Nephrol Dial Transplant (2022) 37: 1833–1843
doi: 10.1093/ndt/gfab253
Advance Access publication 2 September 2021

ORIGINAL ARTICLE

Whole-exome sequencing identifies *FOXL2*, *FOXA2* and *FOXA3* as candidate genes for monogenic congenital anomalies of the kidneys and urinary tract

Bixia Zheng[1,2], Steve Seltzsam[1], Chunyan Wang[1], Luca Schierbaum[1], Sophia Schneider[1], Chen-Han Wilfred Wu[1], Rufeng Dai[1], Dervla M. Connaughton[1], Makiko Nakayama[1], Nina Mann[1], Natasa Stajic[3], Shrikant Mane[4], Stuart B. Bauer[5], Velibor Tasic[6], Hyun Joo Nam[7], Shirlee Shril[1] and Friedhelm Hildebrandt[1]

[1]Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA, [2]Nanjing Key Laboratory of Pediatrics, Children's Hospital of Nanjing Medical University, Nanjing, China, [3]Department of Pediatric Nephrology, Institute for Mother and Child Health Care, Belgrade, Serbia, [4]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA, [5]Department of Urology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA, [6]Medical Faculty of Skopje, University Children's Hospital, Skopje, Macedonia and [7]Department of Biological and Environmental Science, Texas A&M University at Commerce, Commerce, TX, USA

Correspondence to: Friedhelm Hildebrandt; E-mail: friedhelm.hildebrandt@childrens.harvard.edu

ABSTRACT

Background. Congenital anomalies of the kidneys and urinary tract (CAKUT) constitute the most common cause of chronic kidney disease in the first three decades of life. Variants in four Forkhead box (*FOX*) transcription factors have been associated with CAKUT. We hypothesized that other *FOX* genes, if highly expressed in developing kidneys, may also represent monogenic causes of CAKUT.

Methods. We here performed whole-exome sequencing (WES) in 541 families with CAKUT and generated four lists of CAKUT candidate genes: (A) 36 *FOX* genes showing high expression during renal development, (B) 4 *FOX* genes known to cause CAKUT to validate list A, (C) 80 genes that we identified as unique potential novel CAKUT candidate genes when performing WES in 541 CAKUT families and (D) 175 genes identified from WES as multiple potential novel CAKUT candidate genes.

Results. To prioritize potential novel CAKUT candidates in the *FOX* gene family, we overlapped 36 *FOX* genes (list A) with lists C and D of WES-derived CAKUT candidates. Intersection with list C identified a *de novo FOXL2* in-frame deletion in a patient with eyelid abnormalities and ureteropelvic junction obstruction, and a homozygous *FOXA2* missense variant in a patient with horseshoe kidney. Intersection with list D identified a heterozygous *FOXA3* missense variant in a CAKUT family with multiple affected individuals.

Conclusions. We hereby identified *FOXL2*, *FOXA2* and *FOXA3* as novel monogenic candidate genes of CAKUT, supporting the utility of a paralog-based approach to discover mutated genes associated with human disease.

Keywords: congenital anomalies of the kidney and urinary tract, FOXL2, FOXA2, FOXA3, whole-exome sequencing

INTRODUCTION

Congenital anomalies of the kidneys and urinary tract (CAKUT) constitute the most frequent cause of chronic kidney disease in the first three decades of life [1]. Embryonic kidney development is a very complex biological process that is regulated precisely by a network composed of many genes and signaling pathways in time and space. Most gene products that cause CAKUT in humans or mice, if altered, are transcription factors or are otherwise involved in protein–protein interactions that form large transcription complexes. To date, 40 monogenic genes have been identified to cause CAKUT in humans, explaining 15–20% of CAKUT patients [2–4]. This underlines that up to 80% of CAKUT cases are genetically unsolved. Thus, we surmised that novel genes still remain to be discovered.

The forkhead (FH) box (*FOX*) transcription factor family of genes, which is characterized by the presence of an evolutionary conserved 'forkhead' or 'winged-helix' DNA-binding domain, engages in diverse functions during development as well as maintaining homeostasis of adult tissues [5, 6]. FH is originally named after the *Drosophila melanogaster* (Dm) gene *forkhead*, whose absence causes a characteristic 'forked head' appearance resulting from the homeotic transformation of the foregut into a head structure [7]. More than forty *FOX* genes have been identified in humans and mice [5]. The ∼100-residue FH DNA-binding domain of Fox proteins is

KEY LEARNING POINTS

What is already known about this subject?

- Whole-exome sequencing (WES) is a powerful tool that has helped to identify monogenic causes of congenital anomalies of the kidneys and urinary tract (CAKUT).
- Forty monogenic genes have been identified to cause CAKUT, explaining 15–20% of CAKUT patients.
- The forkhead box (*FOX*) transcription factor family of genes, which is characterized by the presence of an evolutionary conserved 'forkhead' or 'winged-helix' DNA-binding domain, engages in diverse functions during development.

What this study adds?

- By overlapping 36 highly expressed *FOX* genes in developmental kidneys with potential CAKUT candidate genes resulting from unbiased WES in 591 CAKUT patients, we identified *FOXL2*, *FOXA2* and *FOXA3* as potential novel candidate genes of monogenic CAKUT.
- We identified a *de novo FOXL2* in-frame deletion variant in a patient with eyelid abnormalities and left ureteropelvic junction obstruction, a homozygous *FOXA2* missense variant in a patient with horseshoe kidney and a heterozygous *FOXA3* missense variant in a CAKUT family with multiple affected individuals.

What impact this may have on practice or policy?

- We identified *FOXA2* and *FOXA3* as novel monogenic candidates of CAKUT. We also provided further evidence for renal phenotypic expansion in *FOXL2* variants related to blepharophimosis, ptosis and epicanthus inversus syndrome.
- Our study supports the utility of a paralog-based approach to discover mutated genes associated with human disease.

remarkably conserved across all members of the *FOX* gene family (Supplementary data, Figure S1).

The Online Mendelian Inheritance in Man (OMIM) database lists 15 *FOX* genes that, if mutated, cause monogenic human diseases in both a dominant and a recessive manner (Supplementary data, Figure S1). Notably, of these 15 genes, variants in 4 genes (*FOXP1*, *FOXC1*, *FOXF1* and *FOXC2*) have been associated with isolated or syndromic CAKUT in humans [2, 8–10]. In addition, according to data in the Mouse Genome Informatics database (http://www.informatics.jax.org/), a knockout mouse model of nine *FOX* transcription factors showed renal and urinary defects, among which six *FOX* genes led to CAKUT phenotypes (Supplementary data, Table S1). In light of these findings and the correlation between *FOX* gene paralogs and human disease phenotypes, we hypothesized that other members of the *FOX* transcription factors that are highly expressed in developmental kidneys may also represent monogenic causes of CAKUT similar to the known human monogenic CAKUT genes (*FOXP1*, *FOXC1*, *FOXF1* and *FOXC2*). By whole-exome sequencing (WES) analyses in 541 families with CAKUT, we identified *FOXA2* and *FOXA3* as novel monogenic candidates of CAKUT. We also provided further evidence for renal phenotypic expansion for *FOXL2* variants in the blepharophimosis, ptosis and epicanthus inversus syndrome (BPES, OMIM#110100).

MATERIALS AND METHODS

Study participants

This study was approved by the institutional review board (IRB) of Boston Children's Hospital as well as the IRBs of institutions at which families with CAKUT were recruited after obtaining and archiving written informed consent from January 2010 to January 2019. About 680 affected individuals from 541 unrelated families were enrolled and had WES performed on their DNA samples. All patients with CAKUT were referred to us by their pediatric nephrologist or urologist, who had made a clinical diagnosis of CAKUT on the basis of renal imaging studies. CAKUT was defined as demonstration of any abnormality of number, size, shape or anatomic position of the kidneys, gonads or other parts of the genital urinary tract that included at least one of the following: renal agenesis, renal hypo/dysplasia, multicystic dysplastic kidney, hydronephrosis, ureteropelvic junction obstruction, hydroureter, vesicoureteral reflux (VUR), ectopic or horseshoe kidney, duplex collecting system, ureterovesical junction obstruction, epi/hypospadias, posterior urethral valves and cryptorchidism.

Whole-exome sequencing and variant calling

WES was performed as previously described [11]. In brief, genomic DNA was isolated from blood lymphocytes or saliva samples and subjected to exome capture using Agilent SureSelect human exome capture arrays (Life Technologies) followed by next-generation sequencing on the Illumina HighSeq sequencing platform. Sequence reads were mapped to the human reference genome assembly (NCBI build 37/hg19), and variants were called using CLC Genomics Workbench™ (version 6.5.2) software (CLC Bio, Aarhus, Denmark).

Variant filtering to identify novel monogenic causes of CAKUT

Variant analysis was performed under recessive, dominant or *de novo* models, as previously published [4]. Variant analysis was performed by geneticists and cell biologists, who had knowledge regarding clinical phenotypes, pedigree structure and genetic mapping, and in line with proposed guidelines [12]. Sequence variants remaining after WES evaluation were examined for segregation. Filtering was performed to retain only alleles with a minor allele frequency (MAF) <0.1%, a widely accepted cutoff for autosomal dominant disorders. MAF was estimated using combined datasets incorporating all available data from the 1000 Genomes Project, the Exome Variant Server (EVS) project, dbSNP145, the Exome Aggregation Consortium and genome aggregation database (gnomAD). We filtered to retain variants with a probability of being loss-of-function intolerant (pLI) score of >0.3 based on a dominant hypothesis. To predict deleteriousness of variants, we used the University of California, Santa Cruz Human Genome Browser for the presence of paralogous genes, pseudogenes or misalignments, then scrutinized all variants with MAF <0.1% within the sequence alignments of the CLC Genomic Workbench™ software program and employed other web-based programs (see 'Web resources'). Variants were confirmed by Sanger sequencing and for segregation of phenotype with genotype.

Homozygosity mapping

Homozygosity mapping (HM) was performed based on WES data. In brief, aligned BAM files were processed using Picard and SAMtools as previously described [11]. Single nucleotide variant calling was performed using the Genome Analysis Tool Kit [13]. The resulting VCF files were used to generate HM data and visual outputs using the program Homozygosity Mapper [14].

Screening for variants in known monogenic causes of CAKUT

We evaluated WES data for causative pathogenic variants in the 40 monogenic genes that are currently known to cause non-syndromic CAKUT, and in the currently known 179 monogenic genes for syndromic CAKUT (Supplementary data, Tables S2 and S3).

Definition of potential novel unique and multiple candidate genes of CAKUT

As previously described [4], if no causative variants were found in a known isolated and syndromic CAKUT gene, an analysis toward identification of potential novel candidate genes for CAKUT was applied by WES based on the hypothesis (pedigree structure; homozygosity). If no single gene per family could be prioritized on the basis of genetic criteria, multiple candidate genes were kept in the family.

Consideration of structural data and evolutionary conservation for variant evaluation

Protein domain structure depictions and evaluation was based on the UniProt (Universal Protein Resource) database. Orthologous proteins used to evaluate evolutionary conservation were obtained from the Ensemble Genome Browser and were aligned using the Clustal Omega multiple sequence alignment tool (EMBL-EBI).

Predictions of FH domain variants on DNA binding and protein stability

FOXL2 mutant p.Asn105del and *FOXA3* mutant p.Arg155Gln were modeled based on the structure model-FOXA3/DNA complex (PDB ID 1VTN) [15]. To understand the effect of the two variants in the FH domain on DNA binding and protein stability, free energies (ΔGs) of intra-protein interaction and protein–DNA interaction were calculated using the program FoldX. Supplementary data, Table S4 presents the calculated ΔΔGs from the wild-type (WT).

Web resources

1000 Genomes Browser, http://browser.1000genomes.org
Clustal Omega, http://www.ebi.ac.uk/Tools/msa/clustal
Ensembl Genome Browser, http://www.ensembl.org
EVS, http://evs.gs.washington.edu/EVS
gnomAD, http://gnomad.broadinstitute.org
HGMD Professional 2016.3, https://portal.biobase-international.com/hgmd
Homozygosity Mapper, http://www.homozygositymapper.org/
Human fetal Kidney Atlas, https://home.physics.leidenuniv.nl/~semrau/humanfetalkidneyatlas/
MutationTaster, http://www.mutationtaster.org
OMIM, http://www.omim.org
Phylogeny.fr, http://www.phylogeny.fr/index.cgi
Polyphen2, http://genetics.bwh.harvard.edu/pph2
Sorting Intolerant from Tolerant (SIFT), http://sift.jcvi.org
UCSC genome browser, http://genome.ucsc.edu/cgi-bin/hgGateway
UniProt Consortium, http://www.uniprot.org/

RESULTS

In order to identify potential novel monogenic CAKUT candidates from unbiased WES evaluations in 680 patients from 541 unrelated CAKUT families, we first evaluated the 43 known human *FOX* gene family members for distinct temporal-spatial single-cell mRNA expression pattern, by searching the single-cell transcriptomics data of human fetal kidney at 16 weeks of gestation [16]. Thirty-six of these 43 genes showed expression levels similar to or higher than the expression levels of the four known CAKUT *FOX* genes in human developing kidneys (Figure 1A, Supplementary data, Figure S2). We set the four known CAKUT *FOX* genes as positive controls in this cohort (Figure 1B). We then validated the renal single-cell transcriptomics expression derived list (A) by showing that all four *FOX* genes known to cause human CAKUT (list B) were part of the list (Figure 1A and B). This strengthened our hypothesis that other *FOX* genes that are also transcription factors are good candidate genes for CAKUT. To prioritize potential novel CAKUT genes in *FOX* gene family members, we then examined these 36 *FOX* genes (Figure 1A) for overlap with two lists (lists

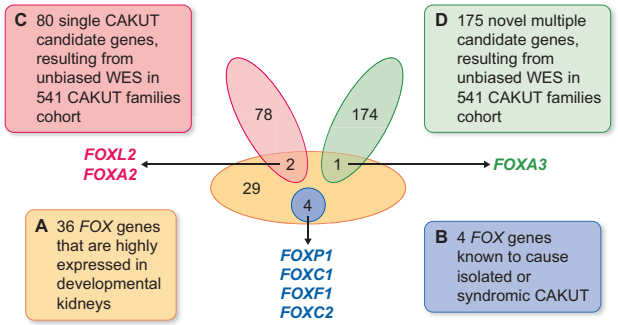

FIGURE 1: Overlapping of *FOX* gene family members (list A) that are highly expressed in human developmental kidneys with candidate genes (lists C and D) resulting from unbiased WES evaluations in 680 patients from 541 unrelated CAKUT families. (**A**) A list of 36 *FOX* gene family members that exhibited high expression in developing kidney in single-cell RNA-seq database [16] was generated as a candidate gene list for CAKUT. (**B**) Four *FOX* genes (*FOXP1*, *FOXC1*, *FOXF1* and *FOXC2*) that are known to cause CAKUT in humans were used as 'positive control' for the candidate hypothesis that highly expressed developmental kidney genes (A, yellow) represent good candidate genes for CAKUT. Note that all four genes (blue) do overlap with the candidates. (**C**) Overview of 80 novel single CAKUT candidate genes (red oval) resulting from unbiased WES in 541 families with CAKUT. Two *FOX* genes (*FOXL2* and *FOXA2*) overlap with the list. (**D**) Overview of 175 novel multiple candidate genes (green oval) resulting from unbiased WES in 541 families with CAKUT. *FOXA3* overlaps with this list.

C and D in Figure 1) of independent CAKUT candidate genes resulting from our unbiased WES evaluations in 541 unrelated CAKUT families.

Intersection with 80 novel single CAKUT candidate genes resulting from unbiased WES evaluations that we performed in 541 unrelated CAKUT families (Figure 1C) identified *FOXL2* as a phenotypic expansion for BPES, and *FOXA2* as a novel recessive candidate (Figure 1A and C). Intersection with 175 novel multiple CAKUT candidate genes resulting from our unbiased WES evaluations that we performed in 541 unrelated CAKUT families (Figure 1D) identified *FOXA3* as a novel dominant candidate (Figure 1A and D).

Identification of a *de novo* FOXL2 in-frame deletion variant in a patient with eyelid abnormalities and CAKUT

In family B3061, by trio WES, we identified a *de novo* in-frame deletion (c.313_315delAAC; p.Asn105del) in *FOXL2* in a patient who had presented with blepharophimosis and ptosis, combined with left ureteropelvic junction obstruction (UPJO) (Table 1, Figure 2). Sanger sequencing of all family members confirmed the variant to be *de novo* (Figure 2D). The variant was absent from the control database gnomAD. The Asn105 in-frame deletion was located in the FH domain and evolutionarily well conserved from *Homo sapiens* to Dm (Figure 2E). To understand the effect of the Asn105 deletion on DNA binding and protein stability, a mutant model was built based on the structure model-FOXA3/DNA complex using FoldX. The computational energy calculation of the mutant model showed that the deletion of Asn105 is expected to highly destabilize the protein and the protein–DNA interaction (Supplementary data, Table S4, Figure 2F). Heterozygous (Het) *FOXL2* gene variants are known to cause BPES (OMIM#110100), which is a rare autosomal dominant disorder.

Identification of a homozygous FOXA2 missense variant in a patient with CAKUT

Individual B998-21 was a boy of Arabic descent, who was diagnosed with horseshoe kidney at age 3 years (Table 1, Figure 3A and B). As his parents were consanguineous and unaffected, we hypothesized that the cause of CAKUT in this patient was a recessively inherited homozygous (Hom) variant (Figure 3A). HM yielded 49 segments of homozygosity with a total cumulative genomic length of 224.8 Mb (Figure 3C). Following WES evaluation, three potentially deleterious Hom variants were detected within the Hom peaks. No Hom truncating variant was identified. Taking into consideration the results from renal expression, mouse model phenotype, implicated pathway and literature review, the gene *FOXA2* was considered as the strongest candidate gene for a potential role in the development of CAKUT in our patient (please see Supplementary data, Table S5). Post-WES, a detailed history and medical chart review revealed that he has extra-renal features including recurrent infection with otitis media and pharyngitis, asthma, hyperopia, astigmatism and glucose-6-phosphate dehydrogenase deficiency (G6PD). At 11 years of age, he was in the 19% percentile for height and 92% percentile for weight. Re-analysis of his WES data identified a hemizygous variant in *G6PD* (NM_000402.3: c.653C>T; p.Ser218Phe) as the cause of his G6PD phenotype, which has been classified as pathogenic in Clinvar.

The Hom missense variant (c.155T>A; p.Met52Lys) in *FOXA2* was identified in exon 2 (NM_021784.4). The Met52 residue is located within the transcriptional activation domain (TAD) of the protein (Figure 3E). This variant was absent from

Table 1. Variants of *FOX* gene family members in three families with CAKUT

Family	Gene	Genomic position	Transcript position	Amino acid change (Het/Hom)	Segregation	gnomAD (Hom/Het/WT allele count)	PPH2	SIFT	Mutation Taster	Mm	Gg	Xt	Dr	Ci	Ce	Dm	Sc	CAKUT phenotype	Extra-renal phenotype	Gender	Ethnicity
B3061	*FOXL2*	chr3:138,665,250-138,665,252	c.313_315delAAC	p.Asn105del (Het)	*De novo*	0/0/Never reported	In-frame deletion			N	N	N	N	N	N	N	/	Left UPJO	Blepharophimosis, bilateral ptosis	M	Albanian
B998	*FOXA2*	chr20:22,563,725	c.155T>A	p.Met52Lys (Hom)	No data	0/0/Never reported	0.99	D	DC	M	M	M	M	/	Y	Q	/	Unilateral Horseshoe kidney	Recurrent infection with otitis media and pharyngitis, asthma, overweight, hyperopia, and astigmatism, G6PD	M	Arabic
A3404	*FOXA3*	chr19:46,375,727	c.464G>A	p.Arg155Gln (Het)	Affected father	0/6/282838	1.00	D	DC	R	R	R	R	R	K	R	K	Left UVJO	None specified	M	Serbian

Ce, *Caenorhabditis elegans*; Ci, *Ciona intestinalis*; D, deleterious; DC, disease causing; F, female; gnomAD (https://gnomad.broadinstitute.org/); Gg, *Gallus gallus*; M, male; Mm, *Mus musculus*; PPH2 score, PolyPhen-2 prediction score (0.0–1.0; i.e. tolerated to deleterious; variants from 0.85 to 1 are more confidently predicted to be damaging) (http://genetics.bwh.harvard.edu/pph2/); SIFT (https://sift.bii.a-star.edu.sg/); T, tolerated; Xt, *Xenopus tropicalis*. Red background represents deleterious prediction by the *in silico* algorithm. Green background indicates the mutated amino acid is well conserved among species by clustal analysis. Black background represents deletion. The genomic coordinates are based on genome build GRCh37 (hg19).

the gnomAD database and yielded predominantly deleterious prediction scores by three algorithms (PolyPhen-2, MutationTaster and SIFT) (Table 1). The conservation of the Met residue in position 52 across evolution shows it is well conserved from *H. sapiens* to *Danio rerio* (Dr; Table 1).

Identification of a Het FOXA3 missense variant in a CAKUT family with multiple affected individuals

Individual A3404-21 was a Serbian boy, who was diagnosed with left ureterovesical junction obstruction (UVJO) at birth (Table 1, Figure 4A). As his father presented with right renal agenesis and left VUR, we hypothesized that the cause of CAKUT in this family was a dominantly inherited variant (Figure 4A). Following trio analysis, we detected four potentially disease-causing Het missense variants dominantly inherited from the affected father (Figure 4B). No Het truncating variants were identified. Taking into consideration the results from renal expression, mouse model phenotype, implicated pathway and literature review, the gene *FOXA3* was considered as the strongest candidate gene for a potential role in the development of CAKUT in this family (please see Supplementary data, Table S6). Post-WES, no extra-renal features were noted in the affected individuals.

The Het missense variant (c.464G>A; p.Arg155Gln) in *FOXA3* was identified in exon 2 (NM_004497.2). The Arg155 residue is located within the FH domain of the protein (Figure 4C). This variant occurred six times heterozygously in the gnomAD database and yielded predominantly deleterious prediction scores by three algorithms (PolyPhen-2, MutationTaster and SIFT) (Table 1). As shown in Figure 4C, the Arg155 residue in FOXA3 is well conserved with Arg or the same positive charged Lys across *H. sapiens* to *Saccharomyces cerevisiae* (Sc). To understand the effect of the missense change in the FH domain on DNA binding and protein stability, a mutant model was built based on the structure model-FOXA3/DNA complex using FoldX. The computational energy calculation showed that the p.Arg155Gln mutant is predicted destabilizing for the protein but does not affect the protein–DNA interaction (Supplementary data, Table S4, Figure 4D).

Screening for variants in known monogenic causes of CAKUT

All families were screened for genes known to cause isolated or syndromic CAKUT, if mutated (Supplementary data, Tables S2 and S3). However, no likely causative Het or biallelic variants were detected in those genes.

DISCUSSION

The *FOX* gene family of transcriptional regulators is an evolutionarily ancient gene family that regulates diverse biological processes during development. Taking the following into consideration: (i) members of paralogous gene families often share molecular functions, (ii) 36 *FOX* genes have overlapping temporal-spatial expression patterns in developing kidney in single-cell transcriptomics data (Supplementary data, Figure S2) and (iii) four *FOX* genes (*FOXP1*, *FOXC1*, *FOXF1*

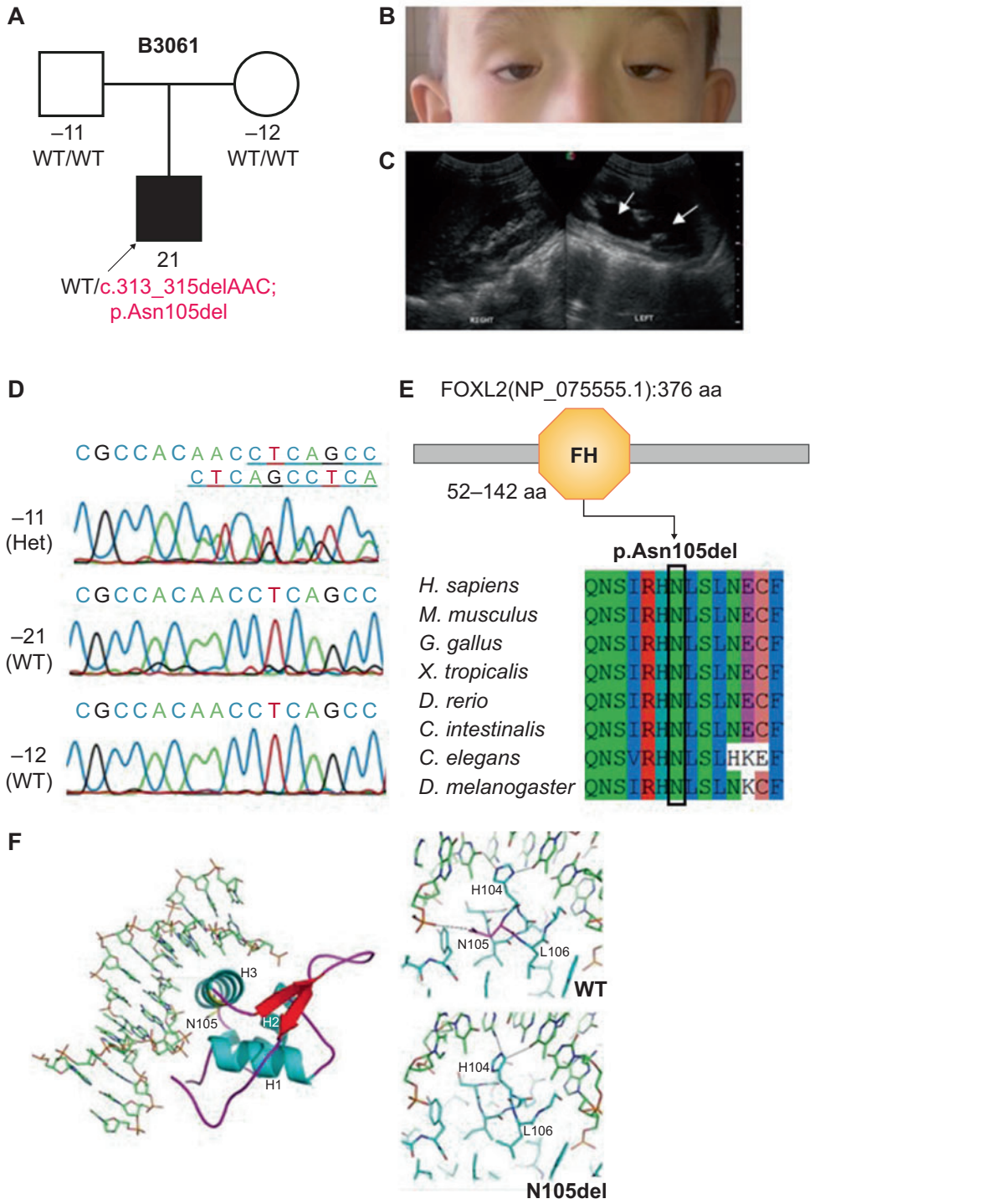

FIGURE 2: Identification of a *de novo FOXL2* in-frame deletion variant in a patient with BPES combining CAKUT. (**A**) Pedigree and genotype information for the affected family B3061. Squares indicate males, circles females, filled symbols are affected individuals and open symbols indicate healthy individuals. Proband (individual -21 of the family) is denoted by a black arrow. The proband carried a Het *de novo FOXL2* in-frame deletion (c.313_315delAAC; p.Asn105del). (**B**) Facial picture of the proband at the age of 9 years with blepharophimosis and ptosis. (**C**) Renal ultrasound suggests left ureteropelvic junction obstruction. (**D**) Sequencing chromatograms of the Het *FOXL2* variant in the proband and WT sequence detected in the parents. The AAC deletion is underlined and marked in red. (**E**) Protein domain content of FOXL2 and multiple sequence alignment of the FOXL2 protein region flanking residue Asn105. The Asn105 in-frame deletion detected in the family is mapped to the FH domain and Asn105 is well conserved from *H. sapiens* to Dm. (**F**) Structural model of FOXL2 FH/DNA-recognition motif complex generated using FOXA3 crystal structure (PDB ID 1VTN). N105 indicated by the stick model is located on the α-helix H3 (left panel). The WT (top right) and the N105del mutant (bottom right) models are shown in stick representation. The lines represent hydrogen bonds.

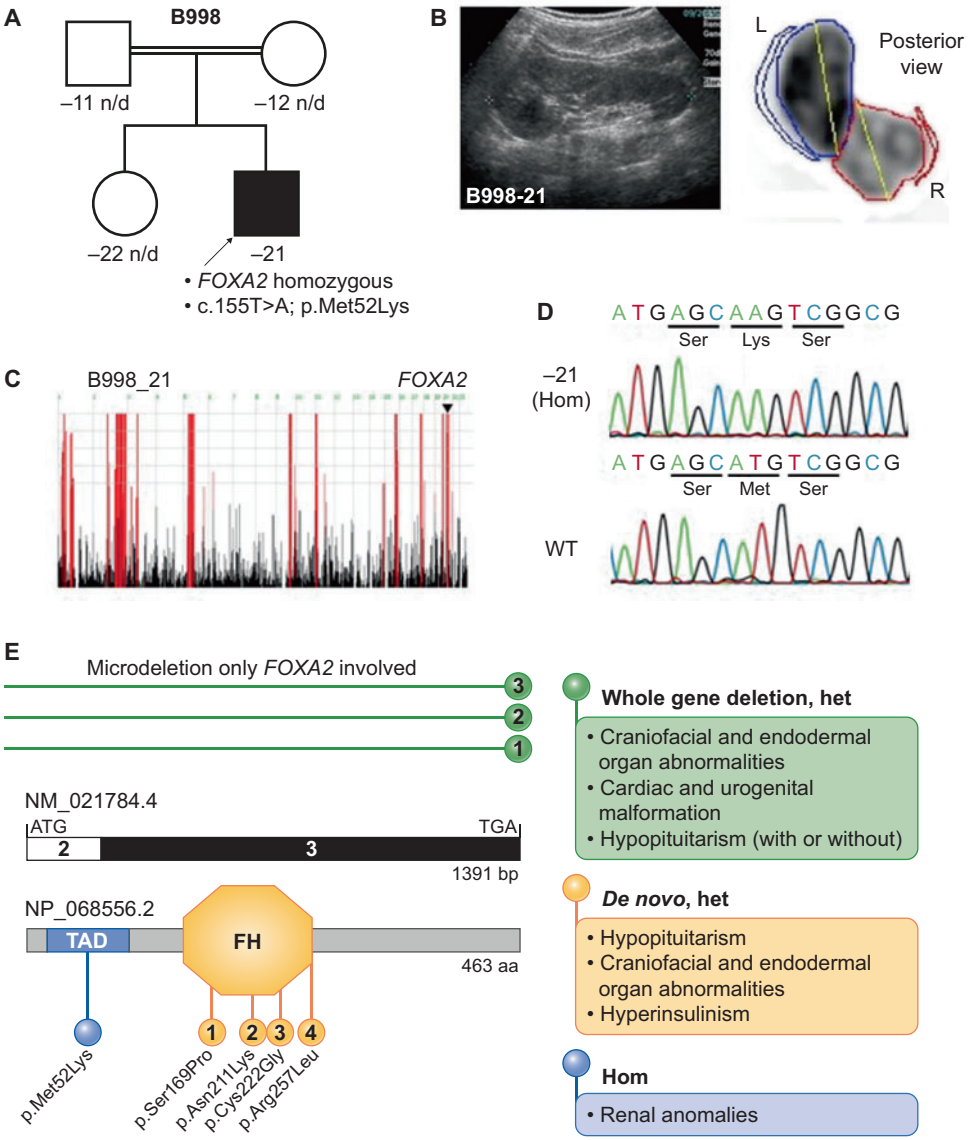

FIGURE 3: WES identifies a Hom *FOXA2* missense variant in a consanguineous family with CAKUT. (**A**) Pedigree and genotype information for members of family B998. Squares indicate males, circles females, filled symbols are affected individuals, open symbols indicate healthy individuals. Proband (individual -21 of the family) is denoted by black arrow. The proband carried a Hom missense variant in *FOXA2* (c.155T>A; p.Met52Lys). (**B**) Diagnostic images of the CAKUT phenotype of the affected individual B998-21. The left and right panels depict a renal ultrasound and a DMSA renal scan respectively. The renal ultrasound shows a horseshoe kidney located predominantly in the left abdomen, with an empty right renal fossa. DMSA images demonstrate abnormal anatomical location of the kidneys on the posterior view. Images are compatible with a horseshoe kidney. (**C**) Genome-wide HM in individual B998-21 identifies Hom peak regions. The gene locus for *FOXA2* that identified a Hom missense variant in this individual is located within a peak region on chromosome 20p11.3-20p12.1 (black arrow head). (**D**) Sequencing chromatograms of the Hom variant c.155T>A (p.Met52Lys) of the *FOXA2* gene in the proband compared to DNA from a healthy control (indicated by arrow head). (**E**) Genotypic and phenotypic information of *FOXA2* patients identified in this study and the literature. Protein schematic shows the functional domains of *FOXA2*, which contains a TAD (amino acids 14–93) and the FH domain (amino acids 59–257). Green lollipops indicate three previously reported patients with microdeletion only encompassing *FOXA2* [17, 18]. A blue lollipop indicates the Hom variant p.Met52Lys detected here in family B998 with renal anomaly. Four yellow lollipops indicate four *de novo* variants previously reported [19–22].

and *FOXC2*) have been previously associated with isolated or syndromic CAKUT in humans [2, 8–10], we tested the hypothesis that 36 *FOX* genes that are highly expressed in developmental kidneys may also represent monogenic causes of CAKUT. By analysis of the WES data in 680 patients from 541 unrelated CAKUT families, we delineated three new monogenic candidates for CAKUT (*FOXL2*, *FOXA2* and *FOXA3*). This also supports the utility of a paralog-based approach applied to discover mutated genes associated with human disease.

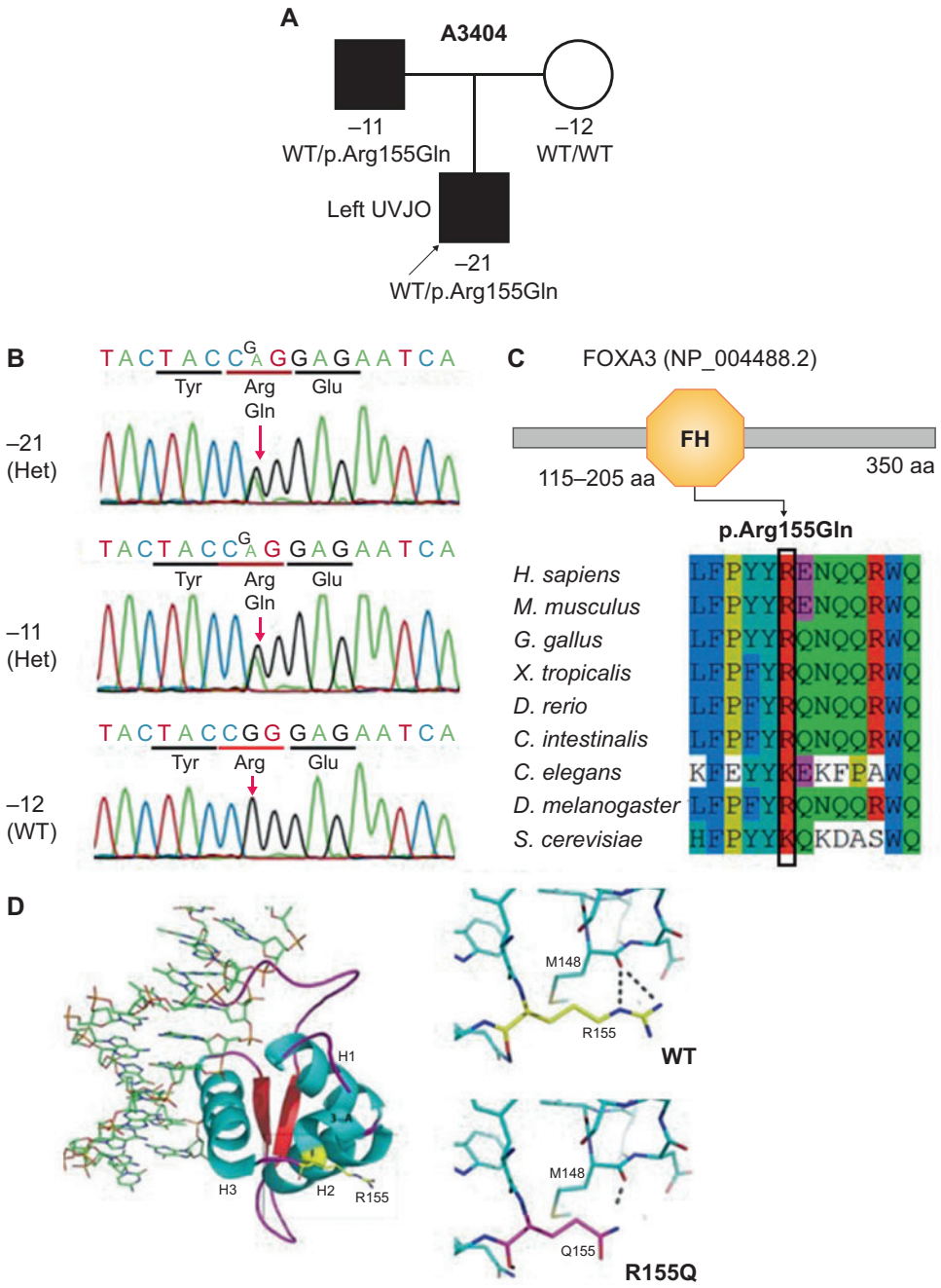

FIGURE 4: WES identifies a Het *FOXA3* missense variant in a CAKUT family with multiple affected individuals. (**A**) Pedigree and genotype information for members of family A3404. Squares indicate males, circles females, filled symbols are affected individuals and open symbols indicate healthy individuals. Proband is denoted by black arrow. The proband (-21) and the affected father (-11) both have CAKUT and carried a Het missense variant in *FOXA3* (c.464G>A; p.Arg155Gln). (**B**) Sequencing chromatograms of the Het variant c.464G>A; p.Arg155Gln in the affected proband and the affected father (-11), and WT sequence (indicated by arrow head) of *FOXA3* detected in the mother (-12). (**C**) Depicts the protein structure of *FOXA3* and multiple sequence alignment of the *FOXA3* protein region flanking residue Arg155. The p.Arg155Gln missense change detected in the family is mapped to the FH domain and well conserved from *H. sapiens* to Dm with the exception of substitution by 'K', which represents also a positively charged amino acid residue. (**D**) Structural model of FOXA3 FH/DNA-recognition motif complex (PDB ID 1VTN). Arg155 is indicated by the stick model on the 3_{10} helix. The WT (top right) and the R155Q mutant (bottom right) models are shown in stick representation. The dotted lines represent hydrogen bonds.

Renal phenotypic expansion in FOXL2-related BPES

We identified a *de novo* in-frame deletion p.Asn105del of *FOXL2* in a patient who presented with BPES and left UPJO. The p.Asn105del variant was previously reported in a familial (two generations) BPES case [23], in which the renal affected status is not mentioned. Interestingly, Gulati *et al.* described one case with co-occurrence of congenital hydronephrosis and *FOXL2*-associated BPES [24]. Here we also report a second

BPES patient with congenital kidney malformations. There are no data regarding the role of this master transcription factor in kidney and urinary tract development. However, according to single-cell RNA sequence data of human fetal kidneys at developmental week 16, *FOXL2* was mainly expressed in pre-tubular aggregate cells and interstitial cells [16]. It is possible that renal anomalies are a low penetrance feature of BPES, or that *FOXL2* is a novel gene that likely contributes to the CAKUT phenotype. We recommend that patients with pathogenic or likely pathogenic variants in *FOXL2* should undergo renal ultrasound or other renal imaging examinations.

FOXA2 genotype and phenotype

Previous studies of mouse models have indicated the critical role of Foxa2 in the development of ventral midline structures [25], as well as development of endoderm-derived organs, including the liver, lung, gastrointestinal tract and pancreas [26–29]. There is no defined human phenotype related to *FOXA2* variants in OMIM. As shown in Figure 3E and Supplementary data, Table S7, four different *de novo FOXA2* variants have been reported to be associated with hypopituitarism, hyperinsulinism, endodermal organ and craniofacial abnormalities [19–22]. Interestingly, all four reported *de novo* variants are located in the well-conserved DNA binding domain (FH). Additionally, three patients with a microdeletion of 20p11.21 (only *FOXA2* involved) presented with endodermal organ and craniofacial abnormalities, urogenital and cardiac malformation, with or without hypopituitarism [17, 18]. In this study, we identified a likely disease-causing Hom missense variant in the TAD in a patient who presented with a renal anomaly without any other extrarenal malformations. FoxA proteins harbor an N-terminal TAD domain that presumably recruits transcriptional co-factors, which in turn can facilitate other factors to enter the chromatin [30]. These observations suggest an allelism that differentiates effects of variants located out of the FH domain is likely responsible for an isolated or milder phenotype. The same allelic dosage was observed in *FOXC1* variants related Axenfeld–Rieger syndrome or anterior segment dysgenesis [8], as well as in *FOXL2* variants related to BPES phenotype [31, 32]. In most of the cases, *FOXL2* causes BPES in a dominant manner. Nallathambi *et al.* reported an in-frame duplication (p.A228_A232dup), which was located outside of the FH domain. The variant is segregated in a large Indian kindred where Het carriers are unaffected, whereas Hom individuals have the typical BPES phenotype [32].

Potential mechanisms for FOXAs in CAKUT

FoxAs have been shown to function as pioneer factors to open chromatin and thus increase the accessibility of other transcription factors to their target genes [33, 34]. There are no data regarding the role of *FOXA2* and *FOXA3* in kidney development. scRNA-seq data analysis of 16-week gestation human fetal kidney shows that *FOXA2* is mainly expressed in renal vesicle/comma-shaped body cells and s-shaped body cells, with lower expression levels seen in the distal tubule/loop of Henle cells (Supplementary data, Figure S2). *FOXA3* is specifically expressed in the ureteric bud/collecting duct cells (Supplementary data, Figure S2). Previous studies showed that *FOXA2* as a marker of a transient urothelial progenitor cell population is a key regulator of embryonic bladder development and patterning [35–37]. In this study, we firstly reported a Hom missense variant in the TAD of *FOXA2* in a patient with renal malformation. Qian and Costa precisely characterized the N-terminal TAD extending from amino acids 14–93 in Foxa2, which can enable the binding of other transcription factors to DNA in chromatin [30]. It is possible that Foxa2 acts, in part, as a transcriptional regulator of other important transcription factors involved in renal development such as *GATA3* and *SOX17*, which are known to cause human CAKUT [38, 39]. A number of studies have demonstrated that *FOXA2*, *GATA3* and *SOX17* were co-localized and probably act in parallel in the specification and formation of endoderm [23, 40, 41].

In conclusion, by unbiased WES analyses in 541 families with CAKUT, we identified *FOXA2* and *FOXA3* as novel monogenic candidates of CAKUT. We also provided further evidence for renal phenotypic expansion for *FOXL2* variants in the BPES (OMIM#110100). We show here the utility of WES for the identification of novel monogenic candidates of families with CAKUT.

SUPPLEMENTARY DATA

Supplementary data are available at ndt online.

ACKNOWLEDGEMENTS

We would like to thank the families and study individuals for their contribution. F.H. is the William E. Harmon Professor of Pediatrics. F.H. is also supported by the Begg Family Foundation. B.Z. is supported by the program of China Scholarships Council (No. 201908320472). S.Seltzsam is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, 442070894). D.M.C. is funded by Health Research Board, Ireland (HPF-206-674), the International Pediatric Research Foundation Early Investigators' Exchange Program and the Amgen® Irish Nephrology Society Specialist Registrar Bursary. She is also funded by the Eugen Drewlo Chair for Kidney Research and Innovation at the Schulich School of Medicine & Dentistry at Western University, London, Ontario, Canada. C.-H.W.W. is supported by funding from the American College of Medical Genetics and Genomics Foundation (ACMG/Takeda Next-Generation Biochemical Genetics Award). N.M. is supported by funding from the National Institutes of Health T32-DK007726 grant. We also thank the Yale Center for Mendelian Genomics for whole-exome sequencing analysis (U54HG006504).

FUNDING

This research was supported by grants from the National Institutes of Health to F.H. (DK076683 and DK088767).

AUTHORS' CONTRIBUTIONS

B.Z. and F.H. designed the study. B.Z., C.W., S. Seltzsam, L.S., S. Schneider, S. Shril and F.H. analyzed the data. B.Z., C.W., S. Seltzsam, S. Schneider, L.S., C.H.W., R.D., D.M.C., M.N., N.M., S. Shril and S.M. performed WES analysis and Sanger sequencing. N.S., S.B. and V.T. collected the clinical data. H.J.N. conducted the crystal structure analysis. B.Z. and F.H. drafted the paper. All authors revised the manuscript and approved the final version.

CONFLICT OF INTEREST STATEMENT

F.H. is a cofounder and SAC member and holds stock in Goldfinch-Bio. All other authors declare that they have no competing financial interests.

DATA AVAILABILITY STATEMENT

The datasets supporting the current study have not been deposited in a public repository, but are available from the corresponding author on request.

REFERENCES

1. Caruana G, Bertram JF. Congenital anomalies of the kidney and urinary tract genetics in mice and men. *Nephrology (Carlton)* 2015; 20: 309–311
2. Bekheirnia MR, Bekheirnia N, Bainbridge MN *et al.* Whole-exome sequencing in the molecular diagnosis of individuals with congenital anomalies of the kidney and urinary tract and identification of a new causative gene. *Genet Med* 2017; 19: 412–420
3. Kohl S, Habbig S, Weber LT *et al.* Molecular causes of congenital anomalies of the kidney and urinary tract (CAKUT). *Mol Cell Pediatr* 2021; 8: 2
4. van der Ven AT, Connaughton DM, Ityel H *et al.* Whole-exome sequencing identifies causative mutations in families with congenital anomalies of the kidney and urinary tract. *J Am Soc Nephrol* 2018; 29: 2348–2361
5. Golson ML, Kaestner KH. Fox transcription factors: from development to disease. *Development* 2016; 143: 4558–4570
6. Zaiss DMW, Coffer PJ. Forkhead box transcription factors as context-dependent regulators of lymphocyte homeostasis. *Nat Rev Immunol* 2018; 18: 703–715
7. Weigel D, Jurgens G, Kuttner F *et al.* The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. *Cell* 1989; 57: 645–658
8. Wu CW, Mann N, Nakayama M *et al.* Phenotype expansion of heterozygous FOXC1 pathogenic variants toward involvement of congenital anomalies of the kidneys and urinary tract (CAKUT). *Genet Med* 2020; 22: 1673–1681
9. Stankiewicz P, Sen P, Bhatt SS *et al.* Genomic and genic deletions of the FOX gene cluster on 16q24.1 and inactivating mutations of FOXF1 cause alveolar capillary dysplasia and other malformations. *Am J Hum Genet* 2009; 84: 780–791
10. Jones GE, Richmond AK, Navti O *et al.* Renal anomalies and lymphedema distichiasis syndrome. A rare association? *Am J Med Genet A* 2017; 173: 2251–2256
11. Braun DA, Sadowski CE, Kohl S *et al.* Mutations in nuclear pore genes NUP93, NUP205 and XPO5 cause steroid-resistant nephrotic syndrome. *Nat Genet* 2016; 48: 457–465
12. MacArthur DG, Manolio TA, Dimmock DP *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014; 508: 469–476
13. Li H, Handsaker B, Wysoker A *et al.*; 1000 Genome Project Data Processing Subgroup. The sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25: 2078–2079
14. Van der Auwera GA, Carneiro MO, Hartl C *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; 43: 11.10.1–11.10.3
15. Clark KL, Halay ED, Lai E *et al.* Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 1993; 364: 412–420
16. Hochane M, van den Berg PR, Fan X *et al.* Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development. *PLoS Biol* 2019; 17: e3000152
17. Tsai EA, Grochowski CM, Falsey AM *et al.* Heterozygous deletion of FOXA2 segregates with disease in a family with heterotaxy, panhypopituitarism, and biliary atresia. *Hum Mutat* 2015; 36: 631–637
18. Elsayed AK, Aghadi M, Ali G *et al.* Generation of a human induced pluripotent stem cell line (QBRIi009-A) from a patient with a heterozygous deletion of FOXA2. *Stem Cell Res* 2020; 42: 101705
19. Giri D, Vignola ML, Gualtieri A *et al.* Novel FOXA2 mutation causes hyperinsulinism, hypopituitarism with craniofacial and endoderm-derived organ abnormalities. *Hum Mol Genet* 2017; 26: 4315–4326
20. Vajravelu ME, Chai J, Krock B *et al.* Congenital hyperinsulinism and hypopituitarism attributable to a mutation in FOXA2. *J Clin Endocrinol Metab* 2018; 103: 1042–1047
21. Boda H, Miyata M, Inagaki H *et al.* FOXA2 gene mutation in a patient with congenital complex pituitary hormone deficiency. *Eur J Med Genet* 2019; 62: 103570
22. Stekelenburg C, Gerster K, Blouin JL *et al.* Exome sequencing identifies a de novo FOXA2 variant in a patient with syndromic diabetes. *Pediatr Diabetes* 2019; 20: 366–369
23. Bunyan DJ, Thomas NS. Screening of a large cohort of blepharophimosis, ptosis, and epicanthus inversus syndrome patients reveals a very strong paternal inheritance bias and a wide spectrum of novel FOXL2 mutations. *Eur J Med Genet* 2019; 62: 103668
24. Gulati R, Verdin H, Halanaik D *et al.* Co-occurrence of congenital hydronephrosis and FOXL2-associated blepharophimosis, ptosis, epicanthus inversus syndrome (BPES). *Eur J Med Genet* 2014; 57: 576–578
25. Weinstein DC, Ruiz I Altaba A, Chen WS *et al.* The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. *Cell* 1994; 78: 575–588
26. Wan H, Dingle S, Xu Y *et al.* Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem* 2005; 280: 13809–13816
27. Gao N, LeLay J, Vatamaniuk MZ *et al.* Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev* 2008; 22: 3435–3448
28. Lee CS, Friedman JR, Fulmer JT *et al.* The initiation of liver development is dependent on Foxa transcription factors. *Nature* 2005; 435: 944–947
29. Burtscher I, Barkey W, Lickert H. Foxa2-venus fusion reporter mouse line allows live-cell analysis of endoderm-derived organ formation. *Genesis* 2013; 51: 596–604
30. Qian X, Costa RH. Analysis of hepatocyte nuclear factor-3 beta protein domains required for transcriptional activation and nuclear targeting. *Nucleic Acids Res* 1995; 23: 1184–1191
31. Kaur I, Hussain A, Naik MN *et al.* Mutation spectrum of fork-head transcriptional factor gene (FOXL2) in Indian Blepharophimosis Ptosis Epicanthus Inversus Syndrome (BPES) patients. *Br J Ophthalmol* 2011; 95: 881–886
32. Nallathambi J, Moumne L, De Baere E *et al.* A novel polyalanine expansion in FOXL2: the first evidence for a recessive form of the blepharophimosis syndrome (BPES) associated with ovarian dysfunction. *Hum Genet* 2007; 121: 107–112
33. Cernilogar FM, Hasenoder S, Wang Z *et al.* Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. *Nucleic Acids Res* 2019; 47: 9069–9086
34. Iwafuchi-Doi M, Donahue G, Kakumanu A *et al.* The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell* 2016; 62: 79–91
35. Boumelhem BB, Fraser ST, Assinder SJ. Differentiation of urothelium from mouse embryonic stem cells in chemically defined conditions. *Methods Mol Biol* 2019; 2029: 103–115

36. Oottamasathien S, Wang Y, Williams K *et al.* Directed differentiation of embryonic stem cells into bladder tissue. *Dev Biol* 2007; 304: 556–566
37. Georgas KM, Armstrong J, Keast JR *et al.* An illustrated anatomical ontology of the developing mouse lower urogenital tract. *Development* 2015; 142: 1893–1908
38. Van Esch H, Groenen P, Nesbit MA *et al.* GATA3 haplo-insufficiency causes human HDR syndrome. *Nature* 2000; 406: 419–422
39. Gimelli S, Caridi G, Beri S *et al.* Mutations in SOX17 are associated with congenital anomalies of the kidney and the urinary tract. *Hum Mutat* 2010; 31: 1352–1359
40. Wang P, Rodriguez RT, Wang J *et al.* Targeting SOX17 in human embryonic stem cells creates unique strategies for isolating and analyzing developing endoderm. *Cell Stem Cell* 2011; 8: 335–346
41. Genga RMJ, Kernfeld EM, Parsi KM *et al.* Single-cell RNA-sequencing-based CRISPRi screening resolves molecular drivers of early human endoderm development. *Cell Rep* 2019; 27: 708–718.e10

Received: 11.6.2021; Editorial decision: 23.8.2021