Please cite this article in press as: Chen et al., Perturbations of genes essential for Müllerian duct and Wölffian duct development in Mayer-Rokitansky-Küster-Hauser syndrome, The American Journal of Human Genetics (2021), https://doi.org/10.1016/j.ajhg.2020.12.014

REPORT

# Perturbations of genes essential for Müllerian duct and Wölffian duct development in Mayer-Rokitansky-Küster-Hauser syndrome

Na Chen,[1,31] Sen Zhao,[2,3,31] Angad Jolly,[4,5,31] Lianlei Wang,[2,3,6,31] Hongxin Pan,[7,31] Jian Yuan,[8] Shaoke Chen,[9] André Koch,[10] Congcong Ma,[1] Weijie Tian,[1] Ziqi Jia,[1] Jia Kang,[1] Lina Zhao,[2,3] Chenglu Qin,[7] Xin Fan,[9] Katharina Rall,[10] Zeynep Coban-Akdemir,[4] Zefu Chen,[2,3] Shalini Jhangiani,[4] Ze Liang,[1] Yuchen Niu,[3] Xiaoxin Li,[3] Zihui Yan,[2,3] Yong Wu,[3] Shuangshuang Dong,[11] Chengcheng Song,[11] Guixing Qiu,[2,3,12] Shuyang Zhang,[13] Pengfei Liu,[4,14] Jennifer E. Posey,[4] Feng Zhang,[11] Guangnan Luo,[7] Zhihong Wu,[3,12,15] on behalf of Deciphering Disorders Involving Scoliosis and COmorbidities (DISCO) study group, Jianzhong Su,[8] Jianguo Zhang,[2,3] Eugenia Y. Chen,[16,17,18] Konstantinos Rouskas,[19,29] Stavros Glentis,[19,30] Flora Bacopoulou,[20] Efthymios Deligeoroglou,[21] George Chrousos,[20] Stanislas Lyonnet,[22] Michel Polak,[22]

*(Author list continued on next page)*

## Summary

Mayer-Rokitansky-Küster-Hauser syndrome (MRKHS) is associated with congenital absence of the uterus, cervix, and the upper part of the vagina; it is a sex-limited trait. Disrupted development of the Müllerian ducts (MD)/Wölffian ducts (WD) through multifactorial mechanisms has been proposed to underlie MRKHS. In this study, exome sequencing (ES) was performed on a Chinese discovery cohort (442 affected subjects and 941 female control subjects) and a replication MRKHS cohort (150 affected subjects of mixed ethnicity from North America, South America, and Europe). Phenotypic follow-up of the female reproductive system was performed on an additional cohort of *PAX8*-associated congenital hypothyroidism (CH) (n = 5, Chinese). By analyzing 19 candidate genes essential for MD/WD development, we identified 12 likely gene-disrupting (LGD) variants in 7 genes: *PAX8* (n = 4), *BMP4* (n = 2), *BMP7* (n = 2), *TBX6* (n = 1), *HOXA10* (n = 1), *EMX2* (n = 1), and *WNT9B* (n = 1), while LGD variants in these genes were not detected in control samples (p = 1.27E−06). Interestingly, a sex-limited penetrance with paternal inheritance was observed in multiple families. One additional *PAX8* LGD variant from the replication cohort and two missense variants from both cohorts were revealed to cause loss-of-function of the protein. From the *PAX8*-associated CH cohort, we identified one individual presenting a syndromic condition characterized by CH and MRKHS (CH-MRKHS). Our study demonstrates the comprehensive utilization of knowledge from developmental biology toward elucidating genetic perturbations, i.e., rare pathogenic alleles involving the same loci, contributing to human birth defects.

Mayer-Rokitansky-Küster-Hauser syndrome (MRKHS [MIM: 277000]), also referred to as Müllerian aplasia, is characterized by congenital absence of the uterus, cervix, and upper part of the vagina in females with a normal karyotype (46, XX).[1] With an incidence of 1 in 4,500–5,000 newborn females, MRKHS is the second most common cause of primary amenorrhea after gonadal dysgenesis.[2] MRKHS is further divided into MRKHS type I (isolated) and MRKHS type II (syndromic) according to the presence of multi-organ involvement.[3]

Formation and morphogenesis of the Müllerian ducts take place during weeks 5–6 of human embryogenesis

[1]Department of Obstetrics and Gynaecology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; [2]Department of Orthopedic Surgery, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; [3]Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Beijing 100730, China; [4]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [5]M.D./Ph.D. Medical Scientist Training Program, Baylor College of Medicine, Houston, TX 77030, USA; [6]Department of Orthopedic Surgery, Qilu Hospital of Shandong University, Jinan, Shandong 250012, China; [7]Department of Obstetrics and Gynaecology, The 3rd Affiliated Hospital of Shenzhen University, Luohu hospital, Shenzhen, Guangdong 518000, China; [8]Institute of Biomedical Big Data, Wenzhou Medical University, Wenzhou 325027, China; [9]Department of Pediatrics, the Second Affiliated Hospital of Guangxi Medical University, Guangxi 530003, China; [10]Department of Gynecology and Obstetrics, Research Centre for Women's Health, Tübingen University Hospital, Tübingen 72076, Germany; [11]Obstetrics and Gynecology Hospital, NHC Key Laboratory of Reproduction Regulation (Shanghai Institute of Planned Parenthood Research), School of Life Sciences, Shanghai 200011, China; [12]Key laboratory of big data for spinal deformities, Chinese Academy of Medical Sciences, Beijing 100730, China; [13]Department of Cardiology, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; [14]Baylor Genetics, Houston, TX 77021, USA; [15]Medical Research Center, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; [16]Baylor College of Medicine, Houston, TX 77030, USA; [17]Departments of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; [18]Department of Obstetrics and Gynecology, Texas Children's Hospital and Baylor College of Medicine, Houston, TX 77030, USA; [19]Institute for Bioinnovation, Biomedical Sciences Research Center Al. Fleming, Vari, Athens 16672, Greece; [20]Center for Adolescent Medicine and UNESCO Chair on Adolescent Health Care, First Department of Pediatrics, School of Medicine, National and

*(Affiliations continued on next page)*

Please cite this article in press as: Chen et al., Perturbations of genes essential for Müllerian duct and Wölffian duct development in Mayer-Rokitansky-Küster-Hauser syndrome, The American Journal of Human Genetics (2021), https://doi.org/10.1016/j.ajhg.2020.12.014

Carla Rosenberg,[23] Irene Dingeldein,[24] Ximena Bonilla,[25] Christelle Borel,[25] Richard A. Gibbs,[4,26] Jennifer E. Dietrich,[17,18,27] Antigone S. Dimas,[19] Stylianos E. Antonarakis,[25,28] Sara Y. Brucker,[10] James R. Lupski,[4,17,26,27,32] Nan Wu,[2,3,4,12,15,32,*] and Lan Zhu[1,32,*]

and are induced by the Wölffian ducts (mesonephric ducts).[4] Knock-out of genes expressed in the Müllerian duct (MD) or the Wölffian duct (WD) such as *Lhx1*,[5] *Pax8*,[6] *Wnt9b*,[7] and *Tbx6*[8] lead to MRKHS-like phenotypes in mice. In addition, knock-out of abdominal B homeobox genes (*Hoxa9*, *Hoxa10*, *Hoxa11*, *Hoxa13*), which are expressed along the anterior-posterior axis in a segmental pattern, can also disrupt the development of the corresponding part of the Müllerian duct, Wölffian duct, and the spine.[9]

Despite numerous candidate genes and pathways, only *WNT4* (MIM: 603490) has been well established to be associated with a clinically distinct subtype of MRKHS characterized by Müllerian aplasia with hyperandrogenism (MIM: 158330).[10] To fully explore the molecular pathogenesis and genetic architecture of MRKHS, we took advantage of the comprehensive knowledge base of genitourinary developmental biology and performed a mutational burden analysis in a cohort of Chinese individuals with MRKHS. We also studied a global multi-center replication cohort consisting of North American, South American, and European individuals with MRKHS, and a Chinese phenotypic follow-up cohort of individuals with congenital hypothyroidism, to further investigate findings from the discovery cohort.

In the discovery cohort, we recruited 442 Chinese individuals of whom 105 derived from trios in which parents were unaffected and 941 female control subjects. The affected individuals consisted of 330 (74.7%) individuals with MRKHS type I and 112 (25.3%) with MRKHS type II. The replication cohort recruited 150 multi-ethnic individuals with MRKHS, including 78 singletons, 66 affected subjects with unaffected familial samples, and 3 families that each have 2 affected members (supplemental methods). This cohort contained 84 (56.0%) with MRKHS type I, 61 (40.7%) with MKRHS type II, and 5 (3.3%) with MRKHS unspecified (Table 1, Figure 1B). Probands and available familial samples from both cohorts underwent exome sequencing (ES, details provided in supplemental methods). The case-control dataset then underwent harmonization processing, where an individual coding sequence site was excluded from the analysis if the absolute difference in percentages of cases compared to controls with at least 10× coverage differed by greater than 8%.[11] Afterward, the mutational burden of likely gene-disrupting (LGD) variants and deleterious missense (D-mis) variants in a set of 19 MD/WD development-associated genes (Table S1) was evaluated. Specifically, LGD variants were defined as variants predicted to cause loss-of-function (LoF) of the gene, and included nonsense, frameshift, and canonical splice site variants. The LGD + D-mis model additionally included missense variants predicted to be damaging by *in-silico* tools (SIFT score[12] $< 0.05$, PolyPhen-2 score[13] $> 0.95$, CADD score[14] $> 15$).

As a result, we identified 12 LGD variants in 7 candidate genes, including *PAX8* (MIM: 167415), *BMP4* (MIM: 112262), *BMP7* (MIM: 112267), *TBX6* (MIM: 602427), *HOXA10* (MIM: 142957), *EMX2* (MIM: 600035), and *WNT9B* (MIM: 602864) (Table 2, Figure 2A), while LGD variants were not detected in any of the candidate genes from the exomes of 941 female control samples ($p = 1.2E{-}06$, SKAT-O). Of the 12 LGD variants, 8 were protein-truncating and 4 affect the canonical splice sites. All of the truncating variants except one in *WNT9B* and one in *BMP4* were predicted to result in unstable mutant RNA susceptible to nonsense-mediated mRNA decay by NMDEscPredictor[15] (Figure S1). A significant burden of LGD + D-mis variants was also observed from cases (35/442) versus controls (38/941) (odds ratio = 1.98, $p = 3.8E{-}03$) in the discovery cohort (Table 2, Figure 2A).

With the identification of four LGD variants (c.156_157dupCG [p.Val53AlafsTer24], c.25+1G>T [splice donor], c.195delC [p.Tyr66ThrfsTer10], and c.322C>T [p.Arg108Ter], GenBank: NM_003466.3), *PAX8* represents the most significant disease-associated gene (false discovery rate [FDR] adjusted $p = 0.01$) (Tables 2, 3, and S2). Notably, $Pax8^{-/-}$ mice recapitulate the human MRKHS phenotype: normal ovaries/oviducts and loss of the uterus.[6] In humans, haploinsufficiency of *PAX8* is associated with congenital hypothyroidism (CH) caused by thyroid dysgenesis.[16] However, all four individuals carrying *PAX8* heterozygous LGD variants presented with isolated MRKHS type I without a clinically observable CH phenotype from the initial visit. Clinical follow up confirmed

Kapodistrian University of Athens, Aghia Sophia Children's Hospital, Athens 11527, Greece; [21]Division of Pediatric-Adolescent Gynecology and Reconstructive Surgery, 2nd Department of Obstetrics and Gynecology, School of Medicine, National and Kapodistrian University of Athens, Aretaieion Hospital, Athens 10679, Greece; [22]Institut Imagine, UMR-1163 INSERM et Universite de Paris, Hospital Universitaire Necker-Enfants Malades, Paris 75015, France; [23]Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of Sao Paulo, Sao Paulo 05508-090, Brazil; [24]Inselspital FrauenKlinik, University of Bern, Bern 3012, Switzerland; [25]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1205, Switzerland; [26]Human Genome Seuencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [27]Texas Children's Hospital, Houston, TX 77030, USA; [28]Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva 1205, Switzerland; [29]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thermi, Thessaloniki 57001, Greece; [30]Division of Pediatric Hematology/Oncology, First Department of Pediatrics, University of Athens, Aghia Sophia Children's Hospital, Athens 11527, Greece
[31]These authors contributed equally
[32]These authors contributed equally
*Correspondence: dr.wunan@pumch.cn (N.W.), zhu_julie@vip.sina.com (L.Z.)
https://doi.org/10.1016/j.ajhg.2020.12.014.

**Table 1. Cohort demographic information and phenotypes**

| Information | Discovery cohort (n = 442) | Replication cohort (n = 150)[a] |
|---|---|---|
| Age of visit, y, mean ± SD | 24.0 ± 5.3 | 17.2 ± 3.8 |
| **Classification, n (%)** | | |
| MRKHS type I | 330 (74.7%) | 84 (56.0%) |
| MRKHS type II | 112 (25.3%) | 61 (40.7%) |
| Unspecified | – | 5 (3.3%) |
| **Complications, n (%)** | | |
| Skeletal anomaly | 90 (20.4%) | 21 (22%) |
| Kidney anomaly | 46 (10.4%) | 32 (33%) |
| Other deformities | 3 (0.7%) | 8 (8%) |

[a]In the replication cohort of 150 samples, data regarding MRKHS classification were available for all the individuals; data regarding age of visit and complications were available for 96 individuals. The percentages in this column were calculated based on individuals with available information. Abbreviations: MRKHS, Mayer-Rokitansky-Küster-Hauser syndrome.

that the thyroid hormone levels of MRK49 (c.25+1G>T) and MRK467 (c.322C>T [p.Arg108Ter]) were within normal range (Table S3). Intriguingly, sequencing of the parental DNA from these two individuals revealed that their *PAX8* LGD variants were both paternally inherited (Figure 2B), consistent with a dominant disease trait with sex-dependent penetrance.

In addition to LGD variants, we also identified three D-mis *PAX8* variants (c.542C>T [p.Ser181Phe], c.266T>C [p.Val89Ala], c.236C>G [p.Ser79Cys]) from three individuals in the discovery cohort (Table 3). We experimentally explored the potential effect on the DNA-binding ability of these three D-mis variants using a luciferase reporter assay and found that one of the three variants (c.236C>G [p.Ser79Cys]) caused decreased transactivation potential of PAX8 protein on its consensus binding sequence (Figure 2C), thus experimentally supporting its characterization as a LoF allele. We followed up this individual with the c.236C>G (p.Ser79Cys) variant (MRK330), and intriguingly, this variant was also found to be inherited from her father (Figure 2B). Moreover, MRK330 and her father both presented with subclinical hypothyroidism (Table S3), supporting the genetic pleiotropy of *PAX8*.

From the replication MRKHS cohort, we identified one nonsense variant in *PAX8* (c.619C>T[p.Arg207Ter]) and two heterozygous D-mis variants in *PAX8* (c.136G>A [p.Asp46Asn] and c.727C>G[p.Gln243Glu]) in three subjects affected with MRKHS type I (Table 3). The c.136G>A (p.Asp46Asn) variant was revealed to cause LoF of PAX8 by luciferase assay (Figure 2C), and this variant was also confirmed to be paternally inherited, consistent with an autosomal-dominant mode of inheritance, with sex-dependent phenotypic expression of the disease trait (Figure 2B). Notably, LGD and D-mis variants in *PAX8* were enriched in the DNA-binding PAX domain of the protein (Figure 2D), implicating an important role of the DNA-binding function of PAX8 in the pathogenesis of MRKHS.

In order to further investigate the genetic pleiotropy of *PAX8*, we enrolled five female individuals ascertained as having pathogenic *PAX8* variants from an independent Chinese cohort with CH.[17] Interestingly, ultrasound and MRI examination of the pelvis during clinical follow-up revealed that one of the individuals (CH123) had previously unobserved aplasia of the uterus, suggesting a clinical diagnosis of MRKHS (Table S4, Figure S2). The *PAX8* variant in this individual was c.68G>T (p.Gly23Val) (Table S3) and was validated to be a LoF allele *in vivo* (Figure 2C). The phenotype of this individual is consistent with a syndromic condition, which we define as congenital hypothyroidism-Mayer-Rokitansky-Küster-Hauser syndrome (CH-MRKHS).

In addition to *PAX8*, two genes encoding bone morphogenetic proteins, namely *BMP4* and *BMP7*, showed a significant mutational burden in the discovery cohort (FDR adjusted p = 0.03 for both *BMP4* and *BMP7*), implicating the contribution of BMP signaling perturbations to MRKHS. *BMP4* and *BMP7* are mainly expressed in the Wölffian duct epithelium and regulate the early formation of the Müllerian duct[18] (Figure 2A). To date, genetic studies have been focused on the function of these genes in the development of the urinary system but not the reproductive system.[19–21]

A heterozygous stop-gain variant c.367G>T (p.Glu123Ter) and a heterozygous splice acceptor variant c.−132−1G>A in *BMP4* (GenBank: NM_001202.3) were identified in MRK644 and MRK166, respectively (Table S2). MRK166 presented with MRKHS type I, and MRK644 had a more complex clinical presentation with idiopathic scoliosis. Interestingly, MRK644 inherited her *BMP4* allele from a mother mosaic for the c.367G>T variant. The *BMP4* variant c.367G>T (p.Glu123Ter) in the proband was detected in 26/55 (47%) variant/total reads by ES, indicating a putative heterozygous variant. In contrast, her mother's blood DNA had Vr/Tr = 13/60 (22%) by ES, consistent with a mosaic variant allele. The mosaicism was further supported by digital droplet PCR (ddPCR), results of which suggested a mutational ratio of around 50% in the proband and 10% in the mother (Figure S3). In the second individual (MRKH166), the splice acceptor variant was paternally inherited, presenting a similar sex-limited inheritance pattern as *PAX8* (Figure S3). ES studies in the replication cohort identified another LGD variant in *BMP4* (c.766C>T [p.Arg256Ter]) from an individual affected with MRKHS type I, further suggesting the association of *BMP4* variants with MRKHS (Table S2). In *BMP7* (GenBank: NM_001719.2), a splice acceptor variant c.1036−2A>G from MRK444 and a frameshift variant c.275_287dupTGGAGGAGGGCGG (p.Pro98GlyfsTer31) from MRK342 were identified (Table S2). Both individuals presented with MRKHS type I, with unknown parental origin.

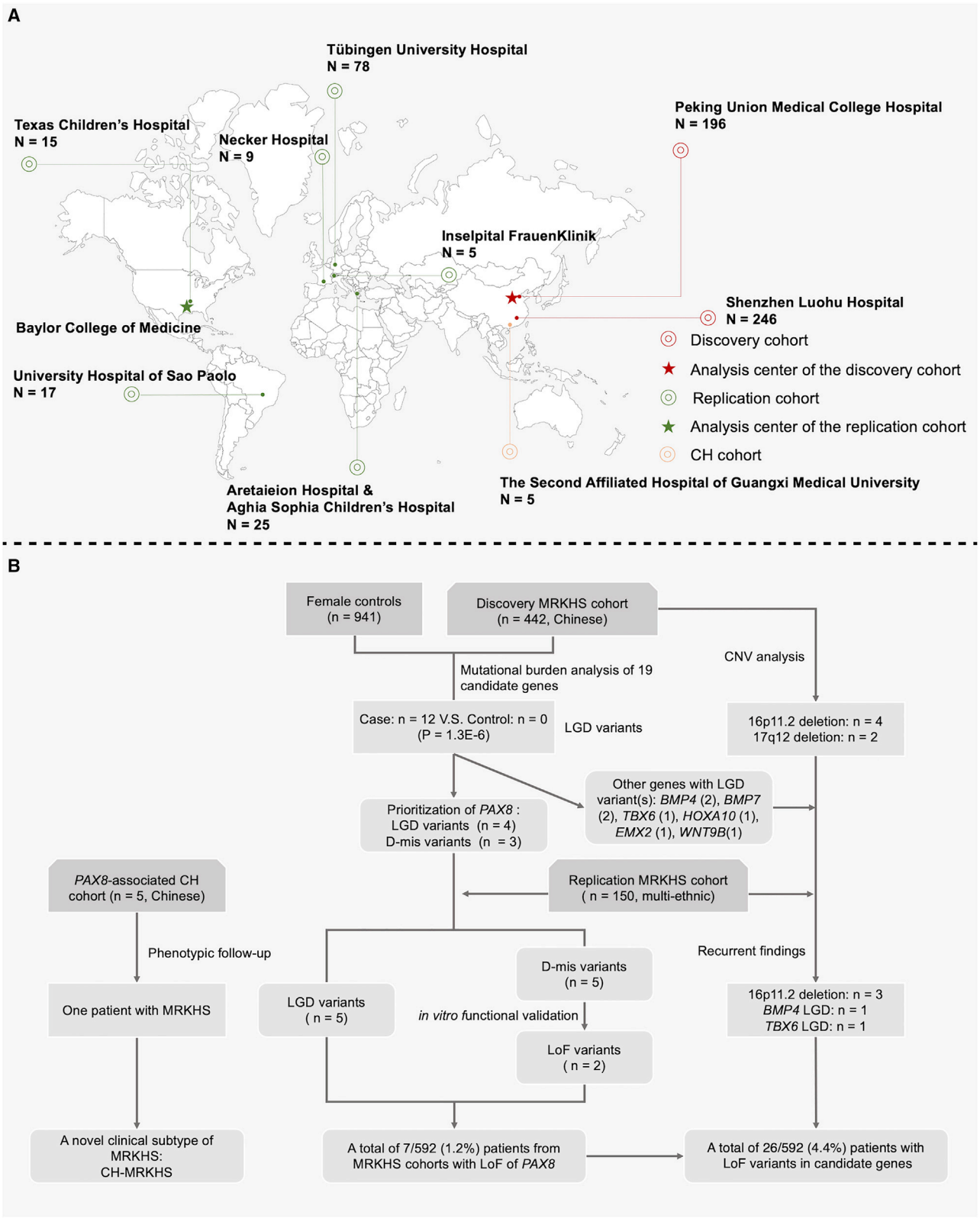


**Figure 1. Study design and workflow of the study**
(A) Centers of recruitment and analysis for MRKHS- and CH-affected individuals.
(B) Workflow and main findings of this study. Abbreviations: MRKHS, Mayer-Rokitansky-Küster-Hauser syndrome; CH, congenital hypothyroidism; LGD, likely gene-disrupting; LoF, loss-of-function; D-mis, deleterious missense.

Please cite this article in press as: Chen et al., Perturbations of genes essential for Müllerian duct and Wölffian duct development in Mayer-Rokitansky-Küster-Hauser syndrome, The American Journal of Human Genetics (2021), https://doi.org/10.1016/j.ajhg.2020.12.014

**Table 2. Mutational burden analysis of 19 candidate genes among 442 affected subjects and 941 female control subjects**

| | | LGD variants | | | LGD + D-mis variants | | | |
|---|---|---|---|---|---|---|---|---|
| **Gene symbol** | **pLI** | **Case (n = 442)** | **Control (n = 941)** | **FDR adjusted p value** | **Case (n = 442)** | **Control (n = 941)** | **OR** | **FDR adjusted p value** |
| *PAX8* | 0.90 | 4 | 0 | 0.01 | 7 | 4 | 3.77 | 0.16 |
| *BMP4* | 0.97 | 2 | 0 | 0.03 | 5 | 1 | 10.77 | 0.05 |
| *BMP7* | 0.98 | 2 | 0 | 0.03 | 4 | 1 | 8.62 | 0.13 |
| *TBX6* | 0.03 | 1 | 0 | 0.08 | 6 | 4 | 3.23 | 0.27 |
| *HOXA10* | 0.68 | 1 | 0 | 0.08 | 3 | 2 | 3.23 | 0.35 |
| *EMX2* | 0.94 | 1 | 0 | 0.08 | 1 | 2 | 1.08 | 0.78 |
| *WNT9B* | 0.05 | 1 | 0 | 0.08 | 3 | 5 | 1.29 | 0.59 |
| *HNF1B* | 1.00 | 0 | 0 | – | 2 | 1 | 4.31 | 0.35 |
| *HOXA9* | 0.00 | 0 | 0 | – | 1 | 1 | 2.15 | 0.59 |
| *GATA3* | 0.88 | 0 | 0 | – | 1 | 2 | 1.08 | 0.78 |
| *LHX1* | 0.29 | 0 | 0 | – | 0 | 0 | – | – |
| *WNT7A* | 0.44 | 0 | 0 | – | 0 | 0 | – | – |
| *HOXA11* | 0.86 | 0 | 0 | – | 1 | 3 | 0.72 | 0.81 |
| *WT1* | – | 0 | 0 | – | 1 | 3 | 0.72 | 0.81 |
| *HOXA13* | – | 0 | 0 | – | 0 | 2 | 0.00 | 0.59 |
| *PAX2* | 0.12 | 0 | 0 | – | 0 | 2 | 0.00 | 0.59 |
| *PBX1* | 0.91 | 0 | 0 | – | 0 | 2 | 0.00 | 0.59 |
| *WNT4* | 0.15 | 0 | 0 | – | 0 | 1 | 0.00 | 0.81 |
| *WNT5A* | 0.97 | 0 | 0 | – | 0 | 2 | 0.00 | 0.76 |
| Total | – | 12 | 0 | 1.2E−06 | 35 | 38 | 1.98 | 3.8E−03 |

Rare variants were analyzed using SNP-set (Sequence) Kernel Association Test-Optimized (SKAT-O) test to determine association of mutational burden in the 19 genes. D-mis variants are defined by missense variants with CADD > 15 and predicted to be deleterious by both SIFT and PolyPhen. Abbreviations: pLI, probability of loss-of-function intolerance from ExAC database; D-mis, damaging missense; FDR, false discovery rate; OR, odds ratio.

We also identified LGD variants in genes previously related to MRKHS, including *TBX6* and *WNT9B*.[22–24] A *TBX6* (GenBank: NM_004608.3) splice donor variant, c.621+1G>A, was identified in individual MRK639 with MRKHS type II. Besides Müllerian aplasia, MRK639 was affected with congenital scoliosis caused by a hemivertebra at L1. Sanger sequencing revealed that this individual carried the hypomorphic T-C-A haplotype (composed of three SNPs, rs2289292-rs3809624-rs3809627[25]) in *trans* with the c.621+1G>A splice variant, which is consistent with the compound inheritance gene dosage model of *TBX6*-associated congenital scoliosis (TACS).[25,26] From the replication cohort, we identified another frameshift variant (c.856_859delAATG [p.His286CysfsTer28]) in a non-canonical transcript of *TBX6* (ENST00000553607) in Mul23, an individual diagnosed with MRKHS type II (Table S2). *WNT9B* is a key signaling molecule which directs early formation of the Müllerian ducts.[4] Missense variants in *WNT9B* were first described in a Chinese cohort of 42 MRKHS-affected individuals.[27] In the present study, we identified a heterozygous nonsense variant c.976C>T (p.Gln326Ter) (GenBank: NM_001320458.2) in MRK57 with isolated MRKHS type I (Tables 2 and S2).

By screening for CNVs that encompass one or more of the 19 candidate genes in the discovery cohort, we identified four deletions of 16p11.2 (encompassing *TBX6*) and two deletions of 17q12 (encompassing *LHX1* [MIM: 601999] and *HNF1B* [MIM: 189907]) (Table S5, Figure S4). Both regions have been linked to MRKHS in previous studies.[28] All four individuals with 16p11.2 deletion were classified as MRKHS type II: one presented with idiopathic scoliosis, and the other three presented with vertebral malformations. Haplotyping of the 16p11.2 locus revealed that the individual with idiopathic scoliosis carried the C-T-C haplotype in the remaining allele, and the other three carried the T-C-A haplotype (Table S5), consistent with the compound inheritance, gene dosage model of TACS.[25,26,29] One of the two individuals with 17q12 deletion presented with MRKHS type I, and the other had a clinical picture complicated by idiopathic scoliosis (Table S5). From the replication cohort, three 16p11.2 deletion CNVs were identified in three individuals (Figure S5), including two individuals with MRKHS

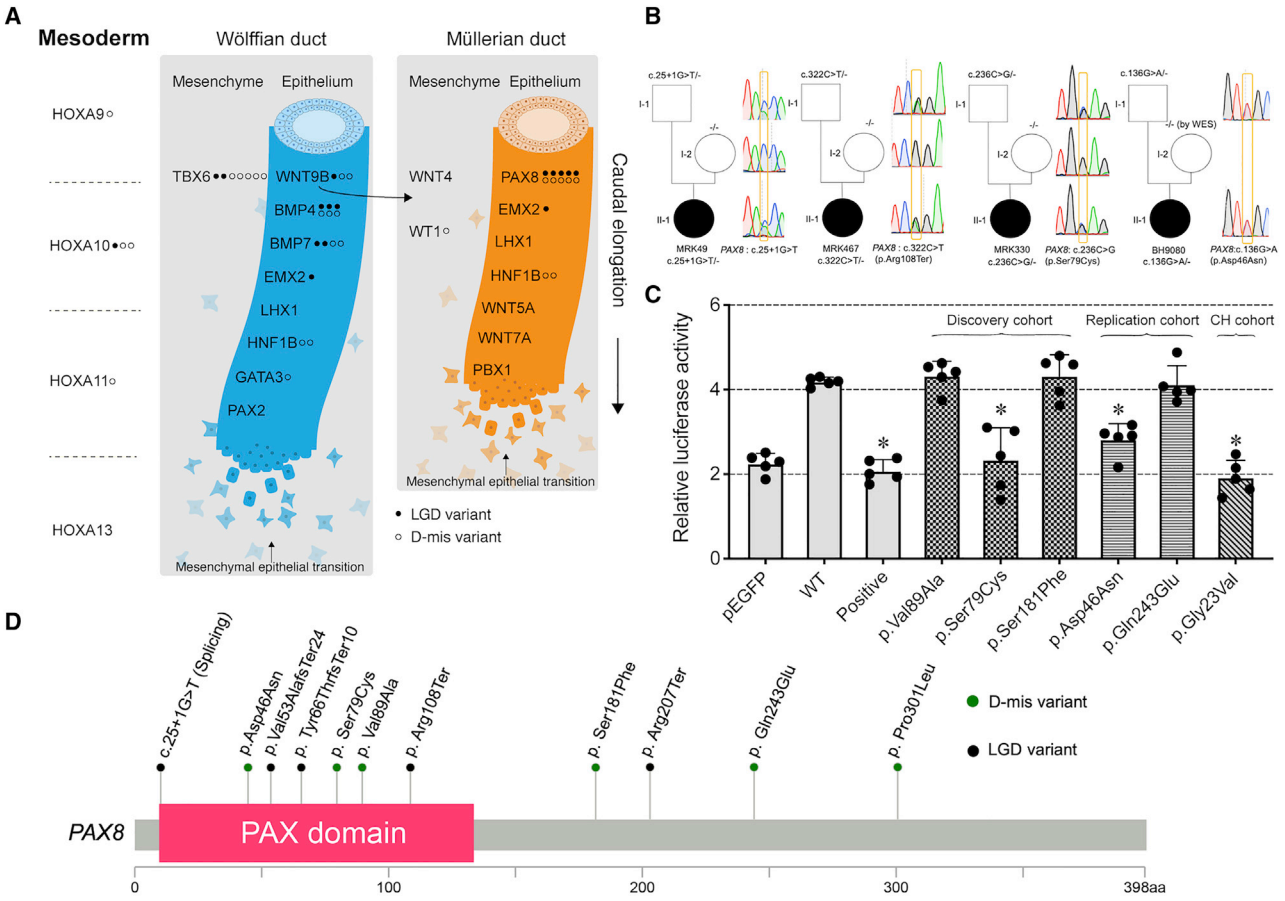


**Figure 2. Identification of variants in genes essential for Müllerian duct (MD) and Wölffian duct (WD) development**
(A) The elongation process of Müllerian duct/Wölffian duct is depicted. Protein located in the corresponding regions and with prioritized variants in them are also presented. Abbreviations: LGD, likely gene-disrupting; D-mis, damaging missense.
(B) Pedigree and Sanger sequencing results of three families with *PAX8* variants from the discovery cohort and one family from the replication cohort.
(C) Luciferase assay on *PAX8* missense variants from the discovery cohort and the replication cohorts. WT, wild type; R31C, a known deleterious variant, was used as the positive control; CH, congenital hypothyroidism. The data are the mean of n = 5 independent experiments. Error bars show one standard deviation (*$p < 0.05$, Dunnett's multiple comparison test).
(D) The mutational spectrum of LGD variants and damaging missense variants in *PAX8*.

type I and one individual with hip dysplasia and thus MRKHS type II. No 17q12 deletion CNVs were found in the replication cohort.

In summary, candidate variants in genes associated with MD/WD development were identified in MRKHS. A sex-limited penetrance with paternal inheritance was observed in multiple families. Of them, *PAX8* represents the most significant gene underlying the etiology of 7/592 (1.2%) individuals and is associated with a syndromic condition characterized by CH and MRKHS (CH-MRKHS). Other candidate genes such as *BMP4* and *BMP7* still warrant further genetic and functional studies. Our study demonstrates the comprehensive utilization of knowledge from developmental biology toward elucidating genetic perturbations, i.e., rare variant pathogenic alleles involving the same loci, contributing to human birth defects.

## Data and Code Availability

The datasets supporting the current study have not been deposited in a public repository due to institutional ethics restrictions but are available from the corresponding author on request.

### Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.12.014.

### Consortia

The members of the DISCO consortium are Guixing Qiu, Zhihong Wu, Terry Jianguo Zhang, Nan Wu, Shengru Wang, Jiaqi Liu, Sen Liu, Yuzhi Zuo, Gang Liu, Chenxi Yu, Lian Liu, Jiashen Shao,

**Table 3. Summary of *PAX8* variants identified from the discovery cohort and the replication cohort**

| Sample ID | MRKHS classification | Ethnicity | Zygosity | CHR | POS | vR | tR | Mutation type | Gene name | cDNA change | Protein change | gnomAD-AF | ExAC-AF | Inheritance | CADD score | Functional validation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Discovery cohort** | | | | | | | | | | | | | | | | |
| MRK51 | type I | Chinese | het | 2 | 114004364 | 22 | 46 | frameshift | *PAX8* | c.156_157dupCG | p.Val53AlafsTer24 | 0 | 0 | unknown | – | – |
| MRK49 | type I | Chinese | het | 2 | 114035946 | 105 | 265 | splice_donor | *PAX8* | c.25+1G>T | – | 0 | 0 | paternal | – | – |
| MRK442 | type I | Chinese | het | 2 | 114002198 | 53 | 101 | frameshift | *PAX8* | c.195delC | p.Tyr66ThrfsTer10 | 0 | 0 | paternal | – | – |
| MRK467 | type I | Chinese | het | 2 | 114002071 | 49 | 89 | stop_gained | *PAX8* | c.322C>T | p.Arg108Ter | 0 | 0 | unknown | – | – |
| MRK236 | type I | Chinese | het | 2 | 113999644 | 33 | 76 | missense | *PAX8* | c.542C>T | p.Ser181Phe | 0 | 0 | unknown | 28.8 | normal |
| MRK283 | type I | Chinese | het | 2 | 114002127 | 58 | 99 | missense | *PAX8* | c.266T>C | p.Val89Ala | 0 | 0 | unknown | 26.7 | normal |
| MRK330 | type I | Chinese | het | 2 | 114002157 | 56 | 112 | missense | *PAX8* | c.236C>G | p.Ser79Cys | 0 | 0 | paternal | 26.5 | LoF |
| **Replication cohort** | | | | | | | | | | | | | | | | |
| SEA13832 | type I | European | het | 2 | 113999286 | 38 | 91 | stop_gained | *PAX8* | c.619C>T | p.Arg207* | 0 | 0 | unknown | – | – |
| BH9080 | type I | European | het | 2 | 114004386 | 84 | 179 | missense | *PAX8* | c.136G>A | p.Asp46Asn | 0 | 0 | paternal | 29.3 | LoF |
| SEA13831 | type I | European | het | 2 | 113999178 | 59 | 224 | missense | *PAX8* | c.727C>G | p.Gln243Glu | 0 | 0 | unknown | 22.4 | normal |

GenBank: NM_003466.3 was used for annotations of cDNA and protein changes. Abbreviations: Chr, chromosome; gnomAD, genome aggregation database; ExAC, Exome Aggregation Consortium; AF, allele frequency; Ref, reference; VR, variant reads; TR, total reads; VAF, variant allele frequency; –, not applicable; LoF, loss-of-function; CADD, combined annotation dependent depletion.

Sen Zhao, Zihui Yan, Hengqiang Zhao, Yuchen Niu, Xiaoxin Li, Huizi Wang, Congcong Ma, Zefu Chen, Bowen Liu, Xi Cheng, Jiachen Lin, Huakang Du, Yaqi Li, Shuang Song, Weijie Tian, Zhixin Xie, Zhengye Zhao, Lina Zhao, Zhi Zhao, Zhifa Zheng, and Yingzhao Huang.

## Acknowledgments

We appreciate all of the patients, their families, and clinical research coordinators who participated in this project. We thank GeneSeeq Inc. for exome sequencing technical support. We thank Beijing Ekitech Co. Ltd. for support in bioinformatic analyses. This work was supported by the National Natural Science Foundation of China (81822030 and 82072391 to N.W., 81772299 and 81930068 to Z.W., 31625015 and 31771396 to F.Z., 81772301 and 81972132 to G.Q., 81801401 to N.C., and 81672123 and 81972037 to J.Z.), Beijing Natural Science Foundation (JQ20032 to N.W., 7191007 to Z.W.), 2016 Milstein Medical Asian American Partnership Foundation Fellowship Award in Translational Medicine (to N.W.), National Science and Technology Support Program (2015BAI13B04 to L. Zhu), CAMS Initiative Fund for Medical Sciences (2016-I2M-3-003 to G.Q. and N.W., 2017-12M-1-002 to L. Zhu, 2016-I2M-2-006 and 2017-I2M-2-001 to Z.W.), the Central Level Public Interest Program for Scientific Research Institute (2018RC31003 to N.W.), the National Key Research and Development Program of China(2016YFC0901501 to S. Zhang, 2018YFC0910506 to N.W. and Z.W.), Sanming Project of Medicine in Shenzhen, Health and Family Planning Commission of Guangdong Province (A2018431), Shenzhen Healthcare Research Project (201601055 and SZLY2017020), Shanghai Medical Center of Key Programs for Female Reproductive Diseases (2017ZZ01016 to F.Z.), and Shanghai Municipal Science and Technology Major Project (2017SHZDZX01 to F.Z.). Also supported by the US National Institutes of Health, National Institute of Neurological Disorders and Stroke (NINDS R35 NS105078 to J.R.L.), National Human Genome Research Institute/National Heart, Lung, and Blood Institute (NHGRI/NHLBI UM1 HG006542 to J.R.L.), the National Human Genome Research Institute (NHGRI K08 HG008986 to J.E.P.), the Foundation Santé Biomedical Research Grant (to A.S.D.), Latsis Foundation Research Grant (to A.S.D.), Stavros Niarchos Foundation (startup funds to A.S.D.), the Deutsche Forschungsgemeinschaft (DFG 351381475), European Research Council (ERC 2199678), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 2015/14821-1, 2017/16283-2), and an Institutional grant from the Faculty of Medicine at the University of Geneva Medical School.

## Declaration of interests

J.R.L. has stock ownership in 23andMe, is a paid consultant for Regeneron Pharmaceuticals and Novartis, and is a co-inventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting. The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from the chromosomal microarray analysis and clinical genomic sequencing offered in the Baylor Genetics Laboratory (http://baylorgenetics.com).

Received: July 11, 2020
Accepted: December 21, 2020
Published: January 11, 2021

## Web Resources

Combined annotation dependent depletion (CADD), https://cadd.gs.washington.edu/snv
ExAC database, http://exac.broadinstitute.org/
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
Genome aggregation database (gnomAD), https://gnomad.broadinstitute.org/
Online Mendelian Inheritance in Man (OMIM), https://www.omim.org/

## Reference

1. Morcel, K., Camborieux, L., Guerrier, D.; and Programme de Recherches sur les Aplasies Müllériennes (2007). Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome. Orphanet J. Rare Dis. *2*, 13.
2. Cheroki, C., Krepischi-Santos, A.C., Rosenberg, C., Jehee, F.S., Mingroni-Netto, R.C., Pavanello Filho, I., Zanforlin Filho, S., Kim, C.A., Bagnoli, V.R., Mendonça, B.B., et al. (2006). Report of a del22q11 in a patient with Mayer-Rokitansky-Küster-Hauser (MRKH) anomaly and exclusion of WNT-4, RAR-gamma, and RXR-alpha as major genes determining MRKH anomaly in a study of 25 affected women. Am. J. Med. Genet. A. *140*, 1339–1342.
3. Grimbizis, G.F., Gordts, S., Di Spiezio Sardo, A., Brucker, S., De Angelis, C., Gergolet, M., Li, T.C., Tanos, V., Brölmann, H., Gianaroli, L., and Campo, R. (2013). The ESHRE-ESGE consensus on the classification of female genital tract congenital anomalies. Gynecol. Surg. *10*, 199–212.
4. Orvis, G.D., and Behringer, R.R. (2007). Cellular mechanisms of Müllerian duct formation in the mouse. Dev. Biol. *306*, 493–504.
5. Kobayashi, A., Shawlot, W., Kania, A., and Behringer, R.R. (2004). Requirement of Lim1 for female reproductive tract development. Development *131*, 539–549.
6. Mittag, J., Winterhager, E., Bauer, K., and Grümmer, R. (2007). Congenital hypothyroid female pax8-deficient mice are infertile despite thyroid hormone replacement therapy. Endocrinology *148*, 719–725.
7. Carroll, T.J., Park, J.S., Hayashi, S., Majumdar, A., and McMahon, A.P. (2005). *Wnt9b* plays a central role in the regulation of mesenchymal to epithelial transitions underlying organogenesis of the mammalian urogenital system. Dev. Cell *9*, 283–292.
8. Nacke, S., Schäfer, R., Habré de Angelis, M., and Mundlos, S. (2000). Mouse mutant "rib-vertebrae" (rv): a defect in somite polarity. Dev. Dyn. *219*, 192–200.
9. Mullen, R.D., and Behringer, R.R. (2014). Molecular genetics of Müllerian duct formation, regression and differentiation. Sex Dev. *8*, 281–296.
10. Biason-Lauber, A., Konrad, D., Navratil, F., and Schoenle, E.J. (2004). A *WNT4* mutation associated with Müllerian-duct regression and virilization in a 46,XX woman. N. Engl. J. Med. *351*, 792–798.
11. Sanna-Cherchi, S., Khan, K., Westland, R., Krithivasan, P., Fievet, L., Rasouly, H.M., Ionita-Laza, I., Capone, V.P., Fasel, D.A., Kiryluk, K., et al. (2017). Exome-wide Association Study Identifies *GREB1L* Mutations in Congenital Kidney Malformations. Am. J. Hum. Genet. *101*, 789–802.
12. Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M., and Ng, P.C. (2016). SIFT missense predictions for genomes. Nat. Protoc. *11*, 1–9.

13. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
14. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
15. Coban-Akdemir, Z., White, J.J., Song, X., Jhangiani, S.N., Fatih, J.M., Gambin, T., Bayram, Y., Chinn, I.K., Karaca, E., Punetha, J., et al.; Baylor-Hopkins Center for Mendelian Genomics (2018). Identifying Genes Whose Mutant Transcripts Cause Dominant Disease Traits by Potential Gain-of-Function Alleles. Am. J. Hum. Genet. *103*, 171–187.
16. Macchia, P.E., Lapi, P., Krude, H., Pirro, M.T., Missero, C., Chiovato, L., Souabni, A., Baserga, M., Tassi, V., Pinchera, A., et al. (1998). *PAX8* mutations associated with congenital hypothyroidism caused by thyroid dysgenesis. Nat. Genet. *19*, 83–86.
17. Fu, C., Chen, R., Zhang, S., Luo, S., Wang, J., Chen, Y., Zheng, H., Su, J., Hu, X., Fan, X., et al. (2015). *PAX8* pathogenic variants in Chinese patients with congenital hypothyroidism. Clin. Chim. Acta *450*, 322–326.
18. Atsuta, Y., and Takahashi, Y. (2016). Early formation of the Müllerian duct is regulated by sequential actions of BMP/Pax2 and FGF/Lim1 signaling. Development *143*, 3549–3559.
19. Dudley, A.T., Lyons, K.M., and Robertson, E.J. (1995). A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye. Genes Dev. *9*, 2795–2807.
20. Winnier, G., Blessing, M., Labosky, P.A., and Hogan, B.L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes Dev. *9*, 2105–2116.
21. Hwang, D.Y., Dworschak, G.C., Kohl, S., Saisawat, P., Vivante, A., Hilger, A.C., Reutter, H.M., Soliman, N.A., Bogdanovic, R., Kehinde, E.O., et al. (2014). Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary tract. Kidney Int. *85*, 1429–1433.
22. Waschk, D.E., Tewes, A.C., Römer, T., Hucke, J., Kapczuk, K., Schippert, C., Hillemanns, P., Wieacker, P., and Ledig, S. (2016). Mutations in *WNT9B* are associated with Mayer-Rokitansky-Küster-Hauser syndrome. Clin. Genet. *89*, 590–596.
23. Bernardini, L., Gimelli, S., Gervasini, C., Carella, M., Baban, A., Frontino, G., Barbano, G., Divizia, M.T., Fedele, L., Novelli, A., et al. (2009). Recurrent microdeletion at 17q12 as a cause of Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome: two case reports. Orphanet J. Rare Dis. *4*, 25.
24. Sandbacka, M., Laivuori, H., Freitas, É., Halttunen, M., Jokimaa, V., Morin-Papunen, L., Rosenberg, C., and Aittomäki, K. (2013). *TBX6*, *LHX1* and copy number variations in the complex genetics of Müllerian aplasia. Orphanet J. Rare Dis. *8*, 125.
25. Wu, N., Ming, X., Xiao, J., Wu, Z., Chen, X., Shinawi, M., Shen, Y., Yu, G., Liu, J., Xie, H., et al. (2015). *TBX6* null variants and a common hypomorphic allele in congenital scoliosis. N. Engl. J. Med. *372*, 341–350.
26. Yang, N., Wu, N., Zhang, L., Zhao, Y., Liu, J., Liang, X., Ren, X., Li, W., Chen, W., Dong, S., et al. (2019). *TBX6* compound inheritance leads to congenital vertebral malformations in humans and mice. Hum. Mol. Genet. *28*, 539–547.
27. Wang, M., Li, Y., Ma, W., Li, H., He, F., Pu, D., Su, T., and Wang, S. (2014). Analysis of *WNT9B* mutations in Chinese women with Mayer-Rokitansky-Küster-Hauser syndrome. Reprod. Biomed. Online *28*, 80–85.
28. Nik-Zainal, S., Strick, R., Storer, M., Huang, N., Rad, R., Willatt, L., Fitzgerald, T., Martin, V., Sandford, R., Carter, N.P., et al. (2011). High incidence of recurrent copy number variants in patients with isolated and syndromic Müllerian aplasia. J. Med. Genet. *48*, 197–204.
29. Liu, J., Wu, N., Yang, N., Takeda, K., Chen, W., Li, W., Du, R., Liu, S., Zhou, Y., Zhang, L., et al.; Deciphering Disorders Involving Scoliosis and COmorbidities (DISCO) study; Japan Early Onset Scoliosis Research Group; and Baylor-Hopkins Center for Mendelian Genomics (2019). *TBX6*-associated congenital scoliosis (TACS) as a clinically distinguishable subtype of congenital scoliosis: further evidence supporting the compound inheritance and *TBX6* gene dosage model. Genet. Med. *21*, 1548–1558.