

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN  
Scienze Mediche Specialistiche**

Ciclo XXVI

**Settore Concorsuale di afferenza: 06/D2**

**Settore Scientifico disciplinare: MED/14**

**Molecular genetics of inherited cystic kidney diseases: new diagnostic approaches**

**Presentata da: Dr.ssa Adelia Aquilano**

**Coordinatore Dottorato:  
Prof. Sandro Mattioli**

**Relatore  
Prof. Gaetano La Manna**

**Correlatore  
Dr.ssa Vilma Mantovani**

**Esame finale anno 2014**

## Table of contents

1. Abstract .....	4
2. Background .....	5
2.1. Inherited cystic kidney diseases .....	5
2.1.1. Nephronophthisis .....	5
2.1.2. Medullary cystic kidney disease 2 and Familial juvenile hiperuricemic nephropathy type1 .....	8
2.1.3. Familial juvenile hyperuricemic nephropathy type2 .....	10
2.1.4. Renal cysts and diabetes syndrome .....	10
2.1.5. Polycystic kidney disease.....	12
2.2. The molecular diagnosis of inherited cystic kidney disease.....	15
2.2.1. Conventional mutation screening .....	15
2.2.2. Next generation sequencing for molecular diagnostics applications .....	19
2.2.3. Ion Torrent PGM platform.....	20
3. Purpose .....	23
4. Methods .....	24
4.1. Diagnostics algorithm for patients selection.....	24
4.2. DNA extraction.....	25
4.3. Conventional methods for mutation screening.....	26
4.3.1. PCR Primers design.....	26
4.3.2. Deletion analysis of <i>NPHP1</i> gene .....	30
4.3.3. <i>NPHP5</i> analysis .....	32
4.3.4. dHPLC analysis and Sanger sequencing of <i>UMOD</i> , <i>REN</i> and <i>HNF1B</i> genes 33	
4.3.5. MLPA analysis of <i>HNF1B</i> gene.....	37
4.4. Next Generation Sequencing of <i>PKHD1</i> gene.....	38
4.4.1. Patients selection .....	38
4.4.2. Assay Design .....	39
4.4.3. DNA Libraries construction.....	42
4.4.4. Template preparation .....	45
4.4.5. Sequencing .....	47
4.4.6. Sequence Analysis.....	47
4.4.7. Validation by Sanger sequencing.....	47
5. Results and Discussion .....	50
5.1. <i>NPHP1</i> deletions .....	50
5.2. <i>NPHP5</i> mutations .....	50

5.3.	<i>REN</i> mutations.....	51
5.4.	<i>HNF1B</i> mutations .....	52
5.5.	Next Generation Sequencing output.....	54
5.5.1.	Annotation and filtration of variants .....	55
5.5.2.	Validation .....	59
5.5.3.	Molecular diagnostics application.....	60
6.	Conclusions.....	62

# 1. Abstract

**Background.** Hereditary cystic kidney diseases are a heterogeneous spectrum of disorders leading to renal failure. Clinical features and family history can help to distinguish the recessive from dominant diseases but the differential diagnosis is difficult due to the phenotypic overlap. The molecular diagnosis is often the only way to characterize the different forms. A conventional molecular screening is suitable for small genes but is expensive and time-consuming for large size genes. Next Generation Sequencing (NGS) technologies enable massively parallel sequencing of nucleic acid fragments.

**Purpose.** The first purpose was to validate a diagnostic algorithm useful to drive the genetic screening. The second aim was to validate a NGS protocol of *PKHD1* gene.

**Methods.** DNAs from 50 patients were submitted to conventional screening of *NPHP1*, *NPHP5*, *UMOD*, *REN* and *HNF1B* genes. 5 patients with known mutations in *PKHD1* were submitted to NGS to validate the new method and a not genotyped proband with his parents were analyzed for a diagnostic application.

**Results.** The conventional molecular screening detected 8 mutations: 1) the novel p.E48K of *REN* in a patient with cystic nephropathy, hyperuricemia, hyperkalemia and anemia; 2) p.R489X of *NPHP5* in a patient with Senior Loken Syndrome; 3) p.R295C of *HNF1B* in a patient with renal failure and diabetes.; 4) the *NPHP1* deletion in 3 patients with medullar cysts; 5) the *HNF1B* deletion in a patient with medullar cysts and renal hypoplasia and in a diabetic patient with liver disease.

The NGS of *PKHD1* detected all known mutations and two additional variants during the validation. The diagnostic NGS analysis identified the patient's compound heterozygosity with a maternal frameshift mutation and a paternal missense mutation besides a not transmitted paternal missense mutation.

**Conclusions.** The results confirm the validity of our diagnostic algorithm and suggest the possibility to introduce this NGS protocol to clinical practice.

## 2. Background

### 2.1. Inherited cystic kidney diseases

Inherited cystic kidney disease is a poorly understood and identified condition and a leading genetic cause of established renal failure. The differential diagnosis of this diseases appears difficult and is hampered by the variability of clinical manifestations, particularly in the early stage of disease and by the genetic heterogeneity.

While the development of cysts and progressive impairment of renal function are common features, other aspects may help in distinguishing different forms: the age of onset, the rate of progression of renal disease and a different extra-renal organ involvement (*Conti G, 2009*).

In their studies, Osathanondh and Potter systematically classified renal cystic diseases into four distinct types. Potter syndrome type I is referred to as autosomal recessive polycystic kidney disease (ARPKD), type II as renal cystic dysplasia, type III as autosomal dominant polycystic kidney disease (ADPKD), and type IV occurs when a longstanding obstruction in either the kidney or ureter leads to cystic kidneys or hydronephrosis. Particularly types II–IV can be part of many syndromes. While this classification still has an impact for pathoanatomical descriptions, it is hardly to be reconciled with clinical and genetic entities and, consequently, it is more and more being replaced by the genetic nomenclature (*Bergmann C, 2014*).

The traditional classification divides these disorders in dominant and recessive diseases in relation to the type of transmission. The identification of genes and the study of their protein products allowed us to shed light on the pathogenic processes underlying the formation of cysts (*T Watnick, 2003; Scolari F, 2004*).

#### 2.1.1. Nephronophthisis

Nephronophthisis (NPH) is an autosomal recessive disease characterized by a chronic tubulo-interstitial nephritis that progresses to terminal renal failure during the second decade in the juvenile form or before the age of 5 years in the infantile form (*Hildebrandt F et al, 2009*).

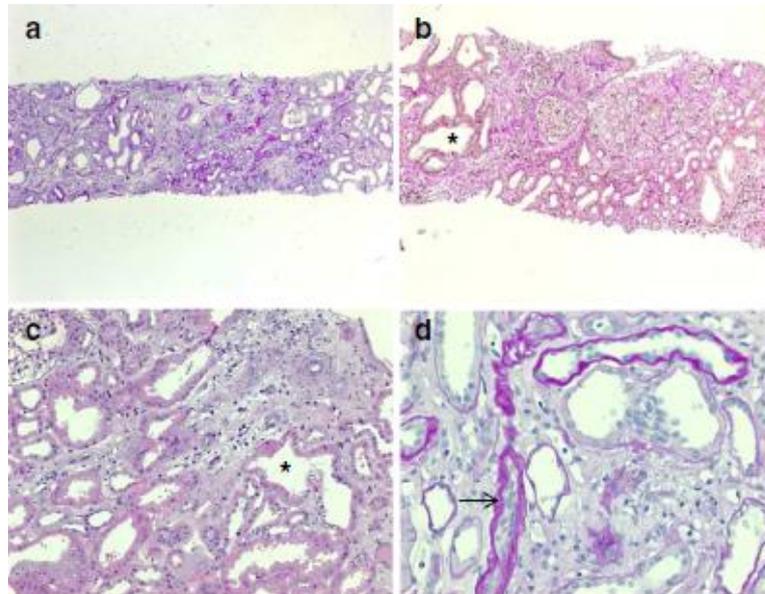
This disease is one of the most common causes of end stage renal disease in children with an incidence of 10-15%, is the third hereditary pediatric disease leading to dialysis and transplantation, and occurs with a prevalence of 1:100,000 individuals (*Hildebrandt F et al, 2007*).

The normal or slightly reduced kidney volume, the loss of cortico-medullary differentiation, the increased echogenicity and the development of cysts with a diameter ranging from 1 to 15 mm at the level of the cortico-medullary junction are typical clinical features (Figure 1).



**Figure 1:** Sonographic features of nephronophthisis. Renal ultrasonography shows the presence of cortico-medullary cysts (*Roslyn JS, 2009*).

Histologically, tubular basement membrane disintegration and thickening, tubular atrophy, and disproportionate tubulointerstitial fibrosis with minimal inflammation are observed (Figure 2).



**Figure 2:** a–d Renal histology of a 14-year-old female with juvenile NPH carrying a homozygous NPHP1 deletion. Chronic tubulo-interstitial alterations with tubular basement membrane disintegration and thickening, focal tubular atrophy (arrow in d), cystic dilated tubuli (asterisk in b and c), and disproportionate tubulo-interstitial fibrosis with some inflammatory cells were observed (*Omran H, 2008*).

To date, about 20 genes have been described for recessive NPH, plus XPNPEP3 in which mutations cause an NPH-like phenotype (NPHP1L) (*Bergamoni C, 2014*). In Table 1 are reported all mutated NPH associated genes.

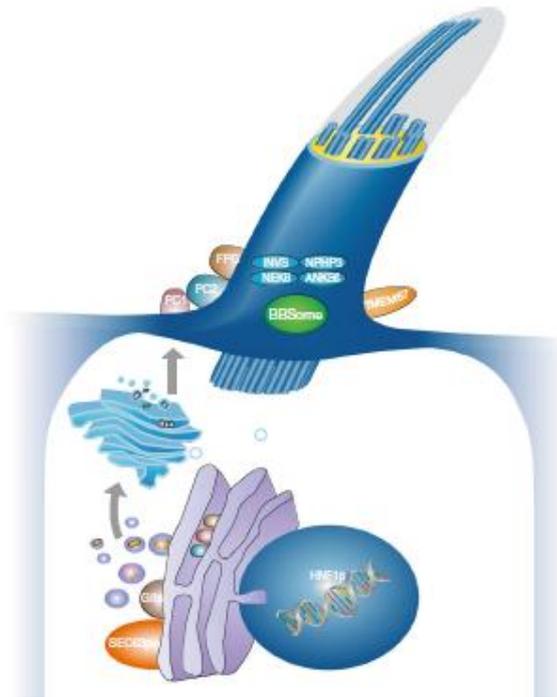
Locus	Gene	Chromosome	Protein	Mutation frequency	Extrarenal features
<b>NPHP1</b>	NPHP1	2q13	Nephrocystin-1	23%	SLS, JS,
<b>NPHP2</b>	INV	9q31	Inversin	1-2%	SLS, HF VSD, situs inversus
<b>NPHP3</b>	NPHP3	3q22.1	Nephrocystin-3	<1%	SLS, HF, MKS, situs inversus
<b>NPHP4</b>	NPHP4	1p36.22	Nephrocystin-4 or nephroretinin	2-3%	SLS
<b>NPHP5</b>	IQCB1	3q21.1	Nephrocystin-5 or IQ motif containing B1	3-4%	SLS
<b>NPHP6</b>	CEP290	12q21.32	Centrosomal protein 290	1%	LCA, SLS, JS, MKS, BBS
<b>NPHP7</b>	GLIS2	16p13.3	GLI similar 2	<0.5%	/
<b>NPHP8</b>	RPGRIP1L	16q12.2	RPGRIP1-like	0.5%	SLS, JS, MKS
<b>NPHP9</b>	NEK8	17q11.1	NIMA-related kinase 8	<0.5%	SLS
<b>NPHP10</b>	SDCCAG8	1q44	Serologically defined colon cancer antigen 8	<0.5%	SLS, BBS-like
<b>NPHP11</b>	TMEM67	8q22.1	Transmembrane protein 67	<0.5%	JS, HF, MKS
<b>NPHP12</b>	TTC21B	2q24.3	Intraflagellar transport protein 139	<1%	JS, MKS, BBS, JATD
<b>NPHP13</b>	WDR19	4p14	WD repeat-containing protein 19	<0.5%	retinal microaneurysm
<b>NPHP14</b>	ZNF423	16q12.1	Zinc finger protein 423	<0.5%	JS
<b>NPHP15</b>	CEP164	11q23.3	Centrosomal protein 164	<0.5%	LCA, SLS
<b>NPHP16</b>	ANKS6	9q22.33	ankyrin repeat and sterile alpha motif domain-containing protein 6	<0.5%	/
<b>NPHP17</b>	IFT172	2p23.3	Intraflagellar transport 172	<0.5%	thoracic dysplasia, polydactyly
<b>NPHPL1</b>	XPNPEP3	22q13	X-prolyl aminopeptidase 3	<0.5%	cardiomyopathy, seizures

**Table 1:** mutated genes in Nephronophthisis and associated extrarenal manifestations. BBS: Bardet-Biedl syndrome; HF: hepatic fibrosis; JATD: Jeune asphyxiating thoracic dystrophy; JS: Joubert syndrome; LCA: Leber's congenital amaurosis; MKS: Meckel-Gruber syndrome; SLS: Senior-Loken syndrome; VSD: ventricular septal defect.

Given that these genes only account for about one-half of all NPH cases, further heterogeneity can be expected. *NPHP1* on chromosome 2q13 is the most commonly mutated gene in NPH. In large series of patients, a homozygous deletion of *NPHP1* was present in 20–40% of cases; heterozygous deletions were found in another 6 % of patients, harboring a concomitant point mutation on the other parental *NPHP1* allele .

Other NPH genes described to date may only contribute a minor part to the total mutational load of typical NPH (*Salomon et al, 2009*).

It is important to note that many NPH genes are pleiotropic and can cause a much broader phenotypic spectrum (with other ciliopathies such as Senior–Loken, Joubert, Meckel, Ivemark, and Jeune syndrome) than just isolated NPH. Mutations in a subset of NPH genes can resemble PKD with enlarged kidneys and sometimes even prenatal manifestation of Potter’s sequence. Obviously, these respective NPHP members engage in a closely assembled protein–protein interaction sub-network (Figure 3).



**Figure 3:** Schematic diagram of a primary cilium and associated processes (*Bergmann C, 2014*)

Positional cloning of these genes and functional characterization of their encoded proteins (nephrocystins) has contributed to a unifying theory that defines cystic kidney diseases as “ciliopathies”. The theory is based on the finding that all the mutated proteins in cystic kidney diseases of humans or animal models are expressed in primary cilia or centrosomes of renal epithelial cells. The ciliary theory explains the multiple organ involvement in NPHP, which includes retinal degeneration, cerebellar hypoplasia, liver fibrosis, situs inversus, and mental retardation.

### **2.1.2. Medullary cystic kidney disease 2 and Familial juvenile hiperuricemic nephropathy type1**

*UMOD*-associated kidney disease (uromodulin-associated kidney disease) is also known as familial juvenile hyperuricemic nephropathy type 1 (FJHN1) and medullary cystic kidney disease type 2 (MCKD2). Clinical findings typically include hyperuricemia and gout (resulting from reduced kidney excretion of uric acid) that usually occur as early as the teenage years. Slowly progressive interstitial kidney disease begins early in life. Elevations in serum creatinine usually occur between ages five and 40 years, leading to end-stage

kidney disease (ESRD) usually between the fourth and seventh decade. The age at ESRD varies both between and within families (*Bleyer AJ, 2013*).

The study of common genomic regions critical, has allowed us to detect the presence of mutations within the same gene locus identified as *UMOD* (*Hart T, 2002; Calado J, 2005*). The *UMOD* gene is located on chromosome 16 and comprises 11 exons. This gene is transcribed only in kidney and encodes for the Uromodulin also known as Tamm-Horsfall protein, a glycoprotein of 95-kDa.

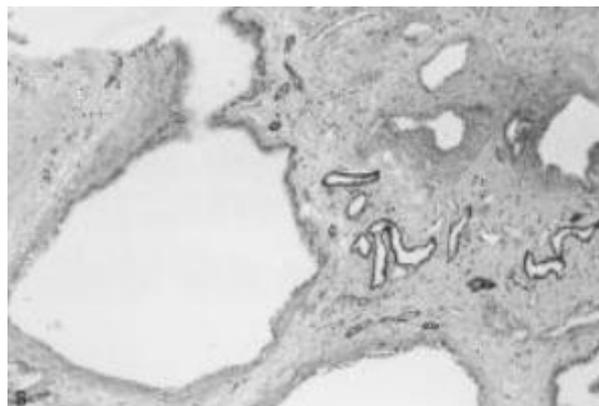
MCKD is often clinically regarded as the autosomal dominant counterpart of NPH (NPHP–MCKD complex) with a usually later onset of renal failure than the recessive forms. The ciliary protein uromodulin (= Tamm–Horsfall glycoprotein), the most abundant protein in the urine of healthy individuals, plays a major role and is encoded by *UMOD* on chromosome 16p12 (*Vylet'al P et al, 2006*).

Mutations in this gene can lead to different tubulointerstitial nephropathies, including MCKD2, glomerulocystic kidney disease, and familial juvenile hyperuricemic nephropathy , which may also be caused by mutations in *TCF2 (HNF1β)* and *REN* (renin). Only recently has *MUC1*, a locus mapped more than a decade ago on chromosome 1q22, been identified as the gene underlying MCKD1 (*Kirby A, 2013*).

*Kiser et al. (2004)* noted that the diagnosis of MCKD is difficult because initial signs and symptoms may be mild or vague, symptoms of frank renal failure occur late, renal cysts may be absent in over 50% of patients, and renal histologic abnormalities are nonspecific.

The MCKD occurs with a prevalence of 1-9/1.000.000 individuals and is mainly characterized by alteration of the urine concentration, hyperuricemia, tubulo-interstitial fibrosis and progressive renal failure (*Scolari F, 2004*).

Frequently observed the presence of renal cysts at the cortico-medullary junction and the histology of the renal parenchyma showing tubular atrophy and appearance of a mild inflammatory infiltrate (*Lhotta K, 2009*) (Figure 4).



**Figure 4:** renal biopsy of a 27 years patient with chronic renal failure and hypertension (*Bisceglia M, 2006*).

Two forms of MCKD have been distinguished in relation to the age of chronic renal failure onset: MCKD1 and MCKD2 with onset of 62 and 32 years respectively.

The clinical features of FJHN1 include hyperuricemia, gout, polyuria, polydipsia, and progressive renal failure from the young age. Rarely renal cysts are reported in patients with FJHN1 (*Izzi C, 2010*).

### **2.1.3. Familial juvenile hyperuricemic nephropathy type2**

The family juvenile hyperuricemic nephropathy type 2 (FJHN2), described for the first time in 1960 by Duncan and Dixon, is an autosomal dominant chronic tubulointerstitial nephritis, (*Duncan H, 1960*).

FJHN2 is a rare disorder (prevalence < 1/1.000.000) that occurs during childhood and progresses to chronic renal failure during the fourth/sixth decade of life. The majority of children show anemia, hyperuricemia (serum uric acid > 6 mg / dl), hematuria and reduced ability to concentrate urine resulting in the development of polyuria within the first year of life (*Bleyer AJ, 2010*).

The size of the kidneys is normal or slightly reduced, with no evidence of renal cysts. Histological examination reveals focal tubular atrophy, secondary glomerular scarring and interstitial fibrosis. FJHN2 is caused by mutations of *REN* gene, located on chromosome 1 and made of 10 exons. Most of the known mutations are missense, small deletions, and small inframe insertions located in the first and second exon. *REN* encodes for renin, a proteolytic enzyme of 37 Kd produced by the liver as Pro- renin and subsequently activated in the juxtaglomerular cells of the kidney. The renin acts by activating the conversion of angiotensinogen to angiotensin I. *REN* mutations result in a reduced expression of protein up to the complete loss of nephrons and progressive renal failure (*Bleyer AJ, 2013*).

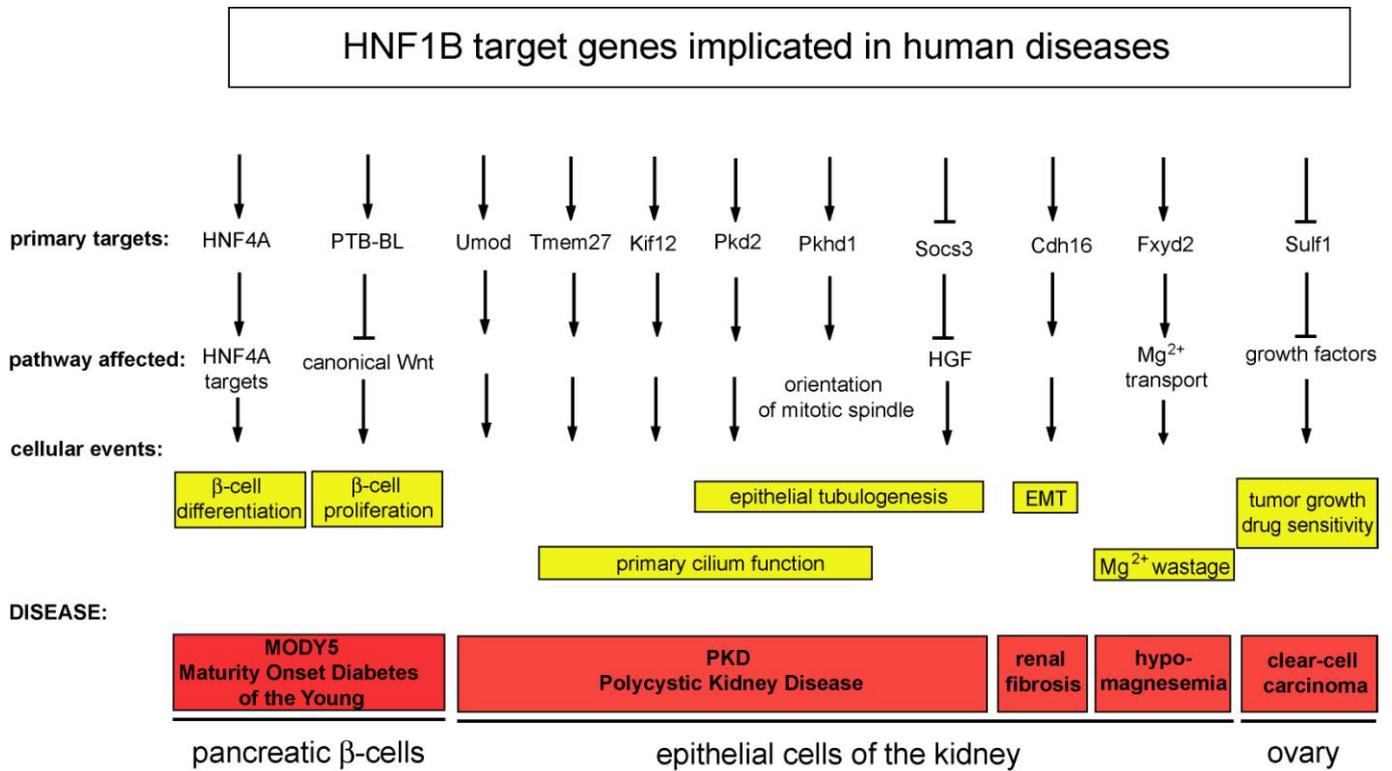
### **2.1.4. Renal cysts and diabetes syndrome**

The term of renal cysts and diabetes syndrome (RCAD) has been recently proposed to substitute for maturity onset diabetes of the young 5 (MODY5) due to the higher prevalence of renal phenotypes than early-onset diabetes. RCAD/MODY5 has been widely reported either in the form of case reports or cohort studies, and the spectrum of *HNF1B* anomalies and clinical phenotypes have been substantially changing over the past decade (*Chen YZ, 2010*)

*HNF1B* gene consists of nine exons and encodes the 557-amino acid hepatocyte nuclear factor 1-beta (*HNF1β*). It is an atypical Pit-1/Oct-1/Unc-86 (POU) transcription factor comprising three distinct domains: an N-terminus dimerization domain, a highly conserved DNA binding domain, and a C-terminus transactivation domain.

In Figure 5 the *HNF1β* target genes that are implicated in human diseases are listed. The genes encoding the transcription factor HNF4A (MODY1) and the protein tyrosine phosphatase-BL (PTP-BL) affect differentiation and proliferation of beta-cells in the pancreas. Dysregulation of both processes are implicated in maturity onset diabetes in the young. In contrast, the gene encoding the GPI anchored uromorulin (Umod) as well as the genes encoding proteins that are preferentially localized in the primary cilium, i. collectrin (Tmem27), kinesin family member 12 (Kif12), polycystin2 (Pkd2), polyductin (Pkh1), are

activated by *HNF1β* in epithelial cells of the kidney. They are involved in primary cilium function and/or epithelial tubulogenesis and dysfunction of both processes lead to polycystic kidney disease. In the case of *Pkhd1* dysfunction a distorted orientation of the mitotic spindle has been reported that leads to renal tubular enlargement and cyst formation (*Gerhart U R, 2009*).



**Figure 5:** HNF1B in human diseases.

Mutation in *HNF1B* was first described by *Horkawa et al* in 1997. Subsequently, different mutation types, such as missense, nonsense, frameshift, splice-site mutations, single exon deletion/duplication, and small in-frame deletion, have been increasingly identified in different domains.

*Thomas et al* first detected *HNF1B* deletion in 2005, and *HNF1B* whole-gene deletion may be the most common mutational mechanism.

Mutations are strikingly located within the DNA binding domain and varied among exons of the DNA binding domain: exons 2 and 4 are the hottest spots, while mutations are sporadically distributed in exon 3. The consistent phenotypes are renal structure anomalies (RSA) (89.6%) and diabetes mellitus (DM) (45.0%). However, the concurrence of RSA and DM is relatively low (27.5%), which hinders the optimal performance of genetic testing and obtainment of timely diagnosis. Other organ involvements are complementary and necessary for the early identification of patients with *HNF1B* anomalies. Analysis of phenotypes of *HNF1B* point mutations shows significant differences in the detection rates of RSA, impaired renal function and DM according to mutation type but not mutation location.

There are so many other features that could provide vital clues to the identity of *HNF1B* anomalies carriers, which include exocrine pancreas dysfunction (6.6%), pancreas structural abnormalities (10.4%), asymptomatic liver dysfunction (15.2%), liver structure anomalies (4.3%), genital tract malformations (13.7%), and intrauterine growth retardation (3.3%) (*Chen YZ, 2010*).

### 2.1.5. Polycystic kidney disease

Polycystic kidney disease (PKD) includes a subset of hereditary kidney diseases characterized by the progressive formation and growth of fluid-filled cysts in both kidneys. Based on its mode of inheritance, the PKD is classified as an autosomal dominant and autosomal recessive form. In Table 2 are reported clinical and pathophysiological aspects of two forms.

Characteristics	ARPKD	ADPKD
<b>Synonyms</b>	Potter type I	Potter type III
<b>Incidence Approx.</b>	1:20.000	1:500–1.000 (approx. 2 % early manifesting)
<b>Pathology of kidneys</b>		
<b>Macroscopy</b>	Massively, symmetrically enlarged kidneys (reniform)	Generally enlarged (also reniform), but usually to a lesser extent
<b>Location of cysts</b>	Dilated collecting ducts and distal tubules	Cysts in all parts of the nephron (including glomerulus)
<b>Ultrasound and diameter of cysts</b>	At onset, typical pepper–salt pattern evident on ultrasound scan, increased echogenicity of renal parenchyma throughout cortex and medulla due to tiny, sometimes invisible cysts (usually <2 mm); with advancing age, cysts up to several centimeters large appear, similar to ADPKD pattern	Cysts in all parts of the nephron (including glomerulus)Cysts of different size in cortex and medulla (usually several larger cysts in adults); at onset often small, however, sometimes already several centimeters early in childhood
<b>Pathology of liver</b>	Mandatory: ductal plate malformation/congenital hepatic fibrosis with hyperplastic biliary ducts and portal fibrosis (may impress as Caroli disease)	“Liver cysts” common in adults, but rare in children. Occasionally, ductal plate malformation/congenital hepatic fibrosis
<b>Associated anomalies</b>	Rarely pancreatic cysts and/or fibrosis; single case reports with intracranial aneurysms	Pancreatic cysts and/or cysts in other epithelial organs; intracranial aneurysms in approx. 8 %, familial clustering
<b>Main clinical manifestations</b>	Peri-/neonatal period: respiratory distress (30–50 %of cases)With prolonged survival, renal insufficiency, portal hypertension, and other variable co-morbidities	General onset 3rd–5th decade with arterial hypertension, proteinuria, hematuria, and/or renal insufficiency; approx. 2 %earlymanifestation in childhood (rarely with perinatal respiratory distress)
<b>Risk for siblings</b>	25%	50 % (except for rare cases of spontaneous mutation with virtually no risk)
<b>Risk for own children</b>	<1 % (unless unaffected parent is related to his/her affected partner, or ARPKD is known in the unaffected partner’s family)	50 % (also for patients with a spontaneous mutation)
<b>Manifestation in affected family members</b>	Often similar clinical course in siblings (however, in approx.20 % extensive intrafamilial variability)	Variable, however, often similar within the same family; in the case of early manifestation approx. 50 % recurrence risk

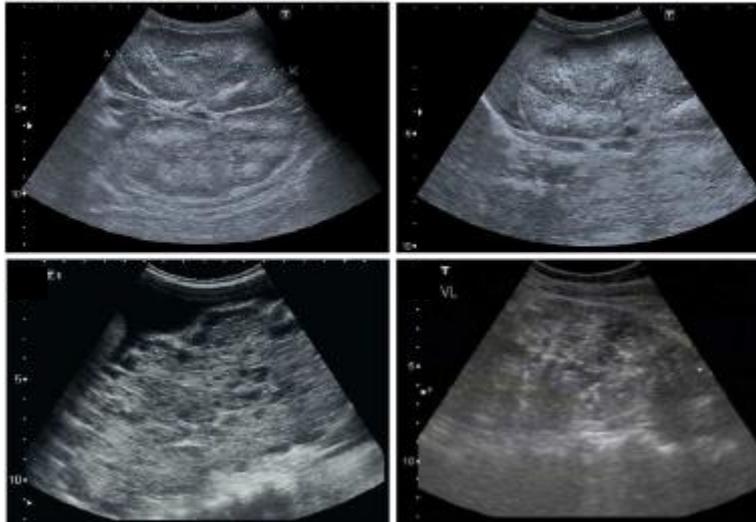
<b>Parental kidneys</b>	No alterations	Except for cases of spontaneous mutation, usually one parent is affected and shows renal cysts (be careful when parents are too young for definite clinical diagnosis, namely, <30– 40 years)
<b>Prognosis</b>	In perinatal cases with respiratory distress, usually poor; for those surviving the neonatal period, much better with renal death in approx. 15–30 % in childhood/early adolescence, often severe complications (e. g., esophageal varices) due to portal hypertension; if possible transplantation (often combined kidney–liver TX)	In early manifesting cases, often better than in ARPKD. In “adult” cases, chronic renal failure in approx. 50 % by age of 60 years; median age of ESRD onset (58.1 vs. 79.9 years in PKD1 vs. PKD2)

**Table 2:** Characteristics of autosomal recessive and autosomal dominant polycystic kidney diseases

Autosomal recessive polycystic kidney disease (ARPKD), although less frequent than the dominant form, is a common, inherited kidney disease of childhood. ARPKD belongs to the family of cilia-related disorders and shows distinct clinical features and genetics. This condition occurs with a frequency ranging from 1:6000 to 1:40,000 live births.

Although the clinical spectrum in ARPKD is much more variable than generally presumed, and even only moderately affected elderly people have been described, then majority of patients are severely affected and ARPKD is identified late in pregnancy or at birth. Affected fetuses display a “Potter” oligohydramnios phenotype with massively enlarged kidneys, pulmonary hypoplasia, a characteristic facies, and contracted limbs with club feet. Approximately 30–50 % of affected neonates die shortly after birth from respiratory insufficiency due to pulmonary hypoplasia and thoracic compression by the excessively enlarged kidneys (*Adeva M et al, 2006*).

Ultrasound scans typically show bilaterally enlarged hyperechoic kidneys with poor corticomedullary differentiation, retained reniform contour, and multiple tiny cysts confined to distal tubules and collecting ducts. With advancing clinical course the kidney structure might increasingly resemble the pattern observed in ADPKD with renal cysts that vary considerably in size and appearance, often also accompanied by some degree of interstitial fibrosis (Figure 6), (*Avni FE et al, 2002*).



**Figure 6** : Renal ultrasound scans of babies and young children with ARPKD. Symmetrically enlarged echogenic kidneys with fusiform dilations of collecting ducts and distal tubules are arranged radially throughout the renal parenchyma from medulla to cortex (*Bergamon c, 2014*)

Arterial hypertension, often difficult to control despite multi-drug treatment, usually develops during the first months of life and affects up to 80 % of children with ARPKD. While the early appearance of ARPKD is typically clearly dominated by renal manifestations and associated co-morbidities, histological liver involvement is present in every ARPKD patient from early embryonic development on, as reflected by the disease term “polycystic kidney and hepatic disease 1 (PKHD1).” These obligatory liver changes are characterized by defective remodeling of the ductal plate with congenital hepatic fibrosis and biliary duct ectasia, defined as ductal plate malformation (DPM) which is also a frequent feature in other ciliopathies such as, for example, Bardet–Biedl, Joubert, Meckel, and Jeune syndrome.

To date, *PKHD1* is the only known gene for classical ARPKD, but there is compelling evidence for locus heterogeneity and phenocopies. Thus, single heterozygous mutations and results that are merely based on linkage need to be interpreted with caution. Patients with two truncating mutations generally display a severe phenotype with peri- or neonatal death, whereas patients surviving the neonatal period usually carry at least one hypomorphic (missense) mutation. Due to allelic heterogeneity and a high level of missense mutations and private changes, mutation analysis for *PKHD1* is still laborious, but it has greatly benefited from the availability of new sequencing techniques.

*PKHD1* is a large gene that extends over a genomic segment of almost 500 kb on chromosome 6p12. The longest open reading frame comprises 66 exons that encode polyductin/fibrocystin, a type I single-pass transmembrane protein of 4,074 amino acids. In common with both ADPKD proteins (polycystin-1 and polycystin-2) and most other cystoproteins, fibrocystin is localized to primary cilia with the highest concentration in the basal body area. This striking pattern of subcellular localization and known interactions with, for example, polycystin-2, place fibrocystin at key sites of microtubule organization. In line with its proposed role as a ciliary-localized membrane protein, an 18-residue motif in the cytoplasmic tail of fibrocystin serves as a ciliary targeting signal.

## **2.2. The molecular diagnosis of inherited cystic kidney disease**

Knowledge of the family history and the clinical picture, the location and morphology of the cysts and any possible extra-renal manifestations should help the physician in the decision-making process. When an effort is made to classify the wide array of different entities with renal cysts, it might be helpful to first distinguish between acquired and inherited forms. An accurate molecular diagnosis of cystic kidney disorders is essential to both the management of patients with cystic kidneys and the counseling for their families.

### **2.2.1. Conventional mutation screening**

The conventional approach to the molecular diagnosis of hereditary cystic kidney diseases consists in a cascade analysis of targeted genes in order from the most probably associated with the disease. To do this, is set a diagnostic decisional algorithm that guides the screening of mutations to the positive outcome of the genetic test.

#### **Sanger sequencing**

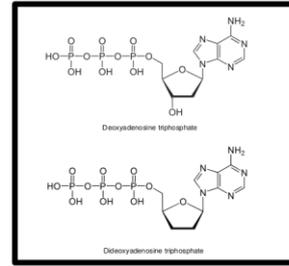
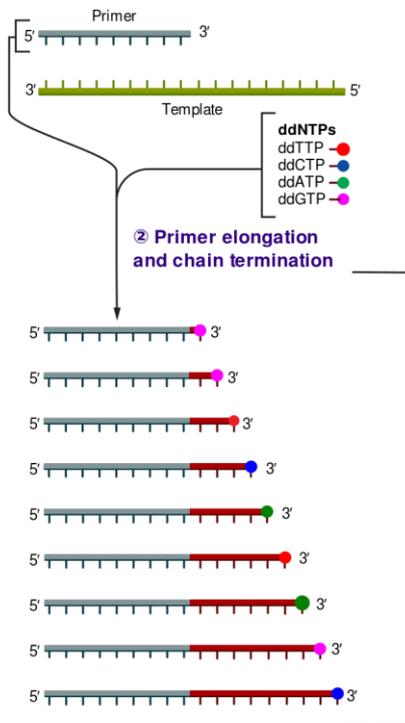
The most widely used method for the detection of point mutations is the Sanger sequencing of the coding regions and the splicing sites. This method devised by Sanger in 1977 is to date the most reliable tool used in genetics laboratories.

Sanger's method, which is also referred as dideoxy sequencing or chain termination, is based on the use of dideoxynucleotides (ddNTP's) in addition to the normal nucleotides (NTP's) found in DNA. Dideoxynucleotides are essentially the same as nucleotides except they contain a hydrogen group on the 3' carbon instead of a hydroxyl group (OH). These modified nucleotides, when integrated into a sequence, prevent the addition of further nucleotides. This occurs because a phosphodiester bond cannot form between the dideoxynucleotide and the next incoming nucleotide, and thus the DNA chain is terminated.

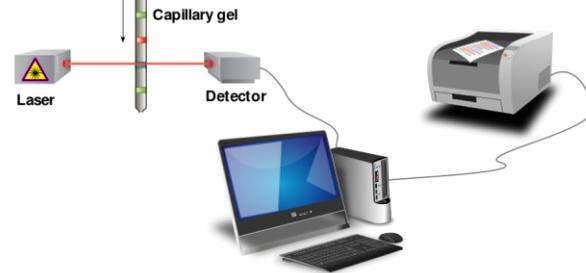
The execution of the original method required a very long time, both for the preparation of the sample, both for the reading of the results, in addition to employing radioactive substances. To date the performance of the analysis have been simplified by the use of automated sequencers and software for data analysis. Even if the execution times are quite long, the Sanger sequencing is still the most reliable method for diagnostics with a sensitivity of 90%.

① Reaction mixture

- ▶ Primer and DNA template
- ▶ DNA polymerase
- ▶ ddNTPs with flouochromes
- ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



③ Capillary gel electrophoresis separation of DNA fragments



④ Laser detection of flouochromes and computational sequence analysis

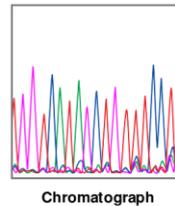
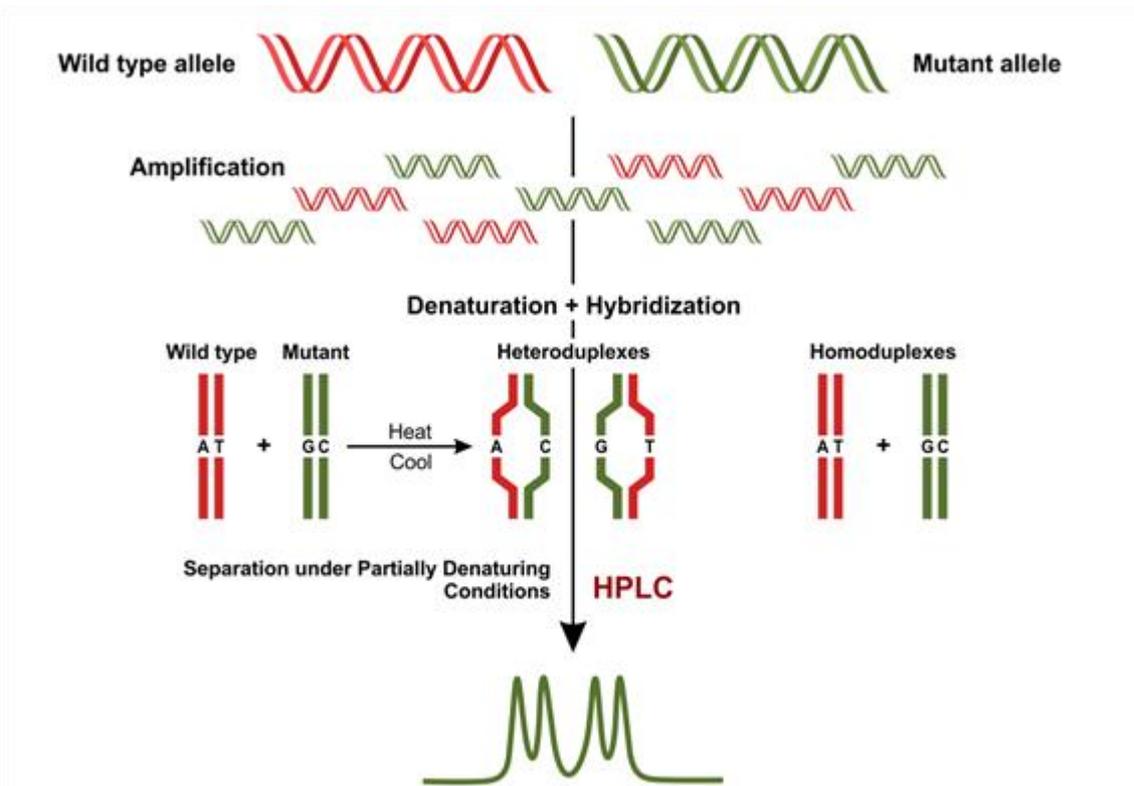


Figure 7: Schematic view of the automated Sanger (chain-termination) method for DNA sequencing.

### dHPLC

The dHPLC (denaturing high performance liquid chromatography), developed in the laboratory of Cavalli-Sforza at Stanford University (USA), is a technique of ion-exchange liquid chromatography in reverse phase.

dHPLC uses heteroduplex formation between wild-type and mutated DNA strands to identify mutations. Heteroduplex molecules are separated from homoduplex molecules by ion-pair, reverse-phase liquid chromatography on a special column matrix with partial heat denaturation of the DNA strands (Figure 8).



**Figure 8:** Mutation screening method by dHPLC. Molecular diagnostics (*Balogh et al, 2011*)

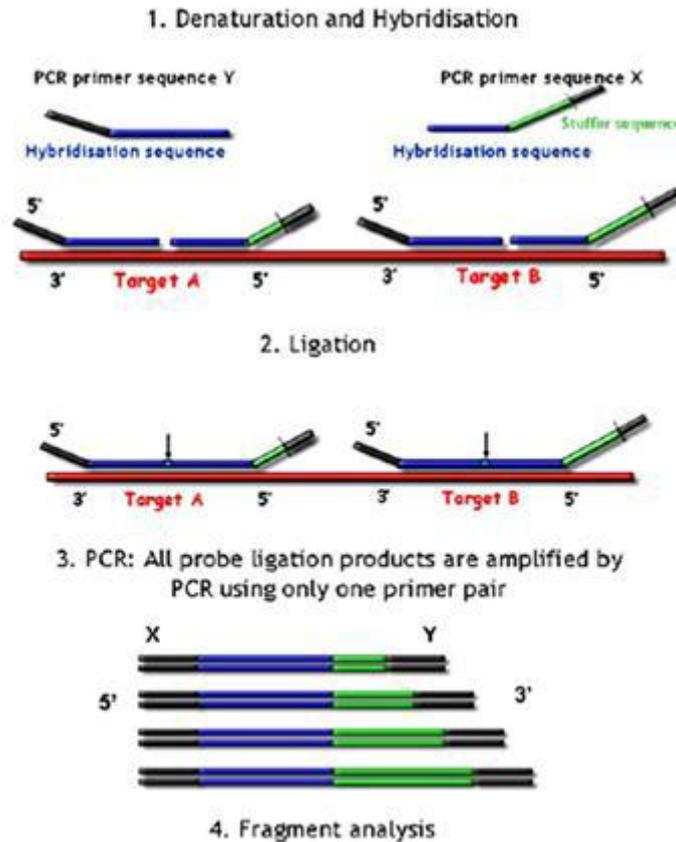
dHPLC is a very useful method for the screening of a large number of samples for mutations. It has mainly been used for the analysis of germline mutations in various inherited diseases. The sensitivity of the technique dHPLC for point mutations is of 95/97% and therefore can be used to analyze DNA fragments of small size, usually between 100 and 700 base pairs (*Sheffield VC, 1993*).

This technique revealing the presence of a mutation or a polymorphism allows us to submit to sequence analysis only positive samples, reducing time and costs of molecular tests.

### **MLPA**

MLPA (Multiplex Ligation-dependent Probe Amplification) is a multiplex PCR method detecting abnormal copy numbers of genomic DNA sequences.

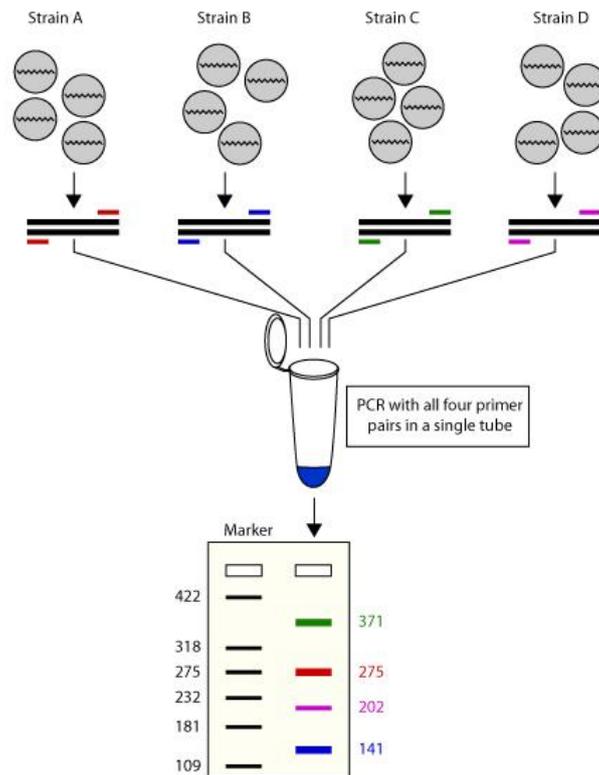
The method is based on the use of oligonucleotide probes complementary to specific gene sequences. Each probe is constituted by one end, defined sequence hybridization, complementary to a target site on the DNA and from one end of the genomic oligonucleotide in which the last twenty base pairs are complementary to a universal primer M13. One of the probes has a sequence nonspecific of precise length that will allow the recognition of the labeled fragments that will be produced, according to their length. Such fragments can be subsequently analyzed by capillary electrophoresis. This technique has been used for the analysis of *HNF1B* rearrangements of RCAD affected patients.



**Figure 9:** outline of MLPA technique: After hybridisation to their target sequence in the sample DNA, the probe oligonucleotides are enzymatically ligated. One probe oligonucleotide contains a non-hybridising stuffer sequence of variable length. Ligation products can be amplified using PCR primer sequences X and Y amplification product of each probe has a unique length (130-480 nt). Amplification products are separated by electrophoresis. (Jiufeng S, 2011).

**Multiplex PCR for large deletion analysis**

The multiple polymerase chain reaction (Multiplex PCR) is a variant of the conventional PCR. Multiplex PCR is a widespread molecular biology technique for amplification of multiple targets in a single PCR experiment. In a multiplexing assay, more than one target sequence can be amplified by using multiple primer pairs in a reaction mixture (Figure 10)



**Figure 10:** Outline of Multiple PCR method.

As an extension to the practical use of PCR, this technique has the potential to produce considerable savings in time and effort within the laboratory without compromising on the utility of the experiment. Multiplex PCR is used for the identification of large homozygous deletions of *NPHP1* gene in patients with clinical suspect of Nephronophthisis; the analysis protocol has been validated by Hildebrandt F (1997).

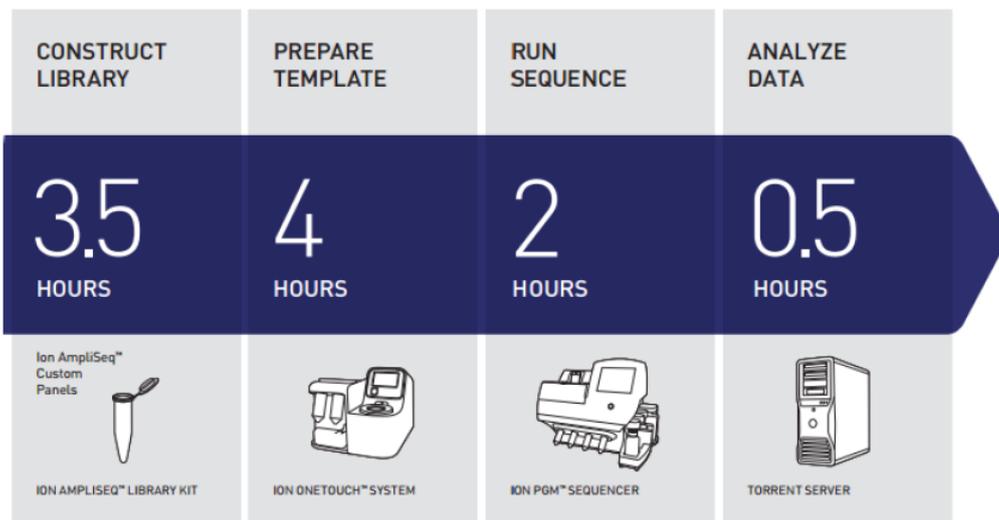
### 2.2.2. Next generation sequencing for molecular diagnostics applications

The introduction of high-throughput techniques (HT) has greatly increased the possibility of analysis of biological molecules. The term High-Throughput refers to the simultaneously ability to perform numerous measurements. The field of genomics has been the first to benefit from the HT technology with the development of so-called high-throughput sequencing (HTS), or next generation sequencing (NGS).

The advent of NGS technology has opened important perspectives in molecular diagnostics and represents a powerful tool for the simultaneous analysis of a large number of coding regions. The application of this technique in molecular analysis of large genes such as *PKHD1* would greatly reduce long analysis times.

### 2.2.3. Ion Torrent PGM platform

The Ion Torrent PGM platform uses a semiconductor technology: the core of the platform is a semiconductor chip containing an array with millions of micro-wells as many sensors. Each micro-well houses a single sphere containing multiple copies of the same strand of ssDNA. This platform works with the “sequencing by synthesis” technology. The incorporation of a new nucleotide during synthesis produces the release of two waste products: the pyrophosphate (PPi) and hydrogen ions (H<sup>+</sup>). Every time a single type dNTP (dATP, dTTP, dCTP or dGTP) is added to the chip and is incorporated in the growing chain, the pH sensitive semiconductor detects the protons increase. The lowering of pH generates an increase of the current on the sensitive surface which is converted in a variation of potential. After the chip is flooded with a single type of dNTP, the signals are recorded; a wash is executed to restore the pH conditions. This step is necessary to detect the addition of the next dNTP. The received signals are reported in a Ionogram. The workflow of Ion PGM system consists of a few basic steps (Figure 11).

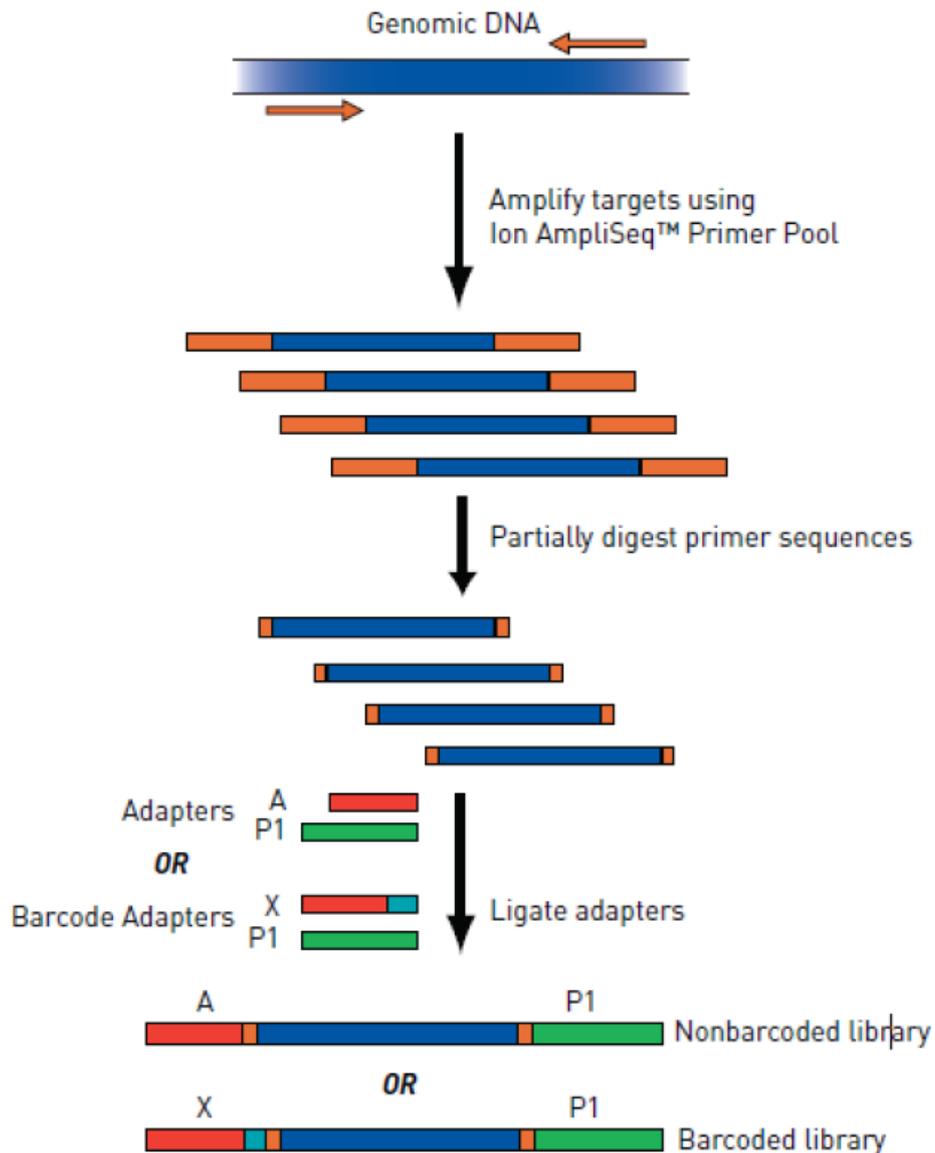


**Figure 11:** Workflow for sequencing protocol by Ion Torrent PGM

- **Library preparation**

During the library preparation the amplicons of 100-400 bp are bound to the adapter sequences.

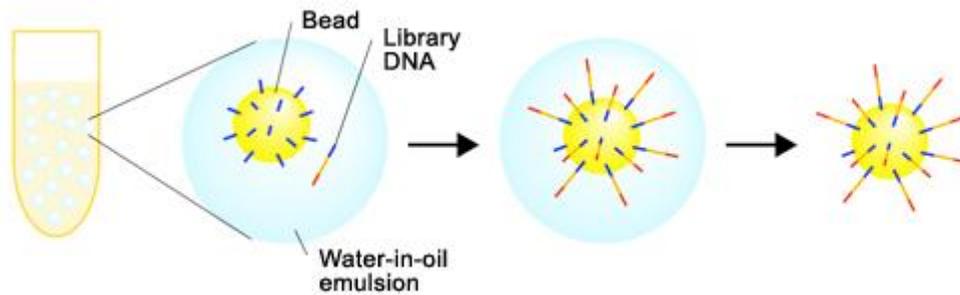
To do this the target sequences are amplified by PCR and the specific primers are submitted to enzymatic digestion. The enzymatic reaction allows the binding of the adapters to 5' or 3' of digested amplicons. Subsequently the barcodes are linked to distinguish different samples. The use of barcodes of 6-10 nucleotides is useful for analyzing multiple samples in the same run (Figure 12).



**Figure 12:** Schematic view of library preparation

- Template preparation

The fragments generated during the library preparation are attached to beads and amplified using emulsion PCR (emPCR). Beads coated with complementary primers are mixed with a dilute aqueous solution containing the fragments to be sequenced along with the necessary PCR reagents (Figure 12). This solution is then mixed with oil to form an emulsion of microdroplets. The concentration of beads and fragments is kept low enough such that each microdroplet contains only one of each. Clonal amplification of each fragment is then performed within the microdroplets. Following amplification the emulsion is ‘broken’ (generally by organic extraction and centrifugation) and the amplified beads are enriched in a glycerol gradient.



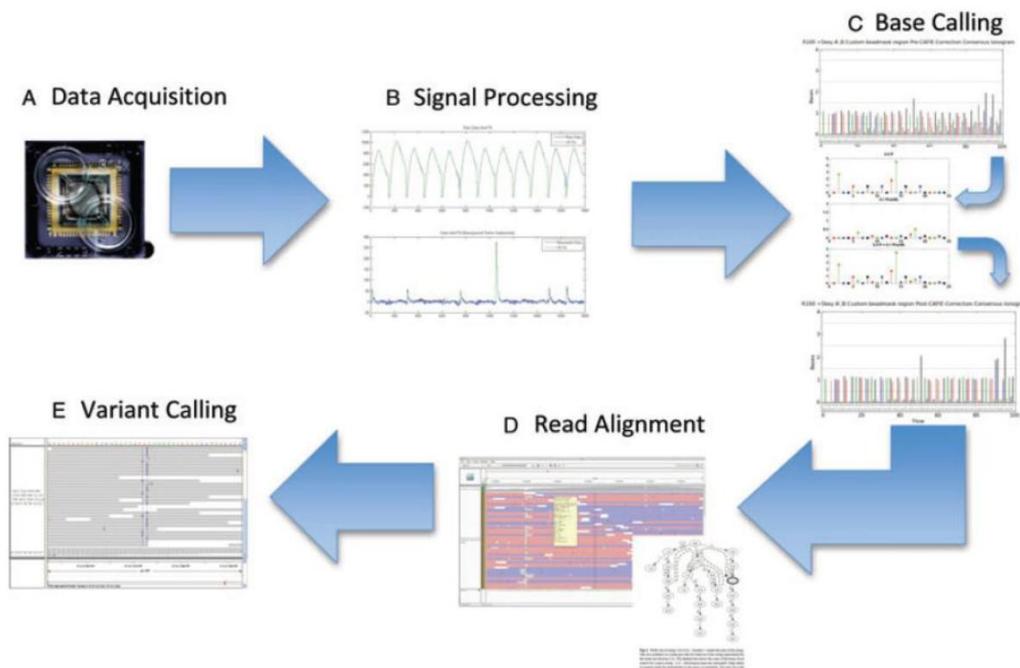
**Figure 13:** Phases of emulsion PCR

- Sequencing

The first step consists in loading the chip on the Ion Torrent PGM. The balls bound to the template and the polymerase are pushed to uniformly fill the wells of the chip by centrifuging. Once the chip is placed on the platform the cycles of washing are alternated with cycles of synthesis. The synthesis and the data acquisition are simultaneously performed. The Ion Torrent PGM machine is direct connected with the Torrent Suite Software and the transmitted data are normalized and converted to the ionogram.

- Data analysis

The Ion Torrent system generates standard output files like FASTQ that are aligned with the reference genome by the Variant Caller. Only reads with high quality score are filtered (error rate <1%). The final output is a .bam file format containing all variants information. The .bam files can be loaded in all the software for the variants analysis (Figure 14).



**Figure 14:** Schematic view from acquisition to data analysis.

### 3. Purpose

The first purpose of this study was to perform a conventional molecular screening of inherited cystic kidney diseases in order to validate a diagnostic algorithm related to clinical phenotypes improving the genetic screening. This diagnostic algorithm drives the identification of the causative variants of *NPHP1*, *NPHP5*, *UMOD*, *REN* and *HNF1B*, starting from the most likely involved gene.

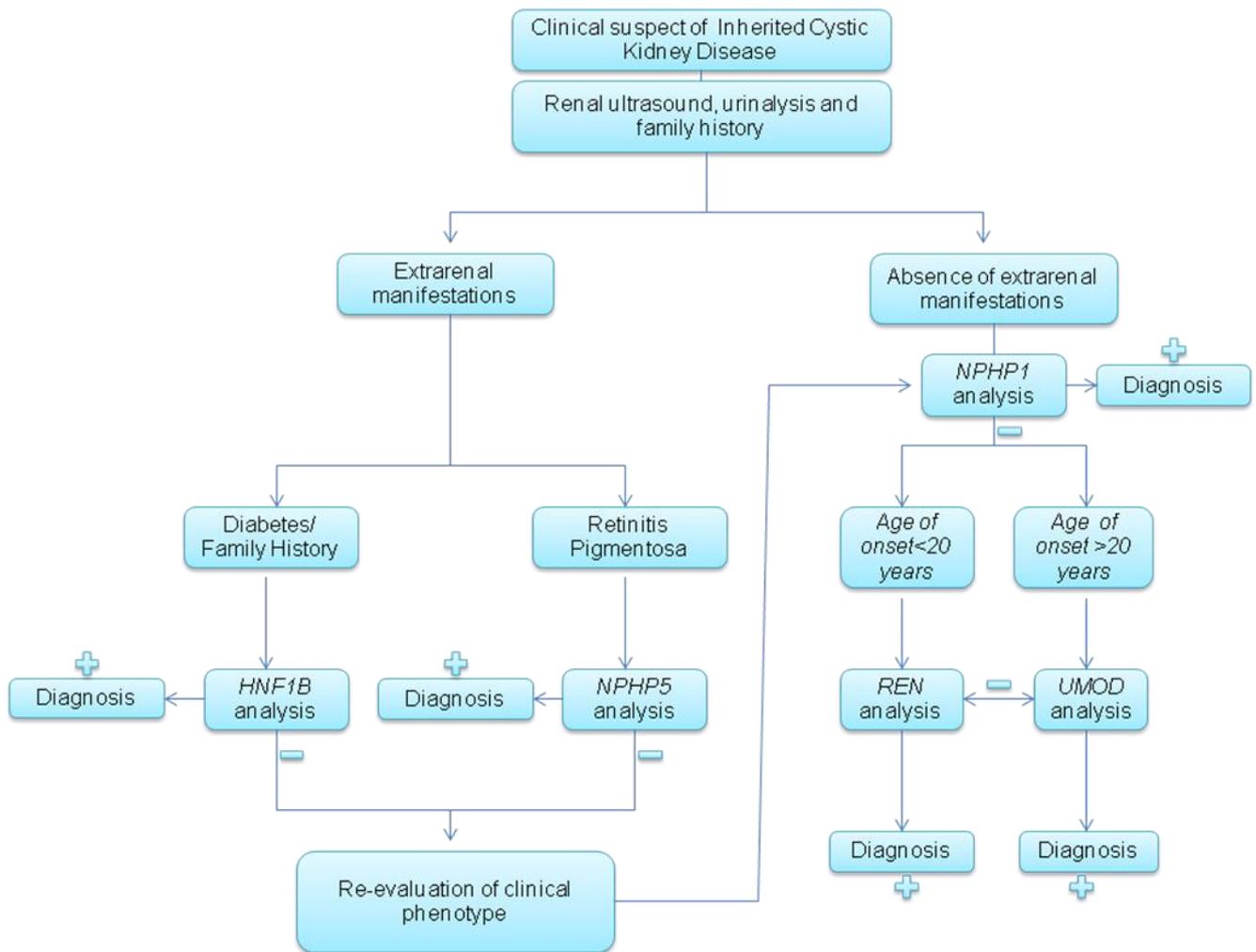
Due the size of the *PKHD1* gene, the second aim of the project was to develop and validate a NGS protocol to enable the mutational screening. The assay is more comprehensive, rapid and inexpensive, compared to the conventional method and could improve the clinical practice allowing the genetic counseling for families.

## 4. Methods

### 4.1. Diagnostics algorithm for patients selection

A diagnostic algorithm related to the clinical phenotypes has been developed to drive the molecular screening.

The diagnostic flowchart used for the patients selection and for the choice of the genes to analyze is an adaptation of the one proposed by *Rèmi Salomon et al* (2009) (Figure 15).



**Figure 15:** Diagnostic algorithm used for patients selection

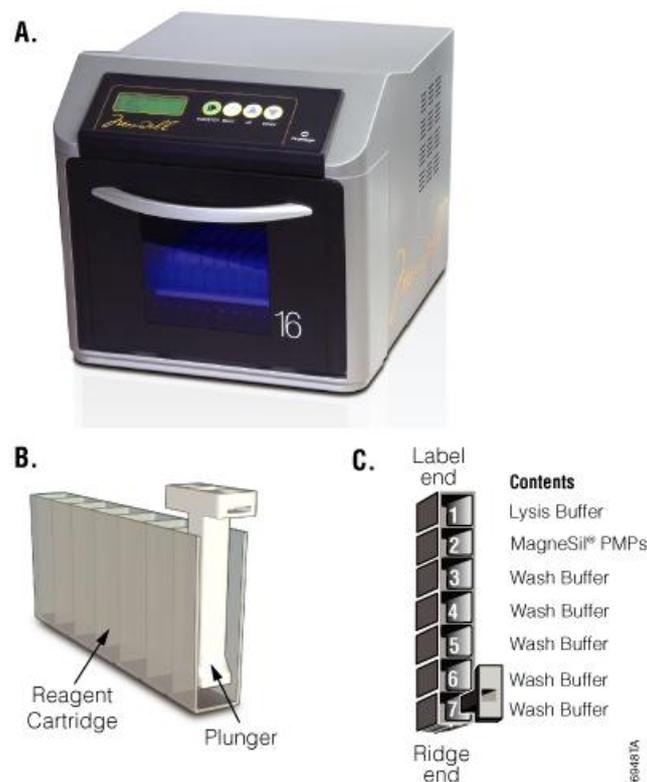
Patients undergoing genetic screening were selected in the departments of Nephrology and Medical Genetics at the S. Orsola-Malpighi Hospital of Bologna, after obtaining informed consent for genetic study. Once ultrasound and urinalysis have excluded polycystic kidney and glomerulonephritis, the presence of extrarenal manifestations has been verified. *NPHP5* and *HNF1B* have been analyzed in patients whom showed Retinitis Pigmentosa and diabetes respectively. If the result was negative the clinical phenotype

was re-evaluated in order to go on with the screening of the genes more frequently associated with the absence of extrarenal manifestations. *NPHP1* analysis was performed in patients in whom the onset of kidney disease occurred before the age of 20 years. In case of negative result the samples have been submitted to the analysis of *REN* and *UMOD* in patients with age of onset less than and greater than 20 years respectively. In patients with clinical suspicion of ADPKD the analysis of gene *PKHD1* was performed using the Ion Torrent PGM platform.

## 4.2. DNA extraction

The DNA extraction from blood cells was performed by a semi-automatic technique using the Maxwell® 16 machine (Promega Corporation, Madison, WI USA). This tool allows to perform the automated extraction up to a maximum of 16 samples. Maxwell 16 reduces the risk of contamination and allows to standardize the DNA concentration.

The Maxwell 16 System used a paramagnetic-particle handling system that processes samples using Maxwell reagent kits (Figure 16).



**Figure 16:** Panel A. The Maxwell 16 Instrument. Panel B. A Maxwell 16 reagent cartridge, cut away to show the plunger. The design of the cartridge allows direct processing of a sample with no need for preprocessing. Panel C. Top view of the Maxwell 16 DNA Purification Cartridge.

The DNA concentration and purity have been measured using a NanoDrop spectrophotometer (Figure 17) (Celbio s.p.a., Milan, Italy). Each sample was normalized to a concentration of 30-50 ng / L. and stored at -20 ° C.



**Figure 17:** Nanodrop Spectrophotometer

### 4.3. Conventional methods for mutation screening

#### 4.3.1. PCR Primers design

The reference sequences were downloaded from three different databases and compared to each other: UCSC Genome Browser, Ensembl, Gene (ge.ucsc.edu), (www.ensembl.org), (www.ncbi.nlm.nih.gov/gene). The PCR primers were designed using the Primer3 Software (<http://primer3.ut.ee>). The presence of self-pairing and the specificity of amplification were tested by the use of Oligo Analysis Tool (<http://www.operon.com/tools/oligo-analysis-tool.aspx>) and UCSC In Silico PCR (<http://ge.ucsc.edu/cgi-bin/hgPcr>) respectively. Primers with more than four consecutive inter and intra primers self-pairings were discarded. The possible presence of SNPs at the site of annealing was verified by SNPmasker (<http://bioinfo.ebc.ee/snpmasker/>).

The sequences of the primers used in PCR reactions are shown in the following tables (Table 3 to 7).

#### **NPHP1**

<b><u>Primer</u></b>	<b><u>Sequence 5'-&gt;3'</u></b>	<b><u>Target (bp)</u></b>	<b><u>Exon</u></b>
LHX9_4F LHX9_4R	ATATGGCTCTGCCTTGCTTC TTGGGCAAAACACACTCTTG	515	Exon 4
LHX9_6F LHX9_6R	ACCCCTAAAAGCCAAGTTGC CCTAATAGTGTCTTTGTCTTCACTGC	180	Exon 6
NPHP1_20F NPHP1_20R	AATGGCACCCCTCCATCCTAC AATCGTGGAGGATCCATCTG	436	Exon 20
NPHP1_5F NPHP1_5R	CACTCATAGCTGGTCTGTTCTTG CAGGTGTACAGGCAGAGTTTTTC	339	Exon 5
NPHP1_7F NPHP1_7R	TGTTTTTACTGGAGGGTTAGGTG CCTAACCTACTTTGATATCCTTTCC	236	Exon 7

**Table 3:** Sequence, target size (bp) and amplified exons for *NPHP1* and *LHX9*.

**NPHP5**

<b><u>Primer</u></b>	<b><u>Sequence 5'-&gt;3'</u></b>	<b><u>Target (bp)</u></b>	<b><u>Exon</u></b>
NPH5_3F NPH5_3R	CATTCTGCCTCTCAAACAAGG GAGCTTCCTAAAGGTGTGATGTT	500	Exon 3
NPH5_4F NPH5_4R	CCTGAGCAGAATGTCCCTGT TGAAGGAAAAGTACAACCAAAACC	567	Exon 4
NPH5_5F NPH5_5R	AGCCAAGTGTCTTTTATTTGCTG CATAGAGATGATTTTTGGAAGGTTG	544	Exon 5
NPH5_6F NPH5_6R	CCTAGTGTATCAACTTGTAGCCCTTA TTCAACATGGTTTCATTTTCAGTG	487	Exon 6
NPH5_7F NPH5_7R	GATGGTCATGCTTTGATTTGG CAATGGTGATGGAACCTCAGC	397	Exon 7
NPH5_8F NPH5_8R	GCACCTTGCACACAGTAGGA TCAGAAAAGTATCTTCCACATGC	598	Exon 8
NPH5_9F NPH5_9R	TTTTCATTTGTTTTTAAGCCAGA TTTTGGGGGTATTTTGCTTG	340	Exon 9
NPH5_10F NPH5_10R	TTGCCTTACCAAGCCTAACA GATTCCTGTCTTACATCCTTTGC	477	Exon 10
NPH5_11F NPH5_11R	CACAACAGCAGCAGATGACA AAAAACTGGTGGGGAGATTG	523	Exon 11
NPH5_12F NPH5_12R	GCTGCATAATCTGGGAAAGC AAAATAAGGCTCAAGAAAATAAGTG	464	Exon 12
NPH5_13F NPH5_13N	CCTGCATTTTGTACAGTCC ACTTCTATGGTAAAGCCAATGC	382	Exon 13
NPH5_14F NPH5_14R	TGATAAAATGAGAGACAGCTAAGCA TTCCTGAGGTTAGGGGATGA	384	Exon 14
NPH5_15AF NPH5_15AR	TCCAGCTTAGGCAATAGAGCA TGTGTGGCTAACGATGAGGA	597	Exon 15

**Table 4:** Sequence, target size (bp) and amplified exons for *NPHP5*.

**UMOD**

<b><u>Primer</u></b>	<b><u>Sequence 5'-&gt;3'</u></b>	<b><u>Target (bp)</u></b>	<b><u>Exon</u></b>
UM1F UM1R	TGCTGTTAGAAGGTGCGAAA CCCCAGTGTCCAAGGTCTTA	384	Exon 1
UM2F UM2R	AGACTGGGATGTTGGTGAGG GACAGGTGCTACATTGCTTCC	473	Exon 2
UM3AF UM3AR	GAATGAGGGAAGGATCTCTGG GGCATACGCACAAGTAGCTG	398	Exon 3
UM3BF UM3BR	TCCTTCTCCTGCGTCTGC CGTGCCATTGAGCCACAT	438	Exon 3
UM3CF UM3CR	AGTACGGGGAGGGCTACG TCACAGGGACAGACAGACAATC	400	Exon 3
UM4F UM4R	GAGTCTCCCCACAGTCCTCA ATATGGCCCAATCTCACAG	296	Exon 4
UM5F UM5R	AGCACTTCCAGATGGTGGTC TCCTCTCAATACACCTGTGGAG	490	Exon 5
UM6F UM6R	GACCAGCAGATTTAGCTTTGAA GGTTAAGGGGTTTGGGGTTA	299	Exon 6
UM7F UM7R	GGAATGCAAATCAGAGAAGG TTTCCTCCATCCAAGTCCAA	397	Exon 7
UM8F UM8R	GCTGAAAGGTGACAGAGCAA AAGAGGGAAACAGGGAAGAAA	351	Exon 8
UM9F UM9R	CATCTATCTAACAAATGGCAGAGC CCTTGGCTCCCAGTTCTTC	379	Exon 9
UM10F UM10R	CTTGCTAAGGGTTGGGACCT CCTTTGGGTTTGTATAAGTTTGTG	499	Exon 10
UM11F UM11R	TAGGAGAGCAGCCCAGAGAG ACACCGTCACAAGTCCCATT	391	Exon 11

**Table 5:** Sequence, target size (bp) and amplified exons for *UMOD*.

**REN**

<b><u>Primer</u></b>	<b><u>Sequence 5'-&gt;3'</u></b>	<b><u>Target (bp)</u></b>	<b><u>Exon</u></b>
RENPF RENPR	ACAGGGCCAAGCCAGATAG GTCGGGAGACCAAAGGTACA	373	Promoter
REN1F REN1R	GGTAATAAATCAGGGCAGAGC ATGTGGAAAAGCCAGGGTGT	300	Exon 1
REN2F REN2R	AGCGACAGAGGTGATTTCCA GCCTTCGTCAAACACAGCTT	399	Exon 2
REN3F REN3R	AAGAGTGAAGCCAGGCAAGA GTGTTGGGCAGGATTGCTC	272	Exon 3
REN4F REN4R	AACCAGCCATACCCACAATG TTGCTCCCCCATAGGTAAGT	393	Exon 4
REN5F REN5R	GAAGCCCAGACCAGACTCC AGATATTGATTGGCTGTCACCT	391	Exon 5
REN6F REN6R	CAGTCCCCCACTTGTTCACT GGGCTTGCTGATGTGAGTTT	290	Exon 6
REN7F REN7R	CACCCTTGCGAACCTTCC GGCAGGATGGTAATGCAGTC	296	Exon 7
REN8F REN8R	CCCACACTCAGGAAGGACAT TGAGGTGAACAAGCGAAGG	344	Exon 8
REN9F REN9R	TCGCTTGTTACCTCACACT GAATCACCTGGGGAAACTGA	491	Exon 9
REN10F REN10R	GTGTGGCTAGGGGAGAGATG TTGTCCTCAAAGCAGGGAAG	460	Exon 10

**Table 6:** Sequence, target size (bp) and amplified exons for *REN*

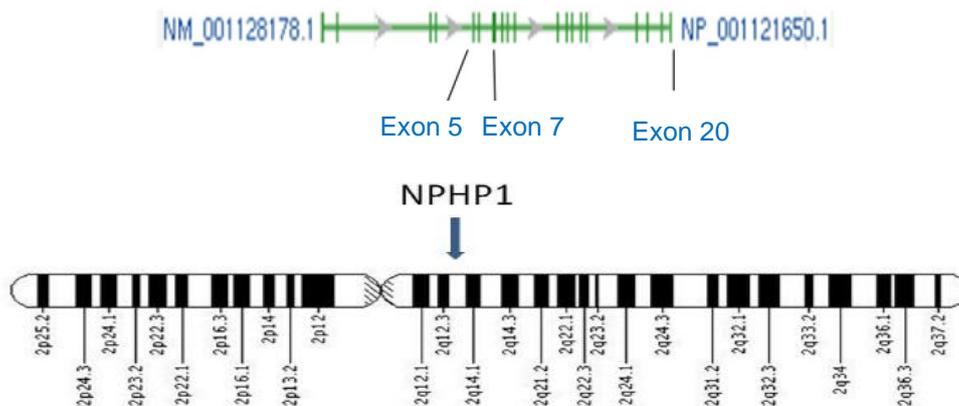
## HNF1B

<u>Primer</u>	<u>Sequence 5'→3'</u>	<u>Target (bp)</u>	<u>Exon</u>
MOD5PF MOD5PR	CCCCAGCGTGAGTACAATG GGATGGCAGCAACTCCTC	479	Promoter
MOD51F MOD51R	ATTTGGGGTTTGCTTGTGAA GACTTCTCTGGTGGGAAACG	547	Exon 1
MOD52F MOD52R	GCACCCCTAGAAAAAGAATG ATCTGCCAAGTGCTCACAAAG	387	Exon 2
MOD53F MOD53R	TGTCTGTCTGCTGAGTGAAGG AGGGTTCCTGGGTCTGTGTA	412	Exon 3
MOD54F MOD54R	CCAACCAAGACTGCTGTGATT CAGATAAGATCCGTGGCAAGA	400	Exon 4
MOD55F MOD55R	ACAAGAGGTGCCGAGTCATT TCTGGACAGCCCTCATTTTC	367	Exon 5
MOD56F MOD56R	CACATCGTGTTGGAACTGC TGGATTTAAGGAGACAGATGAGAA	364	Exon 6
MOD57F MOD57R	TTTAATGCCCATCTCCAACC CGAGAAAGTTCAGACCCAGAG	498	Exon 7
MOD58F MOD58R	AGGAGATGGGAGCTATGGTG AACAAACAGGGAGCCTCAGAA	337	Exon 8
MOD59F MOD59R	GCTGCTCTTTGCTGGTTGAG TGAGAGTGGATTGTCTGAGGTG	299	Exon 9

**Table 7:** Sequence, target size (bp) and amplified exons for *HNF1B*

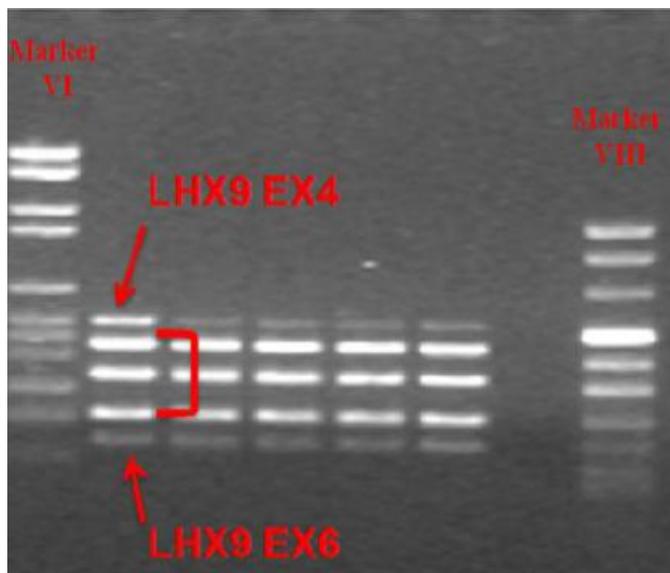
### 4.3.2. Deletion analysis of *NPHP1* gene

For the *NPHP1* analysis the investigation of three large homozygous deletions was performed: *NPHP1* exon 5 (339 bp), *NPHP1* exon 7 (236 bp), and *NPHP1* exon 20 (436 bp) (Figure 18).



**Figure 18:** Schematic view of *NPHP1* and its location on chromosome 2

Each DNA sample was analyzed by multiplex PCR in order to amplify in the same reaction, the three regions potentially delete and two control sequences of the gene LHX9: exon 4 (515bp) and exon 6 (180bp) (Figure 19).



**Figure 19:** Picture of 5 not deleted controls on agarose gel

The PCR was carried out using the AmpliTaq-Gold Polymerase and reagents kit on the GeneAmp PCR system 9700 thermal cycler (Applied Biosystem). The conditions and the program for PCR amplification are reported below (Figure 20).

Primer	Gene-Exon	Oligo Mix %
LHX9_4F	LHX9 Exon4	12.5
LHX9_4R	LHX9 Exon4	12.5
LHX9_6F	LHX9 Exon6	8.33
LHX9_6R	LHX9 Exon6	8.33
NPHP1_20F	NPHP1 Exon 20	8.33
NPHP1_20R	NPHP1 Exon 20	8.33
NPHP1_5F	NPHP1 Exon 5	8.33
NPHP1_5R	NPHP1 Exon 5	8.33
NPHP1_7F	NPHP1 Exon 7	12.5
NPHP1_7R	NPHP1 Exon 7	12.5

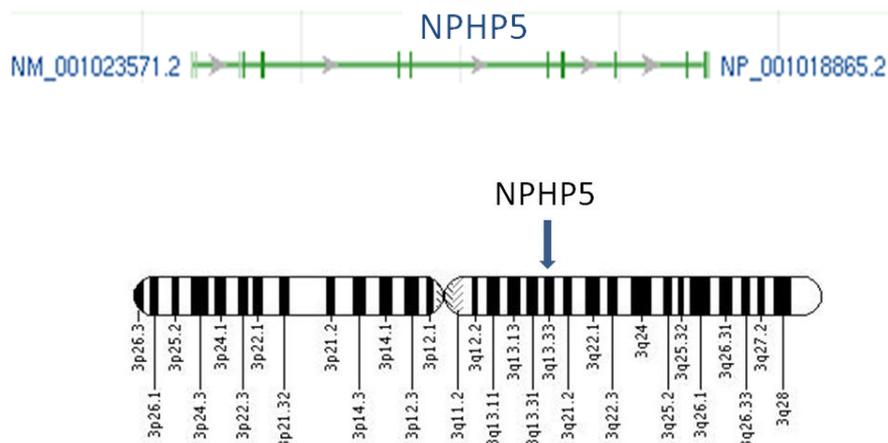
Temperature (°C)	Time (s; m)	Number of cycles
95	10 m	1
95	30 s	} 24
72	30 s	
72	1m	
-0.7°C		
95	30s	} 24
55	30s	
72	1m	
72	7m	1
10	∞	∞

**Figure 20:** Percentage of oligo within the amplification mix and program of amplification

The PCR products were checked by electrophoresis on 2% agarose gel and compared with two different molecular weight markers (DNA Molecular Weight, Markers VI-VIII, Roche Diagnostics GmbH).

### 4.3.3. *NPHP5* analysis

For the analysis of *NPHP5* the DNA samples were subjected to direct sequencing of the coding and flanking regions (Figure 21).



**Figure 21:** Schematic view of *NPHP5* and its location on chromosome 3

PCR were performed using the AmpliTaq Gold DNA Polymerase and reagents kit on the GeneAmp PCR System 9700 thermal cycler (Applied Biosystem) respecting the conditions of amplification of Table 8.

<i>Exon</i>	<i>Annealing Temperature (°C)</i>	<i>Touch-Down</i>	Temperature (°C)	Time (s;m)	Number of cycles
3-4-5-6-7-8-10-11-12-13-14-15	62	No	96	30 s	1
9	62-57	-0.5 °C X 10 cycles	96	10s	} 25
			60	3 m	
			10	∞	∞

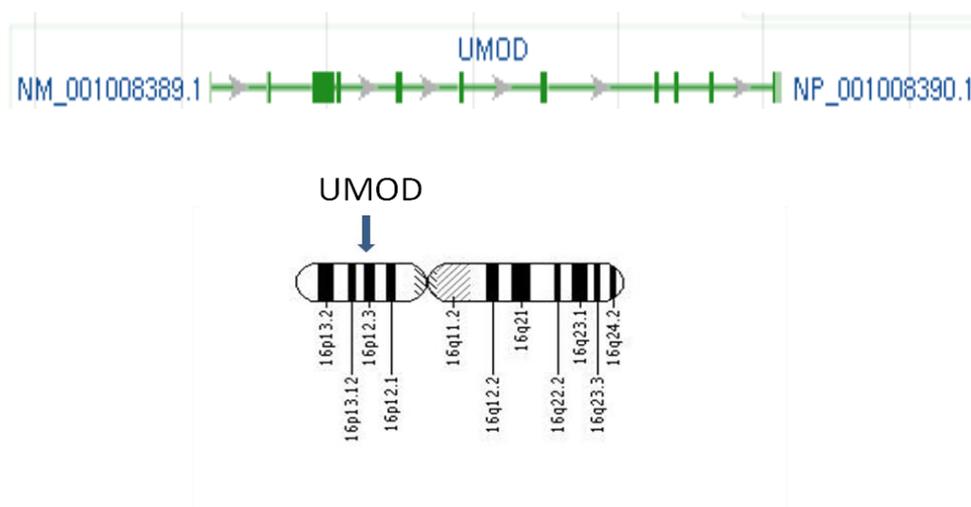
**Table 8:** PCR conditions and program of amplification for *NPHP5* analysis

The PCR products were evaluated by electrophoresis on 2% agarose gel and compared with two markers of molecular weight. Subsequently, the amplified DNA was purified using the QIA quick PCR Purification Kit (250) (QIAGEN Inc.Valencia, CA) The concentration of the purified DNA fragments was assessed by electrophoresis on 2% agarose gel.

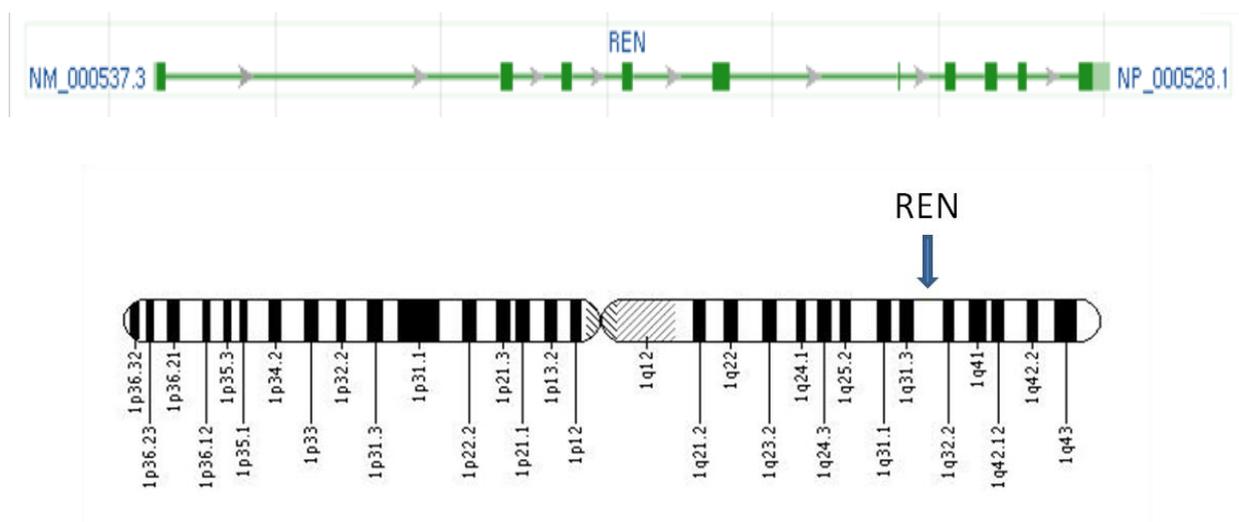
The PCR for sequencing were performed using the BigDye Terminator v1.1 Cycle Sequencing kit (ABI) on the GeneAmp PCR system 9700 thermal cyclcr. PCR products were purified using the Millipore kit plate and loaded on the automatic sequencer 3730 DNA Analyzer, Applied Byosistem (48 capillaries). The results were aligned and compared with a reference sequence by using the Sequencer program (Gene Codes Corp, Ann Arbor, MI).

#### 4.3.4. dHPLC analysis and Sanger sequencing of *UMOD*, *REN* and *HNF1B* genes

The poorly polymorphic fragments of *UMOD* (Figure 22), *REN* (Figure 23) and *HNF1B* (Figure 24) were submitted to the screening of heteroduplex by dHplc. Remaining fragments were submitted to direct sequencing (Table 9).

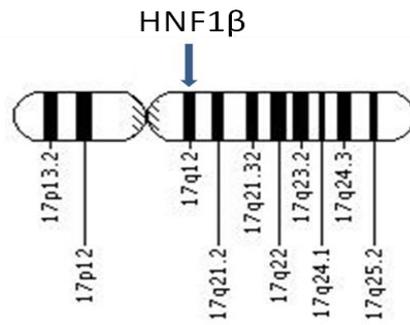


**Figure 22:** Schematic view of *UMOD* and its location on chromosome 16.



**Figure 23:** Schematic view of *REN* and its location on chromosome 1.





**Figure 24:** Schematic view of *HNF1B* and its location on chromosome 17.

Gene	dHPLC	Direct sequencing
<i>UMOD</i>	Exons 1-2-6-7-8-9-10-11	Exons 3- 4- 5
<i>REN</i>	Promoter + Exons 1-2-3-4-5-6-7-8-9-10	/
<i>HNF1B</i>	Exons 2-3-4-5-6-7-8-9	Promoter+ Exon 1

**Table 9:** fragments of *UMOD*, *REN* and *HNF1B* analyzed by dHPLC and direct sequencing.

PCR were performed using the AmpliTaq Gold Polymerase and reagents kit on the GeneAmp PCR System 9700 thermal cycler (Applied Biosystem) respecting the conditions of amplification reported in the following tables (Table 10, Table 11 and Table 12).

### **UMOD**

<b>Fragment</b>	<b>Annealing Temperature (°C)</b>	<b>Touch-Down</b>
Exons 1-2-4-5-6-7-8-10-11	62	No
Exon 3	62-57	-0.5 °C X 10 cicli

**Table 10:** Annealing temperatures set up for PCR amplification of gene fragments of *UMOD*.

## **REN**

<b>Fragment</b>	<b>Annealing Temperature (°C)</b>	<b>Touch-Down</b>
Promoter-Exons 1-2-3- 4-5-6-8-10	62	No
Exons 7-9	63-58	-0.5 °C X 10 cicli

**Table 11:** Annealing temperatures set up for PCR amplification of gene fragments of *REN*.

## **HNF1B**

<b>Fragment</b>	<b>Annealing Temperature (°C)</b>
Exon 1	60
Exons 2-3-4-5-6-7-8-9	62

**Table 12:** Annealing temperatures set up for PCR amplification of gene fragments of *HNF1B*.

The PCR products were evaluated by electrophoresis on 2% agarose gel and compared with two markers of molecular weight.

The amplified samples to analyze by dHPLC were denatured for 10 minutes at 95°C and partially renatured for 10 minutes at room temperature. The optimal temperatures for dHPLC analysis were set up using the Wave Marker dHPLC software. The partially denaturing temperatures for each gene are reported in the following tables (Table 13, Table 14, Table 15).

## UMOD

Fragment	T1 (°C)	T2 (°C)	T3 (°C)	T4 (°C)
<b>Exon 1</b>	57.7	58.2		
<b>Exon 2</b>	57.8	58.9	60.3	
<b>Exon 6</b>	60.9	61.7		
<b>Exon 7</b>	58.6	61.5	62.4	63.3
<b>Exon 8</b>	56.1	59.0	60.2	
<b>Exon 9</b>	60.4	60.9	62.2	
<b>Exon 10</b>	57.3	58.6	59.7	
<b>Exon 11</b>	58.5	59.7	61.6	63.0

**Table 13:** Partially denaturing temperatures for dHPLC analysis of *UMOD*

## REN

Fragment	T1 (°C)	T2 (°C)	T3 (°C)	T4 (°C)
<b>Promoter</b>	61	61,4	-	-
<b>Exon 1</b>	59,4	61,6	62,3	-
<b>Exon 2</b>	59,5	60,3	61	62,4
<b>Exon 3</b>	62,2	62,7	-	-
<b>Exon 4</b>	61,6	63,8 (+0,5)	-	-
<b>Exon 5</b>	61	62,2	-	-
<b>Exon 6</b>	61,1	61,6	-	-
<b>Exon 7</b>	59,9	60,7	61,1	
<b>Exon 8</b>	61,8	62,4	-	-
<b>Exon 9</b>	58,8	61,9	63,1 (+0,5)	-
<b>Exon 10</b>	60,8	62,5	63,3 (+0,3)	-

**Table 14:** Partially denaturing temperatures for dHPLC analysis of *REN*.

## **HNF1B**

Fragment	T1 (°C)	T2 (°C)	T3 (°C)	T4 (°C)
Exon 2	59.2	61.1	62.3	
Exon 3	57.5	59.0	61.3	63.0
Exon 4	62.1	62.9		
Exon 5	58.5	59.8	60.7	61.9
Exon 6	56.8	57.7	58.6	59.8
Exon 7	58.6	61.2	62.4	64.1
Exon 8	60.9	61.9		
Exon 9	60.0	61.5	62.4	63.7

**Table 15:** Partially denaturing temperatures for dHPLC analysis of *HNF1B*.

Positive samples by dHPLC were submitted to sequence analysis.

The PCR products to be sequenced were purified using the QIA quick PCR Purification Kit (250) (QIAGEN Inc.Valencia, CA) The concentration of the purified DNA fragments was assessed by electrophoresis on 2% agarose gel.

The PCR for sequencing were performed using the BigDye Terminator v1.1 Cycle Sequencing kit (ABI) on the GeneAmp PCR system 9700 thermal cycler.

PCR products were purified using the Millipore kit plate and loaded on the automatic sequencer 3730 DNA Analyzer, Applied Biosystem (48 capillaries). The results were aligned and compared with a reference sequence by using the Sequencer program (Gene Codes Corp, Ann Arbor, MI).

### **4.3.5. MLPA analysis of *HNF1B* gene**

The analysis of deletions/duplications of *HNF1B*, was performed using the MLPA P241-D1 MODY kit (MRC-Holland). The P241-D2 MODY probemix contains probes for GCK, HNF1A, HNF1B and HNF4A genes and is therefore specific for MODY 1, 2, 3 and 5. Four healthy controls were included for each session of analysis. The protocol is divided into five phases: 1)Denaturation, 2) Hybridization, 3) Ligation, 4) Amplification, 5) Separation and data analysis.

#### **1) Denaturation**

All DNA samples were normalized in MilliQ water to a concentration of 20 ng / $\mu$ L, transferred within Tube-Strip and denatured at the temperature of 98 ° C for 10 minutes on the Gradient Master Cycler 5331 (Eppendorf).

## **2) Hybridization**

3 µl of Hybridization Mix were added to each sample. The Hybridization Mix was prepared with 1,5 µl of Salsa Probe Mix and 1.5 µl of MLPA Buffer for each sample. The reaction was carried out on the thermal cycler at 95 °C for 1 minute and at 60 °C over night.

## **3) Ligation**

32 µl of Ligase Mix were aliquoted to each sample. For each sample the mix was prepared using 3 µl of MilliQ water, 3 µl of Buffer A, 3 µl of Buffer B and 1 µl of 65 Ligase. The phase of the ligation was carried out at the temperature of 54°C for 15 minutes and at the temperature of 98°C for 5 minutes.

## **4) Amplification**

10 µl of PCR mix prepared by aliquoting 2 µl of SALSA PCR primers and 0.5 µl of SALSA Polymerase were transferred within each Tube-Strip.

PCR reaction was carried out under the following amplifications conditions (Table 16 )

Temperature (°C)	Time (s; m)	Number of cycles
95	30 s	X 35
60	30 s	
72	1 m	
72	20 m	
10	∞	

**Table 16:** Temperature of amplification, time and number of cycles.

## **5) Separation and data analysis**

The MLPA products were subjected to capillary electrophoresis by the use of the Applied Biosystem 3730 DNA Analyzer sequencer. Comparative analysis of samples was performed by Coffalyser.NET.

### **4.4. Next Generation Sequencing of *PKHD1* gene**

#### **4.4.1. Patients selection**

A total of 8 patients were submitted to the molecular screening using the Ion Torrent PGM sequencer. For the validation of the NGS methodology, were included 5 subjects belonging to 4 different families and carrying known mutations in *PKHD1*. These samples were recruited by the Medical Genetics Unit at the S. Orsola-Malpighi University Hospital of Bologna.

Subsequently, an affected proband and his unaffected parents were analyzed. This family was recruited by the Nephrology, Dialysis and Renal Transplant Unit at the S. Orsola-Malpighi University Hospital of Bologna.

#### 4.4.2. Assay Design

The primers design for mutational screening of *PKHD1* gene was performed by the use of Ion AmpliSeq™ Designer. This software allows to design a pool of primers suitable for multiplex PCR. Each primer is synthesized with a 5' modification to allow the partial digestion during the libraries preparation. The software also provide as output the coverage percentage, the DNA concentration for each pool, the length of amplicons and the number of bp to be sequenced. A total of 122 partially overlapping amplicons was provided. The primers were divided into two pools for the construction of two multiplex PCR for sample. A genomic region of 12,245kb corresponding to 99,88% was covered. Only a little region of 15 bp near the 5' of exon 65 was not covered by this design. To amplify this region a couple of primers for conventional sequencing was designed (Table 17).

Amplicon	Pimer Forward	Primer Reverse
PKHD1 ex 65	TTGAACACCTACTATATCCCAGATACT	TTGGGGAAAGAAACAGAATCA

**Table 17:** Primers used to amplify the 5' not covered region of the exon 65.

The result of the design can be viewed directly on the UCSC Genome Browser (Figure 25)

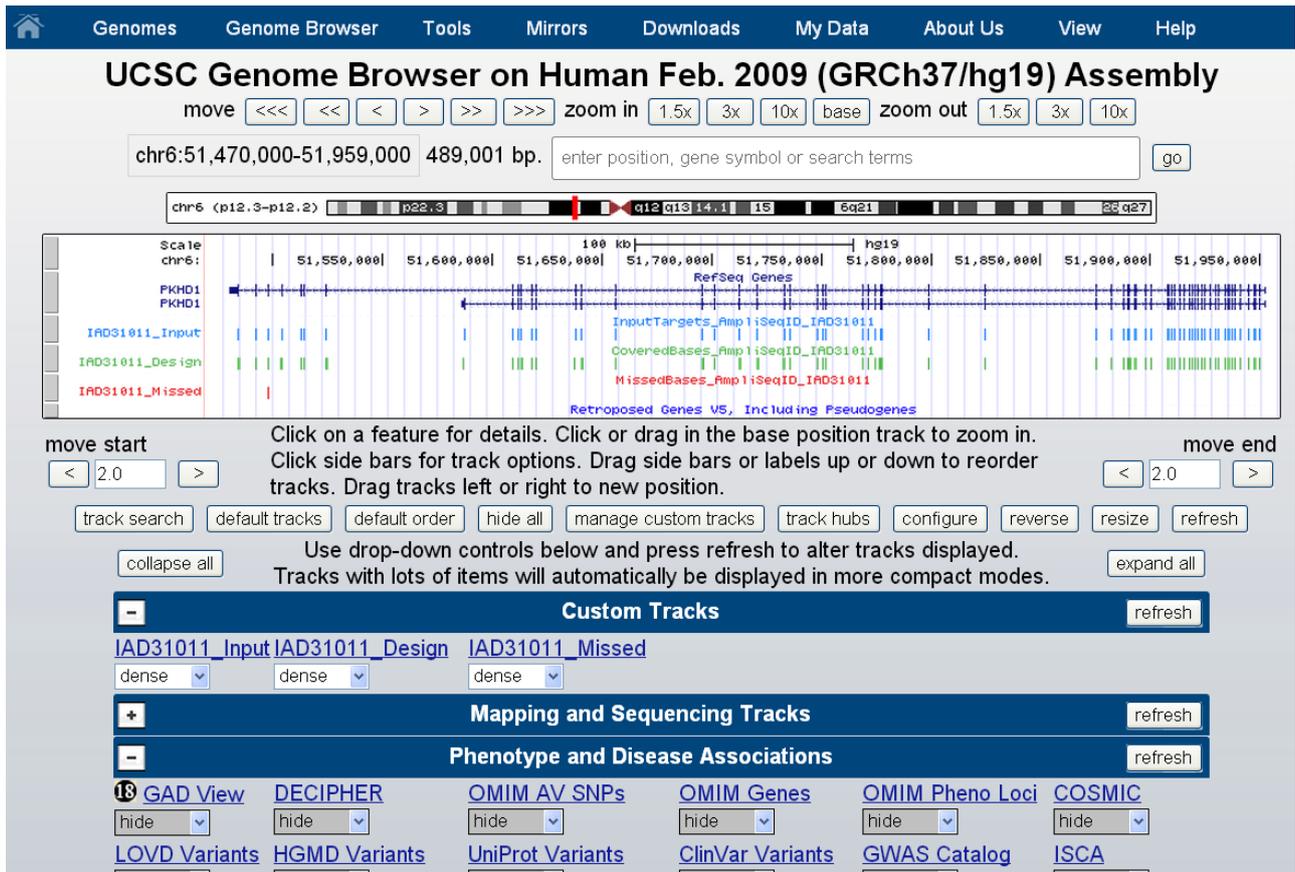


Figure 25: UCSC Genome Browser. *PKHD1* gene is shown in blue, the input sequences are light blue, the output targets are green and missed amplicons are red.

The designed primers are reported in the Table 18.

Amplicon ID	Primer Forward	Primer Reverse	Start	End	Pool
<b>AMPL5757196938</b>	TTCTCAAGGTAACCTAT TGTGTTCTTAC	TGCTCCAATCAAACTG AAAATGCTTTT	51949621	51949801	2
<b>AMPL556530457</b>	CTCATAGTCTTTAGGAT TGTGGGTCAA	TGCAGAAGGTAGTGGT TTGAATCTG	51947911	51948115	1
<b>AMPL574400093</b>	CTCATCCTGTCTGGTCT TCCTATTC	TCTATGCAGCCTGAGAT TTCTTTCC	51947151	51947375	2
<b>AMPL575306224</b>	AACACAAGCACACCCTT AGACTATG	GCTAGCTTTGGGAATTC ATGGTTTT	51944641	51944864	1
<b>AMPL574899571</b>	CCACTCACCTAGGTTTG CAACA	GTAATGCTGTCTGAGG AAAGGCTT	51940999	51941218	2
<b>AMPL585346281</b>	GTAICTCAGCATCAAAAT CAAAAGTTTCCA	ATGACCTGCTTTGCCAT ATTGAGTATA	51938269	51938450	1
<b>AMPL575354245</b>	CCAGCTCATGAGGGAC CATAA	TGTATATGGCTGGATTA TCACTGGAAGAT	51938113	51938326	2
<b>AMPL571859638</b>	AGAATGTGTGTGTGTTG TATCCATGT	AGTTTTGTGAGGAATGT TTATTGGGAGTA	51936840	51937064	1
<b>AMPL574803786</b>	CTTGCAATTGCTTTTTGT CCGGAT	GCTTTCTACTTTCCTGG TTAAAGAGTCA	51935729	51935952	2
<b>AMPL576093592</b>	AGAAAGTAAGCAAGATG AGAGAGATAGGT	GATGATATTGGAGTCTT TGGGCTTATGA	51935127	51935338	1
<b>AMPL573679894</b>	CCAACATTGAGTGAGGC ACAA	CTTTCAGGCAGCCGAC TAAAAA	51934202	51934402	2
<b>AMPL584987583</b>	GTAACCTGGGCAGAATT GTCAAAA	AGGTCATATTCTGGTCT ATATTTGGAAGCT	51930782	51930933	1
<b>AMPL573053414</b>	CAAAATCCCACCTGGTC TCAAC	AGAACAAACATCACAAT TACAGGAGACTT	51930613	51930834	2
<b>AMPL585449128</b>	CCTGGTGGCTTCCTTAC CT	TGTATACCTACACACAC ACACATACAT	51929736	51929897	1
<b>AMPL585434545</b>	AAGCCAGTCAAATAGAT GAGAAAGAAT	GGAAAAGATGTGAGGC TCACCA	51929643	51929786	2
<b>AMPL575263979</b>	CTTTTGATGTGGGCTCC CACT	CGCCCTGTAATGTTTAT GCGATCT	51927281	51927495	1
<b>AMPL575226474</b>	ATGGGTATGGGACTGG CAAC	GTTTCAGTGATTCAGGC CTTGTTA	51924674	51924895	2
<b>AMPL574783531</b>	GCTATCCCATGATGCTC TGCT	GATGCCTGGAAGCTGC ATAGTATT	51923248	51923450	1
<b>AMPL556230311</b>	GCTCCATGGGACTGGA	TGGAGCCATGTACTAC	51923092	51923289	2

	AAGAG	CTGGA			
<b>AMPL575433627</b>	GCATAAAGACCACCCCC AGTT	TGTGTTCTTAGATTGTC TCTCTGGCTTA	51921627	51921847	1
<b>AMPL573311044</b>	AATCAGAATGAAGCCAC GGACA	CCATCTGGTGGGAGAC TTCTTC	51921470	51921694	2
<b>AMPL572981916</b>	CCTACCCACCTGACCCA GAA	GCCTGCCTTTCTATTTT ATCTGAACTG	51920345	51920561	1
<b>AMPL574047440</b>	CCTCCCAGCTGACTGAA TTCC	GACTGCTTCTTGCACTA ATAGAACTGA	51918788	51919012	2
<b>AMPL573589382</b>	ACCTGTTACGTTTGTGT CTGCAA	CAGTTCTGCTCCATGCT GTTTTTC	51917872	51918075	1
<b>AMPL585249067</b>	AAAGGAGGATCACCTGT TCTCT	GGGCCTGTTCTATGTG GATGAAATTATTA	51917764	51917923	2
<b>AMPL574989935</b>	GCGAGCCGTTCCAGAAT CAG	TGTGTAAGTGAATCTGA GTAAGTCAAT	51915065	51915196	1
<b>AMPL574964807</b>	CCTCAAGGCCAACAAGC ATT	GGATCACTGGTCTCTA GTTTCTCAAG	51914896	51915110	2
<b>AMPL574204858</b>	GTGAGTGAGAATATGTG AGTGAGAATTGT	CCCCAACCCAGACGTT AATACT	51913238	51913459	1
<b>AMPL572624939</b>	AAGGTCCACACGTGTTT GTAG	TGAGGATGAAACTCTGT AAGGTGGATTA	51910845	51911055	2

**Table 18:** AmpliSeq Ion Designer output. The Amplicons ID, the sequences of forward and reverse primers, the chromosome position of the start and the end and the reference primers pool are reported.

#### 4.4.3. DNA Libraries construction.

The DNA libraries were constructed using the Ion AmpliSeq™ Library Kit 2.0 (Life Technologies, CA). The workflow consists of five phases:

##### 1) Amplification

A multiplex PCR is prepared for the amplification step. Reagents and conditions of amplification are reported in the Table 19 and Table 20.

5x Ion Ampliseq™ HiFi Master Mix	2µl
2X Ion AmpliSeq™ Primer Pool 1 (o Pool2)	5µl
gDNA (20ng)	1µl
H2O Nuclease-free	2µl
<b>Total volume</b>	<b>10µl</b>

**Table 19:** Reagent mix for sample used for PCR amplification.

Temperature (°C)	Time	Number of cycles
99	2'	1
99	15"	19
60	4'	
10	Hold	

**Table 20:** Program of amplification

## 2) Primers digestion

1 µl of Fupa Reagent was added to each sample. This reagent allows the partial digestion of the primers and the amplicons phosphorylation. The conditions of incubation on thermal cycler are reported in the Table 21.

Temperature (°C)	Time	Number of cycles
99	2'	1
99	15"	19
60	4'	
10	Hold	

**Table 21:** Program of incubation on thermal cycler.

### 3) Adapters and barcodes ligation and purification

5 µl of the mix shown in Table 22 were added to each amplified sample. Subsequently, the samples were incubated on the thermal cycler according to the conditions described in Table 23.

Switch solution	2 µl
Ion P1 adapter	1 µl
IonXpress™ Barcode	1 µl
DNA Ligase	1 µl
H2O Nuclease-free	1 µl
Total volume	5 µl

**Table 22:** Reagent mix prepared for the adapters and barcodes ligation.

Temperature (°C)	Time
22	30'
72	10'
10	Hold

**Table 23:** Incubations program on thermal cycler.

The obtained fragments were purified through a system of magnetic beads, which bind (and then "select") only fragments of more than 100bp. The library of purified fragments was recovered after two washes with ethanol 70%.

### 4) Libraries quantification

The libraries obtained from the previous steps were quantified by qRT-PCR. The quantification was performed by use of the IonQuantification® kit (Life Technology, CA). Three serial dilutions of E.coli DH10B Ion Control Library (68pM) were prepared: 6.8 pM, 0,68 pM and 0,068 pM. This step is necessary for the calibration curve construction. The libraries were diluted 1:100. The qRT-PCR reagents mix was prepared as shown in Table 24 and samples were incubated on StepOne™ Real-Time PCR System (*Applied Biosystems*) according to the conditions shown in the Table 25.

2X Ion TaqMan® Mastermix	10 µl
20X Ion TaqMan® Assay	1µl
Library (1:100) (o standard DH10B (6,8pM, 0,068pM, 0,068pM))	9µl
Total volume	10µl

**Table 24:** Qt-pcr Reagents mix.

Temperature (°C)	Time	Number of cycles
50	2'	Hold
95	20"	Hold
95	1'	40
60	20"	

**Table 25:** Program of incubation on StepOne™ Real-Time PCR System.

### 5) Libraries combination

The different libraries were combined in a single reaction tube at the same concentration of 8 pM.

#### 4.4.4. Template preparation

The preparation of the templates was carried out using the Ion OneTouch™ 200 Template Kit v2 DL (*Life Technology, CA*).

The emulsion PCR was prepared mixing the reagents shown in the Table 26 with mineral oil on the reaction filter. The reaction filter and the amplification plate were loaded on the Ion OneTouch™ 200 (*Life Technology, CA*).

<b>Ion OneTouch™ 2X Reagent Mix</b>	500 µl
<b>Ion OneTouch™ Enzyme Mix</b>	100 µl
<b>Ion OneTouch™ 200 Ion</b>	
<b>Sphere™ Particles</b>	100 µl
<b>Libraries</b>	20 µl
<b>H2O Nuclease-free</b>	280 µl
<b>Total volume</b>	1000 µl

**Table 26:** Emulsion mix

The emulsion PCR products were recovered in two Low-Bind Tube from which the supernatant solution was removed. The two pellets were combined into a unique tube with a volume of 1000 µl of Ion OneTouch™ Wash Solution.

The amplification quality was checked using the Quality Control of Ion OneTouch™200 Ion Sphere™ Particles kit (*Life Technology, CA*). The samples were prepared in according to the conditions shown in the Table 27 and the Table 28.

<b>Annealing Buffer</b>	19 µl
<b>Ion Probe</b>	1 µl
<b>Unenriched sample</b>	2 µl
<b>Total volume</b>	22 µl

**Table 27:** Quality control mix

<b>Temperature (°C)</b>	<b>Time</b>
<b>95</b>	2'
<b>37</b>	2'

**Table 28:** Quality control program on thermal cycler.

Before reading of samples, three washing were executed using 200 µl of Quality Control Buffer. Readings were made using the Qubit 2.0 fluorometer. The percentage yield of linked spheres to the template was obtained from the relationship between the readings made with the Alexa Fluor® 488 and Alexa Fluor® 647. A result between 10-30%

confirms the success of the experiment and the presence of a low amount of polyclonal spheres.

The enrichment process of the emulsion PCR product was performed by the use of the Ion PGM™ Enrichment Beads kit on the Ion OneTouch ES machine. The 8 wells on the machine were filled in order with:

1. 100 µl of emulsion PCR product
2. 130 µl of DynaBeads®MyOne™Streptavidin
3. 300 µl of Ion One Touch™ Wash Solution
4. 300 µl di Ion One Touch™ Wash Solution
5. 300 µl di Ion One Touch™ Wash Solution
6. Empty
7. 300 µl Melt-Off solution (125 mM NaOH + 0,1% Tween 20
8. Empty

During the enrichment, the mechanical arm automatically moves the sample in the different wells, then releasing the enriched sample into a cuvette at the end of the procedure. Subsequently, 5 µl of Ion Sphere™ Particles and 100 µl of Annealing Buffer were added to the half of the enriched sample. After centrifugation and removal of supernatant, 3 µl of sequencing Primers were added to the sample.

#### **4.4.5. Sequencing**

Two consecutive sequencing sessions were performed using in order the Chip 314 v1 during the first session and the Chip 314 v2 during the second session. Each chip was loaded with sample and centrifuged to allow the polymerase and the sphere bounding template to settle in the micro-wells. All the sample solution was removed from chip before the sequencing run on the Ion Torrent PGM.

#### **4.4.6. Sequence Analysis**

The Ion Torrent Suite output files were analyzed using the Ion Reporter™ and Ingenuity online software. The Ion Reporter™ software allows to attribute the chromosomal locations, the protein variations, the type of inheritance, and the number of reads for each variant. The Ingenuity software provides some additional bioinformatics tools, such as SIFT, dbSNPs and Polyphen 2.0 for prediction of the variant pathogenicity.

#### **4.4.7. Validation by Sanger sequencing**

All new mutations found during the NGS analysis were checked by Sanger sequencing. The PCR primers were designed using the Primer3 Software (4.3.1). The primers sequences and the annealing temperatures for PCR amplification are reported in the Table 29.

Amplicon	Pimer forward	Primer Rreverse	Ta
Exon 54 (R2840G)	TGCATGACCATTGACTCCTC	TTGCCTAAAAGGGTGTTTGG	60°C
Exon 61A (c.10856delA)	CACCCACTCTGGTTCAGTCA	GCTCCTTACTGTTGGCGAAT	62°C
Exon 61B (R3482C)	CCATTTGGGCAATTCAGAAG	CCATTTGGGCAATTCAGAAG	58°C
Exon 30 (Y1136C)	TGCATAGGGGTGACTGTGAA	GGGAAAAGAAAATACCTAACTCAA	58°C
Exon 65 (R3842L)	TTGAACACCTACTATATCCCAGATACT	TTGGGGAAAGAAACAGAATCA	60°C

**Table 29:** Primers sequences for PCR amplification

The PCR were carried out in according with the conditions reported in the Table 30 and the Table 31.

Buffer 10x	2,5 µl
MgCl <sub>2</sub> (2mM)	2 µl
dNTPs	4 µl
Primers forward and reverse	0,5 µl
Taq Gold	0,3 µl
Total volume	25 µl

**Table 30:** PCR mix for each sample

Temperature (°C)	Time	Number of cycles
95	10'	1
95	30"	35
See Table 28	30"	
72	30"	
72	7	1

**Table 31:** Amplification program on thermal cycler.

The PCR products were evaluated by electrophoresis on 2% agarose gel and compared with two markers of molecular weight. Subsequently, the amplified DNA was purified using the QIA quick PCR Purification Kit (250) (QIAGEN Inc.Valencia, CA) The concentration of the purified DNA fragments was assessed by electrophoresis on 2% agarose gel.

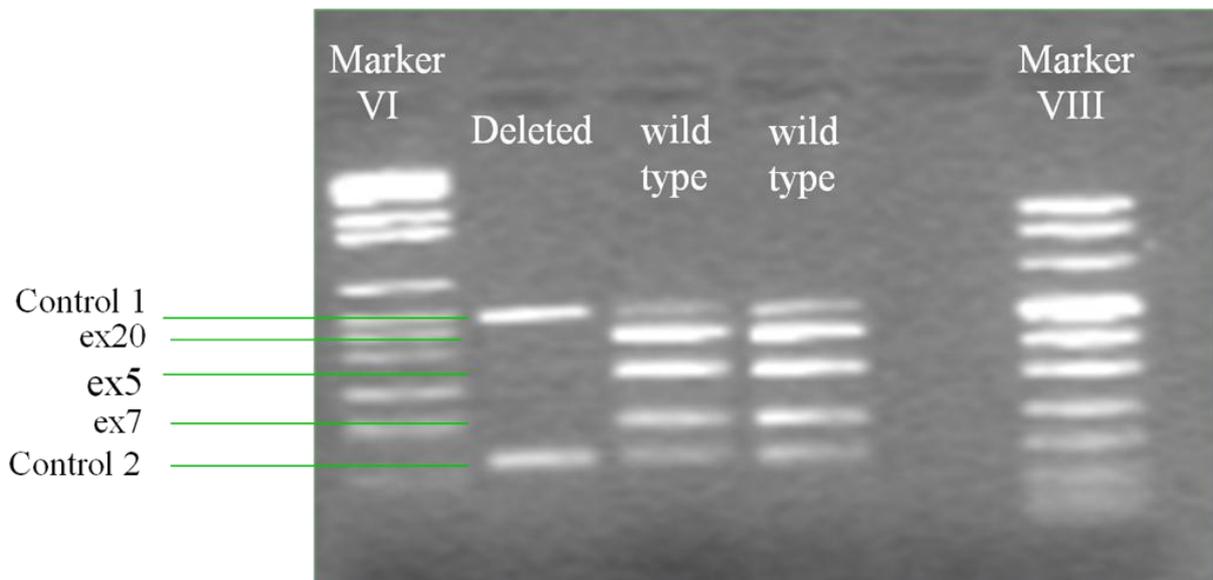
The PCR for sequencing were performed using the BigDye Terminator v1.1 Cycle Sequencing kit (ABI) on the GeneAmp PCR system 9700 thermal cycler.

PCR products were purified using the Millipore kit plate and loaded on the automatic sequencer 3730 DNA Analyzer, Applied Biosystem (48 capillaries). The results were aligned and compared with a reference sequence by using the Sequencer program (Gene Codes Corp, Ann Arbor, MI).

## 5. Results and Discussion

### 5.1. *NPHP1* deletions

A total of 32 patients with medullar kidney cysts were submitted to the deletion analysis of *NPHP1*. In 3 patients the three homozygous deletions of the exons 5, 7 and 20 were identified. The Figure 26 shows the absence of three bands in the positive sample.



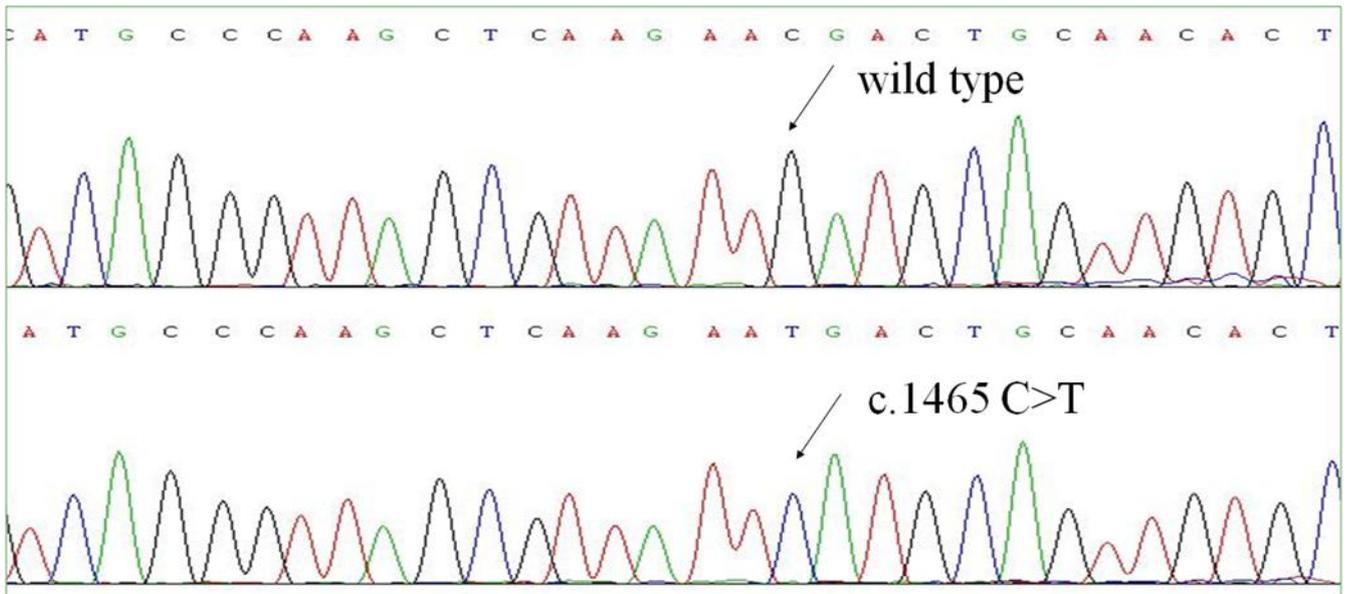
**Figure 26:** 2% gel electrophoresis. Homozygous deletions in affected patient compared with two wild type samples.

These results confirm that the homozygous deletions of *NPHP1* are the most frequent genetic cause of NPH.

### 5.2. *NPHP5* mutations

The sequence analysis of *NPHP5* was performed in 5 patients with retinitis pigmentosa in presence of extrarenal symptoms.

The homozygous c.1465 C>T (p.R489X) non-sense mutation in *NPHP5* gene was detected in a patient with retinitis pigmentosa and recurrent cholangitis, confirming the diagnosis of Senior-Locken Syndrome (Figure 27).

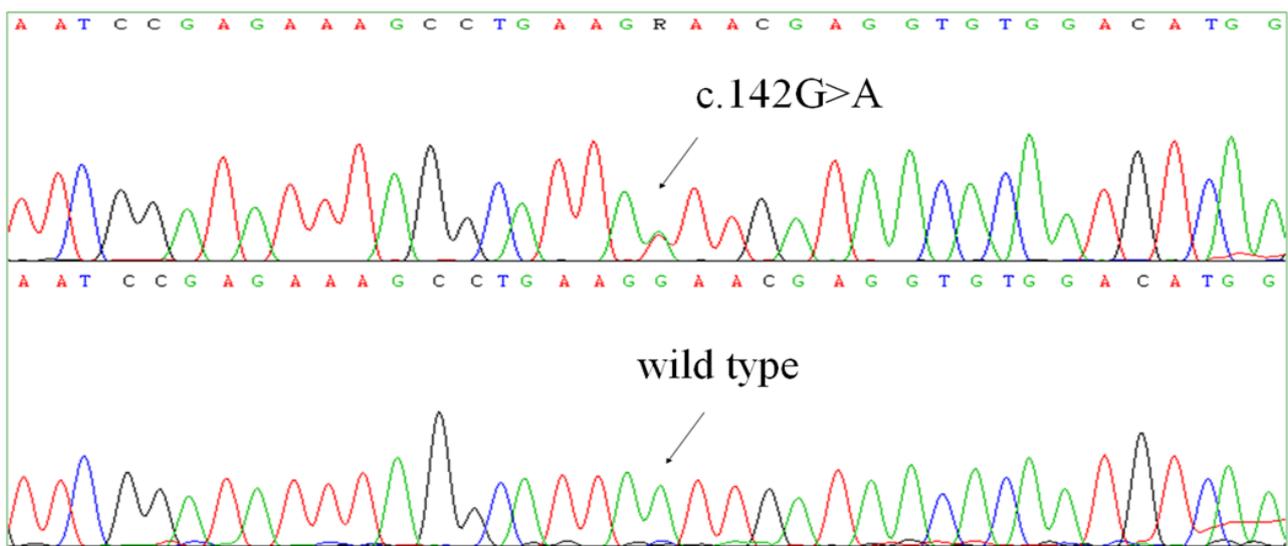


**Figure 27:** electropherogram of the mutated sample in comparison with a wild type sample

This mutation was previously reported in only one Pakistani family by *Otto E A et al* (2008). The mutation c.1465 C>T (p.R489X) is a transition of a cytosine to thymine, which causes the replacement of the 489 arginine in a stop codon. This substitution leads to the production of the truncated not-functional Nephrocystin.

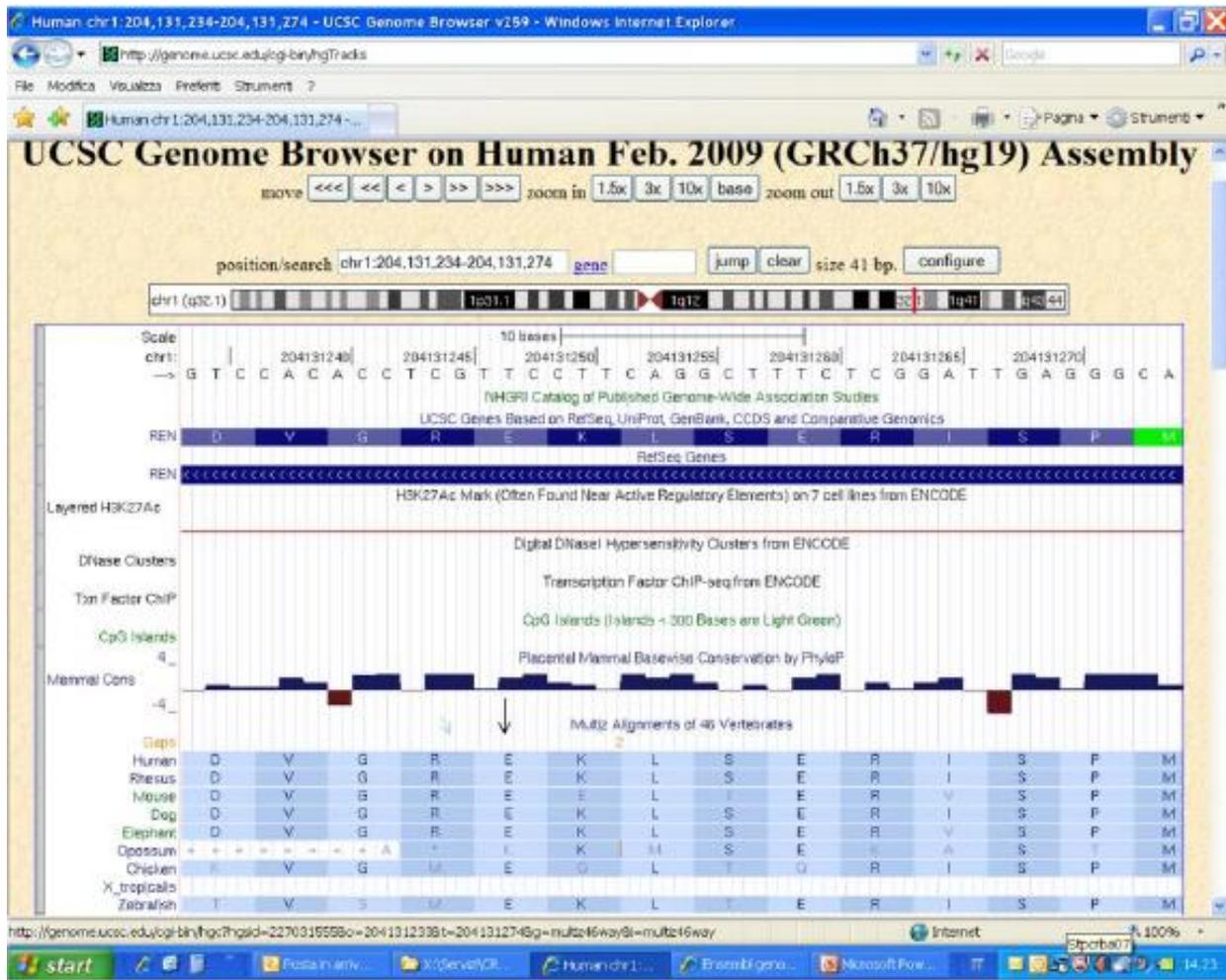
### 5.3. *REN* mutations

The *REN* gene mutation analysis was performed in 25 patients with renal cysts and with an age of onset less than 20 years. The novel heterozygous c.142G>A (p.E48K) missense variant was found in a patient with cystic nephropathy, hyperuricemia, hyperkalemia and anemia (Figure 28).



**Figure 28:** electropherogram of the mutated sample in comparison with a wild type sample.

The c.142G>A is a transition of a guanine in adenine, which causes the modification of the the 142 glutamic acid in lysine. The mutation lies in a conserved position (Figure 29), co-segregated with affected family members and was absent in 50 chromosomes.



**Figure 29:** UCSC Genome Browser conservation score. The E48 amino acid lies in a evolutionary conserved position.

#### 5.4. *HNF1B* mutations

Molecular analysis of *HNF1B* was performed in 31 patients with diabetes and renal failure. Two large rearrangements of *HNF1B* gene were detected in two patients with the clinical features of RCAD. Identified deletions were absent in parents and relatives confirming to be *de-novo* genetic defects.

The first whole gene heterozygous deletion of *HNF1B* was detected in a 17 years old patient with medullary cysts and renal hypoplasia (Figure 30 ).

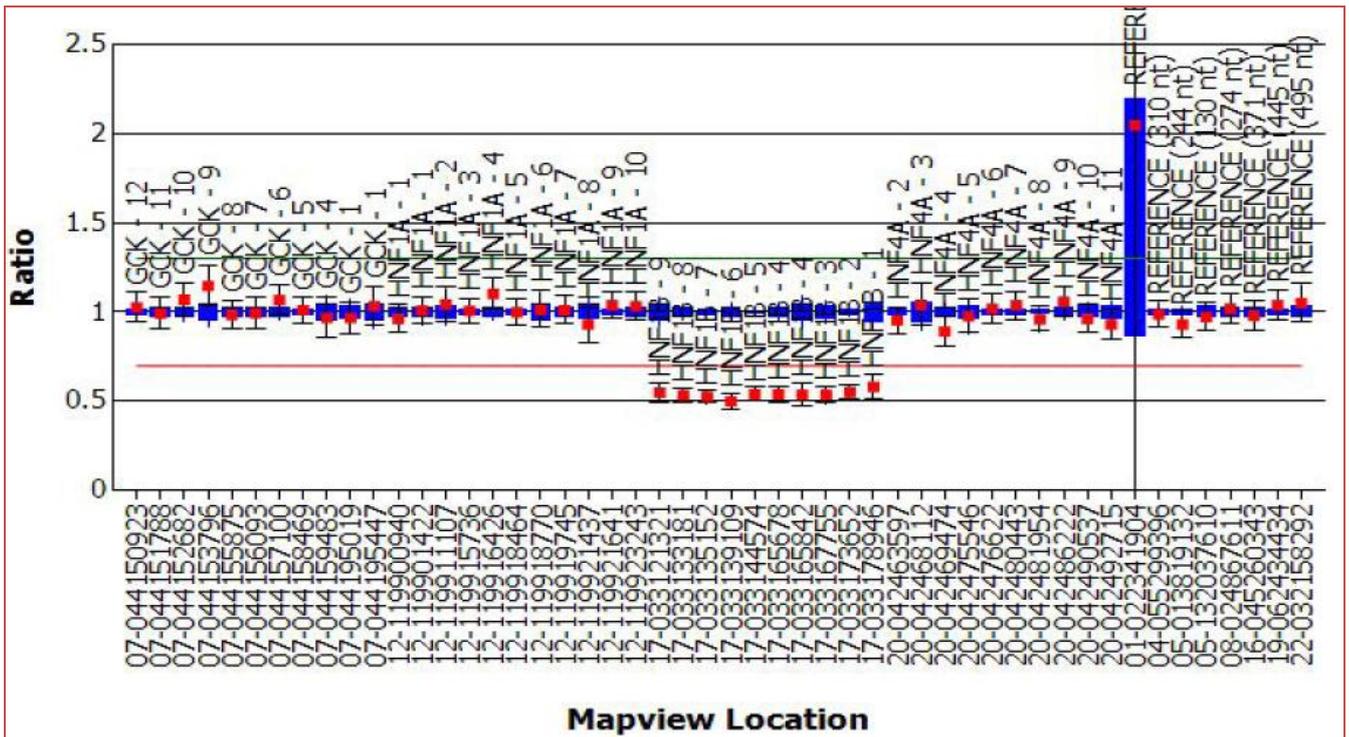


Figure 30: Coffalyser.NET analysis output.

A second whole gene heterozygous deletion of *HNF1B* was found in a 23 years old patient with diabetes and liver disease (Figure 31).

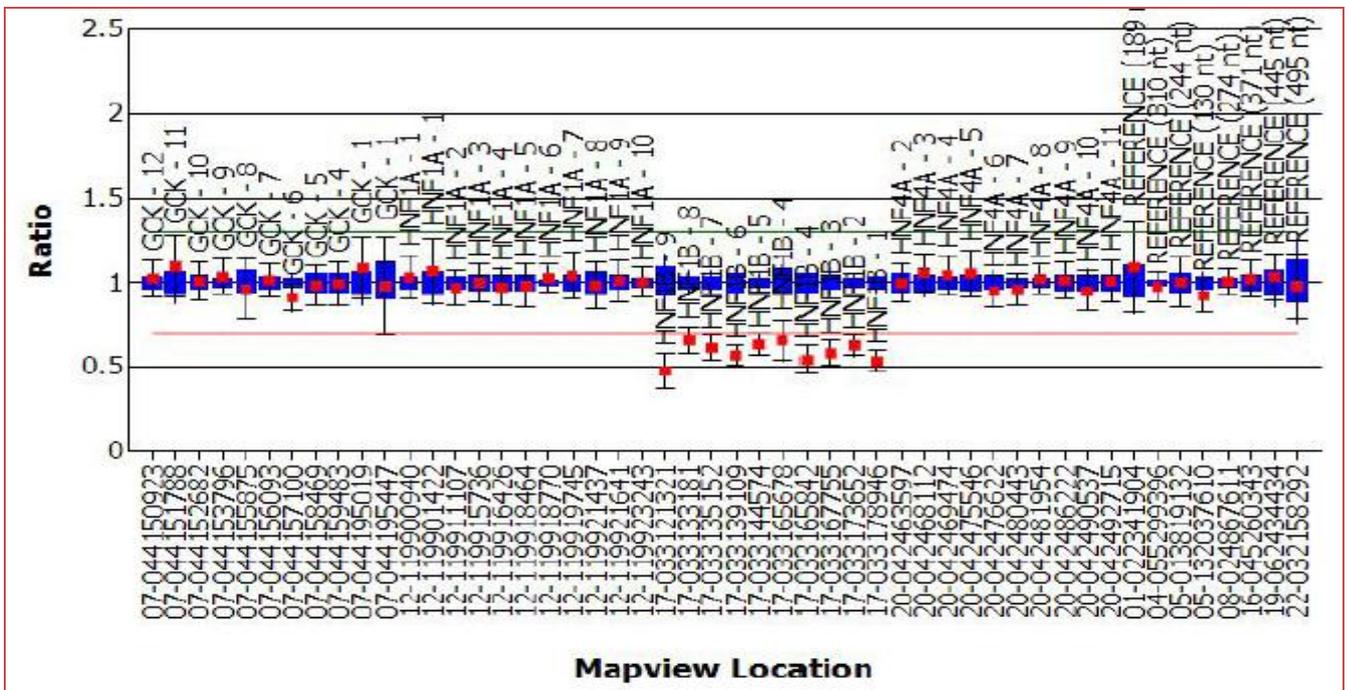
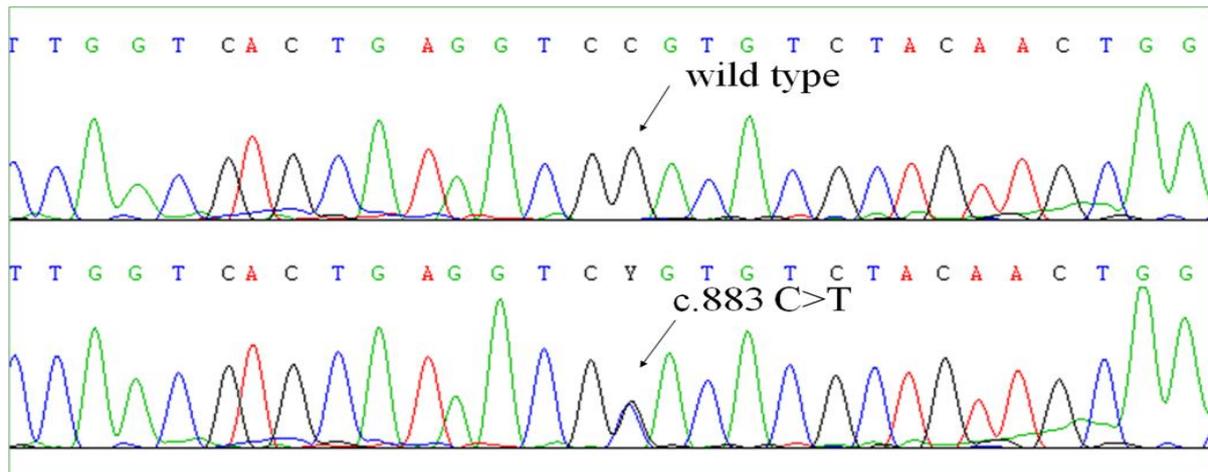


Figure 31: Coffalyser.NET analysis output.

The heterozygous c.883 C>T (pR295C) variant of *HNF1B* was identified in a patient with renal failure and diabetes, confirming the diagnosis of MODY5 (Figure 32).



**Figure 32:** electropherogram of the mutated sample in comparison with a wild type sample.

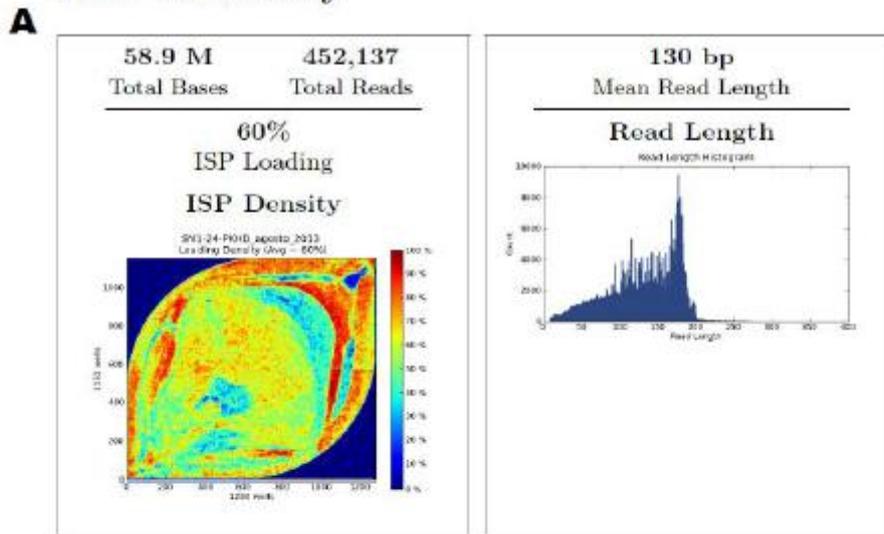
The mutation c.883 C> T is a transition of a cytosine to thymine, which causes a modification of the 295 arginine in cysteine. This substitution leads to the alteration of the transcription factor HNF1 $\beta$  DNA binding domain. This mutation, not previously reported in Italy, was described for the first time in a french study (*Bellanne-Chantelot C, 2005*).

## 5.5. Next Generation Sequencing output

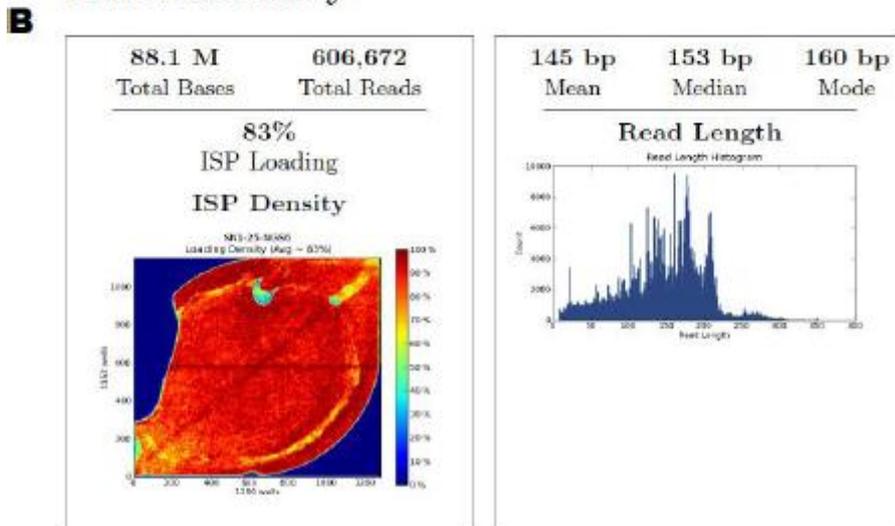
The NGS analysis was carried out in two subsequent experiments: the first experiment was performed with the aim to validate the method and the second experiment as first molecular diagnostics application. A 314.v1 chip (first generation) was used during the first run and a 314.v2 chip (second generation) was used during the second run.

The Torrent Suite output contains a summary of the main run features (Figure 33). The second generation performance are best compared to the first generation chip.

## Run Summary



## Run Summary



**Figure 33:** Torrent Suite software output. 314v.1 chip (A) and 314v2 chip (B). The loading efficiency, the spheres density (ISPs), the total number of sequenced bases and the number of reads are shown in the left panels. The reads length is shown on the right panels.

It took only 3 days of work to complete the two experiments. A total of 8 patients have been sequenced with each chip. The NGS method was much faster compared with the conventional sequencing method. A minimum of 4 weeks to many months would be required to sequence the same number of patients with the Sanger sequencing method.

### 5.5.1. Annotation and filtration of variants

A total of 40 variants have been detected during the experiments. All variants are listed in the Ion Reporter output (Table 30).

Chromosome position	Type	Ref	Var	cDNA	dbSNP ID	Het	Hom	MAF 1000 Genomes
51483961	SNP	T	C	c.12143A>G	rs9381994	3	3	0,44
51484348	SNP	G	A	c.11786-30C>T	rs9395699	3	1	0,19
51491884	SNP	T	C	c.11696A>G	rs4715227	3	3	0,46
51497503	SNP	C	A	c.11525G>T	rs76572975	1	0	0,01
51524067	DEL	CT	C	c.10856_10856delA	/	2	0	/
51524403	SNP	G	A	c.10521C>T	rs34460237	1	0	0,07
51524480	SNP	G	A	c.10444C>T	rs148617572	2	0	/
51586771	SNP	G	A	c.10156+22412C>T	rs9349593	3	3	0,29
51586772	SNP	C	T	c.10156+22411G>A	rs45451196	2	0	0,11
51613177	SNP	C	T	c.9237G>A	rs765525	4	1	0,41
51637561	SNP	T	C	c.8581A>G	rs150925674	1	0	<0,01
51618170	SNP	T	G	c.8798-19delA	rs112525785	3	5	/
51640642	SNP	G	C	c.8518C>G	rs200432861	2	0	/
51640690	SNP	G	A	c.8470C>T	/	1	0	/
51640751	SNP	C	G	c.8441-32G>C	rs3920621	3	2	0,42
51695647	SNP	A	T	c.8302+12T>A	rs1571084	4	2	0,35
51712492	SNP	A	T	c.8107+81T>A	rs9370067	2	0	0,5
51720838	SNP	T	C	c.7764A>G	rs9349603	3	1	0,48
51720872	SNP	A	G	c.7734-4T>C	rs7452724	3	1	0,48
51732628	SNP	G	A	c.7733+33C>T	rs9382044	3	1	0,48
51732807	SNP	C	T	c.7587G>A	rs12210295	3	4	0,4

51750563	SNP	A	T	c.7215+102T>A	rs4711985	3	1	0,47
51776535	SNP	T	C	c.6490+62A>G	rs12196767	1	0	0,11
51824680	INS	G	GT	c.5895_5896insA	/	1	0	/
51875250	SNP	A	C	c.5608T>G	rs2435322	0	8	0,03
51887643	SNP	T	C	c.5336A>G	/	1	0	/
51889738	SNP	G	A	c.4870C>T	/	1	0	/
51890316	SNP	C	T	c.4292G>A	/	1	0	/
51890823	SNP	G	A	c.3785C>T	rs9296669	5	0	0,4
51891011	SNP	T	C	c.3629-32A>G	rs2499480	4	1	0,44
51893107	SNP	T	C	c.3407A>G	rs41273726	1	0	0,01
51914956	SNP	G	A	c.2278C>T	rs9370096	6	0	0,41
51917968	SNP	T	G	c.2046A>C	rs4715271	5	0	0,07
51923409	SNP	A	T	c.1234-10A>T	rs4715272	6	0	0,11
51924774	SNP	A	G	c.1185T>C	rs1896976	0	8	0,03
51938242	SNP	A	G	c.527+19T>C	rs9474140	4	0	0,32
51938417	SNP	G	A	c.449-78C>T	rs12154128	1	0	0,01
51947237	SNP	G	A	c.234C>T	rs9474143	4	0	0,34
51947257	SNP	G	A	c.214C>T	rs6901799	5	0	0,08
51947999	SNP	G	A	c.107C>T	rs137852944	1	0	<0,01

**Table 30:** Ion Reporter output. For each variant the chromosomal location, the type, the wild type bases in comparison with bases found in samples, the position on cDNA, the rs code if reported in dbSNP and the MAF frequency are reported.

Among all 40 variants two frame-shift mutations (c.10856\_10856delA and c.5895\_5896 INS A), a nonsense (c.8470 C> T) and 37 missense, synonymous or intronic variants were

identified. A further filtering step was carried out in according with the following selection criteria :

- Insertions or deletions;
- Non-sense variant;
- Splicing site variant;
- Missense;
- Presence in HGMD;
- $MAF \leq 0,01$  and absence in 1000 Genomes;
- low detection frequency in analyzed samples;
- Absence in dbSNP.

12 filtered variants were checked using the SIFT and PolyPhen pathogenicity prediction software (Table 31). Among 9 missense mutations, 4 were not present in HGMD, then never described for the disease. The frame-shift and nonsense mutations were not subjected to further investigation because certainly pathogenic.

CDS	Protein	Exon	HGMD	SIFT	PolyPhen
<b>c.107C&gt;T</b>	p.T36M	3	present	Damaging	Probably Damaging
<b>c.3407A&gt;G</b>	p.Y1136C	30	present	Tolerated	Possibly Damaging
<b>c.4292G&gt;A</b>	p.C1431Y	32	not present	Damaging	Benign
<b>c.4870C&gt;T</b>	p.R1624W	32	present	Damaging	Possibly Damaging
<b>c.5336A&gt;G</b>	p.N1779S	33	not present	Tolerated	Probably Damaging
<b>c.5895_5896insA</b>	p.L1966TfsX27	36	present	-	-
<b>c.8470C&gt;T</b>	p.Q2824X	54	not present	-	-
<b>c.8518C&gt;G</b>	p.R2840G	54	not present	Tolerated	Benign
<b>c.8581A&gt;G</b>	p.S2861G	55	present	Tolerated	Benign
<b>c.10856_10856delA</b>	p.K3619SfsX7	61	not present	-	-
<b>c.10444C&gt;T</b>	p.R3482C	61	present	Damaging	Possibly Damaging
<b>c.11525G&gt;T</b>	p.R3842L	65	not present	Tolerated	Probably Damaging

**Table 32:** The CDS location, the protein location, the exon, the presence on HGMD and the predictive analysis on Poliphen and Sift are reported for each mutation.

### 5.5.2. Validation

All mutations previously identified by conventional sequencing have been detected by NGS (Table 33). Moreover this NGS protocol has allowed the identification of two additional variants not previously found by Sanger sequencing.

Fam	Samples	Known Mutations	NGS Results	Sanger Results
1	proband	p.R1624W	p.R1624W	Np
		p.R3482C	p.R3482C	Np
2	proband	p.Q2824X	p.Q2824X	Np
		p.N1779S	p.N1779S	Np
3	father	p.T36M	p.T36M	Np
	mother	p.S2861G	p.S2861G p.R3482C	Np p.R3482C
5	proband	p.C1431Y	p.C1431Y	Np
		p.L1966fsX27	p.L1966Tfs	np
			p.Y1136C	p.Y1136C

**Table 33:** mutations identified using the NGS protocol and the known mutations previously identified by Sanger Sequencing; np = not performed

The probands of the families 1 and 2 show the same mutations previously identified by Sanger sequencing. The NGS of the family 3 allowed to confirm the known mutations and also to identify an additional mutation not previously detected. This result indicates that the NGS method is more comprehensive than the conventional sequencing method. Also in the proband of the Family 4 an additional mutation was identified by NGS. In this family the NGS has proved to be more sensitive and useful for diagnostics application and genetic counseling.

### 5.5.3. Molecular diagnostics application

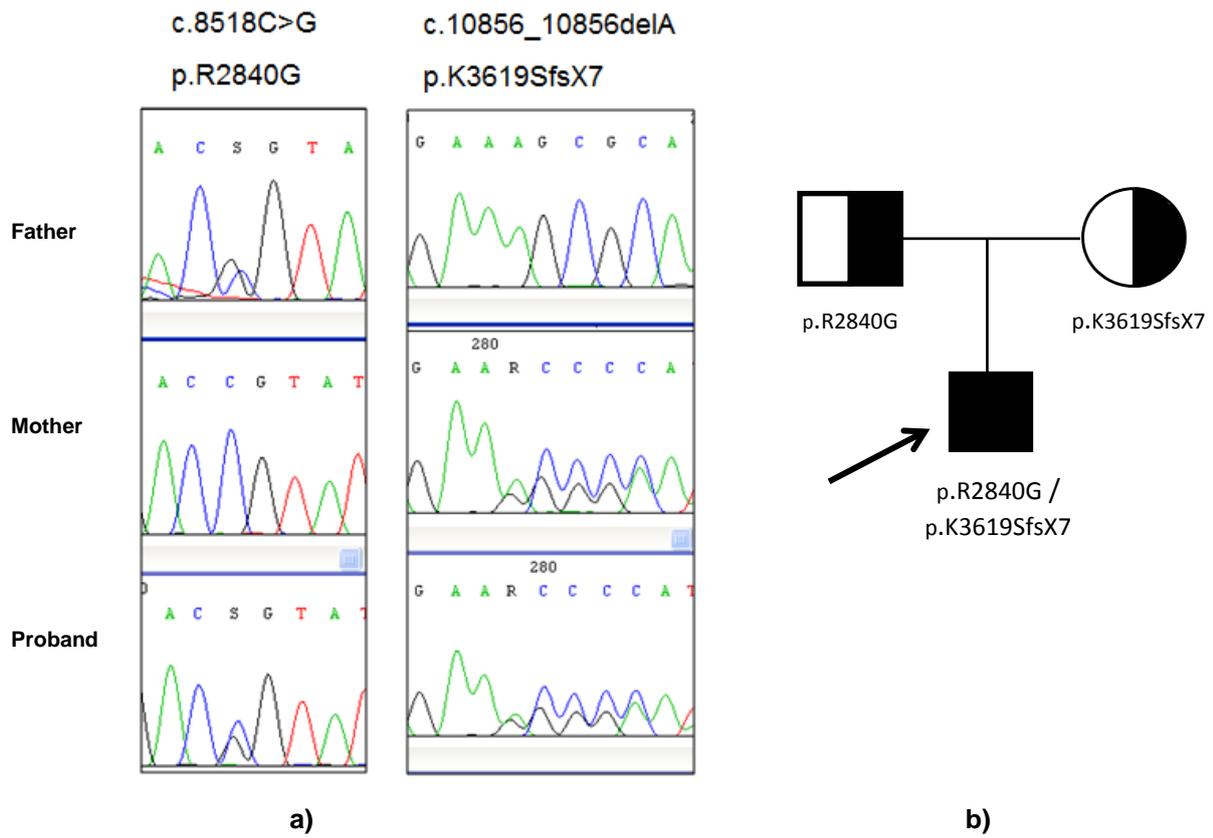
The results obtained in the validation phase of this study, allowed us to proceed with a further molecular analysis of samples not previously characterized. A proband with ARPKD and his healthy parents were submitted to the NGS to evaluate the segregation of the mutations.

The identified variants in the proband and his parents are reported in the Table 34.

Samples	NGS Results	Sanger Results	HGMD	dbSNP	SIFT	Polyphen
<b>Proband</b>	p.K3619SfsX7	p.K3619SfsX7	Not found	rs76572975	_____	_____
	p.R2840G	p.R2840G	Not found	rs150925674		
<b>Father</b>	p.R2840G	p.R2840G	Not found	rs150925674	Tolerated	Benign
	p.R3842L	p.R3842L		rs200432861	Tolerated	Prob. Damaging
<b>Mother</b>	p.K3619SfsX7	p.K3619Sfs	Not found	rs76572975	_____	_____

**Table 34:** The CDS location, the protein location, the exon, the presence on HGMD and the predictive analysis on Polyphen and Sift are reported for each mutation

The frame-shift p.K3619SfsX7 and the missense p.R2840G certainly causative mutations were identified in the proband and respectively in his mother and father. In addition, the father shows a second missense variant p.R3842L not transmitted to the child. The presence of all mutations and the absence of p.R3842L in the proband was confirmed by Sanger sequencing (Figure 34 ).



**Figure 34:** a) electropherograms of the confirmed mutations by Sanger sequencing ; b) variants segregation in the family trio.

## 6. Conclusions

This study provides a contribution to the knowledge of the genetic defects responsible of hereditary cystic kidney diseases, widening the spectrum of known mutations. This genetic screening allows to obtain an accurate molecular diagnosis and to better define the disease outcome for each patient. A careful diagnostic evaluation and an accurate genetic counseling are essential to ensure patient care and could lead to new therapeutic approaches.

The results of this study confirm the validity of the proposed diagnostic algorithm for the conventional molecular screening. This genetic screening gave a molecular diagnosis in about 16% of our patients. The clinical phenotypes of mutation-positive patients were in agreement with the mutated gene. For improving the sensibility of our genetic test, it will be necessary to extend the screening to additional less common genes.

The Ion Torrent PGM system used in this study for NGS screening resulted quite fast allowing to perform the molecular analysis of the *PKHD1* gene in several patients using a single run. Our filtering criteria have been effective and the method resulted more sensitive and comprehensive than the traditional Sanger sequencing, being a powerful tool for molecular diagnostics and genetic counseling.

This molecular screening for inherited cystic kidney diseases can improve the diagnostic capabilities and the genetic counseling for families. Furthermore, the promising performance of NGS suggest the possibility to introduce this technology to clinical practice, extending the analysis to a panel of less common genes involved in cystic diseases.

## Bibliography

- Adeva M, El-Youssef M, Rossetti S. Clinical and molecular characterization defines a broadened spectrum of autosomal recessive polycystic kidney disease (ARPKD). *Medicine (Baltimore)* 2006; 85:1–21.
- Avni FE, Guissard G, Hall M, Janssen F, DeMaertelaer V, Rypens F. Hereditary polycystic kidney diseases in children: changing sonographic patterns through childhood. *Pediatr Radiol* 2002; 32(3):169–174.
- Balogh I, Kappelmayer J, Tózsér J. *Molecular diagnostics 2011. The methodology of molecular diagnostic procedures II. Chapter 13.13.*
- Bellanne-Chantelot C, Chauveau D, Gautier JF, Dubois-Laforgue D, Clauin S, Beaufils S, Wilhelm JM, Boitard C, Noë I LH, Velho G, Timsit J. 'Clinical spectrum associated with hepatocyte nuclear factor-1/3 mutations'. *Ann Intern Med* 2004; 140:510-517.
- Bergmann C. ARPKD and early manifestations of ADPKD: the original polycystic kidney disease and phenocopies. *Pediatr Nephrol.* 2014 Mar 1. DOI 10.1007/s00467-013-2706-2.
- Bisceglia M, Galliani CA, Senger C, Stallone C, Sessa A. 'Renal cystic diseases: a review'. *Adv Anat Pathol* 2006; 13(1): 26-56.
- Bleyer A.J, Hart P.S, Knoch S. 'Hereditary interstitial kidney disease'. *Semin Nephrol* 2010; 30:366–73.
- Bleyer AJ, Hart PS. In: Pagon RA, Adam MP, Bird D, Dolan CR, Fong CT, Smith RJH, Stephens K, editors. *UMOD-Associated Kidney Disease*. GeneReviews. Seattle (WA): University of Washington, Seattle; 1993-2014. 2007 Jan 12 [updated 2013 Sep 12].
- Calado J, Gaspar A, Clemente C, Rueff J. 'A novel heterozygous missense mutation in the UMOD gene responsible for Familial Juvenile Hyperuricemic Nephropathy'. *BMC Med Genet* 2005; Jan 27;6:5.
- Chen Y Z, Gao Q, Zhao X Z, Bennett C L, Xiong X S, Mei C L, Shi Y Q and Chen X M. Systematic review of TCF2 anomalies in renal cysts and diabetes syndrome/maturity onset diabetes of the young type 5. *Chin Med J* 2010;123(22):3326-3333.
- Conti G, Chimenz R, Silipigni L, Chirico V, Vitale A, Fede C. 'Malattie renali cistiche ereditarie-Hereditary cystic kidney diseases'. *Rivista Italiana di Genetica e Immunologia Pediatrica*, 2009.
- Duncan H, Dixon A. 'Gout, familial hyperuricaemia and renal disease'. *Q J Med* 1960; 29: 127 – 136.
- Gerhart U R. Hnf1b *Mus musculus* HNF1 homeobox B. *Transcription Factor Encyclopedia*, updated August 3rd, 2009. <http://www.cisreg.ca/cgi-bin/tfe/articles.pl?tfid=867>.
- Hart TC, Gorry M.C, Hart P.S, Woodard A.S, Shihabi Z, Sandhu J, Shirts B, Xu L, Zhu, H, Barmada M.M, Bleyer A.J. 'Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy'. *J. Med. Genet* 2002; 39: 882-892.

- Hildebrandt F et al. Nephronophthisis: Disease Mechanisms of a Ciliopathy. *J Am Soc Nephrol*, 2009; 20(1): 23-35.
- Hildebrandt F, Otto E, Rensing C, Nothwang HG, Vollmer M, Adolphs J, Hanusch H, Brandis M. 'A novel gene encoding an SH3 domain protein is mutated in nephronophthisis type 1'. *Nat Genet* 1997; 17:149–153.
- Hildebrandt F, Zhou W. 'Nephronophthisis-associated ciliopathies'. *J Am Soc Nephrol* 2007; 18: 1855-71.
- Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, Cockburn BN, et al. Mutation in hepatocyte nuclear factor-1 beta gene (*TCF2*) associated with MODY. *Nat Genet* 1997; 17: 384-385.
- Izzi C, Sottini L, Dallera N, Capistrano M, Foini P, Scolari F. 'Genetica e Nosografia delle malattie renali cistiche'. *G Ital Nefrol* 2010; 27 (S50): S63-S69.
- Jiufeng S, Jin X, Pei L, Qiang M, Yan H, Xiaoli L, Chuanhuan D, Chi L, G S de Hoog, and Xinbing Y. 'Molecular identification of *Clonorchis sinensis* and discrimination with other opisthorchid liver fluke species using multiple Ligation-dependent Probe Amplification (MLPA)'. *Parasit Vectors* 2011; 4: 98.
- Kirby A, Gnirke A, Jaffe DB, Barešová V, Pochet N, Blumenstiel B, Ye C, Aird D, Stevens C, Robinson JT, Cabili MN, Gat-Viks, Kelliher E, Daza R, DeFelice M, Hůlková H, Sovová J, Vylet'al P, Antignac C, Guttman M, Handsaker RE, Perrin D, Steelman S, Sigurdsson S, Scheinman SJ, Sougnez C, Cibulskis K, Parkin M, Green T, Rossin E, Zody MC, Xavier RJ, Pollak MR, Alper SL, Lindblad-Toh K, Gabriel S, Hart PS, Regev A, Nusbaum C, Knoch S, Bleyer AJ, Lander ES, Daly MJ Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in *MUC1* missed by massively parallel sequencing. *Nat Genet* 2013; (3):299–303
- Kleinknecht C et al. Nephronophthisis. *Textbook of Clinical Nephrology*. Oxford University Press, Oxford, 1992; 2188-2197.
- Lhotta K, Gehringer A, Jennings P, Kronenberg F, Brezinka C, Andersone I, Strazdins V. 'Familial juvenile hyperuricemic nephropathy: report on a new mutation and a pregnancy'. *Clin Nephrol* 2009; 71: 80–83.
- Omran H. Nephronophthisis and medullary cystic kidney disease. In: Geary DF, Schaefer F (eds) *Comprehensive pediatric nephrology*, 2008. Elsevier, Amsterdam, pp 143–154.
- Osathanondh V, Potter EL Pathogenesis of polycystic kidneys. Survey of results of microdissection. *Arch Pathol* 1964; 77:510–512.
- Otto E.A, Helou J, Allen S.J, O'Toole J.F, Wise E.L, Ashraf S, Attanasio M, Zhou W, Wolf M.T.F, Hildebrandt F. 'Mutation analysis in nephronophthisis using a combined approach of homozygosity mapping, *cel I* endonuclease cleavage, and direct sequencing'. *Human Mutation* 2008; 29(3),418-426.
- Roslyn J Simms et al. Nephronophthisis. *European Journal of Human Genetics*, 2009; 17, 406–416.
- Salomon R, Saunier S, Niaudet P (2009) Nephronophthisis. *Pediatr Nephrol* 24:2333–2344.

- Scolari F, Caridi G, Rampoldi L, Tardanico R, Izzi C, Pirulli D, Amoroso A, Casari G, Ghiggeri GM. 'Uromodulin storage diseases: clinical aspects and mechanisms'. *Am J Kidney Dis*: 44: 987–999, 2004 Scolari F, Caridi G, Rampoldi L, Tardanico R, Izzi C, Pirulli D, Amoroso A, Casari G, Ghiggeri GM. 'Uromodulin storage diseases: clinical aspects and mechanisms'. *Am J Kidney Dis* 2004; 44: 987–999.
- Sheffield V.C, Beck J.S, Kwitek A.E, Sandstrom D.W, Stone E.M. 'The sensitivity of single strand conformation polymorphism analysis for the detection of single base substations'. *Genomics* 1993; 16: 325-332.
- Thomas CP, Erlandson JC, Edghill EL, Hattersley AT, Stolpen AH. A genetic syndrome of chronic renal failure with multiple renal cysts and early onset diabetes. *Kidney Int* 2008; 74: 1094-1099.
- Tran et al. THM1 negatively modulates mouse Sonic hedgehog signal transduction and affects retrograde intraflagellar transport in cilia. *Nature Genet* 2008; 40: 403-410.
- Vylet'al P, Kublova M, Kalbacova M, Hodanová K, Baresová V, Stibůrková B, Sikora J, Hůlková H, Zivný J, Majewski J, Simmonds A, Fryns JP, Venkat-Raman G, EllederM, Kmoch S. Alterations of uromodulin biology: a common denominator of the genetically heterogeneous FJHN/MCKD syndrome. *Kidney Int* 2006; 70:1155–1169.
- Watnick T, Germino F.. 'From cilia to cyst'. *Nat Genet* 2003; 34: 355-6.