



HAL
open science

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

Lenaïck Gourhant, Ozvan Bocher, Luc de Saint Martin, Thomas Ludwig, Anne Boland, Jean Deleuze, Pierre Dupré, Catherine Lemarié, Francis Couturaud, Cédric Le Maréchal, et al.

► To cite this version:

Lenaïck Gourhant, Ozvan Bocher, Luc de Saint Martin, Thomas Ludwig, Anne Boland, et al.. Whole exome sequencing, a hypothesis-free approach to investigate recurrent early miscarriage. *Reproductive BioMedicine Online*, 2021, 42 (4), pp.789-798. 10.1016/j.rbmo.2021.01.008 . hal-03700566

HAL Id: hal-03700566

<https://hal.science/hal-03700566v1>

Submitted on 24 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1

1 Whole Exome Sequencing, a hypothesis-free approach to investigate Recurrent Early

2

Miscarriage

3 Lenaick Gourhant¹, Ozvan Bocher², Luc De Saint Martin¹, Thomas E Ludwig², Anne Boland³, Jean F Deleuze³,
4 Philippe Merviel^{1,4}, Pierre F Dupré^{2,4}, Catherine A Lemarié^{1,2}, Francis Couturaud^{1,5}, Cédric Le Maréchal^{2,6},
5 Emmanuelle Génin², Elisabeth Pasquier^{1*}

6 ¹ Univ Brest, EA 3878 (GETBO), 29609, Brest, France

7 ² Univ Brest, Inserm, EFS, UMR 1078, GGB, CHU Brest, F-29200 Brest, France

8 ³ Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France

9 ⁴ Brest University Hospital, Division of Gynaecology, 29609, Brest, France

10 ⁵ Univ Brest, CIC-Inserm 1412, 29609, Brest, France

11 ⁶ Univ Brest, Laboratoire de Génétique Moléculaire et d'histocompatibilité, 29609, Brest, France

12

13 *Correspondence : Dr Elisabeth. Pasquier, CHU La Cavale Blanche, département de médecine vasculaire, interne et de
14 pneumologie, rue Tanguy Prigent, 29200, Brest, France.

15 E-mail: elisabeth.pasquier@chu-brest.fr

16

17

19 **Research Question:** Are there genetic determinants shared by unrelated women with unexplained recurrent early
20 miscarriage (REM)?

21 **Design:** We selected 30 REM cases and 30 controls with extreme-phenotype among women from eastern Brittany
22 (France), previously enrolled in an incident case-control study on thrombophilic mutations. Cases and controls were
23 selected based on the number of early miscarriages or live-births, respectively. Peripheral blood was collected for
24 DNA extraction at initial visit. We compared the burden of low frequency variants in the coding part of the genes
25 using Whole Exome Sequencing (WES).

26 **Result(s):** Cases had 3 to 17 early miscarriages (20 cases: ≥ 5 previous losses). Controls had 1 to 4 live births (20
27 controls: ≥ 3 previous live births) and not any miscarriage. WES data were available for 29 cases and 30 controls. A
28 total of 209,387 variants were found (mean variant per patient: 59,073.05) with no difference between groups (P
29 $=0.68$). The top 5 most significantly associated genes were *ABCA4*, *NFAM1*, *TCN2*, *AL078585.1* and *EPS15*. Previous
30 studies suggest the involvement of vitamin B12 deficiency in REM. *TCN2* encodes for vitamin B12 transporter into
31 cells. Therefore, holotranscobalamin (active vitamin B12) was measured for both cases and controls (81.2 ± 32.1 vs.
32 92.9 ± 34.3 pmole/L, respectively, $P=0.186$). Five cases but no control were below 50 pmole/L, ($P=0.052$).

33 **Conclusion(s):** Our study highlights 4 new genes of interest in REM, some of which belong to known networks of
34 genes involved in embryonic development (clathrin-mediated endocytosis and ciliary pathway). Additionally, our
35 study confirms the involvement of *TCN2* (vitamin B12 pathway) in the early first trimester of pregnancy.

36
37 **Key words:** recurrent early miscarriage, whole exome sequencing, *TCN2* gene, holotranscobalamin, extreme
38 phenotype, rare variants

40 Introduction

41 Recurrent early miscarriage (REM) is currently defined as three or more consecutive losses occurring in the first
42 trimester of pregnancy (Delabaere *et al.*, 2014). REM affects 1 to 2% of couples trying to have children. Standard
43 investigations fail to reveal any well-recognized cause in more than 50% of the couples (Branch *et al.*, 2010). To date,
44 on the basis of animal models and clinical studies, there are several hypotheses regarding possible underlying
45 mechanisms of unexplained REM: altered ovarian reserve, progesterone defect, 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, we investigated some possible determinants of REM such as factor V Leiden or mutation G20210A of the prothrombin in both parents (Pasquier *et al.*, 2009), skewed-X chromosome inactivation in the 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 (Zhang *et al.*, 2010; Woolner *et al.*, 2019). However, no evidence for strong genetic determinants of unexplained REM was found, particularly based on findings of hypothesis-based candidate gene studies (Odd 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 were 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). However, none of these genomic regions reached the genome-wide significance level. Moreover, those 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, as fetal akinesia (Rajcan-Separovic *et al.*, 2020). Only a few studies involved early losses. Those 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.

Thus, further studies in REM are needed. Using next generation sequencing, we aim 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, we selected female cases and controls at both ends of the phenotype distribution (i.e. with extreme phenotypes) and compared the burden of low frequency variants in the coding part of the different genes using Whole Exome Sequencing (WES).

72 **Material and Methods**

74 *Selection of cases and controls with extreme-phenotypes (Figure 1)*

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

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

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

93 *Samples (Figure 1)*

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

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 (3µg) was captured using in-solution enrichment methodology (Human All Exon v5 – 50 Mb, Agilent Technologies, CA, USA). Library preparation and exome enrichment protocol (~20.000 targeted genes) were performed automatically on a Bravo liquid handling robot (Agilent Technologies, CA, USA), according to manufacturer's instructions (SureSelect, Agilent Technologies CA, USA). After normalization and quality control, exome enriched libraries were sequenced on a HiSeq 2000 from Illumina (Illumina Inc., CA, USA) as paired-end 100 bp reads. Samples were sequenced in pools of 3 samples on each lane of the HiSeq2000 flow cell, in order to reach at least 90 to 100X of good quality samples. Image analysis and base calling 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 NGS pipeline in line with the Broad Institute 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 GATK Combine GVCF, 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 GATK and only the variants (SNVs 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 SNVs or ≤ 200 for indels, SOR (Symmetric Odds Ratio of 2x2 contingency table to detect strand bias) ≤ 3 for SNVs or ≤ 10 for indels, MQ (overall mapping quality of reads supporting a variant call) ≥ 40 for SNVs or ≥ 10 for indels were kept. When available, the following parameters were also used to filter variants: Inbreeding coefficient (InbreedingCoeff) $\geq (-0.8)$, MQRankSum (Z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities) $\geq (-12.5)$, ReadPosRankSum (Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias) was either $\geq (-8)$ for SNVs

125 or $\geq(-20)$ for indels. Genotypes with a depth <10 or a genotype quality (gq) <20 were set to missing and comparison of
126 missing rates was done between cases and controls. Only the variants with a genotype call rate greater than 90% and
127 not significantly different between cases and controls (p-value of the Fisher's exact test comparing the number of
128 missing genotypes between groups >0.001) were kept. A QC on the allelic balance was also performed to ensure that
129 the mean allelic balance calculated over heterozygous genotypes was within [25%-75%] in each group (cases and
130 controls). Once the set of eligible variants was determined, QC was performed at the individual level by checking the
131 ancestry within Europe (using principal component analysis (PCA) on a pruned subset of common variants ($MAF \geq$
132 0.05), heterozygosity levels and kinship between individuals. All these tests were done with plink1.9 (Chang *et al.*,
133 2015).

134 **Rare variant association tests.** To test the hypothesis that rare variants located in a same gene could be
135 involved in disease susceptibility, we performed rare variant association tests using the SKAT-O method (Lee et al.
136 2012) that combines burden and variance component tests and was shown to be the most powerful under different
137 disease susceptibility models. Only single nucleotide variants (SNV), that passed all QC, had an expected effect on the
138 protein (missense, stoploss, stopgain, canonical splice site variants) and with a minor allele frequency (MAF) below
139 0.05 in either the case or the control group, were included in the analysis. Data extraction and formatting were
140 performed in R using our own scripts implemented in the Ravages package (Bocher *et al.*, 2019) and association tests
141 were performed using the SKAT R package (<https://cran.r-project.org/web/packages/SKAT/>) (Lee et al. 2012).

143 *Active vitamin B12 measurement*

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

149
150 The main characteristics of women and their holo-TC blood levels were compared between the two groups using
151 SPSS Statistics v.19. Student's t-test and chi-squared test were used for continuous variables (ages, BMI, tobacco use,
152 parity ratios by pregnancy (calculated at inclusion in the case-control study and at the end of follow-up), duration of
153 follow-up, holo-TC blood levels) and categorical variables (allelic distribution), respectively. For each woman, the

7

155 parity ratios were calculated as follows: the number of previous live births divided by the number of previous started
156 pregnancies. The odds ratio (OR) was calculated using logistical regression. $P < 0.05$ was considered statistically
157 significant.

158

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

161

162

Results*Case and control characteristics*

We selected 30 cases and 30 controls with an extreme phenotype among the women previously recruited in our master incident case-control study on thrombophilic mutations in Eastern Brittany (France) (Pasquier *et al.*, 2009):

i) Cases had a past history of 3 to 17 early miscarriages at inclusion in the master incident case-control study (median:5, IR: 4-6.25). Twenty of the cases had 5 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, IR: 5-7).

ii) Controls had 1 to 4 previous live births at inclusion in the case control-study (median:3, IR: 2-3). Twenty of the controls had 3 and 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 ($P>0.1$). No controls but 4 cases were lost to follow-up. Cases were significantly younger than controls at the end of follow-up (38.0 ± 5.7 vs. 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 ($P=0.173$).

WES

Only one individual (a case) was removed after the quality control (QC): this individual had a high heterozygosity level and was found to be an outlier on the PCA. Thus, whole exome sequencing (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 controls, t-test $P=0.668$). Among these variants, 195,090 are single nucleotide variants (SNV) and on average an individual carries 54,112.41 SNVs (54,091.93 in average in cases and 54,132.20 in controls, t-test $P=0.683$).

Rare Variant Association Tests

For the rare variant association test, only missense, stoploss, stopgain and canonical splice site SNVs with a minor allele frequency 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 5 list of the most significantly associated genes is given in Table 2. None of the gene reached genome-wide significance (usually defined as a p-value below 2.5×10^{-6} for rare variant association tests). The 5 most significant genes were *ABCA4*, *NFAMI*, *TCN2*, *AL078585.1* and *EPS15*. Interestingly, *TCN2* encodes for a protein involved in the 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 to the GnomAD reference population (Karczewski *et al.*, 2019) (0.190 and 0.033 respectively compared to 0.120 in GnomAD European populations).

Holo-TC measurements

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

The quartiles of the whole distribution (cases and controls together) were as follows: 60.3, 83.45 and 107.8 pmole/L, respectively. The quartiles distribution of Holo-TC blood levels was not statistically different between cases and controls. However, as the normal distribution of Holo-TC blood levels was left-truncated at 50 pmole/L in the controls, we considered a threshold of 50 pmole/L: 5 cases but no controls were below 50 pmole/L, ($P=0.052$).

In the cases, lower Holo-TC blood level 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 levels using regression analyses.

217 *Correlations between Holo-TC measurements and TCN2 mutations*

218 Holo-TC levels from cases were negatively and significantly correlated with rs 9606756 polymorphism (cases:
219 P=0.03). Holo-TC levels were also correlated with rs 9606756 polymorphism when cases and controls were analyzed
220 together (all: P=0.007).

221 The other mutations were not significantly correlated with Holo-TC levels (Figure 4).

222

Discussion

Using a WES strategy free of a-priori hypotheses, we compared 2 groups of fertile women with extreme phenotypes. Exome data was analyzed for 29 of the 30 women included in the REM group. Those women had a REM history of at least 3 losses with a pregnancy failure rate of 84 %. The control group included 30 women who did not suffer any miscarriage and had 2 to 4 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 analyzed. 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 our 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 levels of Holo-TC (TC-vitamin B12 complex) were measured in both cases and controls. At inclusion, the cases displayed lower (but not significantly) Holo-TC blood levels than the controls (81.2 ± 32.1 vs 92.9 ± 34.3 pmole/L, $P=0.186$). Notably, *TCN2* rs9606756 polymorphism was highly and negatively correlated to the blood levels of Holo-TC ($P=0.007$), especially in the cases ($P=0.03$), as previously reported in overall population (Riedel et al., 2011).

Folic acid and the whole vitamin B12/folic acid pathway (Quadros *et al.*, 2013) is known to be essential during early pregnancy and was 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 increase the risk of recurrent abortion (Nelen *et al.*, 2000). Vitamin B12, at serum B12 levels under the threshold of 180 pg/L was highly associated with REM compared to levels found in healthy premenopausal women (Reznikoff-Etievant *et al.*, 2002). Additionally, vitamin B12 was found at serum levels significantly lower in REM women compared to 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 level in the blood, was not statistically different between cases and controls (Hübner *et al.*, 2008). Here, we proposed another method for assessment of functional vitamin B12 through the measurements of Holo-TC. To the best of our knowledge, this method has never been reported in women with REM. Lastly, two case-control studies investigated known *TCN2* polymorphisms in REM women compared to controls (Kim *et al.*, 2014; Hashemi *et al.*, 2018). Kim et al 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 did not find any association between REM and *TCN2* rs180119 (a common polymorphism) or *TCN2* rs2336573 among Iranian women. More recently, Park et al reported an

increased risk for recurrent implantation failure after in vitro fertilization for carriers of the *TCN2* rs9606756 G allele ($P=0.025$ OR: 4.732; 95% CI 1.220-18.356) (Park *et al.*, 2019).

In addition, we are 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 that appears to be important in embryonic development (Offenhäuser *et al.*, 2000; Milesi *et al.*, 2019). *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 that plays a role in the endocytosis of insulin growth factor (IGF)-1 receptors. Notably, EHD1 might be involved in endometrium decidualization and be implied in recurrent implantation failure (Zhou *et al.*, 2019).

ABCA4 is a member of the ATP-binding cassette (ABC) family transport proteins and codes a retina-specific ABC transporter whose substrate is N-retinylidene-PE expressed in retina photoreceptor cells. *ABCA4* is implicated in the 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 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 long-non 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 2 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 signaling and development of B-cells (Ohtsuka *et al.*, 2004).

283 In our study, it is noteworthy that the MAF observed in both groups for these five top candidate genes are quite
284 different than those recorded in reference databases as The Genome Aggregation Database (gnomAD) (Karczewski *et*
285 *al.*, 2019) or the French Exome Project Database (FREX) (Génin *et al.*, 2017). This is probably explained by the
286 extreme phenotype of cases and controls.

287 Apart from the free of a-priori hypotheses approach, one of the strong points of our study lies in the selection of
288 controls from the electoral lists, in the same geographic area (eastern Brittany, a geographic area with low ethnic
289 mixing) and in the same period than the cases.

290 Due to the limited sample size of our study, we were not able to find genome-wide significant associations; i.e.
291 associations that remain significant after correction for multiple testing at the scale of all the genes tested. Regarding
292 common variants, a “suggestive association” is defined by $P < 10^{-5}$. Regarding rare variants, fewer tests are done, since
293 the variants of each gene are grouped together. Consequently, some authors used higher threshold for P value as $< 10^{-3}$
294 (Grant *et al.*, 2018). Nevertheless, the three genes ABCA4, NFAM1 and TCN2 showed interesting signals with P -
295 values $< 5 \cdot 10^{-4}$. Using asymptotical calculations, accounting for the genes estimated effect sizes and observed
296 cumulative MAF, the probability of detecting genome-wide significant associations ranged from 78.5% to 93.1%
297 (78.5% for TCN2, 87.8% for NFAM1 and 93.1% for ABCA4). This confers a good study power and the small sample
298 size per se does not lower the clinical evidence of our results.

299 We acknowledge that the few other WES studies in REM most often involved the trio mother-father-embryo and did
300 not often reveal the same modified genes between the different families. The design of our case-control study focused
301 on shared altered genes among women with REM. As a limitation, our study did not involve the two members of the
302 couples enrolled in our princeps study on thrombophilic mutations (Pasquier *et al.*, 2009). Moreover, miscarriage
303 products were not available. Other limitations of our study warrant consideration. We did not study the functional
304 impact of the gene polymorphism of the other 4 genes that came out from WES study. Regarding *TCN2*, we found a
305 significant association between rs9606756 polymorphism and i) REM ($P=0.03$) as well as ii) Holo-TC levels
306 ($P=0.007$). However, we did not find any significant difference of Holo-TC levels between cases and controls,
307 which may be due to the routine practice of vitamin supplementation in REM women. This could suggest a more
308 complex functional impact of *TCN2* rs9606756 favoring miscarriage than B12 deficiency. Moreover, Holo-TC
309 measurement is not a perfect marker for B12 status or B12 cell intake (Harrington, 2017).

311 In conclusion, despite some limitations, our study confirms the involvement of the vitamin B12 pathway and
312 especially *TCN2* in the early first trimester of pregnancy. This reinforces the rationale for seeking vitamin B12

313 deficiency concomitant with folate deficiency, in women with REM. Furthermore, we highlighted new genes of
314 interest in REM, some of which belong to known networks of genes involved in embryonic development (clathrin-
315 mediated endocytosis and ciliary pathway). Others might be involved in endometrium homeostasis. Regarding these
316 new genes, further studies are required to replicate our results and elucidate the underlying mechanisms leading to
317 REM.

318

319 Acknowledgements

320 The authors thank the GYNOBS 29 network and Zarrin Alavi (CIC-Inserm 1412).

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

323

324 Funding

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

326

327 **Conflict-of-interest disclosure:** None

328

329

330

331

References

- 332
333
- 334 Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin
335 MD, Redett RA, Raymond G, Schwender H, Jin SC, Cooper ME, Dunnwald M, Mansilla MA, Leslie E,
336 Bullard S, Lidral AC, Moreno LM, Menezes R, Vieira AR, Petrin A, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou
337 YH, Chen PK, Wang H, Ye X, Huang S, Yeow V, Chong SS, Jee SH, Shi B, Christensen K, Melbye M,
338 Doheny KF, Pugh EW, Ling H, Castilla EE, Czeizel AE, Ma L, Field LL, Brody L, Pangilinan F, Mills JL,
339 Molloy AM, Kirke PN, Scott JM, Arcos-Burgos M, Scott AF. A genome-wide association study of cleft lip
340 with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat Genet.* 2010;42:525-9.
- 341 Bocher O, Marenne G, Saint Pierre A, Ludwig TE, Guey S, Tournier-Lasserre E, Perdry H, Génin E. Rare variant
342 association testing for multicategory phenotype. *Genet Epidemiol.* 2019;43:646-656.
- 343 Branch DW, Gibson M, Silver RM. Clinical practice. Recurrent miscarriage. *N Engl J Med.* 2010;363:1740-1747.
- 344 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the
345 challenge of larger and richer datasets. *Gigascience.* 2015; 25:4-7
- 346 Delabaere A, Huchon C, Lavoue V, Lejeune V, Iraola E, Nedellec S, Gallot V, Capmas P, Beucher G., Subtil D,
347 Carcopino X, Vialard F, Nizard J, Quibel T, Costedoat-Chalumeau N, Legendre G, Venditelli F, Rozenberg P,
348 Lemery D, Deffieux X. [Definition of pregnancy losses: Standardization of terminology from the French
349 National College of Obstetricians and Gynecologists (CNGOF)]. *J. Gynecol Obstet Biol Reprod (Paris)* 2014
350 Dec; 43: 756–763
- 351 Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, Wright FA, Rieder MJ, Tabor HK,
352 Nickerson DA, Barnes KC; National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing
353 Project; Lung GO, Gibson RL, Bamshad MJ. Exome sequencing of extreme phenotypes identifies DCTN4 as
354 a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat Genet.* 2012;44:886-889.
- 355 Genin E, Redon R, Deleuze JF, Champion D, Lambert JC, Dartigues JF for the FREX Consortium (2017) The
356 French Exome (FREX) Project: A Population-based Panel of Exomes to Help Filter Out Common Local
357 Variants. *Genet Epidemiol* 2017; 41:691-691.

- 358 Grant RC, Denroche RE, Borgida A, Virtanen C, Cook N, Smith AL, Connor AA, Wilson JM, Peterson G, Roberts
359 NJ, Klein AP, Grimmond SM, Biankin A, Cleary S, Moore M, Lemire M, Zogopoulos G, Stein L, Gallinger
360 S. Exome-Wide Association Study of Pancreatic Cancer Risk. *Gastroenterology*. 2018;154:719-722
- 361 Harrington DJ. Laboratory assessment of vitamin B12 status. *J Clin Pathol* 2017;70:168–173.
- 362 Hashemi M, Mokhtari M, Yazdani-Shahrbabaki V, Danesh H, Bizhani F, Taheri M. Evaluation of transcobalamin
363 II rs1801198 and transcobalamin II receptor rs2336573 gene polymorphisms in recurrent spontaneous
364 abortion. *J Obstet Gynaecol*. 2018;38:860-863.
- 365 Hübner U, Alwan A, Jouma M, Tabbaa M, Schorr H, Herrmann W. Low serum vitamin B12 is associated with
366 recurrent pregnancy loss in Syrian women. *Clin Chem Lab Med*. 2008;46:1265-9.
- 367 Karczewski K J, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM., Ganna A,
368 Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM,
369 Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin
370 N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B,
371 Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S,
372 Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N,
373 Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski
374 ME; Genome Aggregation Database Consortium, Neale BM, Daly MJ, MacArthur DG. The mutational
375 constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020 May;581(7809):434-443.
- 376 Kim HS, Lee BE, Jeon YJ, Rah H, Lee WS, Shin JE, Choi DH, Kim NK. Transcobalamin II (TCN2 67A>G and
377 TCN2 776C>G) and transcobalamin II receptor (TCbIR 1104C>T) polymorphisms in Korean patients with
378 idiopathic recurrent spontaneous abortion. *Am J Reprod Immunol*. 2014 ;72:337-46.
- 379 Kolte AM, Nielsen HS, Moltke I, Degen B, Pedersen B, Sunde L, Nielsen FC, Christiansen OB. A genome-wide
380 scan in affected sibling pairs with idiopathic recurrent miscarriage suggests genetic linkage. *Mol Hum Reprod*
381 2011;17:379–85.
- 382 Leclercq E, de Saint Martin L, Bohec C, Le Martelot MT, Roche S, Alavi Z, Mottier D, Pasquier E. Blood anti-
383 Müllerian hormone is a possible determinant of recurrent early miscarriage, yet not conclusive in predicting a
384 further miscarriage. *Reprod Biomed Online*. 2019;39(2):304-311.
- 385 Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA; NHLBI GO Exome Sequencing Project—
386 ESP Lung Project Team, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant

- 387 association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum*
388 *Genet.* 2012;91:224–237.
- 389 Martin RH, Harper TA, Kelso W. Serum-folic acid in recurrent abortions. *Lancet* 1965;1:670–2.
- 390 Merrigan SD, Owen WE, Straseski JA. Performance characteristics of the ARCHITECT Active-B12
391 (Holo-transcobalamin) assay. *Clin Lab.* 2015;61:283-8.
- 392 Milesi C, Alberici P, Pozzi B, Oldani A, Beznoussenko GV, Raimondi A, Soppo BE, Amodio S, Caldieri G,
393 Malabarba MG, Bertalot G, Confalonieri S, Parazzoli D, Mironov AA, Tacchetti C, Di Fiore PP, Sigismund S,
394 Offenhäuser N. Redundant and nonredundant organismal functions of EPS15 and EPS15L1. *Life Sci Alliance*
395 2019; 2: e201800273.
- 396 Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated
397 cells. *Nucleic Acids Res* 1988;16:1215.
- 398 Morris VE, Hashmi SS, Zhu L, Maili L, Urbina C, Blackwell S, Greives MR, Buchanan EP, Mulliken JB, Blanton
399 SH, Zheng WJ, Hecht JT, Letra A. Evidence for craniofacial enhancer variation underlying nonsyndromic
400 cleft lip and palate. Evidence for craniofacial enhancer variation underlying nonsyndromic cleft lip and palate.
401 *Hum Genet.* 2020; doi: 10.1007/s00439-020-02169-9.
- 402 Nelen WL, Blom HJ, Steegers EA, Den Heijer M, Thomas CM, Eskes TK. Homocysteine and folate levels as risk
403 factors for recurrent early pregnancy loss. *Obstet Gynecol* 2000;95:519–24.
- 404 Offenhäuser N, Santolini E, Simeone A, Di Fiore PP. Differential patterns of expression of Eps15 and Eps15R
405 during mouse embryogenesis. *Mech Dev.* 2000 Jul;95:309-12.
- 406 Ohtsuka M, Arase H, Takeuchi A, Yamasaki S, Shiina R, Suenaga T, Sakurai D, Yokosuka T, Arase N, Iwashima
407 M, Kitamura T, Moriya H, Saito T. NFAM1, an immunoreceptor tyrosine-based activation motif-bearing
408 molecule that regulates B cell development and signaling. *Proc Natl Acad Sci U S A.* 2004;101:8126–8131.
- 409 Park HS, Kim JO, An HJ, Ryu CS, Ko EJ, Kim YR, Ahn EH, Lee WS, Kim JH, Kim NK. Genetic polymorphisms
410 of the cobalamin transport system are associated with idiopathic recurrent implantation failure. *J Assist*
411 *Reprod Genet.* 2019;36:1513-1522.
- 412 Pasquier E, Bohec C, De Saint Martin L, Le Maréchal C, Le Martelot MT, Roche S, Laurent Y, Férec C, Collet M,
413 Mottier D. Strong evidence that skewed X-chromosome inactivation is not associated with recurrent
414 pregnancy loss: an incident paired case control study. *Hum Reprod.* 2007;22:2829-33.

- 415 Pasquier E, Bohec C, Mottier D, Jaffuel S, Mercier B, Férec C, Collet M and de Saint Martin L. Inherited
416 thrombophilias and unexplained pregnancy loss: an incident case–control study. *J Thromb Haemost.* 2009; 7:
417 306-11
- 418 Pereza N, Ostojić S, Kapović M, Peterlin B. Systematic review and meta-analysis of genetic association studies in
419 idiopathic recurrent spontaneous abortion. *Fertil Steril.* 2017.;107:150–159
- 420 Qiao Y, Wen J, Tang F, Martell S, Shomer N, Leung PC, Stephenson MD, Rajcan-Separovic E. Whole exome
421 sequencing in recurrent early pregnancy loss. *Mol Hum Reprod.* 2016;22:364-72
- 422 Quadros EV, Sequeira JM. Cellular uptake of cobalamin: transcobalamin and the TCbIR/CD320 receptor.
423 *Biochimie.* 2013;95:1008–18
- 424 Quintero-Ronderos P, Mercier E, Fukuda M, Gonzalez R, Suarez CF, Patarroyo MA, Vaiman D, Gris JC, Laissue
425 P. Novel genes and mutations in patients affected by recurrent pregnancy loss. *PLoS One* 2017; 12 (10),
426 e0186149.
- 427 Rajcan-Separovic E. Next generation sequencing in recurrent pregnancy loss-approaches and outcomes. *Eur J Med*
428 *Genet.* 2020;63(2):103644.
- 429 Reznikoff-Etievant MF, Zittoun J, Vaylet C, Pernet P, Milliez J. Low vitamin B12 level as a risk factor for very
430 early recurrent abortion. *Eur J Obstet Gynecol Reprod Biol.* 2002;104:156–9.
- 431 Shastry BS. Evaluation of the common variants of the ABCA4 gene in families with Stargardt disease and
432 autosomal recessive retinitis pigmentosa. *Int J Mol Med.* 2008;21:715–720.
- 433 Riedel BM, Molloy AM, Meyer K, Fredriksen A, Ulvik A, Schneede J, Nexø E, Hoff G, Ueland PM.
434 Transcobalamin polymorphism 67A->G, but not 776C->G, affects serum holotranscobalamin in a cohort of
435 healthy middle-aged men and women. *J Nutr.* 2011;141:1784–1790.
- 436 Sütterlin M, Bussen S, Ruppert D, Steck T. Serum levels of folate and cobalamin in women with recurrent
437 spontaneous abortion. *Hum Reprod.* 1997 Oct;12(10):2292-6.
- 438 Thumberger T, Hagenlocher C, Tisler M, Beyer T, Tietze N, Schweickert A, Feistel K, Blum M. Ciliary and non-
439 ciliary expression and function of PACRG during vertebrate development. *Cilia.* 2012;1:13.
- 440 van Bergen En Henegouwen PM. Eps15: a multifunctional adaptor protein regulating intracellular trafficking. *Cell*
441 *Commun Signal.* 2009.8;7:24.
- 442 Velázquez-Aragón JA, Alcántara-Ortigoza MA, Estandia-Ortega B, Reyna-Fabián ME, Méndez-Adame CD,
443 González-Del Angel A. Gene Interactions Provide Evidence for Signaling Pathways Involved in Cleft
444 Lip/Palate in Humans. *J Dent Res.* 2016;95(11):1257-1264.

- 445 Wang Li, Zeng Chan Wang, Cui Xie, Xiao Feng Liu, Mao Sheng Yang. Genome-wide screening for risk loci of
446 idiopathic recurrent miscarriage in a Han Chinese population: a pilot study. *Reprod Sci* 2010;17:578–84.
- 447 Woolner AMF, Raja E-A, Bhattacharya S, Danielian P, Bhattacharya S. Inherited susceptibility to miscarriage: a
448 nested case-control study of 31565 women from an intergenerational cohort. *Am J Obstet Gynecol* 2019. pii:
449 S0002-9378(19)31013-0.
- 450 Xu Q, Duan H, Gan L, Liu X, Chen F, Shen X, Tang YQ, Wang S. MicroRNA-1291 promotes endometrial fibrosis
451 by regulating the ArhGAP29-RhoA/ROCK1 signaling pathway in a murine model. *Mol Med Rep.*
452 2017;16(4):4501-4510.
- 453 Yuan Q, Blanton SH, Hecht JT. Association of ABCA4 and MAFB with non-syndromic cleft lip with or without
454 cleft palate. *Am J Med Genet Part A.* 2011; 155:1469–1471.
- 455 Zhang BY, Wei YS, Niu JM, Li Y, Miao ZL, Wang ZN. Risk factors for unexplained recurrent spontaneous
456 abortion in a population from southern China. *International Journal of Gynecology and Obstetrics* 2010;
457 108:135–138
- 458 Zhou Q, Yan G, Ding L, Liu J, Yu X, Kong S. EHD1 impairs decidualization by regulating the Wnt4/ β -catenin
459 signaling pathway in recurrent implantation failure. *EbioMedicine* 2019; 50: 343–354.

460
461

462

	CASES	CONTROLS	P value
N	30 (29)	30	
Age at inclusion, y (mean±SD)	32.9±4.6 (32.9±4.7)	34.3±3.2	0.170
Median [min-max]	33.7[20.3-41.8]	34.7[28.1-42]	(0.162)
At inclusion, pack Smoking -years (mean±SD)	3.3±4.6 (3.3±4.6)	4.5±7.0	0.467 (0.467)
BMI	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, y (mean±SD)*	5.7±3.0 (5.5±3.0)	6.8±1.6	0.117 (0.067)
Age at the end of follow-up, y (mean±SD)*	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)*	0.16±0.13 (0.16±0.13)	1	<0.001

463

464

465

466

467

Table 1. Main characteristics of cases compared to the 30 controls, given either for all the 30 cases or the 29 cases with available WES (*italic*). *4 missing data among cases (omitted from the analyses)

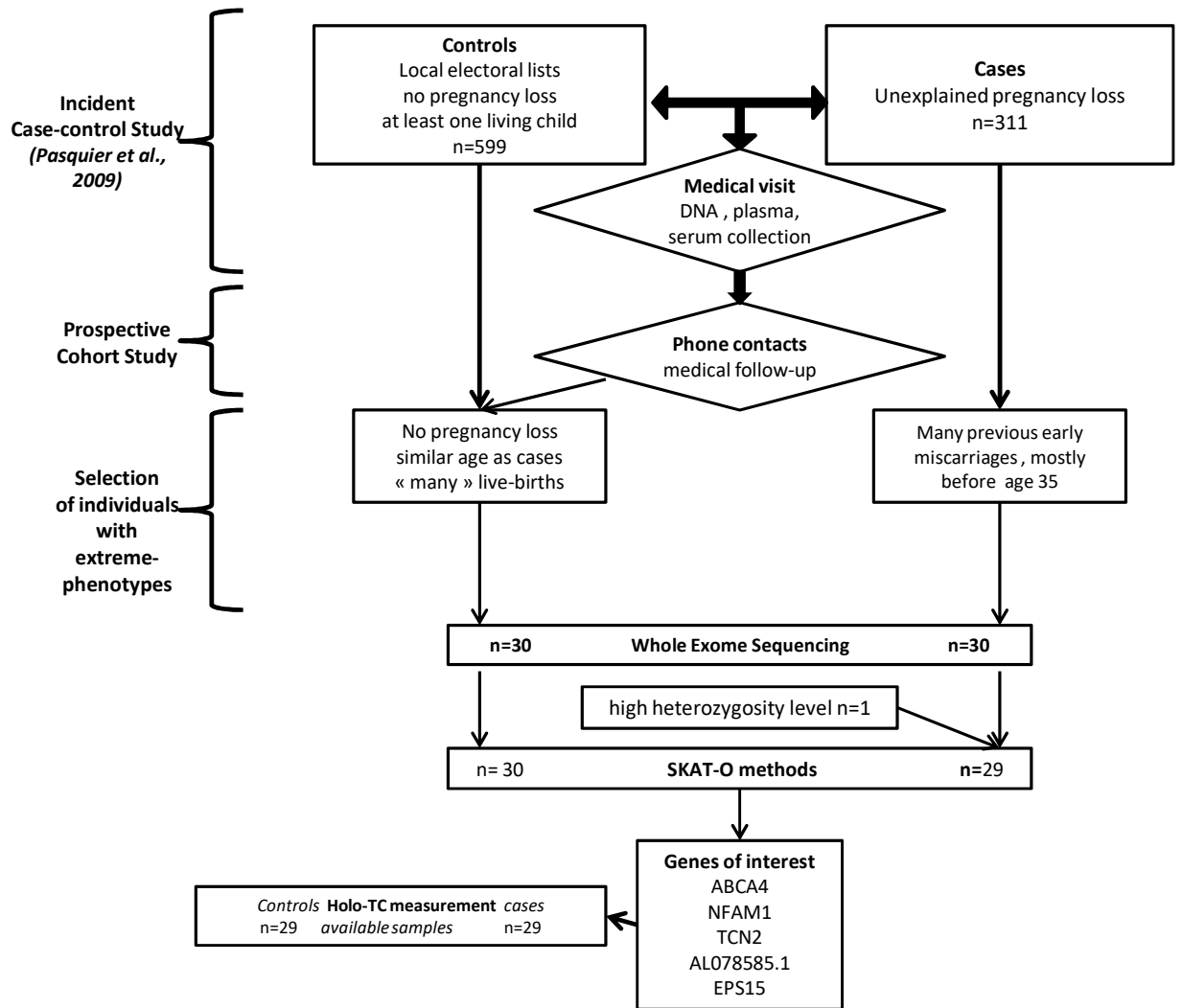
Gene	P-value of SKAT-0 test	Number of Variants	cMAF	
			Cases	controls
ABCA4	6.74e-05	7	0.259	0.017
NFAM1	9.97e-05	2	0.483	0.033
TCN2	2.86e-04	6	0.534	0.083
AL078585.1	5.69e-04	1	0.224	0.017
EPS15	8.928e-04	3	0.190	0.017

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]

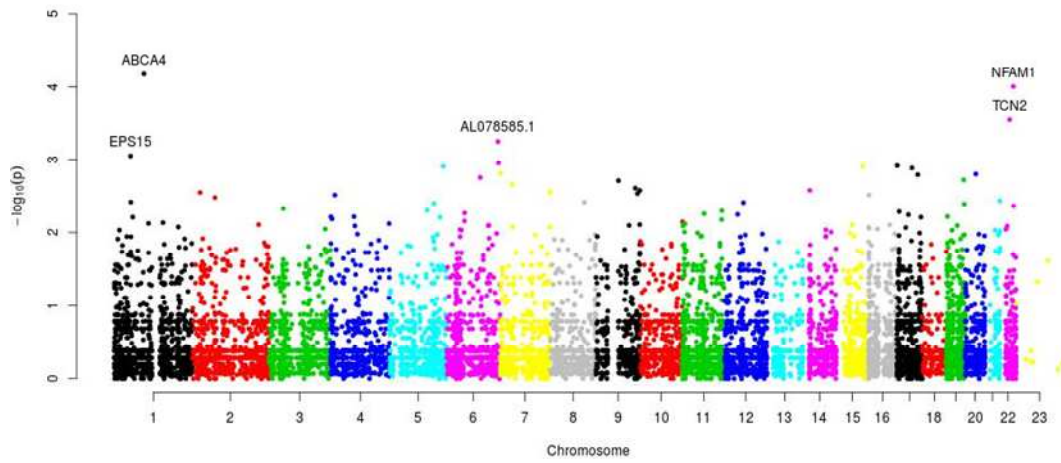
470 **Table 2. Five most associated genes in the SKAT-0 rare variant association test.** The p-value of the test is given as
471 well as the number of variants and their cumulative minor allele frequency (cMAF) in cases and controls. P-values,
472 Odds Ratios (OR) and their 95% confidence interval are also given using the burden test WSS.

Gene	CHR	POS	RS	REF	ALT	WORST_ANNOT	FreqALT_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

475 **Table 3. List of variants from the top 5 genes included in the rare variant association test.** Genes are ranked from
476 the most significant to the less significant. For each variant, its position, rs name, reference and alternative alleles and
477 the frequencies of the alternative allele in the cases, controls and in GnomAD total database and GnomAD Non-
478 Finnish European (NFE) population are reported.



483 **Figure 2:** Manhattan plot of the rare variant association test SKAT-O. For each gene ordered by chromosome and
484 position on chromosome, the $-\log_{10}p$ -value of the SKAT-O test is given. The top 5 most associated genes are named.

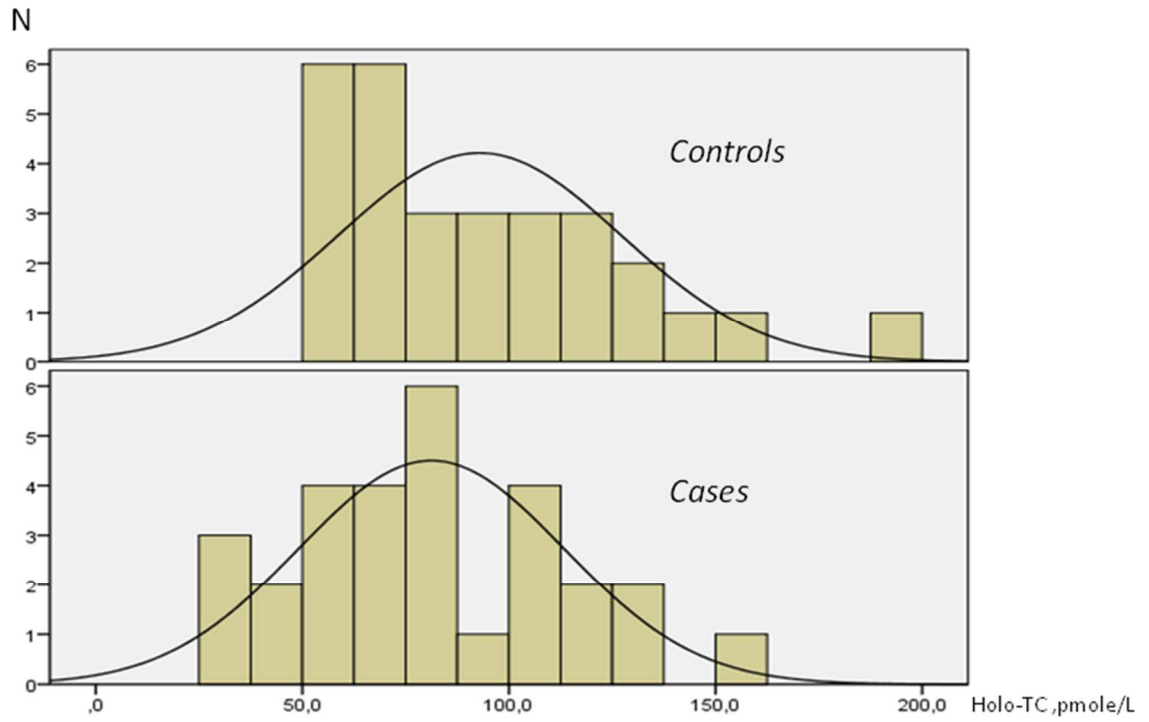


485

486

487 **Figure 3:** Distribution of the Holo-TC blood levels among the 29 cases and 29 controls.

488

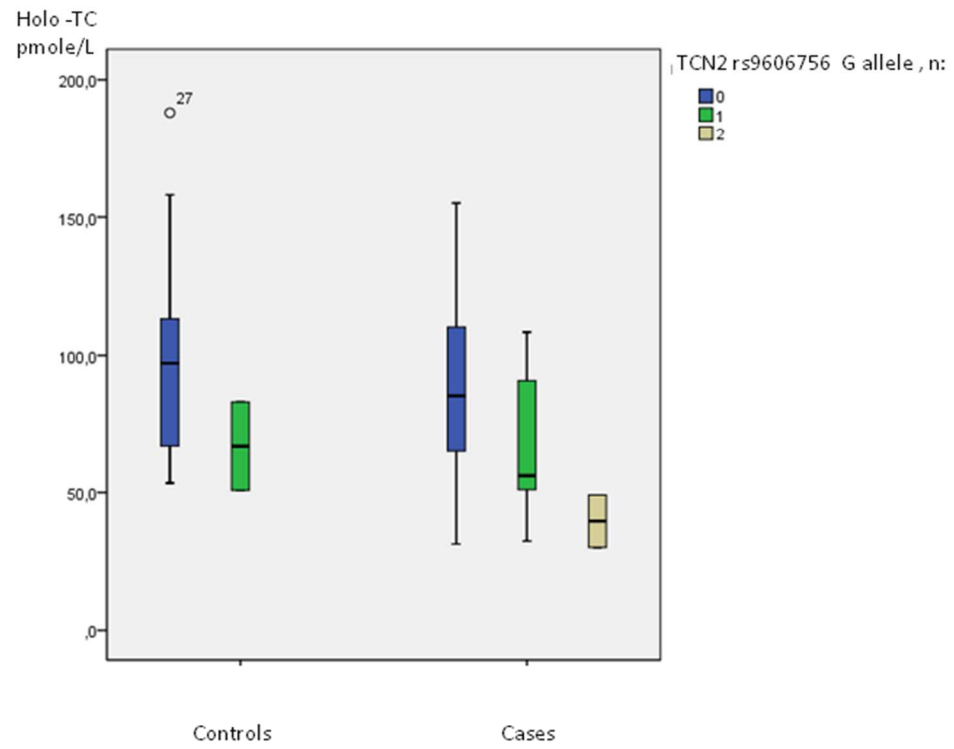


489

490

491 **Figure 4:** Distribution of the Holo-TC blood levels among the 29 cases and 29 controls according to rs9606756
492 allelic distribution

493



494