A	G	missense_variant	0.190	0.033	0.120	0.117
		31008867	rs35915865	T	C	missense_variant	0.017	0.000	0.013	0.020
		31013419	rs9621049	C	T	missense_variant	0.172	0.033	0.115	0.116
		31018975	rs1131603	T	C	missense_variant	0.121	0.017	0.042	0.055
		31018998	rs144166182	G	A	missense_variant	0.017	0.000	0.000	0.000
		31019044	rs4820889	G	A	missense_variant	0.017	0.000	0.034	0.030
AL078585.1	6	1.64E+08	rs1124951	C	G	missense_variant	0.776	0.983	0.838	0.838
EPS15	1	51868160	rs34957183	C	G	missense_variant	0.034	0.000	0.009	0.015
		51873967	rs41292521	G	A	missense_variant	0.052	0.000	0.013	0.019
		51888878	rs72696106	A	G	Splice_donor_variant	0.103	0.017	0.019	0.031

Genes are ranked from the most significant to the less significant. For each variant, its position, rs name, reference and alternative alleles and the frequencies of the alternative allele in the cases, controls and in GnomAD total database and GnomAD Non-Finnish European (NFE) population are reported.

The quartiles of the whole distribution (cases and controls together) were as follows: 60.3, 83.45 and 107.8 pmol/l, respectively (data not shown). The quartiles distribution of holo-TC blood concentrations was not statistically different between cases and controls ($P = 0.553$). However, as the normal distribution of holo-TC blood concentrations was left-truncated at 50 pmol/l in the controls, a threshold of 50 pmol/l was considered: five cases but no controls were below 50 pmol/l ($P = 0.052$).

In the cases, lower holo-TC blood concentration did not predict a further miscarriage. Additionally, the parity ratios by pregnancy (at inclusion and at the end of follow-up) were not significantly correlated with the holo-TC blood concentrations using regression analyses ($P = 0.214$ and $P = 0.414$).

Correlations between holo-TC measurements and TCN2 mutations

Holo-TC concentrations from cases were negatively and significantly correlated with rs9606756 polymorphism (cases: $P = 0.03$). Holo-TC concentrations were also correlated with rs9606756 polymorphism when cases and controls were analysed together (all: $P = 0.007$) (FIGURE 4).

The other mutations were not significantly correlated with holo-TC concentrations (data not shown).

DISCUSSION

Using a WES strategy free of *a priori* hypotheses, this study compared two groups of fertile women with extreme phenotypes. Exome data were analysed for 29 of the 30 women included in the REM group. Those women had a REM history of at least three losses with a pregnancy failure rate of 84%. The control group included 30 women who did not suffer any miscarriage and had two to four previous live births at the end of follow-up. Only damaging variants with a MAF of less than 5% either in cases or controls were analysed. This highlighted five potential candidate genes for REM: *ABCA4*, *NFAM1*, *TCN2*, *AL078585.1* and *EPS15,* although none reaching genome-wide significance. *TCN2* was the only gene from a gene network that has already been studied in this clinical setting. Notably, *TCN2* rs9606756 polymorphism was significantly associated with REM in this study ($P = 0.03$; OR 6.0; 95% CI 1.17–30.72). *TCN2* codes for a protein (TC) that binds vitamin B12 and is responsible for up to 99% of vitamin B12 transport into tissues. Therefore, the blood concentrations of holo-TC (transcobalamin–vitamin B12 complex) were measured in both cases and controls. At inclusion, the cases displayed lower (but not significantly) holo-TC blood concentrations than the controls (81.2 ± 32.1 vs 92.9 ± 34.3 pmol/l, $P = 0.186$). Notably, *TCN2* rs9606756 polymorphism was highly and negatively correlated to the blood concentrations of holo-TC ($P = 0.007$), especially in the cases ($P = 0.03$), as previously reported in the overall population (*Riedel et al., 2011*).

Folic acid and the whole vitamin B12/folic acid pathway (*Quadros et al., 2013*) are known to be essential during early pregnancy and were consequently suspected to participate in REM pathogenesis. Serum folate was first investigated in women with REM in 1965 (*Martin et al., 1965*). Years later, hyperhomocysteinemia was reported to

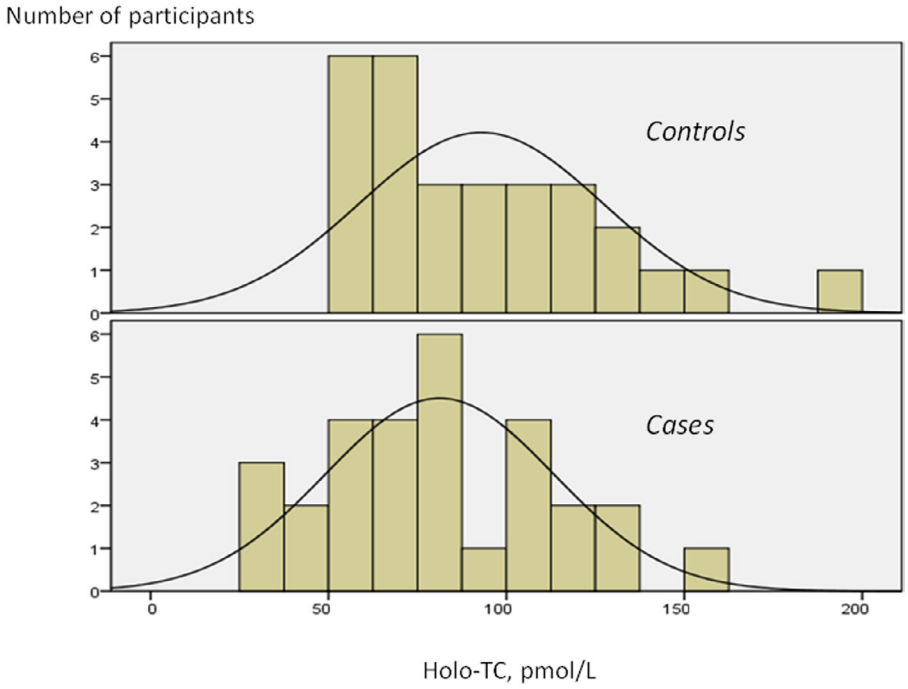

FIGURE 3 Distribution of the holo-TC blood concentrations were compared between the 29 cases and 29 controls (P = 0.186). The curves show the normal distribution.

increase the risk of recurrent abortion (*Nelen et al., 2000*). Vitamin B12, at serum B12 concentrations under the threshold of 180 pg/l, was highly associated with REM compared with concentrations found in healthy premenopausal women (*Reznikoff-Etievant et al., 2002*). Additionally, vitamin B12 was found at serum concentrations significantly lower in REM women compared with healthy pregnant women (*Hübner et al., 2008*). These findings were in contrast with the results of a case–control study (*Sütterlin et al., 1997*). Interestingly, functional vitamin B12 assessed by the measurements of methyl malonic acid concentration in the blood was not statistically different between cases and controls (*Hübner et al., 2008*).

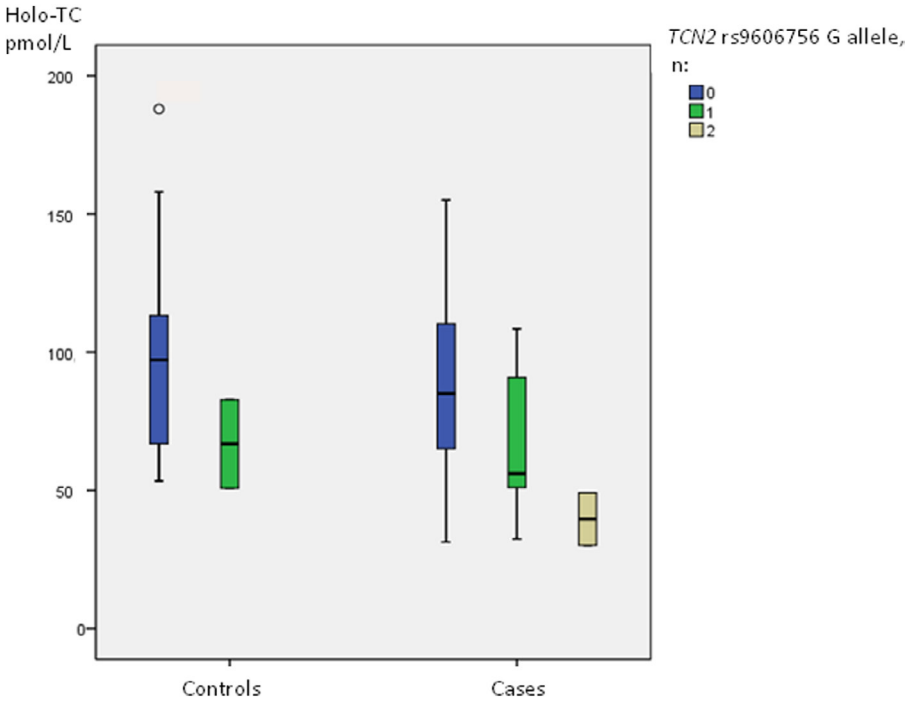

FIGURE 4 Distribution of the holo-TC blood concentrations among the 29 cases and 29 controls according to rs9606756 allelic distribution (box plot: median, IQR, highest and lowest values which are no greater than 1.5 times the IQR). Holo-TC concentrations were significantly correlated with rs9606756 polymorphism (P = 0.007). Note that a control had a higher holo-TC value than 1.5 times the IQR and no controls had two G alleles.

Here, another method was proposed for assessment of functional vitamin B12 through the measurements of holo-TC. It is thought that this method has never been reported in women with REM. Lastly, two case–control studies investigated known *TCN2* polymorphisms in REM women compared with controls (*Hashemi et al., 2018*; *Kim et al., 2014*). *Kim et al. (2014)* reported a possible relationship between REM and *TCN2* rs9606756 among Korean women ($P = 0.019$; OR: 2.918; 95% CI 1.193–7.137) whereas *Hashemi et al. (2018)* did not find any association between REM and *TCN2* rs180119 (a common polymorphism) or *TCN2* rs2336573 among Iranian women. More recently, *Park et al. (2019)* reported an increased risk for recurrent implantation failure after IVF for carriers of the *TCN2* rs9606756 G allele ($P = 0.025$; OR: 4.732; 95% CI 1.220–18.356).

In addition, this study is the first to report the possible implication of *ABCA4*, *NFAM-AL078585.1* and *EPS15* in REM. *EPS15*, located at chromosome 1, is especially attractive. EPS15 is implicated in internalization of the epithelial growth factor receptor (EGFR) (*van Bergen en Henegouwen, 2009*) and in clathrin-mediated endocytosis, which appears to be important in embryonic development (*Milesi et al., 2019*; *Offenhaüser et al., 2000*). EPS15 is organized into four domains containing an evolutionary conserved motif: EPS15 homology (EH domains) that is present in three copies at the NH2-terminus of EPS15. This motif is shared by a wide range of endocytic proteins such as EHD1, which plays a role in the endocytosis of insulin growth factor (IGF)-1 receptors. Notably, EHD1 might be involved in endometrium decidualization and be a factor in recurrent implantation failure (*Zhou et al., 2019*).

ABCA4 is a member of the ATP-binding cassette (ABC) family of transport proteins and codes a retina-specific ABC transporter whose substrate is N-retinylidene-PE expressed in retina photoreceptor cells. *ABCA4* is implicated in Stargardt disease (form of macular degeneration) and the related retinopathies (*Shastry et al., 2008*). More interestingly, a genome-wide association study identified *ABCA4* as one of two new genes associated with cleft lip and palate (*Beaty et al., 2010*). This was confirmed in other studies despite the absence of biological evidence supporting the role of ABCA4 in craniofacial morphogenesis (*Yuan et al., 2011*). Some mechanisms were suggested as strong linkage disequilibrium with true causal variants in other genes (*Beaty et al., 2010*) or presence of craniofacial enhancers within *ABCA4* intronic regions affecting the expression of other genes such as ArhGAP29 (*Morris et al.,2020*). Interestingly, in connection with REM, the expression of *ArhGAP29* contributes to the regulation of the development of intrauterine adhesions caused by lesions of the endometrium (*Xu et al., 2017*). Additionally, an ABCA4–folic acid consumption interaction was identified in cleft lip and palate (*Velázquez-Aragón et al., 2016*). Similar mechanisms should be sought in REM.

AL078585.1 is a subcategory of the PACRG Antisense RNA 3 (*PACRG-AS3*) gene. This is an RNA gene, affiliated with the longnon-coding RNA class. The protein is uncharacterized; no data are available regarding *AL078585.1* function. However, *PACRG* was originally identified as a gene related to Parkinson's disease in humans and has ciliary functions that are pivotal during early vertebrate embryogenesis (*Thumberger et al., 2012*). Mutations in DYNC2H1 (another gene involved in cilia biogenesis) were identified in miscarriages from two families with unexplained REM (*Qiao et al., 2016*).

Finally, *NFAM* codes for a protein that contains an immunoreceptor tyrosine-based activation and is thought to regulate the signalling and development of B-cells (*Ohtsuka et al., 2004*).

In this study, it is noteworthy that the MAF observed in both groups for these five top candidate genes are quite different from those recorded in reference databases such as the Genome Aggregation Database (gnomAD) (*Karczewski et al., 2020*) or the French Exome Project Database (FREX) (*Génin et al., 2017*). This is probably explained by the extreme phenotype of cases and controls.

Apart from the free of *a priori* hypotheses approach, one of the strong points of this study lies in the selection of controls from the electoral lists, in the same geographic area (Eastern Brittany, a geographic area with low ethnic mixing) and in the same period as the cases.

Due to the limited sample size of the current study, it was not possible to find genome-wide significant associations, i.e. associations that remain significant after correction for multiple testing at the scale of all the genes tested. Regarding common variants, a 'suggestive association' is defined by $P < 10^{-5}$. Regarding rare variants, fewer tests are done, because the variants of each gene are grouped together. Consequently, some authors used a higher threshold for *P*-values, such as $<10^{-3}$ (*Grant et al., 2018*). Nevertheless, the three genes ABCA4, NFAM1 and TCN2 showed interesting signals with *P*-values $<5 \times 10^{-4}$. Using asymptotical calculations, accounting for the estimated effect sizes and observed cumulative MAF of the genes, the probability of detecting genome-wide significant associations ranged from 78.5% to 93.1% (78.5% for TCN2, 87.8% for NFAM1 and 93.1% for ABCA4). This confers a good study power and the small sample size *per se* does not reduce the clinical evidence of the results.

It is acknowledged that the few other WES studies in REM most often involved the mother–father–embryo trio and did not often reveal the same modified genes between the different families. The design of this case–control study focused on shared altered genes among women with REM. As a limitation, this study did not involve the two members of the couples enrolled in the PRINCEPS study on thrombophilic mutations (*Pasquier et al., 2009*). Moreover, miscarriage products were not available. Other limitations of the current study warrant consideration. The functional impact of the gene polymorphism of the other four genes that came out of the WES study were not studied. Regarding *TCN2*, a significant association was found between rs9606756 polymorphism and (i) REM ($P = 0.03$) as well as (ii) holo-TC concentrations ($P = 0.007$). However, no significant difference in holo-TC concentrations between cases and controls was found, which may be due to the routine practice of vitamin supplementation in REM women. This could suggest a more complex functional impact of *TCN2* rs9606756 favouring miscarriage than B12 deficiency. Moreover, holo-TC measurement is not a perfect marker for B12 status or B12 cell intake (*Harrington, 2017*).

In conclusion, despite some limitations, this study confirms the involvement of the vitamin B12 pathway and especially

TCN2 in the early first trimester of pregnancy. This reinforces the rationale for seeking vitamin B12 deficiency concomitant with folate deficiency in women with REM. Furthermore, the study highlights new genes of interest in REM, some of which belong to known networks of genes involved in embryonic development (clathrin-mediated endocytosis and ciliary pathway). Others might be involved in endometrium homeostasis. Regarding these new genes, further studies are required to replicate these results and elucidate the underlying mechanisms leading to REM.

ACKNOWLEDGEMENTS

The authors thank the GYNOBS 29 network and Zarrin Alavi (CIC INSERM 1412).

This work was carried out within the framework of a collaboration between the Centre National de Recherche en Génomique Humaine and Brest University Hospital.

This study was conducted with financial support from the French Health Office (PHRCi 2012-204).

REFERENCES

Beaty, T.H., Murray, J.C., Marazita, M.L., Munger, R.G., Ruczinski, I., Hetmanski, J.B., Liang, K.Y., Wu, T., Murray, T., Fallin, M.D., Redett, R.A., Raymond, G., Schwender, H., Jin, S.C., Cooper, M.E., Dunnwald, M., Mansilla, M.A., Leslie, E., Bullard, S., Lidral, A.C., Moreno, L.M., Menezes, R., Vieira, A.R., Petrin, A., Wilcox, A.J., Lie, R.T., Jabs, E.W., Wu-Chou, Y.H., Chen, P.K., Wang, H., Ye, X., Huang, S., Yeow, V., Chong, S.S., Jee, S.H., Shi, B., Christensen, K., Melbye, M., Doheny, K.F., Pugh, E.W., Ling, H., Castilla, E.E., Czeizel, A.E., Ma, L., Field, L.L., Brody, L., Pangilinan, F., Mills, J.L., Molloy, A.M., Kirke, P.N., Scott, J.M., Arcos-Burgos, M., Scott, A.F. **A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4.** Nat. Genet. 2010; 42: 525–529

Bocher, O., Marenne, G., Saint Pierre, A., Ludwig, T.E., Guey, S., Tournier-Lasserve, E., Perdry, H., Génin, E. **Rare variant association testing for multicategory phenotype.** Genet. Epidemiol. 2019; 43: 646–656

Branch, D.W., Gibson, M., Silver, R.M. **Clinical practice. Recurrent miscarriage.** N. Engl. J. Med. 2010; 363: 1740–1747

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J. **Second-generation PLINK: rising to the challenge of larger and richer datasets.** Gigascience 2015; 25: 4–7

Delabaere, A., Huchon, C., Lavoue, V., Lejeune, V., Iraola, E., Nedellec, S., Gallot, V., Capmas, P., Beucher, G., Subtil, D., Carcopino, X., Vialard, F., Nizard, J., Quibel, T., Costedoat-Chalumeau, N., Legendre, G., Venditelli, F., Rozenberg, P., Lemery, D., Deffieux, X. **[Definition of pregnancy losses: Standardization of terminology from the French National College of Obstetricians and Gynecologists (CNGOF)].** J. Gynecol. Obstet. Biol. Reprod. (Paris) 2014; 43: 756–763

Emond, M.J., Louie, T., Emerson, J., Zhao, W., Mathias, R.A., Knowles, M.R., Wright, F.A., Rieder, M.J., Tabor, H.K., Nickerson, D.A., Barnes, K.C. **National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project.**, Lung, G.O., Gibson, R.L., Bamshad, M.J. **Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis.** Nat. Genet. 2012; 44: 886–889

Genin, E., Redon, R., Deleuze, J.F., Campion, D., Lambert, J.C. **Dartigues JF for the FREX Consortium (2017) The French Exome (FREX) Project: A Population-based Panel of Exomes to Help Filter Out Common Local Variants.** Genet. Epidemiol. 2017; 41: 691

Grant, R.C., Denroche, R.E., Borgida, A., Virtanen, C., Cook, N., Smith, A.L., Connor, A.A., Wilson, J.M., Peterson, G., Roberts, N.J., Klein, A.P., Grimmond, S.M., Biankin, A., Cleary, S., Moore, M., Lemire, M., Zogopoulos, G., Stein, L., Gallinger, S. **Exome-Wide Association Study of Pancreatic Cancer Risk.** Gastroenterology 2018; 154: 719–722

Harrington, D.J. **Laboratory assessment of vitamin B12 status.** J. Clin. Pathol. 2017; 70: 168–173

Hashemi, M., Mokhtari, M., Yazdani-Shahrbabaki, V., Danesh, H., Bizhani, F., Taheri, M. **Evaluation of transcobalamin II rs1801198 and transcobalamin II receptor rs2336573 gene polymorphisms in recurrent spontaneous abortion.** J. Obstet. Gynaecol. 2018; 38: 860–863

Hübner, U., Alwan, A., Jouma, M., Tabbaa, M., Schorr, H., Herrmann, W. **Low serum vitamin B12 is associated with recurrent pregnancy loss in Syrian women.** Clin. Chem. Lab. Med. 2008; 46: 1265–1269

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E. **Genome Aggregation Database Consortium, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans.** Nature 2020; 581: 434–443

Kim, H.S., Lee, B.E., Jeon, Y.J., Rah, H., Lee, W.S., Shin, J.E., Choi, D.H., Kim, N.K. **Transcobalamin II (TCN2 67A>G and TCN2 776C>G) and transcobalamin II receptor (TCblR 1104C>T) polymorphisms in Korean patients with idiopathic recurrent spontaneous abortion.** Am. J. Reprod. Immunol. 2014; 72: 337–346

Kolte, A.M., Nielsen, H.S., Moltke, I., Degn, B., Pedersen, B., Sunde, L., Nielsen, F.C., Christiansen, O.B. **A genome-wide scan in affected sibling pairs with idiopathic recurrent miscarriage suggests genetic linkage.** Mol. Hum. Reprod. 2011; 17: 379–385

Leclercq, E., de Saint Martin, L., Bohec, C., Le Martelot, M.T., Roche, S., Alavi, Z., Mottier, D., Pasquier, E. **Blood anti-Müllerian hormone is a possible determinant of recurrent early miscarriage, yet not conclusive in predicting a further miscarriage.** Reprod. Biomed. Online 2019; 39: 304–311

Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A. **NHLBI GO Exome Sequencing Project—ESP Lung Project Team, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies.** Am. J. Hum. Genet. 2012; 91: 224–237

Martin, R.H., Harper, T.A., Kelso, W. **Serum-folic acid in recurrent abortions.** Lancet 1965; 1: 670–672

Merrigan, S.D., Owen, W.E., Straseski, J.A. **Performance characteristics of the ARCHITECT Active-B12 (Holotranscobalamin) assay.** Clin. Lab. 2015; 61: 283–288

Milesi, C., Alberici, P., Pozzi, B., Oldani, A., Beznoussenko, G.V., Raimondi, A., Soppo, B.E., Amodio, S., Caldieri, G., Malabarba, M.G., Bertalot, G., Confalonieri, S., Parazzoli, D., Mironov, A.A., Tacchetti, C., Di Fiore, P.P., Sigismund, S., Offenhäuser, N. **Redundant and nonredundant organismal functions of EPS15 and EPS15L1.** Life Sci. Alliance 2019; 2e201800273

Miller, S.A., Dykes, D.D., Polesky, H.F. **A simple salting out procedure for extracting DNA from human nucleated cells.** Nucleic Acids Res. 1988; 16: 1215

Morris, V.E., Hashmi, S.S., Zhu, L., Maili, L., Urbina, C., Blackwell, S., Greives, M.R., Buchanan, E.P., Mulliken, J.B., Blanton, S.H., Zheng, W.J., Hecht, J.T., Letra, A. **Evidence for craniofacial enhancer variation underlying nonsyndromic cleft lip and palate.** Hum. Genet. 2020. doi:10.1007/s00439–020–02169–9

Nelen, W.L., Blom, H.J., Steegers, E.A., Den Heijer, M., Thomas, C.M., Eskes, T.K. **Homocysteine and folate levels as risk factors for recurrent early pregnancy loss.** Obstet. Gynecol. 2000; 95: 519–524

Offenhäuser, N., Santolini, E., Simeone, A., Di Fiore, P.P. **Differential patterns of expression of Eps15 and Eps15R during mouse embryogenesis.** Mech. Dev. 2000; 95: 309–312

Ohtsuka, M., Arase, H., Takeuchi, A., Yamasaki, S., Shiina, R., Suenaga, T., Sakurai, D., Yokosuka, T., Arase, N., Iwashima, M., Kitamura, T., Moriya, H., Saito, T. **NFAM1, an immunoreceptor tyrosine-based activation motif-bearing molecule that regulates B cell development and signaling.** Proc. Natl. Acad. Sci. U. S. A. 2004; 101: 8126–8131

Park, H.S., Kim, J.O., An, H.J., Ryu, C.S., Ko, E.J., Kim, Y.R., Ahn, E.H., Lee, W.S., Kim, J.H., Kim, N.K. **Genetic polymorphisms of the cobalamin transport system are associated with idiopathic recurrent implantation failure.** J. Assist. Reprod. Genet. 2019; 36: 1513–1522

Pasquier, E., Bohec, C., De Saint Martin, L., Le Maréchal, C., Le Martelot, M.T., Roche, S., Laurent, Y., Férec, C., Collet, M., Mottier, D. **Strong evidence that skewed X-chromosome inactivation is not associated with recurrent pregnancy loss: an incident paired case control study.** Hum. Reprod. 2007; 22: 2829–2833

Pasquier, E., Bohec, C., Mottier, D., Jaffuel, S., Mercier, B., Férec, C., Collet, M., de Saint Martin, L. **Inherited thrombophilias and unexplained pregnancy loss: an incident case–control study.** J. Thromb. Haemost. 2009; 7: 306–311

Pereza, N., Ostojić, S., Kapović, M., Peterlin, B. **Systematic review and meta-analysis of genetic association studies in idiopathic recurrent spontaneous abortion.** Fertil. Steril. 2017; 107: 150–159

Qiao, Y., Wen, J., Tang, F., Martell, S., Shomer, N., Leung, P.C., Stephenson, M.D., Rajcan-Separovic, E. **Whole exome sequencing in recurrent early pregnancy loss.** Mol. Hum. Reprod. 2016; 22: 364–372

Quadros, E.V., Sequeira, J.M. **Cellular uptake of cobalamin: transcobalamin and the TCblR/CD320 receptor.** Biochimie 2013; 95: 1008–1018

Quintero-Ronderos, P., Mercier, E., Fukuda, M., Gonzalez, R., Suarez, C.F., Patarroyo, M.A., Vaiman, D., Gris, J.C., Laissue, P. **Novel genes and mutations in patients affected by recurrent pregnancy loss.** PLoS One 2017; 12e0186149

Rajcan-Separovic, E. **Next generation sequencing in recurrent pregnancy loss-approaches and outcomes.** Eur. J. Med. Genet. 2020; 63103644

Reznikoff-Etievant, M.F., Zittoun, J., Vaylet, C., Pernet, P., Milliez, J. **Low vitamin B12 level as a risk factor for very early recurrent abortion.** Eur. J. Obstet. Gynecol. Reprod. Biol. 2002; 104: 156–159

Shastry, B.S. **Evaluation of the common variants of the ABCA4 gene in families with Stargardt disease and autosomal recessive retinitis pigmentosa.** Int. J. Mol. Med. 2008; 21: 715–720

Riedel, B.M., Molloy, A.M., Meyer, K., Fredriksen, A., Ulvik, A., Schneede, J., Nexø, E., Hoff, G., Ueland, P.M. **Transcobalamin polymorphism 67A->G, but not 776C->G, affects serum holotranscobalamin in a cohort of healthy middle-aged men and women.** J. Nutr. 2011; 141: 1784–1790

Sütterlin, M., Bussen, S., Ruppert, D., Steck, T. **Serum levels of folate and cobalamin in women with recurrent spontaneous abortion.** Hum. Reprod. 1997; 12: 2292–2296

Thumberger, T., Hagenlocher, C., Tisler, M., Beyer, T., Tietze, N., Schweickert, A., Feistel, K., Blum, M. **Ciliary and non-ciliary expression and function of PACRG during vertebrate development.** Cilia. 2012; 1: 13

van Bergen En Henegouwen, P.M. **Eps15: a multifunctional adaptor protein regulating intracellular trafficking.** Cell. Commun. Signal 2009; 8: 24

Velázquez-Aragón, J.A., Alcántara-Ortigoza, M.A., Estandia-Ortega, B., Reyna-Fabián, M.E., Méndez-Adame, C.D. **González-Del Angel A. Gene Interactions Provide Evidence for Signaling Pathways Involved in Cleft Lip/Palate in Humans.** J. Dent. Res. 2016; 95: 1257–1264

Wang, Li, Zeng, Chan Wang, Cui, Xie, Xiao, Feng Liu, Mao, Sheng Yang **Genome-wide screening for risk loci of idiopathic recurrent miscarriage in a Han Chinese population: a pilot study.** Reprod. Sci. 2010; 17: 578–584

Woolner, A.M.F., Raja, E.-A., Bhattacharya, S., Danielian, P., Bhattacharya, S. **Inherited susceptibility to miscarriage: a nested case-control study of 31565 women from an intergenerational cohort.** Am. J. Obstet. Gynecol. 2019

Xu, Q., Duan, H., Gan, L., Liu, X., Chen, F., Shen, X., Tang, Y.Q., Wang, S. **MicroRNA-1291 promotes endometrial fibrosis by regulating the ArhGAP29-RhoA/ROCK1 signaling pathway in a murine model.** Mol. Med. Rep. 2017; 16: 4501–4510

Yuan, Q., Blanton, S.H., Hecht, J.T. **Association of ABCA4 and MAFB with non-syndromic cleft lip with or without cleft palate.** Am. J. Med. Genet. Part A 2011; 155: 1469–1471

Zhang, B.Y., Wei, Y.S., Niu, J.M., Li, Y., Miao, Z.L., Wang, Z.N. **Risk factors for unexplained recurrent spontaneous abortion in a population from southern China.** International Journal of Gynecology and Obstetrics 2010; 108: 135–138

Zhou, Q., Yan, G., Ding, L., Liu, J., Yu, X., Kong, S. **EHD1 impairs decidualization by regulating the Wnt4/β-catenin signaling pathway in recurrent implantation failure.** EbioMedicine 2019; 50: 343–354

Received 29 July 2020; received in revised form 19 December 2020; accepted 12 January 2021.