GREB1L variants in familial and sporadic hereditary urogenital adysplasia and Mayer-Rokitansky-Kuster-Hauser syndrome

Adeline Jacquinet, Bouchra Boujemla, Corinne Fasquelle, Jerôme Thiry, Claire Josse, Aimé Lumaka, Elise Brischoux-Boucher, Christèle Dubourg, Véronique David, Laurent Pasquier, et al.

GREB1L variants in familial and sporadic hereditary urogenital adysplasia and Mayer-Rokitansky-Kuster-Hauser syndrome

Running title: *GREB1L,* urogenital adysplasia and MRKH

Authors and affiliations :

Jacquinet Adeline [1,2], Boujemla Bouchra [2], Fasquelle Corinne [1], Thiry Jerôme [2], Josse Claire [2,3], Lumaka Aimé [2], Brischoux-Boucher Elise [4], Dubourg Christèle [5,6], David Véronique [5,6], Pasquier Laurent [7], Lehman Anna [8], Morcel Karine [5], Guerrier Daniel [5], Bours Vincent [1,2].

[1] Center for Human Genetics, Centre Hospitalier Universitaire, Liège, Belgium.

[2] Human Genetic Laboratory, GIGA Institute, University of Liège, Liège, Belgium

[3] Medical Oncology, Centre Hospitalier Universitaire CHU Liege, Liège, Belgium

[4] Centre de Génétique Humaine, Université de Franche-Comté, Besançon, France.

[5] Univ. Rennes, CNRS, IGDR - UMR 6290, F-35000 Rennes, France.

[6] Univ. Rennes, CHU Rennes, Department of Molecular Genetics and Genomics, F-35000 Rennes, France.

[7] Univ. Rennes, CHU Rennes, Department of Medical Genetics, CLAD Ouest, F-35000 Rennes, France.

[8] Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; BC Children's Hospital and BC Women's Hospital, Vancouver, BC, Canada.

Corresponding author

adeline.jacquinet@chuliege.be

ORCID: 0000-0002-8342-0427

Acknowledgments

We would like to thank the families for their participation to this study and the members of both the MAIA (http://www.maia-asso.org) and Syndrome de Rokitansky-MRKH (http://asso.orpha.net/MRKH) associations for their contribution. We are also indebted to the physicians involved in the French national PRAM network (Programme de Recherches sur les Aplasies Müllériennes), the GACUA and the GARD research programs. We thank the GIGA Genomics Platform for technical assistance with NGS data generation. This research was supported by funding from the CHU Liège (FIRS), the Walloon Region (WALGEMED project), the University of Liège and the Fond Léon Frédéricq. AJ is supported by the National Fund for Scientific Research, Belgium.

Conflict of Interest statement

The authors declare that they have no conflict of interest.

Ethics approval

The GARD (genetic alterations in rare diseases) protocol was approved by the UBC Children's and Women's Research Ethics Board (Vancouver, Canada). The GACUA (Identification of genetic factors involved in uterine development by exome sequencing of individuals with hereditary or syndromic uterine and kidney malformation) protocol was approved by the ethics committee from the University Hospital and the University of Liège, Belgium. The PRAM project was approved by the local institutional review board (Comité de Protection des Personnes), and is registered with the French Ministry of Health. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Abstract

Congenital uterine anomalies (CUA) may have major impacts on the health and social well-being of affected individuals. Their expressivity is variable, with the most severe end of the spectrum being the absence of any fully or unilaterally developed uterus (aplastic uterus), which is a major feature in Mayer-Rokitansky-Kuster-Hauser syndrome. To date, etiologies of CUA remain largely unknown. As reports of familial occurrences argue for strong genetic contributors in some cases, we performed whole-exome sequencing in nine multiplex families with recurrence of uterine and kidney malformations, a condition called hereditary urogenital adysplasia. Heterozygous likely causative variants in the gene *GREB1L* were identified in four of these families, confirming *GREB1L* as an important gene for proper uterine and kidney development. The apparent mode of inheritance was autosomal dominant with incomplete penetrance. The four families included fetuses with uterovaginal aplasia and bilateral renal agenesis, highlighting the importance to investigate *GREB1L* in such phenotypes. Subsequent sequencing of the gene in a cohort of 68 individuals with Mayer-Rokitansky-Kuster-Hauser syndrome or uterine malformation (mostly sporadic cases) identified six additional variants of unknown significance. We therefore conclude that heterozygous *GREB1L* variants contribute to MRKH syndrome and this likely requires additional genetic or environmental factors for full penetrance.

Keywords: Mayer Rokitansky Kuster Hauser syndrome, Uterine Anomalies, Renal and Mullerian duct hypoplasia,Mullerian aplasia, Urogenital abnormalities, Renal Adysplasia, Hereditary renal agenesis, *GREB1L*

1. Introduction

Congenital uterine anomalies (CUA) affects 3 % of women with variable expression and severity [1]. Formation, fusion or septal absorption defects during the embryonic development result in aplastic uterus/hemi-uterus, bicorporeal uterus and septate uterus, respectively [2]. The most severe form of CUA is the congenital absence of the uterus, cervix, and the upper part of the vagina in an otherwise phenotypically normal 46, XX female, which characterizes the Mayer-Rokitansky-Kuster-Hauser syndrome (MRKH). In its typical, complete and isolated form, this anomaly is referred to as MRKH type 1. Atypical forms (incomplete or symptomatic because of functional endometrial tissues) and/or the presence of additional malformations (most frequently renal or vertebral) define MRKH type 2 [3]. MRKH and CUA are mainly sporadic, suggesting multifactorial/polygenic inheritance. Recently, models of semi-cloning technology in mice have provided first evidence for digenic/oligogenic inheritance [4]. In familial cases, strong genetic contributors might create high susceptibility with the requirement of additional genetic, epigenetic or environmental factors to explain the incomplete penetrance. To date, a few underlying genetic factors have been identified in the overall CUA. Recurrent copy number variants in 17q12, 22q11.2, 16p11.2 and 1q21.1 were implicated in both Müllerian fusion anomalies and MRKH with 7 to 14% of girls with CUA or MRKH shown to be carriers for deletion/duplication in one of these susceptibility loci [5,6]. Variants of unknown significance were reported in *LHX1, WNT9B* and *TBX6* genes in a significant but still low number of individuals [3] and association with functional polymorphisms in *TBX6* was suggested in previous studies [7,8]. CUA have also been reported in a high number of syndromes [3] suggesting that multiple developmental genes are involved in uterine development. Prime examples of syndromes exhibiting CUA as a major feature include Mullerian aplasia and hyperandrogenism (OMIM #158330) due to *WNT4* heterozygous pathologic variants and the RCAD (renal cyst and diabetes) syndrome (OMIM: #137920) caused by *HNF1B* heterozygous mutation.

Focusing on familial cases might help identifying new genes involved in uterine development. Whereas recurrence of isolated uterine malformation is rare, several families have been reported with developmental defects of the kidney and/or urinary tracts and anomalies of the Mullerian ducts [9], a condition referred to as hereditary urogenital adysplasia by Schimke and King in 1980 [10]. The close temporal and spatial proximity of Mullerian and Wolfian ducts during development in early embryonic life suggested a common pathology basis for both defects. As it is the case for familial congenital anomalies of kidney and urinary tract (CAKUT) [11], high genetic heterogeneity is likely in CUA as well and some genes have been shown to be involved in both uterine and renal malformations in humans. First reported in families with CAKUT, *GREB1L* heterozygous variants were also identified in individuals with CUA and MRKH syndrome in some affected families [12–14]. The function of this gene and its role during kidney and female genital tract development is not yet understood. *GREB1L* (for growth regulation by estrogen in breast cancer 1 like) is part of a chromatin complex with retinoic acid and steroid hormone receptors, and its role as a coactivator in retinoic acid-mediated transcription has been suggested [15]. Animal models have confirmed its causality with absence of kidneys and genital tracts demonstrated in E13.5 knockout mice embryos [12].

In order to identify new genes involved in CUA and MRKH, we performed whole exome sequencing in nine multiplex families with recurrence of uterine and kidney malformations. Herein, we describe four additional families with *GREB1L* variants. Subsequently, we sequenced the gene in a separate cohort of 68 individuals with MRKH or CUA. We confirm *GREB1L* as being a frequent genetic cause for hereditary urogenital adysplasia and a significant contributor in MRKH syndrome, more precisely in MRKH type 2 subgroup.

2. Materials and Methods

2.1 Subjects

DNA from 39 individuals from nine families was collected after written informed consent

through three different research study protocols: the GARD (genetic alterations in rare diseases) protocol, in the Department of Medical Genetics, Children's and Women's Health Center of British Columbia, Vancouver, Canada; the French national PRAM network (Programme de recherches sur les aplasies Mullériennes); the GACUA (genetic alterations in congenital uterine anomalies) protocol, CHU of Liège, Belgium. Affected relatives presented congenital uterine anomalies and/or kidney malformations. When possible, DNA from unaffected relatives was also collected. Clinical data for each affected relative are described in supplemental table 1.

DNA extracted from saliva samples, tissue samples (from autopsy) or blood samples was available for at least two affected relatives per family. Normal chromosomal microarray performed at least in one affected family member (Agilent SNP array 180k (Amadid 029830)) was a prerequisite.

Pedigrees are described in figure 1 (Families 5,6,7 and 9) and in supplemental figure 1 (Families 1 to 4 and 8).

DNA from 63 additional individuals with MRKH syndrome or uterovaginal aplasia and five with a hemi-uterus or bicorporeal uterus was collected through the PRAM network and the GACUA protocol (Supplemental data table 2). The cohort included 3 fetal samples. DNA was extracted from blood samples or, in case of fetuses, from tissue samples. 7/68 individuals had a family history of one relative with kidney or uterine malformation (DNA was not available for the relatives). Clinical data for the cohort are summarized in table 1.

2.2 WES

After library preparation using the Ion AmpliSeq Exome Kit (Family 1 and 2), the Agilent SureSelect V4 kit (Family 3 and 4) and the Agilent SureSelect Human All Exon V6 capture technology (Family 5 to 9), whole exome sequencing was performed on Ion Proton System (Family 1 and 2), Nextseq500 (Family 7) or HiSeq 2000/2500 Illumina sequencers (Family 3

to 6, 8 and 9), generating 150bp reads (paired-end with Illumina platforms) that were aligned to the hg19 reference Human Genome. The average exome coverage was above 50X. Variants were annotated and filtered using Alissa Interpret v2.1 (Agilent Technologies). Based on the hypothesis of an autosomal dominant inheritance with incomplete penetrance, at first, only affected relatives were sequenced (2 or 3 for each family). Shared heterozygous variants were filtered based on a minor allele frequency <0.1% in the population database Gnomad[16], their presence in exonic regions +/-20bp of exon boundaries and a predicted change at the protein level (synonymous variants excluded). The functional impact of single nucleotide variants was evaluated using several in silico prediction tools (PolyPhen-2, SIFT, PROVEAN, Mutation Taster, Mutation Assessor for missense variants; Human splicing Finder and Alamut Visual splicing Predictions tools to assess splicing effect) (Supplemental data table 3). For each variant of interest, we investigated the CADD scaled score, an integrative score based on the output of multiple in silico programs and genomic annotations that rank the variant deleteriousness relative to all single nucleotide variants in the genome (a score between 20 and 30 meaning that the variant is estimated to be in the 1% of the most deleterious variants in the genome)[17] We considered a cut-off of 25 for the CADD scaled score as indicative of a possible deleterious effect[18]. The GERP++RS score, estimating the evolutionary constraint for individual alignment site, was used to assess the level of conservation of modified nucleotides. A positive score ≥ 2 (with a maximum of 6.18) indicates evolutionary constraint and conservation[19]. Conservation of the aminoacid was evaluated through the multiple alignment track for sequences of 100 vertebrates on UCSC genome browser (Multiz Alignments of 100 Vertebrates) [20]. Finally, we looked at the missense substitution analysis algorithm Align GVGD, which assigns an ordered grade ranging from Class C0 (most likely neutral) to Class C65 (most likely deleterious) based on biophysical distance between aminoacids and evolutionary aminoacid conservation. We considered C45 as a cut-off for a likely deleterious effect [21]. Clinical interpretation of the variants was based on ACMG guidelines [22] For Family 9, recessive (homozygous or

compound heterozygous) and *de novo* inheritance with parental mosaicism were also evaluated.

Variants were confirmed by Sanger sequencing and were investigated by Sanger sequencing in additional relatives with available DNA.

2.3 cDNA synthesis

RNA was extracted from blood collected on a Paxgene tube for individual II:2 in Family 9 and an unrelated control. cDNA was synthesized (Thermo ScientificTM RevertAidTM First Strand cDNA Synthesis Kit) and then used as template for amplification. PCR was conducted using the following primers for *GREB1L* (NM_001142966.2): Forward (in exon 22)-5'- ACCACGCTGACTATAGCAACCAG -3', Reverse (at the junction between exon 23 and 24)-5'- CACTGCTTTCATTGTGATCGGT -3'. PCR products for the patient and the control was run on a 2,5% agarose gel electrophoresis. The bands were then extracted and Sanger sequenced.

2.4 Targeted *GREB1L* sequencing

Regions of interest were captured using the single molecule Molecular Inversion Probes (smMIPs) technology [23]. The smMIPS (EasySeqNGS-customized Targeted Capture Kit, Nimagen) were designed to target the exons +/- 20 nucleotides of intronic/exonic boundaries and the 5'UTR region of *GREB1L* (NM_001142966.2). Massive parallel sequencing on a Nextseq 500 Illumina sequencer generated 150bp paired-end reads. More than 95% of nucleotides were covered at least 40X and more than 91% of the targeted region, including the entirety of the exons 2 to 33 (with exception for 56 nucleotides within the exon 22) were covered more than 100X. Exon 1 (non-coding exon) had a lower depth of coverage. The gap in exon 22 (Hg19, chr18:g.19079859 to chr18:g.19079915) was filled in with Sanger sequencing (primers for *GREB1L* (NM_001142966.2): Forward-5'-TGGAGAATGGAGTGAGCTCTTCCA-3', Reverse-5'- TACCTGTGGGGGCCCTGTCA-3').

Single nucleotide variations were interrogated using the Seqnext (SeqPilot) ® software. A lower threshold of 2% was used for variant allele detection. All rare variants identified by targeted sequencing were confirmed by Sanger sequencing.

3. Results

In four families, heterozygous deleterious variants were identified in the gene *GREB1L (Genebank:* NM_001142966.2) (Figure 1, figure 2 and table 2).

In Family 5, WES performed in individuals III:5 and II:1 identified a novel variant consisting in a two base pair deletion (c.2787_2788del) leading to a frameshift (p.(Asp930Profs*12)). Individual III:5 was the mother of a female fetus with bilateral renal agenesis (BRA) and uterovaginal aplasia (DNA not available for this fetus). She was not known to have any kidney or uterine anomalies. Individual II:1, the maternal uncle of III:5, had unilateral renal agenesis (URA). The daughter of II:1 (individual III:2) presented URA and she had a male fetus (IV:1) with URA and contralateral multicystic kidney disease (MCKD). Presence of the variant in individual III:2 and fetus IV:1 was confirmed by Sanger sequencing.

In Family 6, WES identified a novel variant, a one base pair deletion (c.2227del) leading to a frameshift and a premature stop codon (p.(Gln743Argfs*10)) in a female fetus (IV:1) with uterovaginal aplasia and bilateral congenital renal anomalies (URA and contralateral MCKD), and individual III:5 (fetus (IV:1)'s mother cousin) affected by MRKH type 2, URA and mild scoliosis. The presence of the frameshift variant was confirmed by Sanger sequencing in III:2 and III:4, respectively the mother and maternal aunt of fetus IV:1. Individual III:2 was unaffected (renal ultrasound was normal, no uterine malformation reported) whereas individual III:4 had URA and MRKH type 2. Besides the *GREB1L* variant, an heterozygous missense variant (c.2846T>C, p.(Ile949Thr)) was identified in the gene *ROBO2* (NM_001290040.1) in individuals III:2,III:4,III:5 and IV:1. This variant was previously described in one family with CAKUT (vesicoureteral reflux and hypoplastic kidneys) [24]. The

variant affects a highly conserved nucleotide (GERP++RS Score: 6.16) and a moderately conserved aminoacid. The variant is present in a very low frequency in the population database Gnomad (MAF:0.00001219). However, although a contribution of this variant to the phenotype in Family 6 cannot be ruled out, there is no existing functional data for this *ROBO2* variant. In addition, *in silico* predictions (PolyPhen-2, PROVEAN, SIFT, Mutation assessor, Align GVGD and the CADD score) are rather suggestive of a benign/non deleterious effect.

In Family 7, a novel heterozygous missense variant (c.5198A>G, p.(Asn1733Ser)) was identified by WES in a female fetus (IV:1) with uterovaginal aplasia, BRA, streak gonads, ureter and bladder aplasia, as well as in the sister (IV:2) and in the father (III:1) both presenting URA. Familial analyses by Sanger sequencing identified the variant in one unaffected paternal aunt (individual III:4) (normal renal ultrasound) and the paternal grandmother (II:2) (not known to be affected). The niece of II:2 (individual III:6) had URA and a hemi-uterus but she was not available for testing. The identified missense variant was absent from the population database Gnomad and was predicted to be deleterious or likely deleterious by five *in silico* programs (PolyPhen-2, PROVEAN, SIFT, Mutation Taster and Mutation assessor). Additional in silico evidence for deleteriousness were the CADD score at 26.1 an the Align-GVGD class at C45. The variant affects a highly conserved nucleotide (GERP++RS score of 5.89) and a highly conserved amino acid, from zebrafish to human.

In Family 9, an intronic variant (c.3970-20A>G), was identified in two fetuses with partial uterovaginal aplasia, bilateral renal agenesis, ureter and bladder aplasia (III:1 and III:2) as well as in their asymptomatic father (II:2) (normal renal ultrasound). In addition, the fetus III:2 presented unilateral hexadactyly. This variant was located 20 nucleotides before the boundary between intron 22 and exon 23 and was predicted to create a new acceptor splice site (Human splicing finder, Alamut Visual splicing Predictions tools).

Amplification of the fragment between exon 22 and exon 24 was performed on cDNA from the father (II:2) and an unrelated control. Variability in amplification of the targeted region was observed, likely due to very low level of *GREB1L* expression in blood (according to GTEx Portal, an open access database on tissue-specific gene expression[25]). For the control, sequencing of the PCR products after extraction of the electrophoretic band confirmed the normal sequence. For the case, aberrant electrophoretic bands were variably detected (Figure 3). Sequencing of the PCR products showed the expected normal sequence as well as an alternative sequence with 19 additional nucleotides between exon 22 and 23 (Figure 3), leading to a frameshit and a premature stop codon (p.(Val1324Leufs*34)). Overlap of the normal and alternative cDNA sequences was observed for the third electrophoretic band. Low *GREB1L* expression in blood as well as random and variable amplification of each DNA strand (the wild type strand and the strand with the variant) might explain why we did not consistently observe the alternative transcript. Moreover, the alternative splicing might not be complete or lead to an unstable mRNA, thus reducing the level of aberrant mRNA in the blood.

Finally, based on the hypothesis that most severely affected individuals would have inherited additional genetic factors from their second parent, the exome data for fetuses in Families 7 and 9 were filtered looking for additional rare variants, i.e. in Family 7, present in the fetus IV:1 but absent in the father III:1 and the sister IV:2, and in Family 9, present in both fetuses III:1 and III:2, but absent in the father II:2. We focused on genes previously involved in CAKUT or MRKH, in the GREB1/ER interactome or in the Wnt pathway. These variants are described in supplemental table 4. No likely pathogenic or pathogenic variant could be identified.

Targeted sequencing of *GREB1L* in a cohort of 68 unrelated individuals with uterine malformations (including 3 fetuses) presented on Table 1, led to identify rare variants of

unknown significance (four missense and one synonymous) in three individuals with MRKH type 2 (U113, U125, U147), one fetus with uterovaginal aplasia and bilateral kidney malformation (U042), and one individual with MRKH type 1 (U098) (Table 2). For most individuals, parental DNA was not available to check for *de novo* occurrence. The variant identified in fetus U042 was inherited from the asymptomatic father (normal renal ultrasound). Finally, the variant (c.277G>A, p.(Glu93Lys)) (rs185578147) was identified in two additional unrelated individuals (U120 and U149) as well as in Family 9 (II:2; III:1;III:2). Considering the entire cohort (9 familial cases and 68 sporadic cases (3/77)), the allele frequency for this variant was 0.019 (3/154) which was 3.85 times higher than the minor allele frequency reported in Gnomad (MAF: 0.004939 in European (non-Finnish) population). At present, the association between this variant and the phenotype cannot be asserted. Finally, we did not identify any rare mosaic variant considering a minimum allele frequency of 2%.

4. Discussion

GREB1L disease-causing variants have been previously reported in families and sporadic cases of congenital anomalies of kidneys and optionally of the uterus [11–15,26,27]. Haploinsufficiency is the suggested mechanism for pathogenicity [15] with ten loss-of-function variants out of the 39 variants reported thus far. CUA were noted in ten of these 41 families and sporadic cases [12–14]. We describe four additional familial cases and three sporadic cases with uterovaginal aplasia or MRKH syndrome type 2 and convincing *GREB1L* variants.

Amongst the previously reported *GREB1L* families and herein, formation defects were the most frequently observed developmental defect of the uterus, with aplastic uterus or MRKH type 2 being reported in 15 individuals with *GREB1L* convincing variants and hemi-uterus in three. Fusion and septal absorption defects were uncommon with bicorporeal and arcuate uterus each described in 1 individual (Table 3). Additional anomalies of the female genital tract and ovaries included in some cases agenesis of the Fallopian tubes and ovaries, streak

gonads, hemi-vagina and absence of the uterine left artery (Table 3). The CAKUT phenotypes were most of the time renal agenesis, renal dysplasia/hypoplasia, vesicoureteral reflux and MCKD (Table 3). Bladder/ureter agenesis/hypoplasia were described in six fetuses with bilateral renal agenesis. Megaureter, duplication of the ureters and ectopic kidney were each reported in two cases. In individuals with MRKH or uterovaginal aplasia, renal agenesis (uni or bilateral) or MCKD were consistently associated whereas with the other types of uterine anomalies, kidney anomalies were not always present. Conversely, uterine anomalies were not always noted in females in case of renal anomalies.

In our cohort, we also identified one novel missense variant (c.1936T>C, p.(Cys646Arg)) , not reported in Gnomad, in one individual with MRKH type 1 (U098). However, conflicting *in silico* predictions make this variant less likely to be causative, most in silico tools (PolyPhen-2, PROVEAN, SIFT, Mutation assessor, CADD, Align GVGD) supporting a neutral/non deleterious effect, whereas only Mutation Taster and Human Splicing Finder states for a possible damaging effect (Table 2). If *GREB1L* seems to contribute to isolated CUA in some individuals, sequencing in additional cohorts of MRKH individuals with accurate clinical examination would be necessary to determine its possible involvement in MRKH type 1.

Variability in the phenotype among our patients across or within families could not be solely explained by the type or the location of *GREB1L* variants. These were scattered throughout the gene and included loss-of-function as well as deleterious missense mutations in both individuals with isolated renal anomalies or renal and uterine anomalies, with no hotspot for uterine malformation (Figure 2). The incomplete penetrance and intrafamilial variability observed in our study, and previously reported [20,21], may result either from stochastic differences during development, the influence of modifier genes or additional variants in *cis* or *trans*, variation in allele dosage, epigenetic factors or environmental exposures.

Maternal bias in transmission was previously suggested as it was observed that the mother was the transmitting parent in most families [12,13]. In our four families, we did not confirm

parental bias and could not explain non-penetrance by the gender of the transmitting parent. Amongst the affected individual carriers of *GREB1L* variants for whom the status of the parents was known (by genotyping or because they were obligate carriers), five individuals inherited the variant from their mother and five from their father. The severity of the phenotype was not correlated with the gender of the transmitting parent. Indeed, for the five affected fetuses with BRA (or URA+MCKD) +/- uterovaginal aplasia, two inherited the variant from their mother and three from their father. Finally, three unaffected females inherited the variant from their mother. In families 5, 6 and 7, a three-generation pedigree indicated a more severe phenotype for fetuses at the third generation, as also previously reported in a single family [13]. This is likely due to ascertainment bias rather than anticipation as families are sooner referred to geneticists in case of severe and lethal congenital anomalies. Also, pregnancy losses in prior generation may not have been assigned to bilateral renal agenesis.

In *GREB1L* families and sporadic cases with congenital anomalies of kidneys and uterus, additional clinical features were sometimes present, expanding the clinical spectrum. Rib anomalies, heart defects, anomalies of the adrenal glands, polydactyly and ear anomalies were reported in more than one individual, and some other only once (Table 3). Moreover, in a study on non-syndromic inner ear malformation, two *de novo* loss-of-function *GREB1L* variants were identified in two male individuals with sensorineural hearing loss, absent cochleae and 8th cranial nerve malformations without any known kidney anomaly [29]. Abnormal sensory epithelial innervation was demonstrated in zebrafish for the *greb1l* homozygous mutant sa16374 [29]. As both variants were loss-of-function, haploinsufficiency was also suggested as the mechanism of pathogenicity. Interestingly, hearing loss is reported in 10-25% of individuals with MRKH syndrome [30]. In this series, no anomalies of the auditory system were reported in individuals with *GREB1L* variant, but hearing tests were not systematically performed.

Amongst the nine families investigated herein, plausible *GREB1L* variants were detected in 4/9 (44%), and the gene is currently the most frequent one associated with familial congenital

uterine malformations. *GREB1L* sequencing in all reported CUA/MRKH families would help to better define the frequency in families with recurrence. All our families with *GREB1L* variants included individuals with MRKH type 2/uterovaginal aplasia and fetuses with bilateral renal anomalies (renal agenesis or MCKD). The diagnostic yield for *GREB1L* mutation was previously shown to be around 0.86% in isolated CAKUT (2/232) [11], 2.8% in renal agenesis or renal hypodysplasia [14] and much higher (25.8%) in fetuses with BRA [12]. The presence of uterine malformation or MRKH might thus increase even more significantly the chance to find *GREB1L* mutation. We therefore suggest investigating this gene in all families with uterovaginal aplasia or MRKH syndrome, with renal agenesis.

This study is the first to interrogate the presence of *GREB1L* mutations in sporadic MRKH/uterovaginal aplasia. In our cohort of 63 individuals, *GREB1L* heterozygous variants were identified in 7.9% (5/63) with most convincing variants in three (4.8%), all with a MRKH type 2/non isolated uterovaginal aplasia phenotype. This suggests that *GREB1L* might also be a significant genetic contributor in sporadic MRKH/uterovaginal aplasia, at least when renal anomalies are present (MRKH type 2). The presence of somatic mosaicism to explain sporadic cases was not supported by our results. Analysis of affected tissues (i.e. rudimentary uterine tissue) would, though, be more appropriate to investigate this hypothesis. In addition, the apparent phenotypic continuum of CUA and MRKH syndrome associated with various malformations raises the possibility of a digenic/oligogenic origin, which could help understand the complexity of the disease and the still unexplained cases.

References

1. Chan YY, Jayaprakasan K, Zamora J, Thornton JG, Raine-Fenning N, Coomarasamy A. The prevalence of congenital uterine anomalies in unselected and high-risk populations: a systematic review. *Hum Reprod Update*. 2011;17(6):761-771. doi:10.1093/humupd/dmr028

2. Grimbizis GF, Gordts S, Di Spiezio Sardo A, et al. The ESHRE-ESGE consensus on the classification of female genital tract congenital anomalies. *Gynecol Surg*. 2013;10(3):199-212. doi:10.1007/s10397-013-0800-x

3. Jacquinet A, Millar D, Lehman A. Etiologies of uterine malformations. *Am J Med Genet A*. 2016;170(8):2141-2172. doi:10.1002/ajmg.a.37775

4. Wang L, Zhang Y, Fu X, et al. Joint utilization of genetic analysis and semi-cloning technology reveals a digenic etiology of Müllerian anomalies. *Cell Res*. October 2019. doi:10.1038/s41422-019-0243-7

5. Ledig S, Tewes AC, Hucke J, et al. Array-comparative genomic hybridization analysis in patients with Müllerian fusion anomalies. *Clin Genet*. 2018;93(3):640-646. doi:10.1111/cge.13160

6. Nik-Zainal S, Strick R, Storer M, et al. High incidence of recurrent copy number variants in patients with isolated and syndromic Müllerian aplasia. *J Med Genet*. 2011;48(3):197-204. doi:10.1136/jmg.2010.082412

7. Sandbacka M, Laivuori H, Freitas É, et al. TBX6, LHX1 and copy number variations in the complex genetics of Müllerian aplasia. *Orphanet J Rare Dis*. 2013;8:125. doi:10.1186/1750-1172-8-125

8. Tewes A-C, Rall KK, Römer T, et al. Variations in RBM8A and TBX6 are associated with disorders of the müllerian ducts. *Fertil Steril*. 2015;103(5):1313-1318. doi:10.1016/j.fertnstert.2015.02.014

9. Herlin M, Højland AT, Petersen MB. Familial occurrence of Mayer-Rokitansky-Küster-Hauser syndrome: a case report and review of the literature. *Am J Med Genet A*. 2014;164A(9):2276-2286. doi:10.1002/ajmg.a.36652

10. Schimke RN, King CR. Hereditary urogenital adysplasia. *Clin Genet*. 1980;18(6):417-420.

11. van der Ven AT, Connaughton DM, Ityel H, et al. Whole-Exome Sequencing Identifies Causative Mutations in Families with Congenital Anomalies of the Kidney and Urinary Tract. *J Am Soc Nephrol JASN*. 2018;29(9):2348-2361. doi:10.1681/ASN.2017121265

12. De Tomasi L, David P, Humbert C, et al. Mutations in GREB1L Cause Bilateral Kidney Agenesis in Humans and Mice. *Am J Hum Genet*. 2017;101(5):803-814. doi:10.1016/j.ajhg.2017.09.026

13. Herlin MK, Le VQ, Højland AT, et al. Whole-exome sequencing identifies a GREB1L variant in a three-generation family with Müllerian and renal agenesis: a novel candidate gene in Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome. A case report. *Hum Reprod Oxf Engl*. August 2019. doi:10.1093/humrep/dez126

14. Sanna-Cherchi S, Khan K, Westland R, et al. Exome-wide Association Study Identifies GREB1L Mutations in Congenital Kidney Malformations. *Am J Hum Genet*. 2017;101(6):1034. doi:10.1016/j.ajhg.2017.11.003

15. Brophy PD, Rasmussen M, Parida M, et al. A Gene Implicated in Activation of Retinoic Acid Receptor Targets Is a Novel Renal Agenesis Gene in Humans. *Genetics*. 2017;207(1):215-228. doi:10.1534/genetics.117.1125

16. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv*. April 2020:531210. doi:10.1101/531210

17. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894. doi:10.1093/nar/gky1016

18. van der Velde KJ, de Boer EN, van Diemen CC, et al. GAVIN: Gene-Aware Variant INterpretation for medical sequencing. *Genome Biol*. 2017;18(1):6. doi:10.1186/s13059-016-1141-7

19. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLOS Comput Biol*. 2010;6(12):e1001025. doi:10.1371/journal.pcbi.1001025

20. Blanchette M, Kent WJ, Riemer C, et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*. 2004;14(4):708-715. doi:10.1101/gr.1933104

21. Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452

BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet*. 2006;43(4):295-305. doi:10.1136/jmg.2005.033878

22. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med Off J Am Coll Med Genet*. 2015;17(5):405-424. doi:10.1038/gim.2015.30

23. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res*. 2013;23(5):843-854. doi:10.1101/gr.147686.112

24. Lu W, van Eerde AM, Fan X, et al. Disruption of ROBO2 is associated with urinary tract anomalies and confers risk of vesicoureteral reflux. *Am J Hum Genet*. 2007;80(4):616-632. doi:10.1086/512735

25. GTEx Portal. https://gtexportal.org/home/gene/GREB1L. Accessed April 9, 2020.

26. Boissel S, Fallet-Bianco C, Chitayat D, et al. Genomic study of severe fetal anomalies and discovery of GREB1L mutations in renal agenesis. *Genet Med Off J Am Coll Med Genet*. October 2017. doi:10.1038/gim.2017.173

27. Rasmussen M, Sunde L, Nielsen ML, et al. Targeted gene sequencing and whole-exome sequencing in autopsied fetuses with prenatally diagnosed kidney anomalies. *Clin Genet*. 2018;93(4):860-869. doi:10.1111/cge.13185

28. Pallotta R, Bucci I, Celentano C, Liberati M, Bellati U. The "skipped generation" phenomenon in a family with renal agenesis. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol*. 2004;24(5):586-587. doi:10.1002/uog.1111

29. Schrauwen I, Kari E, Mattox J, et al. De novo variants in GREB1L are associated with non-syndromic inner ear malformations and deafness. *Hum Genet*. 2018;137(6-7):459-470. doi:10.1007/s00439-018-1898-8

30. Fontana L., Gentilin B., Fedele L., Gervasini C., Miozzo M. Genetics of Mayer-Rokitansky-Küster-Hauser (MRKH) Syndrome. Clinical genetics. 2017; 91(2):233-246. doi: 10.1111/cge.12883

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Figure legends

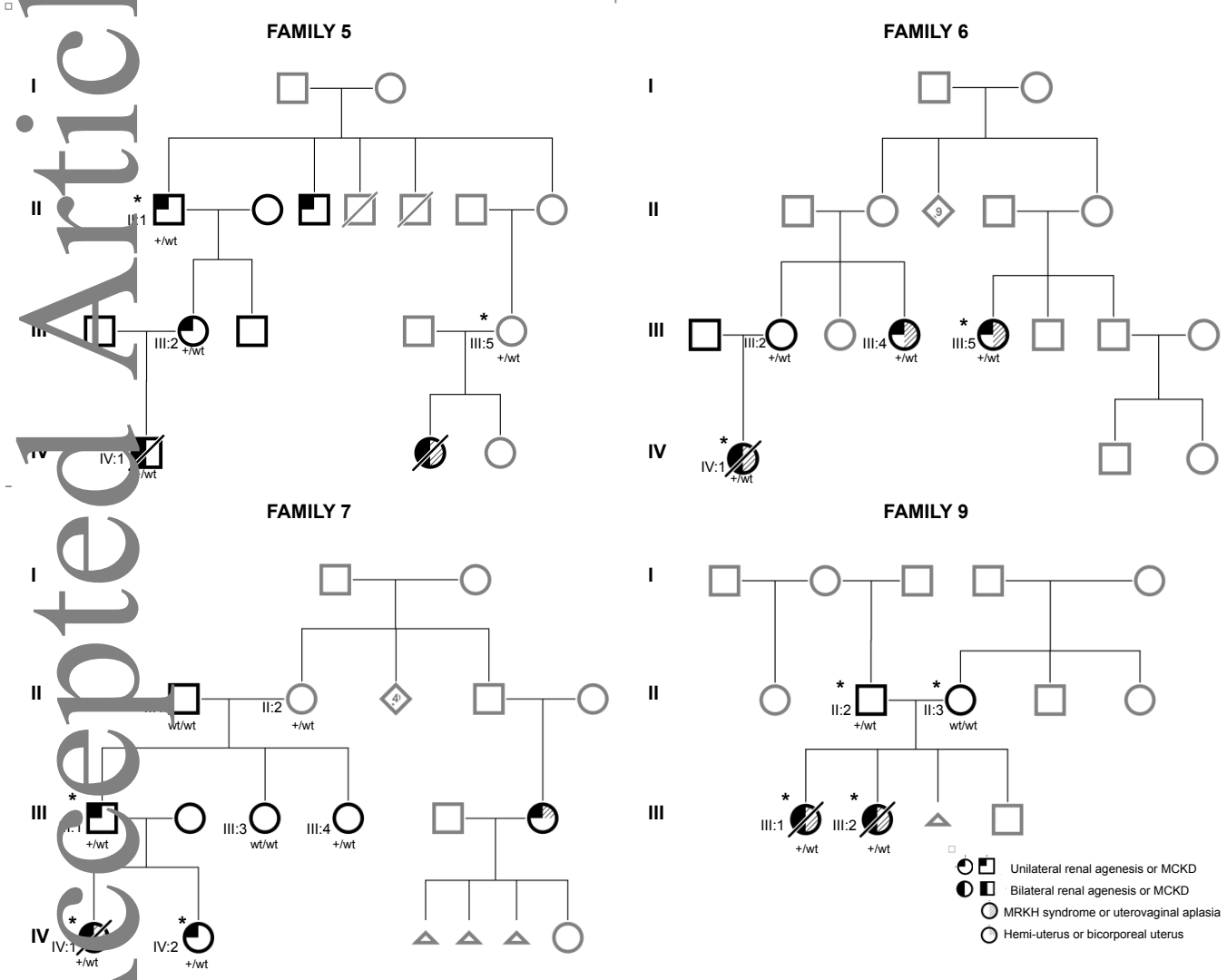
Fig. 1 Pedigree of the four families with heterozygous *GREB1L* variants. Male and female symbols are black when the phenotype was known for the individual (renal US performed, no known uterine malformation). Male and female symbols are grey when the individual's phenotype was unknown. Symbol numbers indicates individuals with available DNA. An asterisk indicates the relatives in whom WES was performed. Sanger sequencing results are indicated for the individuals who were tested as follow: +/wt: heterozygous variant present ; wt/wt: wildtype. Pedigrees were generated using PedigreeXP (http://www.pedigreexp.com).

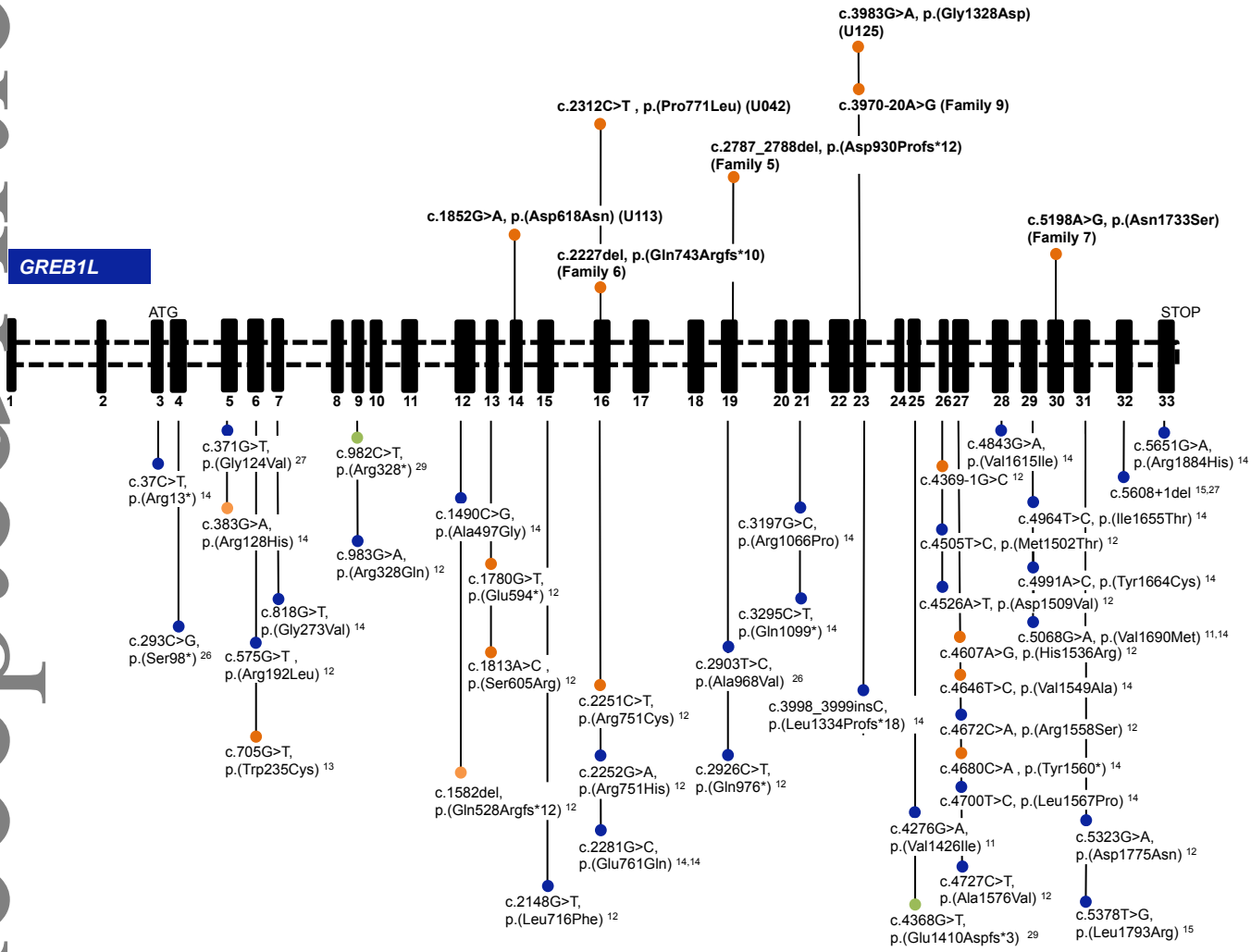
Fig.2 Variants in *GREB1L* (NM_001142966.2) reported in the literature. On the top, variants from this publication and underneath, the variants previously reported. Variants associated with congenital uterine (+/- renal) anomalies are pointed in orange. Variants associated with only congenital renal anomalies are pointed in blue. Variants associated with non-syndromic inner ear malformation are pointed in green.

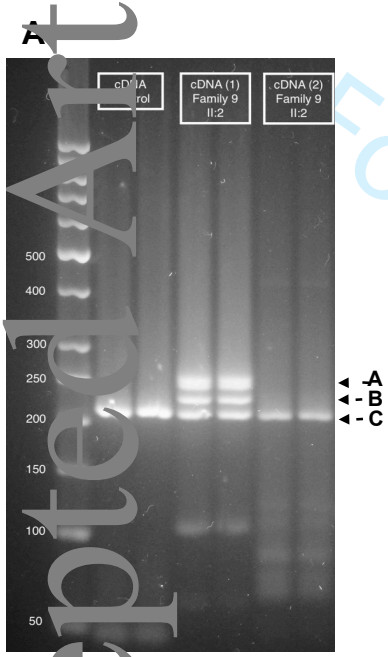
Fig. 3 Results of agarose gel electrophoresis of cDNA (Family 9 II:2) *GREB1L* PCR products (A) and sequence analysis (B) for the novel potential splice-site mutation c.3970-20A>G. Three bands (around 250bp (band A), at 229bp (band B) and at 210bp (band C)) were discernible on agarose gel electrophoresis. Sequence analyses demonstrated that band C was the sequence from the wildtype transcript and that the alternative sequence (band B) resulted from insertion of the last 19bp of intron 22. Sequencing of band A showed an overlap between both wildtype and alternative sequences.

Supplemental Fig.1 Pedigree of the five families without any identified variants in *GREB1L*. Male and female symbols are black when the phenotype was known for the individual (renal

US performed, no known uterine malformation). Male and female symbols are grey when the individual's phenotype was unknown. An asterisk indicates the relatives in whom WES was performed. Symbol numbers indicates individuals with available DNA. Pedigrees were generated using PedigreeXP (http://www.pedigreexp.com).







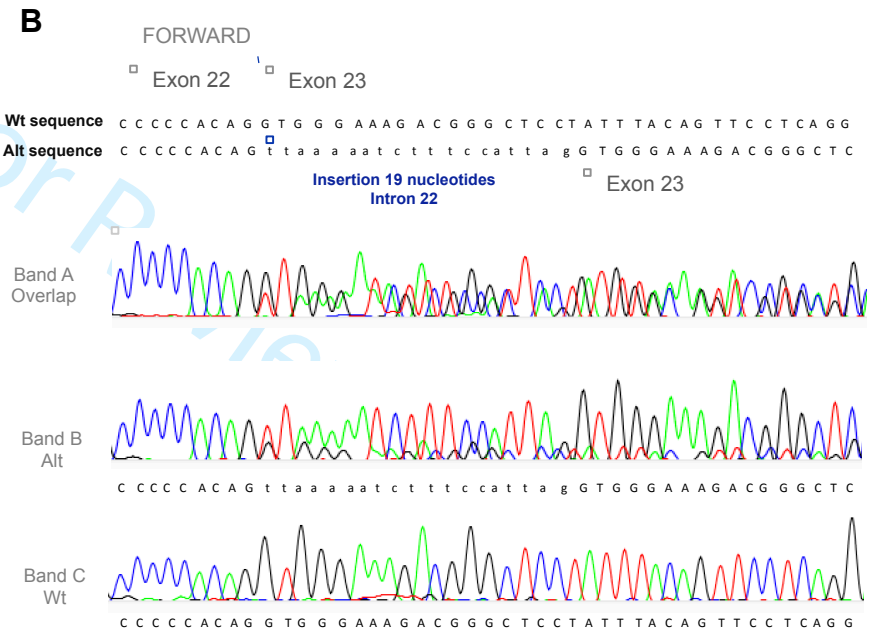


Table 1 : Clinical data for the cohort of sporadic cases

Clinical data	**N=68**
Uterine phenotype	
Adult cases	65
MRKH type 1	24
MRKH type 2	33
Other	
Uterovaginal aplasia and abnormal development of secondary sexual characteristics or streak ovaries or gonadal dysgenesis	4
Hemi-uterus	2
Bicorporeal uterus	2
Fetuses	3
Uterovaginal aplasia (non isolated)	2
Bicorporeal uterus	1
Associated kidney malformation	**29/68**
Family history of renal/uterine malformations	**7/68**

Table 2. Information on *GREB1L* variants identified in familial and sporadic cases.

Identification	Phenotype	NGS	Nucleotide change	Aminoacid change	State	Conservation GERP++RS score	PP2 score SIFT MT	Align GVGD	Scaled CADD score	Splicing in silico prediction (HSF)	Gnomad MAF (Hm/Hz/total)	ACMG classification
Variants of Clinical Interest												
Family 5		WES	c.2787_2788del	p.(Asp930Profs*12)	Hz	5.71	-/-/D.C	-	-	-	NR	Pathogenic
IV :1#	URA and MCKD											
III :2	URA											
II :1	URA											
III :5	Asymptomatic (fetus with uterovaginal aplasia and BRA)											
Family 6		WES	c.2227del	p.(Gln743Argfs*10)	Hz	6.08	-/-/D.C	-	-	-	NR	Pathogenic
IV :1#	URA and MCKD											
III :5	MRKH type 2,URA,mild scoliosis											
III :4	MRKH type 2,URA											
Family 7		WES	c.5198A>G	p.(Asn1733Ser)	Hz	5.89	0.994/D./D.C.	C45	26.1	No significant alteration	NR	Uncertain significance
III :1	URA											
IV :2	URA											
IV :1#	Uterovaginal aplasia,BRA,streak gonads, ureter and bladder aplasia											
Family 9		WES	c.3970-20A>G	p.(Val1324Leufs*34)	Hz	0 (na)	-/-/Pm	-	0.490	New acceptor site	NR	Likely pathogenic
III :1#	Uterovaginal aplasia,BRA Ureter and bladder aplasia											
III :2#	Uterovaginal aplasia,BRA Ureter and bladder aplasia Unilateral hexadactyly											
II :2	Asymptomatic											
U113	Ectopic kidney VUR Duplicated ureter MRKH type 2 Unilateral polydactyly Facial asymetry	smMIPS	c.1852G>A	p.(Asp618Asn)	Hz	5.68	0.998/T/D.C.	C0	25.1	No significant alteration	0.00002108 (0/4/189764)	Uncertain significance
U042#	Uterovaginal aplasia Streak ovaries URA-MCKD 11 pairs of ribs Cervical hemivertebrae	smMIPS	c.2312C>T	p.(Pro771Leu)	Hz	6.08	0.999/D/D.C.	C65	31	No significant alteration	0.00003187 (0/1/31378)	Uncertain significance

Identification	Phenotype	NGS	Nucleotide change	Aminoacid change	State	Conservation GERP++RS score	PP2 score SIFT MT	Align GVGD	Scaled CADD score	Splicing in silico prediction (HSF)	Gnomad MAF (Hm/Hz/total)	ACMG
Variants of Clinical Interest (continued)												
U125	MRKH type 2 URA	smMIPS	c.3983G>A	p.(Gly1328Asp)	Hz	4.68	1/D/D.C.	C65	27,4	No significant alteration	NR	Uncertain significance
Other variants												
U098	MRKH type 1	smMIPS	c.1936T>C	p.(Cys646Arg)	Hz	4.67	0.001/T/D.C	C0	20.9	Alt. exonic ESE site	NR	Uncertain significance
U147	MRKH type 2 Ectopic kidney	smMIPS	c.3492G>T	p.(Gly1164=)	Hz	0 (na)	-/T/Pm	-	0.028	Alt. exonic ESE site	NR	Uncertain significance
U120	MRKH type 1	smMIPS	c.277G>A	p.(Glu93Lys)	Hz	5.35	0.291/D/D.C.	C0	24.6	Alt. exonic ESE site	0.002813 (1/532/189090)	Uncertain significance
U149	MRKH type 2 Mild caliectasis Vertebral anomalies Imperforate anus	smMIPS	c.277G>A	p.(Glu93Lys)	Hz	5.35	0.291/D/D.C.	C0	24.6	Alt. exonic ESE site	0.002813 (1/532/189090)	Uncertain significance
Family 9 (II :2 ;III :1# ; III:2#)	Phenotype described above	smMIPS	c.277G>A	p.(Glu93Lys)	Hz	5.35	0.291/D/D.C.	C0	24.6	Alt. exonic ESE site	0.002813 (1/532/189090)	Uncertain significance

Splicing prediction: at least 3 algorithms for ESE(Exonic Splicing Enhancer)/ESS (Exonic Splicing Silencer)alteration on Human Splicing Finder ; at least 2 algorithms for new donor/acceptor site. URA : unilateral renal agenesis, BRA : bilateral renal agenesis, MCKD : multicystic kidney disease, MRKH : Mayer-Rokitansky-Kuster-Hauser syndrome ; na : data not available, NR :not reported,#fetus ; D :damaging, T :tolerated, D.C. : disease causing, Pm : polymorphism ; PolyPhen-2, SIFT, Mutation Taster, CADD, Align GVGD (v2007), and the GERP++ RS score were used for prediction.

Table 3. Phenotypic features in heterozygous carriers of convincing *GREB1L* variants

Phenotype	**Our study** 20 individuals (from 7 families)	**Literature** 96 individuals (from 54 families)	**Total** 116 individuals (from 61 families)
Renal			
Unilateral renal agenesis	10/20	28/96[11–15,27]	**38/116**
Bilateral renal agenesis	3/20	27/96[12–15,26,27]	**30/116**
Multicystic kidney disease	3/20	7/96[12–14]	**10/116**
Vesicoureteral reflux	1/20	7/96[11,12,14]	**8/116**
Renal hypoplasia/dysplasia		7/96[11,12,14]	**7/116**
Bladder hypo/aplasia	3/20	3/96[15,27]	**6/116**
Ectopic kidney	1/20	1/96[12]	**2/116**
Duplication of the ureter	1/20	1/96[14]	**2/116**
Megaureter		2/96[14]	**2/116**
Megaurethra		1/96[12]	**1/116**
Horseshoe kidney		1/96[12]	**1/116**
Multilocular cyst		1/96[12]	**1/116**
Clear cell renal carcinoma		1/96[13]	**1/116**
Genital			
Uterus/uterovaginal aplasia	4/20	5/96[12,14]	**9/116**
MRKH type 2	4/20	2/96[13]	**6/116**
Hemi-uterus		3/96[12,14]	**3/116**
Streak gonads	2/20		**2/116**
Unique follapian tube and/or ovary		2/96[12,14]	**2/116**
Uterus anomaly (not described)		1/96[14]	**1/116**
Blind ending hemi-vagina and bicornuated uterus		1/96[12]	**1/116**
Arcuate uterus		1/96[13]	**1/116**
Absence of uterine left artery		1/96[12]	**1/116**
Ovarian hernia		1/96[12]	**1/116**
Undifferentiated external genitalia		1/96[15,27]	**1/116**
Heart			
Thickened left ventricular wall		2/96[12]	**2/116**
Aortic stenosis		1/96[12]	**1/116**
Retro-esophageal subclavian artery		1/96[12]	**1/116**
Adrenal			
Adrenal gland hypoplasia		1/96[12]	**1/116**
Adrenal cytomegaly		1/96[12]	**1/116**
Liver			
Hepatic portal fibrosis		1/96[12]	**1/116**
Skeletal anomalies			
Ribs anomalies	1/20	3/96[12]	**4/116**
Cervical hemivertebrae	1/20		**1/116**
Mild scoliosis	1/20		**1/116**
Anomalies of the extremities			
Polydactyly	2/20		**2/116**
Clinodactyly		2/96[12]	**2/116**

Single transverse palmar crease		1/96[11]	**1/116**
Brachydactyly		1/96[11]	**1/116**
Genu valga		1/96[14]	**1/116**
Flat feet		1/96[14]	**1/116**
Face			
Facial dysmorphism		1/96[11]	**1/116**
Facial asymetry	1/20		**1/116**
Short neck		1/96[11]	**1/116**
Ear			
Neurosensorial hypoacusia		3/96[14,29]	**3/116**
Auricular tag		2/96[12]	**2/116**
Congenital anomalies of the inner ears		2/96[29]	**2/116**
Absent 8th nerve		2/96[29]	**2/116**
Lop ear		1/96[12]	**1/116**
Eye			
Iris anomaly		1/96[12]	**1/116**
Other			
Enlarged thymus		1/96[12]	**1/116**
Henoch Schonlein Purpura		1/96[14]	**1/116**
Diabetes		1/96[12]	**1/116**
Supernumerary nipple		1/96[11]	**1/116**
No phenotype			
Asymptomatic (Normal renal US, no known uterine anomaly)	4/20	9/96[12,13,15,27]	**13/116**
Not ascertained by renal US	1/20	7/96[12,14]	**8/116**