

Genome sequencing and transcriptome profiling in twins discordant for Mayer-Rokitansky-Küster-Hauser syndrome

Rebecca Buchert^{1*}, Elisabeth Schenk^{1*}, Thomas Hentrich¹, Nico Weber², Katharina Rall^{3,4}, Marc Sturm¹, Oliver Kohlbacher^{2,6,7}, André Koch⁵, Olaf Rieß^{1,4}, Sara Y. Brucker^{3,4*}, Julia M. Schulze-Hentrich^{1,6*+}

1 Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany

2 Applied Bioinformatics, Dept. of Computer Science, University of Tübingen, Tübingen, Germany

3 Department of Women's Health, University of Tübingen, Tübingen, Germany

4 Rare Disease Center Tübingen, University of Tübingen, Tübingen, Germany

5 Research Institute for Women's Health, University of Tübingen, Tübingen, Germany

6 Institute for Bioinformatics and Medical Informatics, University of Tübingen, Germany

7 Institute for Translational Bioinformatics, University Hospital Tübingen, Germany

+ Corresponding author

Corresponding author: julia.schulze-hentrich@med.uni-tuebingen.de, +49 7071-29-72276

Keywords: MRKH syndrome, monozygotic discordant twins, genome sequencing, transcriptome analysis, Müllerian ducts

Abstract

Objective

Mayer-Rokitansky-Küster-Hauser syndrome (MRKH) is a rare congenital disease manifesting with aplasia or severe hypoplasia of uterine structures. Even though extensive studies have been performed, for the majority of cases the etiology remains unclear. In this study, we sought to identify genetic causes in discordant monozygotic (MZ) twins using genome sequencing of blood of both twins as well as uterine tissue of the affected twin. In addition, we profiled the endometrial transcriptome of affected twins to compare perturbations with those of sporadic MRKH cases.

Results

First, analyzing the data under the assumption that a variant solely identified in the affected twin or affected tissue could cause the phenotype, we identified a mosaic variant in *ACTR3B* with a high allele frequency in affected tissue, a low allele frequency in blood of the affected twin and almost absent in the blood of the unaffected twin. Since *ACTR3B* has not been reported for genitourinary anomalies before, clinical relevance of the variant needs to be clarified.

Second, examining the data for candidate genes previously implied in MRKH, we detected a pathogenic variant in *GREB1L* in one twin pair and their unaffected mother showing a reduced phenotypic penetrance. Furthermore, two variants of unknown clinical significance in *PAX8* and *WNT9B* were identified. Analysis for copy number and structural variants revealed no discordant variants in the twins or variants in candidate genes or regions.

Third, we conducted transcriptome analysis of affected tissue and observed widespread perturbations largely similar to those in sporadic MRKH cases. These shared transcriptional changes were enriched for terms associated with estrogen and its receptors pointing at a key role of estrogen in MRKH pathology.

Conclusion

Our study on genome sequencing of blood and uterine tissue of discordant twins is the most extensive study performed on twins discordant for MRKH so far. Nevertheless, no clear pathogenic differences in the twins or between blood and tissue samples were detected. This points towards a more complex etiology of MRKH less dependent on genetic differences and more determined by epigenetic changes or environmental factors. Our transcriptome data showed a clear overlap with gene expression data of sporadic MRKH cases, indicating that the etiology for MRKH in discordant twins and sporadic cases is largely similar.

Introduction

Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome is a rare congenital disease present in approximately one in 5,000 live female births manifesting as aplasia or severe hypoplasia of structures related to the development of the Müllerian duct including the uterus and upper part of the vagina [1]. MRKH can either occur as an isolated form (type 1) or with additional extragenital phenotypes (type 2) such as renal or skeletal malformations [2].

Females affected by MRKH syndrome usually have a normal 46.XX karyotype, normal ovarian function and normal secondary sexual characteristics. Besides various sporadic cases, several familial occurrences have been observed suggesting a potential genetic origin of the syndrome [3]. Multiple genes have been subject to intensive investigation but could only partly be linked to a subset of MRKH cases in individual cohorts. Among them are genes essential for Müllerian and Wolffian duct development which might play a key role for MRKH etiology [4]. Copy Number Variants (CNVs) in 1q21.1, 16p11.2, 17q12, and 22q11 with the corresponding genes *LHX1* and *HNF1B*, WNT signaling pathway genes such as *WNT4*, *WNT9B* and the HOX family as well as some others have been pivotal candidates [5]. Recently, WES was used to identify *GREB1L* as a potential candidate for MRKH type 2 cases with kidney anomalies [6], but in general the multifaceted phenotype of the type 2 MRKH syndrome makes interpretation of genetic findings complex. Until recently, previous work only focused on genome wide CNV detection using array-CGH assays or WES and SNP arrays [7, 8] limiting detection to exonic variants and therefore discarding potential mutations in regulatory factors. In 2019, Pan *et al.* [9] published results based on WGS for detection of *de novo* variants in nine MRKH type 1 patients, which still remains, up to our knowledge, the only publication with a whole genome approach for MRKH patients complementing previously performed work.

As MRKH occurs mostly sporadic, this syndrome could also be explained by *de novo* dominant mutations in the germline of the offspring, or somatic mutations in the affected tissue. Those somatic variants which might occur in early development could also explain cases of monozygotic twins discordant for the MRKH syndrome [10-12]. Up until now, we are not aware of any publication analyzing and discussing tissue specific variants in MRKH patients which leads to certain unknowns concerning potential genetic mosaics in uterus rudiments as a potential source for MRKH type 1 and type 2 malformations.

To better understand the genetic contribution to MRKH etiology, we herein sequence and analyze the whole genome of blood and affected uterus rudiment of five twin pairs discordant for MRKH in an attempt to include regulatory features as well as potential tissue-specific gene alteration. In addition, we extend our tissue-specific endometrial transcriptome analysis from sporadic cases [13] to monozygotic discordant twins and evaluate perturbations in gene expression underlying MRKH.

Materials and Methods

Patients and Ethical Approval

From five MRKH-discordant pairs of MZ twins, blood samples and clinical data were collected during routine clinical visits at the Department of Obstetrics and Gynecology, Tübingen University Hospital, Tübingen, Germany. The study was approved by the Institutional Review Board of the University of Tübingen (approval number: 205/2014BO1) and informed consent for genetic studies was obtained from each patient before recruitment. Participants' or their legal representatives' consent to take part in the present study and have the results of this research work published.

Of the five patients with MRKH, one had type I and four had type II MRKH with associated renal malformations, one of whom had ureter abnormalities. Except for one pair, where both had renal agenesis, the corresponding co-twins were not affected by genital, renal, skeletal, or other malformations (Table 1). Tissue from uterine rudiments was collected from the affected twin during laparoscopic creation of a neovagina.

DNA isolation and sequencing

DNA from blood was isolated using the DNeasy purification kit and automated on the QIA-cube (Qiagen). DNA from rudimentary uterine tissue was isolated with the QiaAmp DNA Mini Kit (Qiagen). Genome sequencing was performed using the TruSeq PCR-free kit (Illumina) on an Illumina platform (NovaSeq6000). Generated sequences were processed using the megSAP analysis pipeline (<https://github.com/imgag/megSAP>). megSAP performs quality control, read alignment, various alignment post-processing steps, variant detection, as well as comprehensive annotation of variants. Diagnostic analysis was performed using the GSvar clinical decision support system (<https://github.com/imgag/ngs-bits>). We used the GRCh38 reference genome assembly for the analysis.

Genome analysis and variant calling

The resulting data was then analyzed implementing different filter criteria and sample constellations.

Initially, the affected twin's tissue and blood samples were compared to the respective healthy twin's blood sample using a multi-sample dominant filter (allele frequency $\leq 0.10\%$, coding or near splice variants, keep pathogenic variants of HGMD or ClinVar, quality filter (quality: 250 | depth: 0 | mapq: 55 | strand bias: 20 | allele balance: 80)) and filtered for variants present in affected tissue or affected twin while absent in blood or unaffected twin, respectively. Later, the multi-sample dominant filter was modified by including all coding or non-coding variants in MRKH target regions. This filter contained 612 regions and 752 different genes already described in literature to be involved in uterine development and tissue genesis pathways.

The second analysis focused on comparing the affected twin's tissue and blood using once more the modified multi-sample dominant filter and target region filter.

Lastly, we analyzed for rare ($>1\%$ allele frequency) Single Nucleotide Variants (SNVs), CNVs or structural variants in the MRKH target region in the affected twin's tissue

All remaining variants of the previous analyses were then manually evaluated, with respect to quality, conservation, the gene's biological functions and previous association to MRKH.

Validation using Sanger sequencing

To validate the variant found in *ACTR3B*, primer pairs for Sanger sequencing were designed with Geneious Prime, a platform utilizing the widely used Primer3 algorithm. The resulting forward- (*ACTR3B*-F: TACTCTCAGGAGGCTCCACC) and reverse-primer (*ACTR3B*-R: CCAGGGAGAGTGTGAAGCTG) were produced by metabion international AG (Steinkirchen, Germany). PCR was then performed using the Taq DNA Polymerase Kit (Qiagen, Santa Clarita, CA, USA) with a final volume of 40 μ l containing 0.4 μ l of Taq DNA Polymerase, 0.8 μ l dNTPs, 4 μ l QIAGEN PCR Buffer, 8 μ l Q-Solution, 21.6 μ l Ampuwa® (Fresenius), 2 μ l genomic DNA ($c=36\text{ng}/\mu\text{l}$ (blood samples); $110\text{ng}/\mu\text{l}$ (tissue sample)) and 1.6 μ l of each primer (10 μM). The resulting PCR amplicons were purified using the QIAquick® PCR Purification Kit (Qiagen, Santa Clarita, CA, USA) and later sequenced with the DCTS Quick Start Kit and CEQ™ 8000 Genetic Analysis System both by Beckman Coulter Inc. (Brea, CA, USA). All procedures were performed according to the manufacturer's manual.

RNA isolation and sequencing

Total RNA from endometrium of rudimentary uterine tissue was isolated using the RNeasy Mini Kit (Qiagen) and used for paired-end RNA-seq. Quality was assessed with an Agilent 2100 Bioanalyzer. Samples with high RNA integrity number (RIN > 7) were selected for library construction. Using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina and 100 ng of total RNA for each sequencing library, poly(A) selected paired-end sequencing libraries (101 bp read length) were generated according to the manufacturer's instructions. All libraries were sequenced on an Illumina NovaSeq 6000 platform at a depth of around 40 mio reads each. Library preparation and sequencing procedures were performed by the same individual, and a design aimed to minimize technical batch effects was chosen.

RNA quality control, alignment, and differential expression analysis

Read quality of RNA-seq data in fastq files was assessed using *FastQC* (v0.11.4) [14] to identify sequencing cycles with low average quality, adaptor contamination, or repetitive sequences from PCR amplification. Reads were aligned using *STAR* (v2.7.0a) [15] allowing gapped alignments to account for splicing against the *Ensembl* H. sapiens genome v95. Alignment quality was analyzed using *samtools* (v1.1) [16]. Normalized read counts for all genes were obtained using *DESeq2* (v1.26.0) [17]. Transcripts covered with less than 50 reads (median of all samples) were excluded from the analysis leaving 15,131 genes for determining differential expression. We set $|\log_2 \text{fold-change}| \geq 0.5$ and BH-adjusted p -value ≤ 0.05 to call differentially expressed genes. Gene-level abundances were derived from *DESeq2* as normalized read counts and used for calculating the \log_2 -transformed expression changes underlying the expression heatmaps for which ratios were computed against mean expression in control samples. The *sizeFactor*-normalized counts provided by *DESeq2* also went into calculating nRPKM (normalized Reads Per Kilobase per Million total reads) as a measure of relative gene expression [18]. Upstream regulators as well as predicted interactions among DEGs were derived from *Ingenuity Pathway Analysis* (IPA, v01–16, Qiagen). *Cytoscape* was used for visualizing networks [19]. Cell type-specific endometrial marker genes based on single-cell data [20].

Results

Case Reports

For our study, we recruited five pairs of discordant monozygotic (MZ) twins of which one sister each was affected with MRKH while the other sister did not show vaginal and uterus aplasia (Table 1). We obtained EDTA blood from both twins, their parents as well as unaffected siblings willing to participate. We were also able to obtain tissue from the uterine rudiment of the affected twin (except twin 3). Additionally, malformations of the kidneys were observed in four affected individuals as well as one twin sister not affected with MRKH. Additional phenotypic features such as skeletal, heart or eye malformations were observed in three affected individuals. Besides one twin sister none of the twins or siblings showed any of these additional phenotypes.

Multi-sample analysis of twin genomes for discordant variants

In a first analysis step, we performed multi-sample analyses of blood and tissue of the affected twin against blood of the unaffected twin as a control (Supplementary Table 1 and 2). This filter step gave us an average of 16 variants in the coding region as well as seven variants in the non-coding regions. Since this filtering step enriched for technical artefacts, variant quality was assessed manually using IGV and resulted in only one mosaic variant of good quality in *ACTR3B* in twin 2 (Figure 1, Table 2).

This variant (*ACTR3B* (ENST00000256001.13):c.1066G>A, p.Gly356Arg) had an approximate allele frequency of 39% in affected tissue, 10% in blood of the affected twin and only one read in blood of the unaffected twin 2-2 (Figure 1). *ACTR3B* encodes a member of the actin-related proteins and plays a role in the organization of the actin cytoskeleton. It is highly expressed in the brain and many other tissues (<https://www.proteinatlas.org/ENSG00000133627-ACTR3B/tissue>). The glycine at position 365 is highly conserved and within the actin domain. Variants in *ACTR3B* have not been implied in MRKH or genitourinary malformation before and the functional consequences of this variant remain unclear.

For the other twin pairs, multi-sample analysis of tissue against blood of the affected twin did not reveal any discordant variants.

Furthermore, we performed analysis for copy number variants and structural variants only present in the affected twin. Unfortunately, no variant with sufficient quality could be identified.

Analysis of genome data for rare conserved variants in MRKH candidate genes

Under the assumption that the MRKH phenotype might not be fully penetrant, genome data were further analyzed for rare conserved single nucleotides as well as copy number or structural variants (Supplementary Table 3) and revealed one pathogenic variant in both twins carrying a stop variant in *GREB1L* (ENST00000269218.10):c.4665T>A, p.Tyr1555* (Table 2). *GREB1L* is a target gene in the retinoic acid signaling pathway, which is highly expressed in the developing fetal human kidney and involved in early metanephros and genital development [21]. Dominant variants in *GREB1L* have been previously described for renal hypodysplasia/aplasia 3 (OMIM #617805) including uterine abnormalities and MRKH. Interestingly, the twin sister not affected with MRKH had unilateral renal agenesis. Segregation analysis showed that this variant was inherited from the healthy mother. Reduced penetrance is well known for *GREB1L* variants [21, 22]; thus, we classified this variant as pathogenic.

In twin pair 1, a heterozygous *PAX8* variant (ENST00000263334.9: c.1315G>A, p.Ala439Thr) present in both twins was identified inherited from the unaffected mother (Table 2). *PAX8* has been postulated in MRKH before [4]. The variant affects a well conserved amino acid within the paired-box protein 2C terminal domain of *PAX8* and is not listed in gnomAD. According to ACMG criteria, we classified this variant as a variant of unknown significance.

In twin pair 2, a missense variant in *WNT9B* (ENST00000290015.7:c.205C>T, p.Arg69Trp) was detected which is also present in the unaffected sister and was inherited from the father (Table 2). Variants in *WNT9B* have been implicated in MRKH before [23, 24]. This variant affects a conserved amino acid within the WNT domain and is listed in gnomAD [25] with a frequency of 0.0002. Among the individuals carrying this variant in gnomAD there is no shift between female and male carriers. Nevertheless, we cannot exclude an effect of this variant on the MRKH phenotype potentially in combination with other variants. Thus, we classified this variant as variant of unknown significance.

MRKH twins and sporadic cases showed similar endometrial transcriptome changes

To test whether the observed variants in *ACTR3B*, *GREB1L*, *PAX8*, and *WNT9B* altered their gene expression levels, we performed RNA-sequencing of uterine rudiments obtained from three affected twins (affected twin from pair 1, 2, and 4). Compared to previously published controls and sporadic

MRKH patients [13], no significant change in gene expression was observed for these genes (Supplementary Figure 1), indicating that the variants did not alter the underlying expression.

As discordant MRKH twins might be based on a distinct etiology when compared to sporadic cases, we next asked whether the overall endometrial transcriptome differs between discordant twin and sporadic cases. Therefore, we compared the three twin transcriptomes to previously published transcriptomes of 35 sporadic patients (19 type 1 and 16 type 2) as well as 25 controls [13]. Principal component analysis separated MRKH samples from controls and the three twin samples clustered together with those of sporadic cases, indicating a high similarity of transcriptomic signatures (Supplementary Figure 2). Based on single-cell data from endometrial tissue [20], cell type-specific expression of marker genes from ciliated and unciliated epithelial cells was highly similar between twins and sporadic cases (Supplementary Figure 3).

To compare these commonalities in more detail, differential expression changes were determined between MRKH twins and control samples. When compared to controls, a total of 449 differentially expressed genes (DEGs) with 235 being up- and 214 being down-regulated were identified (Figure 2A). Of those, 174 were also significantly changed in MRKH sporadic cases (Figure 2B), which in total included 2121 DEGs as previously reported [13]. Furthermore, these shared 174 DEGs showed highly similar expression changes in twins as well as sporadic samples when compared to controls (Figure 2C).

To better understand the underlying biology of the endometrial perturbances in MRKH twins, enrichment analysis was applied to identify molecules upstream of the observed DEGs. The most significant molecule observed was fulvestrant (Figure 3A), a selective estrogen receptor degrader (SERD), used as medication to treat hormone receptor (HR)-positive metastatic breast cancer in postmenopausal women with disease progression as well as HR-positive, HER2-negative advanced breast cancer in combination with palbociclib in women with disease progression after endocrine therapy [26, 27]. In line, the estrogen receptor itself was found as a significant upstream regulator (Figure 3A). Visualizing a network of DEGs around *fulvestrant* connects several key genes such as *WNT4* previously linked to MRKH (Figure 3B). Individually plotting gene expression changes of these candidates, further points at the similarity of endometrial gene expression changes between MRKH twins and sporadic cases (Figure 3C).

Taken together, our transcriptome analysis indicates that MRKH twins do not differ from other MRKH cases and that in general observation made in twins can be transferred to sporadic cases as well.

Discussion

Etiology of MRKH syndrome

The etiology of MRKH syndrome is complex, multifactorial and, to date, not well understood. Even though a number of candidate genes has been proposed, many lack functional evidence. Additionally, many familial cases show an autosomal dominant inheritance with incomplete penetrance and variable expressivity [3, 28], which complicates genetic diagnosis using segregation analysis. Especially missense variants in candidate genes are very hard to interpret in this context but also nonsense variants pose difficulties in this regard. While some studies showed that pathogenic loss of function variants in candidate genes like *PAX8*, *BMP4*, and *BMP7* are predominantly inherited by the father [4], there is no sex difference seen for loss of function variants in gnomAD for these genes. This points towards reduced penetrance and further mechanisms involved in MRKH. Studying

discordant twins poses an opportunity to investigate variants differing between the twins as well as looking for potential epigenetic marks and transcriptomic changes between these twins.

Genomic differences of monozygotic twins

Even though monozygotic twins developed from the same zygote, they do not have the exact same genome. During development a number of variants occur and may differ in variant allele frequency (VAF) between both twins depending on the timepoint and cell population of twinning. Some of these variants may be nearly constitutional with a VAF of >0.45, while others are mosaic with lower VAFs. A recent study of 381 monozygotic twins and two triplets estimated the average number of early developmental variants differing in twins to be around 5.2 [29]. The number of discordant variants observed for each twin pair had a huge variation ranging from more than 100 to no variant observed. In our study, we could only identify one mosaic variant in a total of five twin pairs and no nearly constitutional variant differing in the twins. One explanation could be that our cohort size is much smaller and may be biased towards a lower number of discordant variants. Another explanation could be that while we had an average coverage of about 43x, the study by Jonsson and colleagues had a coverage of 152x for many of their genomes leading to a higher accuracy of variant and mosaic calling.

Mosaic variant in *ACTR3B*

The only discordant variant we could identify in our twin cohort was a mosaic variant c.1066G>A, p.Gly356Arg in *ACTR3B*. *ACTR3B* encodes for an actin-related protein (ARP) involved in the organization of the cytoskeleton which has not been implicated in MRKH or any other disorder so far. GTEx expression data for several tissues shows a predominant expression of *ACTR3B* in brain and a rather low expression in uterus (<https://www.proteinatlas.org/ENSG00000133627-ACTR3B/tissue>). In contrast, protein levels of *ACTR3B* are relatively high in female tissue. During murine development it is mainly expressed in embryonal ectoderm and not mesoderm or endoderm (derived from GXD [30]), which are more crucial to the development of the genital tract. *ACTR3B* is highly intolerant to loss of function variants with a pLI-score of and the glycine at position 356 is highly conserved within the actin domain. Still, without further functional evidence and the identification of further individuals with variants in *ACTR3B* and genitourinary anomalies, the connection between this mosaic variant and the observed phenotype remains elusive.

Pathogenic *GREB1L* variant

Our analysis for rare conserved variants in candidate genes for MRKH identified a heterozygous loss of function variant c.4665T>A, p.Tyr1555* in *GREB1L*. *GREB1L* has been previously associated with a variety of genitourinary disorders (OMIM #617805) including MRKH [6, 31-33]. Many affected families present with a pedigree with autosomal dominant inheritance and reduced penetrance and variable expressivity. This is also seen in the family presented here. The mother shows no symptoms of genitourinary malformations while one daughter has MRKH and a pelvic kidney (MRKH type 2) and her twin sister has unilateral kidney agenesis, hip dysplasia and scoliosis with no evidence of MRKH. In previous reports on pathogenic *GREB1L* variants, the authors noted a bias in transmission towards maternal inheritance which might be caused by imprinting of the paternal allele [21, 32]. The mothers transmitting the variant were usually unaffected or only showed a mild phenotype. This could also be an explanation why the mother of one twin pair is unaffected but both her daughters show symptoms of *GREB1L* haploinsufficiency.

Variants of unknown significance

WNT9B

WNT9B encodes a secretory glycoprotein and paracrine factor that is part of the canonical Wnt signaling pathway, one of the main pathways responsible for organization of the mammalian urogenital system by regulating embryonic development, adult tissue homeostasis and differentiation [34]. In mice, it is mostly expressed in the developing kidney and epithelium of the Wolffian duct in both sexes, therefore playing an essential role in renal development [34].

Even though the gene is not highly expressed in the Müllerian duct, which later forms the uterus, oviduct and upper part of the vagina, posterior Müllerian duct elongation is only made possible by Wnt9b mediated interactions with the epithelium of the Wolffian duct [34, 35]. A study with *Wnt9b*^{-/-} knock-out mice corroborates the abovementioned mechanism, as males did not develop a vas deferens and epididymis and females lacked a uterus, oviduct and upper vagina, whereas the gonads appeared normal, as is the case in MRKH [34].

Multiple studies found that some patients with Müllerian duct anomalies or MRKH did indeed have variants in *WNT9B* [23, 24]. Furthermore, SNPs in this gene were linked to a higher MRKH risk [36]. However, a different study, comparing *WNT9B* variants of patients with Müllerian abnormalities to healthy controls, ruled out *WNT9B* variants as the only causative factor for Müllerian anomalies, as most of them could be found in a small number of healthy controls as well [37]. This phenomenon could also be observed in twin pair 2, which implies that *WNT9B* is not the only causative factor of MRKH, but might still be involved in its, yet to be understood, pathogenesis.

PAX8

The *PAX8* gene encodes a homeodomain signaling molecule, strongly expressed in the Müllerian duct, Wolffian duct and kidney during embryonic development [38]. Together with *PAX2*, a gene closely related and partly redundant in function to *PAX8*, they initiate the mesenchymal-epithelial transition which later leads to the formation of the Wolffian duct [39]. The same mechanism is believed to be involved in the Müllerian duct formation [35]. Animal models corroborate this theory as *Pax2*^{-/-} mice do not have kidneys nor a reproductive tract [39], only initially developing the anterior part of the Wolffian and Müllerian duct [40]. In *Pax2*^{-/-}*Pax8*^{-/-} embryos there is no Wolffian nor Müllerian duct formation at all.

The studies on *Pax8*^{-/-} knock-out mice are quite controversial. Whereas Mittag *et al.* observed only remnants of myometrial tissue instead of a uterus and the absence of a vaginal opening in *Pax8*^{-/-} mice [41], Mansouri *et al.* reported no urogenital malformations with the same knockout model [38]. Even if it is not yet clear whether a variant in *PAX8* alone is enough to disrupt the urogenital development, the gene is certainly involved in this delicate process.

Furthermore, multiple *PAX8* variants and deletions have already been described in MRKH patients [4, 42], indicating a connection between this gene and the MRKH pathogenesis.

Transcriptome analysis of MRKH twins

In addition, to our genome analysis, we also used the chance to better understand gene expression perturbations in monozygotic twins and profiled the endometrial transcriptome of the affected twins. Interestingly, the observed perturbations highly agreed with changes previously described in sporadic cases [13]. The main chemical associated with the shared DEGs was fulvestrant, a selective estrogen receptor degrader (SERD), used as medication to treat hormone receptor (HR)-positive metastatic breast cancer in postmenopausal women with disease progression as well as HR-positive,

HER2-negative advanced breast cancer in combination with palbociclib in women with disease progression after endocrine therapy [26, 27]. In addition, the estrogen receptor itself was found as a significant upstream regulator and the gene encoding for estrogen receptor 1, *ESR1*, was perturbed in affected twins and sporadic cases in line with our previous findings [43-45].

Agreeing with previous literature, estrogens are necessary for the embryonic development of the female reproductive tract. The pivotal role of *ESR1* in female reproductive tract development has been previously shown by disrupting the corresponding gene in mouse, subsequently leading to hypoplastic uterine and vaginal tissue [46]. Intriguingly, prenatal exposures of fetuses to synthetic estrogen such as diethylstilbestrol (DES) can also alter HOX gene expression in the developing Müllerian system thereby disrupting the development of the female reproductive tract [47]. In the 1970's, DES was prescribed to pregnant women to prevent miscarriages until it was realized that daughters exposed to DES in utero showed a higher incidence of Müllerian anomalies [48] and a higher prevalence of MRKH [49].

Taken together, our transcriptome analysis confirms previous transcriptome perturbations in MRKH and points at a high similarity of endometrial gene expression changes between twins and sporadic cases.

Conclusion

In this study, we examined the genetic differences of discordant twins as well as blood and affected tissue for MRKH using genome sequencing. To our knowledge, no previous study conducted genome sequencing for discordant MRKH twins or on affected tissue. Even though, our intention was to shed light on the genetic causes of MRKH by thoroughly checking for small copy number variants or structural variants missed by previous array analysis as well as identifying potential intronic or intergenic pathogenic variants, we were not able to identify any such causes neither in tissue nor as discordant variants nor as variants in candidate regions. The only clearly pathogenic variant we were able to identify was a loss of function variant in *GREB1L*. This shows that most likely previous studies of exome sequencing of blood did not miss out on a significant portion of pathogenic genetic variants. One explanation could be that the etiology of MRKH is far more complex, and incomplete penetrance and variable expressivity hamper with the identification of causative variants. Another explanation is that genetics play a minor role in the development of MRKH and epigenetic changes as well as environmental and microbiological factors should be taken more into account when looking into the etiology of MRKH.

Acknowledgements

We thank all patients who participated in the study. We thank the team of the NGS Competence Center Tübingen, foremost Nicolas Casadei and Sven Poths, for their support.

Funding

This study was supported by a project grant of the Deutsche Forschungsgemeinschaft (DFG 351381475; BR 5143/5-1, AOBJ: 639534; KO 2313/7-1, AOBJ: 639535; RI 682/15-1, AOBJ: 639536) and through funding of the NGS Competence Center Tübingen (NCCT-DFG, project 407494995). JMS-H was supported by the solveRD consortium (European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 779257).

Author contribution

R.B. and E.S. performed the genome analysis and variant interpretation. R.B. and E.S. validated variants using Sanger sequencing. T.H. and J.SH. performed the bioinformatic analysis of RNA-sequencing data. K.R. and S.Y.B. obtained patient samples and were responsible for the clinical input. A.K. processed these samples. J.SH, S.Y.B., O.K. and O.R. conceived and designed the project. N.W. and M.S. set-up the analysis pipeline. R.B, E.S., and J.SH. wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Data availability

Genome and RNA-sequencing data that support the findings of this study have been deposited in the European Genome-phenome Archive (EGA) (primary accession number pending).

References

1. Herlin, M.K., M.B. Petersen, and M. Brannstrom, *Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome: a comprehensive update*. Orphanet J Rare Dis, 2020. **15**(1): p. 214.
2. Rall, K., et al., *Typical and Atypical Associated Findings in a Group of 346 Patients with Mayer-Rokitansky-Kuester-Hauser Syndrome*. J Pediatr Adolesc Gynecol, 2015. **28**(5): p. 362-8.
3. Herlin, M., A.T. Hojland, and M.B. Petersen, *Familial occurrence of Mayer-Rokitansky-Kuster-Hauser syndrome: a case report and review of the literature*. Am J Med Genet A, 2014. **164A**(9): p. 2276-86.
4. Chen, N., et al., *Perturbations of genes essential for Mullerian duct and Wolffian duct development in Mayer-Rokitansky-Kuster-Hauser syndrome*. Am J Hum Genet, 2021. **108**(2): p. 337-345.
5. Mikhael, S., et al., *Genetics of agenesis/hypoplasia of the uterus and vagina: narrowing down the number of candidate genes for Mayer-Rokitansky-Kuster-Hauser Syndrome*. Hum Genet, 2021. **140**(4): p. 667-680.
6. Kyei Barffour, I. and R. Kyei Baah Kwarkoh, *GREB1L as a candidate gene of Mayer-Rokitansky-Kuster-Hauser Syndrome*. Eur J Med Genet, 2021. **64**(3): p. 104158.
7. Backhouse, B., et al., *Identification of Candidate Genes for Mayer-Rokitansky-Kuster-Hauser Syndrome Using Genomic Approaches*. Sex Dev, 2019. **13**(1): p. 26-34.
8. Ledig, S. and P. Wieacker, *Clinical and genetic aspects of Mayer-Rokitansky-Kuster-Hauser syndrome*. Med Genet, 2018. **30**(1): p. 3-11.
9. Pan, H.X., et al., *Detection of de novo genetic variants in Mayer-Rokitansky-Kuster-Hauser syndrome by whole genome sequencing*. Eur J Obstet Gynecol Reprod Biol X, 2019. **4**: p. 100089.
10. Lischke, J.H., C.H. Curtis, and E.J. Lamb, *Discordance of vaginal agenesis in monozygotic twins*. Obstet Gynecol, 1973. **41**(6): p. 920-4.
11. Regenstein, A.C. and A.S. Berkeley, *Discordance of mullerian agenesis in monozygotic twins. A case report*. J Reprod Med, 1991. **36**(5): p. 396-7.
12. Steinkampf, M.P., S.P. Dharia, and R.D. Dickerson, *Monozygotic twins discordant for vaginal agenesis and bilateral tibial longitudinal deficiency*. Fertil Steril, 2003. **80**(3): p. 643-5.
13. Hentrich, T., et al., *The Endometrial Transcription Landscape of MRKH Syndrome*. Front Cell Dev Biol, 2020. **8**: p. 572281.
14. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*. 2010.

15. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner*. *Bioinformatics*, 2013. **29**(1): p. 15-21.
16. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. *Bioinformatics*, 2009. **25**(16): p. 2078-9.
17. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. *Genome Biol*, 2014. **15**(12): p. 550.
18. Srinivasan, K., et al., *Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses*. *Nat Commun*, 2016. **7**: p. 11295.
19. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. *Genome Res*, 2003. **13**(11): p. 2498-504.
20. Wang, W., et al., *Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle*. *Nat Med*, 2020. **26**(10): p. 1644-1653.
21. De Tomasi, L., et al., *Mutations in GREB1L Cause Bilateral Kidney Agenesis in Humans and Mice*. *Am J Hum Genet*, 2017. **101**(5): p. 803-814.
22. Schrauwen, I., et al., *Autosomal Dominantly Inherited GREB1L Variants in Individuals with Profound Sensorineural Hearing Impairment*. *Genes (Basel)*, 2020. **11**(6).
23. Wang, M., et al., *Analysis of WNT9B mutations in Chinese women with Mayer-Rokitansky-Kuster-Hauser syndrome*. *Reprod Biomed Online*, 2014. **28**(1): p. 80-5.
24. Waschke, D.E., et al., *Mutations in WNT9B are associated with Mayer-Rokitansky-Kuster-Hauser syndrome*. *Clin Genet*, 2016. **89**(5): p. 590-6.
25. Karczewski, K.J., et al., *The mutational constraint spectrum quantified from variation in 141,456 humans*. *Nature*, 2020. **581**(7809): p. 434-443.
26. Carlson, R.W., *The history and mechanism of action of fulvestrant*. *Clin Breast Cancer*, 2005. **6 Suppl 1**: p. S5-8.
27. Nathan, M.R. and P. Schmid, *A Review of Fulvestrant in Breast Cancer*. *Oncol Ther*, 2017. **5**(1): p. 17-29.
28. Kyei-Barffour, I., et al., *The Embryological Landscape of Mayer-Rokitansky-Kuster-Hauser Syndrome: Genetics and Environmental Factors*. *Yale J Biol Med*, 2021. **94**(4): p. 657-672.
29. Jonsson, H., et al., *Differences between germline genomes of monozygotic twins*. *Nat Genet*, 2021. **53**(1): p. 27-34.
30. Ringwald, M., et al., *A database for mouse development*. *Science*, 1994. **265**(5181): p. 2033-4.
31. Brophy, P.D., et al., *A Gene Implicated in Activation of Retinoic Acid Receptor Targets Is a Novel Renal Agenesis Gene in Humans*. *Genetics*, 2017. **207**(1): p. 215-228.
32. Herlin, M.K., et al., *Whole-exome sequencing identifies a GREB1L variant in a three-generation family with Mullerian and renal agenesis: a novel candidate gene in Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome. A case report*. *Hum Reprod*, 2019. **34**(9): p. 1838-1846.
33. Sanna-Cherchi, S., et al., *Exome-wide Association Study Identifies GREB1L Mutations in Congenital Kidney Malformations*. *Am J Hum Genet*, 2017. **101**(5): p. 789-802.
34. Carroll, T.J., et al., *Wnt9b plays a central role in the regulation of mesenchymal to epithelial transitions underlying organogenesis of the mammalian urogenital system*. *Dev Cell*, 2005. **9**(2): p. 283-92.
35. Kobayashi, A. and R.R. Behringer, *Developmental genetics of the female reproductive tract in mammals*. *Nat Rev Genet*, 2003. **4**(12): p. 969-80.
36. Ma, W., et al., *Associations of Polymorphisms in WNT9B and PBX1 with Mayer-Rokitansky-Kuster-Hauser Syndrome in Chinese Han*. *PLoS One*, 2015. **10**(6): p. e0130202.
37. Tang, R., et al., *WNT9B in 542 Chinese women with Mullerian duct abnormalities: mutation analysis*. *Reprod Biomed Online*, 2014. **28**(4): p. 503-7.
38. Mansouri, A., K. Chowdhury, and P. Gruss, *Follicular cells of the thyroid gland require Pax8 gene function*. *Nat Genet*, 1998. **19**(1): p. 87-90.
39. Bouchard, M., P. Pfeffer, and M. Busslinger, *Functional equivalence of the transcription factors Pax2 and Pax5 in mouse development*. *Development*, 2000. **127**(17): p. 3703-13.

40. Torres, M., et al., *Pax-2 controls multiple steps of urogenital development*. *Development*, 1995. **121**(12): p. 4057-65.
41. Mittag, J., et al., *Congenital hypothyroid female pax8-deficient mice are infertile despite thyroid hormone replacement therapy*. *Endocrinology*, 2007. **148**(2): p. 719-25.
42. Smol, T., et al., *Mayer-Rokitansky-Kunster-Hauser syndrome due to 2q12.1q14.1 deletion: PAX8 the causing gene?* *Eur J Med Genet*, 2020. **63**(4): p. 103812.
43. Brucker, S.Y., et al., *Decidualization is Impaired in Endometrial Stromal Cells from Uterine Rudiments in Mayer-Rokitansky-Kuster-Hauser Syndrome*. *Cell Physiol Biochem*, 2017. **41**(3): p. 1083-1097.
44. Brucker, S.Y., et al., *Sequence variants in ESR1 and OXTR are associated with Mayer-Rokitansky-Kuster-Hauser syndrome*. *Acta Obstet Gynecol Scand*, 2017. **96**(11): p. 1338-1346.
45. Rall, K., et al., *A combination of transcriptome and methylation analyses reveals embryologically-relevant candidate genes in MRKH patients*. *Orphanet J Rare Dis*, 2011. **6**: p. 32.
46. Masse, J., et al., *The developing female genital tract: from genetics to epigenetics*. *Int J Dev Biol*, 2009. **53**(2-3): p. 411-24.
47. Block, K., et al., *In utero diethylstilbestrol (DES) exposure alters Hox gene expression in the developing mullerian system*. *FASEB J*, 2000. **14**(9): p. 1101-8.
48. Kaufman, R.H., et al., *Continued follow-up of pregnancy outcomes in diethylstilbestrol-exposed offspring*. *Obstet Gynecol*, 2000. **96**(4): p. 483-9.
49. Wautier, A., et al., *Genital tract and reproductive characteristics in daughters of women and men prenatally exposed to diethylstilbestrol (DES)*. *Therapie*, 2020. **75**(5): p. 439-448.

Figure Legends

Figure 1. Mosaic missense variant in *ACTR3B* identified in blood and tissue of twin 2-1

IGV reads show that the variant c.1066G>A, p.Gly356Arg in *ACTR3B* has an allele frequency of about 39% in uterine tissue and an allele frequency of about 11% in blood of the affected twin 2-1, while blood of twin 2-2 only has 1 supporting read.

Figure 2. Endometrial transcriptomes of MRKH twins and sporadic cases show similar changes.

- (A) Schematic diagram of comparisons between MRKH monozygotic twins (MZT), MRKH sporadic cases (SP) and unaffected women (controls) indicating number of differentially expressed genes (DEGs). DEGs between sporadic cases and controls based on previous work [13]. Fold change and significance cut-offs below.
- (B) Venn diagram showing number of common DEGs MRKH twins and sporadic cases each compared to controls.
- (C) Expression profiles (\log_2 expression change relative to *Ctrl* group) of 174 DEGs (common DEGs indicated in Figure 2B) across all samples. Rows hierarchically clustered by Euclidian distance and *ward.D2* method. Cycle information (proliferative or secretory) and patient type (monozygotic twin, sporadic, or control) on top.

Figure 3. Gene expression changes in MRKH monozygotic twins point at regulators linked to estrogen receptor.

- (A) Predicted upstream regulators for common 174 DEGs (from Figure 1B) based on Ingenuity Pathway Analysis. Top five significant regulators shown.
- (B) Network of DEGs associated with upstream regulator *fulvestrant*. Interactions based on Ingenuity Pathway Analysis with line width indicating number of curated interactions. Genes color-coded by mean expression change observed in MZT / CTRL on top and SP / CTRL separately for type 1 and type 2 below.
- (C) Expression levels for the selected DEGs plotted as individual data points with mean \pm SEM.

Supplementary Figure 1. No significant change in genes expression observed for *ACTR3B*, *GREB1L*, *PAX8*, and *WNT9B*

Expression levels for the selected DEGs plotted as individual data points with mean \pm SEM.

Supplementary Figure 2. Both MRKH twin and sporadic samples separated from control tissue in unaffected individuals.

Principal component analysis of endometrial gene expression profiles for three affected MRKH twins, 35 sporadic MRKH patients (19 type 1 and 16 type 2) as well as 25 unaffected controls. Axis percentages indicate variance contribution.

Supplementary Figure 3. Gene expression of cell type-specific markers genes similar between MRKH twins, sporadic cases and unaffected controls.

Cell type-specific gene expression per sample for unciliated and ciliated epithelium. Boxplots show geometric mean as well as 10th, 25th, 75th, and 90th quantile of expression values for all genes classified based on single-cell data of the human endometrium [20]. Number of genes per cell type in brackets.

Table 1 Phenotype of individuals in this study

Twin pair	Individual	Age at surgery (years) indicated as range	MRKH type	Kidney malformation	Skeletal malformation
1	1-1	26-30	MRKH 2	yes	yes
2	2-1	16-20	MRKH 2	yes	
3	3-1	31-35	MRKH 2	yes	
3	3-2	-	-	yes	yes
4	4-1	16-20	MRKH 1	-	
5	5-1	16-20	MRKH 2	yes	

Table 2 Identified variants in SNV analysis

Analysis type	Sample	MRKHS classification	Zygoty	Transcript (ENST-number)	Variant reads	Total reads	Gene	cDNA change	Protein change	phyloP	gnomAD-allele frequency	Inheritance
Single sample analysis	Tissue MRKH-twin 1	MRKH-I	het	ENST00000263334.9	14	37	PAX8	c.1315G>A	p.Ala439Thr	42.050	0	AD
Single sample analysis	Tissue MRKH-twin 2	MRKH-I	het	ENST00000290015.7	20	31	Wnt9B	c.205C>T	p.Arg69Trp	27.030	0.0002180	AD
Single sample analysis	Tissue MRKH-twin 3	MRKH-II	het	ENST00000269218.10	33	51	GREB1L	c.4665T>A	p.Tyr1555Ter	- 0.7730	0	AD
Multi sample analysis	Tissue MRKH-twin 2	MRKH-I	het	ENST00000256001.13	18	46	ACTR3B	c.1066G>A	p.Gly356Arg	75.720	0	AD
	Blood MRKH-twin 2	MRKH-I	het		6	56						
	Blood healthy-twin 2	MRKH-I	wt		1	47	-	-	-	-	-	-

This table illustrates the SNVs found in the various analyses.

The abbreviations are as follows: AD - autosomal dominant | het – heterozygous | n/a - not annotated | wt - wild type

Figure 1

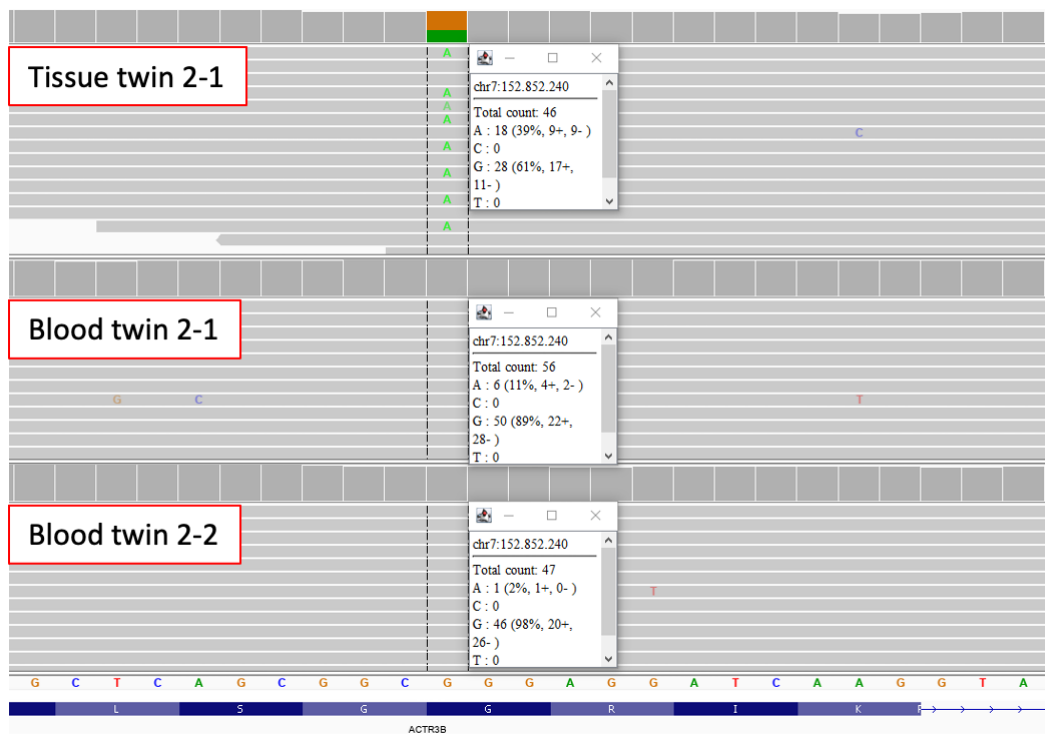
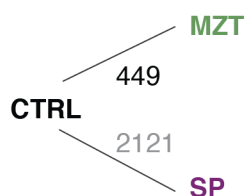


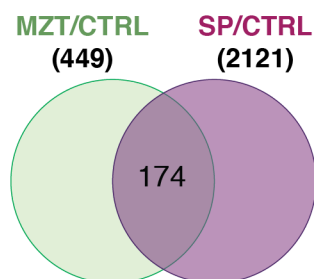
Figure 2

A



$$P_{adj} \leq 0.05 \wedge |\log_2 FC| \geq 0.5$$

B



C

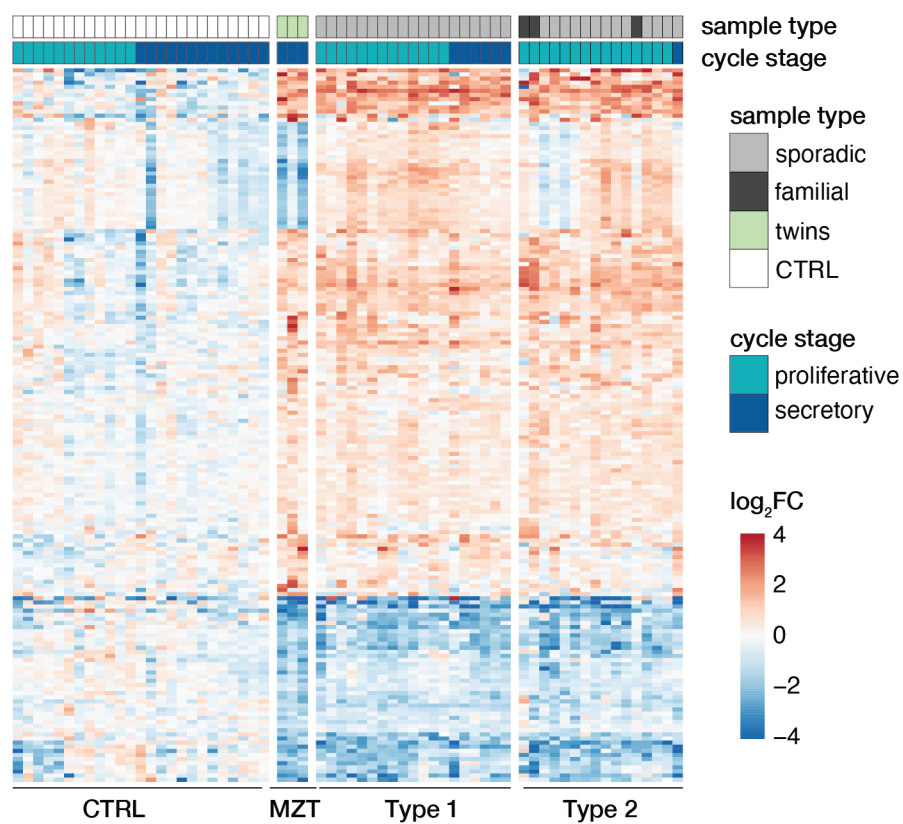
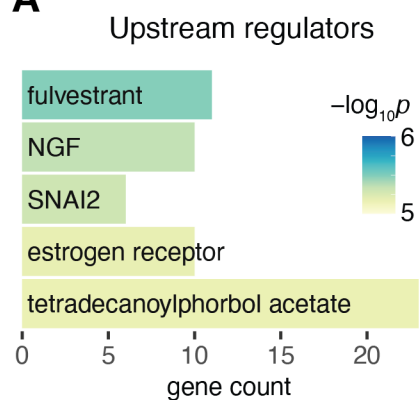
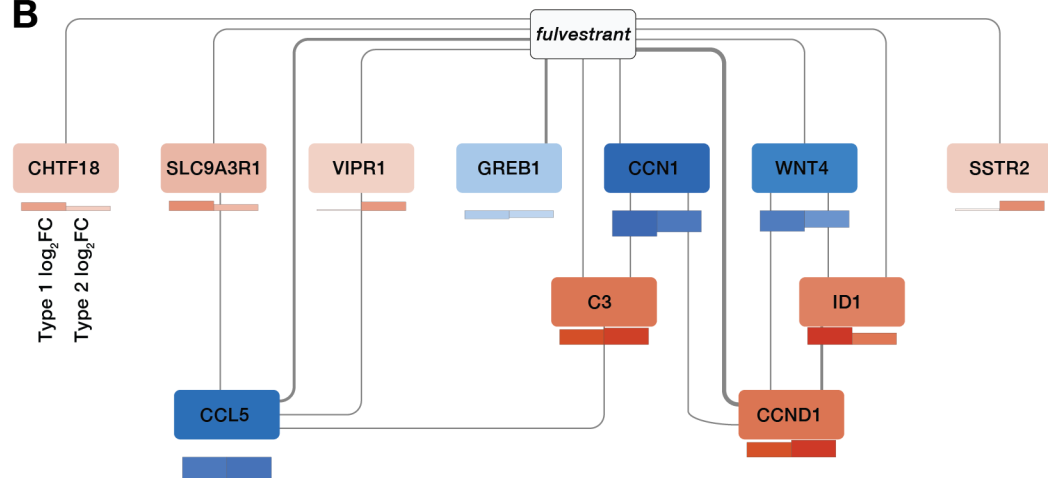


Figure 3

A



B



C

