

Genetic Studies of Familial Vesicoureteral Reflux

Zsuzsa Ingulf Bartik

Department of Pediatrics
Institute of Clinical Sciences
Sahlgrenska Academy, University of Gothenburg



UNIVERSITY OF GOTHENBURG

Gothenburg 2018

Cover illustration: Searching for the VUR gene by Eszter Kónya

Genetic Studies of Familial Vesicoureteral Reflux

© Zsuzsa Ingulf Bartik 2018

zsuzsa.bartik@vgregion.se

ISBN 978-91-7833-091-1 (PRINT)

ISBN 978-91-7833-092-8 (PDF)

Printed in Gothenburg, Sweden 2018

Printed by BrandFactory

To my wonderful family

Genetic Studies of Familial Vesicoureteral Reflux

Zsuzsa Ingulf Bartik

Department of Pediatrics, Institute of Clinical Sciences
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden

ABSTRACT

Vesicoureteral reflux (VUR) is a common congenital anomaly with a high risk of recurrent urinary tract infections (UTI) and, as a consequence, scarring of the renal parenchyma. Additionally, high-grade reflux is often associated with congenital renal damage (hypodysplasia). A clear heredity is seen, although genetic factors are only known for a minority of cases. The aim of this thesis was to study the heritability and genetic contribution as well as to compare the differences between familial and sporadic VUR.

Study I compared clinical data from familial VUR with sporadic cases. Out of the 726 children with reflux that have been treated at Queen Silvia Children's Hospital between 1990 and 2004, 99 individuals (from 66 families) have reported relatives with VUR. A strong overrepresentation of maternal transmission of VUR was seen. The phenotype of VUR did not differ between familial and non-familial cases.

Study II investigated the contribution of *ROBO2* and *SLIT2* genes in familial VUR through mutation screening by direct sequencing in 54 unrelated patients with primary VUR. Six sequence variants were observed in *ROBO2* gene in the exon–intron boundary area, two of which were new, but none of them altered gene splicing. One *SLIT2* missense mutation was detected and predicted to alter the secondary structure of the protein. However, this variant did not segregate with VUR in the family. Gene variants in *ROBO2* and *SLIT2* are rare causes of VUR in humans.

Study III investigated 14 families from south-western Sweden with 3 or more affected members with primary VUR for shared genomic regions, possibly inherited from a common ancestor, and for recurrent copy-number variants in the families. A high-density SNP array was used for genotyping affected individuals and four controls. We found no unique haplotype region shared by most of the families, thus common founder mutation was excluded.

However, subset of families shared different regions, six of them corresponding to previous linkage studies. We presented the genes and non-coding elements relevant for urinary tract development that are located within these regions. One CNV, a deletion at 5q31.1, segregated with VUR and hypodysplasia in one of the investigated families.

Study IV analysed 13 of the above-mentioned 14 families by whole-exome sequencing (WES) in order to find disease causing gene mutations. The findings were confirmed with segregation analysis based on Sanger sequencing in the whole family. We identified three novel variants that might affect function, in *LAMC1*, *KIF26B* and *LIFR* genes, in three families. *SALL1*, *ROBO2* and *UPK3A* gene variants, predicted to be deleterious, were excluded by segregation analysis. In all, we demonstrated likely causal gene mutation in 23% of the families.

In conclusion, severity of the disease did not differ between familial and non-familial VUR. Our studies show that VUR is a genetically highly heterogeneous malformation. WES in combination with a segregation study is a useful tool when it comes to confirming variants in known candidate genes and identifying new genes that might be involved in the pathogenesis of VUR.

Keywords: Vesicoureteral reflux, Heredity, Genetic heterogeneity, Phenotype, Renal hypodysplasia, Renal development, Genome-wide association studies, Single nucleotide polymorphism, Haplotype sharing, DNA copy number variations, Whole exome sequencing

ISBN 978-91-629-7833-091-1 (PRINT)

ISBN 978-91-629-7833-092-8 (PDF)

<http://hdl.handle.net/2077/56353>

SAMMANFATTNING PÅ SVENSKA

Bakgrund och syfte

Vesikoureteral reflux (VUR) är en medfödd missbildning i urinvägarna som innebär backflöde av urin från blåsan tillbaka till njuren. Reflux förekommer hos 1-2% av alla barn. Den graderas I-V där grad V är den allvarligaste med mest vidgning av urinledaren och mest påverkan på njuren. Barn med VUR har en ökad risk för att få urinvägsinfektioner och som följd av dessa, ärr på njuren (fokala njurskador). Vid höggradig reflux finns njurförändring (generella njurskador) i större delen av fallen redan vid födelsen. Hos en del barn försvinner refluxen spontant, men hos andra finns den kvar vilket gör att njurskadan kan förvärras av upprepade urinvägsinfektioner.

Det finns en klar ärftlighet vid reflux; risken att drabbas är 30-50% för syskon till barn med VUR och för avkommor är risken så hög som 66%. Trots det tydliga nedärvningsmönstret är den genetiska bakgrunden till reflux inte fullständigt kartlagd, de genetiska faktorerna är kända i endast en bråkdel av fallen. Syftet med denna avhandling, som bygger på fyra delstudier, var att allmänt undersöka ärftligheten vid VUR i ett patientmaterial från Västsverige och att undersöka vilka gener som orsakar VUR vid familjär reflux. Visionen är att i framtiden med genanalys från blodprov eller salivprov kunna avgöra sjukdomens svårighetsgrad och prognos, samt leta efter andra sjukdomsfall i familjen. På detta sätt kan upprepade röntgenundersökningar, som innebär urinkateter och strålning, undvikas.

Metod och resultat

Studie I utgår från alla 726 patienter som har behandlats för VUR vid Drottning Silvias barn och ungdomssjukhus mellan 1990 och 2004. Av dessa uppgav 99 personer (ur 66 familjer) att de har släktingar med reflux. Vid jämförelse av kliniska data mellan gruppen med familjär reflux och gruppen utan släktingar med reflux, fann vi ingen skillnad i sjukdomsbilden. Nedärvning av sjukdomen var vanligare från moderns sida.

I **studie II** undersöktes DNA från 54 av dessa 66 familjer avseende förändringar i 2 gener, *SLIT2* och *ROBO2*, som hos musfoster har funnits ha viktig roll i bildningen av urinledaren. Vi fann ett fåtal förändringar strax utanför den delen av arvsmassan som översätts till *ROBO2* protein, som vi bedömde att inte vara sjukdomsframkallande. Potentiellt sjukdomsorsakande varianter var även ovanliga i *SLIT2*, där vi bara fann en förändring som bidrar till modifiering av *SLIT2* proteinet. Den detekterades hos en sjuk individ,

men inte hos de övriga sjuka i familjen. Vi anser därför att förändringar i *SLIT2* och *ROBO2* generna är sällsynta orsaker till VUR hos människor.

Till studie III och IV rekryterades familjer från Västsverige med minst tre sjuka. I **studie III** undersöktes 14 storfamiljer för att hitta områden i arvsmassan innehållande möjliga sjukdomsgener, som nedärvts från en gemensam anfader. Vi har inte funnit något område som är gemensamt för alla. I subgrupper av familjer har vi däremot ringat in gemensamma områden med gener som är relevanta för urinvägsbildningen. De bör undersökas vidare som eventuella kandidatgener för VUR.

I **studie IV** undersöktes 13 storfamiljer för att identifiera möjliga sjukdomsframkallande förändringarna framför allt i gener med funktion i utvecklingen av urinvägarna. För detta användes helexomsekvensering, som analyserar alla generna i arvsmassan samtidigt. Vi har identifierat tre nya varianter (i generna *LAMC1*, *KIF26B* respektive *LIFR*) med potentiell roll i uppkomst av VUR hos tre familjer (23 % av alla undersökta familjerna).

Sammanfattningsvis skiljer sig inte familjär och icke-familjär reflux från varandra vad gäller sjukdomens svårighetsgrad. Våra studier bekräftar att VUR är en heterogen missbildning som kan bero på varianter i en rad olika gener. Studierna har i nuläget endast hittat genetiska orsaken till VUR i en bråkdel av fallen. Någon uppsättning av gener som skulle ge en förklaring i de flesta fallen existerar inte. Därför krävs det mer forskning innan en smidig metod som helexomsekvensering kan ersätta merparten av för patienterna besvärliga röntgenundersökningarna.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

- I. **Bartik Z I**, Nordenskjöld A, Sjöström S, Sixt R, Sillén U. Hereditary Vesicoureteral Reflux: A Study of 66 Families. *Open Journal of Pediatrics* 2015; 5: 304-313
- II. Zu S, **Bartik Z**, Zhao S, Sillen U, Nordenskjöld A. Mutations in the *ROBO2* and *SLIT2* genes are rare causes of familial vesico-ureteral reflux. *Pediatric Nephrology* 2009; 24: 1501-1508
- III. **Bartik Z I**, Sillén U, Östensson M, Fransson S, Djos A, Sjöberg R-M, Martinsson T. A genome-wide scan locating candidate regions of familial vesicoureteral reflux. *Manuscript*
- IV. **Bartik Z I**, Sillén U, Martinsson T, Djos A, Lindholm A, Fransson S. Whole exome sequencing identifies *KIF26B*, *LIFR* and *LAMC1* mutations in familial vesicoureteral reflux. *Manuscript*

Article II reprinted by permission from Springer Nature: Springer, Pediatric Nephrology, Shulu Zu, © 2009

CONTENT

ABBREVIATIONS	IV
DEFINITIONS IN SHORT	VI
1 INTRODUCTION	1
1.1 Heredity of VUR	3
1.2 Embryology of the urinary tract	4
1.3 Pathophysiology of VUR	6
1.4 Basic genetics	9
The human genome	9
Genetic variations	11
Epigenetics	12
Genome projects and databases	13
1.5 Genetics behind VUR	14
2 AIM	21
3 PATIENTS AND METHODS	23
3.1 Patients	24
3.2 Methods	27
Phenotyping the familial and non-familial VUR	27
Febrile urinary tract infections	27
Voiding cystourethrography	27
Renal scintigraphy	28
⁵¹ Cr-EDTA-clearance and other GFR estimates	29
DNA isolation	29
Polymerase Chain Reaction (PCR)	30
Sanger sequencing	32
Whole-exome sequencing (WES)	34
Data processing WES	36
SNP microarray	37
Data processing SNP microarray - the Colour Method	39

Data processing SNP microarray - Copy number detection	40
3.3 Statistics.....	42
3.4 Ethical Considerations.....	44
4 RESULTS.....	45
4.1 Familial aggregation of VUR	45
4.2 Phenotype of familial and non-familial VUR	46
4.3 Mutation screening of <i>ROBO2</i> gene.....	48
4.4 Mutation screening of <i>SLIT2</i> gene	48
4.5 Testing for IBD haplotype with the Colour method.....	49
4.6 Testing for disease variant excluding the common haplotypes.....	51
4.7 Copy-number variants in VUR families.....	53
4.8 Gene mutation screening using whole-exome sequencing.....	53
5 DISCUSSION	59
5.1 On heritability and phenotype	59
5.2 On shared chromosomal regions	60
5.3 On shared copy-number variants.....	61
5.4 On gene mutations.....	62
5.5 Methodological considerations.....	64
6 CONCLUSION.....	66
7 CLINICAL IMPLICATIONS AND FUTURE PERSPECTIVES.....	67
8 WEB RESOURCES	69
ACKNOWLEDGEMENTS	71
REFERENCES	73

ABBREVIATIONS

bp	base pair
CAKUT	congenital anomalies of the kidney and urinary tract
CNV	copy-number variation
⁵¹ Cr-EDTA	⁵¹ Chromium-ethylenediaminetetraacetic (edetic) acid
DMSA	dimercapto-succinic acid
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
ddNTP	dideoxynucleotide
dsDNA	double-stranded DNA
GFR	glomerular filtration rate
GWAS	genome-wide association study
IBD	identical by descent
IBS	identical by state
kb	kilobase pairs, thousand base pairs
lncRNA	long non-coding RNA
MAG-3	mercaptoacetyltriglycine
Mb	megabase pairs, million of base pairs
MET	mesenchymal-epithelial transition
MM	metanephric mesenchyme

mRNA	messenger ribonucleic acid
ND	nephric or Wolffian duct
NGS	next-generation sequencing
RNA	ribonucleic acid
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
UB	ureteric bud
UCSC	University of California, Santa Cruz
UVJ	ureterovesical junction
UTI	urinary tract infection
VCUG	voiding cystourethrography
VUR	vesicoureteral reflux
WES	whole-exome sequencing
WGS	whole-genome sequencing

DEFINITIONS IN SHORT

Allele	A variant form of a gene
Genotype	Genetic constitution of an individual / set of alleles present at one or more specific loci
Haplotype	A set of DNA variations, or polymorphisms on the same chromosome, which tend to be inherited together
Heterozygous	Two different alleles at a genetic locus
Homozygous	Two identical alleles at a particular locus
Hypodysplasia	Abnormally small and malformed organ
Phenotype	Physical traits expressed in an individual

1 INTRODUCTION

Primary vesicoureteral reflux (VUR) is a congenital urinary tract defect in which a dysfunctional vesicoureteral junction allows the retrograde flow of urine from the bladder to the kidneys. Secondary VUR occurs due to anatomical or functional bladder outflow obstruction, e.g. posterior urethral valves or neurogenic bladder. Galen, a Greek physician and surgeon in the second century, and Leonardo Da Vinci in the fifteenth century were the first ones to describe and illustrate the oblique entry of the ureter into the bladder that constitutes one of the main features of the anti-reflux mechanism [1].

VUR is diagnosed by radiological techniques with catheterisation as in voiding cystourethrography (VCUG) or by indirect radionuclide techniques with reduced radiation dose and avoidance of a bladder catheter for the price of somewhat lower sensitivity [2]. The presently used grading system (grade I-V) was introduced in the International Reflux Study [3] (Figure 1, Table 1).

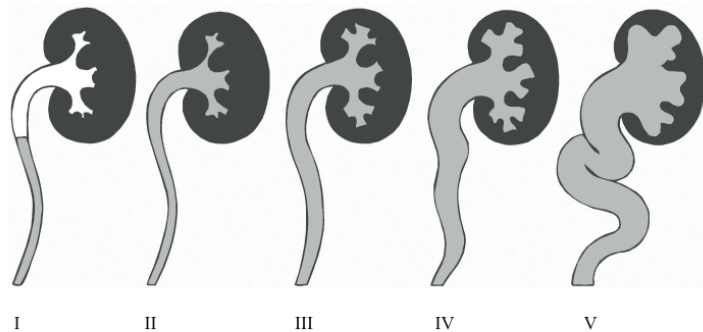


Figure 1. Reflux grading according to the International Reflux Study in Children, image courtesy of Sverker Hansson

Table 1. The definitions of the International Grades of vesicoureteral reflux

Grade	Definition
I	Appearance of contrast in the ureter only
II	Appearance of contrast in the ureter and pelvis, no dilatation, no blunting of fornices
III	Mild/moderate dilatation of the ureter and pelvis, no blunting of fornices
IV	Moderate dilatation of the ureter and pelvis, blunting of fornices but preserved papillary impressions
V	Severe calyceal dilatation with ureteral tortuosity, loss of papillary impressions

An often-quoted figure regarding VUR prevalence in children is 1-2% [4], which is only estimation because large-scale population screening using invasive diagnostic tests is unethical to perform. Between 1949 and 1981 a series of studies were published describing the prevalence of VUR in children without a predisposing condition [5]. In these studies the sample size varied from 24 to 722, age span was from zero to 74 years and wildly different diagnostic methods were used. The frequency of VUR ranged from 0% to 28,2% as a result of very different study populations. It is well known that the prevalence of reflux is inversely correlated with the age of the study population, as spontaneous resolution of reflux often occurs with growth [4].

Based on epidemiological data, on the incidence of urinary tract infections (UTI) and on the prevalence of VUR in children with UTI, the prevalence of VUR can be estimated in the general paediatric population. The cumulative incidence for UTI until the age of seven years was found to be 5% (8,4% for girls and 1,7% for boys) [6]. The incidence of first time UTI was highest during the first year of life and decreased markedly for boys beyond the age of one, although the reduction was also evident for girls [7]. When children with UTI underwent VCUG, 25% to 40% of them were diagnosed with VUR [5, 7]. This translates to VUR prevalence of ~ 1,6% in children (2,7% in girls and 0,6% in boys), which is probably a low estimate, as not necessarily all children with VUR develop UTI.

Although subjects with VUR may present similar initial symptoms, some cases have a benign natural course with no recurrent UTIs, no progressive renal damage and a high rate of spontaneous resolution of the reflux. Others, on the other hand, have recurrent UTIs, deterioration of renal status and persistent reflux [8]. These two phenotypes may have different aetiological and genetic backgrounds.

This thesis focuses on the heredity of primary vesicoureteral reflux and its genetic background in order to better understand the aetiology and to improve diagnostics and prediction of prognosis. Here follows further introduction to the topic with description of heredity, anatomy, embryology, pathophysiology, basic genetics and the genetics behind VUR.

1.1 HEREDITY OF VUR

Reflux has a proven hereditary character. Siblings of affected children are at higher risk of reflux than the general population, with reported prevalence between 27 and 51% [9-12]. Among multiple gestation births the concordance is higher in monozygotic twins than in dizygotic ones (80% vs. 35%) [13]. Moreover, the risk for offspring of parents with previously diagnosed VUR of having VUR themselves is reported to be 66% [14]. The differences in prevalence between studies on familial VUR can be due to differences in study populations and diagnostic methods used for detecting VUR.

This familial clustering of VUR implies that genetic factors play an important role in its pathogenesis. Defect in molecular regulation of embryological development of kidney and urinary tract can be inherited. A variety of inheritance pattern have been observed in families with VUR. Autosomal dominant inheritance with reduced penetrance is most widely accepted [15-19], although some authors point toward a possible recessive [20], X-linked [21] or complex polygenic model [22]. Figure 2 shows the inheritance pattern in two of the families with VUR participating in our studies.

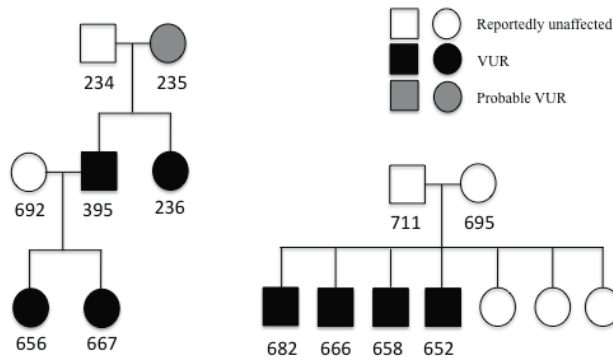


Figure 2. Pedigrees depicting two of the families enrolled in the study. **Squares** males, **circles** females, **black** symbols indicate diagnosis confirmed by voiding cystourethrography, **grey** symbols indicate strong history of VUR but no available radiological investigations

1.2 EMBRYOLOGY OF THE URINARY TRACT

The identification of animal models that mimic human urinary tract malformations has helped to elucidate the key morphological and molecular events underlying urinary tract morphogenesis. These can be grouped around three major developmental phases: 1) Wolffian (or nephric) duct formation, 2) ureteric budding and kidney induction and 3) distal ureter maturation. Each phase of renal development is dependent on the expression of different genes. So far, more than 400 genes have been identified as playing a role in renal development [23, 24].

Starting from the fourth week of gestation, three embryonic excretory organs develop from the intermediate mesoderm in a temporally and spatially distinct order from rostral to caudal end of the embryo: the pronephros, mesonephros and metanephros (Figure 3). The pronephros regresses rapidly without forming any nephrons in humans. The intermediate mesoderm also gives rise to the nephric duct (ND) on each side of the embryo, it extends caudally and connects to the cloaca. Mesonephric tubules are formed along the ND, which are primitive nephrons that empty individually into the ND. This primitive mesonephric kidney functions between the fourth and tenth week of gestation in humans. The final structure, the metanephros, forms the permanent adult kidney and it is derived from the ureteric bud (UB) and a specific region of the intermediate mesoderm called the metanephric mesenchyme (MM). The formation of the UB starts with a swelling on the ND, close to the junction with the cloaca, in the fifth gestational week. While this swelling enlarges, the UB emerges from the ND [25]. The precise location of a single budding, the angle and direction of this budding are crucial for normal development and future function. Too rostral UB leads to ectopic ureteric insertion within the neck of the bladder, causing distal ureteral obstruction with hydroureter. Too caudal budding positions the ureter too far laterally within the bladder wall, causing VUR [26].

When the UB elongates and invades the MM, UB starts branching at the tip as a response to signals from MM. On the other hand, signalling from the ureteric tip stimulates MM to form the kidney through mesenchymal-epithelial transition (MET). Thus, the UB gives rise to the collecting ducts, renal pelvis, ureter and ureterovesical junction, while the MM differentiates into distal and proximal tubules, loop of Henle, glomeruli and renal stroma. The development of nephrons continues until the 36 week of gestation, but their maturation proceeds during the first two years of life. [25].

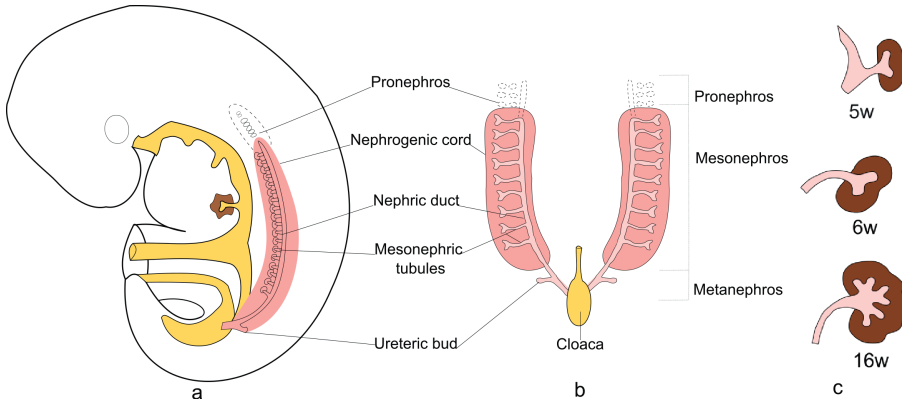
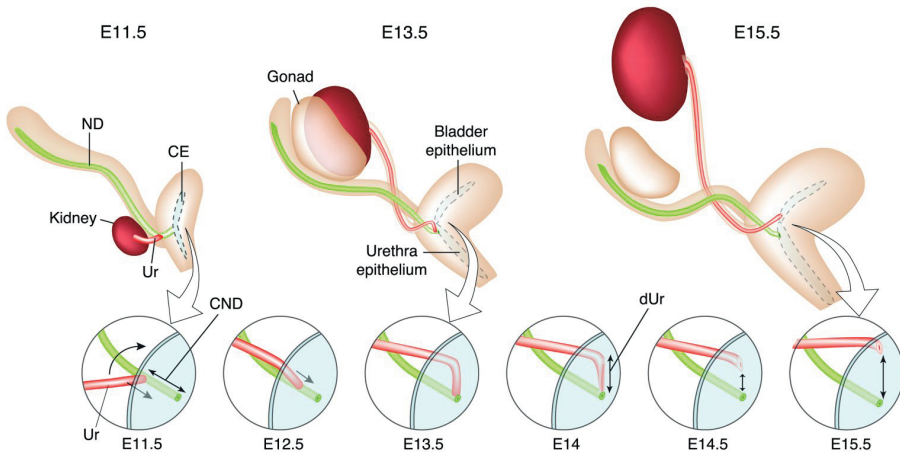


Figure 3. Embryonic kidney and urinary tract development during the 5th week. a, lateral view of the three sets of nephric systems; b, ventral view, the mesonephric tubules are drawn laterally for better picturing; c, successive development of the ureteric bud from fifth to sixteenth weeks, the metanephric mesenchyme (the caudal part of the nephrogenic cord) condensates around the ureteric bud to form the nephrons and connective tissue of the kidney. Illustration by Bernadett Pakucs



*Figure 4. Model of distal ureter rearrangement in mouse during embryonic day E11.5–E15.5. At E12.5 the CND undergoes apoptosis. At E13.5–E14, the distal ureter has completely laid down against the bladder and undergoes apoptosis, so the ureter separates from the nephric duct. Growth of the bladder allows further separation of the two orifices and at E15.5, the ND drains into the urethra. Bl, bladder; CE, cloaca epithelium; CND, common nephric duct; dUr, distal ureter; Ur, ureter. Republished with permission of American Society for Clinical Investigation, from Noriko Uetani et al., *J. Clin. Invest.* 119:924–935 (2009); permission conveyed through Copyright Clearance Center, Inc.*

Distal ureter maturation is the process by which the ureter is displaced from the nephric duct to its final position within the bladder wall (Figure 4). This process was the last to be morphologically characterised and identified as another crucial step for normal function [27]. Contrary to previous belief, the common nephric duct (CND), which is the distal ND between the cloaca and the nephric duct branching point towards the ureter, is not incorporated into the urogenital sinus to form the bladder trigone, but it undergoes apoptosis instead. The distal ureter descends, comes into contact with the urogenital sinus, which is the ventral part of the cloaca, and undergoes a 180 degrees rotation around the axis of the ND. Ureter rotation is crucial and provides an understanding of Weigert–Meyer law, which describes that the upper system of a duplex kidney is drained in the bladder inferiorly to the lower system. Following the rotation, the bifurcation and the common nephric duct lies down against the urogenital sinus and is eliminated by apoptosis. The ureteral orifice shifts rostrally, dorsally and laterally by the expansion of the surrounding urogenital sinus tissue during the development of the urinary bladder [28]. The remaining part of the ND forms the epididymis, ductus deferens and seminal vesicles.

Interference in the interaction between the UB and the MM can result in both renal parenchymal dysgenesis and urinary tract malformation. To emphasise this association, the term CAKUT (congenital anomalies of the kidney and urinary tract) was coined [29]. Invasion of the UB is necessary for the survival of the MM. Failure of UB outgrowth leads to apoptosis of the MM and consequent renal agenesis. Meanwhile, ectopic UB outgrowth or abnormal ureteric tree formation during the branching morphogenesis leads to dysplasia, hydroureter and duplex kidney. If one process is disrupted, it will disrupt the other as well [25]. Even the process of distal ureter maturation can be disturbed resulting in malformation of the urinary tract.

1.3 PATHOPHYSIOLOGY OF VUR

The ureteral orifices at the ureterovesical junctions (UVJ) are normally located in the lateral trigonal corners of the lower part of the bladder. UVJ is an important area that separates the low-pressure upper urinary tract from the variable pressure, urine storing lower urinary tract. The ureters enter the bladder at a sharp angle, run obliquely through the muscular layer of the bladder wall and end in a submucosal tunnel that forms a flap-valve (a passive anti-reflux mechanism, Figure 5). The competence of the valve is influenced by the diameter of the ureteral orifice, the intramural length and

course of the ureter. The length of the intravesical ureter increases markedly during the last trimester to about 3-4 mm in length at birth [30].

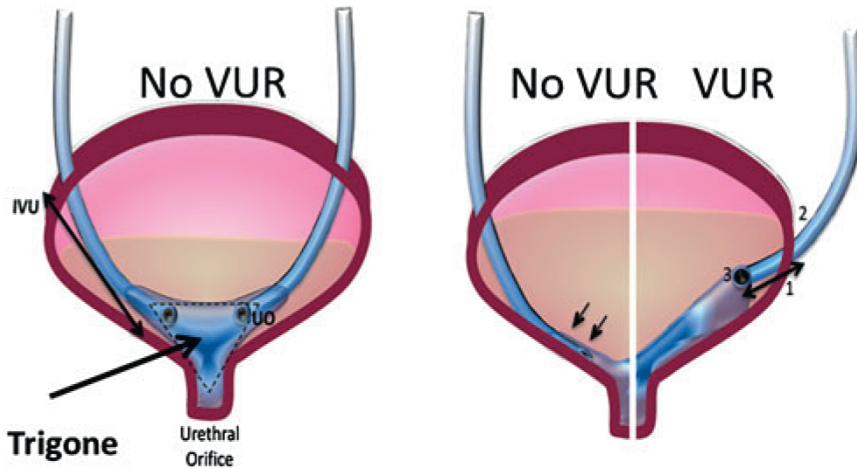


Figure 5. Illustration of normal UVJ (left picture and left panel of the picture on the right, No VUR) and abnormal UVJ (right panel, VUR) showing 1) short IVU, 2) ectopic ureter and 3) golf hole orifice. **IVU** intravesical ureter, **UO** ureteral orifice, **arrows** flap-valve (UO closed by the filled bladder). Reprinted by permission from Springer Nature: Springer, *Pediatr Nephrol, Vesicoureteric reflux and reflux nephropathy: from mouse models to childhood disease*, Fillion, © 2014

The muscular wall of the ureter consists of an inner longitudinal and an outer circular muscle layer in the upper two thirds, and an additional outer longitudinal muscle layer in the distal third before entering the bladder wall. These muscles squeeze the urine into the bladder by peristalsis. According to some authors, they also form a physiologic sphincter at UVJ, that contracts in response to vesical contraction and relaxes on external urethral sphincter contraction [31]. In fact, the longitudinal muscle fibres of the intramural ureter are surrounded by the detrusor muscle of the bladder forming a functional entity, and contractions in the UVJ are probably not isolated but rather connected to detrusor contractions or peristalsis in the ureter.

Primary VUR is the consequence of a congenital abnormality, or delayed maturation of the UVJ. In case of congenital anomaly, an ectopic ureter is sometimes seen, which appears as a "golf hole", with short intravesical ureter (Figure 5). Furthermore, the muscular wall of the distal end of refluxing ureter has been found to be degraded, disorganised, with deprivation of the intramural nerve supply [30]. VUR is detected most commonly during

voiding (VCUG), when intravesical pressure rises, but may occur any time in the filling/voiding cycle, particularly when bladder function is abnormal.

In many cases, VUR resolves spontaneously during childhood due to growth of the bladder and elongation of the submucosal tunnel, which lead to better functioning UVJ. Neonatal high-grade reflux should be regarded as a different entity with a much higher rate of spontaneous resolution during the infant year than the resolution rate for reflux found in older children [32]. This could be explained, in addition to bladder and ureter growth, by improvement of the inadequate sphincter relaxation and immature bladder dyscoordination seen during infancy, especially in boys.

The morbidity seen in children with VUR is often related to recurrent UTI, with a risk of progressive renal damage. Reflux of urine with bacterial contamination is a risk factor for pyelonephritis, which may cause focal renal damage, acquired reflux nephropathy. The reflux in itself, without bacterial contamination and with low pressure in the bladder, has not been documented as damaging. High-grade VUR in infants (often males) is frequently associated with congenital generalised renal damage, renal hypodysplasia. A proportion of them are diagnosed following prenatal hydronephrosis, prior to any pyelonephritis [33]. Thus, the respective pathophysiological mechanisms for focal and generalised renal damage are likely different. It is now well accepted that congenital hypodysplasia is a consequence of maldevelopment of the ureteric bud, which not only causes VUR but disturbs the kidney morphogenesis from MM as well.

The renal damage may in the long run lead to hypertension, pregnancy complications (UTI and hypertension) and renal insufficiency. Thanks to the prompt treatment of UTIs and close clinical supervision throughout childhood, the prognosis has improved over the years and most children with VUR do well [34-39]. The risk of developing end stage renal disease due to VUR is considered small in Scandinavian countries but is still a reality in other parts of the world [40].

Studies on familial VUR report a large number of asymptomatic cases. Among siblings with radiological reflux only 15% have a history of VUR [11]. This fact, and the possible natural course of spontaneous resolution during childhood, make reflux a difficult abnormality to study in terms of heredity from one generation to another and of genetic aetiology.

1.4 BASIC GENETICS

The human genome

The human genome, which contains all our genetic information, is built by double sets of 23 deoxyribonucleic acid (DNA) molecules compactly organized into chromosomes in the cell nuclei: 22 autosomal and one sex chromosome (X or Y). The DNA molecules consist of two polynucleotide chains in different combinations of the four nucleotide bases: adenine (A), cytosine (C), guanine (G) and thymine (T). The paired DNA strands are held together by complementary hydrogen bonds between A-T and C-G base pairs (bp) and form a double helix (Figure 6). This model was first described by James Watson and Francis Crick in 1953 based upon the X-ray crystallography images of DNA by Rosalind Franklin and Maurice Wilkins [41]. In 1962 Watson, Crick and Wilkins received the Nobel Prize in Physiology or Medicine for their discovery that formed the basis for molecular genetics and modern biotechnology.

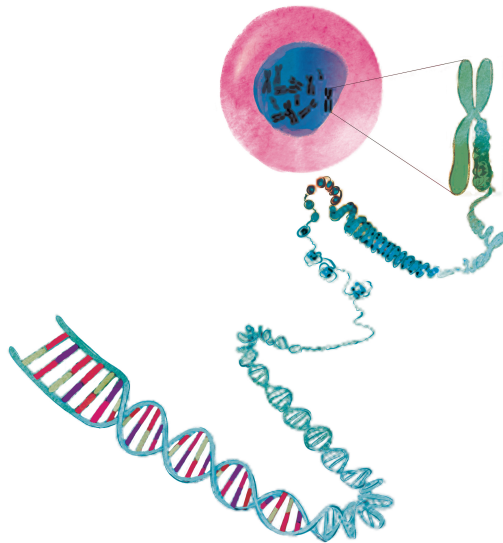


Figure 6. The organisation of DNA within the chromosome. The human genome contains approximately 3 billion base pairs, which reside in the 22 pairs of chromosomes and 2 sex chromosomes within the nucleus of all our cells. The double helix of the paired DNA strands is coiled around histone octamers, which is compacted into chromatin fibre that is condensed into long loops. Further compaction into transcriptionally inactive heterochromatin occurs during cell division. This highly condensed form of DNA can be visualised as chromosomes in light microscope. Illustration by Eszter Kónya

In 1958, Crick defined the function of the DNA in the "Central Dogma" of molecular biology [42]. The genetic information flows from the DNA to ribonucleic acid (RNA) through transcription, which is then translated into protein (Figure 7). Later it was discovered that there are even reverse transcriptases that can generate DNA from RNA template, used mainly by viruses. The bases in the messenger RNA (mRNA, the DNA transcript) are arranged in groups of three, called codon, which represent a single amino acid. Each of the 20 amino acids that occur in human proteins is represented by at least one codon and there are three stop codons that terminate translation. To become a fully functional protein, the amino acid chain is folded into a three-dimensional structure and often undergoes post-translational modification.

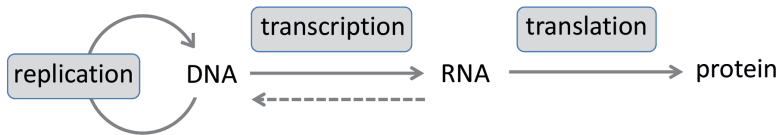


Figure 7. The Central Dogma of molecular biology

There are about 19,900 protein-coding genes in the human genome with the coding parts corresponding to only 1-2 % of the entire genome (<https://www.encodegenes.org/stats/archive.html#a28>). A typical gene consists of several exons (coding sequences) interrupted by introns (non-coding sequences). The first and last exons commonly also contain untranslated regions (UTRs), which are important for mRNA stability and translation, and the last exon ends with a polyadenylation site (AATAA), important for cleavage of RNA. Both exons and introns are initially transcribed into a pre-mRNA, which then undergoes splicing to remove the intronic sequences but also other modifications in order to form the final mature mRNA. Transcription is initiated from the promoter region, located upstream of the gene. The promoter consists of DNA elements that bind RNA polymerase and transcription factors, which are required for transcription initiation. However, there can be additional regulatory elements also affecting transcription, such as enhancers and silencers, which can be located at a large distance from the gene of interest.

Currently, a gene is defined, according to the Guidelines for Human Gene Nomenclature, as "a DNA segment that contributes to phenotype/function. In the absence of demonstrated function, a gene may be characterized by sequence, transcription or homology" [43]. Besides protein coding genes, the human genome contains a large number of non-coding RNAs, pseudogenes

and intergenic regions with (at the moment) unknown function. Furthermore, alternative splicing and post-translational modification contribute to the complexity and diversity of the RNA and protein products.

Genetic variations

Identifying DNA variants that contribute to disease is a central aim in human genetics. At the vast majority of approximately 3 billion genomic sites, each human carries the same base residue on both chromosomal homologs. The remaining positions account for the diversity among humans. Most of these differences occur naturally in the populations, so-called polymorphisms, and are not disease causing. Single-nucleotide polymorphism (SNP) is a variation in the DNA sequences in which one nucleotide differs between individuals. There are roughly 10 million SNPs in the human genome [44]. Usually a SNP is biallelic, meaning that two different nucleotide residues could be seen at the same genomic position in a population, although three and four-allelic SNPs also exist. An allele is one of several possible forms of a DNA sequence for a specific locus. An individual possesses two alleles at each locus of the 22 autosomal chromosomes and can be heterozygous, which means having two different allele variants, or homozygous with two alleles of the same variant. Although most SNPs are non-pathogenic, they may help to predict an individual's response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases. SNPs can also serve as genetic markers for identifying disease genes by linkage studies in families, linkage disequilibrium in isolated populations, association analysis of patients and controls, and in loss of heterozygosity studies in tumours [45].

Any new change in the DNA could be referred to as a mutation, including also a base pair exchange in a non-coding region. However, the term "mutation" is more commonly used for pathogenic alterations and "polymorphism" for naturally occurring, not disease causing DNA sequence variants. Mutations (disease-causing DNA variants) are often found in the coding region of the gene but may also be found in regulatory elements of the gene. A nucleotide exchange in the coding region that results in no change of amino acid is called synonymous variant. The opposite is the non-synonymous variant that can either be missense, resulting in a different amino acid, or nonsense, resulting in a premature stop codon. Mutations can, beside nucleotide substitution, also result from deletions (where one or more bases are lost) or insertions (where one or more bases are inserted). Deletion or insertion will cause a shift in the reading frame if the number of involved nucleotides is not a multiple of three, and will most likely result in a

truncated protein. Mutations could also affect splicing by either creating or destroying splice site signals.

There is also a form of large-scale polymorphism that involves DNA copy number variation (CNV) caused by deletion or duplication of DNA sequence longer than 1 kb (1000 bp). The chromosomal segment displaying CNV could contain a single gene or a set of genes and thus affect expression dosage of affected genes. Similarly to SNPs, CNVs can be found in the normal population and are important contributors to human genetic variation [46]. However, studies have shown that some CNVs can be associated with a variety of birth defects, common diseases or susceptibility to diseases [44, 47].

Epigenetics

Heritable, but reversible, changes in gene expression can occur without alterations in DNA sequence. DNA packing is one way to regulate gene expression. The double helix is wrapped around an octamer of histones approximately two turns to form units called nucleosomes (Figure 6). A nucleosome contains about 200 bp DNA sequence. Depending on chemical modification of the histone-tails, a genomic segment can be more or less accessible for binding of DNA- or chromatin binding proteins as well as transcription factors, which in turn form a complex with RNA polymerase, which is doing the actual transcription. An active gene is associated with acetylation of the histones. On the other hand, histone deacetylation makes the DNA strand more condensed and thereby less active. Methylation of histones can either increase or decrease transcription of genes, depending on which amino acids in the histone tails are methylated, how many methyl groups are attached and the presence of acetyl groups in the vicinity. Other posttranslational modifications of the histones are phosphorylation, glycosylation or ubiquitination. More recently discovered modifications influencing nucleosome dynamics are succinylation and malonylation [48]. The combination of all of these modifications constitute a “histone code”, which defines the status of the chromatin structure [49].

Additional epigenetic mechanisms are DNA methylation causing gene silencing (like seen in genomic imprinting or X-chromosome inactivation) or small non-coding RNA, which binds mRNA and impedes translation. The epigenetic mechanisms play a major role during embryonic development. Moreover, epigenetics is implicated in pathogenesis of complicated disorders in human (cancer, autoimmune disorders, memory, addiction,

neurodegenerative and psychological disorders) and in individuals' response to environment changes such as nutrition, stress, toxicity, exercise and drugs.

Genome projects and databases

Large-scale DNA sequencing efforts of many public and private organisations, including the Human Genome Project enables genetic research. The Human Genome Project (even called HUGO-project) was an international research effort; collaboration of the National Institute of Health (NIH), U.S. Department of Energy, numerous universities in the United States, the United Kingdom, France, Germany, Japan and China. The project was launched in 1990 and completed in April 2003 [50].

Most of the research results in this field are freely available on Internet. Powerful computer programs have been designed to permit searching of DNA and protein sequences in databases for matching the sequence under investigation. GeneBank (<https://www.ncbi.nlm.nih.gov/genbank>) and European Molecular Biology Laboratory (EMBL) are frequently used nucleotide sequence databases. The "basic local alignment sequence tool" (BLAST) is one of the most useful algorithms for sequence searching [51]. This program is available through different platforms, such as the National Center for Biotechnology Information (NCBI: <https://www.ncbi.nlm.nih.gov>) and the European Bioinformatics Institute (EBI: <https://www.ebi.ac.uk>). Ensembl (<https://www.ensembl.org>), a cooperative project between EMBL - EBI and the Sanger Institute, also provides up-to-date information on the human genome. The University of California, Santa Cruz human genome browser (UCSC: <https://genome.ucsc.edu>) is yet another updated site providing an enormous amount of useful information with good search functions. It offers an even newer, faster and more accurate sequence searching algorithm called the "BLAST-like alignment tool" (BLAT) [52].

The Database of Genomic Variants (<http://dgv.tcag.ca/dgv/app/home>) and the Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (<https://decipher.sanger.ac.uk/>) are public efforts aiming to comprehensively catalogue all human CNVs (and other forms of structural variation) similar to the governmental project dbSNP for SNPs (<http://www.ncbi.nlm.nih.gov/projects/SNP>). The SweGen dataset provides data on SNP frequency in a Swedish cohort (<https://swegen-exac.nbis.se>).

In this thesis we have frequently used the UCSC genome browser for DNA alignment, BLAT searches and in-silico PCR.

1.5 GENETICS BEHIND VUR

Various approaches have been used to identify candidate genes that explain VUR susceptibility. These include genetic analyses of patients with syndromic VUR; studies of animal models; gene expression studies; association-, linkage- and sequencing studies of candidate genes; genetic mapping studies in families with multiple affected individuals; genome-wide association studies; and the latest technological advance, whole-exome or whole-genome sequencing.

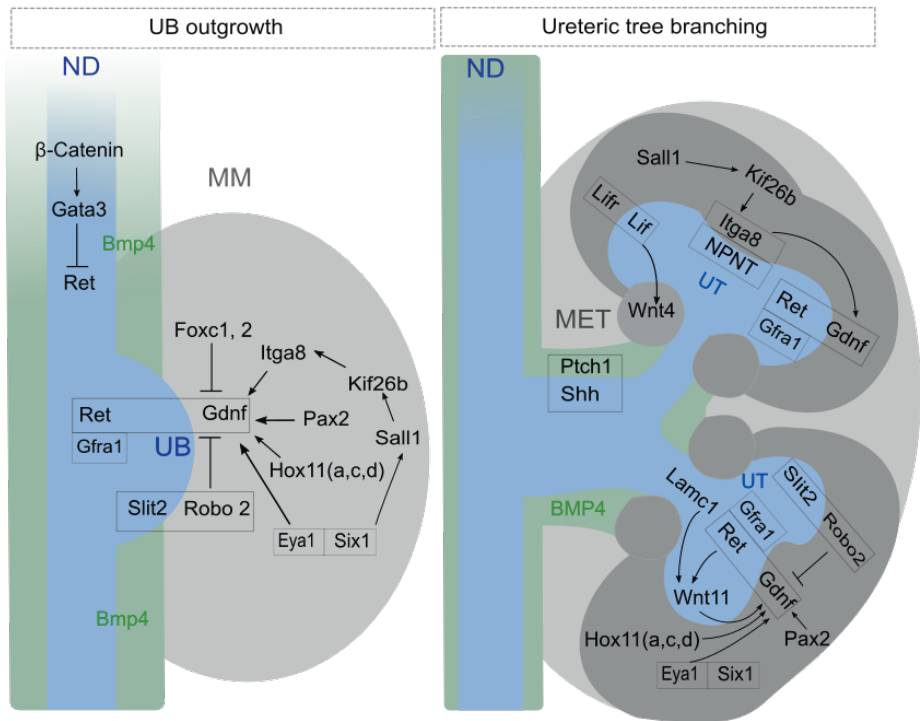


Figure 8. Genes regulating ureteric bud outgrowth and branching morphogenesis. The illustrations show a selection of genes involved in the ureteric bud initiation, outgrowth (left) and branching (right). The MM condenses around the branching UB and forms nephrons through MET at the "armpit region" of the ureteric tree. MET, mesenchymal-epithelial transition; MM, metanephric mesenchyme; ND, nephric duct; UB, ureteric bud; UT, ureteric tree. Illustration by Bernadett Pakucs

Key players in urinary tract development

Use of human homolog gene knock-out mouse models has identified a series of genes crucial for the urinary tract development. The *in vivo* modelling

shows that the developmental process is highly complex, where GDNF-RET/GFR α 1 as well as Wnt-signalling is essential for proper development (Figure 8). A number of genes associated to these pathways have been investigated through mouse models in the context of VUR: *Gdnf*, *Ret*, *Gfra1*, *Robo2*, *Slit2*, *Upk2*, *Upk3*, *Agr2*, β -catenin, *Bmp4*, *Fgfr2*, *Foxc1*, *Foxc2*, *Gdf11*, *Hox11(a,c,d)*, *Kif26b*, *Osr1*, *Shh*, *Spry1* and *Wnt11* [19, 25, 29, 53-59]. The central role of the GDNF-RET/GFR α 1 pathway is supported by the fact that the majority of the genes implicated in the UB formation are regulators of *Gdnf* or *Ret* expression. The GDNF ligand, secreted from the MM, signals to the receptor RET and coreceptor GFR α 1 complex expressed in the ND [60]. Proliferation of the responding cells facilitates the migration of the UB towards the MM. *Bmp4* is expressed in the mesenchyme along the ND and around the outgrowing UB to restrict the site of UB outgrowth to one location and inhibit premature branching before reaching the target tissue, the MM [61, 62]. FOXC1/FOXC2 are additional transcription regulators blocking GDNF [63]. Yet another signalling complex that restricts the domain of GDNF expression is by the secreted factor SLIT2 and its receptor ROBO2 [19, 53]. *Slit2* is expressed mainly along the ND and *Robo2* along the nephrogenic mesenchyme as well as at a lower level in the ND, at the budding site and anterior to it. The ROBO2/SLIT2 complex functions as chemorepellent that causes migrating cells to turn away from it. These negative regulators of GDNF/RET pathway are required for the outgrowth of a single UB at the correct position. Inactivation of either *Slit2* or *Robo2* leads to supernumerary UB development and abnormal maintenance of *Gdnf* expression in anterior nephrogenic mesenchyme [53].

Kif26b is essential for the adhesion of mesenchymal cells surrounding ureteric bud as it interacts with the cytoskeleton [56]. *Kif26b* acts downstream of *Sall1* but upstream of *Gdnf*. In *Kif26b*-deficient mice the ureteric buds were attracted close to the mesenchyme, but failed to invade and branch into the mesenchyme [64]. These mutant mice even showed impaired integrin α 8 (*Itga8*) expression, which led to failure of *Gdnf* maintenance. The *Kif26b*-null mice died shortly after birth due to bilateral kidney agenesis or unilateral agenesis in combination with hypoplasia of the other kidney [56]. Subsequent to UB invasion of MM, signals from MM start off branching of the UB tip and signalling from the UB tip stimulates MM to form the kidney through MET. Just as GDNF/RET signalling is central to UB outgrowth and branching, Wnt signalling is key to induction of MET [65]. Numerous additional genes are involved in MET and segmentation of nephrons [25] and mutation of these genes can cause renal dysplasia without VUR [66].

Urinary tract malformations have also been associated with mutations in the uroplakin family of proteins. Uroplakins are integral membrane proteins at the luminal surface of the urothelium, which lines the renal pelvis, the ureter and the bladder. Their functions are to strengthen the urothelium during filling, prevent bacterial adherence and contribute to the permeability barrier. In knock-out mouse models, the UPK3-depleted urothelium in the bladder is thick, is leaking and the ureteral orifice resembles a large golf hole [54]. Similar changes in the urothelium, as well as hydronephrosis due to either VUR or distal ureteral obstruction, were seen in homozygous *Upk2* knockout mice [55].

Nevertheless, the entire repertoire of genes involved in the development of kidney and urinary tract is still unknown. Multiple studies have been performed in order to analyze expression patterns at different stages of animal development or in organ cultures as well as the effect of the absence of the respective protein on the organ development [67]. These studies indicate that an intricate orchestrating in time and place of multiple genes is essential for proper embryonic development of the UB and MM. The experimental models also suggest that a mutation affecting a single gene may result in different phenotypes and mutations of different genes can result in the same disease [1].

Syndromes associated with VUR

VUR is a feature of numerous complex syndromes such as papillorenal syndrome, branchio-oto-renal syndrome, hyperparathyroidism-deafness-renal dysplasia syndrome and Townes-Brocks syndrome. Genetic studies of these syndromes have revealed several candidate genes that are active during renal and urinary tract development. *PAX2* mutations represent one of the main genetic abnormalities of renal coloboma syndrome, also known as papillorenal syndrome, which is associated with optic nerve abnormalities, hypodysplastic kidneys and in 26% of the cases, VUR [68, 69]. *Pax2* is expressed in the MM and plays a role in initiation and maintenance of *Gdnf*/*Ret* signalling [70]. *EYA1*, *SIX1* and *SIX5* mutations are associated with branchio-oto-renal syndrome, which is characterised by malformations of the outer, middle and inner ear, branchial fistulae and cysts, as well as renal malformations, e.g. renal hypodysplasia and VUR [71-73]. Mutations in *Six1* and *Six5* genes affect the interaction SIX1-EYA1 and SIX5-EYA1, respectively. The SIX1-EYA1 complex binds the *Gdnf* promoter as a transcriptional regulator [74]. Gene ablation of *Eya1* or *Six1*, normally expressed by the MM, leads to loss of *Gdnf*, failure of UB outgrowth and apoptosis of MM [75]. The urinary manifestation of the hyperparathyroidism-

deafness-renal dysplasia syndrome includes VUR, renal hypoplasia or aplasia. This disorder is caused by mutation of *GATA3* [76]. *Gata3* is expressed in the ND; it prevents ectopic and supernumerary ureter budding by transcriptional regulation of *Ret* expression [77]. The Townes-Brocks syndrome is a triad of imperforate anus, dysplastic ears and thumb malformations, but renal impairment is also frequent, including VUR. Mutations of the *SALL1* gene cause this syndrome [78]. *Sall1* acts downstream of *Six1* in the MM, and is essential for the invasion of the UB into the MM [79-81]. Although strongly indicated through the syndrome aetiology, mutations in these known syndromic genes are rarely seen in nonsyndromic VUR.

Sequencing studies of candidate genes

Sequencing the coding and regulatory regions of specific genes, implicated through VUR-associated syndromes and other mouse models to participate in renal and urinary tract development, have been performed on patient material in multiple studies. These studies detected mutations in only a small proportion of all cases, which mainly were patients with CAKUT: *PAX2* [82], *UPK3A* [83, 84], *UPK2* [85], *ROBO2* [19, 86, 87], *RET* and *GDNF* [88], *SIX2* and *BMP4* [89], *SOX17* [90] and *TNXB* [87, 91].

Linkage analysis

Rather than looking at specific, known genes, the whole genome can be scanned using genetic markers in order to identify specific chromosomal region(s) associated with VUR – regions that potentially harbour the genes that participate in VUR pathogenesis.

Linkage analysis is performed in families with several affected individuals in order to find chromosomal segments that are co-inherited with the trait under investigation [92]. Previously, panels of microsatellite markers (short-tandem repeats, a sequence of two to four nucleotides that is repeated) were used for linkage analysis. Nowadays it is more common to use SNP markers to map the genome, as SNPs are much more abundant than microsatellite markers and use of SNP-microarrays allows rapid, inexpensive and comprehensive genotyping [93]. The mode of inheritance has to be specified prior to parametric linkage analysis, whereas non-parametric linkage analysis looks for allele or chromosome segments that are shared only by affected relatives without knowing the mode of inheritance. The result of an analysis is given in a LOD score (logarithm to the base 10 of odds) comparing the likelihood of obtaining the data if the chromosome region is linked to the disease to the likelihood of observing the data by chance. Traditionally LOD score ≥ 3 is

regarded as a significant evidence of linkage and LOD score < -2 excludes linkage. The results in-between these values are inconclusive and should be validated by additional studies. For genome-wide significance level of 5% the LOD score threshold used should rather be raised to 3.3, to take into account the problem of multiple testing (see Section 3.3) [94].

The results of the published linkage studies highlight different chromosomal regions in familial nonsyndromic VUR and only a few findings have been confirmed by subsequent studies. These studies either include a large number of small families with ≥ 2 affected members, often siblings [66, 95-98] or a small number of large families with many affected cases [16, 18, 99, 100]. The majority of the studies with positive findings show suggestive linkage, however, Briggs et al. demonstrate significant linkage to chromosomes 5q14.2, 13q33.3 and 18q21.1 by nonparametric linkage analysis of 150 affected sib-pairs from 98 families [96]. According to the published LOD scores, suggestive linkage have been observed to loci on chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16, 19, 20, 21, 22 and X [16, 20, 95, 97, 100]*. Among these findings special attention should be given to the following regions as they received support from different linkage studies: 3p12.3–3q21.2 (containing *ROBO2* at 3q12.3) [95, 100], 3q26.31 [95, 97], 6q26 [95, 97], 10q25.3–10q26.13 (containing *FGFR2*, *EMX2*, *GFRa1* and close to *PAX2*) [95, 97], 13q33.3 [16, 95, 96] and 22q11.22–22q12.1 [16, 100].

These findings support the hypothesis that VUR is highly heterogeneous. The significant LOD scores at loci where there are VUR genes can be cancelled by negative scores due to lack of allele sharing at those loci in families with different mutations. Alleles can be shared just by chance at loci where no VUR genes lie. Thus significant linkage can be missed and false linkage peaks can be created [66].

Association studies

Linkage analysis has successfully mapped many monogenic diseases and a few disorders with locus heterogeneity, whereas it does not perform as well in complex diseases where the effects are usually too small to be detected by co-segregation within the families [92]. Therefore, focus has lately shifted towards association studies, which compare the frequencies of particular

* Studies showing suggestive linkage to loci on respective chromosome: 1 [95, 100], 2 [16, 95, 97], 3 [16, 95, 97, 100], 4 [100], 5 [97], 6 [95, 97], 7 [95], 8 [16], 9 [16], 10 [95, 97], 11 [97], 12 [20], 13 [16, 95], 16 [95], 19 [97], 20 [16, 95], 21 [97], 22 [16, 100] and X [16]

alleles between affected and unaffected individuals in case-control setup. Association studies can also be performed within families with at least one affected child by analysing case-parent trios [101]. Using allele-sharing methods, risk genes are identified by searching for loci where heterozygous parents overtransmit one of the two alleles [92].

Cordell et al. examined 262,264 SNPs in 172 VUR cases and 2938 controls, and found association to single markers on chromosome 10, 11 and 18. In family-based analysis of case-parent trios from 320 families, three loci (on chromosome 3, 10 and 11) were associated with VUR in the whole material. Association was shown to SNP on chromosome 18 in a subset of these cases and to chromosome 5 and a different marker on chromosome 18 in another subpopulation [97]. Darlow et al. have analysed 582,923 SNPs in 500 VUR cases and 851 controls and found no association that reached genome-wide significance level but suggestive significance for association to 3 adjacent markers on 5p15.2 as well as to single markers on chromosome 1, 3, 4, 8 and 13. Additionally, a few loci approached genome-wide significance level in their family-based association analyses of 643,691 SNPs in 410 affected child-parent trios (on chromosome 3, 7 and 17). The most interesting was the association to 23 adjacent SNPs on 3q25.32 [66]. However, there is no concordance in the results from these two genome-wide association studies (GWAS).

There are many genes, all over the genome, known to be involved in the development of the urinary tract. If we randomly choose a point anywhere on the genome, we would come close to one of these genes by chance. Therefore, we can never be sure whether these numerous, scattered, small linkage peaks or weak associations are real unless we identify the pathogenic mutations, which are not necessarily located in the coding regions [66].

Next-generation sequencing

In recent years, next-generation sequencing (NGS) has revolutionised genomic research. NGS provides rapid detection of DNA variants and an opportunity to arrive at a molecular diagnosis with a single test. The NGS studies performed in this field so far, investigated cohorts with different CAKUT diagnoses where only a fraction of the study subjects had VUR. In line with previous approaches, a variety of candidate genes were detected without much overlap between the studies of different populations.

Hwang et al examined an international cohort of 749 cases with CAKUT from 650 unrelated families, of which 288 individuals had VUR (68 familial

and 220 sporadic VUR). Mutations were found in 12 of 17 analysed CAKUT genes in 6% of the families in total: *BMP7*, *CHC5L*, *CHD1L*, *EYA1*, *GATA3*, *HNF1B*, *PAX2*, *RET*, *ROBO2*, *SALL1*, *SIX2* and *SIX5* [102]. In another well-done study of 453 CAKUT cases (of which 53 patients had VUR) 208 candidate genes were analysed. They prioritized 148 variants in 82 genes in 151 patients, however, from a burden test using 498 controls, none of the genes reached exome-wide significance [103]. Other studies have detected mutations in *SLIT2* and *SRGAP1* [104], *TBC1D1* [105], *LIFR* [106], *EYA1*, *HNF1B*, *PAX2* and *FOXPI* [107], *ANOS1*, *EYA1*, *CHD7*, *GATA3*, *HNF1B*, *KIF14*, *PAX2*, *PBX* and *SIX1* [24], *EYA1*, *DSTYK*, *HNF1B*, *RET*, *SIX5*, *SALL1* and *WNT4* [108] and *GREB1L* [109, 110]. Although rare mutations in multiple genes have been shown to cause CAKUT, causative genomic variants remain unknown for the majority of all CAKUT cases.

Copy-number variation (CNV)

Genomic copy-number variation (microdeletions and microduplications) was demonstrated to be an important pathogenic mechanism of a variety of common diseases [111]. In studies on CAKUT array-based comparative genomic hybridisation (array-CGH) [112, 113] or SNP microarrays [114-116] were used for genotyping, but CNVs can also be detected by whole-genome sequencing [117]. Sanna-Cherchi et al. detected pathogenic CNVs in 16.6% of individuals with renal hypodysplasia, of which the most common was at chromosome 17q12, (a region containing *HNF1B*, a gene involved in the development of epithelia in various organs as well as ureteric bud branching, initiation of nephrogenesis and nephron segmentation) [114, 118]. Caruana et al. analysed 178 cases with CAKUT, of whom 29 had VUR. Four VUR cases (14%) had either a pathogenic genomic disorder or a CNV of unknown significance [116]. Siomou et al. investigated seven children with CAKUT from three families, of whom six cases had VUR and no pathogenic CNVs. The seventh, a boy with ureterovesical junction obstruction (a second cousin of two boys with VUR) had a deletion on 17q12, which included *HNF1B* [113]. Although the results from these initial investigations are interesting, further studies on larger cohorts are needed.

2 AIM

The overall aim of this thesis was to improve the diagnostics, risk assessment and treatment of VUR through better understanding the genetic background of this congenital malformation. The ultimate goal for the future is that genetic analyses of blood sample or buccal smear may replace VUCG as the screening method for relatives of VUR patients. Furthermore, these analyses will hopefully identify patients at risk, by distinguishing severe cases that require prompt treatment and frequent follow-up from those with benign course that will resolve spontaneously.

Specific aims of the four studies:

Study I

- To study the heritability of VUR with epidemiological methods in a cohort of VUR patients from western Sweden.
- To establish whether the inherited (familial) form of VUR represents the same disease as the sporadic cases (when only one individual in the family has VUR) or whether it has a more aggressive course.

Study II

- To perform mutational analysis of two candidate genes for VUR, *ROBO2* and *SLIT2*, in nonsyndromic familial VUR.

Study III

- To search for common chromosomal areas (haplotypes), in high-density SNP arrays, in families with inherited VUR, in order to identify a unique haplotype associated with the disease.
- To evaluate shared haplotypes, even in subsets of families, for genes, coding and non-coding, of interest for the VUR abnormality.
- To investigate CNVs for association with VUR and hypodysplasia

Study IV

- To identify potential disease-causing gene mutations in familial primary nonsyndromic VUR using whole-exome sequencing.
- To evaluate whether one candidate gene causes the disease in all or some of the families or, if this is not the case, do the members of a family share the same candidate gene variant or they don't.

3 PATIENTS AND METHODS

Table 2. Patients and methods in the four studies at a glance

	Patients	Methods
I	66 families with VUR (66 index cases and 55 relatives) and 358 controls (sporadic, non-familial VUR cases) treated at Queen Silvia Children's Hospital between 1990 – 2004	Construction of pedigrees, review of medical records for clinical data and investigation results For comparisons between groups the Chi-square test, Kruskal-Wallis test, Fisher's exact test, Mantel-Haenszel chi-square test and Mann-Whitney U-test were used.
II	52 unrelated patents with familial VUR treated at Queen Silvia Children's Hospital between 1990 – 2004 2 patients with familial VUR treated at Astrid Lindgren Children's Hospital 96 controls, healthy voluntary blood donors	DNA extracted from blood Sanger sequencing of all exons of <i>ROBO2</i> and <i>SLIT2</i> genes Alignment to the reference sequence with bioinformatics software, evaluation of the identified gene variants using prediction programs
III	14 families with ≥ 3 affected members with primary VUR from the south-western part of Sweden; all cases confirmed by VCUG 4 in-house negative controls	DNA extracted from blood or buccal cells All affected individuals were genotyped with Affymetrix 250K SNP NspI arrays. Data analysed by the Colour method to identify shared haplotypes between VUR families, possibly harbouring the disease gene Bioinformatics software was used for copy number detection. Review of databases for common variants and visualisation of results using UCSC genome browser
IV	13 families with ≥ 3 affected members with primary VUR from the south-western part of Sweden; all cases confirmed by VCUG	DNA extracted from blood for whole-exome sequencing, and from blood or buccal cells for Sanger sequencing The latter was performed to verify WES results with Segregation analysis Extensive literature search for previous studies on genes involved in kidney and urinary tract development, and review of databases for common gene variants

3.1 PATIENTS

Between 1990 and 2004, 726 children were treated for VUR at Queen Silvia Children's Hospital, Gothenburg, Sweden. The hospital is a tertiary referral centre from a region with a population of 1.8 million people. Figure 9 illustrates the data selection process for **study I** and **II**. Letters were sent to all but 13 index cases, inquiring as to whether there were other members of the family or close relatives with VUR. The 99 cases that reported more than one individual with VUR belonged to 66 families. These families were interviewed by telephone enquiring as to which members of the family had diagnosed VUR or had symptoms indicating such a problem, i.e. recurrent UTIs, bladder function symptoms or kidney problems. We also asked about possible consanguinity in the family. Familial VUR cases were defined as patients with one or more first, second or third degree relatives with VUR. For inclusion of family members in the group of relatives with VUR, a previously performed voiding cystourethrography (VCUG) showing reflux was mandatory, except for older relatives diagnosed in an era when VCUG was not in general use. A history of recurrent febrile UTIs during childhood with or without renal damage suggesting high probability for VUR was accepted in these cases. Patients with VUR secondary to urethral valves, myelomeningocele or high anal atresia with neurogenic bladder were excluded from the study.

The 66 interviewed families were invited to participate in our study and they received oral and written information. Before entering the study, all subjects and/or their parents signed informed consent for genetic screening and review of medical records regarding their VUR and kidney status. In **study I** we compared the index cases with the group of relatives with VUR and the non-familial cases. The included families were also asked to provide blood samples for a genetic study. For **study II** 52 families donated blood samples. The remaining 14 families who accepted to participate could not provide blood samples for different reasons. Additional two large families with hereditary VUR were included from Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm. These two families were treated and followed up by our co-researchers. A total of 54 unrelated patients were analysed and compared with a control group consisting of anonymous DNA samples of 96 healthy voluntary blood donors at Karolinska University Hospital.

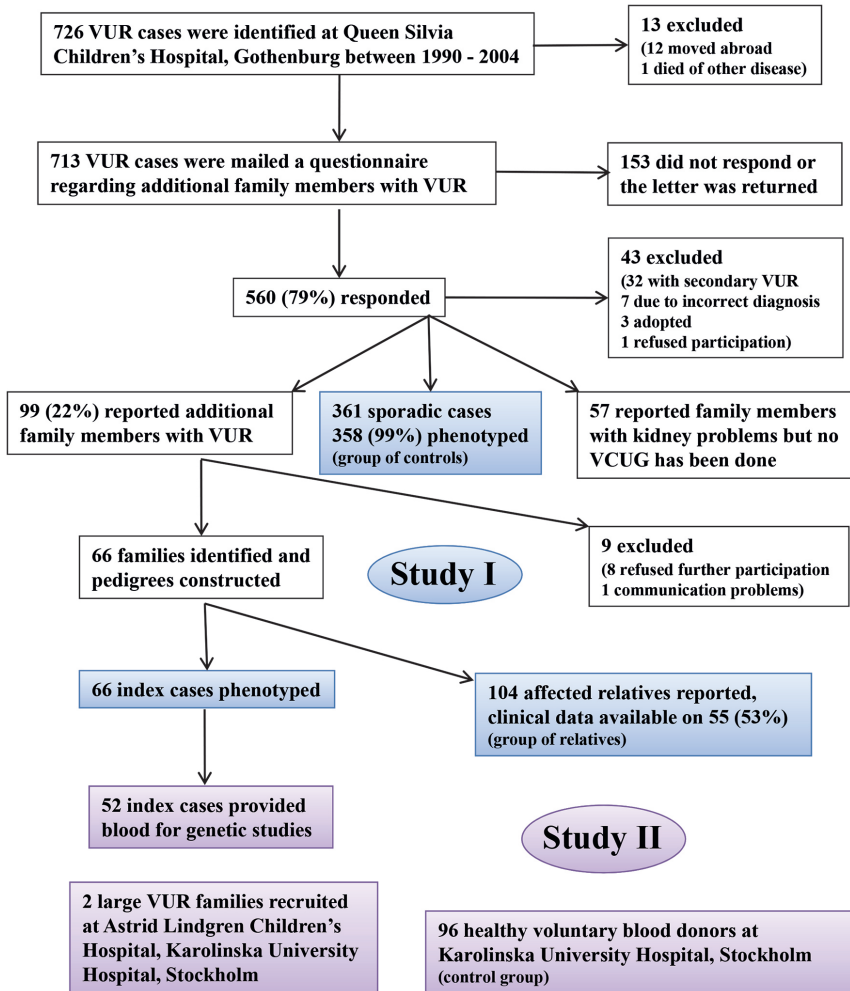


Figure 9. Flow chart showing the collection of cases included in study I and II

In **study III and IV** we investigated families from the south-western part of Sweden, with three or more affected members with primary VUR. All cases were confirmed by VCUG. Figure 10 illustrates the data selection process for **study III and IV**. Only seven of the previously participating families met selection criteria, therefore we extended the study period until 2012 for the retrospective data collection and recruited additional 5 large families. Thanks to awareness of the on-going study and personal knowledge about the family situation of the outpatients at the Paediatric Urologic Centre, additional two large families were identified and they accepted participation in the study in 2014. Consequently, fourteen families from the south-western

part of Sweden with ≥ 3 affected members with primary VUR were recruited for **study III** at Queen Silvia Children's Hospital. DNA samples were collected from all families except those who already provided that for **study II**. Four healthy in-house controls were included in the analyses. In the last study (**IV**) we further analysed thirteen out of the fourteen large VUR families. In the remaining family, the collected DNA quality did not meet the requirements for the last tests, but the family was not able to provide blood samples due to the decease of the proband. Therefore, they do not participate in **study IV**, which includes thirteen families.

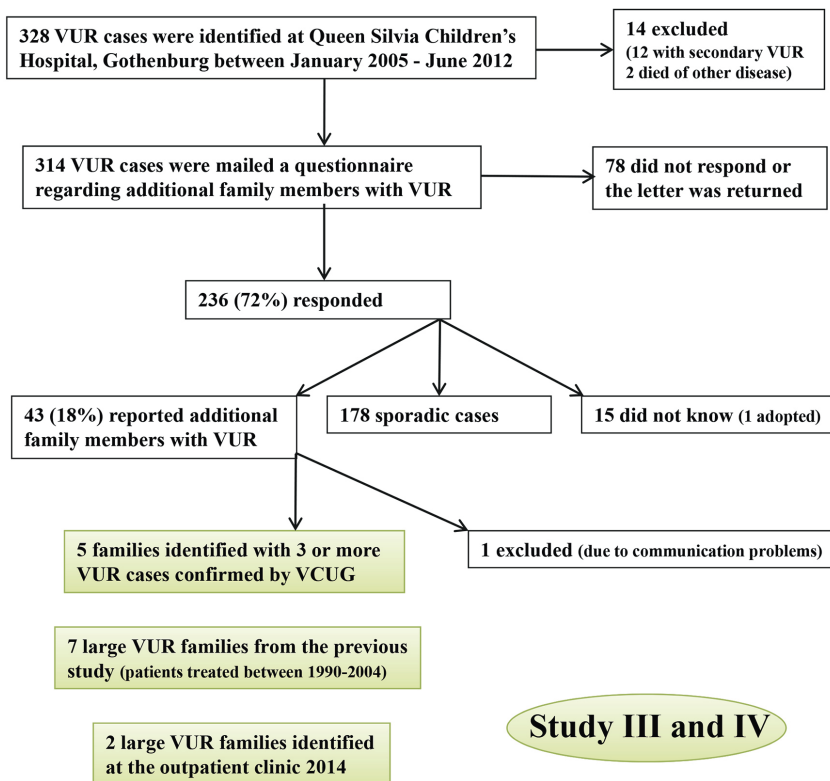


Figure 10. Flow chart showing the collection of families included in study III and IV

3.2 METHODS

Phenotyping the familial and non-familial VUR

Pedigrees were constructed for all families with hereditary VUR in order to clarify the relationship and analyse the pattern of inheritance. VUR in all affected individuals was characterised regarding presenting symptom, grade of reflux, recurrent UTIs, kidney damage, natural course of the reflux, surgical treatment and additional anomalies. Clinical data were obtained from medical records. Data was also collected on subjects who did not report any relatives with VUR (controls), a total of 361 individuals. In these patients we only recorded the presenting symptoms, grade of reflux at presentation, recurrent UTIs (yes or no) and renal damage.

Febrile urinary tract infections

When recording the recurrent UTIs, only the febrile infections ($>38.5^{\circ}\text{C}$) with positive urinary culture (at least 100,000 colony forming units/ml in urine obtained by midstream or bag technique) were counted in the hereditary groups. In the group of controls such data about symptoms and urinary culture were not always available.

Voiding cystourethrography

Grade of reflux was registered from VCUG investigations (Figure 11) according to the international grading system, both at presentation and at the last follow up prior to surgical treatment [3]. This investigation requires a transurethral catheter for bladder filling with iodine-based contrast medium. Spot fluoroscopy was performed with anterior images during filling and micturition. For boys, lateral projection was also mandatory to detect anomalies of the urethra, especially posterior urethral valves. Postvoid image was taken for all children. The radiation burden of VCUG is calculated as approximately 0,1-0,3 mSv/examination (the higher number accounts for investigation in boys). Thus, a VCUG corresponds to 1-3 months of background radiation.

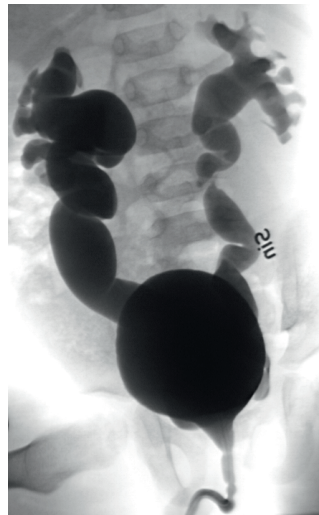


Figure 11. VCUG in a 20 months old boy with bilateral VUR grade V

Antibiotic prophylaxis was given in conjunction with the examination and sedation with midazolam was optional. The images were assessed and graded I-V. Resolution of VUR was defined as grade II or less. In case of bilateral VUR, the kidney with the highest grade of reflux was used to characterise the patient.

Renal scintigraphy

Evaluation of renal damage was performed by static or dynamic renal scintigraphy according to European standards [119, 120]. DMSA scan was the first method of choice, as it is the most appropriate method available for static cortical imaging, with a high sensitivity for detection of acute and chronic cortical abnormalities in the kidney (Figure 12). It detects focal renal parenchymal lesions in acute pyelonephritis, as well as renal sequelae 6 months after acute infection. Furthermore, this method can also diagnose associated anomalies: small, dysplastic and ectopic kidney. However, in the presence of dilated renal pelvis and calices, when obstruction or poor drainage could be expected, MAG3 scintigraphy, a dynamic method with fast tracer excretion, was preferred. Split function (side distribution of renal function in percent of the total renal activity uptake) can be estimated with both scintigraphic methods.

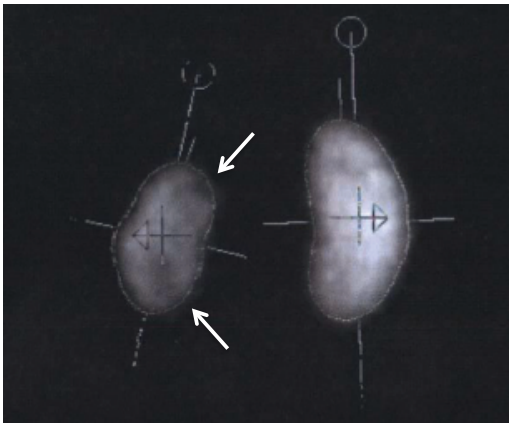


Figure 12. DMSA image of generalised renal damage in a small left kidney with split function of 30% of the total uptake in an 18 months old child. In addition, there is focal damage cranially and caudally in the same kidney (arrows) after a recent pyelonephritis.

Venflon catheter was required for both investigations for intravenous injection of the radionuclide. The radiation burden of DMSA scintigraphy is approximately 1 mSv/examination [119], while the MAG3 scintigraphy gives an effective dose of 0.2-0.4 mSv/examination [120]. To get a simple idea of what this corresponds to, it can be compared to the effective dose we receive from background radiation (from the ground and food), which is approximately 1-1.5 mSv per year.

On reviewing the scintigrams, focal damage was defined as one or more areas with reduced uptake or indentation of the renal outline, which is caused by postnatally acquired renal scarring. A small kidney with reduced tracer uptake or diffuse parenchymal anomaly was classified as generalised damage referred to as congenital renal hypodysplasia (Figure 12). The relative split function in normal kidneys has been shown to be $50 \pm 5\%$ (mean ± 2 SD) [119]. A kidney without uptake defect and a split function of 45% or more was classified as normal, whereas a kidney with reduced or absent uptake in one or more areas or a relative function of $< 45\%$ was considered abnormal. On analysis, the kidney with the most pronounced damage was used to characterize the case. Deterioration of renal status in an individual at follow-up was determined as new lesions on the renal scans or loss of $\geq 7\%$ in split function.

⁵¹Cr-EDTA-clearance and other GFR estimates

Measurement of glomerular filtration rate (GFR) provides an overall estimate of renal function. Measuring GFR requires the infusion of a substance that is filtered freely and is not reabsorbed, secreted, or metabolized by the kidneys. In this study ⁵¹Chromium-ethylenediaminetetraacetic acid (⁵¹Cr-EDTA), was used, given intravenously through a Venflon catheter. The radiation burden for ⁵¹Cr-EDTA is approximately 0.011 mSv/examination [121].

The reference value used for normal GFR was $110 \text{ mmol/l/1.73m}^2$ after two years of age according to Bröchner-Mortensen and GFR $< 80\%$ (< 2 SD) of expected was considered subnormal [122]. Deterioration of total renal function over time was defined as loss of $\geq 12\%$ of GFR. Before two years of age, steady state of GFR is not achieved, and for this age group we used the equation developed by Winberg for estimation of expected clearance [123]. In the cases where no GFR measurement was performed, estimated GFR (eGFR) was calculated according to the formula of Schwartz [124] from serum creatinine and the height (cm) of the patient.

DNA isolation

For **study II**, genomic DNA was isolated from blood lymphocytes according to standard procedures at the laboratory of the Centre for Molecular Medicine, (Karolinska University Hospital, Stockholm). Fourteen out of 66 eligible families who accepted participation could not provide blood samples for various reasons. All affected individuals in the family as well as both parents of the proband were asked to donate blood samples. Some families found it difficult to visit the laboratory during working hours and others refused enrolment in the study because of the child's fear of hospital and of

potentially painful procedures. Therefore, during **study III**, we tested different alternatives to blood samples for the collection of DNA, in order to minimize the inconvenience for participating children and their families. Of the tested saliva and buccal smear kits, Isohelix Buccal Swabs yielded good DNA quality in laboratory settings. The families received the sampling kit at home by post and could return the biological samples the same way. DNA was then extracted using a Maxwell 16 Buccal Swab LEV DNA Purification Kit (Promega, Madison, WI). DNA was quantified by NanoDrop Spectrophotometer (DeNovix DS-11 Spectrophotometer). However, in clinical settings, this method yielded DNA of lower quantity and quality than in laboratory evaluation, although the sampling kit was sent with attached sampling instructions. Despite suboptimal starting material, quality was still sufficient for SNP microarray and for Sanger sequencing methods used in **study III** and **IV**. However, analysis of whole-exome sequencing performed in **study IV** required better DNA quality and the individuals to be tested had to be asked to deliver blood samples instead. Qiagen DNeasy Blood & Tissue Kit (Qiagen, Aarhus, Denmark) was used for DNA extraction from blood and the DNA was evaluated through fluorometric and spectrophotometric quantitation before downstream applications.

Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction (PCR) technique, a fundamental tool for many genome analyses, was discovered by Kary Mullis in 1985 who was awarded the Nobel Prize in 1993 [125]. This rapid, sensitive, specific and reliable method is used to amplify a well-defined target DNA sequence present within a source of DNA to amounts required for further analysis. The selective amplification of a specific sequence is determined by oligonucleotide primers (15-25 bp long) designed to bind to complementary sequences flanking the target gene region. The synthesis of new complementary DNA strands is performed in presence of the four deoxynucleotides (dNTA, dNTC, dNTG and dNTT), a stabilising buffer with MgCl₂, the primers that bind the denatured DNA and a thermostable DNA polymerase (*Taq* polymerase). The *Taq* polymerase is derived from a thermophilic bacterium *Thermus aquaticus*, which lives in hot springs. PCR is called chain reaction because the newly synthesized DNA strands will act as template for further DNA synthesis in subsequent PCR cycles. The PCR cycle consists of three steps: 1) denaturation of the dsDNA template usually at $\geq 94^{\circ}\text{C}$, 2) annealing of primers to templates at a primer-specific annealing temperature and 3) elongation where the DNA is synthesised, commonly at 72°C for most DNA polymerases (Figure 13). PCR is performed by simply changing the temperature in a cyclic manner with the target DNA amplicon

accumulating exponentially, theoretically to 10^9 copies after 30 cycles. When the reaction components are consumed, the reaction slows down and the product starts to degrade. Therefore, after an adequate number of cycles (25-40), the reaction is stopped by decreasing the temperature and holding it at 4°C . In traditional PCR the products are then evaluated by electrophoresis on agarose gel after being dyed through Ethidium bromide (EtBr) or other fluorescent stains such as GelRed® in order to estimate size, specificity and concentration.

PCR is a fundamental step for many different methods in molecular genetics and in this thesis we have frequently used PCR as base for mutation detection methods (**study II** and **IV**) and SNP genotyping with 250K microarray (**study III**). Primer sequences were selected using either Primer 3 program (<http://primer3.wi.mit.edu>) [126] or ExonPrimer (<https://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html>). In **study II** the PCR reactions were performed either with AmpliTaq Gold (Applied Biosystems, Foster City, CA) or with DyNAzyme EXT DNA Polymerase (Finnzyme, Espoo, Finland). Titanium DNA Amplification Kit (Clontech Laboratories, Mountain View, CA) was used in **study III** and AmpliTaq Gold 360 Master Mix (Applied Biosystems) in **study IV**.

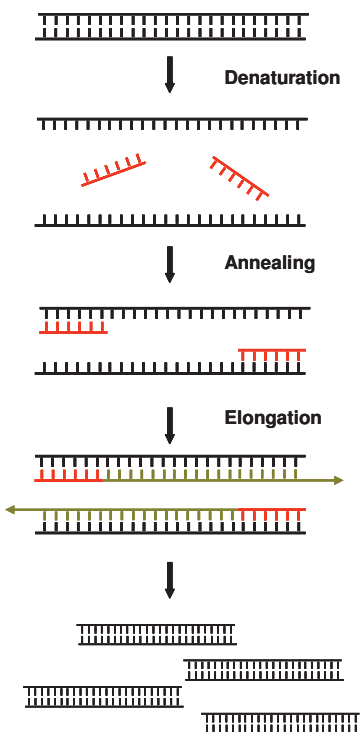


Figure 13. Schematic illustration of PCR amplification. The double-stranded DNA molecule is denatured at high temperature ($\geq 94^\circ\text{C}$). When the temperature is decreased (to $50\text{-}60^\circ\text{C}$ depending on the primer) during the annealing phase, the primers (in red) bind to the single-stranded template. This is followed by elongation where the DNA is synthesised (green sequence) at slightly higher temperatures (72°C). These three steps are repeated in consecutive rounds (25-40 cycles). Reprinted by permission from Susanne Fransson

The disadvantage of the PCR technique is that replication errors can occur and might be amplified during the process, thus they can be misinterpreted as mutations in following sequence analyses. Moreover, the method is highly sensitive to external contamination with DNA from other sources, which requires strict protocols to minimize contamination during the process.

Sanger sequencing

DNA sequencing is the determination of the precise nucleotide sequence in a DNA template. The common method for this was developed by Sanger et al in 1977 and is called the Sanger method, dideoxy method or first-generation sequencing [127]. Frederick Sanger received his second Nobel Prize in Chemistry in 1980, this time a shared prize, for his contribution to the determination of base sequences in nucleic acids. The Sanger method relies on a cycle reaction similar to PCR, but in contrast to a normal PCR reaction, only one primer is used, either forward or reverse, thereby leading to linear amplification. In addition to the normal deoxynucleotides (dNTPs), fluorescence-labelled base-specific dideoxynucleotides (ddNTPs) are added, one for each base. The ddNTPs lacks the hydroxyl group at the 3' end, which is required for the phosphodiester bond which connects the next nucleotide in the elongation process. Therefore, when they are randomly incorporated into the growing DNA chain, they cause an abrupt termination of the DNA synthesis. Hence, the sequence reaction products will consist of nucleotide chains of different lengths with a fluorescence-labelled nucleotide at the end. These nucleotide chains are precipitated, re-suspended in Formamide solution to keep them single-stranded and then separated on a gel in an automated sequencer. The fragments are separated according to their size and the emitted fluorescence signal from each excited fluorescent dye at the end of each fragment determines the identity of the nucleotide in the original DNA template. The subsequent nucleotide sequence is presented by a software program as peaks on an electropherogram (Figure 14).

In this thesis, Big Dye Terminator ddNTP chemistry (Applied Biosystems) was used to sequence PCR products from genomic DNA. Before setting up the sequence reaction in **study II**, the PCR products were purified from remaining PCR primers with ExoSap-IT enzyme (USB Europe, Staufen, Germany). The PCR fragments were sequenced in both directions using the BigDye Terminator v3.1 kit (Applied Biosystems) and analysed in ABI Prism 3730 sequencer (ABI, Foster City, CA). Sequence analysis was performed with the program SeqScape v2.5 (Applied Biosystems). In **study IV**, PCR products were submitted to GATC Biotech (Constance, Germany) for purification and sequencing with a Big Dye Terminator Cycle Sequencing

Ready Reaction Kit (PE Applied Biosystems). All amplicons were analysed in two separate sequence reactions using forward and reverse primer respectively. The sequencer used by GATC Biotech was ABI 3730xl DNA Analyser systems (Applied Biosystems) and results were visualized for manual inspection with SnapGene software (GSL Biotech, Chicago, IL).

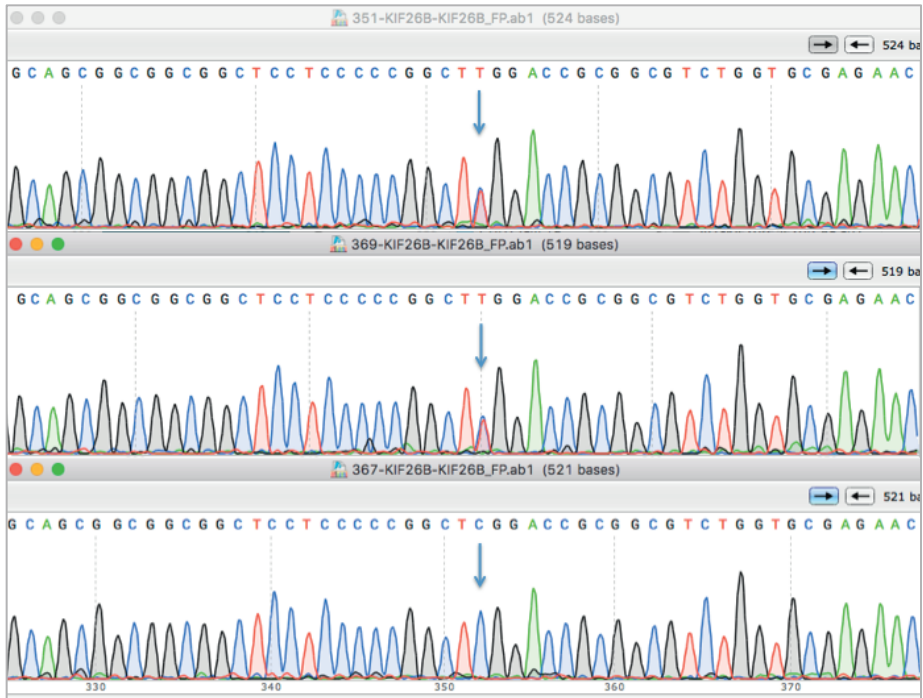


Figure 14. Electropherogram of a section of the *KIF26B* gene in members of Family 17. *VUR* affected individuals 351 (upper panel) and 369 (middle panel) show a novel heterozygous single nucleotide variant (SNV, indicated by overlapping red and blue peaks, corresponding to T and C respectively) at locus Chr1:245,319,888, marked by arrow. Non-affected individual 367 (lower panel) is homozygous for C (blue peak only) at this locus.

DNA sequencing detects all subtle mutations, even a point mutation where a single nucleotide base is changed. Yet, some single heterozygous nucleotide substitutions can be hard to detect if one of the base-peaks in the electropherogram is very low. Moreover, depending on the number of PCR-fragments and on the number of samples to analyse, this method can be very time-consuming and therefore expensive.

Whole-exome sequencing (WES)

The field of sequencing underwent a tremendous development lately. The massively parallel sequencing technology, also known as next-generation sequencing (NGS) or second-generation sequencing, has revolutionized biological science and it is making its way from research laboratories into clinics. With NGS the entire human genome can be sequenced within a single day, as the sequencing occurs along hundreds of millions of short templates in a massively parallel fashion. Since 2004, three principal NGS technologies are available and these platforms differ in engineering configuration and sequencing chemistry [128]. We used the reversible termination method on the Illumina sequencing platform. First the sequencing library is prepared by random cleavage of the DNA sample into short fragments (approximately 300-400 bp each) and by ligation of adaptors to both ends of the fragments. The fragments are hybridised to a glass slide, a flow cell (Figure 15), through surface-bound adaptors that are complementary to the added library adaptors.

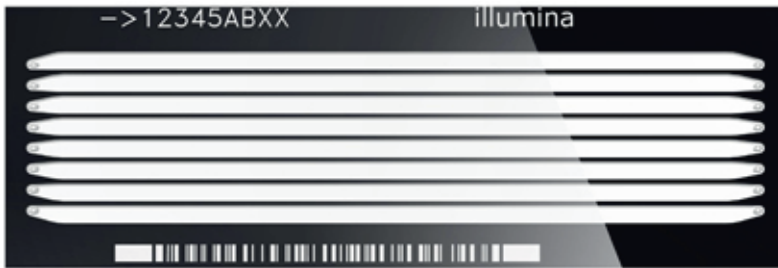


Figure 15. HiSeq 3000/4000 Flowcell, image courtesy of Illumina, Inc.

PCR is carried out to amplify each read, creating a cluster with multiple identical copies of the same read. All amplified molecules are then separated into single strands before sequencing. A sequencing primer that binds to the adaptor is added to the flow cell before it is flooded with nucleotides and a DNA polymerase. These nucleotides are fluorescently labelled, with a different colour corresponding to each base. They also have a terminator, so that only one base can be added at a time. In each cycle, the clusters are excited by a light source as a fluorescent signal is emitted and captured from each cluster. In each cluster location, there will be a fluorescent signal indicating which base has been added. After scanning the clusters for fluorescent signals, the flow cell is rinsed from unincorporated nucleotides and the 3'hydroxyl-blocking group is removed before the cycle starts over again. This process is called sequencing by synthesis. By following the emitted signals after every cycle for all clusters, the sequence of each specific fragment attached to the flow cell can be noted (see Figure 16 for NGS steps).

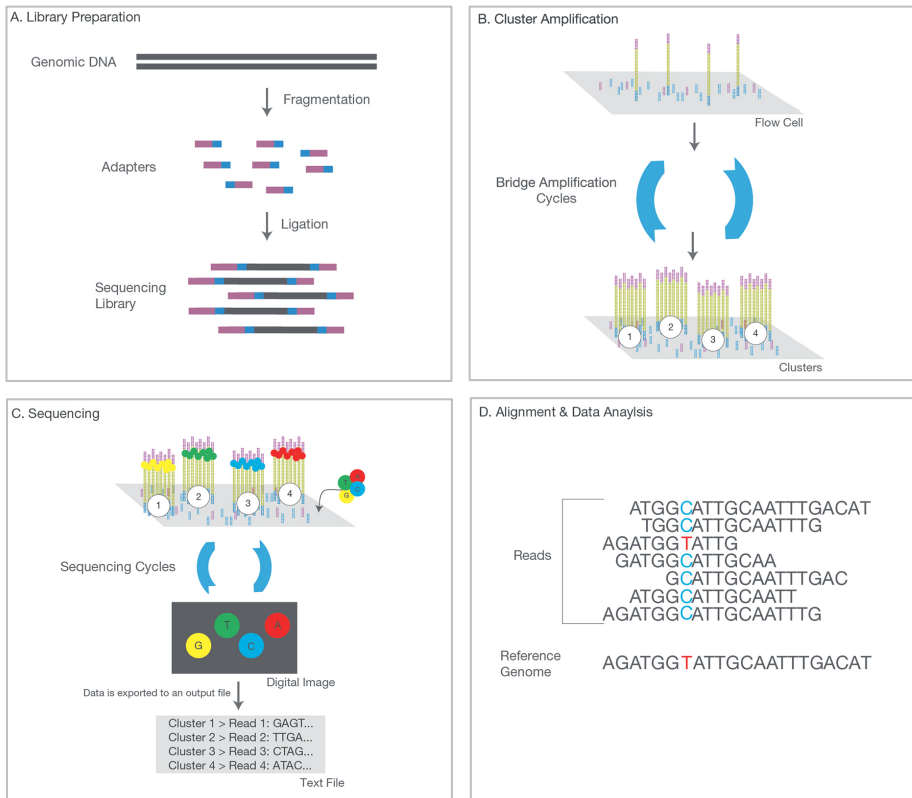


Figure 16. Overview of the Next-generation sequencing procedure according to Illumina. (A) NGS library is prepared by fragmentation of a genomic DNA sample and ligation of specialized adapters to both ends of the fragment. (B) The library is loaded to a flow cell where the fragments hybridize to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification. (C) Sequencing reagents, including fluorescently labelled nucleotides, are added and the first base is incorporated. The flow cell is imaged and emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. The cycle is repeated "n" times to create a read of "n" bases. (D) The generated reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified. Image courtesy of Illumina, Inc.

The sequence constructed from a single cluster is commonly referred to as a "read". All of the sequence reads will be the same length, as the read length depends on the number of cycles carried out. Pair-end sequencing is used to improve the accuracy of reads mapping onto a reference genome, that is, a single fragment is sequenced consecutively from both ends. After completion of the first read, the molecules are flipped on the flow cell, by bending over and bridging to another surface-bound adaptor, and the second sequencing primer is added. The flow cell is flooded again with marked nucleotides and DNA polymerase and the sequencing of the reverse strand starts. The

contiguous sequences (forward and reverse strands) are mapped back to the human reference genome, in our studies GRCh37/hg19, for subsequent variant identification.

NGS can be performed in different scales, from studying individual exons to whole-genome sequencing (WGS). Whole-exome sequencing (WES) targets the protein coding regions, corresponding to about 2% of the genome. These genomic regions of interest (e.g. exons) are captured by enrichment process. Labelled probes hybridise with the selected regions, after which the hybridised DNA fragments can be pulled down and washed from excess material. Only sequencing a fraction of the whole genome reduces the cost for sequencing, storage and analysis and thereby makes it feasible to increase the number of samples to be sequenced. Since the coding region of the genome has been characterized to a higher degree, it is easier to interpret identified variants. The drawback is that WES omits regulatory regions such as promoters and enhancers. In addition, WGS but not WES allows high-resolution examination of copy number variations (CNVs), as it covers both coding and non-coding regions of the genome.

In **study IV** genomic DNA was subjected to whole-exome sequencing (WES). The DNA samples were submitted to GATC Biotech (Constance, Germany) for sequencing on Illumina instrumentation (Illumina, San Diego, CA) after DNA enrichment using Agilent SureSelect human All exon v6 (Agilent technologies, Santa Clara, CA). Primary analyses (e.g. image analysis and base calling) were generated by the Illumina pipeline using default settings.

Sanger sequencing is still the gold standard method when it comes to verifying NGS results or sequencing of single genes with few exons. Nevertheless, NGS is superior to the Sanger method in the case of simultaneous interrogation of a large number of genes/genomic regions at the same time or when analysing samples with low-input DNA. The advantages of next generation sequencing technologies combined with the rapid decrease in the cost of sequencing has now made NGS competitive with microarrays for most of the assays (e.g. transcription factor binding analysis, gene expression analysis, DNA methylation analysis) with the possible exception of SNP genotyping. DNA arrays are being rapidly replaced by sequencing for nearly every assay that has been previously performed on microarrays.

Data processing WES

The analysis of the output data from NGS includes base calling, quality control, alignment and variants calling. Sequencing platforms (Illumina, in

this particular case) usually have integrated base calling software. Base quality score, a measurement of uncertainty to each base call from image analysis, is typically given in the standard Phred quality score (Phred score 20 corresponds to 1% error rate in base calling). Read trimming (removing poor quality reads), mapping to the human reference genome, hg19, and variant calling were performed using CLC Biomedical Genomics Workbench software (Qiagen, Aarhus, Denmark). The accuracy of a variant calling depends on mapping quality, read depth and allele balance. Only high-quality called variants with a minimum of 15% allele frequency and a total read coverage of ten were considered for further analysis. Coverage (read depth) is influenced by low-complexity sequences, when a high proportion of reads share identical start sites, resulting in alignment artefacts. Variants with an allele frequency above 1% in either SweGen dataset (<https://swegen-exac.nbis.se>), 1000 genomes (<http://www.internationalgenome.org/>), Exome Aggregation Consortium (ExAC, <http://exac.broadinstitute.org>), Genome Aggregation Database (gnomAD, <http://gnomad.broadinstitute.org>) or NHLBI Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>) were discarded, as well as all synonymous variants or variants in non-coding regions except those affecting canonical splice sites. Remaining variants were assessed manually through the Integrative Genomics Viewer (IGV) for the removal of calls due to mapping artefacts and paralogues [129]. PolyPhen 2, SIFT and CADD were used to predict the functional relevance of called single nucleotide variants (SNVs) and possible relevance to the biological disease context was assessed using the Ingenuity variant analysis tool (<https://www.qiagenbioinformatics.com/>).

The remaining potential rare causal gene variants were further filtered via an extensive literature search, where we focused on gene function of interest (e.g. involvement in signalling pathways in the embryological development of the kidney and the urinary tract), associated animal models and tissue expression. Genes associated with syndromes were included if a connection to kidney development or VUR was stated, but syndromes with other CAKUT phenotypes were excluded.

SNP microarray

A DNA microarray (DNA chip or biochip) is a collection of DNA oligonucleotides (probes) attached at known locations to a solid surface. The array technique was initially developed for high throughput genotyping, but now it exists in a variety of forms and has different areas of use including analysis of gene expression or DNA methylation. In high-density oligonucleotide SNP arrays, hundreds of thousands of probes are arrayed on a

small gene chip, allowing for many SNPs to be interrogated simultaneously. The first SNP genotyping chip was prototyped by Wang et al in 1998 and allowed simultaneous genotyping of 500 SNPs [45]. Over the past two decades, the synergistic relationship between advances in biological understanding, computational methodology and the technological development in the arrays themselves, spurred great progress of the method. Subsequent versions increased stepwise from 10 000 to 500 000, and finally to over one million SNPs in the current release. Different versions of SNP chips contain somewhat different sets of SNPs. Each probe on the chip is 25 nucleotides long and is designed to be complementary, or very nearly complementary, to a portion of the DNA sequence harbouring the SNP site. Commonly, biallelic SNPs are used and thus, two different probes must be utilized for each SNP position in order to detect both alleles. The two alleles of an SNP are often designated as A and B rather than referring to the specific bases associated to the respective alleles. Since each individual inherits one allele from each parent for most genomic loci, the individual's genotype at a SNP site could be AA, AB or BB. The DNA sample under investigation is fragmented and labelled with fluorescent dye prior hybridization to the microarray with immobilised oligonucleotide probes. A detection system records and interprets the hybridisation signals from all probe positions.

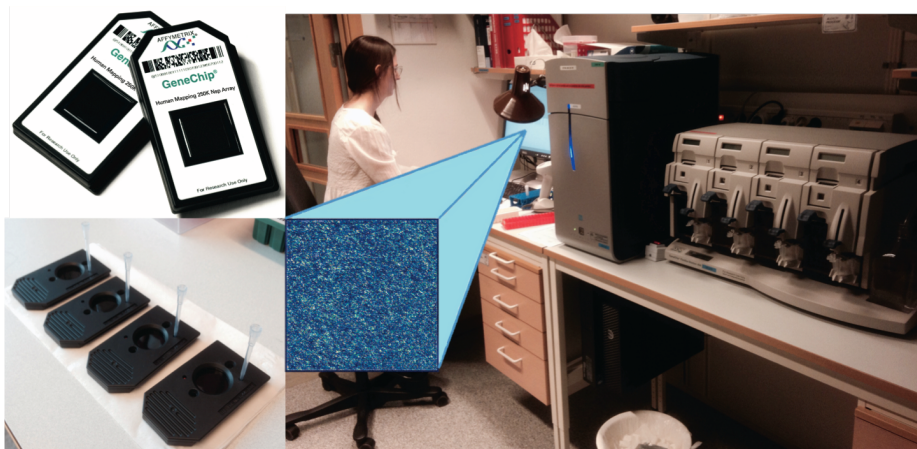


Figure 17. SNP microarray. Upper left panel of the picture: GeneChip® Human Mapping 250K NspI Array. Actual size of the hybridization surface of the GeneChip® array is 1.28 x 1.28 cm. Lower left panel: Preparing to load the biotinylated PCR fragments on the GeneChip® for hybridisation overnight. Picture to the right: Scanning the GeneChip®. Affymetrix Fluidics Station 450 on the right of the GeneChip Scanner 3000. Central panel: GeneChip® Array output, on the computer workstation. The upper left and central image courtesy of Affymetrix, Inc., Santa Clara, CA, USA

In addition to SNPs, copy number variants (CNVs) are also important contributors to human genetic variation. These submicroscopic chromosomal imbalances (deletions or duplications) appear to be associated with a variety of common diseases such as autism, schizophrenia, Crohn's disease, rheumatoid arthritis, type 1 diabetes, obesity and numerous developmental diseases [111]. The SNP arrays can be used to detect CNVs. When the fluorescent labelled fragments hybridise with the probes, expected probe intensity increases with increased quantity of DNA harbouring the region interrogated by the probe, and vice versa with loss of genomic material. Thus, fluorescent intensity for each probe also enables the copy number to be inferred from the array after *in silico* normalisation against controls. Thereby the SNP-microarrays can be used for both genotyping and copy number calling.

In **study III** we used Affymetrix 250K SNP NspI arrays (Affymetrix Inc., Santa Clara, CA), which detect ~262,000 SNPs. The genomic DNA was digested with the NspI restriction enzyme and ligated to adaptors. After ligation, the template was subjected to PCR amplification using a generic primer that recognises the adaptor sequence. The amplified DNA was fragmented with DNase I, labelled with biotin and hybridised to a GeneChip® Human Mapping 250 K array (Figure 17). The hybridised probes were washed using the Affymetrix Fluidics Station 450 and marked with streptavidin-phycoerythrin. The arrays were scanned using a confocal laser scanner, GeneChip Scanner 3000 (Affymetrix) (Figure 17). Primary data analysis was performed using GDAS (GeneChip® DNA Analysis software) and GTYPE (Affymetrix) for the extraction of genotype calls.

This method for SNP genotyping has been highly successful and is widely used. Its genotyping accuracy is well over 99.5%. Nevertheless, the method has some limitations. Arrays provide an indirect measure of relative concentration. There might be homologous sequences in the genome that bind to the same probe. Finally, DNA array can only detect sequences that the array was designed to detect, meaning that genes that have not yet been annotated in a genome will not be represented on the array.

Data processing SNP microarray - the Colour Method

SNP genotype data for individuals were analysed using the Colour Method, previously described by Östensson and Ohlsson et al. [92, 130] This method can be applied for dominant traits in families where several of the subjects are affected. In the original Colour Method, it was assumed that all affected individuals have a common ancestor and share the causal variant IBD

(identical by descent). Similar to the homozygosity mapping, we can identify shared haplotypes, which are candidates for the causal variant of the disease, using different subsets of the affected subjects or families. The aim is to identify regions free from incompatibilities. For each SNP locus, individuals can have either genotype calls 'AA', 'AB' or 'BB'. A locus where at least one affected individual is 'AA' and at least one other affected individual is 'BB' is scored as an incompatibility. A locus of this kind is therefore an 'incompatibility' and cannot by definition be included in the correct disease gene haplotype. In contrast, a continuous large region of SNP loci, without any incompatibilities among affected individuals, may include a unique disease haplotype and, consequently, also the disease gene. Genotypes for one affected individual per family were compared and incompatibilities as defined above for all the 260,000 SNP loci were scored and plotted against the genome position for each locus. Corresponding genotype data generated by the Affymetrix 250K array in healthy control individuals were entered one by one into the analysis. Given that an affected study subject and an unaffected control do not have the same disease phenotype, their DNA must by definition be different. The regions of the genome that were identical by state (IBS) in both VUR cases and controls were therefore excluded. After testing the controls one after the other, the shared region that remains would be the region identical by descent (IBD).

Additionally, a second strategy was tested. The controls were used to rule out common haplotypes within the general population. We only excluded regions that were identical for VUR cases and all four controls. The haplotype region unique to the affected individuals and not present in all four controls might be specific to the disease.

All genomic positions for SNPs are given relative to the February 2009 – GRCh37/hg19 genome assembly. The UCSC genome browser (<https://genome.ucsc.edu>) was used to visualise gene regions, which theoretically may contain the disease gene. We recorded the genes, both coding and non-coding DNA sequences, in these regions. In addition, we examined their expression and role during kidney development using GUDMAP (the GenitoUrinary Development Molecular Anatomy Project data; <https://www.gudmap.org>) and via an extensive literature search (<https://www.ncbi.nlm.nih.gov/pubmed>).

Data processing SNP microarray - Copy number detection

The SNP microarray data was also used for detection of copy number variants. Figure 18 shows an example of CNV from an unrelated study. The

R package ‘aroma.affymetrix’ was used for copy number detection. We used the CRMA v2 method [131] for copy number estimation of the intensity values obtained from the CEL files for each sample. For copy number segmentation we used the Circular Binary Segmentation (CBS) method [132].

The ‘aroma.affymetrix’ package is an R package for analysing small to extremely large Affymetrix data sets. It allows you to analyse any number of arrays of various chip types, e.g. 10,000s of expression arrays, SNP chips, exon arrays and so on. Criteria for inclusion of CNVs were: 1) number of SNPs per CNV > 10; 2) CNV size >50 kb but < 3 Mb; 3) CNV frequency <1% in the general population. They were also filtered by the log₂ values: those in-between -0.2 – +0.2 were excluded. We searched for identical CNVs within the families. CNVs were defined as identical if they had the same CNV state, showed ≤ 30% difference in length, and overlapped > 70%. The significance of each CNV detected was determined by comparison with public CNV databases: Database of Genomic Variants (DGV, <http://dgv.tcag.ca/dgv/app/home>) and Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER, <https://decipher.sanger.ac.uk/>). Analysis was performed using UCSC Genome Browser February 2009 – GRCh37/hg19 assembly.

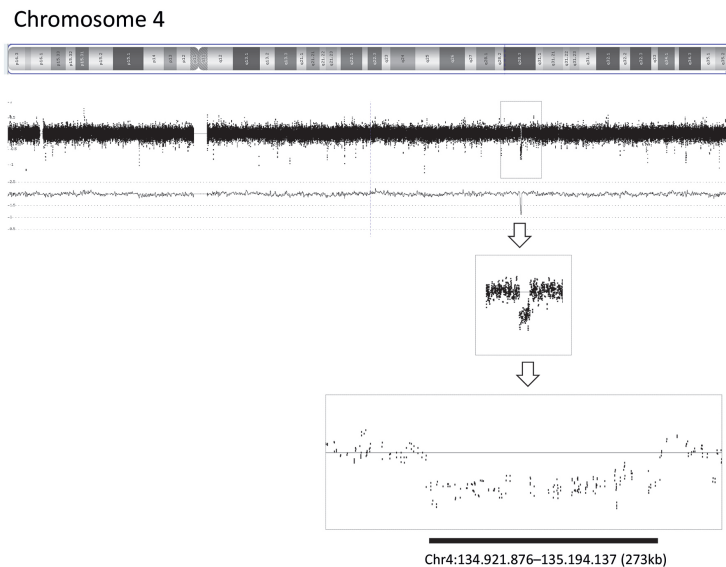


Figure 18. SNP microarray profile of chromosome 4 and subsequent magnifications of a genomic region harbouring a copy-number variation (deletion) of 273kb. The profile has been generated using Affymetrix CytoScan High-density Array (with 2,7 million probes). The dots show signal intensity for each probe plotted against the position of the probe in the genome. Reprinted by permission from Tommy Martinsson

3.3 STATISTICS

In order to interpret collected data different statistical methods are used. Studies often start with a descriptive part and continue with the analytic part in order to draw conclusion from the observations. The nature of the data determines what statistical analysis to apply. Parametric tests are used for normally distributed numeric data, whereas non-parametric tests make no assumption of the distribution of the data and are applied for numeric as well as categorical variables. The p-value shows the probability of detecting a difference just by chance, when in fact there is no real difference (i.e. the risk of making a type I error). The commonly chosen significance level is 5% ($p=0.05$) [133].

The studies in this thesis are observational. Study I is a retrospective cross-sectional case-control study comparing familial VUR with non-familial VUR as well as a longitudinal study on familial VUR. The data collected are mainly *categorical* (qualitative) such as gender, presenting symptoms for VUR, recurrent UTIs, renal abnormalities. Some of these are dichotomous (binominal), such as gender, recurrent UTIs (yes/no) or renal function (normal/subnormal). Renal damage/abnormality can be regarded as dichotomous if the point is if there is damage or not, but it can be an ordered variable (*ordinal* data) when it is categorised as normal, focal damage and general damage. VUR grade I–V is another ordered categorical variable. A non-ordered categorical variable (*nominal* data) is for example presenting symptom of VUR. *Numeric* (quantitative) data can be *continuous* or *discrete*, however, these data can also be categorised as *interval* or *ratio*. Age at presentation (in months) is an example of continuous ratio data. Temperature measured in °C is continuous interval data, whereas measured in K (kelvin) is continuous ratio data. The number of UTIs is a discrete variable.

For categorical variables n (%) is presented, and for continuous variables Median (Min - Max). For comparisons between groups the chi-square test was used for non-ordered categorical variables, while Kruskal-Wallis test was used for continuous variables. For pairwise comparison between groups, Fisher's exact test was used for dichotomous variables, Mantel-Haenszel chi-square test was used for ordered categorical variables, chi-square test was used for non-ordered categorical variables and Mann-Whitney U-test for continuous variables. All tests were 2-tailed and conducted at the 5% significance level.

In Study II-IV public databases and programs were used to evaluate the relevance of our findings, e.g. NCBI (<https://www.ncbi.nlm.nih.gov>),

SweGen (<https://swegen-exac.nbis.se>), ExAC (<http://exac.broadinstitute.org>), 1000 genomes project (<http://www.internationalgenome.org/>), gnomAD (<http://gnomad.broadinstitute.org/about>), UCSC (<https://genome.ucsc.edu>), NHLBI ESP exome (<http://evs.gs.washington.edu/EVS/>), NNPREDICT (University of California), PolyPhen (<http://genetics.bwh.harvard.edu/pph2>) as described earlier in the Methods section. Only variants with an allele frequency below 1% in all populations in public databases (SweGen, ExAC, 1000 genome, gnomAD and NHLBI ESP exome) were considered for further assessment.

The more significance tests are made on one set of data, the more likely it is that a significant finding is due to chance (type I error). For example, in case of 100 tests at the $p < 0.05$ level, about five statistically significant results will be false positives. One way to correct for multiple testing is using the Bonferroni method, i.e. to divide the significance threshold by the number of tests conducted. This method assumes that the individual tests are independent of each other and could thereby be too strict in many multiple testing settings. It may lead to a very high rate of false negatives (type II error). When using high throughput methods such as DNA microarrays for genotyping hundreds of thousands of genetic markers and NGS for mutation screening of thousands of genes the issue of multiple testing should be addressed [134].

Our sample size is too small for significance testing, thus further, larger cohorts and additional studies are needed to validate the findings.

For analysing this enormous amount of data and possible combinations as in study III and IV, bioinformatics resources and assistance of a biostatistician are prerequisites.

3.4 ETHICAL CONSIDERATIONS

The Regional Ethical Review Board in Gothenburg approved the studies (Dnr 589-05). Each family received written and oral information about the study and gave their written consent to participate. During the phone interview the parents had the possibility to ask questions and they could contact us for further information when needed. Participation was voluntary and they could leave the study at any time.

New technologies in genomics raise important ethical issues, such as how to deal with the vast amount of the acquired information. Should the individuals in these studies be allowed to have access to their sequencing information? The investigations may lead to unexpected, accidental findings that are not related to the actual study, e.g. finding a risk factor for a serious illness later in life. Is such knowledge of any clinical benefit for the patient? This area of genomics still remains a challenge and researchers are looking into how to address these questions. Paramount importance must be given to how personal data with such sensitive information can be handled safely.

4 RESULTS

4.1 FAMILIAL AGGREGATION OF VUR

Of 560 individuals with VUR who responded to the questionnaire regarding relatives with VUR (for **study I and II**), 99 reported additional family members, 361 reported no relatives with VUR and 57 reported diffuse urinary tract problems in family members but without diagnosed reflux. Due to the uncertainty of whether or not this latter group had hereditary reflux, they were excluded. Thus, 22% (99/460) had relatives with reflux. These 99 patients belonged to 66 families. The child that during the study period was the first among the siblings to be diagnosed with VUR was denoted as index/proband. A total of 104 relatives with confirmed or highly probable VUR were reported and these affected relatives were analysed as a separate group in comparison with index and non-familial cases. The distribution of relatives with a history of VUR was: 38 siblings, 20 parents (15 of whom were mothers), 19 cousins and second cousins, 15 aunts/uncles and 12 grandparents. The proportion of girls among siblings and cousins was 22/38 (58%) and 10/17 (59%) respectively, which does not differ from the proportion in the entire material (58%).

Table 3. Sixty-six families with hereditary VUR; number of affected subjects/family and maternal/parental transmission

No of VUR patients/family	No. of families	Maternal transmission	Paternal transmission	Unknown transmission
2 affected/family	41	12	7	1 unknown side, 21 only sibs
3 affected/family	18	7	8	3 both sides
4 affected/family	4	4		
5 affected/family				
6 affected/family	3	2		1 both sides
Total number	66	25	15	26

The numbers of affected members in the families are shown in Table 3. 25/66 (38%) families had three or more affected members, whereas 41 only had two. In this latter group with 2 affected individuals, 21 of 41 were sibs. In total, VUR was inherited from the mother's family in 25 cases while paternal inheritance was seen in 15 cases.

4.2 PHENOTYPE OF FAMILIAL AND NON-FAMILIAL VUR

For analysis, the material was divided into index patients (n=66), relatives with VUR with available clinical data (n=55) and controls with non-familial VUR (n=358). The results are presented in Table 4.

Table 4. Demographic data, VUR grades, renal abnormalities and function split by hereditary (index and relatives) and non-hereditary (controls) VUR

Variable	Index n=66	Relatives n=55	Controls n=358	Test between groups: p-value		
				Index vs relatives	Index vs controls	Relatives vs controls
Gender						
Female	38 (58%)	37 (67%)	201 (56%)			
Male	28 (42%)	18 (33%)	157 (44%)	0.365	0.925	0.152
Presenting symptom						
Pyelonephritis	61 (94%)	36 (69%)	308 (90%)			
Screening ¹	2 (3%)	14 (27%)	25 (7%)			
Others	2 (3%)	2 (4%)	9 (3%)	0.001	0.449	<0.0001
Age at presentation ²	8 (0.1 - 84)	7 (0.1 - 98)	8 (0 - 141)	0.674	0.637	0.893
Grade of reflux						
I	2 (3%)	4 (8%)	9 (3%)			
II	7 (11%)	11 (23%)	42 (12%)			
III	21 (32%)	22 (45%)	123 (34%)			
IV	24 (36%)	8 (16%)	140 (39%)			
V	12 (18%)	4 (8%)	44 (12%)	0.002	0.436	0.0003
Grade at follow up ³						
0 - II	18 (32%)	18 (47%)				
III - V	38 (68%)	18 (53%)		0.136		
Recurrent UTIs						
No	27 (44%)	20 (47%)	168 (50%)			
Yes	35 (56%)	23 (53%)	171 (50%)	0.919	0.465	0.831
Renal damage						
No	18 (28%)	13 (36%)	90 (26%)			
Yes	46 (72%)	23 (64%)	255 (74%)	0.543	0.840	0.276
Total renal function						
Normal	50 (82%)	34 (79%)	226 (87%)			
Subnormal	11 (18%)	9 (21%)	33 (13%)	0.900	0.378	0.235

Categorical variables, n (%); Continuous variables, median (range); ¹ Pre- or postnatal screening; ² Age in months; ³ Grade of reflux at the last VCUG prior to eventual surgical treatment, follow-up time median 32 months (range: 5 - 117)

There was an overrepresentation of females in all three groups, with no significant difference between the groups. The presenting symptom was pyelonephritis in the majority of cases. However, VUR in relatives had been detected to a significantly higher extent by pre- or postnatal screening than in index cases and controls. Fourteen of 52 (27%) relatives were diagnosed by screening in comparison with only 2 of 65 (3%) index patients and 25 of 342 (7%) controls ($p < 0.0001$). Nine of the 14 screened relatives were examined due to reflux in sibling. Age at presentation was median 8 months (range: 0 - 141), with no difference between the groups. The number of individuals with recurrent UTIs was high in all three groups.

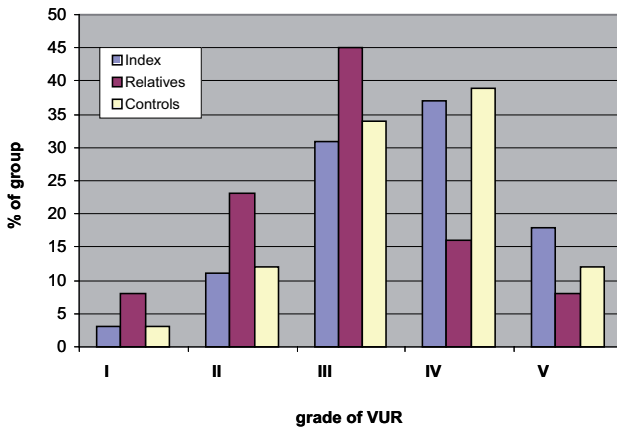


Figure 19. Grade of VUR according to study group, hereditary (index patients and relatives) and non-hereditary (controls).

The grade of reflux at presentation was significantly lower among relatives than among index patients ($p = 0.002$) and controls ($p = 0.0003$) (Table 4, Figure 19). Bilateral reflux was more common than unilateral, with no differences between the groups. Nearly 2/3 of relatives and 3/4 of index patients and controls had renal damage (Table 4). The distribution between focal renal damage and general hypodysplasia is shown in Figure 20. Most of the renal damage was unilateral.

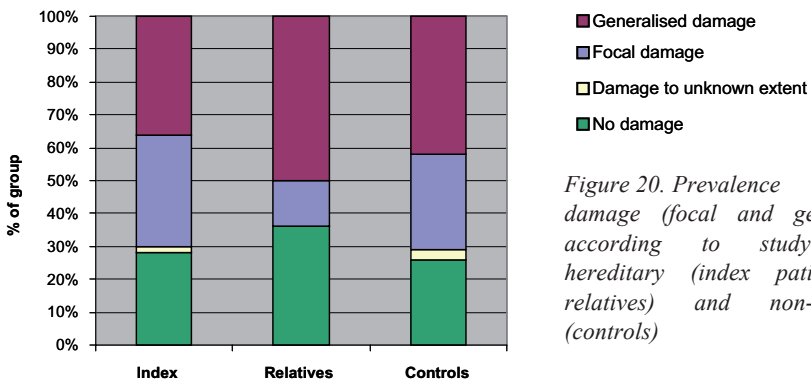


Figure 20. Prevalence of renal damage (focal and generalised) according to study group, hereditary (index patients and relatives) and non-hereditary (controls)

4.3 MUTATION SCREENING OF *ROBO2* GENE

Of the 66 families 52 families donated blood samples for **study II**. Two large families with hereditary VUR were additionally included from Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, by our co-researchers.

A total of 54 unrelated patients with familial VUR were analysed and compared with a control group consisting of anonymous DNA samples of 96 healthy voluntary blood donors at Karolinska University Hospital. All 26 exons of the *ROBO2* gene were normal on sequencing of the 54 cases. Six sequence variants were observed in the exon-intron boundary area of which two (IVS13+43C>T and IVS21-40 T>C) have not been described earlier. The IVS13+43C>T sequence variant was found in heterozygous form in an affected sib-pair, it was inherited from an apparently healthy mother and was not detected in the control DNA samples. The sequence variant IVS21-40 T>C was detected in one patient in heterozygous form, in an affected sib pair in homozygous form and in one of the control DNA samples in heterozygous form. Four patients presented the heterozygous sequence variant IVS3-3 C>T, however it did not co-segregate with VUR in the families, and two of the control DNA samples also showed this variant. Bertoli-Avella et al have also detected this variant and regarded it as possible polymorphism [86]. We predicted that none of the above sequence variants would alter gene splicing.

4.4 MUTATION SCREENING OF *SLIT2* GENE

The 37 exons of the *SLIT2* gene were subjects to Sanger sequencing. Through this analysis we detected twenty single nucleotide variants, whereof two were novel. One variant, c.4253C>T, which was found in two families, leads to an amino acid substitution in a relatively well-conserved amino acid, p.Ala1418Val, which was predicted to cause an altered secondary structure but to have little impact on the three-dimensional structure. This missense variant was not found in any control subjects but it did not segregate with VUR in the two families (family 50 and 53, Figure 21). Out of the 20 variants, additional three sequence changes mapped within the exons, but they were synonymous. One variant in the exon-intron boundary, IVS7+81C>G, has not been described previously. A total of 11 cases and 21 controls presented this variant, mostly in heterozygous form but some in homozygous form. Consequently, it may represent a non-pathogenic variant. None of the detected sequence variants were predicted to cause any splicing alteration or protein structure modifications.

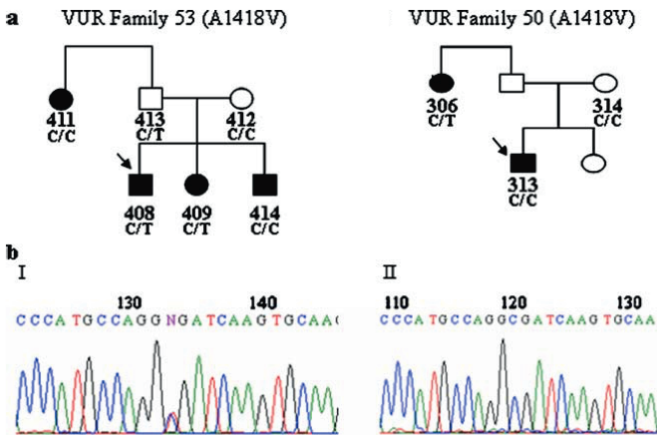


Figure 21. *SLIT2* missense variant in familial VUR. **a** Pedigree of family 53 and 50 with the nucleotide changes of the exon 36 (c.4253C→T), variant indicated below each individual. Arrow, proband; filled symbols, patients with VUR **b** Chromatograms showing C→T change in heterozygous form in exon 36 of family 53 and 50 (I) and a normal sequence (II)

4.5 TESTING FOR IBD HAPLOTYPE WITH THE COLOUR METHOD

Fourteen families with three or more affected members with primary VUR were included in **study III**. All cases were confirmed by VCUG. All affected family members who provided DNA samples were genotyped with Affymetrix 250K SNP NspI arrays. SNP genotype data for one individual from each family were analysed using the Colour Method, previously described by Ostensson and Ohlsson et al [92, 130], in order to identify a unique disease haplotype IBD (identical by descent), which is inherited from a common ancestor. Corresponding genotype data generated by the Affymetrix 250K array in four healthy control individuals were entered one by one into the analysis. Given that an affected study subject and an unaffected control do not have the same disease phenotype, their DNA must by definition be different. The regions of the genome that were identical by state (IBS) in both VUR cases and controls were therefore excluded. After testing the controls one after the other, the shared region that remains in the VUR cases would be the region identical by descent.

In the analyses without controls, the most frequent haplotype was present in 13 of 14 VUR families (illustrated with the highest peak on Figure 22). When tested with four controls, excluding all haplotypes found in any of the controls, the number of families sharing a haplotype decreased from 13 to 7 and the number of haplotypes shared by ≥ 7 families decreased from 34 to only one (this locus is marked ** on Figure 22). As a result, we did not find a unique haplotype that was shared by most or all the families, as the regions shared by many families were also seen in one or more of the controls.

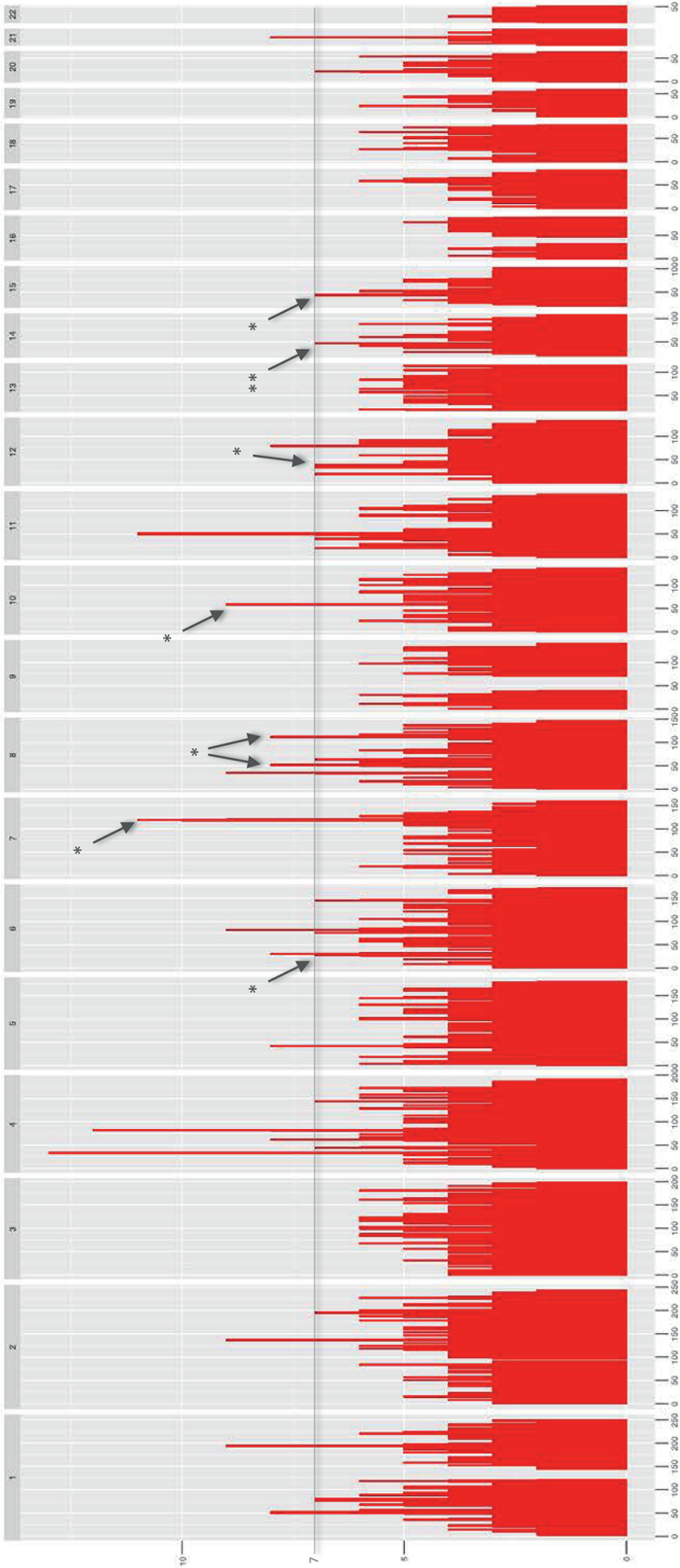


Figure 22. Genomic location of shared haplotype region in relation to the number of families sharing the locus. X axis genomic locations on the 22 chromosomes (Mb). Y axis number of families sharing the region. There are 34 peaks at and above the horizontal line that marks 7 families, denoting the haplotype regions shared by ≥ 7 families, before including the control. The peak marked ** is the only haplotype shared by ≥ 7 VUR families that remained after exclusion of haplotypes found in any of the four controls. The peaks marked * denote haplotypes that were excluded and the unmarked peaks show haplotypes that remained when controls were used to exclude common haplotypes in the general population (i.e. the haplotype was present in all four controls).

4.6 TESTING FOR DISEASE VARIANT EXCLUDING THE COMMON HAPLOTYPES

Our next strategy was to use the controls to rule out common haplotypes within the general population. We only excluded regions that were identical for VUR cases and all four controls (marked * on Figure 22). The haplotype region unique to the affected individuals and not present in all four controls might be specific to the disease. A total of 27 haplotype regions shared by seven to 13 families were identified with a total sum of 36.16 Mb, corresponding to 1.2% of the genome. Fourteen regions were shared by seven families, six regions by eight families and four regions by nine families, whereas only one haplotype was shared by 11, 12 and 13 families each. Six of our findings corresponded to the results of previously published studies (Figure 23).

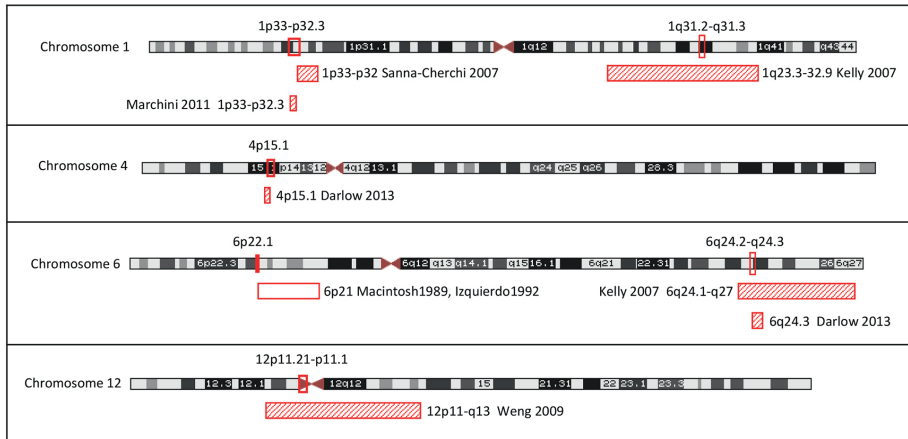


Figure 23. Haplotype regions shared by ≥ 7 families in relation to previous studies. Common haplotypes in the general population have been excluded by the use of controls. Results in the present study are indicated through red boxes on the chromosome ideogram with positions given above; Rectangles below the ideogram indicates results from previous studies; unfilled box, position uncertain; Sanna-Cherchi et al. [99], Marchini et al. [98], Kelly et al. [95], Darlow et al. [66], Mackintosh et al. [135], Izquierdo et al. [136], Weng et al. [20]

In the 27 candidate regions, we searched for possible genes, coding and non-coding, of interest for the embryological development of the kidney and urinary tract (UT). The haplotype shared by 13 of the 14 families at 4p15.1 is 1.84 Mb in size and includes 156 SNP markers. Darlow et al. also reported this position in a genome-wide case-control association study comprising 500 VUR patients [66]. The region did not contain any coding elements and the non-coding RNAs were not expressed in the post-developmental kidney or

urinary tract, but information about expression in the foetal kidney was not available. The largest shared haplotype region, 3.65 Mb, was identified at cytogenetic band 1p33-1p32.3 and was shared by eight families. This locus overlaps with findings of Sanna-Cherchi et al. and Marchini et al. in their genome-wide scans on familial renal hypodysplasia and VUR, respectively [98, 99]. This region contains the *ZFYVE9* gene, which is expressed in the metanephros and has an established function in the UB and MM (GUDMAP). This protein interacts directly with *SMAD2* and *SMAD3* in the transforming growth factor-beta (TGFB) signalling pathway, both important genes in the embryogenesis of the kidney and ureter [137].

The haplotype region at 4q21.21 was shared by 12 families. In this region, tentative candidate genes were identified, such as *BMP3* and *FGF5*, both expressed in the metanephros and both with known functions in UB and MM [138-140]. *BMP3* encodes a ligand to TGFB receptors and thus also participates in the TGF-Beta signalling pathway. Seven families shared a haplotype region at chromosome 12p12.3, which contains the gene *PLEKHA5*. Mutations in this gene have been found in patients with Kallmann syndrome – hypogonadotropic hypogonadism, occasionally twinned with kidney malformations, mainly renal agenesis.

Furthermore, there are also genes in the shared haplotype regions that are expressed in the metanephros, according to GUDMAP (<https://www.gudmap.org>), but without a specific known function during kidney development: *SEPP1* (5p13.1), *ZNRDT* and *TRIM26* (6p22.11), *AEBP2* (12p12.3), *PAX1* (20p11.23-p11.22) and *UBP16* (21q21.3).

Non-coding RNA was seen in almost all haplotype regions, often with expression in many of the displayed tissues (GTEx Portal, <https://www.gtexportal.org/home/>). The expression of a few lncRNA is exclusive or almost exclusive to the post-natal kidney and urinary tract (1p31.1, 2q21.2-3, 14q21.2-3). We searched for data regarding foetal expression in the kidney (ENCODE- HaploReg 4.1) but without success.

A large number of haplotype regions were shared by six families each. Four interesting genes, known to participate in UB and MM development, were identified in these regions. These genes are *WNT2* (7q31.2-3) [141], *TRPS* (8q23.3) [142, 143], *PTCH1* (9q22.32) and *MMPs* (11q22.3) [144].

4.7 COPY-NUMBER VARIANTS IN VUR FAMILIES

A large number of CNVs were detected among all analysed individuals. We searched for recurrent identical CNVs shared by affected individuals within the families, and identified such chromosomal imbalances in five of the families. A deletion at 5q31.1 in Family 32, containing the *FSTL4* gene, was the only CNV present in all affected relatives. The other four CNVs were only partially shared, meaning that some but not all affected family members were carriers of that specific CNV. In addition, eighteen CNVs were shared by ≥ 2 unrelated individuals among the families. Common CNVs in the general population were excluded by search in the available databases (DGV <http://dgv.tcag.ca/dgv/app/home>, DECIPHER <https://decipher.sanger.ac.uk/>).

4.8 GENE MUTATION SCREENING USING WHOLE-EXOME SEQUENCING

All families from **study III**, except for one, were further analysed in **study IV**. Thus, thirteen large families with hereditary VUR were included in this study. The most severely affected available family member, preferably with renal hypodysplasia in addition to high-grade VUR, was chosen for exome sequencing. In three families, WES was performed on one additional individual, the most distant affected relative of the proband observed in their respective pedigree (aunt, uncle and cousin). WES on these 16 individuals from 13 families resulted in a total of 117,216 high-quality variants in 18,722 genes in comparison to the GRCh37/hg19 human reference sequence (RefSeq, <https://www.ncbi.nlm.nih.gov/refseq>). After multistep variant filtering, as described in the Data processing section of the Methods chapter, we retained 40 candidate variants for pathogenicity in 32 genes previously associated with VUR or nephrogenesis. All our findings were heterozygous. Three of these genes (*FREM2*, *ROR2* and *FRAS1*) are associated with autosomal recessive diseases. Therefore, in the absence of additional variants that may result in compound heterozygosity, *FREM2* and *ROR2* were not examined further.

As part of our efforts to elucidate whether the members of the same family inherited the same VUR gene, WES was performed in a second family member in three families. In one family (Fam. 32) with severe VUR and renal hypodysplasia, two rare DNA variants in possible causal genes (*LIFR*, *CLDN3*) were shared between the analysed patients and, in the second family

(Fam. 17) a novel *KIF26B* variant was the common detected variant (Table 5). In the third family (Fam. 82), we could not detect any common variant in kidney-associated genes, in spite of their astonishingly similar phenotype.

Segregation analysis was performed for all candidate variants (except the genes with autosomal recessive inheritance mentioned previously) in all relatives with available DNA samples (Table 5). Sanger sequencing validated three possibly disease-causing heterozygous dominant DNA variants in three different nephrogenesis-related genes (*KIF26B*, *LAMC1* and *LIFR*) in three of our 13 studied families with hereditary VUR (Table 6). The variants in these genes showed segregation consistent with dominant autosomal inheritance as they were seen only in affected individuals. Thirty-four of 37 analysed variants did not segregate with the phenotype in all the families it concerned.

Table 5. Results of Sanger sequencing used for segregation analysis in 13 families with hereditary VUR

Fam.	Genes	Protein change	Investigated (WES)		Investigated (Sanger sequencing)			
9			221 F		220 M	204 F	218^cM	219 F
	<i>SALL2</i>	p.P168L	+		+	-	+	-
	<i>SIM1</i>	p.G254K	+		+	-	-	-
17			351 F	369 F	347 F	355^pF	368^pF	367 M
	<i>KIF26</i>	p.S123L	+	+	+	+	+	-
	<i>UPK2</i>	SS	+	-	+	+	-	-
19			357 M		329 F	233 F	348 M	
	<i>SALL1</i>	p.G1265E	+			-	+	
	<i>CHD7</i>	p.L2984F	+		+	+	-	
	<i>LIFR</i>	p.D816G	+		-	+	-	
30			250 M		251 F	253 F	252^cF	248 M
	<i>MDM4</i>	p.K374Q	+		-	+	+	-
	<i>CLDN3</i>	p.P134L	+		+	+	-	Hom
	<i>SALL2</i>	p.T47N	+		-	+	-	+
32			236 F	656 F	395 M	235^pF	234 M	
	<i>LIFR</i>	p.V487A	+	+	+	+	-	
	<i>CLDN</i>	p.P134L	+	+	+	+	-	
	<i>GLI3</i>	p.R114K	-	+		-	-	
	<i>CHD7</i>	p.L2984F	-	+	-	-	-	
46			364 M		362 M	363 M	366^cF	365 M
	<i>MMP9</i>	p.R24C	+		+	+	Hom	-
	<i>SALL2</i>	p.P168L	+		-	+	-	+
	<i>TGFBR</i>	p.F435S	+		-	-	-	+

Table 5. Continued

Fam.	Genes	Protein change	Investigated (WES)	Investigated (Sanger sequencing)				
49	<i>GATA3</i>	p.P154S	+	-	+	-		
	<i>PYGO1</i>	p.N250I	+	-	+	-		
				391 M	392 M	390 ^c F	389 M	
76	<i>ROBO2</i>	p.I598T	+	-	+	-	+	
	<i>FRAS1</i>	p.M2129V	+	-	-	-	+	
	<i>LAMC</i>	p.K646fs*3	+	+	+	+	-	
	<i>GREB</i>	p.E93K	+	+	+	+	-	
				650 F	648 M	660 F	644 ^c F	653 M
77	<i>BMP7</i>	p.N321S	+	-	-			
	<i>WNT3A</i>	p.A172T	+		-			
	<i>POSTN</i>	p.Q71K	+		-			
	<i>KIF26</i>	p.S1218F	+	-	+	-		
				645 M	649 F	690 F	651 M	
79	<i>FRAS1</i>	p.Y1758C	+	+	-	-		
	<i>NRTN</i>	p.V125L	+					
	<i>TGFBR</i>	SS	+	+	?	-		
				715 M	647 F	693 M	659 F	
80	<i>SLIT3</i>	p.S629N	+	-	+	-	+	-
				682 M	652 M	658 M	666 M	695 F 711 M
82	<i>UPK3A</i>	p.W182*	+	-	-	+		
	<i>CHD1L</i>	p.G700R	+	-				
	<i>MMP9</i>	p.R24C	-	+	-	-		
	<i>TGFBR</i>	p.H155R	-	+	-	-		
				655 M	705 F	710 F	691 M	
83	<i>DSTYK</i>	SS	+	+	-			
	<i>MDM4</i>	p.K374Q	+	-	-			
	<i>GREB1</i>	p.E93K	+	-	-			
				698 M	670 F	657 F		

bold digit, affected family members; **bold gene symbol**, the gene variant segregates with the phenotype in the family; F, female; Fam., family; hom, homozygous variant; M, male; SS, splice site loss; WES, whole-exome sequencing; +, variant present in heterozygous form; -, variant missing; ?, Sanger sequencing failed, chromatogram not assessable

^c Probable carrier according to the pedigree

^p Probable VUR, strong history of VUR but no available radiological investigations

The variants that did not show consistent co-occurrence with a VUR phenotype included predicted deleterious, missense variants in the known VUR genes, *SALL1* (p.G1265E) and *ROBO2* (p.I598T), which were inherited from healthy fathers in the respective family. A truncating *UPK3A* mutation (p.W182*) fell out on the same exclusion criteria. Splice site loss variant in *UPK2* (c.418+1G>A) segregated in the nuclear but not the extended family. Splice site loss variant in *DSTYK* (c.654+1G>A) segregated in the two youngest generations but not the oldest known affected individual in the family. The *GREB1L* variant allele (p.E93K) that segregated with VUR in Family 76 was also detected in the youngest individual in Family 83, but was not seen in other affected relatives. The deleterious *CLDN3* variant allele (p.P134L) fully segregated in Family 32, but the same variant was also detected in Family 30, where the unaffected father was homozygous for this variant. One *MMP9* variant was detected in affected individuals from two different families (Family 46 and 82), although it only segregated in one of the two families, where one unaffected family member was homozygous for the variant (individual 366 in Family 46). A variant in *CHD7* was detected in two families, but it only segregated in Family 19.

Table 6. Three possibly pathogenic variants identified in nephrogenesis-related genes in three families with hereditary VUR

Gene	Family-individual	Renal phenotype	Extrarenal phenotype	Nucleotide change ^a	Amino-acid change	Impact	SIFT ^b	PP2 ^c	CADD ^d	MAF % SweGen ^e
<i>KIF26B</i>	17-351 17-369	B VUR, U FRD U VUR	Scoliosis, MI, JIA, Marfan?	c.368C>T	p.S123L	M	D	0.952	26.3	Novel
<i>LAMC1</i>	76-650	B VUR, U RHD		c.1935delG	p.K646fs*3	F	NA	NA	NA	Novel
<i>LIFR</i>	32-656 32-236	B VUR, U RHD, SubnRF U VUR, U RHD		c.1460T>C	p.V487A	M	A	0	<10	Novel

A, activating; B, bilateral; D, damaging; F, frameshift; FRD, focal renal damage; JIA, juvenile idiopathic arthritis; M, missense; MI, mitral insufficiency; NA, no prediction available; RHD, renal hypodysplasia; SubnRF, subnormal total renal function; T, tolerated; U, unilateral; VUR, vesicoureteral reflux. Nucleotide change numbering refers to the cDNA position of the following transcripts: NM_018012.4 (*KIF26B*), NM_002293.3 (*LAMC1*), NM_002310.5 (*LIFR*).

^a All mutations are heterozygous.

^b Sorting Intolerant From Tolerant (<http://sift.bii.a-star.edu.sg>)

^c PolyPhen-2 prediction score ranges from 0 (=benign) to 1 (=probably damaging) (<http://genetics.bwh.harvard.edu/pph2/>)

^d Combined annotation-dependent depletion score (<http://cadd.gs.washington.edu>)

^e Minor allele frequency of the variant in SweGen dataset if present, otherwise in 1,000 Genomes and gnomAD databases

5 DISCUSSION

The genetic background of primary VUR has been investigated for the last four decades, however, there are many questions left to answer. Our genetic epidemiological studies were intended to fill in some missing pieces in this complex puzzle. Firstly, we explored whether the phenotype of the familial reflux differs from the non-familial cases. Secondly, we searched for chromosomal regions associated with hereditary VUR. Thirdly, we searched for possibly pathogenic variants in genes previously associated with VUR or genes involved in urinary tract morphogenesis. In the last two studies (**study III** and **IV**) we were focusing on the severe end of the VUR phenotype to try to define the causal pathway.

5.1 ON HERITABILITY AND PHENOTYPE

Twenty-two percent of the VUR cases treated at our hospital between 1990-2004 from among those who responded to our inquiry were classified as hereditary (**study I**). This is a low estimate, as relatives of patients with VUR were not routinely screened in Sweden. We have also excluded 11 % of the possible familial cases (those who reported family members with vague kidney problems) due to uncertainty about the phenotype of the relatives. There was strong overrepresentation of maternal transmission of VUR. In 20 families, VUR was inherited from an affected parent, and in 15 of these 20 cases the inheritance was maternal. In addition, when all affected family members (except siblings) were taken into account, maternal inheritance was seen in 25 families and paternal inheritance in 15 families. This overrepresentation of maternal inheritance has been noted in other studies but it has not been attributed any particular significance [16, 100, 145]. It can of course be associated with the fact that reflux is more often observed in females, 58% in the present study. Another possible explanation could be selection bias, e.g. the mothers of the index patients rather than the fathers were interviewed on telephone, thereby reflecting the health problems of the maternal relatives to a higher degree. Naseri et al. suggested linkage to the X-chromosome in a report describing a large family [21], but no other genetic study has confirmed the involvement of the X-chromosome in transmission of VUR. Imprinting mechanism is yet another possible explanation.

When comparing familial and non-familial VUR, there were no significant differences in gender, presenting symptoms, severity of VUR, frequency of breakthrough infections, renal damage or total renal function. However, differences were noted within the hereditary group between the index group

and the group with relatives. As compared to index patients, the relatives were more often diagnosed through screening (27% vs 3%) and had significantly lower grades of VUR. Wan et al. has also shown similar differences in VUR grades in screened siblings of children with reflux [9]. Yet, the prevalence of renal damage was similar in all our study groups. The lower-grade reflux among relatives is not explained by the higher screening frequency in siblings, because it applied to the parents, cousins and aunts/uncles too.

We have found an overall high frequency of renal damage (64-74%), which is very similar to the figures shown by the Swedish reflux trial [146] but much higher than the 30% presented by Menezes et al., comparing index patients with their refluxing siblings [147]. It seems that fewer children with low-grade reflux are diagnosed and treated in Scandinavia. Jerkins et al. has shown that the siblings of index patients with established renal damage had a much higher percentage of reflux and also renal damage than the siblings of index patients without renal damage [10]. This can be interpreted as a sign that the heritability of VUR with co-occurrent renal damage is higher than that of VUR only. This might explain why we found no statistically significant difference at all between the index group and the group of relatives in the prevalence of kidney damage although the latter had lower-grade reflux.

5.2 ON SHARED CHROMOSOMAL REGIONS

There are numerous association and linkage studies searching for the ultimate region that explains VUR inheritability. However, most studies report different results, with few peak overlaps [20, 66, 95, 98]. Mutation analyses of hundreds of candidate genes have revealed pathogenic gene variants in fewer than 10% of VUR patients [102, 103, 107]. Therefore, we used a different approach in **study III**. In this study performed on 14 families with three or more individuals with primary VUR from the south-western part of Sweden we focused on individuals with high-grade VUR and renal hypodysplasia. We used the in-house Colour Method, a variant of linkage analysis for genome-wide association (GWA) data designed to identify a dominantly inherited disease gene haplotype assumed to come from a common ancestor [92, 130]. We found no haplotype region shared by most affected families with hereditary vesicoureteral reflux and renal hypodysplasia. These results suggest that all of the 14 families from the south-western part of Sweden with ≥ 3 affected members have no common ancestral founder mutation explaining the inheritance, which is in line with

the study results in recent decades in other populations [5, 148]. However, subsets of the families share regions as shown in the results section and in Figure 22.

By now knowing that VUR is a genetically heterogeneous disease, we present haplotype regions shared by seven or more families, which is warranted since at least six of our findings (on chromosomes 1, 4, 6 and 12) are in agreement with loci identified in other studies (Figure 23) [20, 66, 95, 98, 99, 135, 136]. We investigated these regions for genes (coding and non-coding) involved in nephrogenesis by searching in databases. In the regions that were overlapping with previously published findings, *ZFYVE9* (1p32.3) stands out as a gene of interest. This gene is expressed in the metanephros and plays a significant role in TGF-mediated signalling, which is one of the many pathways involved in the UB outgrowth and UB-MM interaction [137]. In our study, the haplotype region at 4q21.21 was shared by 12 families and contains the *BMP3* and *FGF5* genes. Both of them are expressed and have an established role in the metanephros, *BMP3* through the above-mentioned TGF-beta signalling pathway and *FGF5* via RET signalling [139, 140]. On chromosome 12 we have identified the *PLEKHA5* gene (12p12.3), which is involved in Kallmann syndrome including kidney malformations, mainly agenesis [149].

In addition to our findings in the coding regions, non-coding RNAs were seen in almost all shared haplotype regions, of which three were almost exclusive to the post-natal kidney and urinary tract (1p31.1, 2q21.2-3, 14q21.2-3). More research is needed to understand the regulatory function of these elements on the genes involved during embryogenesis.

5.3 ON SHARED COPY-NUMBER VARIANTS

The CNVs detected in these 14 families with hereditary VUR were examined for inheritance pattern. A deletion was seen at 5q31.1 in all three affected members of Family 32, which was the only CNV segregating in the entire family in our material. This segment of chromosomal imbalance includes a part of the *FSTL4* gene with no known function in the UB and MM, but with detected expression later, during renal tubuli development. This CNV position has not been reported in previous publications regarding CNV in CAKUT [112-117]. Although the majority of the detected CNVs did not segregate fully with disease in the families, we searched for possibly pathogenic de novo CNVs by comparing our data with previously published findings. Three previously presented CNVs corresponded to our data: 7p22.1 and 12q24 detected by Sanna-Cherchi et al [114] and 8q24.13 found by

Caruana et al [116]. The common abnormality in the patients showing these CNVs in the studies by Sanna-Cherchi and Caruana was renal hypodysplasia. The CNV at 7p22.1 is also connected to the 7p interstitial duplication syndrome that includes developmental delay and intellectual disability. However, individual 652 in Family 80 displaying this duplication has VUR without kidney damage and no extrarenal phenotype. The CNV at 12q2 is described as large pathogenic de novo duplication connected to congenital kidney malformations [114].

5.4 ON GENE MUTATIONS

All genes that regulate UB formation and urinary tract development are highly relevant candidates as causal factors in VUR pathogenesis. Studies on mice and mutation screening in two large cohorts of families with VUR have revealed *ROBO2* and its ligand *SLIT2* to be two of those candidates [19, 53, 86]. Our investigation of 54 unrelated individuals with familial VUR, by sequencing all exons of the *ROBO2* and *SLIT2* genes, has identified gene variants in both (**study II**). Six intron–exon boundary sequence variants were found in the *ROBO2* gene, but none of them were predicted to influence gene splicing. One sequence variant not described before (IVS13+43C>T) was found in an affected sib-pair and in their apparently healthy mother. The other variants were also found in controls and four of them did not co-segregate with VUR. These results are in line with findings of Conte et al. who detected in a genome-wide analysis linkage to a large area on chromosome 3 including locus 3p12.3 but found no mutation in *ROBO2*, which maps there [100].

One non-synonymous missense variant was found in *SLIT2* (p.Ala1418Val), causing alteration of the secondary structure that may affect the SLIT-ROBO interaction. However, based on the analysis of multiple protein alignment and three-dimensional structures by PolyPhen program, the impact of this variant was predicted as benign. In addition, the variant did not co-segregate with VUR in the two families in which it was detected. The intronic *SLIT2* variants could not be regarded as disease causing either, as they either were common SNPs, new variants that were present in controls too or they did not co-segregate with VUR.

Thus, according to our study, *ROBO2* and *SLIT2* gene variants are rare causes of VUR.

In **study IV** we have expanded the search from mutation screening of single genes to whole-exome sequencing. In the 13 large families with primary

VUR we identified 40 heterozygous variants in 32 genes, all being promising candidates of pathogenicity due to evidence on their involvement in the renal and urinary tract development and/or expression in the UB – MM. However, segregation analysis could only validate three of these gene variants (*KIF26B*, p.S123L; *LAMC1*, p.K646fs*3; *LIFR*, p.V487A) in three (23%) of the families. Recent studies using WES detected significant mutations in a much smaller fraction of cases (3.2%, 4.8%, 6.3% and 17.6% in the studies by Nicolaou et al., Bekheirnia et al., Hwang et al. and Heidet et al. respectively) [24, 102, 103, 107]. Even though their studies were large, most patients were non-hereditary cases with primary focus on CAKUT and not only the VUR/renal hypodysplasia complex. Our families all had three or more individuals with the VUR, which has higher heritability than CAKUT in general.

LAMC1 (Laminin Subunit Gamma 1) has not been reported earlier as possible candidate gene in patients with VUR or other CAKUT, however previous studies have showed that laminins have an important role in kidney development. In one of the first published studies on this topic, no effect was observed in mice with a heterozygous *Lamc1* mutation on the kidney phenotype, while all homozygous mice died lacking kidneys (one or most commonly both) and had ectopic ureters [150]. In a more recent study, Yang et al. have proven that *Lamc1* regulates branching morphogenesis: inactivation of *Lamc1* in the UB resulted in either small kidneys or a total lack of both kidneys and ureters [151]. There is probably a laminin concentration threshold above which UB penetration is enabled, determining the dual fate of kidney agenesis or hypodysplasia.

KIF26B regulates the adhesion of mesenchymal cells in contact with ureteric buds and it is thus essential for the UB invasion of MM and UB branching [56]. The novel variant p.S123L segregated with the phenotype in family 17, but the damaging, missense, rare variant p.S1218F in family 77 did not co-segregate.

We also identified a heterozygous novel activating missense *LIFR* variant (p.V487A) in a family with high-grade VUR and unilateral renal hypodysplasia (Family 32). Another missense variant (p.D816G) in *LIFR* was seen in an additional family (Family 19), but that variant did not co-segregate with the phenotype and was predicted as tolerated. Receptor *LIFR* in the MM promotes MET when it binds to its ligand, LIF, secreted by the UB [152, 153]. Kosfeld et al. recently demonstrated heterozygous *LIFR* mutations in 3.3% of CAKUT patients [106].

Our findings also excluded some variants that were previously reported as disease causing, as they did not segregate with the phenotype within the families. For example, *ROBO2* variant p.I598T was inherited by two of the three affected children from a healthy father and not from the mother with kidney problems. The truncating *UPK3A* variant (p.W182*, stop gain) was also inherited from a healthy father and not the affected mother. Jiang et al. concluded that major uroplakin mutations (such as truncation) are not found in patients with VUR, as they probably cause severe renal abnormalities that are not compatible with life [154]. We can now refute this claim, as it seems that one healthy allele of the uroplakin gene may be functionally adequate in humans.

The diversity of the findings in different studies supports the hypothesis that primary VUR genetically is a very heterogeneous disease. Reduced penetrance and variable expressivity due to genetic, epigenetic and environmental factors make the genetic study of familial VUR challenging.

5.5 METHODOLOGICAL CONSIDERATIONS

Spontaneous regression of dilating reflux is common [32, 155]. An overwhelming majority (73%) of siblings with VUR are asymptomatic including 60% of siblings renal damage [10]. VUR being a non-visible malformation in asymptomatic individuals and with a possible natural course of spontaneous resolution during childhood makes reflux a difficult abnormality to study in terms of heredity from one generation to another.

One of the methodological limitations of **study I** was the lack of VCU in almost half of the relatives with VUR, owing to the fact that this radiological examination was not performed prior to the 1960s. In addition, relatives of patients with VUR were not routinely screened in Sweden, which could be responsible for the lower prevalence of familial VUR in our material. Relatives with VUR could be underreported in both the hereditary and control groups. The missed cases probably presented with less symptoms, given the availability of the Swedish health care. We agree that VCU is the gold standard method to detect VUR. However, it is a highly invasive investigation, which limits its use in asymptomatic relatives.

Although most people were positive to the study when they received the invitation to participate, we had recruitment problems when they were asked to donate blood samples for analysis in **study II**. Using buccal smear kits sent home by post minimized the inconvenience for children and their families, and increased the willingness to participate. However, in clinical setting this

method yielded DNA of suboptimal quantity and quality, insufficient for whole-exome sequencing.

Small sample sizes may yield insufficient power to detect a difference, even if it is there. On one hand, our sample size limited the potential of our study in finding novel pathogenic variants. On the other hand, our study populations consisted of familial cases, and this design was aimed at increasing the chance of finding genetic causes of VUR.

Some previous studies show conflicting findings, and this is partly a result of study design differences. Another explanation lies in the differences between study populations. However, the problem of genetic association studies with non-replication may also be caused by failure to fully account for the consequences of making multiple comparisons [134]. Reports of positive associations may reflect type I error, i.e. false positive results, or confounding for alternative associations. Failure to identify an association to markers on a gene does not rule out other sites within the gene or regulatory regions of the gene [5]. Our material in **study II-IV** is too small for significance testing. Our aim was to identify interesting loci and gene variants, which need to be further validated and fine-mapped.

6 CONCLUSION

Vesicoureteral reflux is a highly heterogeneous malformation. We have found a possible increased inheritance of VUR via the mother, a result that is worthy of further examination both regarding inheritance and genetics. No difference was found between the phenotype of familial and non-familial reflux, whereas there was a difference between the index and relative groups, such as lower grade of reflux in the latter group.

We did not identify a unique haplotype for all 14 large families with VUR in the south-western part of Sweden containing a common ancestral founder mutation. However, we have identified common haplotype regions in subsets of families on chromosomes 1p, 4q and 12p, containing possible candidate genes with known functions in the embryogenesis of the kidney and UT. For determining the effect of non-coding elements of these shared regions on the genes involved in embryogenesis, further research is needed.

On analysing the detected CNVs, all affected individuals in one of the families inherited a deletion at 5q31.1. This chromosomal imbalance disrupts the *FSTL4* gene, which is probably involved in renal tubuli development.

Variants in *ROBO2* and *SLIT2* genes are rare causes of VUR in humans. We have identified possibly pathogenic, heterozygous novel mutations in *KIF26B*, *LAMC1* and *LIFR*, genes associated with renal development, in 23% of families with hereditary VUR. Next-generation sequencing technology is a useful tool to understand the pathogenesis and improve the molecular diagnostics of this complex disorder.

7 CLINICAL IMPLICATIONS AND FUTURE PERSPECTIVES

Clinical implications

We have still not reached our ultimate goal for the future, i.e. to recommend genetic analysis of blood or saliva as screening method and prognostic tool for patients with VUR hypodysplasia complex. As far as we know, some family members share a gene variant that may cause or contribute to their hereditary disorder, while in other families different genes are involved. In addition, there are many unresolved families with yet no detected mutation. Unless there is an agreement on the most common gene variants in VUR and hypodysplasia, and this list of genes covers the vast majority of cases, routine diagnostics by mutation analysis is not yet possible.

After decades of research in this field, we are just at the beginning of understanding the genetic background of VUR. Genetic techniques are improving and expanding very quickly, consequently our knowledge on complex diseases will also extend. The major advantages of these studies are seen on another level. We highly value the cooperation developed between the Department of Clinical Genetics at the University of Gothenburg and the Department of Paediatric Surgery at Queen Silvia Children's Hospital. This means that important genetic expertise is now available for research at our unit, where many congenital anomalies are taken care of.

Future perspectives

Another important objective was to use mutation analysis for identification of patients at risk, by distinguishing severe cases that require prompt treatment and frequent follow-up from those with benign course that will resolve spontaneously. One possible further study to address this question is to evaluate the differences between genders.

The incidence and phenotype for VUR is different in men and women. Reflux is more common in females, which was also seen in the present study. In boys, reflux is usually diagnosed early during infancy, it is of high-grade and often associated with congenital generalised renal damage, hypodysplasia. In girls, on the other hand, the peak for diagnosis is between 1 and 2 years of age, the reflux is of moderate or low grade and when renal damage is present it is often focal, thus acquired. Therefore, a gender-specific genetic difference may well be possible, which was actually recently shown

in a study of sib-pairs with familial VUR: in males significant peaks were found on chromosome 1 and 5 and in females on 3, 13 and 15 [98].

The question of the difference in genetic background between genders cannot be answered in the present studies due to the limited number of cases. As future study, a genetic analysis of a large cohort of both hereditary and non-hereditary VUR cases with stratification for gender is suggested. To gather an adequate number of patients, cooperation with other centers is a necessary prerequisite. Using whole-genome sequencing, mutation analysis can be performed in both coding and non-coding areas, while it would also provide an opportunity to explore copy-number variations.

8 WEB RESOURCES

1000 Genomes, <http://www.internationalgenome.org/>

CADD, Combined Annotation Dependent Depletion,
<http://cadd.gs.washington.edu/>

Clinical Genome Resource, <https://www.clinicalgenome.org/>

CNV CHOP, The Copy Number Variation project at the Children's Hospital of Philadelphia, <http://cnv.chop.edu/>

dbGAP, International Standards for Cytogenomic Arrays (ISCA) Consortium, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000205.v6.p2

DECIPHER, <https://decipher.sanger.ac.uk/>

DGV, Database of Genomic Variants, <http://dgv.tcag.ca/dgv/app/home>

Ensembl Genome Browser, <http://www.ensembl.org/index.html>

GeneCards, The Human Gene Database, <https://www.genecards.org/>

GUDMAP, Genitourinary Database Molecular Anatomy Project,
<http://www.gudmap.org/>

GTEEx Portal, Genotype-Tissue Expression project,
<https://www.gtexportal.org/home/>

KEGG Pathway Database, <https://www.genome.jp/kegg/pathway.html>

National Center for Biotechnology Information,
<http://www.ncbi.nlm.nih.gov/>

OMIM, Online Mendelian Inheritance in Man, <http://www.omim.org/>

PolyPhen-2, Polymorphism Phenotyping v2,
<http://genetics.bwh.harvard.edu/pph2/>

PubMatrix, <https://pubmatrix.irp.nia.nih.gov/cgi-bin/index.pl>

RVIS, Residual Variation Intolerance Score, <http://genic-intolerance.org/>

SIFT, Sorting Intolerant From Tolerant, <http://sift.bii.a-star.edu.sg/>

SweGen, <https://swefreq.nbis.se/dataset/SweGen/browser>

UCSC Genome Browser Home, <http://www.genome.ucsc.edu>

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to those who have, in one way or another, contributed to this work:

All **children and their families** for their willingness to participate, which made this project possible.

Ulla Sillén, my supervisor, for her tremendous support, excellent guidance in the world of science, for sharing her great experience and for the many hours we have spent together trying to decipher the secret code of VUR.

Agneta Nordenskjöld, my co-supervisor, for leading me into the world of genomics in her lab at CMM 02 at Karolinska Institutet; for her big heart, positive spirit and great hospitality, for her quick assistance whenever I needed.

To our beacon in the dark waters of genomics, **Susanne Fransson**, my co-supervisor, for all input and help, for constructive comments on the manuscripts and on this thesis, for her excellent and enthusiastic teaching.

Sofia Sjöström, my co-supervisor for all great input, the fine words of encouragement and constructive comments.

Tommy Martinsson, my co-supervisor, for support and guidance in genetics, for great scientific discussions and for always being so optimistic and encouraging.

Kate Abrahamsson, head of the Department of Pediatric Surgery, for always believing in me and supporting me.

Alice Andersson, **Eva Johansson**, **Monica Doroszkiewicz** and **Tina Linnér** for their invaluable assistance in collecting patients, samples and data. **Maria Lönn**, **Ing-Marie Åkesson** and **Jenny Bergström** for more practical help.

Anna Djos, for all her great assistance in the lab work, for her friendship and the pleasant time spent together in the lab. **Rose-Marie Sjöberg**, for laboratory assistance, for her cheerfulness and enthusiasm. **Christina Nyström** and **Fredrik Lundberg** at CMM 02 for teaching me the basic molecular biology laboratory techniques.

Sverker Hansson, Svante Swerkeresson and Per Brandström for generous support and for sharing their great knowledge on paediatric nephrology. **Georgios Vasilogiannakis** and **Anna Lindholm**, for help with recruiting families. **Shulu Zu**, first author of paper II, for fruitful cooperation.

Malin Östenson, for skilful assistance with bioinformatics for paper III, for making great figures and graphics. **Matthias Molin** at Statistiska konsultgruppen for help with statistical calculation in paper I. **Staffan Nilsson** for all his good advice on statistical methods in genetics.

My old friend **Nonó**, alias **Noémi László**, for correcting the linguistic aspects of this thesis, even though receiving some of it during her vacation! **Jeanette Kliger** and **Linda Schenck**, for linguistic revision of the papers.

Soffi Pettersson, infomaster SU/Utvecklingsenheten, for professional help with graphics and pictures.

Lennart Sjöholm, my supervisor and mentor during my clinical studies, for sharing his knowledge and vast experience, not only at our hospital but also in Jinka, Ethiopia.

All **colleagues and friends** at the Department of Pediatric Surgery, the wards and the operating theatre at Queen Silvia Children's Hospital for making everyday work enjoyable. Special thanks to the "gastro team" – **Vladi, Linus, Helena, Anders, Sigge, Cathrine** and **Matilda** – for the helpful atmosphere, for the extra workload they carried while I was working on this thesis.

My parents, **Mária** and **László**, for endless love and practical support, for making our life so much easier. My brother **Lóránt** and sister-in-law **Bernadett** for being there whenever needed and sharing all the happy family events. Thanks Betti for your help with the illustrations! **Eszter Kónya**, my very talented niece, for the cover illustration.

My beloved **Olof**, for the life we share, for always being so calm and supporting. **Vera** and **Artur**, for their unconditioned love, for their patience and understanding with me when my work is taking over. You are the sunshine of my life, you make me happy!

This study was financially supported by government grants through the Agreement on Medical Education and Research.

REFERENCES

1. Murer L, Benetti E, Artifoni L (2007) Embryology and genetics of primary vesico-ureteric reflux and associated renal dysplasia. *Pediatr Nephrol* 22:788-797.
2. Gordon I, Peters AM, Morony S (1990) Indirect radionuclide cystography: a sensitive technique for the detection of vesico-ureteral reflux. *Pediatr Nephrol* 4:604-606.
3. Lebowitz RL, Olbing H, Parkkulainen KV, et al. (1985) International system of radiographic grading of vesicoureteric reflux. *International Reflux Study in Children. Pediatr Radiol* 15:105-109.
4. Sargent MA (2000) What is the normal prevalence of vesicoureteral reflux? *Pediatr Radiol* 30:587-593.
5. Williams G, Fletcher JT, Alexander SI, et al. (2008) Vesicoureteral reflux. *J Am Soc Nephrol* 19:847-862.
6. Hellstrom A, Hanson E, Hansson S, et al. (1991) Association between urinary symptoms at 7 years old and previous urinary tract infection. *Arch Dis Child* 66:232-234.
7. Hansson S, Bollgren I, Esbjorner E, et al. (1999) Urinary tract infections in children below two years of age: a quality assurance project in Sweden. The Swedish Pediatric Nephrology Association. *Acta Paediatr* 88:270-274.
8. Sjostrom S, Sillen U, Jodal U, et al. (2010) Predictive factors for resolution of congenital high grade vesicoureteral reflux in infants: results of univariate and multivariate analyses. *J Urol* 183:1177-1184.
9. Wan J, Greenfield SP, Ng M, et al. (1996) Sibling reflux: a dual center retrospective study. *J Urol* 156:677-679.
10. Jerkins GR, Noe HN (1982) Familial vesicoureteral reflux: a prospective study. *J Urol* 128:774-778.
11. Noe HN (1992) The long-term results of prospective sibling reflux screening. *J Urol* 148:1739-1742.
12. Parekh DJ, Pope JCT, Adams MC, et al. (2002) Outcome of sibling vesicoureteral reflux. *J Urol* 167:283-284.
13. Kaefer M, Curran M, Treves ST, et al. (2000) Sibling vesicoureteral reflux in multiple gestation births. *Pediatrics* 105:800-804.
14. Noe HN, Wyatt RJ, Peeden JN, Jr., et al. (1992) The transmission of vesicoureteral reflux from parent to child. *J Urol* 148:1869-1871.
15. Chapman CJ, Bailey RR, Janus ED, et al. (1985) Vesicoureteric reflux: segregation analysis. *Am J Med Genet* 20:577-584.
16. Feather SA, Malcolm S, Woolf AS, et al. (2000) Primary, nonsyndromic vesicoureteric reflux and its nephropathy is genetically heterogeneous, with a locus on chromosome 1. *Am J Hum Genet* 66:1420-1425.
17. Eccles MR, Jacobs GH (2000) The genetics of primary vesico-ureteric reflux. *Ann Acad Med Singapore* 29:337-345.

18. Sanna-Cherchi S, Reese A, Hensle T, et al. (2005) Familial vesicoureteral reflux: testing replication of linkage in seven new multigenerational kindreds. *J Am Soc Nephrol* 16:1781-1787.
19. Lu W, van Erde AM, Fan X, et al. (2007) Disruption of ROBO2 is associated with urinary tract anomalies and confers risk of vesicoureteral reflux. *Am J Hum Genet* 80:616-632.
20. Weng PL, Sanna-Cherchi S, Hensle T, et al. (2009) A recessive gene for primary vesicoureteral reflux maps to chromosome 12p11-q13. *J Am Soc Nephrol* 20:1633-1640.
21. Naseri M, Ghiggeri GM, Caridi G, et al. (2010) Five cases of severe vesicoureteric reflux in a family with an X-linked compatible trait. *Pediatr Nephrol* 25:349-352.
22. de Vargas A, Evans K, Ransley P, et al. (1978) A family study of vesicoureteric reflux. *J Med Genet* 15:85-96.
23. Uy N, Reidy K (2016) Developmental Genetics and Congenital Anomalies of the Kidney and Urinary Tract. *J Pediatr Genet* 5:51-60.
24. Heidet L, Moriniere V, Henry C, et al. (2017) Targeted Exome Sequencing Identifies PBX1 as Involved in Monogenic Congenital Anomalies of the Kidney and Urinary Tract. *J Am Soc Nephrol* 28:2901-2914.
25. Little M, Georgas K, Pennisi D, et al. (2010) Kidney development: two tales of tubulogenesis. *Curr Top Dev Biol* 90:193-229.
26. Mackie GG, Stephens FD (1975) Duplex kidneys: a correlation of renal dysplasia with position of the ureteral orifice. *J Urol* 114:274-280.
27. Uetani N, Bouchard M (2009) Plumbing in the embryo: developmental defects of the urinary tracts. *Clin Genet* 75:307-317.
28. Batourina E, Tsai S, Lambert S, et al. (2005) Apoptosis induced by vitamin A signaling is crucial for connecting the ureters to the bladder. *Nat Genet* 37:1082-1089.
29. Ichikawa I, Kuwayama F, Pope JCT, et al. (2002) Paradigm shift from classic anatomic theories to contemporary cell biological views of CAKUT. *Kidney Int* 61:889-898.
30. Radmayr C, Schwentner C, Lunacek A, et al. (2009) Embryology and anatomy of the vesicoureteric junction with special reference to the etiology of vesicoureteral reflux. *Ther Adv Urol* 1:243-250.
31. Shafik A (1997) Study of the effect of external urethral sphincter contraction on the mechanical activity of the ureterovesical junction and urinary bladder: recognition of the sphinctero-ureterovesical reflex. *Urology* 50:949-952.
32. Sjostrom S, Sillen U, Bachelard M, et al. (2004) Spontaneous resolution of high grade infantile vesicoureteral reflux. *J Urol* 172:694-698; discussion 699.
33. Peters C, Rushton HG (2010) Vesicoureteral reflux associated renal damage: congenital reflux nephropathy and acquired renal scarring. *J Urol* 184:265-273.
34. Smellie JM, Prescod NP, Shaw PJ, et al. (1998) Childhood reflux and urinary infection: a follow-up of 10-41 years in 226 adults. *Pediatr Nephrol* 12:727-736.

35. Martinell J, Jodal U, Lidin-Janson G (1990) Pregnancies in women with and without renal scarring after urinary infections in childhood. *Bmj* 300:840-844.
36. Martinell J, Lidin-Janson G, Jagenburg R, et al. (1996) Girls prone to urinary infections followed into adulthood. Indices of renal disease. *Pediatr Nephrol* 10:139-142.
37. Wennerstrom M, Hansson S, Hedner T, et al. (2000) Ambulatory blood pressure 16-26 years after the first urinary tract infection in childhood. *J Hypertens* 18:485-491.
38. Wennerstrom M, Hansson S, Jodal U, et al. (2000) Renal function 16 to 26 years after the first urinary tract infection in childhood. *Arch Pediatr Adolesc Med* 154:339-345.
39. Jodal U, Smellie JM, Lax H, et al. (2006) Ten-year results of randomized treatment of children with severe vesicoureteral reflux. Final report of the International Reflux Study in Children. *Pediatr Nephrol* 21:785-792.
40. Silva JM, Santos Diniz JS, Marino VS, et al. (2006) Clinical course of 735 children and adolescents with primary vesicoureteral reflux. *Pediatr Nephrol* 21:981-988.
41. Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171:737-738.
42. Crick FH (1958) On protein synthesis. *Symp Soc Exp Biol* 12:138-163.
43. Wain HM, Bruford EA, Lovering RC, et al. (2002) Guidelines for human gene nomenclature. *Genomics* 79:464-470.
44. LaFramboise T (2009) Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res* 37:4181-4193.
45. Wang DG, Fan JB, Siao CJ, et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077-1082.
46. Iafrate AJ, Feuk L, Rivera MN, et al. (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36:949-951.
47. Southard AE, Edelmann LJ, Gelb BD (2012) Role of copy number variants in structural birth defects. *Pediatrics* 129:755-763.
48. Yokoyama A, Katsura S, Sugawara A (2017) Biochemical Analysis of Histone Succinylation. *Biochem Res Int* 2017:8529404.
49. Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293:1074-1080.
50. (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931-945.
51. Altschul SF, Gish W, Miller W, et al. (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410.
52. Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12:656-664.
53. Grieshammer U, Le M, Plump AS, et al. (2004) SLIT2-mediated ROBO2 signaling restricts kidney induction to a single site. *Dev Cell* 6:709-717.

54. Hu P, Deng FM, Liang FX, et al. (2000) Ablation of uroplakin III gene results in small urothelial plaques, urothelial leakage, and vesicoureteral reflux. *J Cell Biol* 151:961-972.
55. Kong XT, Deng FM, Hu P, et al. (2004) Roles of uroplakins in plaque formation, umbrella cell enlargement, and urinary tract diseases. *J Cell Biol* 167:1195-1204.
56. Uchiyama Y, Sakaguchi M, Terabayashi T, et al. (2010) Kif26b, a kinesin family gene, regulates adhesion of the embryonic kidney mesenchyme. *Proc Natl Acad Sci U S A* 107:9240-9245.
57. Xu J, Liu H, Chai OH, et al. (2016) Osr1 Interacts Synergistically with Wt1 to Regulate Kidney Organogenesis. *PLoS One* 11:e0159597.
58. Murawski IJ, Gupta IR (2006) Vesicoureteric reflux and renal malformations: a developmental problem. *Clin Genet* 69:105-117.
59. Fillion ML, Watt CL, Gupta IR (2014) Vesicoureteric reflux and reflux nephropathy: from mouse models to childhood disease. *Pediatr Nephrol* 29:757-766.
60. Costantini F, Shakya R (2006) GDNF/Ret signaling and the development of the kidney. *Bioessays* 28:117-127.
61. Miyazaki Y, Oshima K, Fogo A, et al. (2000) Bone morphogenetic protein 4 regulates the budding site and elongation of the mouse ureter. *J Clin Invest* 105:863-873.
62. Nishinakamura R, Sakaguchi M (2014) BMP signaling and its modifiers in kidney development. *Pediatr Nephrol* 29:681-686.
63. Kume T, Deng K, Hogan BL (2000) Murine forkhead/winged helix genes *Foxc1* (*Mf1*) and *Foxc2* (*Mfh1*) are required for the early organogenesis of the kidney and urinary tract. *Development* 127:1387-1395.
64. Nishinakamura R, Uchiyama Y, Sakaguchi M, et al. (2011) Nephron progenitors in the metanephric mesenchyme. *Pediatr Nephrol* 26:1463-1467.
65. Carroll TJ, Park JS, Hayashi S, et al. (2005) Wnt9b plays a central role in the regulation of mesenchymal to epithelial transitions underlying organogenesis of the mammalian urogenital system. *Dev Cell* 9:283-292.
66. Darlow JM, Dobson MG, Darlay R, et al. (2014) 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.
67. McMahon AP, Aronow BJ, Davidson DR, et al. (2008) GUDMAP: the genitourinary developmental molecular anatomy project. *J Am Soc Nephrol* 19:667-671.
68. Murawski IJ, Myburgh DB, Favor J, et al. (2007) Vesico-ureteric reflux and urinary tract development in the Pax2^{1Neu+/-} mouse. *Am J Physiol Renal Physiol* 293:F1736-1745.
69. Okumura T, Furuichi K, Higashide T, et al. (2015) Association of PAX2 and Other Gene Mutations with the Clinical Manifestations of Renal Coloboma Syndrome. *PLoS One* 10:e0142843.

70. Clarke JC, Patel SR, Raymond RM, Jr., et al. (2006) Regulation of c-Ret in the developing kidney is responsive to Pax2 gene dosage. *Hum Mol Genet* 15:3420-3428.
71. Abdelhak S, Kalatzis V, Heilig R, et al. (1997) Clustering of mutations responsible for branchio-oto-renal (BOR) syndrome in the eyes absent homologous region (eyaHR) of EYA1. *Hum Mol Genet* 6:2247-2255.
72. Ruf RG, Xu PX, Silvius D, et al. (2004) SIX1 mutations cause branchio-oto-renal syndrome by disruption of EYA1-SIX1-DNA complexes. *Proc Natl Acad Sci U S A* 101:8090-8095.
73. Hoskins BE, Cramer CH, Silvius D, et al. (2007) Transcription factor SIX5 is mutated in patients with branchio-oto-renal syndrome. *Am J Hum Genet* 80:800-804.
74. Li X, Oghi KA, Zhang J, et al. (2003) Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. *Nature* 426:247-254.
75. Xu PX, Zheng W, Huang L, et al. (2003) Six1 is required for the early organogenesis of mammalian kidney. *Development* 130:3085-3094.
76. Fukami M, Muroya K, Miyake T, et al. (2011) GATA3 abnormalities in six patients with HDR syndrome. *Endocr J* 58:117-121.
77. Grote D, Boualia SK, Souabni A, et al. (2008) Gata3 acts downstream of beta-catenin signaling to prevent ectopic metanephric kidney induction. *PLoS Genet* 4:e1000316.
78. Botzenhart EM, Bartalini G, Blair E, et al. (2007) Townes-Brocks syndrome: twenty novel SALL1 mutations in sporadic and familial cases and refinement of the SALL1 hot spot region. *Hum Mutat* 28:204-205.
79. Nishinakamura R, Matsumoto Y, Nakao K, et al. (2001) Murine homolog of SALL1 is essential for ureteric bud invasion in kidney development. *Development* 128:3105-3115.
80. Chai L, Yang J, Di C, et al. (2006) Transcriptional activation of the SALL1 by the human SIX1 homeodomain during kidney development. *J Biol Chem* 281:18918-18926.
81. Kiefer SM, Robbins L, Stumpff KM, et al. (2010) Sall1-dependent signals affect Wnt signaling and ureter tip fate to initiate kidney development. *Development* 137:3099-3106.
82. Nishimoto K, Iijima K, Shirakawa T, et al. (2001) PAX2 gene mutation in a family with isolated renal hypoplasia. *J Am Soc Nephrol* 12:1769-1772.
83. Jenkins D, Bitner-Glindzicz M, Malcolm S, et al. (2005) De novo Uroplakin IIIa heterozygous mutations cause human renal adysplasia leading to severe kidney failure. *J Am Soc Nephrol* 16:2141-2149.
84. Schonfelder EM, Knuppel T, Tasic V, et al. (2006) Mutations in Uroplakin IIIA are a rare cause of renal hypodysplasia in humans. *Am J Kidney Dis* 47:1004-1012.
85. Jenkins D, Bitner-Glindzicz M, Malcolm S, et al. (2006) Mutation analyses of Uroplakin II in children with renal tract malformations. *Nephrol Dial Transplant* 21:3415-3421.

86. Bertoli-Avella AM, Conte ML, Punzo F, et al. (2008) ROBO2 gene variants are associated with familial vesicoureteral reflux. *J Am Soc Nephrol* 19:825-831.
87. Elahi S, Homstad A, Vaidya H, et al. (2015) Rare variants in tenascin genes in a cohort of children with primary vesicoureteric reflux. *Pediatr Nephrol*.
88. Skinner MA, Safford SD, Reeves JG, et al. (2008) Renal aplasia in humans is associated with RET mutations. *Am J Hum Genet* 82:344-351.
89. Weber S, Taylor JC, Winyard P, et al. (2008) SIX2 and BMP4 mutations associate with anomalous kidney development. *J Am Soc Nephrol* 19:891-903.
90. Gimelli S, Caridi G, Beri S, et al. (2010) Mutations in SOX17 are associated with congenital anomalies of the kidney and the urinary tract. *Hum Mutat* 31:1352-1359.
91. Gbadegesin RA, Brophy PD, Adeyemo A, et al. (2013) TNXB mutations can cause vesicoureteral reflux. *J Am Soc Nephrol* 24:1313-1322.
92. Ostensson M (2012) *Statistical Methods for Genome Wide Association Studies*. Mathematical Sciences. Chalmers University of Technology, Göteborg, Sweden.
93. Allen-Brady K, Camp NJ (2011) Genetic distance and markers used in linkage mapping. *Methods Mol Biol* 713:43-53.
94. Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241-247.
95. Kelly H, Molony CM, Darlow JM, et al. (2007) A genome-wide scan for genes involved in primary vesicoureteric reflux. *J Med Genet* 44:710-717.
96. Briggs CE, Guo CY, Schoettler C, et al. (2010) A genome scan in affected sib-pairs with familial vesicoureteral reflux identifies a locus on chromosome 5. *Eur J Hum Genet* 18:245-250.
97. Cordell HJ, Darlay R, Charoen P, et al. (2010) Whole-genome linkage and association scan in primary, nonsyndromic vesicoureteric reflux. *J Am Soc Nephrol* 21:113-123.
98. Marchini GS, Onal B, Guo CY, et al. (2012) Genome gender diversity in affected sib-pairs with familial vesico-ureteric reflux identified by single nucleotide polymorphism linkage analysis. *BJU Int* 109:1709-1714.
99. Sanna-Cherchi S, Caridi G, Weng PL, et al. (2007) Localization of a gene for nonsyndromic renal hypodysplasia to chromosome 1p32-33. *Am J Hum Genet* 80:539-549.
100. Conte ML, Bertoli-Avella AM, de Graaf BM, et al. (2008) A genome search for primary vesicoureteral reflux shows further evidence for genetic heterogeneity. *Pediatr Nephrol* 23:587-595.
101. Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506-516.
102. Hwang DY, Dworschak GC, Kohl S, et al. (2014) Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary tract. *Kidney Int* 85:1429-1433.

103. Nicolaou N, Pulit SL, Nijman IJ, et al. (2016) Prioritization and burden analysis of rare variants in 208 candidate genes suggest they do not play a major role in CAKUT. *Kidney Int* 89:476-486.
104. Hwang DY, Kohl S, Fan X, et al. (2015) Mutations of the SLIT2-ROBO2 pathway genes SLIT2 and SRGAP1 confer risk for congenital anomalies of the kidney and urinary tract. *Hum Genet* 134:905-916.
105. Kosfeld A, Kreuzer M, Daniel C, et al. (2016) Whole-exome sequencing identifies mutations of TBC1D1 encoding a Rab-GTPase-activating protein in patients with congenital anomalies of the kidneys and urinary tract (CAKUT). *Hum Genet* 135:69-87.
106. Kosfeld A, Brand F, Weiss AC, et al. (2017) Mutations in the leukemia inhibitory factor receptor (LIFR) gene and Lifr deficiency cause urinary tract malformations. *Hum Mol Genet* 26:1716-1731.
107. Bekheirnia MR, Bekheirnia N, Bainbridge MN, et al. (2017) 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* 19:412-420.
108. Wu H, Xu Q, Xie J, et al. (2017) Identification of 8 Novel Mutations in Nephrogenesis-Related Genes in Chinese Han Patients with Unilateral Renal Agenesis. *Am J Nephrol* 46:55-63.
109. Sanna-Cherchi S, Khan K, Westland R, et al. (2017) Exome-wide Association Study Identifies GREB1L Mutations in Congenital Kidney Malformations. *Am J Hum Genet* 101:789-802.
110. De Tomasi L, David P, Humbert C, et al. (2017) Mutations in GREB1L Cause Bilateral Kidney Agenesis in Humans and Mice. *Am J Hum Genet* 101:803-814.
111. Zarrei M, MacDonald JR, Merico D, et al. (2015) A copy number variation map of the human genome. *Nat Rev Genet* 16:172-183.
112. Weber S, Landwehr C, Renkert M, et al. (2011) Mapping candidate regions and genes for congenital anomalies of the kidneys and urinary tract (CAKUT) by array-based comparative genomic hybridization. *Nephrol Dial Transplant* 26:136-143.
113. Siomou E, Mitsioni AG, Giapros V, et al. (2017) Copy-number variation analysis in familial nonsyndromic congenital anomalies of the kidney and urinary tract: Evidence for the causative role of a transposable element-associated genomic rearrangement. *Mol Med Rep* 15:3631-3636.
114. Sanna-Cherchi S, Kirylyuk K, Burgess KE, et al. (2012) Copy-number disorders are a common cause of congenital kidney malformations. *Am J Hum Genet* 91:987-997.
115. Verbitsky M, Sanna-Cherchi S, Fasel DA, et al. (2015) Genomic imbalances in pediatric patients with chronic kidney disease. *J Clin Invest* 125:2171-2178.
116. Caruana G, Wong MN, Walker A, et al. (2015) Copy-number variation associated with congenital anomalies of the kidney and urinary tract. *Pediatr Nephrol* 30:487-495.

117. Nagano C, Nozu K, Morisada N, et al. (2018) Detection of copy number variations by pair analysis using next-generation sequencing data in inherited kidney diseases. *Clin Exp Nephrol* 22:881-888.
118. Ferre S, Igarashi P (2018) New insights into the role of HNF-1beta in kidney (patho)physiology. *Pediatr Nephrol*.
119. Piepsz A, Colarinha P, Gordon I, et al. (2001) Guidelines for 99mTc-DMSA scintigraphy in children. *Eur J Nucl Med* 28:Bp37-41.
120. Gordon I, Piepsz A, Sixt R (2011) Guidelines for standard and diuretic renogram in children. *Eur J Nucl Med Mol Imaging* 38:1175-1188.
121. Piepsz A, Colarinha P, Gordon I, et al. (2001) Guidelines for glomerular filtration rate determination in children. *Eur J Nucl Med* 28:Bp31-36.
122. Brochner-Mortensen J, Haahr J, Christoffersen J (1974) A simple method for accurate assessment of the glomerular filtration rate in children. *Scand J Clin Lab Invest* 33:140-143.
123. Winberg J (1959) The 24-hour true endogenous creatinine clearance in infants and children without renal disease. *Acta Paediatr* 48:443-452.
124. Schwartz GJ, Brion LP, Spitzer A (1987) The use of plasma creatinine concentration for estimating glomerular filtration rate in infants, children, and adolescents. *Pediatr Clin North Am* 34:571-590.
125. Mullis K, Faloona F, Scharf S, et al. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51 Pt 1:263-273.
126. Untergasser A, Cutcutache I, Koressaar T, et al. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res* 40:e115.
127. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463-5467.
128. Moorthie S, Mattocks CJ, Wright CF (2011) Review of massively parallel DNA sequencing technologies. *Hugo J* 5:1-12.
129. Robinson JT, Thorvaldsdottir H, Winckler W, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24-26.
130. Ohlsson M, Hedberg C, Bradvik B, et al. (2012) Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin. *Brain* 135:1682-1694.
131. Bengtsson H, Wirapati P, Speed TP (2009) A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6. *Bioinformatics* 25:2149-2156.
132. Olshen AB, Venkatraman ES, Lucito R, et al. (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5:557-572.
133. Gellerstedt M (2004) M^{12} : medicinsk statistik. Sollentuna : Merck Sharp & Dohme, Sollentuna.
134. Qu HQ, Tien M, Polychronakos C (2010) Statistical significance in genetic association studies. *Clin Invest Med* 33:E266-270.

135. Mackintosh P, Almarhoos G, Heath DA (1989) HLA linkage with familial vesicoureteral reflux and familial pelvi-ureteric junction obstruction. *Tissue Antigens* 34:185-189.
136. Izquierdo L, Porteous M, Paramo PG, et al. (1992) Evidence for genetic heterogeneity in hereditary hydronephrosis caused by pelvi-ureteric junction obstruction, with one locus assigned to chromosome 6p. *Hum Genet* 89:557-560.
137. Gunaratne A, Chan E, El-Chabib TH, et al. (2015) aPKC alters the TGFbeta response in NSCLC cells through both Smad-dependent and Smad-independent pathways. *J Cell Sci* 128:487-498.
138. Vukicevic S, Helder MN, Luyten FP (1994) Developing human lung and kidney are major sites for synthesis of bone morphogenetic protein-3 (osteogenin). *J Histochem Cytochem* 42:869-875.
139. Takahashi H, Ikeda T (1996) Transcripts for two members of the transforming growth factor-beta superfamily BMP-3 and BMP-7 are expressed in developing rat embryos. *Dev Dyn* 207:439-449.
140. Cancilla B, Ford-Perriss MD, Bertram JF (1999) Expression and localization of fibroblast growth factors and fibroblast growth factor receptors in the developing rat kidney. *Kidney Int* 56:2025-2039.
141. Lin Y, Liu A, Zhang S, et al. (2001) Induction of ureter branching as a response to Wnt-2b signaling during early kidney organogenesis. *Dev Dyn* 222:26-39.
142. Gui T, Sun Y, Gai Z, et al. (2013) The loss of Trps1 suppresses ureteric bud branching because of the activation of TGF-beta signaling. *Dev Biol* 377:415-427.
143. Gai Z, Zhou G, Itoh S, et al. (2009) Trps1 functions downstream of Bmp7 in kidney development. *J Am Soc Nephrol* 20:2403-2411.
144. Pohl M, Sakurai H, Bush KT, et al. (2000) Matrix metalloproteinases and their inhibitors regulate in vitro ureteric bud branching morphogenesis. *Am J Physiol Renal Physiol* 279:F891-900.
145. Scott JE, Swallow V, Coulthard MG, et al. (1997) Screening of newborn babies for familial ureteric reflux. *Lancet* 350:396-400.
146. Brandstrom P, Esbjorner E, Herthelius M, et al. (2010) The Swedish reflux trial in children: I. Study design and study population characteristics. *J Urol* 184:274-279.
147. Menezes M, Puri P (2009) Familial vesicoureteral reflux--is screening beneficial? *J Urol* 182:1673-1677.
148. Carvas F, Silva A, Nguyen HT (2010) The genetics of primary, nonsyndromic vesicoureteral reflux. *Curr Opin Urol* 20:336-342.
149. Quaynor SD, Bosley ME, Duckworth CG, et al. (2016) Targeted next generation sequencing approach identifies eighteen new candidate genes in normosmic hypogonadotropic hypogonadism and Kallmann syndrome. *Mol Cell Endocrinol* 437:86-96.

150. Willem M, Miosge N, Halfter W, et al. (2002) Specific ablation of the nidogen-binding site in the laminin gamma1 chain interferes with kidney and lung development. *Development* 129:2711-2722.
151. Yang DH, McKee KK, Chen ZL, et al. (2011) Renal collecting system growth and function depend upon embryonic gamma1 laminin expression. *Development* 138:4535-4544.
152. Barasch J, Yang J, Ware CB, et al. (1999) Mesenchymal to epithelial conversion in rat metanephros is induced by LIF. *Cell* 99:377-386.
153. Plisov SY, Yoshino K, Dove LF, et al. (2001) TGF beta 2, LIF and FGF2 cooperate to induce nephrogenesis. *Development* 128:1045-1057.
154. Jiang S, Gitlin J, Deng FM, et al. (2004) Lack of major involvement of human uroplakin genes in vesicoureteral reflux: implications for disease heterogeneity. *Kidney Int* 66:10-19.
155. Esbjorner E, Hansson S, Jakobsson B (2004) Management of children with dilating vesico-ureteric reflux in Sweden. *Acta Paediatr* 93:37-42.