HHS Public Access

Author manuscript

Hum Genet. Author manuscript; available in PMC 2022 June 21.

Published in final edited form as:

Hum Genet. 2021 April ; 140(4): 667–680. doi:10.1007/s00439-020-02239-y.

Genetics of agenesis/hypoplasia of the uterus and vagina: narrowing down the number of candidate genes for Mayer-Rokitansky-Küster-Hauser Syndrome

Sasha Mikhael[1],

Sonal Dugar[1],

Madison Morton[1],

Lynn P. Chorich[1],

Kerlene Berwick Tam[1],

Amy C. Lossie[2],

Hyung-Goo Kim[3],

James Knight[4,5],

Hugh S. Taylor[6],

Souhrid Mukherjee[7],

John A. Capra[7],

John A. Phillips III[8],

Michael Friez[9],

Lawrence C. Layman[1,10]

[1] Section of Reproductive Endocrinology, Infertility, and Genetics, Department of Obstetrics and Gynecology, Neuroscience Program, Medical College of Georgia, Augusta University, Augusta, GA 30912, USA

[2] Beautiful You MRKH Foundation, Inc., 13301 Clifton Rd, Silver Spring, MD 20904, USA

[3] Neurological Disorders Research Center, Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Doha, Qatar

[4] Department of Genetics, Yale University School of Medicine, New Haven, CT, USA

[5] Yale Center for Genome Analysis, Yale University, New Haven, CT, USA

[6] Department of Obstetrics and Gynecology, Yale University School of Medicine, New Haven, CT, USA

[7] Department of Biological Sciences, Center for Structural Biology, Vanderbilt University, Nashville, TN 37232, USA

[8] Division of Medical Genetics and Genomic Medicine, Department of Pediatrics, Vanderbilt University Medical Center, 1211 Medical Center Drive, Nashville, TN 37232, USA

[9] Greenwood Genetic Center, 106 Gregor Mendel Circle, Greenwood, SC 29646, USA

✉Sasha Mikhael smikhael@augusta.edu; Lawrence C. Layman lalayman@augusta.edu.

[10] Department of Neuroscience and Regenerative Medicine, Department of Physiology, Medical College of Georgia, Augusta University, Augusta, GA 30912, USA

Abstract

Purpose—Mayer–Rokitansky–Küster–Hauser (MRKH) syndrome consists of congenital absence of the uterus and vagina and is often associated with renal, skeletal, cardiac, and auditory defects. The genetic basis is largely unknown except for rare variants in several genes. Many candidate genes have been suggested by mouse models and human studies. The purpose of this study was to narrow down the number of candidate genes.

Methods—Whole exome sequencing was performed on 111 unrelated individuals with MRKH; variant analysis focused on 72 genes suggested by mouse models, human studies of physiological candidates, or located near translocation breakpoints in t(3;16). Candidate variants (CV) predicted to be deleterious were confirmed by Sanger sequencing.

Results—Sanger sequencing verified 54 heterozygous CV from genes identified through mouse (13 CV in 6 genes), human (22 CV in seven genes), and translocation breakpoint (19 CV in 11 genes) studies. Twelve patients had ≥ 2 CVs, including four patients with two variants in the same gene. One likely digenic combination of *LAMC1* and *MMP14* was identified.

Conclusion—We narrowed 72 candidate genes to 10 genes that appear more likely implicated. These candidate genes will require further investigation to elucidate their role in the development of MRKH.

Introduction

Human sex development begins similarly in male and female embryos. In males, *SRY* on the short (p) arm of the Y chromosome initiates a cascade of events resulting in the formation of the testes, where Sertoli cells produce Anti-Müllerian hormone (AMH), which removes the Müllerian system and facilitates differentiation of Wolffian tissues into the male reproductive tract. In females, who lack this AMH surge, the Wolffian ducts regress, and the Müllerian tissues develop into the uterus, cervix, upper portion of the vagina, and fallopian tubes. Mayer–Rokitansky–Küster–Hauser (MRKH) syndrome, a.k.a. Müllerian aplasia, is the most severe Müllerian uterovaginal anomaly.

People with MRKH have a 46,XX karyotype and functioning ovaries, resulting in typical female secondary sex characteristics with normal breast development. However, their Müllerian-derived tissues are missing or underdeveloped. MRKH affects approximately 1/5000 females and constitutes the second most common cause of primary amenorrhea (Reindollar et al. 1981). MRKH can be classified as type I (isolated Müllerian defect; MIM 277,000) or type II (with an associated anomaly; MIM 601,076). These anomalies most commonly include renal dysgenesis, skeletal anomalies, cardiac defects, and deafness (Morcel et al. 2007). A subset of type II MRKH consists of Müllerian abnormalities with associated renal and cervicothoracic defects known as MURCS (Morcel et al. 2007; Oppelt et al. 2006).

Embryonic development of the urogenital tract in mouse and human proceeds in a cranio-caudal fashion. Paired Müllerian ducts emerge from invaginations of the coelomic epithelium of the intermediate mesoderm at ~ 7 embryonic weeks (E11.75 in mice) (Masse et al. 2009; Hashimoto 2003). The ducts progress caudally and fuse at 8 embryonic weeks (E13.5 in mice), giving rise to the uterus, fallopian tubes, and upper 2/3 of the vagina. The mature metanephric kidney originates from the caudal mesoderm and is initially connected to the Wolffian duct. This will later become the ureter at the bladder insertion. The close interaction of the Wolffian and Müllerian ducts with renal development helps to explain why abnormal development in one system is often associated with defects in the other. Ultimately, failed development of the Müllerian ducts is believed to be responsible for MRKH.

Emerging evidence indicates that MRKH has a genetic component. Families with > 1 member with MRKH have been reported. (Herlin et al. 2014) Additionally, first-degree relatives can display MRKH-associated anomalies, but without Müllerian anomalies (Herlin et al. 2014; Williams et al. 2017). To date, two genes are supported by in vitro analyses and genetic studies—*WNT4* (Biason-Lauber et al. 2004) and *HNF1B.* (Lindner et al. 1999) A subset of people with MRKH type I or II have copy-number variants (CNV) that most commonly involve 17q12,16p 11, and 22q11. (Layman 2014) However, these genomic regions contain large numbers of genes, so determining the exact causative gene(s) has been difficult.

Candidate genes from animal models and preliminary human findings indicate that at least 23 genes are critical to Müllerian duct formation in mice, including *Wnt4*, *Pbx2*, *Lhx1*, and *Emx2* for the upper, and *Wnt9b*, *Pax2*, and *Wnt5a* for the caudal portion of the Müllerian system. (Masse et al. 2009; Kurita 2011) Other genes important in murine Müllerian development include: *Rarg*, *Rxr*, *P63*, *Wnt7a*, *Lamc1*, and *Hoxa9-13* (Masse et al. 2009; Kurita 2011) (complete list in Table 1A). These genes represent reasonable candidate genes for study in humans with MRKH.

Human studies of physiological candidate genes for MRKH include: *WT1*, *CFTR*, *WNT7A*, *GALT*, *HOXA7*, *PBX1*, *HOXA13*, *PAX2*, *HOXA10*, *AMH*, *AMHR*, *RARG*, *RXRA*, *CTNNB1*, *LAMC1*, *DLGH1*, and *SHOX* (references in Layman (Layman 2014)), but the small sample sizes with no proven causal variants' cloud interpretation (42 genes are listed in Table 1B, 14 of which are common to mouse). Heterozygous variants in *LHX1*, (Ledig et al. 2012) *TBX6*, (Sandbacka et al. 2013) *WNT9B*, (Waschk et al. 2016) and *RBM8A* have been identified, but no functional analyses or sequencing of other family members was reported. In contrast to *WNT4* (MIM 158,330) and *HNF1B* (MIM 137,920), Online Mendelian Inheritance of Man (OMIM) has not included these four genes with MRKH phenotypes. Although all of the listed human genes are attractive candidates, they can only be considered presumptive at this time.

Balanced chromosomal translocations have provided clues for the pathogenesis of many Mendelian disorders, and several have been reported in patients with MRKH. (Williams et al. 2016) However, only one translocation in a proband with MRKH has been mapped utilizing molecular techniques. (Williams et al. 2016) Although the t(3;16) (p22.3;p13.3)

translocation did not disrupt a gene, nearby genes (n = 21) on either side of der(3) and der(16) chromosomes remain potential candidates for MRKH Table 1C.

The purpose of this study was to analyze candidate variants (CV) from whole exome sequencing (WES) of 111 patients with MRKH. We focused on variants found in 72 candidate genes from three different perspectives: 1) genes involved in Müllerian development in mouse; 2) presumptive human genes requiring confirmation, and 3) genes near the breakpoints of both derivative chromosomes in our t(3;16) translocation. The primary objective was to reduce the large number of possible CV associated with MRKH to a more manageable number for future study.

Methods

Ethics statement

This study was approved by the Augusta University Institutional Review Board (HAC_0904264) and each patient signed a consent form.

Patient samples

We recruited 111 unrelated participants with type I (n = 82) or II (n = 29) MRKH. MRKH was defined in a person with a 46,XX karyotype who had normal breast development and an absent vagina with an absent or hypoplastic uterus on physical exam, supported by imaging (ultrasound and/or an MRI) or surgery.

Whole exome sequencing

De-identified DNA samples were subjected to whole exome sequencing (WES). DNA (2–3 μg/subject) was analyzed at the Yale Center for Genome Analysis for WES as previously described (Theisen et al. 2019). DNA was sheared to create double-stranded fragments ~ 220 bp in length by focused acoustic energy (Covaris E220; Woburn, Massachusetts). Blunt ends of the fragments were created followed by phosphorylation with T4 DNA polymerase and T4 polynucleotide kinase. Adapters were added to the fragments using T4 DNA ligase prior to amplification.

Following polymerase chain reaction (PCR) amplification, custom biotinylated oligonucleotides (IDT xGen Exome Panel; Coralville, Iowa) were synthesized and hybridized to genomic DNA at 65 °C for 16 h following PCR amplification. The captured fragments were then purified with AMPure XP beads. The Illumina NovaSeq 6000 S4 platform was used for exome sequencing to create 100 bp reads. The reads were aligned using BWA MEM. GATK best practices were used to call variants, and variants were annotated using Annovar and Variant Effect Predictor.

Selection of candidate genes

A literature review of the genetics of MRKH was performed to obtain a list of potential candidate genes for analysis. Key words such as “genetics” or “genetic mutation” and, “MRKH” or “Mayer–Rokitansky–Kuster–Hauser syndrome” or “mullerian aplasia” or “congenital absence of the uterus and vagina” were used in search engines including

PubMed and Google Scholar. Additional studies were obtained from literature citations and required an accurate diagnosis of MRKH (confirmed by a 46,XX karyotype and physical exam with ultrasonography, magnetic resonance imaging, laparoscopy, or any combination of the three). Following next generation sequencing, Sanger sequencing confirmation was required. Loss of function mouse models, usually knockout, were also included. Case reports were excluded, as well as studies solely investigating large copy-number variants (CNV).

Filtering variants

For this study, we analyzed CV in known or putative candidate genes ($n = 72$) based on evidence in mouse ($n = 23$), genes studied in humans ($n = 42$; 14 of which overlapped with mouse), and genes near breakpoints of derivative chromosomes 3 and 16 ($n = 21$) (Table 1). Variants were filtered based on: (1) allelic frequency < 0.01 in the gnomAD and 1000 genome databases, (2) a Combined Annotation-Dependent Depletion (CADD) score ≥ 20, predicted to be in the top 1% likely deleterious variants including frameshift, splice site, nonsense, and missense variants (Richards et al. 2015; Kircher et al. 2014).

Confirmation of selected variants

Variants identified by WES were confirmed by Sanger sequencing (Table 2a–c). For each variant, PCR primers were designed using Primer3 (v. 0.4.0). PCR conditions consisted of 30 cycles of 95^0C for 60 s, 55 °C for 30–45 s, 72 °C for 30–45 s, and completed by a 7-min extension at 72 °C. PCR products were confirmed by agarose gel electrophoresis, and then subjected to ethanol precipitation and DNA sequencing reactions. After purification, the sequencing reactions were analyzed using the ABI 310 Sequencer or the SeqStudio Sequencer to validate CV and confirmed in both forward and reverse directions.

RT-PCR

Candidate genes near breakpoints of der(3) and der(16) were studied by reverse transcription PCR (RT-PCR) with One-Step Invitrogen RT-PCR using human RNA to determine expression in three MRKH-dependent tissues (kidney, heart, and uterus). Expression was confirmed by agarose gel electrophoresis for the presence or absence of the gene of interest for each cell type.

DiGePred

Two or more CV were identified in some patients, and these were analyzed by DiGePred. DiGePred is a random forest machine-learning classifier for identifying candidate digenic disease gene pairs using features derived from biological networks, genomics, evolutionary history, and functional annotations (Mukherjee et al. 2020). The DiGePred classifier was trained using DIDA, the largest available database of known digenic disease causing gene pairs (PMID: 26,481,352), and several sets of non-digenic gene pairs, including pairs derived from unaffected relatives of Undiagnosed Diseases Network (UDN) patients. We applied DiGePred to all human gene pairs.

Results

We conducted WES on DNA from 111 people with MRKH and detected a total of 205,056 variants. After initial filtering for CADD score, allelic frequency, and variant type (frameshift, splice site, nonsense and missense), 218 variants remained. We narrowed these further to 63 CV and verified 54 by Sanger sequencing Table 2. These 54 CV are included in our results.

Genes relevant to mouse Müllerian development

At least 23 genes are important for Müllerian development in mouse. We identified 13 heterozygous missense CV, which were predicted to be deleterious by a CADD score ≥ 20, in six human orthologues, including *HOXA10*, *LAMC1*, *PAX2*, *RARA*, *WNT4*, and *WNT9B* (Table 2A). We also found three previously unreported heterozygous missense CV in *WNT4*, a known pathogenic gene for MRKH. In addition, four unique *LAMC1* missense CV, three RARA missense CV, and one each in *PAX2*, *HOXA10* and *WNT9B* were identified. None of these genes, with the exception of *WNT4*, have an MIM number in OMIM, suggesting no causative variants identified to date. The three *RARA* variants were found in two subjects with MRKH including one individual who had two *RARA* variants (Table 3). Fourteen genes were common to mouse and human, but arbitrarily analyzed with the mouse data.

Candidate genes incompletely studied in humans

There are 42 human genes implicated in MRKH, 14 of which have mouse models. We identified 22 heterozygous CV in seven of these genes (Table 2B), including *AMH* (four missense), *CFTR* (one frameshift, one nonsense, and four missense), *TBX6* (three missense), *LRP10* (four missense), *MMP14* (two missense), *SHOX* (two missense), and *GALT* (one missense). Although some genes have MIM numbers associated with other disorders (*AMH*, *CFTR*, *GALT*, *TBX6*, and *SHOX*), none are related to MRKH in OMIM, and two genes (*LRP10* and *MMP14*) have no MIM number. CFTR variants are known to cause cystic fibrosis (CF), and four of six variants are known to cause CF or CF-associated pulmonary and/or pancreatic phenotypes (Chang et al. 2015; Girodon et al. 1997; Chillon et al. 1994; Iso et al. 2019; Lee et al. 2003). One *CFTR* variant (p.E831X) was associated with congenital bilateral absence of the vas deferens (CBAVD) and another (p.R31C) was identified in males with azoospermia and oligospermia without CBAVD (Hinzpeter et al. 2010; Gallati et al. 2009). Finally, one of the CF-causing variants (p.R74W) was part of a triple-mutant allele reported to cause CBAVD (Claustres et al. 2004).

Candidate genes near translocation breakpoints for t(3;16)

Of 21 genes located near der(3) and der(16) chromosomes, 11 genes had 19 heterozygous missense CV (Table 2C). Most frequently observed was *CNOT1* ($n = 4$), *MEFV* ($n = 3$), and *OR1F1* ($n = 2$), while the others had either one or two CV. *CNOT1*, *CRTAP*, *MEFV*, and *TRIM71* are associated with other disorders, while *OR1F1*, *ZNF200*, *ZNF205*, *ZNF263*, and *ZSCAN10* had no associated MIM number. *TMPPE* was not listed in OMIM, but was in NCBI. Thirteen genes were shown to be expressed in ≥ 1 MRKH-related tissue (Supplemental Table 1).

Patients with variants in two or more genes

Twelve patients with MRKH had CV in ≥ 2 genes or had two CV in the same gene (Table 3). Three patients had three CV in either two or three candidate genes. The significance of these findings is unclear, but it is interesting that more than one CV in the same gene in *RARA*, *TBX6*, and *LRP10* were identified. In one individual with three CV, two were in MEFV. We then applied DiGePred to all human gene pairs (Mukherjee et al. 2020). The *LAMC1* and *MMP14* gene pair with variants in one individual was predicted to have higher digenic potential than 99.5% of human gene pairs based on their proximity in pathway, interaction, and co-expression networks (Mukherjee et al. 2020).

Discussion

MRKH consists of a constellation of phenotypes, and genetic studies indicate that the genetics of MRKH will be similarly complex. Herlin et al. (2014) collected 67 families with MRKH from the literature, and showed that about half had more than one family member with MRKH and half consisted of one proband and one relative with MRKH-associated anomalies. We reported 147 individuals with MRKH, unselected for family history, and found no cases with any affected family members (Williams et al. 2017). However, 8/58 (14%) of North American probands had a relative with an MRKH-related nonreproductive phenotype, while 0/41 Turkish probands reported having any family member with MRKH or MRKH-related anomalies. This is much more complicated than other reproductive disorders such as hypogonadotropic hypogonadism (where > 50 genes contain variants in 50% of patients) (Cangiano et al. 2020) or hypergonadotropic hypogonadism (> 20 genes are involved in 25%) (Yatsenko and Rajkovic 2019). Uterine hypoplasia/aplasia complicates genetic analyses, as studies of vertical transmission are limited to people who have used assisted reproductive technologies. We also realize that MRKH could have an environmental, epigenetic, or somatic cell etiology in some patients. (Ma et al. 2011).

In this WES study of 111 unrelated people with MRKH, we analyzed 72 candidate genes ascertained from mouse studies, human analyses, and genes near the translocation breakpoints of der(3) and der(16). Of 63 CV predicted to be deleterious in these 72 candidates, 54 of 63 (86%) CV detected by WES were confirmed by Sanger sequencing. From 23 putative mouse genes, 13 variants in six genes (*HOXA10*, *LAMC1*, *PAX2*, *RARA*, *WNT4*, and *WNT9B*) were identified—all of which had CADD scores > 20 Table 2A. Our analysis for orthologues of mouse genes in people with MRKH revealed four missense CV in *LAMC1*, as well as three missense CV in *RARA* and *WNT4.* Previously reported *WNT4* variants are causative for MRKH based on functional analysis (Biason-Lauber et al. 2004). Our three confirmed CV p.R339W, p.P277T, and p.R247C in *WNT4* have not yet been described and are likely to be within the top 1% of deleterious variants, as reflected by CADD scores > 20. These promising *WNT4* variants will require in vitro analysis. Even if these variants demonstrate functional effects in vitro, they are predicted to have a low prevalence (2–3%) among people with MRKH.

The multiple CV present in *LAMC1*, encoding γ-laminin, make it a very attractive MRKH candidate (Table 2A). *LAMC1* encodes an extracellular matrix glycoprotein, which is a major component of the basement membrane. Mouse knockout (KO) of *Lamc1*, generated

by removing the nidogen-binding domain (located in domain III), resulted in uterine aplasia and renal agenesis, both phenotypic features of MRKH (Willem et al. 2002). Likely pathogenic variants have not yet been demonstrated in this gene. Sanger sequencing of the entire *LAMC1* gene was performed in 12 people with MRKH by Ravel et al., who found four synonymous and five non-synonymous missense variants, which were likely polymorphisms (Ravel et al. 2012). Interestingly, two of our four verified *LAMC1* missense CV were located in exons 16 and 17, which could potentially impact the conformation of the protein and thus affect the nidogen-binding domain (Lossl et al. 2014). Mouse models implicate *HOXA10*, *PAX2*, and *WNT9B* genes in MRKH, however, in our large cohort, only one CV in each gene was detected. *RARA* CV are also attractive candidates, since gene disruption studies in mice impair Müllerian development (Masse et al. 2009); in humans, somatic translocations involving *RARA* have been found only in acute promyelocytic leukemia (MIM 612,376).

Although 42 human genes have been studied in MRKH, only *WNT4* and *HNF1B* are known to be causative by in vitro analysis (Layman 2014). From 111 people with MRKH, we verified 22 variants (1 frameshift, 1 nonsense, and 20 missense) in 7 genes including *AMH*, *CFTR*, *GALT*, *LRP10*, *MMP14*, *SHOX*, and *TBX6. AMH* variants have been associated with persistent Müllerian structures in males, (Lang-Muritano et al. 2001) but a role in MRKH is unclear. Small studies of *AMH* in humans have not yielded causative variants (Layman 2014). *AMH* variants would have to be activating in a 46,XX female to result in MRKH.

CFTR mutations cause CF (MIM 219,700) and CBAVD (MIM 277,180), while *GALT* mutations result in galactosemia (MIM 230,400), all of which are autosomal recessive conditions. Both genes have been suggested to be involved in MRKH (Layman 2014). To date, no variants in the *CFTR* or *GALT* genes have been associated with MRKH (Layman 2014; Timmreck et al. 2003). *CFTR* was originally investigated due to CFTR's involvement in CBAVD, since, in theory, the same genes could be involved in both Müllerian and Wolffian duct development (Timmreck et al. 2003). We found six *CFTR* CV in our patients, all located in different exons. Variants p.R31C, p.E217C, and p.A120T are established causative mutations for CF, so are unlikely to be involved in MRKH (Chang et al. 2015; Chillon et al. 1994; Iso et al. 2019; Lee et al. 2003; Gallati et al. 2009). Two of the six CV are associated with CBAVD. Missense variant p.R74W is part of a triple-mutant allele that includes p.V201M and p.D1270N; however, as an isolated variant, it does not appear to cause any CF-related phenotype (Claustres et al. 2004). Additionally, our p.E831X CV is associated with CF, but also CBAVD in compound heterozygotes containing p.D110H, c.1545_1546delTA, and p.F508del mutations (Hinzpeter et al. 2010). However, the unaffected mother and sister are heterozygous for the p.E831X variant, indicating heterozygosity for a *CFTR* allele is not likely involved in MRKH. It is possible that variants leading to CBAVD could contribute to MRKH in females; however, they would not be expected to be seen with the pulmonary–pancreatic phenotype of CF. One novel *CFTR* frameshift mutation p.R1301fs was verified and is likely deleterious, serving as a good candidate for functional analysis. Since we only found heterozygous variants, it is currently unclear even if the variants impair function in vitro, whether a heterozygous variant could be involved in the pathogenesis of MRKH.

Biallelic *GALT* mutations cause galactosemia. Previous associations between the p.N314D allele in the *GALT* gene and congenital absence of the vagina and uterus have been suggested as a possible etiology for the phenotype (Klipstein et al. 2003; Cramer et al. 1987). We identified a single missense CV in the *GALT* gene absent in the gnomAD database that was predicted to be deleterious (Table 2B). This interesting *GALT* variant has not been associated with galactosemia, and requires further study. Females with *GALT* variants are usually infertile due to ovarian insufficiency, but these are biallelic variants (Kaufman et al. 1979). Although possible, it is unknown whether heterozygous *GALT* variants could be associated with MRKH.

LRP10 and *MMP14* have been implicated in MRKH in two previous studies (Backhouse et al. 2019; Rall et al. 2015). *LRP10* plays a role in the Wnt/B-catenin pathway during Müllerian development; however, causation in humans is still lacking (Backhouse et al. 2019; Rall et al. 2015). Rall et al. (2015) found that one set of twins from five twin pairs studied was discordant for a duplication containing both *LRP10* and *MMP14* in rudimentary Müllerian tissue, but not saliva, suggestive of somatic mosaicism (Rall et al. 2015). No differences were found in the other four twin pairs. In our current study, we also detected two *LRP10* missense and two *MMP14* missense variants in five individuals, but none had variants in both genes. One person possessed two *LRP10* CV which could suggest an autosomal recessive form. The role of these genes remains unknown, since *LRP10* does not have a MIM number, and *MMP14* is associated with Winchester syndrome (MIM 277,950), which is characterized by severe osteolysis in the hands and feet, generalized osteoporosis, and bone thinning.

SHOX on chromosome Xp22 is associated with Langer mesomelic dysplasia (MIM 249,700), Leri–Weill dyschondrosteosis (MIM 127,300), and familial idiopathic short stature (MIM 300,582), and likely plays a role in Turner syndrome. In a study by Gervasini et al., two sisters with MRKH shared a common intragenic *SHOX* duplication, also present in their father, suggesting sex specific X-linked dominant inheritance (Gervasini et al. 2010). This region was hypothesized to be a hotspot for genomic rearrangements, as MRKH genetic studies identified 290–300 kb duplications in three additional unrelated people with MRKH (Gervasini et al. 2010; Ledig et al. 2011; McGowan et al. 2015). We also found two additional *SHOX* missense CV, which warrant further study.

TBX6, a known MRKH candidate gene, is localized to chromosome 16p11.2, and is contained within a common MRKH CNV. We detected three missense CV in exons 4–6 of *TBX6*; one proband had two *TBX6* CV and the other had a single heterozygous CV. Sandbacka et al. (Sandbacka et al. 2013) reported two missense and one splice site variant in *TBX6* in a study of 112 people with MRKH. These alleles had CADD scores ≥ 20 and were not present in 200 controls. However, their variants, like ours, were not studied in vitro. These investigators also identified five large CNVs in 16p11.2 from 112 patients with MRKH that included *TBX6*, but also many other genes. Our findings corroborate those of Sandacka et al (Sandbacka et al. 2013) and strongly suggest *TBX6* as an attractive candidate for further study.

Translocations may disrupt genes or regulatory elements, and can aid in mapping new disease genes (Kim et al. 2010). We previously reported a patient with type I MRKH who had a karyotype of 46,XX,t(3;16)(p22.3;p13.3) with no deletions or duplications at the breakpoints detected by aCGH (Williams et al. 2016). Genes within 13.6 kb at 3p22.3 and 1.9 kb at 16p13.3 of the breakpoints were considered reasonable candidate genes (Table 2C). *TRIM71* and *CNOT1* at 3p22.3 were sequenced in 51 people with MRKH, while *ZNF200*, *OR1F1*, *ZNF213*, and *ZNF205* at 16p13.3 were analyzed in 27 probands (Williams et al. 2016). No likely causative CV were identified in any of these genes; however more people with MRKH needed to be tested. RT-PCR from lymphoblast RNA of the 20 genes located near the breakpoints demonstrated statistically significant downregulation of *CMTM7* and *CCR4*, as well as significant upregulation of *IL32* and *MEFV* (Williams et al. 2016). We would expect that only one of these genes near the translocation breakpoint could result in monogenic MRKH by a position effect, but 21 variants in 11 of these candidate genes were identified, which complicates the analysis (Table 2C). In this study, we showed that most of the genes near these breakpoints are expressed in MRKH-related tissues—uterus, heart, and kidney (Supplemental Table 1).

Some of the listed genes near translocation breakpoints have known associated genetic diseases, lessening the likelihood for involvement in MRKH. These include: *CNOT1* (holoprosencephaly 12 with or without pancreatic agenesis; MIM 618,500); *CRTAP* (osteogenesis imperfecta type VII; MIM 610,682); *TRIM71* (hydrocephalus, communicating, 1; MIM 618,667); and *MEFV* (familial Mediterranean fever, autosomal dominant [MIM 134610] and autosomal recessive [MIM 249100]; acute febrile neutrophilic dermatosis [MIM 608068]). Novel CV were confirmed in *CNOT1* ($n = 4$), *MEFV* ($n = 3$); *CRTAP* ($n = 2$), and *TRIM71* ($n = 1$). None of the seven other genes around the translocation breakpoints are associated with any genetic disease in OMIM. These include *OR1F1*, *TIGD7*, and *TMPPE*, as well as four genes encoding zinc finger proteins (*ZNF200*, *ZNF205*, *ZNF263*, and *ZSCAN10*).

Twelve patients had ≥ 2 CV (Table 3) including four who had two variants in the same gene (*MEFV*, *LRP10*, *TBX6*, and *RARA*). Unlike hypogonadotropic hypogonadism where digenic disease may be found in 2–20% of affected individuals, this occurs in genes already known to be associated with the disorder (Boehm et al. 2015). With MRKH, this is not the case, because only *WNT4* and *HNF1B* variants are the only currently known causative genes in humans. Variants in *WNT4* but not *HNF1B* occurred along with variants in *MEFV* and *CNOT* in two individuals. Using DiGePred, we also identified a potential digenic combination of the *LAMC1* and *MMP14* gene pair based on their proximity in a variety of networks (pathway, interaction, and co-expression). Digenic/oligogenic inheritance is possible, but will require further study. Machine learning approaches appear promising to dissect the intricacies of digenic disease (Mukherjee et al. 2020).

In this study, we began with 72 candidate genes for MRKH, and by WES of DNAs from 111 subjects, we found CV in 24 genes, which narrowed down the number by two-thirds. Variants were identified in six of the 23 genes known to affect Müllerian function in mice. In fact, one gene—*WNT4*—plays a known pathologic role in human MRKH, and we found three CV in this gene. *LAMC1* (four variants) and *RARA* (three variants) were found in

several individuals and are particularly good candidates. We cannot discount *HOXA10*, *PAX2*, and *WNT9B*, although we found only one CV in each of these three genes.

There are some limitations of our study. In genes affecting reproductive function, variant frequency is commonly 1–4%, (Trofimova et al. 2017) so we probably did not have power in 111 patients to detect additional rare variants in other mouse or human genes. However, the major limitation is the lack of family members for analysis. We are now trying to collect all available family members, particularly to analyze trios, which may be much more informative. Nevertheless, WES of these 111 unrelated subjects helped to narrow down the number of candidate genes.

In the 42 human genes that have been studied as physiologic candidates, we found CV in seven genes, but four are known to be associated with other genetic diseases. This leaves *TBX6*, which is a very attractive candidate as one patient had heterozygous CV and another had two *TBX6* CV. *SHOX* is also a possible candidate based upon duplications found in patients with MRKH with potential X-linked dominant inheritance, supported by our finding of two missense CV. *MMP14* and *LRP10* are also possible candidates, but further investigation is necessary. Finally, in the balanced translocation, we would only expect at most, one of the genes to be involved in the pathogenesis of MRKH. Of the 21 positional candidates, CV were found in 11, but 4 are associated with other genetic diseases, which lessens their importance. That still leaves 7 other genes with no known associated genetic disease.

Nearly all of the CV we identified in these candidate genes were heterozygous, which could indicate autosomal dominant inheritance. Since we did not study family members, we cannot determine inheritance until segregation analysis is completed. From the available literature and our own families, we hypothesize that persons with variants in genes associated with MRKH would have heterozygous, de novo variants, or heterozygous variants inherited in an autosomal dominant fashion with reduced penetrance and/or variable expressivity. A sex-limited phenotype to females is also possible. Biallelic variants suggesting autosomal recessive inheritance appear to be less likely because affected siblings are very unusual in our cohorts (Williams et al. 2017). However, we recognize that others have described families with more than one MRKH member (Herlin et al. 2014).

In summary, we performed WES on DNA from 111 individuals with MRKH. From the initial 72 candidate genes studied, we provide evidence for rare CV that narrows down the number of candidate genes to 10. We found CV in six genes studied in mouse (*WNT4*, *LAMC1*, *RARA*, *HOXA10*, *PAX2*, and *WNT9B*) and in four candidate genes based on human studies (*TBX6*, *SHOX*, *MMP14*, and *LRP10*). All of these are predicted to be deleterious given their CADD scores ≥ 20. It is also likely that one gene near the translocation breakpoint is also possibly involved. These promising candidate genes will require further investigation to demonstrate inheritance and causation, which should facilitate our understanding of the molecular basis of MRKH.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was funded by NICHD 1R01 HD092505 (LCL).

References

Backhouse B, Hanna C, Robevska G, van den Bergen J, Pelosi E, Simons C, Koopman P, Juniarto AZ, Grover S, Faradz S, Sinclair A, Ayers K, Tan TY (2019) Identification of candidate genes for Mayer-Rokitansky-Kuster-Hauser syndrome using genomic approaches. Sex Dev 13(1):26–34 [PubMed: 30504698]

Biason-Lauber A, Konrad D, Navratil F, Schoenle EJ (2004) A WNT4 mutation associated with Mullerian-duct regression and virilization in a 46,XX woman. N Engl J Med 351(8):792–798 [PubMed: 15317892]

Boehm U, Bouloux PM, Dattani MT, de Roux N, Dode C, Dunkel L, Dwyer AA, Giacobini P, Hardelin JP, Juul A, Maghnie M, Pitteloud N, Prevot V, Raivio T, Tena-Sempere M, Quinton R, Young J (2015) Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism-pathogenesis, diagnosis and treatment. Nat Rev Endocrinol 11(9):547–564 [PubMed: 26194704]

Cangiano B, Swee DS, Quinton R, Bonomi M (2020) Genetics of congenital hypogonadotropic hypogonadism: peculiarities and phenotype of an oligogenic disease. Hum Genet. 10.1007/s00439-020-02147-1

Chang MC, Jan IS, Liang PC, Jeng YM, Yang CY, Tien YW, Wong JM, Chang YT (2015) Cystic fibrosis transmembrane conductance regulator gene variants are associated with autoimmune pancreatitis and slow response to steroid treatment. J Cyst Fibros 14(5):661–667 [PubMed: 25869325]

Chillon M, Casals T, Gimenez J, Ramos MD, Palacio A, Morral N, Estivill X, Nunes V (1994) Analysis of the CFTR gene confirms the high genetic heterogeneity of the Spanish population: 43 mutations account for only 78% of CF chromosomes. Hum Genet 93(4):447–451 [PubMed: 7513293]

Claustres M, Altieri JP, Guittard C, Templin C, Chevalier-Porst F, Des Georges M (2004) Are p.I148T, p.R74W and p.D1270N cystic fibrosis causing mutations? BMC Med Genet 5:19 [PubMed: 15287992]

Cramer DW, Ravnikar VA, Craighill M, Ng WG, Goldstein DP, Reilly R (1987) Müllerian aplasia associated with maternal deficiency of galactose-1-phosphate uridyl transferase. Fertil Steril 47(6):930–934 [PubMed: 3036608]

Gallati S, Hess S, Galie-Wunder D, Berger-Menz E, Bohlen D (2009) Cystic fibrosis transmembrane conductance regulator mutations in azoospermic and oligospermic men and their partners. Reprod Biomed Online 19(5):685–694 [PubMed: 20021716]

Gervasini C, Grati FR, Lalatta F, Tabano S, Gentilin B, Colapietro P, De Toffol S, Frontino G, Motta F, Maitz S, Bernardini L, Dallapiccola B, Fedele L, Larizza L, Miozzo M (2010) SHOX duplications found in some cases with type I Mayer-Rokitansky-Kuster-Hauser syndrome. Genet Med 12(10):634–640 [PubMed: 20847698]

Girodon E, Cazeneuve C, Lebargy F, Chinet T, Costes B, Ghanem N, Martin J, Lemay S, Scheid P, Housset B, Bignon J, Goossens M (1997) CFTR gene mutations in adults with disseminated bronchiectasis. Eur J Hum Genet 5(3):149–155 [PubMed: 9272738]

Hashimoto R (2003) Development of the human Mullerian duct in the sexually undifferentiated stage. Anat Rec Part A Discov Mol Cel Evol Biol 272(2):514–519

Herlin M, Hojland AT, Petersen MB (2014) Familial occurrence of Mayer-Rokitansky-Kuster-Hauser syndrome: a case report and review of the literature. Am J Med Genet A 164A(9):2276–2286 [PubMed: 24975471]

Hinzpeter A, Aissat A, Sondo E, Costa C, Arous N, Gameiro C, Martin N, Tarze A, Weiss L, de Becdelievre A, Costes B, Goossens M, Galietta LJ, Girodon E, Fanen P (2010) Alternative splicing at a NAGNAG acceptor site as a novel phenotype modifier. PLoS Genet 6(10):e1001153. 10.1371/journal.pgen.1001153 [PubMed: 20949073]

Iso M, Suzuki M, Yanagi K, Minowa K, Sakurai Y, Nakano S, Satou K, Shimizu T, Kaname T (2019) The CFTR gene variants in Japanese children with idiopathic pancreatitis. Hum Genome Var 6:17 [PubMed: 30992994]

Kaufman FR, Kogut MD, Donnell GN, Goebelsmann U, March C, Koch R (1979) Hypergonadotropic hypogonadism in female patients with galactosemia. N Engl J Med 304:994–998

Kim HG, Ahn JW, Kurth I, Ullmann R, Kim HT, Kulharya A, Ha KS, Itokawa Y, Meliciani I, Wenzel W, Lee D, Rosenberger G, Ozata M, Bick DP, Sherins RJ, Nagase T, Tekin M, Kim SH, Kim CH, Ropers HH, Gusella JF, Kalscheuer V, Choi CY, Layman LC (2010) WDR11, a WD protein that interacts with transcription factor EMX1, is mutated in idiopathic hypogonadotropic hypogonadism and Kallmann syndrome. Am J Hum Genet 87(4):465–479 [PubMed: 20887964]

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46(3):310–315 [PubMed: 24487276]

Klipstein S, Bhagavath B, Topipat C, Sasur L, Reindollar RH, Gray MR (2003) The N314D polymorphism of the GALT gene is not associated with congenital absence of the uterus and vagina. Mol Hum Reprod 9(3):171–174 [PubMed: 12606594]

Kurita T (2011) Normal and abnormal epithelial differentiation in the female reproductive tract. Differ Res Biol Divers 82(3):117–126

Lang-Muritano M, Biason-Lauber A, Gitzelmann C, Belville C, Picard Y, Schoenle EJ (2001) A novel mutation in the anti-mullerian hormone gene as cause of persistent mullerian duct syndrome. Eur J Pediatr 160(11):652–654 [PubMed: 11760020]

Layman LC (2014) The genetics of mullerian aplasia. Expert Rev Endocrinol Metab 9:411–419 [PubMed: 30763999]

Ledig S, Schippert C, Strick R, Beckmann MW, Oppelt PG, Wieacker P (2011) Recurrent aberrations identified by array-CGH in patients with Mayer-Rokitansky-Kuster-Hauser syndrome. Fertil Steril 95(5):1589–1594 [PubMed: 20797712]

Ledig S, Brucker S, Barresi G, Schomburg J, Rall K, Wieacker P (2012) Frame shift mutation of LHX1 is associated with Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome. Hum Reprod 27(9):2872–2875 [PubMed: 22740494]

Lee JH, Choi JH, Namkung W, Hanrahan JW, Chang J, Song SY, Park SW, Kim DS, Yoon JH, Suh Y, Jang IJ, Nam JH, Kim SJ, Cho MO, Lee JE, Kim KH, Lee MG (2003) A haplotype-based molecular analysis of CFTR mutations associated with respiratory and pancreatic diseases. Hum Mol Genet 12(18):2321–2332 [PubMed: 12952861]

Lindner TH, Njolstad PR, Horikawa Y, Bostad L, Bell GI, Sovik O (1999) A novel syndrome of diabetes mellitus, renal dysfunction and genital malformation associated with a partial deletion of the pseudo-POU domain of hepatocyte nuclear factor-1beta. Hum Mol Genet 8(11):2001–2008 [PubMed: 10484768]

Lossl P, Kolbel K, Tanzler D, Nannemann D, Ihling CH, Keller MV, Schneider M, Zaucke F, Meiler J, Sinz A (2014) Analysis of nidogen-1/laminin gamma1 interaction by cross-linking, mass spectrometry, and computational modeling reveals multiple binding modes. PLoS ONE 9(11):e112886 [PubMed: 25387007]

Ma J, Qin Y, Liu W, Duan H, Xia M, Chen ZJ (2011) Analysis of PBX1 mutations in 192 Chinese women with Mullerian duct abnormalities. Fertil Steril 95(8):2615–2617 [PubMed: 21575942]

Masse J, Watrin T, Laurent A, Deschamps S, Guerrier D, Pellerin I (2009) The developing female genital tract: from genetics to epi-genetics. Int J Dev Biol 53(2–3):411–424 [PubMed: 19412895]

McGowan R, Tydeman G, Shapiro D, Craig T, Morrison N, Logan S, Balen AH, Ahmed SF, Deeny M, Tolmie J, Tobias ES (2015) DNA copy number variations are important in the complex genetic architecture of mullerian disorders. Fertil Steril 103(4): 1021–1030.e1 [PubMed: 25707337]

Morcel K, Camborieux L, Guerrier D, Mullerieness PDRSLA (2007) Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome. Orphanet J Rare Dis 2:13 [PubMed: 17359527]

Mukherjee S, Cogan JD, Newman JH, Phillips JAIII, Hamid R, Capra JA, Network UDJM (2020) Identifying digenic disease genes using machine learning in the undiagnosed diseases network. bioRxiv. 10.1101/2020.05.31.125716

Oppelt P, Renner SP, Kellermann A, Brucker S, Hauser GA, Ludwig KS, Strissel PL, Strick R, Wallwiener D, Beckmann MW (2006) Clinical aspects of Mayer-Rokitansky-Kuester-Hauser syndrome: recommendations for clinical diagnosis and staging. Hum Reprod 21(3):792–797 [PubMed: 16284062]

Rall K, Eisenbeis S, Barresi G, Ruckner D, Walter M, Poths S, Wallwiener D, Riess O, Bonin M, Brucker S (2015) Fertil Steril 103(2):494–502 e3 [PubMed: 25492683]

Ravel C, Bashamboo A, Bignon-Topalovic J, Siffroi JP, McElreavey K, Darai E (2012) Polymorphisms in DLGH1 and LAMC1 in Mayer-Rokitansky-Kuster-Hauser syndrome. Reprod Biomed Online 24(4):462–465 [PubMed: 22377151]

Reindollar RH, Byrd JR, McDonough PG (1981) Delayed sexual development: a study of 252 patients. Am J Obstet Gynecol 140(4):371–380 [PubMed: 7246652]

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA (2015) 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 17(5):405–424 [PubMed: 25741868]

Sandbacka M, Laivuori H, Freitas E, Halttunen M, Jokimaa V, Morin-Papunen L, Rosenberg C, Aittomaki K (2013) TBX6, LHX1 and copy number variations in the complex genetics of Mullerian aplasia. Orphanet J Rare Dis 8:125 [PubMed: 23954021]

Theisen JG, Sundaram V, Filchak MS, Chorich LP, Sullivan ME, Knight J, Kim HG, Layman LC (2019) The Use of Whole Exome Sequencing in a Cohort of Transgender Individuals to Identify Rare Genetic Variants. Scientific reports 9(1):20099 [PubMed: 31882810]

Timmreck LS, Gray MR, Handelin B, Allito B, Rohlfs E, Davis AJ, Gidwani G, Reindollar RH (2003) Analysis of cystic fibrosis transmembrane conductance regulator gene mutations in patients with congenital absence of the uterus and vagina. Am J Med Genet A 120A(1):72–76 [PubMed: 12794695]

Trofimova T, Lizneva D, Suturina L, Walker W, Chen YH, Azziz R, Layman LC (2017) Genetic basis of eugonadal and hypogonadal female reproductive disorders. Best Pract Res Clin Obstet Gynaecol 44:3–14 [PubMed: 28576390]

Waschk DE, Tewes AC, Romer T, Hucke J, Kapczuk K, Schippert C, Hillemanns P, Wieacker P, Ledig S (2016) Mutations in WNT9B are associated with Mayer-Rokitansky-Kuster-Hauser syndrome. Clin Genet 89(5):590–596 [PubMed: 26610373]

Willem M, Miosge N, Halfter W, Smyth N, Jannetti I, Burghart E, Timpl R, Mayer U (2002) Specific ablation of the nidogen-binding site in the laminin gamma1 chain interferes with kidney and lung development. Development 129(11):2711–2722 [PubMed: 12015298]

Williams LS, Kim HG, Kalscheuer VM, Tuck JM, Chorich LP, Sullivan ME, Falkenstrom A, Reindollar RH, Layman LC (2016) A balanced chromosomal translocation involving chromosomes 3 and 16 in a patient with Mayer-Rokitansky-Kuster-Hauser syndrome reveals new candidate genes at 3p22.3 and 16p13.3. Mol Cytogenet 9:57 [PubMed: 27478502]

Williams LS, Demir Eksi D, Shen Y, Lossie AC, Chorich LP, Sullivan ME, Phillips JA III, Erman M, Kim HG, Alper OM, Layman LC (2017) Genetic analysis of Mayer-Rokitansky-Kuster-Hauser syndrome in a large cohort of families. Fertil Steril 108(1):145–151e2 [PubMed: 28600106]

Yatsenko SA, Rajkovic A (2019) Genetics of human female infertility. Biol Reprod 101(3):549–566 [PubMed: 31077289]

Table 1

List of candidate genes involved in Müllerian development and/or resulting in uterine aplasia in (A) mouse orthologs, (B) humans, and (C) genes near the breakpoint of the t(3;16) translocation identified in an MRKH patient

(A)		
Genes in mouse orthologs	**Author**	**PMID**
*Ctnnb**	Arango et al. (2005)	16,256,976
Dach1	Davis et al. (2008)	18,395,837
Dach2	Davis et al. (2008)	18,395,837
*Emx2**	Miyamoto et al. (1997)	9,165,114
*Hnf1b**	Bingham et al. (2002)	11,918,730
*Hoxa10**	Satokata et al. (1995)	8,532,869
Hoxa11	Satokata et al. (1995)	8,532,869
*Hoxa13**	Warot et a. (1997)	9,428,414
*Hoxa7**	Burel et al. (2006)	1,444,933
Hoxa9	Burel et al. (2006)	1,444,933
*Lamc1**	Willem et al. (2002)	12,015,298
*Lhx1**	Kobayashi et al. (2004)	14,695,376
*Pax2**	Kobayashi et al. (2004)	14,695,376
*Pbx1**	Massé et al. (2009)	19,412,895
Rara	Mendelsohn et al. (1994)	7,607,068
Rarb	Mendelsohn et al. (1994)	7,607,068
*Rarg**	Mendelsohn et al. (1994)	7,607,068
Rxra	Mendelsohn et al. (1994)	7,607,068
Rxrg	Mendelsohn et al. (1994)	7,607,068
*Wnt4**	Vainio et al. (1999)	9,989,404
*Wnt7a**	Miller et al. (1998)	9,767,131
*Wnt9b**	Carroll et al. (2005)	16,054,034
Wnt5a	Mericskay et al. (2004)	15,073,149
(B)		
Human genes	Author	PMID

(A)		
Genes in mouse orthologs	**Author**	**PMID**
BAMH	Resendes et al. (2001)	11,223,848
AMHR	Resendes et al. (2001)	11,223,848
CFTR	Timmreck et al. (2003)	12,794,695
*CTNNB1**	Drummond et al. (2008)	18,979,915
DERL2	Ledig et al. (2017)	29,068,465
DOCK4	Ledig et al. (2017)	29,068,465
DLGH1	Ravel et al. (2012)	22,377,151
*EMX2**	Liu et al. (2015)	25,577,462
ESR1	Brucker et al. (2017)	28,815,558
GALT	Klipstein et al. (2003)	12,606,594
*HNF1B**	Lindner et al. (1999)	10,484,768
HNRNPCL1	Chen et al. (2015)	25,924,657
*HOXA10**	Lalwani et al. (2008)	17,482,600
*HOXA13**	Burel et al. (2006)	16,556,301
*HOXA7**	Guerrier et al. (2006	16,441,882
ITIH5	Morcel et al. (2011)	23,141,121
*LAMC1**	Ravel et al. (2012)	22,377,151
*LHX1**	Cheroki et al. (2008)	18,039,948
LRP10	Rall et al. (2015)	25,492,683
MMP14	Rall et al. (2015)	25,492,683
MSL3	Ledig et al. (2017)	29,068,465
NAALADL2	Ledig et al. (2017)	29,068,465
OR2T2	Chen et al. (2015)	25,924,657
OR4M2	Chen et al. (2015)	25,924,657
OXTR	Brucker et al. (2017)	28,815,558
*PAX2**	Wang et al. (2012)	22,683,154
PAX8	Ma et al. (2014)	25,484,916
*PBX1**	Burel et al. (2006)	16,556,301
PDE11A	Chen et al. (2015)	25,924,657
PLCH1	Ledig et al. (2017)	29,068,465

(A)		
Genes in mouse orthologs	**Author**	**PMID**
*RARG**	Cheroki et al. (2006	16,691,591
RBM8A	Cheroki et al. (2008)	18,039,948
RSPO4	Ledig et al. (2017)	29,068,465
SHOX	Gervasini et al. (2010)	20,847,698
SNAP29	Ledig et al. (2017)	29,068,465
TBX6	Sandbacka et al. (2011)	23,954,021
*WNT4**	Biason-Lauber et al. (2004)	15,317,892
*WNT7A**	Timmreck et al. (2003)	12,794,695
*WNT9B**	Wang et al. (2014)	24,581,601
WT1	Van Lingen et al. (1998)	9,757,958
ZNF277	Ledig et al. (2017)	29,068,465
ZNF816	Chen et al. (2015)	25,924,657

(C)		
t(3;16) Genes	Author	PMID
CASP16P	Williams et al. (2016)	27,478,502
CCR4	Williams et al. (2016)	27,478,502
CMTM6	Williams et al. (2016)	27,478,502
CMTM7	Williams et al. (2016)	27,478,502
CMTM8	Williams et al. (2016)	27,478,502
CNOT10	Williams et al. (2016)	27,478,502
CRTAP	Williams et al. (2016)	27,478,502
DYNC111	Williams et al. (2016)	27,478,502
GLB1	Williams et al. (2016)	27,478,502
IL32	Williams et al. (2016)	27,478,502
MEFV	Williams et al. (2016)	27,478,502
OR1F1	Williams et al. (2016)	27,478,502
TIGD7	Williams et al. (2016)	27,478,502
TMPPE	Williams et al. (2016)	27,478,502
TRIM71	Williams et al. (2016)	27,478,502

(A)		
Genes in mouse orthologs	**Author**	**PMID**
ZNF200	Williams et al. (2016)	27,478,502
ZNF205	Williams et al. (2016)	27,478,502
ZNF213	Williams et al. (2016)	27,478,502
ZNF263	Williams et al. (2016)	27,478,502
ZNF75A	Williams et al. (2016)	27,478,502
ZSCAN10	Williams et al. (2016)	27,478,502

It should be noted that 14 genes, indicated by an asterisk, are common to mouse and human. Note that WNT4 and HNF1B have been shown to be causative for MRKH by functional analysis.

Table 2

Filtered variants called by WES in mouse (A), human (B), and translocation breakpoint (C) candidate genes verified by Sanger sequencing

Gene	Chromosome	Exonic change	Protein change	Reference transcript	CADD score	gnomAD allelic frequency v.2.1.1	gnomAD allelic frequency v.3	gnomAD variant ID	#Patients with variant	MRKH type
(A)										
HOXA10	7	missense; exon 1 c.T188C	p.V63A	NM_018951	20.2	0	0	–	1	I
LAMC1	1	missense; exon 2 c.C713T	p.P238L	NM_002293	26.9	3.18E-05	1.40E-05	1–183072757-C-T	1	I
LAMC1	1	missense; exon 11 c.T1978C	p.Y660H	NM_002293	24.2	0	0	–	1	I
LAMC1	1	missense; exon 16 c.G2854A	p.G952S	NM_002293	34.0	7.96E–06	1.40E–05	1–183095307-G-A	1	II
LAMC1	1	missense; exon 17 c.G3032A	p.R1011H	NM_002293	29.6	4.35E–04	0	1–183096448-G-A	1	I
PAX2	10	missense; exon 3 c.G394A	p.V132I	NM_000278	32	1.20E–05	1.40E-05	10–102510632-G-A	1	I
RARA	17	missense; exon 1 c.127 T	p.P43S	NM_001024809	20.3	3.14E–04	2.93E–04	17–38499083-C-T	1	I
RARA	17	missense; exon 3 c.A320G	p.Q107R	NM_001145302	23.5	0	0	–	1	I
–*RARA*	17	missense; exon 5 c.A611G	p.Q204R	NM_000964	24	0	6.98E–06	17–40352051-A-G	1	I
WNT4	1	missense; exon 5 c.C1015T	p.R339W	NM_030761	33	3.18E–05	1.40E–05	1–22446584-G-A	1	I
WNT4	1	missense; exon 5 c.C829A	p.P277T	NM_030761	26.4	1.42E–05	5.58E–05	1–22446770-G-T	1	I
WNT4	1	missense; exon 5 c.C739T	p.R247C	NM_030761	32	3.18E–05	1.40E–05	1–22446860-G-A	1	I
WNT9B	17	missense; exon 3 c.G454A	p.E152K	NM_001320458	26.4	5.38E–04	4.53E–04	17–44952586-G-A	1	I
(B)										
AMH	19	missense; exon 1 c.A236G	p.Y79C	NM_000479	23.3	0	0	–	1	I
AMH	19	missense; exon 2 c.C451T	p.P151S	NM_000479	23.0	4.50E–05	9.07E–05	19–2250374-C-T	1	II
AMH	19	missense; exon 5 c.C864G	p.D288E	NM_000479	22.9	1.27E–03	1.53E–03	19–2251137-C-G	1	I

Gene	Chromosome	Exonic change	Protein change	Reference transcript	CADD score	gnomAD allelic frequency v.2.1.1	gnomAD allelic frequency v.3	gnomAD variant ID	#Patients with variant	MRKH type
AMH	19	missense; exon 5 c.A974G	p.Q325R	NM_000479	22.8	5.54E–03	9.48E–03	19–2251247-A-G	1	I
CFTR	7	stop-gained; exon 15 c.G2491T	p.E831X	NM_000492	55	0	0	–	1	I
CFTR	7	frameshift; exon 24 c.3903dupA	p.R1301fs	NM_000492	n/a	0	0	–	1	II
CFTR	7	missense; exon 2 c.C91T	p.R31C	NM_000492	27.3	1.64E–03	1.19E–03	7–117144344-C-T	1	I
CFTR	7	missense; exon 3 c.C220T	p.R74W	NM_000492	24	1.72E–03	4.33E–03	7–117149143-C-T	3	I
CFTR	7	missense; exon 4 c.G358A	p.A120T	NM_000492	27.1	1.38E–04	4.19E–05	7–117171037-G-A	1	I
CFTR	7	missense; exon 6 c.A650G	p.E217G	NM_000492	24.5	4.58E–03	3.29E–03	7–117175372-A-G	2	I & II
GALT	9	missense; exon 4 c.A236G	p.Q79R	NM_001258332	25.3	0	0	–	1	II
LRP10	14	missense; exon 3 c.C142T	p.R48W	NM_001329226	23.4	1.09E–03	3.28E–04	14–23342582-C-T	1	I
LRP10	14	missense; exon 5 c.C691T	p.R231W	NM_001329226	33.0	2.35E–03	7.21E–03	14–23344848-C-T	1	I
LRP10	14	missense; exon 6 c.G1458C	p.E486D	NM_001329226	28.1	1.13E–04	2.37E–04	14–23345931-G-C	1	II
LRP10	14	missense; exon 7 c.G1685A	p.R562H	NM_014045	26.6	7.02E–03	6.01E–03	14–23346279-G-A	2	I
MMP14	14	missense; exon 8 c.G1237A	p.D413N	NM_004995	28.5	1.59E–05	0	14–23313925-G-A	1	I
MMP14	14	missense; exon 10 c.C1484T	p.P495L	NM_004995	34.0	1.19E–05	1.40E–05	14–23314983-C-T	1	I
SHOX	X	missense; exon 5 c.G547A	p.V183I	NM_000451	22.8	0	1.26E–04	X-641001-G-A	1	I
SHOX	X	missense; exon 6 c.C698T	p.A233V	NM_000451	29.0	0	9.77E–05	X-644455-C-T	1	II
TBX6	16	missense; exon 6 c.G815A	p.R272Q	NM_004608	34.0	9.61E–04	7.68E–04	16–30099890-C-T	1	II
TBX6	16	missense; exon 5 c.C718T	p.R240C	NM_004608	24.6	2.83E–05	4.19E–05	16–30100064-G-A	1	I
TBX6	16	missense; exon 4 c.G484A	p.G162S	NM_004608	33.0	4.54E–03	3.97E–03	16–30100401-C-T	1	II

Gene	Chromosome	Exonic change	Protein change	Reference transcript	CADD score	gnomAD allelic frequency v.2.1.1	gnomAD allelic frequency v.3	gnomAD variant ID	#Patients with variant	MRKH type
(C) *CNOT1*	3	missense; exon 1 c.C122T	p.S41L	NM_001256742	24.3	1.77E–03	1.86E–03	3–32737239-C-T	1	I
CNOT1	3	missense; exon 12 c.T1586C	p.L529S	NM_001256742	28.8	1.80E–04	1.26E–04	3–32776360-T-C	1	II
CNOT1	3	missense; exon 12 c.C1634T	p.A545V	NM_001256742	23.7	4.60E–05	4.19E–05	3–32776408-C-T	1	I
CNOT1	3	missense; exon 14 c.A1802T	p.Y601F	NM_001256742	28.9	6.56E–03	6.56E–03	3–32800976-A-T	2	II
CRTAP	3	missense; exon 2 c.C514G	p.L172V	NM_006371	23.9	2.14E–05	1.19E–04	3–33161878-C-G	1	II
CRTAP	3	missense; exon 5 c.C1039T	p.L347F	NM_006371	29.0	1.99E–03	2.44E–03	3–33174163-C-T	1	I
MEFV	16	stop-gained; exon 1 c.C202T	p.Q68X	NM_001198536	29.3	0	0	–	1	I
MEFV	16	missense; exon 9 c.G1777A	p.A593T	NM_001198536	21.6	3.98E–06	0	16–3293875-C-T	1	I
MEFV	16	missense; exon 8 c.G1319A	p.R440H	NM_001198536	21.6	0	0	–	1	I
OR1F1	16	missense; exon 1 c.C172T	p.P58S	NM_012360	25.1	7.78E–05	4.19E–05	16–3254418-C-T	1	I
OR1F1	16	missense; exon 1 c.G538C	p.D180H	NM_012360	24.8	2.47E–05	2.79E–05	16–3254784-G-C	1	II
TIGD7	16	frameshift; exon 2 c.143delA	p.N48fs	NM_033208	–	0	0.00E + 0	0	1	II
TMPPE	3	missense; exon 2 c.G628A	p.V210M	NM_001039770	25.6	1.63E–04	1.40E–05	3–33135060-C-T	1	I
TRIM71	3	missense; exon 2 c.G965A	p.R322Q	NM_001039111	25.0	5.35E–05	5.59E–05	3–32915422-G-A	1	I
ZNF200	16	missense; exon 4 c.T347A	p.V116E	NM_001145446	21.6	3.89E–05	2.79E–05	16–3282540-A-T	1	I
ZNF205	16	missense; exon 7 c.C925T	p.R309W	NM_001042428	23.1	3.28E–05	2.09E–05	16–3169586-C-T	1	I
ZNF205	16	missense; exon 7 C.T1549A	p.S517T	NM 001042428	23.2	2.68E–03	2.73E–03	16–3170210-T-A	2	II
ZNF263	16	missense; exon 4 C.G653T	p.S218I	NM_005741	24.9	3.93E–04	1.95E–04	16–3336033-G-T	1	II

Gene	Chromosome	Exonic change	Protein change	Reference transcript	CADD score	gnomAD allelic frequency v.2.1.1	gnomAD allelic frequency v.3	gnomAD variant ID	#Patients with variant	MRKH type
ZSCAN10	16	missense; exon 5 C.G568T	p.G190W	NMJXH282415	26.5	0	0	–	1	II

Table 3

MRKH patients with two or more variants

Patient	Gene	Chromosome	Exonic change	Protein change	Reference transcript	CADD score	gnomAD allelic frequency	gnomAD variant ID
*MRKH*212	*AMH*	19	missense; exon 5 c.A974G	p.Q325R	NM_000479	22.8	5.54E–03	19–2251247-A-G
	HOXA10	7	missense; exon 1 c.T188C	p.V63A	NM_018951	20.2	0	–
MRKH468	*AMH*	19	missense; exon 2 c.C451T	p.P151S	NM_000479	23.0	4.50E–05	19–2250374-C-T
	CRTAP	3	missense; exon 2 c.C514G	p.L172V	NM_006371	23.9	2.14E–05	3–33161878-C-G
MRKH144	*AMH*	19	missense; exon 1 c.A236G	p.Y79C	NM_000479	23.3	0	–
	MEFV	16	missense; exon 9 c.G1777A	p.A593T	NM_001198536	21.6	3.98E–06	16–3293875-C-T
	MEFV	16	missense; exon 8 c.G1319A	p.R440H	NM_001198536	21.6	0	–
MRKH225	*CFTR*	7	missense; exon 6 c.A650G	p.E217G	NM_000492	24.5	4.58E–03	7–117175372-A-G
	SHOX	X	missense; exon 6 c.C698T	p.A233V	NM_000451	29.0	0	–
	ZNF263	16	missense; exon 4 c.G653T	p.S218I	NM_005741	24.9	3.93E–04	16–3336033-G-T
MRKH148	*CNOT1*	3	missense; exon 12 c.C1634T	p.A545V	NM_001256742	23.7	4.60E–05	3–32776408-C-T
	WNT4	1	missense; exon 5 c.C739T	p.R247C	NM_030761	32	3.18E–05	1–22446860-G-A
MRKH604	*MEFV*	16	stop-gained; exon 1 c.C202T	p.Q68X	NM_001198536	29.3	0	–
	WNT4	1	missense; exon 5 c.C829A	p.P277T	NM_030761	26.4	1.42E–05	1–22446770-G-T
MRKH214	*LAMC1*	1	missense; exon 2 c.C713T	p.P238L	NM_002293	26.9	3.18E–05	1–183072757-C-T
	MMP14	14	missense; exon 8 c.G1237A	p.D413N	NM_004995	28.5	1.59E–05	14–23313925-G-A
	ZNF205	16	missense; exon 7 c.C925T	p.R309W	NM_001042428	23.1	3.28E–05	16–3169586-C-T
MRKH215	*LRP10*	14	missense; exon 3 c.C142T	p.R48W	NM_001329226	23.4	1.09E–03	14–23342582-C-T
	SHOX	X	missense; exon 5 c.G547A	p.V183I	NM_000451	22.8	0	–
MRKH222	*LRP10*	14	missense; exon 5 c.C691T	p.R231W	NM_001329226	33.0	2.35E–03	14–23344848-C-T
	LRP10	14	missense; exon 7 c.G1685A	p.R562H	NM_014045	26.6	7.02E–03	14–23346279-G-A
MRKH589	*RARA*	17	missense; exon 3 c.A320G	p.Q107R	NM_001145302	23.5	0	–
	RARA	17	missense; exon 5 c.A611G	p.Q204R	NM_000964	24	0	–
MRKH608	*TBX6*	16	missense; exon 4 c.G484A	p.G162S	NM_004608	33.0	4.54E–03	16–30100401-C-T
	TBX6	16	missense; exon 6 c.G815A	p.R272Q	NM_004608	34.0	9.61E–04	16–30099890-C-T
MRKH235	*TRIM71*	3	missense; exon 2 c.G965A	p.R322Q	NM_001039111	25.0	5.35E–05	3–32915422-G-A
	WNT9B	17	missense; exon 3 c.G454A	p.E152K	NM_001320458	26.4	5.38E–04	17–44952586-G-A

Three patients had two variants in the same gene including MRKH222 in the *LRP10* gene, MRKH589 in the *RARA* gene and MRKH608 in *TBX6*. In addition, MRKH144 has two *MEFV* variants and one *AMH* variant