REVIEWS

Genetic, environmental, and epigenetic factors involved in CAKUT

Nayia Nicolaou, Kirsten Y. Renkema, Ernie M. H. F. Bongers, Rachel H. Giles and Nine V. A. M. Knoers

Abstract | Congenital anomalies of the kidney and urinary tract (CAKUT) refer to a spectrum of structural renal malformations and are the leading cause of end-stage renal disease in children. The genetic diagnosis of CAKUT has proven to be challenging due to genetic and phenotypic heterogeneity and incomplete genetic penetrance. Monogenic causes of CAKUT have been identified using different approaches, including single gene screening, and gene panel and whole exome sequencing. The majority of the identified mutations, however, lack substantial evidence to support a pathogenic role in CAKUT. Copy number variants or single nucleotide variants that are associated with CAKUT have also been identified. Numerous studies support the influence of epigenetic and environmental factors on kidney development and the natural history of CAKUT, suggesting that the pathogenesis of this syndrome is multifactorial. In this Review we describe the current knowledge regarding the genetic susceptibility underlying CAKUT and the approaches used to investigate the genetic basis of CAKUT. We outline the associated environmental risk factors and epigenetic influences on CAKUT and discuss the challenges and strategies used to fully address the involvement and interplay of these factors in the pathogenesis of the disease.

Introduction

Congenital anomalies of the kidney and urinary tract (CAKUT) collectively refer to a diverse group of structural malformations that are characterized by defects in embryonic kidney development (Figure 1). CAKUT constitute ~20–30% of all congenital malformations and their prevalence has been estimated to range between three and six per 1,000 births.[1–3] The most common malformation is ureteropelvic junction obstruction (~20%);[3] other phenotypes include kidney agenesis, multicystic dysplastic kidneys, kidney dysplasia, kidney hypoplasia, vesicoureteral reflux (VUR), megaureter, ectopic ureter, horseshoe kidney, duplex collecting system, and posterior urethral valves, although whether posterior urethral valves share a common aetiology with the other anomalies is debated (Figure 1). Some anomalies occur concurrently in CAKUT, such as VUR and duplex collecting system.[4] Successful detection of CAKUT *in utero* has substantially improved in conjunction with increased availability and utility of prenatal ultrasonography, but many cases still remain undiagnosed until after birth when overt symptoms, such as hydronephrosis, cause urinary tract infections.[3]

Kidney development is a multi-stage process that begins with induction of the ureteric bud from the nephric duct, followed by mesenchymal-to-epithelial transition and branching morphogenesis, and terminates with the completion of nephron patterning and elongation.[4] Disturbances to normal nephrogenesis, due to exposure to environmental risk factors or the dysfunction of genes that direct this process, can lead to CAKUT (Figure 2).[5] Our current understanding of the mechanisms involved in the pathogenesis of CAKUT and the genes involved in renal development and nephrogenesis has mostly been derived from mouse models. These models have led to the identification of numerous candidate genes that might cause CAKUT in humans. Further information regarding renal development that has been gained from knockout mouse models has been detailed elsewhere.[5–7]

Data from national and regional renal registries in Europe show that CAKUT are the leading cause of end-stage renal disease (ESRD) in children, accounting for 41.3% of children that receive renal replacement therapy (RRT).[8] Data from the European Renal Association–European Dialysis and Transplant Association Registry consisting of >200,000 patients undergoing RRT, revealed that the estimated median age of RRT initiation was 31 and 61 years in patients with and without CAKUT, respectively.[9] The age of onset of RRT was lowest in patients with renal dysplasia (16 years) and generally varied within CAKUT sub-phenotypes.[9] Treatment outcomes and survival rates are better for patients with CAKUT than for patients with other kidney diseases, such as type 1 diabetic nephropathy, because of lower associated cardiovascular mortality.[9] The number of adults with ESRD due to CAKUT is reported to be 2.2%, but this figure may well be an underestimation.[9] Some patients with CAKUT

Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, PO Box 85090, 3508 AB, Utrecht, Netherlands (**N.N.**, **K.Y.R.**, **N.V.A.M.K.**). Department of Human Genetics, Radboud University Medical Center, PO Box 9101, 6500 HB, Nijmegen, Netherlands (**E.M.H.F.B.**). Department of Nephrology and Hypertension, University Medical Center Utrecht, PO Box 85500, 3508 AB, Utrecht, Netherlands (**R.H.G.**).

Correspondence to: N.V.A.M.K. v.v.a.knoers@ umcutrecht.nl

Competing interests
The authors declare no competing interests.

Key points

- Approaches to discover genetic causes of congenital anomalies of the kidney and urinary tract (CAKUT) include candidate gene and whole exome sequencing, and genome-wide linkage and copy number variant (CNV) analyses
- The majority of sporadic cases of CAKUT cannot yet be explained by monogenic causes
- Certain CNVs are associated with an elevated risk of CAKUT, indicating that CNV analysis should become part of the diagnostic procedure
- Chromosomal imbalances and single nucleotide variants in non-coding regions contribute to congenital malformations, suggesting that genomic regulatory elements might also function in the pathogenesis of CAKUT
- Epigenetic and gestational environmental risk factors can influence kidney development and/or fibrosis and might also increase susceptibility to CAKUT
- Collaborative efforts are needed to collect large cohorts of patients with CAKUT and to integrate the data from epidemiologic, clinical, genomic, epigenomic, transcriptomic, and proteomic studies

are asymptomatic during childhood and remain undiagnosed, but might develop hypertension, proteinuria, and renal impairment in adulthood.[9,10] Disease diagnosis undoubtedly needs to be optimized and should rely both on clinical information and genetic and molecular findings in order to improve prognostic counselling and personalized disease management.

A genetic predisposition to CAKUT is supported by the co-occurrence of these anomalies with multi-organ defects and the frequency of familial cases.[5] CAKUT can occur either in an isolated form (when only CAKUT are present), or might occasionally develop in association with other congenital anomalies. An estimated 34% of infants born with CAKUT over a 26-year period in North-Eastern France had associated congenital anomalies.[3,11] The most frequent syndromes associated with CAKUT included vertebral defects, anal atresia, cardiac defects, tracheo-oesophageal fistulae, renal anomalies, limb abnormalities, and Prune belly, Meckel–Gruber, renal-coloboma, renal cysts and diabetes, branchio-oto-renal, and Fraser syndromes.[11] A family history of renal abnormalities is present in ~10% of CAKUT patients.[12] An epidemiological study found that 51% of index Turkish patients with CAKUT—diagnosed on the combined basis of a renal ultrasound examination and family history—had at least one first-degree relative with CAKUT.[13] The most common anomaly in first-degree relatives was a duplex collecting system (Figure 1). The high frequency of a positive family history in this cohort, however, might be biased by parental consanguinity.[13]

In this Review, we summarize the key findings from genetic studies of CAKUT, epidemiological research on environmental risk factors for CAKUT, and data regarding the known epigenetic factors (Figure 2). Furthermore, we discuss the application of recent genetic techniques to identify causative mutations in patients with CAKUT, as well as the challenges for the clinical interpretation of genetic abnormalities and genetic counselling.

Genetic causes of CAKUT

Multiple lines of evidence support the assumption that CAKUT can be caused by a defect in a single gene. Firstly, multiplex families with CAKUT have been described in the literature, suggesting an autosomal dominant mode of inheritance with reduced penetrance.[14–17] Secondly, well known, rare genetic syndromes characterized by a phenotype associated with CAKUT and extrarenal anomalies caused by single-gene defects also indicate that CAKUT can have a monogenic etiology.[5] Mutation screening studies conducted in a series of patients with CAKUT have resulted in a list of candidate genes that might cause CAKUT in an autosomal dominant or, less frequently, recessive pattern of inheritance.[5] These investigations revealed that the clinical phenotype and severity of CAKUT can vary markedly among patients, both within and between families with the same underlying mutation, leading to the hypothesis of an underlying oligogenic model of inheritance.[5]

In 1995, the first single gene defect described as being causative of CAKUT was a frameshift deletion in *PAX2* in a family with optic nerve colobomas, renal hypoplasia, and VUR.[18] *HNF1B* was the second gene to be implicated in CAKUT following the discovery of a heterozygous mutation in two siblings with renal cysts and diabetes syndrome.[19] Both of these findings were followed by several studies with larger patient cohorts that confirmed *PAX2* and *HNF1B* mutations as important causes of syndromic and isolated CAKUT.[20–22] These genes are estimated to explain ~15% cases of CAKUT, subsequently rendering them the first and most important genes to screen for diagnostic purposes. Mutations in *PAX2* are more frequently associated with renal dysplasia or hypoplasia, whereas mutations in *HNF1B* are more frequently associated with cystic kidneys.[20–22] Several additional genes underlying syndromic forms of CAKUT have been identified (Table 1). The detection of a pathogenic mutation in syndromic cases is particularly important to facilitate early recognition and/or monitoring and timely treatment of clinically significant co-morbidities.

Approaches to study genetic susceptibility

Approaches to discover new genes that cause CAKUT have so far included candidate gene studies, genome-wide linkage analyses, whole exome sequencing, and copy number variation analyses. Genetic studies followed by functional tests using any of the *in vitro* and *in vivo* models available for CAKUT provide an initial framework for the classification of mutations that may potentially cause disease.

Candidate gene studies

Gene discovery for non-syndromic CAKUT has been less successful than for syndromic CAKUT, but some candidate genes proposed from knockout mouse models have been validated by the identification of mutations in patients (Table 1). Mutation screening of candidate genes has identified disease-causing mutations in 6–20% of patients with CAKUT in a heterozygous state. Mutations have so far been reported in genes including *BMP7*, *CHD1L*, *CDC5L*, *EYA1*, *GATA3*, *RET*, *ROBO2*, *SALL1*, *SIX2*, *SIX5*, *FRAS1,* and *FREM2*.[20,23,24] The majority of studies, however, have lacked co-segregation analyses, statistically rigorous methods, and functional evidence

Figure 1 | Illustrative 3D models of congenital abnormalities of the kidney and urinary tract (CAKUT) that were created using the modelling tools ZBrush (Pixologic, Inc.) and 3ds Max (Autodesk, Inc.), to illustrate the phenotypic spectrum of CAKUT. The black arrows indicate the site of obstruction in ureteropelvic junction obstruction and posterior urethral valves. The red arrows indicate the abnormal flow of the urine from the bladder to the ureter or kidney occurring in vesico-ureteral reflux.


to support the claim of causality or association with CAKUT. The number of genes and variants that have a true pathogenic effect in CAKUT, therefore, is probably overestimated and the detection of a predicted deleterious variant in any of the above genes should not by itself imply pathogenicity. In addition, databases assembled from large-scale sequencing projects in the general population, such as the Exome Aggregation Consortium,[25] serve as useful reference sets of variants and reveal that loss-of-function variants in previously reported disease-causing genes for CAKUT are also present in the general population. Such resources will be useful for comparing the frequency of candidate variants observed in patients with CAKUT with a control dataset.

Overexpression of *SIX2* and *BMP4* human RNA in zebrafish, containing heterozygous mutations found in patients with CAKUT, has helped validate the pathogenic effect of these mutations on kidney development.[26] Similarly, *in vitro* analyses of mutations in *SOX17*, *WNT4,* and *RET* have revealed differences in signalling activity between the variant and wild-type proteins.[27–29] Although these genes have a well characterized function in kidney development, some parents in affected pedigrees were healthy carriers of mutations, suggesting that the degree of penetrance of these mutations can be low and as a result, risk stratification of carriers is difficult and/or that the pathogenesis of CAKUT is multifactorial.[26]

Linkage studies
Linkage analyses have been successful for mapping rare Mendelian diseases and susceptibility loci implicated in common diseases, such as Alzheimer disease, insulin-dependent diabetes mellitus, and breast cancer, by demonstrating co-segregation of genetic markers with a disease in a family-based approach.[30–32] This approach is particularly powerful for identifying candidate genomic regions that include variants with a large effect, but has not been as straightforward or successful for mapping candidate regions in families with CAKUT. Families with traits such as renal agenesis, hypoplasia, or dysplasia are particularly hard to ascertain because they occur less frequently than abnormalities of the urinary tract, and most cases are either truly sporadic or familial with incomplete penetrance.[3] Moreover, due to ethical and financial barriers, systematic renal ultrasound screening of relatives is not routinely performed to accurately distinguish between unaffected and asymptomatic family members.

The majority of published linkage studies on CAKUT have focused on familial VUR, which occurs in ~1% of children.[33] Linkage analyses performed to date have assumed an autosomal dominant model of inheritance with reduced penetrance (70–85%). Linkage has been
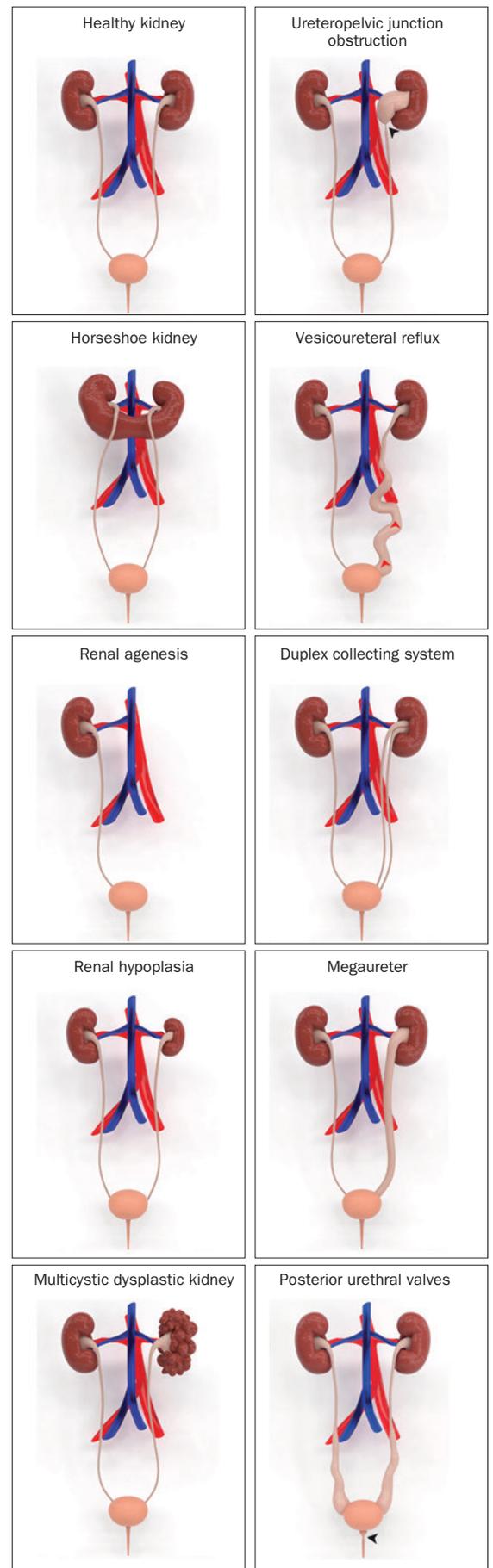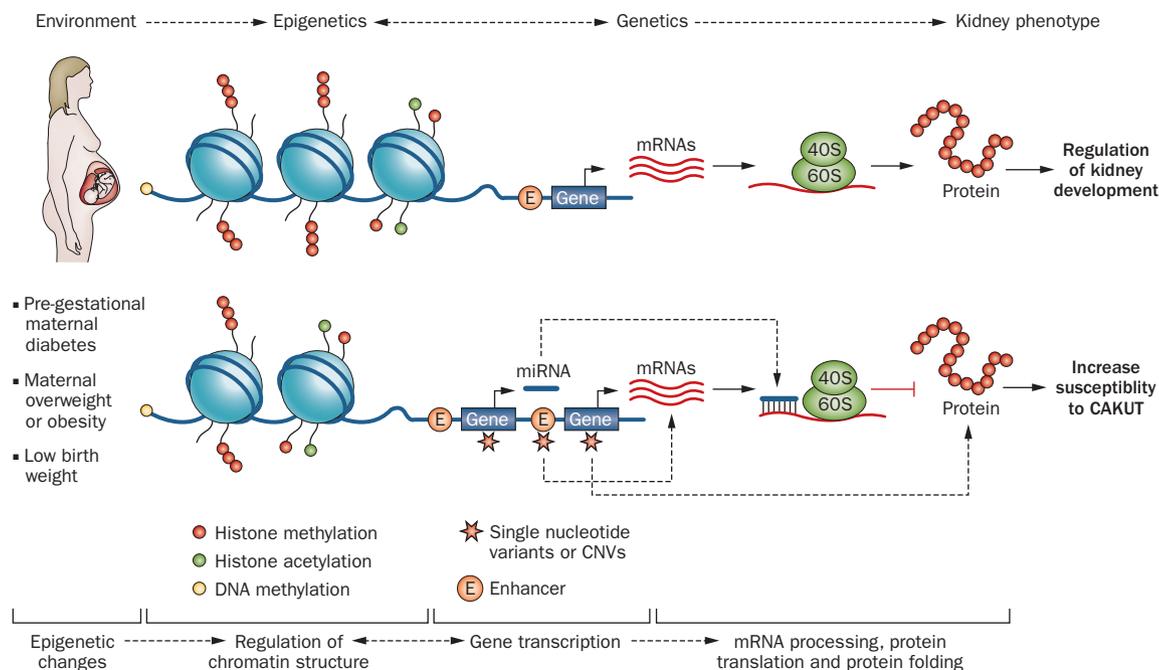
Figure 2 | The potential effect of the interplay between environmental, epigenetic, and genetic factors on kidney development. At the environmental level, changes in the *in utero* environment are postulated to influence epigenetic mechanisms that can modulate gene expression by changing the composition of chromatin. At the epigenetic level, key epigenetic marks include DNA methylation and tri-methylation of lysine 27 of histone 3 (H3K27me3) at repressed genes, and acetylation of the latter residue (H3K27me1) and mono-methylation of lysine 4 at histone 3 (H3K4me1) at active genes. Changes in epigenetic marks might repress or activate genes involved at different stages of renal development. At the genetic level, genetic variants (indicated by stars), including single nucleotide variants or CNVs could occur in regulatory regions and affect the transcription of genes involved in renal development; within coding regions, leading to changes in protein expression or function; or in genes producing non-coding miRNAs that can regulate protein translation at a post-transcriptional level. Genes that are essential for kidney development, therefore, are subject to environmental, epigenetic, and genetic modifications, which could disrupt their regulation and result in increase susceptibility to CAKUT. Abbreviations: CAKUT, congenital anomalies of the kidney and urinary tract; CNV, copy number variant; mRNA, microRNA.

found at multiple loci, with few loci reaching genome-wide significance and almost none of them confirmed in independent families. A genome-wide linkage scan in 12 families with VUR showed linkage in 60% of the families on chromosome 12p11–q13 under a recessive model, but the specific gene remains elusive.[34] A large Somalian family with eight of 12 children presenting with CAKUT showed linkage on chromosome 8q24 under an autosomal recessive model of inheritance. Subsequent screening of the only predicted gene within the interval, *KHDRBS3*, did not reveal any mutations.[35] Seven Italian multi-generation pedigrees with CAKUT have also been studied, with evidence of linkage for renal hypodysplasia on chromosome 1p32–33 under an autosomal dominant model with reduced penetrance in three of the seven families.[36] Follow-up studies on this locus have yet to be performed.

Linkage analysis in CAKUT might be further complicated by incomplete genetic penetrance, which obscures the true mode of inheritance; clinical and genetic heterogeneity within families; family members with the underlying genetic cause who appear unaffected or are insufficiently phenotyped; linkage peaks that have not yet been thoroughly sequenced; and false-positive reported linkage peaks. Moreover, CAKUT

may be caused by a number of low-penetrant genes that contribute to disease risk in combination with environmental risk factors. Linkage analysis, however, is not sufficiently powerful to detect susceptibility loci with a small effect size. In addition, linkage signals might point to genetic defects within non-coding regions that require additional characterization.

Targeted next-generation sequencing
Technological advancements in high-throughput sequencing methods have enabled next-generation sequencing (NGS) and consequent identification of genetic causes of sporadic cases of CAKUT. Analysis of multiple genes in parallel using NGS in patients with CAKUT has demonstrated that <10% patients with isolated CAKUT carry variants in previously implicated genes, such as *HNF1B*, *PAX2*, *EYA1*, *SIX5*, and *RET*.[23,24,37] This finding implies that the majority of causes of CAKUT are still unknown. Meanwhile, the list of novel variants requiring functional characterization to determine whether they are true risk factors or harmless polymorphisms has expanded. Interestingly, this list now includes missense variants in *FRAS1*, *FREM2*, and *GRIP1*, in which recessive mutations have been previously characterized as causative of Fraser syndrome.[38] This overlap raises the question as to whether

Table 1	Genes implicated in syndromic or non-syndromic CAKUT	
Gene	Disease	MIM#
Autosomal recessive mode of inheritance		
ACE	Renal tubular dysgenesis[111]	267430
AGT	Renal tubular dysgenesis[111]	267430
AGTR1	Renal tubular dysgenesis[112]	267430
CHRM3	Prune-belly-like syndrome[113]	1188494
FRAS1	Fraser syndrome[24]	219000
FREM1	Bifid nose with or without renal agenesis and anorectal malformations[114]	608980
FREM2	Fraser syndrome[24]	219000
GRIP1	Fraser syndrome[115]	219000
HPSE2	Urofacial syndrome[116]	236730
ITGA8	Renal agenesis[39]	191830
LRP4	Cenani–Lenz syndrome associated with renal agenesis[117]	212780
REN	Renal tubular dysgenesis[111]	267430
ROR2	Robinow syndrome with hydronephrosis[118]	268310
Autosomal dominant mode of inheritance		
BICC1	Multicystic kidney dysplasia[119]	601331
BMP4	CAKUT[26,120]	NA
DSTYK	CAKUT[46]	612666
EYA1	Branchio-oto-renal syndrome and renal hypoplasia[20]	602588
FAM58A	STAR syndrome[121]	300707
FGF8	Hypogonadotropic hypogonadism[122]	612702
FGFR1	Kallmann syndrome[123]	147950
GATA3	Hypoparathyroidism, sensorineural deafness and renal dysplasia[124]	146255
GLI3	Pallister–Hall syndrome[125]	146510
HNF1B	Multicystic kidney dysplasia, renal hypoplasia, renal cysts and diabetes syndrome[126,127]	137920
JAG1	Alagille syndrome[128]	118450
KAL1	Kallmann syndrome[129]	308700
NOTCH2	Alagille syndrome with renal anomalies[130]	610205
PAX2	Renal coloboma syndrome and CAKUT[131]	120330
PAX8	Hypothyroidism and renal agenesis[132]	218700
RET	Renal agenesis and Hirschsprung disease[27]	191830
ROBO2	Vesicoureteral reflux[133]	610878
SALL1	Townes–Brocks syndrome[134]	107480
SEMA3A	Kallmann syndrome[135]	614897
SIX1	Branchio-oto-renal syndrome and branchiootic syndrome[136]	608389 and 608389
SIX2	Renal hypodysplasia[26]	NA
SIX5	Branchio-oto-renal syndrome[136]	610896
SOX17	CAKUT[28]	613674
TFAP2A	Branchio-oto-renal syndrome[138]	113620
UPK3A	Renal dysplasia[139]	NA
WNT4	Mullerian aplasia and hyperandrogenism[29]	158330

Abbreviations: CAKUT, congenital anomalies of the kidney and urinary tract; MIM, Mendelian inheritance in man; NA, not applicable; STAR, toe syndactyly, telecanthus, and anogenital and renal malformations.

loss-of-function mutations in genes associated with Fraser syndrome could cause subtler phenotypes, such as isolated CAKUT, when present in a heterozygous state.[24,38]

Whole exome sequencing in a candidate-free approach has also led to the identification of new genes involved in CAKUT, including *ITGA8* and *TRAP1*, providing further evidence that CAKUT might in some cases be an autosomal recessive disease.[39,40] Recessive mutations in *ITGA8* have been identified in two families with bilateral renal agenesis. Furthermore, mutations in *TRAP1* have been detected at a frequency of 0.15% in CAKUT and 0.6% in CAKUT with VACTERL (vertebral anomalies, anal atresia, cardiac defects, tracheoesophageal fistula and/or esophageal atresia, renal and radial anomalies and limb defects).[41,42] Recessive mutations in both genes are, therefore, a novel but rare cause of these conditions.

The identification of autosomal dominant causes of CAKUT by whole exome sequencing is challenging unless multiple family members are affected, which can help refine the large number of identified variants to only those that are shared by affected individuals. Databases, such as PhenomeCentral,[41] DECIPHER,[42] ClinGen,[43] LOVD,[44] and Matchmaker Exchange[45] are being assembled to enable the discovery of patients with similar phenotypes who have had their genome sequenced at different centres but who share the same candidate variants. The advantage of co-segregation analysis was illustrated when genome-wide linkage analysis was used in a family with seven affected members together with whole exome sequencing in two of the family members, in order to identify a new candidate gene for CAKUT.[46] A heterozygous variant in *DSTYK* was the only variant that fulfilled the following filtering criteria: absent from public databases; mapped within the linkage intervals; and shared by all of the affected individuals. Variants in *DSTYK* were subsequently found in 2.3% of an independent cohort with CAKUT.[46] *Dstyk* is expressed in the developing mouse nephron and is essential for the development of multiple organs in zebrafish.[46] The combination of genome-wide linkage analysis, whole exome sequencing, and functional studies has, therefore, provided substantial evidence for the involvement of *DSTYK* in CAKUT.

A *de novo* mutation paradigm in more commonly occurring disorders has been supported by exome sequencing of patient–parent trios, where an increased exonic *de novo* mutation rate in clinically and genetically heterogeneous disorders, such as intellectual disability, schizophrenia, and autism spectrum disorders has been revealed.[47–49] Approximately 50% of mutations in *PAX2*, and microdeletions that span *HNF1B* are estimated to occur *de novo*, but evidence for *de novo* mutations in other genes is currently lacking.[50] The interpretation of the phenotypic consequences of *de novo* mutations in patients with CAKUT still requires considerable effort. Nonetheless, a model or comprehensive catalogue of *de novo* mutations across hundreds of genes that potentially contribute to kidney development and/or function could also explain why CAKUT is relatively common and mostly sporadic.

Copy number variant analysis

Chromosomal imbalances, such as large genomic deletions and duplications, can give rise to complex phenotypes that usually involve intellectual disability and multiple congenital anomalies.[51] These imbalances lead to the loss or gain of the affected genomic regions.[51] Renal anomalies are also associated with syndromes caused by genomic deletions, such as Williams, Alagille, branchio-oculo-facial, Kallmann, and renal cysts and diabetes syndromes.[52–55] Traditional techniques for the detection of large chromosomal deletions include standard karyotyping or fluorescent *in situ* hybridization, which can detect deletions between 500 kb and 5 Mb in length.[56] Smaller copy number variants (CNVs) can now be detected using high-resolution microarrays, such as the array comparative genomic hybridization (array-CGH) and single nucleotide polymorphism (SNP) microarrays. Both types of array offer advantages regarding throughput, cost, and genomic resolution, especially with the availability of customized arrays that can detect CNVs of at least 500 bp.[56] Coding CNVs that contain three or more exons can also be detected by whole exome sequencing.[57]

As previously discussed, microdeletions of chromosome 17q12 spanning *HNF1B* recurrently occur *de novo* in patients with CAKUT and renal disease, with or without diabetes mellitus.[55] Genomic deletions in this region associate with an earlier onset of the renal phenotype than in patients with point mutations in *HNF1B*.[55] In addition, a large deletion identified by array-CGH affecting *PAX2* was recently described for the first time as a rare cause of renal coloboma syndrome in addition to already known smaller gene deletions and point mutations.[58] A genome-wide CNV analysis using a SNP microarray in 522 individuals with renal agenesis or hypoplasia revealed rare CNVs >100 kb that disrupted gene-coding regions in 16.6% of patients.[59] When 192 patients with renal hypodysplasia of European ancestry were compared to 4,733 matched controls, the burden of large (>100 kb) rare CNVs was found to be significantly higher in cases of renal hypodysplasia ($P = 4.8 \times 10^{-11}$).[59] A total of 34 known CNVs previously associated with genomic disorders were identified in 10.5% of this cohort. The most frequent of these CNVs comprised deletions at the 17q12 and 22q11 loci that cause renal disease and diabetes syndrome and DiGeorge syndrome, respectively.[59] In a subsequent study, CNVs >200 kb were found in 10.1% of patients with a broad spectrum of CAKUT conditions, including renal hypoplasia, dysplasia, agenesis, multicystic dysplasia, duplex collecting system, VUR, vesicoureteric or ureteropelvic junction obstruction and posterior urethral valves (Figure 1).[60] These studies suggest that CNVs contribute to the aetiology of CAKUT and implicate the genes within the CNV loci (for example, *KIF26B* and *PBX1* that are associated with kidney defects in mice) and the non-coding regions that might function as essential regulatory elements in kidney development, as contributors to disease.[59,61,62] We conclude that nephrogenesis is sensitive to variation in gene dosage, which underscores the importance of detecting CNVs in patients with congenital abnormalities as part of the diagnostic procedures.

Public databases, such as ECARUCA,[63] DECIPHER,[42] ISCA,[64] PubMed, and the Database of Genomic Variants,[65] are available to check whether newly identified CNVs already exist in healthy or diseased individuals. A systematic analysis of CNVs recorded in the DECIPHER database showed that CNV-associated phenotypes overlap more frequently than expected by chance with those of Mendelian diseases caused by single genes that are located within or adjacent to the reported CNVs.[66] In cases where identified CNVs have not been associated with a genetic disease before or do not contain a gene of potential relevance, the interpretation of their effect in any disease is extremely challenging. Generally, a pathogenic effect is suggested when CNVs are absent from healthy individuals and when there is phenotypic resemblance among patients with overlapping CNVs.[67] From here, further investigation is still required to determine the gene(s) that contributes to the phenotype, which is particularly difficult when the CNV affects the dosage of multiple genes. In this case, extensive functional characterization—such as RNA expression analysis, or functional studies in model organisms—should be undertaken before any claims of causality are made.[67] Functional analyses have been performed in five individuals with multi-organ defects (including renal anomalies) who harbour overlapping microdeletions on 8q24.3, which encompasses only three genes.[68] Genetic knockdown of the three genes within this region in zebrafish revealed that knockdown of *scrib* (a planar cell polarity effector) contributed to coloboma and renal defects, whereas knockdown of the splicing factor *puf60* induced cardiac anomalies.[68] The combined abnormal dosage of *scrib* and *puf60* exacerbated other phenotypes, including craniofacial defects and short stature. Chromosomal imbalances spanning *SCRIB* could, therefore, be a potential cause for renal coloboma syndrome in humans, but whether loss-of-function mutations in *SCRIB* could also cause a similar phenotype in humans still needs to be determined.

Whole genome sequencing

Genetic studies in unexplained diseases are generally shifting from CNV analysis and whole exome sequencing to whole genome sequencing (WGS). Whole exome sequencing can miss interesting variants located in exons that have not been captured and variants in non-coding regions. WGS has become more affordable since its development and enables the detection of all classes of genetic change, including single nucleotide variants, CNVs, and other structural variations, such as translocations.[69] Extensive clinical and genetic investigation of a cohort of 50 patients with severe intellectual disability, involving targeted gene analysis, genomic microarray analysis, and whole exome sequencing, failed to identify any pathogenic mutations.[70] Subsequent utility of WGS in the same patients and their unaffected parents identified *de novo* coding single nucleotide variants and CNVs that affected known candidate genes for intellectual disability in 21 of the patients.[70] As illustrated by this example, not only could WGS improve genetic

diagnosis in CAKUT, but it would also enable the detection of genetic abnormalities in non-coding regions that might confer a pathogenic effect.

Genomic regulatory regions

Chromosomal imbalances and deleterious mutations can affect non-coding regions of DNA that contain regulatory elements, such as enhancers or silencers, which can result in altered gene expressions critical for kidney development and/or CAKUT. This hypothesis is supported by the detection of pathogenic mutations in regulatory genomic regions in patients with congenital malformations. The activity of an enhancer in driving the expression of *PTF1A* was abolished by homozygous or compound heterozygous mutations that were identified by WGS in familial cases of isolated pancreatic agenesis.[71] One of the familial cases had a 7.6 kb deletion that included the entire putative enhancer. *PTF1A* coding mutations were already known to cause syndromic pancreatic agenesis.[71] In the case of congenital heart diseases, putative enhancers within the *TBX5* locus were sequenced in 260 Brazilian children born with isolated congenital heart defects.[72] A homozygous SNP was identified in an enhancer downstream of *TBX5* and was shown to abrogate the ability of the enhancer to drive expression within the heart in transgenic mouse and zebrafish models.[72] Coding mutations in *TBX5* are known to cause Holt–Oram syndrome, which is characterized by heart and limb defects,[73] but the enhancer mutation is associated with isolated cardiac anomalies without limb involvement.[72]

Alterations in microRNA (miRNA) expression might also be involved in the pathogenesis of developmental abnormalities. miRNAs are small non-coding RNAs of ~21 nucleotides that regulate gene expression in a post-transcriptional manner. They hybridize to complementary sites in the 3'-untranslated region of their target mRNAs through their 5'-proximal seed sequence in order to direct mRNA degradation or inhibit protein translation.[74] To date, dysregulation of miRNAs by chromosomal imbalances has mainly been linked to cancers, such as Wilms tumour, where loss of heterozygosity at 2q37 results in a heterozygous deletion of miR-562, which targets *EYA1*, a renal developmental gene.[75] Autosomal dominant progressive hearing loss was the first example of a Mendelian disease associated with a point mutation in an miRNA. Causal mutations were found in the seed sequence of miR-96, which is expressed in the hair cells of the inner ear.[76] Studies on the function of miRNAs in kidney development are currently in their infancy. Upon inhibiting the pathway of miRNA biogenesis in murine renal tubules, multiple defects arise, such as hydroureter, hydronephrosis, tubular and glomerular cysts, low nephron endowment, and ureteropelvic junction obstruction (Figure 1).[77,78] Similarly, podocyte-specific inhibition of miRNAs leads to proteinuria, tubular and glomerular injury, and enhances apoptosis of nephron progenitors.[79] The genes involved in autosomal dominant polycystic kidney disease, *PKD1* and *PKD2*, are regulated by miR-17.[80,81] Cyst growth was successfully ameliorated in a mouse model of polycystic kidney disease upon deletion of the miR-17~92 locus in developing renal tubules. Finally, the miR-92a directly targets the 3'-untranslated region of *HNF1B* and its expression negatively correlates with *Hnf1b* levels in murine kidneys.[81]

Deep RNA-sequencing has identified 51 miRNAs in murine embryonic kidneys, but their function remains undefined.[82] These data suggest that miRNAs are important in fine-tuning the expression of renal developmental genes and their misexpression may contribute to the phenotypic variability observed in CAKUT. RNA sequencing enables the analysis of transcription at a genome-wide level, including the detection of differential expression of miRNAs.[83] NGS is, therefore, becoming the method of choice for investigating the association of any type of genetic factor in CAKUT.

Complex genetic aetiology

CNVs and single gene mutations do not explain the majority of sporadic cases of CAKUT, rather, complex interactions of multiple genetic and environmental factors probably account for a substantial proportion of cases. Association studies using a large number of patients and ethnicity-matched controls have facilitated major advances in unravelling the genetic component of numerous complex diseases. Here, we discuss the advances that have been made in the past decade by association studies in CAKUT and chronic kidney disease (CKD).

Gene-based or genome-wide association studies

Only a limited number of association studies have been published for CAKUT, the majority of which have focused on candidate genes. An association test, known as the transmission disequilibrium test (TDT), provides a family-based method for the identification of risk variants by testing the transmission of alleles from heterozygous parents to affected offspring. A TDT analysis was performed for six genes (*AGTR2*, *HNF1B*, *PAX2*, *RET*, *ROBO2,* and *UPK3A*) in affected patients with VUR from the UK and Slovenia, but no evidence for an association between common variants in these genes and VUR was found.[84] Another approach was performed in a Dutch population of 207 patients with VUR and 554 controls, where the association between VUR and common genetic variants in 44 genes that are assumed to contribute to ureteric budding was investigated.[85] The top 14 SNPs that reached significance were subsequently included in a follow-up analysis in 202 patients and 892 controls but none reached genome-wide significance.[85] The top genes in this study were *GREM1*, *EYA1*, and *ROBO2* for VUR, and *EYA1* and *UPK3A* for a duplex collecting system. Single gene-based association studies found polymorphisms in *MMP3*, *BMP4,* and *AGTR2* to be associated with an increased risk of CAKUT.[86–88]

High-throughput genotyping technologies allow the rapid genotyping of >1 million SNPs across the whole genome and enable genome-wide association studies (GWAS) in a hypothesis free, unbiased manner. The only study to date that has exploited these advantages conducted a genome-wide TDT analysis in 410 patient–parent trios from Ireland using 643,691 SNPs,

as well as a GWAS in 500 unrelated patients with VUR and 851 controls using 580,000 SNPs. No SNP was significantly associated with VUR using either approach.[89] A GWAS performed in 494 controls and 436 cases with hypospadias (a congenital malformation of the male external genitalia that affects ~1 in 750 births) showed a significant association between common variants in *DGKK* and hypospadias, a result that was confirmed in two replication cohorts.[90] Association studies might not have been as successful in CAKUT because specific sub-phenotypes, such as VUR, are common malformations that need larger sample sizes in order to increase the statistical power to detect genome-wide significance. Increases in sample size might facilitate GWAS to identify associated loci that confer risk of CAKUT, as shown with recent GWAS in CKD, where cohorts of >50,000 individuals with CKD enabled the detection of risk variants in *APOL1*,[91–92] *UMOD*, *SHROOM3,* and *STC1*.[93,94]

Environmental factors

The complex aetiology of CAKUT implies that both genetic and environmental factors contribute to the natural history of disease; we thus also need to understand the effects of several environmental factors before and during pregnancy to elucidate the pathogenesis (Figure 2).[95] A case-control study in Colorado consisting of 189 infants with renal agenesis and 940 control infants, investigated the association of renal agenesis with maternal and paternal age, maternal weight gain, maternal diabetes mellitus, infant birth weight, and time since last delivery.[96] Renal agenesis was significantly associated with pre-gestational maternal diabetes mellitus (OR = 4.98; 95% CI 1.08–22.93) after adjusting for maternal age, ethnicity, and alcohol exposure. Children born to African American women were also more likely to develop renal agenesis (OR = 2.19; 95% CI 1.26–3.95). None of the other measured factors, such as age or alcohol consumption, reached significance.[96] A larger case-control study in Washington involved 20,032 control participants and 1,994 patients who were diagnosed with renal dysplasia and/or agenesis and obstructive uropathy, and developed CKD by 21 years of age.[97] A range of maternal characteristics were measured, including age, race, education, urban or rural residence, smoking, previous pregnancies, diabetes mellitus, and body mass index (BMI). The adjusted odds ratios for CKD associated with maternal characteristics were 1.54 for pre-gestational diabetes mellitus (95% CI 1.13–2.09), 1.24 for maternal overweight (95% CI 1.05–1.48), 1.26 for maternal obesity (95% CI 1.05–1.52), and 2.88 for low birth weight (95% CI 2.28–3.63).[97] Consistent with these results, a Canadian study published an additional evaluation of maternal diabetes mellitus associated with CAKUT, where it was refined into pre-gestational and gestational diabetes mellitus depending on whether it was diagnosed before or after 20 weeks in pregnancy, respectively. Early exposure to diabetes mellitus increased the risk of CAKUT, with an odds ratio of 1.67 (95% CI 1.14–2.46).[98] The effect of pre-gestational diabetes mellitus warrants monitoring of glycaemia in pregnant women and long-term follow-up of renal function in the offspring. Animal studies also support the effect of maternal diabetes mellitus on CAKUT risk, as hyperglycaemia is negatively correlated with nephron endowment in murine models.[99] These models have also indicated an association of calorie and protein intake, and gestational hypoxia with reduced nephron endowment, hypertension, and microalbuminuria.[100,101]

Epigenetics

Epigenetic modifications have been postulated as a mechanism that facilitates the interaction between environmental factors during development with the genome, and its impact on disease susceptibility. Epigenetics refers to heritable changes that are not caused by alterations in the nucleotide sequence itself. The most common studied mechanisms of epigenetic modification include DNA methylation at CpG dinucleotides, histone acetylation, histone phosphorylation, and histone methylation. These modifications modulate the structure of the chromatin and change its accessibility to transcription factors, allowing certain regions to become accessible (or inaccessible) to transcription factors and consequently for genes to be expressed or silenced (Figure 2). As previously discussed, gene expression levels during development can be affected by CNVs or single nucleotide changes located within enhancers and miRNAs. In contrast, epigenetic modifications are reversible and susceptible to environmental stress, thus enabling developmental and temporal variability in gene expression patterns.[102]

The mechanisms underlying the effects of environmental factors on epigenetic marks are not yet fully understood, but several examples exist that support such an interplay. Maternal depression has been observed to influence the epigenetic marks on genes that are crucial during human development.[103] The methylation status of ~1,400 genetic regions were shown to vary across 237 neonates and 75% of the variation could be explained by differences in the *in utero* environment, such as maternal smoking, depression, BMI, and infant birth weight, whereas the genotype alone could explain only 25% of the variation.[104] In kidney fibrosis, TGF-β1 increases the expression of histone methyltransferases, leading to an increase in histone methylation at H3K4 at critical pro-fibrotic gene promoters and consequently increasing their expression.[105] A genome-wide cytosine methylation survey in samples of tubules from healthy individuals and patients with CKD showed that differentially methylated regions mainly overlap with putative enhancer regions and contain binding motifs for kidney-specific transcription factors, such as *SIX2*.[106] The subsequent differences in gene expression correlated with changes in the methylation pattern and affected numerous genes in the TGF-β pathway.[106]

Epigenetic changes might also contribute to the development of CKD by influencing developmental and pro-fibrotic pathways.[106] Interestingly, *PAX2* functions in an epigenetic network that determines tissue specificity by recruiting PTIP, a ubiquitous nuclear protein that acts as a H3K4 methyltransferase cofactor, thereby promoting the assembly of a histone H3K4 methyltransferase

complex at DNA target sites of *PAX2*.[107] Podocyte-specific loss of *Ptip* in mice resulted in glomerular sclerosis and mesangial expansion later in life.[108] Regulation of histone methylation in a spatial and temporal manner can therefore be achieved via developmental and tissue-specific DNA-binding proteins, such as *PAX2,* which induce the intermediate mesoderm to become kidney epithelial cells. Altogether, these data suggest that epigenetic changes could be instructive and predetermine disease susceptibility, or they could activate certain pathways in response to renal injury.

Phenotypically discordant monozygotic twins have become an important model in epigenetic studies as they offer the opportunity to control for potential confounders, such as differences in genetic background or *in utero* environment.[109] Genetic differences between monozygotic twins might occur due to somatic mosaicism, as well as epigenetic changes due to environmental differences. A pair of 30-year-old monozygotic twins who were discordant for congenital renal agenesis, were analyzed by whole exome sequencing, CNV analysis based on exome data, and reduced-representation bisulfite sequencing, in order to identify differential patterns in DNA methylation.[110] The analyses revealed 514 differentially methylated regions between the twins, but no differential single nucleotide variants or CNVs. Although regulatory regions of the genome were not checked for differences at the sequence level, this study suggests that environmental factors can contribute to the pathogenesis of CAKUT via epigenetic modifications.[110]

Based on these data, epigenetic studies in CAKUT present novel opportunities to understand disease mechanisms as they might explain the interaction between genetic and environmental risk factors and how variable genetic penetrance might arise. An investigation into the association between epigenetic changes and CAKUT in a large case-control study should now be performed. In addition, elucidating the patterns of epigenetic modifications in fetal and adult kidney tissues will provide insight into their potential role in kidney development and function and will provide information about the genes that are controlled by epigenetic machinery in response to environmental exposures. Understanding the epigenetic basis of kidney development might provide new insights into the pathogenesis of CAKUT. The ultimate goal is to improve clinical practice by identifying epigenetic biomarkers that can serve as predictive indicators and can be modified by therapeutic drugs. The heterogeneity of epigenetic alterations between tissue types and epigenetic variations between development and adulthood, however, make it difficult to collect appropriate samples, thus creating a considerable challenge for epigenetic studies on CAKUT in humans.

Conclusions

The development of new genetic techniques, such as NGS, array–CGH, and microarrays for CNV detection have enabled the identification of plausible pathogenic variants and CNVs in CAKUT and have revealed new pathways involved in the pathogenesis of structural renal abnormalities. These genetic findings provide an opportunity to develop an ideal diagnostic CAKUT genetic test in clinical practice to facilitate early diagnosis, better management of the disease, and genetic counselling. Several challenges must first be overcome to reach these aims. As the list of identified variants continue to grow, the development of high-throughput functional assays is becoming more vital for the discrimination of rare polymorphisms from truly disease-causing variants. Genome-wide studies with larger numbers of population-matched affected and control individuals are required in order to improve the power for detection of variants throughout the genome and to allow for statistical testing of the association of variants in new genes with CAKUT. Sharing samples and sequencing data with other laboratories would facilitate this required increase in patient numbers and statistical power. In addition, the combination of traditional techniques with more modern techniques, such as linkage analysis, array–CGH, methylome profiling, and WGS, should be the preferred choice to study familial cases of CAKUT with an unexplained cause. Furthermore, to fully comprehend the contribution of disease-associated variants to the development of the disease and understand how disease progression is influenced by the environment and epigenetic mechanisms, multi-disciplinary efforts are required to collect and integrate data from epidemiologic, genomic, epigenomic, transcriptomic, and proteomic studies. Such efforts will not only help unravel the pathogenesis of CAKUT, but will also identify new avenues for attractive therapeutic targets, such as drugs that can target epigenetic mechanisms that can be altered in individual cell types. Despite the complexity in understanding the genetics of CAKUT, the clinical application of such findings will be extremely valuable in patient and family counselling, symptom management, prevention of CKD, and the design of new therapeutic agents.

1. Brown, T., Mandell, J. & Lebowitz, R. L. Neonatal hydronephrosis in the era of sonography. *Am. J. Roentgenol.* **148,** 959–963 (1987).
2. Queisser-Luft, A., Stolz, G., Wiesel, A., Schlaefer, K. & Spranger, J. Malformations in newborn: results based on 30,940 infants and fetuses from the Mainz congenital birth defect monitoring system (1990–1998). *Arch. Gynecol. Obstet.* **266,** 163–167 (2002).
3. Sanna-Cherchi, S. *et al.* Genetic approaches to human renal agenesis/hypoplasia and dysplasia. *Pediatr. Nephrol.* **22,** 1675–1684 (2007).
4. Pope, J. C. 4th, Brock, J. W. 3rd, Adams, M. C., Stephens, F. D. & Ichikawa, I. How they begin and how they end: classic and new theories for the development and deterioration of congenital anomalies of the kidney and urinary tract, CAKUT. *J. Am. Soc. Nephrol.* **10,** 2018–2028 (1999).
5. Vivante, A., Kohl, S., Hwang, D. Y., Dworschak, G. C. & Hildebrandt, F. Single-gene causes of congenital anomalies of the kidney and urinary tract (CAKUT) in humans. *Pediatr. Nephrol.* **29,** 695–704 (2014).
6. Schedl, A. Renal abnormalities and their developmental origin. *Nat. Rev. Genet.* **8,** 791–802 (2007).
7. Blake, J. & Rosenblum, N. D. Renal branching morphogenesis: morphogenetic and signaling mechanisms. *Semin. Cell Dev. Biol.* **36,** 2–12 (2014).
8. Chesnaye, N. *et al.* Demographics of paediatric renal replacement therapy in Europe: a report of the ESPN/ERA–EDTA registry. *Pediatr. Nephrol.* **29,** 2403–2410 (2014).
9. Wuhl, E. *et al.* Timing and outcome of renal replacement therapy in patients with congenital malformations of the kidney and urinary tract. *Clin. J. Am. Soc. Nephrol.* **8,** 67–74 (2013).
10. Kerecuk, L., Schreuder, M. F. & Woolf, A. S. Renal tract malformations: perspectives for

nephrologists. *Nat. Clin. Pract. Nephrol.* **4**, 312–325 (2008).

11. Stoll, C., Dott, B., Alembik, Y. & Roth, M. P. Associated nonurinary congenital anomalies among infants with congenital anomalies of kidney and urinary tract (CAKUT). *Eur. J. Med. Genet.* **57**, 322–328 (2014).

12. Winyard, P. & Chitty, L. S. Dysplastic kidneys. *Semin. Fetal Neonatal Med.* **13**, 142–151 (2008).

13. Bulum, B. *et al.* High frequency of kidney and urinary tract anomalies in asymptomatic first-degree relatives of patients with CAKUT. *Pediatr. Nephrol.* **28**, 2143–2147 (2013).

14. Monn, E. & Nordshus, T. Hereditary renal adysplasia. *Acta. Paediatr. Scand*. **73**, 278–280 (1984).

15. McPherson, E. *et al.* Dominantly inherited renal adysplasia. *Am. J. Med. Genet.* **26**, 863–872 (1987).

16. Kaplan, B. S., Milner, L. S., Jequier, S., Kaplan, P. & de Chadarevian, J. P. Autosomal dominant inheritance of small kidneys. *Am. J. Med. Genet.* **32**, 120–126 (1989).

17. Doray, B., Gasser, B., Reinartz, I. & Stoll, C. Hereditary renal adysplasia in a three generations family. *Genet. Couns.* **10**, 251–257 (1999).

18. Sanyanusin, P. *et al.* Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. *Nat. Genet.* **9**, 358–364 (1995).

19. Horikawa, Y. *et al.* Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat. Genet.* **17**, 384–385 (1997).

20. Weber, S. *et al.* Prevalence of mutations in renal developmental genes in children with renal hypodysplasia: results of the ESCAPE study. *J. Am. Soc. Nephrol.* **17**, 2864–2870 (2006).

21. Thomas, R. *et al.* HNF1B and PAX2 mutations are a common cause of renal hypodysplasia in the CKiD cohort. *Pediatr. Nephrol.* **26**, 897–903 (2011).

22. Madariaga, L. *et al.* Severe prenatal renal anomalies associated with mutations in HNF1B or PAX2 genes. *Clin. J. Am. Soc. Nephrol.* **8**, 1179–1187 (2013).

23. Hwang, D. Y. *et al.* Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary tract. *Kidney Int.* **85**, 1429–1433 (2014).

24. Saisawat, P. *et al.* Identification of two novel CAKUT-causing genes by massively parallel exon resequencing of candidate genes in patients with unilateral renal agenesis. *Kidney Int.* **81**, 196–200 (2012).

25. Exome Aggregation Consortium (ExAC), http://exac.broadinstitute.org [June 2015 accessed].

26. Weber, S. *et al.* SIX2 and BMP4 mutations associate with anomalous kidney development. *J. Am. Soc. Nephrol.* **19**, 891–903 (2008).

27. Skinner, M. A., Safford, S. D., Reeves, J. G., Jackson, M. E. & Freemerman, A. J. Renal aplasia in humans is associated with RET mutations. *Am. J. Hum. Genet.* **82**, 344–351 (2008).

28. Gimelli, S. *et al.* Mutations in SOX17 are associated with congenital anomalies of the kidney and the urinary tract. *Hum. Mutat.* **31**, 1352–1359 (2010).

29. Vivante, A. *et al.* Renal hypodysplasia associates with a WNT4 variant that causes aberrant canonical WNT signaling. *J. Am. Soc. Nephrol.* **24**, 550–558 (2013).

30. Coon, K. D. *et al.* A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J. Clin. Psychiatry* **68**, 613–618 (2007).

31. Julier, C. *et al.* Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. *Nature* **354**, 155–159 (1991).

32. Hall, J. M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).

33. Williams, G., Fletcher, J. T., Alexander, S. I. & Craig, J. C. Vesicoureteral reflux. *J. Am. Soc. Nephrol.* **19**, 847–862 (2008).

34. Weng, P. L. *et al.* A recessive gene for primary vesicoureteral reflux maps to chromosome 12p11-q13. *J. Am. Soc. Nephrol.* **20**, 1633–1640 (2009).

35. Ashraf, S. *et al.* Mapping of a new locus for congenital anomalies of the kidney and urinary tract on chromosome 8q24. *Nephrol. Dial. Transplant.* **25**, 1496–1501 (2010).

36. Sanna-Cherchi, S. *et al.* Localization of a gene for nonsyndromic renal hypodysplasia to chromosome 1p32–33 *Am. J. Hum. Genet.* **80**, 539–49 (2007).

37. Chatterjee, R. *et al.* Traditional and targeted exome sequencing reveals common, rare and novel functional deleterious variants in RET-signaling complex in a cohort of living US patients with urinary tract malformations. *Hum. Genet.* **131**, 1725–1738 (2012).

38. Kohl, S. *et al.* Mild recessive mutations in six Fraser syndrome-related genes cause isolated congenital anomalies of the kidney and urinary tract. *J. Am. Soc. Nephrol.* **25**, 1917–1922 (2014).

39. Humbert, C. *et al.* Integrin alpha 8 recessive mutations are responsible for bilateral renal agenesis in humans. *Am. J. Hum. Genet.* **94**, 288–294 (2014).

40. Saisawat, P. *et al.* Whole-exome resequencing reveals recessive mutations in TRAP1 in individuals with CAKUT and VACTERL association. *Kidney Int.* **85**, 1310–1317 (2014).

41. PhenomeCentral Version 1.0 Milestone-7. *A hub for secure data sharing within the rare disorder community* [online], https://phenomecentral.org (2015).

42. Swaminathan, G. J. *et al.* DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. *Hum. Mol. Genet.* **21**, R37–44 (2012).

43. Rehm, H. L. *et al.* ClinGen - The Clinical Genome Resource. *N. Engl. J. Med.* (2015).

44. Fokkema, I. F. *et al.* LOVD v.2.0: the next generation in gene variant databases. *Hum. Mutat.* **32**, 557–563 (2011).

45. Matchmaker Exchange. *Genomic discovery through the exchange of phenotypic & genotypic profiles* [online], http://matchmakerexchange.org (2015).

46. Sanna-Cherchi, S. *et al.* Mutations in DSTYK and dominant urinary tract malformations. *N. Engl. J. Med.* **369**, 621–629 (2013).

47. Vissers, L. E. *et al.* A *de novo* paradigm for mental retardation. *Nat. Genet.* **42**, 1109–1112 (2010).

48. Girard, S. L. *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nat. Genet.* **43**, 860–863 (2011).

49. O'Roak, B. J. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. *Nat. Genet.* **43**, 585–589 (2011).

50. Bower, M. *et al.* Update of PAX2 mutations in renal coloboma syndrome and establishment of a locus-specific database. *Hum. Mutat.* **33**, 457–466 (2012).

51. Serra-Juhe, C. *et al.* Contribution of rare copy number variants to isolated human malformations. *PLoS ONE* **7**, e45530 (2012).

52. Anad, F. *et al.* Alagille syndrome and deletion of 20p. *J. Med. Genet.* **27**, 729–737 (1990).

53. Davies, A. F. *et al.* Delineation of two distinct 6p deletion syndromes. *Hum. Genet.* **104**, 64–72 (1999).

54. Kirk, J. M. *et al.* Unilateral renal aplasia in X-linked Kallmann's syndrome. *Clin. Genet.* **46**, 260–262 (1994).

55. Mefford, H. C. *et al.* Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am. J. Hum. Genet.* **81**, 1057–1069 (2007).

56. Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12**, 363–376 (2011).

57. de Ligt, J. *et al.* Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum. Mutat.* **34**, 1439–1448 (2013).

58. Hoefele, J. *et al.* A novel interstitial deletion of 10q24.2q24.32 in a patient with renal coloboma syndrome. *Eur. J. Med. Genet.* **55**, 211–215 (2012).

59. Sanna-Cherchi, S. *et al.* Copy-number disorders are a common cause of congenital kidney malformations. *Am. J. Hum. Genet.* **91**, 987–997 (2012).

60. Caruana, G. *et al.* Copy-number variation associated with congenital anomalies of the kidney and urinary tract. *Pediatr. Nephrol.* **30**, 487–495 (2015).

61. Uchiyama, Y. *et al.* Kif26b, a kinesin family gene, regulates adhesion of the embryonic kidney mesenchyme. *Proc. Natl Acad. Sci. USA* **107**, 9240–9245 (2010).

62. Selleri, L. *et al.* Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* **128**, 3543–3557 (2001).

63. Vulto-van Silfhout, A. T. *et al.* An update on ECARUCA, the European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations. *Eur. J. Med. Genet.* **56**, 471–474 (2013).

64. Riggs, E. R., Jackson, L., Miller, D. T. & Van Vooren, S. Phenotypic information in genomic variant databases enhances clinical care and research: the International Standards for Cytogenomic Arrays Consortium experience. *Hum. Mutat.* **33**, 787–796 (2012).

65. MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* **42**, D986–992 (2014).

66. Ibn-Salem, J. *et al.* Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* **15**, 423 (2014).

67. Kloosterman, W. P. & Hochstenbach, R. Deciphering the pathogenic consequences of chromosomal aberrations in human genetic disease. *Mol. Cytogenet.* **7**, 100 (2014).

68. Dauber, A. *et al.* SCRIB and PUF60 are primary drivers of the multisystemic phenotypes of the 8q24.3 copy-number variant. *Am. J. Hum. Genet.* **93**, 798–811 (2013).

69. Francioli, L. C. *et al.* Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–825 (2014).

70. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).

71. Weedon, M. N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* **46**, 61–64 (2014).

72. Smemo, S. *et al.* Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* **21**, 3255–3263 (2012).

73. Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat. Genet.* **15**, 21–29 (1997).

74. Kawahara, Y. Human diseases caused by germline and somatic abnormalities in microRNA and microRNA-related genes. *Congenit. Anom. (Kyoto)* **54**, 12–21 (2014).

75. Drake, K. M. *et al.* Loss of heterozygosity at 2q37 in sporadic Wilms' tumor: putative role for miR-562. *Clin. Cancer Res.* **15**, 5985–5992 (2009).

76. Mencia, A. *et al.* Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat. Genet.* **41**, 609–613 (2009).

77. Patel, V. *et al.* MicroRNAs regulate renal tubule maturation through modulation of Pkd1. *J. Am. Soc. Nephrol.* **23**, 1941–1948 (2012).

78. Bartram, M. P. *et al.* Conditional loss of kidney microRNAs results in congenital anomalies of the kidney and urinary tract (CAKUT). *J. Mol. Med.* **91**, 739–748 (2013).

79. Ho, J. *et al.* Podocyte-specific loss of functional microRNAs leads to rapid glomerular and tubular injury. *J. Am. Soc. Nephrol.* **19**, 2069–2075 (2008).

80. Sun, H. *et al.* MicroRNA-17 post-transcriptionally regulates polycystic kidney disease-2 gene and promotes cell proliferation. *Mol. Biol. Rep.* **37**, 2951–2958 (2010).

81. Patel, V. *et al.* miR-17~92 miRNA cluster promotes kidney cyst growth in polycystic kidney disease. *Proc. Natl Acad. Sci. USA* **110**, 10765–10770 (2013).

82. Thiagarajan, R. D. *et al.* Refining transcriptional programs in kidney development by integration of deep RNA-sequencing and array-based spatial profiling. *BMC Genomics* **12**, 441 (2011).

83. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).

84. Cordell, H. J. *et al.* Whole-genome linkage and association scan in primary, nonsyndromic vesicoureteric reflux. *J. Am. Soc. Nephrol.* **21**, 113–123 (2010).

85. van Eerde, A. M. *et al.* Genes in the ureteric budding pathway: association study on vesico-ureteral reflux patients. *PLoS ONE* **7**, e31327 (2012).

86. Djuric, T. *et al.* MMP-1 and -3 haplotype is associated with congenital anomalies of the kidney and urinary tract. *Pediatr. Nephrol.* **29**, 879–884 (2014).

87. Reis, G. S. *et al.* Study of the association between the BMP4 gene and congenital anomalies of the kidney and urinary tract. *J. Pediatr. (Rio J.)* **90**, 58–64 (2014).

88. Miranda, D. M. *et al.* Association of angiotensin type 2 receptor gene polymorphisms with ureteropelvic junction obstruction in Brazilian patients. *Nephrology (Carlton)* **19**, 714–720 (2014).

89. Darlow, J. M. *et al.* A new genome scan for primary nonsyndromic vesicoureteric reflux emphasizes high genetic heterogeneity and shows linkage and association with various genes already implicated in urinary tract development. *Mol. Genet. Genomic Med.* **2**, 7–29 (2014).

90. van der Zanden, L. F. *et al.* Common variants in DGKK are strongly associated with risk of hypospadias. *Nat. Genet.* **43**, 48–50 (2011).

91. Genovese, G. *et al.* Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329**, 841–845 (2010).

92. Kopp, J. B. *et al.* Clinical features and histology of apolipoprotein L1-associated nephropathy in the FSGS clinical trial. *J. Am. Soc. Nephrol.* (2015).

93. Kottgen, A. *et al.* Multiple loci associated with indices of renal function and chronic kidney disease. *Nat. Genet.* **41**, 712–717 (2009).

94. Kottgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat. Genet.* **42**, 376–384 (2010).

95. Woroniecki, R., Gaikwad, A. B. & Susztak, K. Fetal environment, epigenetics, and pediatric renal disease. *Pediatr. Nephrol.* **26**, 705–711 (2011).

96. Parikh, C. R., McCall, D., Engelman, C. & Schrier, R. W. Congenital renal agenesis: case-control analysis of birth characteristics. *Am. J. Kidney Dis.* **39**, 689–694 (2002).

97. Hsu, C. W., Yamamoto, K. T., Henry, R. K., De Roos, A. J. & Flynn, J. T. Prenatal risk factors for childhood CKD. *J. Am. Soc. Nephrol.* **25**, 2105–2111 (2014).

98. Dart, A. B., Ruth, C. A., Sellers, E. A., Au, W. & Dean, H. J. Maternal diabetes mellitus and congenital anomalies of the kidney and urinary tract (CAKUT) in the child. *Am. J. Kidney Dis.* **65**, 684–691 (2015).

99. Amri, K., Freund, N., Vilar, J., Merlet-Benichou, C. & Lelievre-Pegorier, M. Adverse effects of hyperglycemia on kidney development in rats: *in vivo* and *in vitro* studies. *Diabetes* **48**, 2240–2245 (1999).

100. Hoppe, C. C., Evans, R. G., Bertram, J. F. & Moritz, K. M. Effects of dietary protein restriction on nephron number in the mouse. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **292**, R1768–R1774 (2007).

101. Wilkinson, L. J. *et al.* Renal developmental defects resulting from *in utero* hypoxia are associated with suppression of ureteric beta-catenin signaling. *Kidney Int.* **87**, 975–983 (2015).

102. Feil, R. & Fraga, M. F. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* **13**, 97–109 (2011).

103. Oberlander, T. F. *et al.* Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics* **3**, 97–106 (2008).

104. Teh, A. L. *et al.* The effect of genotype and *in utero* environment on interindividual variation in neonate DNA methylomes. *Genome Res.* **24**, 1064–1074 (2014).

105. Sun, G. *et al.* Epigenetic histone methylation modulates fibrotic gene expression. *J. Am. Soc. Nephrol.* **21**, 2069–2080 (2010).

106. Ko, Y. A. *et al.* Cytosine methylation changes in enhancer regions of core pro-fibrotic genes characterize kidney fibrosis development. *Genome Biol.* **14**, R108 (2013).

107. Patel, S. R., Kim, D., Levitan, I. & Dressler, G. R. The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. *Dev. Cell* **13**, 580–592 (2007).

108. Lefevre, G. M., Patel, S. R., Kim, D., Tessarollo, L. & Dressler, G. R. Altering a histone H3K4 methylation pathway in glomerular podocytes promotes a chronic disease phenotype. *PLoS Genet.* **6**, e1001142 (2010).

109. Castillo-Fernandez, J. E., Spector, T. D. & Bell, J. T. Epigenetics of discordant monozygotic twins: implications for disease. *Genome Med.* **6**, 60 (2014).

110. Jin, M. *et al.* Genomic and epigenomic analyses of monozygotic twins discordant for congenital renal agenesis. *Am. J. Kidney Dis.* **64**, 119–122 (2014).

111. Gribouval, O. *et al.* Spectrum of mutations in the renin-angiotensin system genes in autosomal recessive renal tubular dysgenesis. *Hum. Mutat.* **33**, 316–326 (2012).

112. Gribouval, O. *et al.* Mutations in genes in the renin-angiotensin system are associated with autosomal recessive renal tubular dysgenesis. *Nat. Genet.* **37**, 964–968 (2005).

113. Weber, S. *et al.* Muscarinic acetylcholine receptor M3 mutation causes urinary bladder disease and a prune-belly-like syndrome. *Am. J. Hum. Genet.* **89**, 668–674 (2011).

114. Alazami, A. M. *et al.* FREM1 mutations cause bifid nose, renal agenesis, and anorectal malformations syndrome. *Am. J. Hum. Genet.* **85**, 414–418 (2009).

115. Vogel, M. J. *et al.* Mutations in GRIP1 cause Fraser syndrome. *J. Med. Genet.* **49**, 303–306 (2012).

116. Daly, S. B. *et al.* Mutations in HPSE2 cause urofacial syndrome. *Am. J. Hum. Genet.* **86**, 963–969 (2010).

117. Li, Y. *et al.* LRP4 mutations alter Wnt/beta-catenin signaling and cause limb and kidney malformations in Cenani-Lenz syndrome. *Am. J. Hum. Genet.* **86**, 696–706 (2010).

118. Tufan, F. *et al.* Clinical and molecular characterization of two adults with autosomal recessive Robinow syndrome. *Am. J. Med. Genet. A* **136**, 185–189 (2005).

119. Kraus, M. R. *et al.* Two mutations in human BICC1 resulting in Wnt pathway hyperactivity associated with cystic renal dysplasia. *Hum. Mutat.* **33**, 86–90 (2012).

120. Tabatabaeifar, M. *et al.* Functional analysis of BMP4 mutations identified in pediatric CAKUT patients. *Pediatr. Nephrol.* **24**, 2361–2368 (2009).

121. Unger, S. *et al.* Mutations in the cyclin family member FAM58A cause an X-linked dominant disorder characterized by syndactyly, telecanthus and anogenital and renal malformations. *Nat. Genet.* **40**, 287–289 (2008).

122. Trarbach, E. B. *et al.* Nonsense mutations in FGF8 gene causing different degrees of human gonadotropin-releasing deficiency. *J. Clin. Endocrinol. Metab.* **95**, 3491–3496 (2010).

123. Dode, C. *et al.* Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. *Nat. Genet.* **33**, 463–465 (2003).

124. Ali, A. *et al.* Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into mechanisms of DNA binding by the GATA3 transcription factor. *Hum. Mol. Genet.* **16**, 265–275 (2007).

125. Johnston, J. J. *et al.* Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister-Hall syndromes: robust phenotype prediction from the type and position of GLI3 mutations. *Am. J. Hum. Genet.* **76**, 609–622 (2005).

126. Heidet, L. *et al.* Spectrum of HNF1B mutations in a large cohort of patients who harbor renal diseases. *Clin. J. Am. Soc. Nephrol.* **5**, 1079–1090 (2010).

127. Barbacci, E. *et al.* HNF1beta/TCF2 mutations impair transactivation potential through altered co-regulator recruitment. *Hum. Mol. Genet.* **13**, 3139–3149 (2004).

128. Guegan, K., Stals, K., Day, M., Turnpenny, P. & Ellard, S. JAG1 mutations are found in approximately one third of patients presenting

with only one or two clinical features of Alagille syndrome. *Clin. Genet.* **82**, 33–40 (2012).

129. Albuisson, J. *et al.* Kallmann syndrome: 14 novel mutations in KAL1 and FGFR1 (KAL2). *Hum. Mutat.* **25**, 98–99 (2005).

130. Kamath, B. M. *et al.* NOTCH2 mutations in Alagille syndrome. *J. Med. Genet.* **49**, 138–144 (2012).

131. Favor, J. *et al.* The mouse Pax2(1Neu) mutation is identical to a human PAX2 mutation in a family with renal-coloboma syndrome and results in developmental defects of the brain, ear, eye, and kidney. *Proc. Natl Acad. Sci. USA* **93**, 13870–13875 (1996).

132. Meeus, L. *et al.* Characterization of a novel loss of function mutation of PAX8 in a familial case of congenital hypothyroidism with in-place, normal-sized thyroid. *J. Clin. Endocrinol. Metab.* **89**, 4285–4291 (2004).

133. Bertoli-Avella, A. M. *et al.* ROBO2 gene variants are associated with familial vesicoureteral reflux. *J. Am. Soc. Nephrol.* **19**, 825–831 (2008).

134. Yoshida, Y. *et al.* Increased levels of pigment epithelium-derived factor in aqueous humor of patients with uveitis. *Br. J. Ophthalmol.* **91**, 149–150 (2007).

135. Hanchate, N. K. *et al.* SEMA3A, a gene involved in axonal pathfinding, is mutated in patients with Kallmann syndrome. *PLoS Genet.* **8**, e1002896 (2012).

136. Ruf, R. G. *et al*. SIX1 mutations cause branchio-oto-renal syndrome by disruption of EYA1-SIX1-DNA complexes. *Proc. Natl Acad. Sci. USA* **101**, 8090–8095 (2004).

137. Hoskins, B. E. *et al.* Transcription factor SIX5 is mutated in patients with branchio-oto-renal syndrome. *Am. J. Hum. Genet.* **80**, 800–804 (2007).

138. Li, H., Sheridan, R. & Williams, T. Analysis of TFAP2A mutations in Branchio-Oculo-Facial Syndrome indicates functional complexity within the AP-2α DNA-binding domain. *Hum. Mol. Genet*. **22**, 3195–3206 (2013).

139. Jenkins, D. *et al. De novo* Uroplakin IIIa heterozygous mutations cause human renal adysplasia leading to severe kidney failure. *J. Am. Soc. Nephrol.* **16**, 2141–2149 (2005).

REVIEWS