Association of FOXD1 variants with adverse pregnancy outcomes in mice and humans
Paul Laissue, Besma Lakhal, Magalie Vatin, Frank Batista, Gaëtan Burgio, Eric Mercier, Esther Santos, Christophe Buffat, Diana Sierra-Diaz, Gilles Renault, et al.

To cite this version:
Paul Laissue, Besma Lakhal, Magalie Vatin, Frank Batista, Gaëtan Burgio, et al.. Association of FOXD1 variants with adverse pregnancy outcomes in mice and humans. Open Biology, Royal Society, 2016, 6 (10), pp.160109. 10.1098/rsob.160109 . inserm-01399338

HAL Id: inserm-01399338
https://www.hal.inserm.fr/inserm-01399338
Submitted on 18 Nov 2016

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.
Association of FOXD1 variants with adverse pregnancy outcomes in mice and humans

Paul Laissue1,2,3, Besma Lakhal4,1, Magalie Vatin1,2,†, Frank Batista5, Gaëtan Burgio6,7, Eric Mercier8, Esther Dos Santos9,10, Christophe Buffat1, Diana Carolina Sierra-Diaz3, Gilles Renault1,2, Xavier Montagutelli6, Jane Salmon12, Philippe Monget13, Reiner A. Veitia1,2, Céline Méhats1,2, Marc Fellous1,2, Jean-Christophe Gris8, Julie Cocquet1,2 and Daniel Vaiman1,2

1Institut Cochin, Université Paris Descartes, CNRS (UMR 8104), Paris, France
2Inserm, U1016 Paris, France
3Centro de Investigación en Genética y Genómica (CIGGUR), Grupo GENIUS–Escuela de Medicina y Ciencias de la Salud, Universidad del Rosario, Bogotá, Colombia
4Department of Cytogenetics and Reproductive Biology, Farhat Hached University Teaching Hospital, Sousse, Tunisia
5Department of Cell Biology, Albert Einstein College of Medicine, New York, NY 10461, USA
6Institut Pasteur, Unité de Génétique des Mammifères, Paris, France
7Department of Immunology and Infectious Disease, The John Curtin School of Medical Research, the Australian National University, 131 Garran Road, Canberra 2601, Australian Capital Territory, Australia
8Department of Haematology, University Hospital, Nîmes. Faculty of Pharmacy and Research Team EA 2992, University of Montpellier, Montpellier, France
9GéG-EA 7404, Université de Versailles-Saint Quentin en Yvelines, Unité de Formation et de Recherche des Sciences de la Santé Simone Veil, 78180 Montigny-le-Bretonneux, France
10Service de Biologie Médicale, Centre Hospitalier de Poissy-Saint-Germain, 78300 Poissy, France
11Centre National de la Recherche Scientifique UMR 7278, IRD198, INSERM U1095, Aix-Marseille Université, Marseille, France
12Department of Medicine, Weill Medical College of Cornell University, New York, NY 10021, USA
13INRA – CNRS, Université de Tours—Haranas Nationaux, IFR 135, 37380 Nouzilly, France

DV, 0000-0002-1915-0717

Recurrent spontaneous abortion (RSA) is a common cause of infertility, but previous attempts at identifying RSA causative genes have been relatively unsuccessful. Such failure to describe RSA aetiological genes might be explained by the fact that reproductive phenotypes should be considered as quantitative traits resulting from the intricate interaction of numerous genetic, epigenetic and environmental factors. Here, we studied an interspecific recombinant congenic strain (IRCS) of Mus musculus from the C57BL6/J strain of mice harbouring an approximate 5 Mb DNA fragment from chromosome 13 from Mus spretus mice (66H-MMU13 strain), with a high rate of embryonic resorption (ER). Transcriptome analyses of endometrial and placental tissues from these mice showed a deregulation of many genes associated with the coagulation and inflammatory response pathways. Bioinformatics approaches led us to select Foxd1 as a candidate gene potentially related to ER and RSA. Sequencing analysis of Foxd1 in the 66H-MMU13 strain, and in 556 women affected by RSA and 271 controls revealed non-synonymous sequence variants. In vitro assays revealed that some led to perturbations in FOXD1 transactivation properties on promoters of genes having key roles during implantation/placentation, suggesting a role of this gene in mammalian implantation processes.

1. Introduction

Human infertility represents a public health concern affecting 10–15% of all couples [1]. Despite advances in diagnosis and treatment, approximately 30% of...
cases are still considered idiopathic [2]. Recurrent spontaneous abortion (RSA) is clinically defined by at least three pregnancy losses prior to the 20th week of gestation and is a common cause of infertility, because it affects 1% of all pregnancies [3–6]. Approximately 50% of those are considered idiopathic, thereby underlining the disease’s potential genetic and epigenetic causes. Unfortunately, previous attempts at identifying RSA causative genes have been relatively unsuccessful. Several candidate genes (e.g. AMN, TM, EPCR, VEGF, p53, eNOS, JAK2, MTHFR, WNT6) have been studied, especially using Sanger sequencing, but only a few variants and genotypes have been associated with the phenotype [7–11]. Although some genetic markers indicative of an elevated risk of being affected by RSA have been proposed, functional evidence is rare. This has restricted their efficient use in clinical studies. Genome-wide scan-based studies have been reported, although they have not reached the classical accepted statistical threshold for significance and have apparently failed to identify specific genes [12,13]. Such failure to identify RSA aetiologies might be explained by the fact that reproduction’s inherent complexity theoretically implies that mutations in hundreds of candidate genes may be responsible for the phenotype [10,14]. Furthermore, the genetic study (e.g. via classical genetic linkage analysis or GWAS) of families affected by RSA is particularly challenging owing to their rarity. Indeed, causative variants related to reproductive fitness are under strong negative selection. Compared with genetic analysis of ovarian infertility, which allowed discovery of interesting genes [15,16], pertinent candidates genes were seldom found when the infertility was linked to placental/endometrial defects, such as RSA [14].

It is worth stressing that mammalian reproductive phenotypes (and their inherent molecular mechanisms) should be considered as quantitative traits resulting from the intricate interaction of numerous genetic, epigenetic and environmental factors. With that in mind, our group took advantage of a particular mouse interspecific recombinant congenic strains (IRCS) model allowing the identification of quantitative trait loci (QTL) related to complex phenotypes [17–21]. Phenotyping of IRCS animals enabled us to map several QTL related to embryonic resorption (ER) and lethality [20,22]. Murine ER might be caused by molecular disturbances regarding implant and fetal-placental unit function, and it is highly plausible that causal genes play similar roles in human diseases, such as RSA, preeclampsia and/or fetal growth restriction [23–25].

In a previous study, we showed that the 66H-MMU13 IRCS strain is affected by a high ER rate (14.7% versus 4.6% observed in C57BL/6J females, p < 0.01). This strain contains a unique approximately 5 Mb spruetus fragment on chromosome MMU13 (between rs120693734 and D13Mit47 polymorphic markers) encompassing 31 genes [22].

This study includes whole transcriptome analysis of endometrial and placental tissues from the IRCS 66H-MMU13 presenting with a high ER rate. Transcriptomics and bioinformatics approaches led us to select Foxd1 (a forkhead transcription factor located in the critical fragment) as a candidate gene potentially related to ER and RSA. Sequencing analysis of Foxd1 coding regions in the 66H-MMU13 strain, Mus spruetus animals, 556 women affected by RSA and in a control population revealed many non-synonymous sequence variants. In vitro assays revealed that some of them (e.g. Foxd1-p.Thr152Ala, FOXD1-Ala356Gly, FOXD1-Ile364Met and FOXD1-429AlaAla) had a functional effect as they led to perturbations in FOXD1 transactivation properties on promoters of the Placental Growth Factor (PGF) and the complement component gene (C3) having key roles during implantation/placentation [26–28]. Finally, with our study, we found that women with FOXD1 mutations have a statistically high risk (10.3 relative risk) of suffering RSA.

Taken together, our results showed that FOXD1 is a major actor in mammalian reproduction as sequence variants generated ER and RSA in mice and humans, respectively. We propose that the FOXD1-p.429AlaAla mutation might be used as an RSA molecular marker while FOXD1 p.Ala88Gly variant might have a protective effect.

2. Material and methods

2.1. Expression microarrays from uterine and placental mouse tissues

Eight 66H-MMU13 females were crossed with C57BL6/J males, following a previously described mating protocol [22]. Female mice were euthanized by cervical dislocation at E12.5, in accordance with Paris Descartes University, the Cochin Institute and the Guidelines for Biomedical Research Involving Animals policies (no. 13-020: ‘Implication of FOXD1 in ER and RSA’ reference no. 00175.01). Placentas were dissected from live embryos and total RNA extracted using Trizol. Total RNA was also extracted from uterine tissue located between contiguously implanted (normal development) embryos. A pool of four placentas and four uteri from each mouse was used. Four micrograms from each RNA pool were sent to a NimbleGen expression array platform for DNA end-labelling, hybridization, scanning and data normalization, thereby providing the final data file. The data are available from Annotare 2.0 (accession number: E-MTAB-4643).

2.2. Bioinformatics analysis of microarray data

Placenta and uterus were compared between C57BL6/J mice and 66H-MMU13 mice, using the GSEA tool (http://software.broadinstitute.org/gsea/index.jsp), in order to identify the most relevant gene sets, first against the hallmark dataset collection, and then against the c2.all.v5.1.symbol.gmt dataset. The hallmark dataset encompasses only 50 gene sets with very clear characteristics summarizing large biological functions. The other dataset encompasses 4276 gene sets and allows a much more detailed analysis of the transcriptome.

2.3. Bioinformatics and biostatistics

Statistical analysis (Student’s t-test, followed by Bonferroni correction for multiple testing) identified putative transcription factor binding sites (TFBS) using the Genomatix Gene2Promoter option. Putative TFBS in the promoters of the 50 most upregulated genes were compared with the 40 most downregulated ones for such analysis. Student’s t-test was used for assessing binding site statistical representation by comparing their occurrence in two groups of promoters for each putative binding site identified. The t-test was corrected to take multiple testing into account. The putative binding sites identified by promoter were classified according to their degree of discrepancy (in terms of frequency) between both groups of promoters. Electronic supplementary material, table S1 outlines this analysis for Forkhead Binding sites that were the most significantly
and exonuclease I. All expression constructs were sequenced to confirm the presence of the expected variants and to discard PCR-induced artefacts. These fragments were purified and cloned, using T4 DNA ligase (Invitrogen), into a previously digested PG3L3-basic luciferase vector (Promega).

The human PGF promoter amplicon consisted of 720 bp (−1144 to −424 bp upstream of the initial ATG start codon) or 1653 bp upstream of the ATG start codon for the long version of the promoter. The human C3 promoter amplicon consisted of 729 bp (−792 to −63 bp upstream of the initial ATG start codon). These fragments were amplified from genomic DNA from a control individual having non-mutant sequences and then compared to AC_000146.1 and NC_000019.8 (NCBI). Amplicons were introduced into the pCR4-topoTA cloning vector (Invitrogen) and sequenced to exclude PCR-induced mutations. The PGF and C3 promoter regions were then extracted. These fragments were purified and cloned, using T4 DNA ligase (Invitrogen), into a previously digested PG3L3-basic luciferase vector (Promega).

2.8. Cell culture and luciferase assays

The calcium phosphate method was used for co-transfecting either COS-7 or KGN cells with 400 ng of any FOXD1 construct and 570 ng of any PG3L3 plasmid containing the target genes’ promoter regions (COS-7 cells were used due to their high capacity of being transfected, and used as a classical cell model; KGN are ovary cells, whose expression profile may be closer to the reproductive expression characteristics of the female genetic tract). Experimental controls consisting of PG3L3-basic and pcDNA3.1/CT-GFP empty vectors were included for each condition. The cell medium was replaced by a fresh one 28 h after transfection. The FOXD1 mutants’ transcriptional activity was assessed 44 h after transfection using the Dual-Luciferase Reporter Assay System (Promega, Madison, WI, USA). Each experiment was performed at least twice in triplicates. The firefly activity observed for each replicate was divided by the activity recorded for the Renilla luciferase vector. Student’s t-test was used for estimating statistical significance. Transfection with empty pcDNA expression vector was used to normalize a potential effect on each promoter. STATISTIX add-on software for EXCEL was used for statistical analysis (ANOVA), followed by post hoc tests (Student–Newmann–Keuls).

2.9. Patients and controls

The RSA patient group consisted of 556 women who had enrolled in a matched case-control study of unexplained pregnancy loss in the Nimes Obstetricians and Haematologists (n = 2175 patients) [26]. The control group consisted of 271 women who had at least one live birth with no history of pregnancy loss. All these subjects (RSA patients and controls) were Caucasian.

3. Results

3.1. Transcriptomics and gene ontology analyses reveal downregulation of the coagulation cascade and the inflammatory response in placenta and uterus of 66H-MMU13 strain

To investigate the mechanism underlying the ER phenotype observed in the 66H-MMU13 strain, we undertook a comparative analysis of endometrial and placental tissues
We performed a similar GSEA analysis on genes found deregulated in uteri of 66H-MMU13 versus C57/BL6. Only one hallmark entitled ‘PANCREAS_BETA.Cells’ was found significant in the upregulated genes list (FWER = 0.008), suggesting that alterations of uterine gene expression resemble those found in the pancreas (the genes involved in the detection of this cluster were Ins1m, Chga, Isl1, Pkh1, Sst, Gpc2, Pces2, Neurod1 and Nkx6-1, which induced 5.1-, 2.3-, 2.1-, 1.7-, 1.6-, 1.5-, 1.4-, 1.3- and 1.25-fold, respectively). By contrast, six hallmarks were significant in downregulated genes, which are INFLAMMATORY_RESPONSE, INTERFERON_-GAMMA_RESPONSE, TNFA_SIGNALING_VIA_NFKB, IL2_STAT5_SIGNALING, IL6_JAK_STAT3_SIGNALING and ALLOGRAFT_REJECTION. The first of them is represented in figure 3 along with the most downregulated genes. On the whole, the uterus of the 66H-MMU13 mice was characterized by a strong downregulation of genes involved in inflammation and immunity. The analysis of the C2 gene set from GSEA confirmed a great contrast between the number of upregulated gene sets (48 with an FDR < 0.25) and downregulated gene sets (711 gene sets with a FDR < 0.25). In the downregulated clusters, a vast quantity of inflammatory/immune pathways are present (related to viral infection, IFN targets, CTLA4 pathway, Graft-versus host disease, etc.). In the upregulated groups, beta cell development is present, consistent with the hallmarks previously identified.

3.2. Promoter analysis of deregulated genes identifies Foxd1 as the most relevant candidate gene to explain 66H-MMU13 phenotype

The list of genes present in the Mus spretus fragment has previously been published [22]. Several genes in this region are putatively involved in the phenotype observed, such as F2r1l, F2r. Bioinformatics analysis of putative TFBS using Genomatix showed that transcription factors (TFs) encompassing forkhead binding sites (FKHD) in placental tissue were at the top of the list (statistical discrepancy between promoters of repressed and induced genes). More specifically, repressed genes contained twice as many FKHD than promoters of induced genes (2.54 versus 1.27, p = 0.0009; electronic supplementary material, table S1 and figure S1). The fragment of spretus chromosome 13 fixed in the 66H-MMU13 strain contains only one forkhead domain transcription factor: Foxd1. It is therefore a strong candidate to explain the gene deregulation associated with the elevated ER in 66H-MMU13 line. Because Foxd1 was not modified in the array, our hypothesis was that the spretus and musculus versions of Foxd1 had different effects on gene regulation, rather than one version being differentially expressed compared with the other.

3.3. Spretus version of Foxd1 presents sequence variants which have functional consequences on its target genes

Direct sequencing of the complete Foxd1 coding region from Mus spretus and 66H-MMU13 animals revealed five non-synonymous variants relative to the C57B6/J version: p.Asp73Glu, p.Asn126Glu, p.Thr152Ala, p.Asp76_Leu77In-sAsp and p.Pro319del. One specific variant, Foxd1-Thr152Ala, drew our attention because it is located in the protein’s
forkhead DNA-binding domain (DBD) and threonine in position 152 is strictly conserved during evolution except in *spretus* (figure 4). The SIFT software predicted a highly deleterious effect of a change of this threonine into an alanine (as found in *spretus*; score: 0.00) as this mutation triggers the disappearance of a predicted phosphorylation hot-spot. We next tested the functional effect of p.Thr152Ala variant on promoters of known Foxd1 target genes which were among the deregulated genes in microarray analyses of 66H-MMU13 placenta/uterus *Pgf* and C3.

In luciferase assays, the *spretus* version of Foxd1 (Foxd1-ms) was able to transactivate *Mus musculus* *Pgf* (Pgff-mm) promoter but to a lesser extent (approx. threefold reduction, *p*-value < 0.001) than Foxd1 *Mus musculus* version (Foxd1-mm). Similarly, Foxd1-mm transactivation of the *spretus* *Pgf* (Pgff-ms) promoter was reduced compared with its effects on Pgff-mm promoter. However, Foxd1-mm transactivation of Pgff-ms promoter was significantly higher than that of *spretus* Foxd1 version (Foxd1-ms), suggesting that Foxd1 *Mus musculus* version leads to a higher activation of Pgff promoter whether from *musculus* or *spretus* (figure 5). The Foxd1-Thr152Ala mutation alone significantly decreased Foxd1 ability to transactivate Pgff promoter. Conversely, Foxd1-ms overexpression led to higher C3-mm promoter transcription activity than that induced by the Foxd1-mm version while the Foxd1-T152A mutation alone had no significant differences regarding control (figure 5). On the whole, these data demonstrate that Foxd1 *spretus* and *musculus* versions have different effects on their target genes, in agreement with a model in which the gene deregulation and associated ER phenotype in 66H-MMU13 strain are mainly due to sequence variants in Foxd1 *spretus* version, compared with *musculus*.

### 3.4. FOXD1 sequencing in humans show polymorphisms with effects associated with adverse pregnancy outcomes

As the results we obtained in mice were encouraging, we move to human patients and sequenced FOXD1 in 556 women affected by RSA and 271 women with normal fertility, and found a total of 27 sequence variants (table 2). Nine of them were present in both RSA and control individuals, 18 (10 non-synonymous) variants were identified only in RSA patients and one was present exclusively in control individuals. Sequence variants found in RSA women concerned 33 patients (5.9%). All non-synonymous nucleotide changes exclusive to the RSA group (not present in control individuals) were identified in heterozygous state. A contingency χ²-test revealed highly significant (*p* = 0.0006) statistical comparison of the exclusive variants in the patient population to those in the control group. Relative risk calculated using all these ‘private’ mutations was 10.3 (95% confidence interval: [1.4–77.2]). The statistical significance was achieved using a single sequence variant (p.429AlaAla) in which two alanine residues were inserted at protein position 429. This mutation was found 12 times exclusively in the patient group (2.2%, *p* = 0.015). The p.Ala88Gly variant was found in 8.5% of the women from the control group (3.6% in the RSA group).

In luciferase assays, FOXD1-p.Ala356Gly and FOXD1-p.429AlaAla forms were completely unable to activate the Pgff promoter while the FOXD1-Ile364Met variant retained a transactivation capacity similar to that of the WT form. Regarding the effect on C3 promoter activity, the FOXD1-Ala356Gly versions did not yield to any activation of C3 promoter, contrary to the WT form, while the Ile364Met form induced an approximately sixfold transactivation. The FOXD1-429AlaAla variant also tended to induce the expression level (approx. threefold) of C3 promoter than the WT form (figure 6).

### 4. Discussion

Human fertility, like most biological processes in mammals, is assumed to be the result of subtle interaction of gene variants located in different genomic regions, having a quantitative effect and thus called QTL for quantitative trait loci. The expression of these gene variants, in interaction with

**Figure 2.** The major gene set of downregulated gene in the placenta is composed of genes involved in the regulation of coagulation. (a) The list of the most strongly downregulated in this pathway, from the GSEA analysis (presented in part (b) of the figure). NES reflects the Normalized Enrichment Score, indicating a very significant enrichment of genes involved in this pathway.
environmental factors, results in significant quantitative phenotypic differences between individuals (e.g. at the organ or at the molecule levels). Classical genetic approaches (e.g. genetic linkage analysis in families, direct sequencing of candidate genes) have been relatively unsuccessful for discovering genes modulating fertility in the endometrial/placental context, probably because in this case the failure results in a very complex interaction between two genetically different partners (the placenta with half of the alleles coming from the father, on the one hand, and the uterus, on the other hand), involving highly complex and strongly regulated immune tolerance mechanisms. The genetics of infertility QTL was often studied in mouse models, leading for instance to identification of regions of the X chromosome implied in male infertility in mice [34,35]. However, these approaches have generally not permitted gene identification [22]. Similarly, in species where the economic importance of controlling fertility is huge, such as in cattle, genomic regions containing genes involved in fertility have been identified as well, but rarely genes [36].

The IRCS model that we used in this study presents some advantages such as the high genetic and phenotypic variability between parental strains (Mus musculus and Mus spretus). Because the average length of the spreptus chromosomal fragments fixed in the Mus musculus genetic background is small, it allows a relatively precise identification of the QTL location, albeit the ultimate cloning of relevant genes remains very challenging [37].

A previous study showed that the 66H-MMU13 IRCS strain was affected by a high ER rate (14.7% versus 4.6% observed in C57BL/6J females, p < 0.01). This strain contains a unique 5 Mb spreptus fragment on chromosome MMU13 (between rs12069373 and D13Mit47 polymorphic markers) encompassing 31 genes [22]. In this study, the classical genetic approaches are combined with cDNA microarray assays in relevant tissues (endometrium and placenta) for evaluating potential placental and uterine transcriptional differences between 66H-MMU13 and control animals that could explain the phenotype. This hypothesis proved reliable

Table 1. Highly significant placental gene sets identified as enriched between 66H-MMU13 and C57B6/J mice. The gene sets (obtained from GSEA, http://software.broadinstitute.org/gsea/msigdb/collections.jsp) were systematically tested against the placental transcriptome comparing the two mouse lines under scrutiny. When a name starts the geneset it refers to the researcher that published the dataset. Size refers to the number of genes present in the geneset. Positive values for ES and NES (light grey) stands for gene clusters of upregulated genes, while the negative values (dark grey) account for clusters of downregulated genes. NOM p-value is the nominal (non-corrected p-value). FDR is the false discovery rate and FWER is the family-wise error rate. The RANK at MAX refers to the position of the last gene inside the up- or downregulated group, among the complete classified list of genes from the microarray experiment. For instance, this means that statistically there are 2.45-fold more genes than expected in the present study that are correlated to upregulated genes following tretinoin treatment.

<table>
<thead>
<tr>
<th>name of the geneset</th>
<th>size</th>
<th>ES</th>
<th>NES</th>
<th>NOM p-value</th>
<th>FDR q-value</th>
<th>FWER p-value</th>
<th>rank at max</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARTENS_TRETINOIN_RESPONSE_UP</td>
<td>581</td>
<td>0.46</td>
<td>2.45</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2550</td>
</tr>
<tr>
<td>MIKKELSEN_MCV6_HCP_WITH_H3K27ME3</td>
<td>357</td>
<td>0.42</td>
<td>2.29</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2228</td>
</tr>
<tr>
<td>KEGG_BASAL_CELL_CARCINOMA</td>
<td>53</td>
<td>0.55</td>
<td>2.27</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>3494</td>
</tr>
<tr>
<td>MEISSNER_BRAIN_HCP_WITH_H3K27ME3</td>
<td>225</td>
<td>0.43</td>
<td>2.22</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2220</td>
</tr>
<tr>
<td>HSIAO_LIVER_SPECIFIC_GENES</td>
<td>192</td>
<td>-0.67</td>
<td>-2.55</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1362</td>
</tr>
<tr>
<td>SU_LIVER</td>
<td>40</td>
<td>-0.79</td>
<td>-2.44</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>728</td>
</tr>
<tr>
<td>KEGG_COMPPLEMENT_AND_COAGULATION_CASCADES</td>
<td>48</td>
<td>-0.74</td>
<td>-2.40</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>892</td>
</tr>
<tr>
<td>OHGUCHI_LIVER_HNF4A_TARGETS_DN</td>
<td>111</td>
<td>-0.62</td>
<td>-2.29</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1899</td>
</tr>
<tr>
<td>REACTOME_RESPONSE_TO_ELEVATED_PLATELET_CYTOSOLIC_CA2_</td>
<td>66</td>
<td>-0.66</td>
<td>-2.25</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1230</td>
</tr>
<tr>
<td>VANDELIUS_COMM61_TARGETS_GROUP_3_UP</td>
<td>76</td>
<td>-0.65</td>
<td>-2.23</td>
<td>0.0</td>
<td>1.1 × 10^-4</td>
<td>0.001</td>
<td>1023</td>
</tr>
<tr>
<td>REACTOME_LIPIDPROTEIN_METABOLISM</td>
<td>24</td>
<td>-0.77</td>
<td>-2.18</td>
<td>0.0</td>
<td>9.9 × 10^-5</td>
<td>0.001</td>
<td>280</td>
</tr>
<tr>
<td>REACTOMEFORMATION_OF_FIBRIN_CLOT_CLOTTING_CASCADE</td>
<td>21</td>
<td>-0.79</td>
<td>-2.16</td>
<td>0.0</td>
<td>9.0 × 10^-5</td>
<td>0.001</td>
<td>892</td>
</tr>
<tr>
<td>MILLI_PSEUDOPODIA_HAPTOTAXIS_UP</td>
<td>369</td>
<td>-0.55</td>
<td>-2.16</td>
<td>0.0</td>
<td>1.6 × 10^-4</td>
<td>0.002</td>
<td>4426</td>
</tr>
<tr>
<td>SERVITIA_ISLET_HNF1A_TARGETS_DN</td>
<td>95</td>
<td>-0.61</td>
<td>-2.16</td>
<td>0.0</td>
<td>1.5 × 10^-4</td>
<td>0.002</td>
<td>643</td>
</tr>
<tr>
<td>REACTOME_CHYLOMICRON_MEDIATED_LIPID_TRANSPORT</td>
<td>15</td>
<td>-0.84</td>
<td>-2.15</td>
<td>0.0</td>
<td>2.1 × 10^-4</td>
<td>0.003</td>
<td>280</td>
</tr>
<tr>
<td>PID_INTEGRIN2_PATHWAY</td>
<td>25</td>
<td>-0.75</td>
<td>-2.14</td>
<td>0.0</td>
<td>2.6 × 10^-4</td>
<td>0.004</td>
<td>406</td>
</tr>
<tr>
<td>PID_UPA_UPAR_PATHWAY</td>
<td>39</td>
<td>-0.69</td>
<td>-2.13</td>
<td>0.0</td>
<td>2.5 × 10^-4</td>
<td>0.004</td>
<td>444</td>
</tr>
<tr>
<td>CAIRO_LIVER_DEVELOPMENT_DN</td>
<td>199</td>
<td>-0.55</td>
<td>-2.11</td>
<td>0.0</td>
<td>5.2 × 10^-4</td>
<td>0.009</td>
<td>1737</td>
</tr>
<tr>
<td>PID_HNF3B_PATHWAY</td>
<td>36</td>
<td>-0.69</td>
<td>-2.10</td>
<td>0.0</td>
<td>6.6 × 10^-4</td>
<td>0.012</td>
<td>348</td>
</tr>
<tr>
<td>KEGG_STEROID_HORMONE_BIOSYNTHESIS</td>
<td>30</td>
<td>-0.70</td>
<td>-2.09</td>
<td>0.0</td>
<td>8.4 × 10^-4</td>
<td>0.016</td>
<td>1454</td>
</tr>
<tr>
<td>LEE_LIVER_CANCER_E2F1_DN</td>
<td>53</td>
<td>-0.64</td>
<td>-2.09</td>
<td>0.0</td>
<td>9.0 × 10^-4</td>
<td>0.018</td>
<td>2318</td>
</tr>
<tr>
<td>BIOCARTA_AMI_PATHWAY</td>
<td>16</td>
<td>-0.82</td>
<td>-2.08</td>
<td>0.0</td>
<td>8.5 × 10^-4</td>
<td>0.018</td>
<td>1182</td>
</tr>
<tr>
<td>YAMASHITA_LIVER_CANCER STEM_CELL_DN</td>
<td>66</td>
<td>-0.60</td>
<td>-2.05</td>
<td>0.0</td>
<td>2.1 × 10^-3</td>
<td>0.045</td>
<td>1575</td>
</tr>
<tr>
<td>BIOCARTA_INTRINSIC_PATHWAY</td>
<td>16</td>
<td>-0.81</td>
<td>-2.04</td>
<td>0.0</td>
<td>2.0 × 10^-3</td>
<td>0.045</td>
<td>1182</td>
</tr>
</tbody>
</table>
because it has been previously shown that incompatibilities between separate chromosome regions in the IRCS model induce strong gene expression alterations, due to interspecific genomic divergence between coding (e.g. TFs) and regulatory regions (e.g. promoters). Indeed, in the IRCS mice, two genomes of two different species that diverged around 1 Ma are collapsed in a few generations, leading to a phenomenon named ‘transcriptomic shock’ [37,38].

Comparative analysis of endometrial and placental tissues’ transcriptomes between 66H-MMU13 and C57BL/6J animals showed deregulation of gene expression throughout the whole genome. Some genes show systematic opposite deregulation ('mirror effect') between placenta and uterus, suggesting an alteration of the bidirectional molecular dialogue normally occurring between these tissues. Such widespread modification in expression profiles might be due to alterations of one or several TFs located on the 66H-MMU13 strain's *spretus* fragment. Following such an assumption, a *Mus spretus*-TF might have abnormal transactivation properties on multiple *Mus musculus* target promoter sequences. Bioinformatics analysis of the promoter content in TF-binding sites in the deregulated genes led us to propose Foxd1 as a strong candidate. Direct sequencing of 66H-MMU13 and *spretus* animals demonstrated that, during evolution, five sequence variants were fixed in Foxd1 coding regions of *Mus spretus* subspecies. *In silico* analysis underlined the strong interest of one of them (Foxd1-Thr152Ala), because it affected the forkhead DBD in a highly conserved position among mammalian species and the bioinformatics tools predicted a functional effect (figure 4).

A luciferase reporter assay was then carried out to investigate target gene transactivation capacity regarding both Foxd1 and the Thr152Ala variant, using two...
genes’ promoter constructs (placenta growth factor, \textit{Pgf} and complement component, \textit{C3}). These were strongly modified in the microarrays and are known to play key roles in mammalian implantation \[26,27,39,40\]. The \textit{Pgf} transcript (\textit{Pgf} is a previously validated Foxd1 direct target gene \[28\]) was 4.3-fold less abundant in the 66H-MMU13 uterus compared with the C57BL/6 J control, while \textit{C3} was upregulated 5.7-fold in the 66H placenta compared to C57BL/6 J control.

Luciferase reporter assays using the mouse Foxd1 coding sequence and target promoters (\textit{C3} and \textit{Pgf}) showed that even though Foxd1-ms was able to stimulate \textit{Pgf}-mm (interspecific) expression, activation became strongly and significantly altered (approx. 3-fold reduction) compared with that for Foxd1-mm on the \textit{Pgf}-mm promoter (intraspecific). These results corroborated their biological pattern as observed in transcriptome assays.

Reciprocally, it was observed that Foxd1-mm transactivation capacity regarding the \textit{Pgf}-ms promoter (interspecific) became reduced compared with its effects on the \textit{Pgf}-mm promoter (intraspecific). This could have been due to natural genomic polymorphism between \textit{musculus} and \textit{spretus} genomes, observed in terms of nucleotide substitution \[38\]. Interestingly, our experiments showed that the Foxd1-mm version was

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Transactivation properties of distinct mouse Foxd1 versions on Pgf and C3 promoters. (a) Luciferase activity on the mouse Pgf promoter in two cell models (COS-7 and KGN). In black are the effects of the \textit{Mus musculus} and \textit{Mus spretus} versions of Foxd1 on the \textit{Mus musculus} promoter. In red are the effects of the same TFs on the \textit{Mus spretus} promoter. The expression profiles are similar, despite significant differences in the cell lines. Overall, the activation of the promoter appears more efficient with the \textit{Mus musculus} FoxD1. (b) The effects of specific variants of Foxd1, including the T152A mutation on the Pgf and C3 promoters. The \textit{Mus spretus} variant appears more efficient on the C3 promoter and less efficient on the \textit{Mus musculus} version of the Pgf promoter. These results are consistent with an overexpression of complement cascades (negative for implantation) and underexpression of a major actor of placental angiogenesis, Pgf. \textit{p}-values * < 0.05, ** < 0.01 and *** < 0.001.}
\end{figure}
significantly more efficient in transactivating the Pgf-ms promoter than the Foxd1-ms version itself. Hypothetically, these features could be consistent with the idea that the Mus spretus Foxd1 variants would modulate the amount of progeny to an optimal (and not maximal) number. On this point, it is worth noting that spretus litter size is naturally smaller than the musculus one (5.3 ± 1.4, versus approx. 7.6 in C57BL/6 J) [39,41]. We also tested the effect of the Foxd1-Thr152Ala mutation alone regarding Pgf-mm expression. As expected, this mutation significantly (albeit moderately) decreased Foxd1 transactivation capacity concerning the Pgf promoter, thereby contributing to downregulation of the Pgf levels observed in the 66H-MMU13 uterine tissue. Furthermore, it indicates that the other Foxd1 spretus variants also participated in this phenomenon (figure 5).

This led to investigating spretus Foxd1’s functional impact on Mus musculus complement C3, a gene playing a central role in ER, as shown by the CBA x DBA cross [42]. Foxd1 spretus overexpression induced higher C3-mm expression levels than those induced by the Foxd1-musculus version (figure 5), while there were no significant differences between the isolated Foxd1-Thr152Ala mutation and control concerning this promoter. These in vitro experiments were consistent with the existing transcriptome data in which C3 is 5.7-fold more abundant in the 66H-MMU13 placenta than in the C57BL/6 J tissue. Our experiments suggest that C3 is a direct target of Foxd1, an observation which has not been reported before.

In sum, Foxd1 seems to regulate the expression of two crucial genes implicated in pregnancy maintenance. Pgf is highly expressed in the placenta, where it regulates vascular endothelial differentiation [43]. Uterine NK cells (uNK), an endometrial lymphocyte cell subset transiently found during endometrial decidualization and essential for immune dialogue with trophoblasts, express Pgf and Vegf [40,44,45]. Pgf<sup>-/-</sup> mice display defects during early differentiation of binucleate uNK cells [37]. Concerning C3, its activation is required for antiphospholipid-induced pregnancy loss that can be reverted by administering heparin, which blocks complement cascade activation [26]. The placenta appears to be subjected to a complement-mediated immune attack at the maternal–fetal interface during normal pregnancy. An appropriate complement inhibition is required for physiological gestation and, as has been thoroughly demonstrated in mice, the deficiency of complement regulatory proteins progressively leads to embryonic lethality [46,47]. Indeed, excessive local complement C3 production may overwhelm complement regulatory mechanisms, thereby leading to pregnancy loss. These findings suggest

Table 2. FOXD1 open reading frame sequencing in RSA patients and control individuals. Mutations tested for their functional impact are in bold.

<table>
<thead>
<tr>
<th>DNA</th>
<th>protein</th>
<th>patients (n = 556)</th>
<th>controls (n = 271)</th>
<th>p-values &lt; 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.69G&gt;C</td>
<td>p.Gly23Gly</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>c.237G&gt;A</td>
<td>p.Leu79Leu</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.300C&gt;T</td>
<td>p.Ala100Ala</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.324C&gt;T</td>
<td>p.Gly108Gly</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.612G&gt;A</td>
<td>p.Glu204Glu</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>c.903C&gt;A</td>
<td>p.Ala301Ala</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.1007 C&gt;T</td>
<td>p.Ala336Val</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.1248G&gt;C</td>
<td>p.Val416Val</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.1297 GCC&gt;GG</td>
<td>p.Ala432Ala</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.1308A&gt;G</td>
<td>p.Ser436Ser</td>
<td>318</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td>c.1055 C&gt;G</td>
<td>p.Arg532Pro</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.1007 C&gt;T</td>
<td>p.Ala336Val</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.164G&gt;C</td>
<td>p.Arg55Pro</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.263G&gt;C</td>
<td>p.Ala88Gly</td>
<td>20</td>
<td>23</td>
<td>0.004</td>
</tr>
<tr>
<td>c.326_327InsGCCG</td>
<td>p.Ins109Gly</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.683C&gt;T</td>
<td>p.Pro228Leu</td>
<td>0</td>
<td>2</td>
<td>0.043</td>
</tr>
<tr>
<td>c.721G&gt;C</td>
<td>p.Ala241Pro</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.976G&gt;A</td>
<td>p.Ala326Thr</td>
<td>21</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>c.1067 C&gt;G</td>
<td>p.Ala356Gly</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.1092 C&gt;G</td>
<td>p.Ile364Met</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.909_1165del256</td>
<td>p.FS&gt;STOP376</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.1146-1160del</td>
<td>p.Gln383_Ala387del</td>
<td>27</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>c.1169_1170insGGCGG</td>
<td>p.Ins391ProPro</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>c.1187C&gt;T</td>
<td>p.Pro396Leu</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>c.1285_1286insGGCGG</td>
<td>p.Ins429AlaAla</td>
<td>12</td>
<td>0</td>
<td>0.016</td>
</tr>
<tr>
<td>c.1309G&gt;A</td>
<td>p.Val437Ile</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c.1324G&gt;T</td>
<td>p.Ala442Ser</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
that strong C3 expression disturbance (activation or inhibition) is related to embryo abortion.

Concerning FOXD1 screening for mutations in women affected by RSA mutations, our statistical results showed that women having FOXD1 mutations were at a high risk (10.3 relative risk) of suffering RSA. To the best of our knowledge, this value is the highest reported to date concerning idiopathic forms of RSA. The functional effects of three FOXD1 mutations were assessed to establish a direct functional link between them and RSA aetiology. These mutations were predicted to be damaging, because they were exclusively found in RSA women, and most were not previously described in SNP databases, while their affected residues are conserved among mammalian species. The FOXD1 wild-type (WT) and mutant versions (FOXD1-Ala356Gly, FOXD1-Ile364Met and FOXD1-429AlaAla) were tested regarding their PGF and C3 promoter transactivation properties (figure 6). FOXD1-Ala356Gly and FOXD1-429AlaAla were completely unable to activate the PGF promoter, while the A706 and 2-ALA variants trigger overexpression of C3 (significant only for A706: p.Ile364Met). The A706 mutant seems to be able to provoke an overexpression of the mouse Pgf promoter (c). Letters reflect significant differences compared with transfections with the empty expression vector. Asterisks relate to the comparison relative to the WT induction: p-values *<0.05, **<0.01 and ***<0.001, compared to WT. Error bars represent standard error.

Figure 6. Transactivation properties of human WT and mutant FOXD1 versions on human PGF and C3 promoters. (a,b) The two mutants found in RPl patients Ala356Gly pFOXD1-2ALA are unable to activate the PGF promoter efficiently, while the A706 and 2-ALA variants trigger overexpression of C3 (significant only for A706: p.Ile364Met). The A706 mutant seems to be able to provoke an overexpression of the mouse Pgf promoter (c). Letters reflect significant differences compared with transfections with the empty expression vector. Asterisks relate to the comparison relative to the WT induction: p-values *<0.05, **<0.01 and ***<0.001, compared to WT. Error bars represent standard error.
It has been shown that C3 levels in humans are higher in patients suffering a third consecutive pregnancy loss than in women having a successful pregnancy after two idiopathic abortions [48]. Moderate C3 expression levels in mice are necessary for successful gestation because C3 knock-out (KO) animals have a much higher resorption rate 15 days post coitum than their WT counterparts, due to trophoblast dysfunction and labyrinth development defects [49]. Our in vitro experiments suggested that FOXD1 mutations decreasing or increasing C3 expression are deleterious. This indicates that a subtle tuning of C3 expression levels is necessary for optimal fetal-placental interface function.

It should be noted that, up to now, severe defects have been seen in Foxd1 KO animals regarding kidney (only in homozygous state) and optic chiasm development [50, 51]. FOXD1 human mutations or the mouse Foxd1-ms functional substitution described here displayed milder functional effects than those observed in Foxd1 homozygous KO mice. This might be due to the mutation’s intrinsic nature (complete KO versus point mutations) or/and to interspecific genetic/physiological differences.

This led to hypothesizing that more drastic mutations in humans may be related to kidney and/or neurological phenotypes. However, it cannot be ruled out that FOXD1 could be implicated in syndromic forms of RSA which include these and/or other clinical features. It would be of interest to check for the presence of eye phenotype in patients suffering from RSA, which in these cases could be due to FOXD1 variants.

One concern in our work is the fact that the expression data about FOXD1 in the utero-placental unit is relatively scarce. The Protein Atlas database suggests, however, a significant expression of the protein in the human placenta (http://www.proteinatlas.org/ENSG00000251493-FOXD1/issue/placenta). The absence of reliable Foxd1 antibody in mice prevented us, despite several attempts, from observing a specific labelling, even in reference tissues such as the kidney. Furthermore, even the microarray data for this gene may sometimes be questionable due to the extreme richness in CG dinucleotides found in its open reading frame (73.8% versus 41% for the genome, with 264 CG and five NotI digestion sites (GGCGCGGC) in an Open Reading Frame of 1371 nucleotides while one is expected in more than 500,000 nucleotides on average), leading to possible problems in hybridization. It is interesting to note, however, that Sha and co-workers [52] published in 2007 a comparison of endothelial cells in the endometrium in endometriosis and control patients, which revealed a 6.1-fold increase ($p = 0.0015$) of FOXD1 mRNA level in the context of this disease, known to be a major cause of infertility. The expression levels reported in control endometrial endothelial cells from this study for FOXD1 corresponded to the expression of genes mildly expressed [52].

In conclusion, QTL positional cloning has been validated for the first time by genotyping an outbred species (human). The approach outlined here may be useful for other positional cloning projects where a limited set of recombination events is available for mapping purposes (as in IRCS, or more generally in mouse strains). Our results have also shown that FOXD1 is possibly a new molecular actor modulating pregnancy maintenance, with mutations associated with ER in mice and RSA in humans. Our findings argue in favour of FOXD1 p.Ala88Gly conferring a protective effect, because it was frequently encountered in the control group. Finally, we propose that the FOXD1 p.429AlaAla mutation can be considered as an RSA molecular marker, which can be easily tested by PCR/sequencing.

Ethics. Female mice were euthanized by cervical dislocation at E12.5, in accordance with Paris Descartes University, the Cochin Institute and the Guidelines for Biomedical Research Involving Animals policies (no. 13-020: ‘Implication of FOXD1 in embryonic resorption and recurrent spontaneous abortion’ reference no. 00175.01). The Hospital of Nîmes’ institutional review board approved all the study’s clinical steps. All individuals provided written informed consent.

Authors’ contributions. P.L., J.C. and D.V. drafted the manuscript. F.B., J.S., P.M., R.A.V. and M.F. provided very useful advice on the manuscript and helped to render it legible. P.L., B.L., M.V., E.D.S. and J.C. carried out the transfection experiments and luciferase assays. P.L. and J.C. carried out the cloning experiments. X.M. and G.B. were in charge of the mouse work with the IRCS strains; G.R. was in charge of the ultrasonography; E.M. and J.-C.G. contributed the human patient DNAs and precious data on their phenotype, as well as advices on the manuscript. C.B. and D.V. were involved in the human sequencing of FOXD1.

Competing interests. The authors declare that no competing interests exists.

Funding. This work has been partly funded by a ‘MAMMIFERT’ grant from the Agence Nationale de la Recherche (ANR) and by France Life Imaging (grant ANR-11-INBS-0006), and Région Ile-de-France. P.L. has been supported by grants from the Roche Research Foundation and the ANR.

Acknowledgements. The Cochin Institute’s expert sequencing by their sequencing platform is gratefully acknowledged. In vivo imaging was performed at the Life Imaging Facility of Paris Descartes University (Plateforme Imageries du Viant—PIV).

References


