Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE MEDICHE SPECIALISTICHE

Ciclo XIXX

Settore Concorsuale di afferenza: 06/D2

Settore Scientifico disciplinare: MED/14

Dominant and Recessive Polycystic Kidney Disease: a novel molecular diagnostics approach based on Next-Generation Sequencing

Presentata da:
Minardi Raffaella

Coordinatore Dottorato
Prof.re Gaetano Domenico Gargiulo

Relatore:
Prof.re Gaetano La Manna

Correlatore:
Dott.ssa Vilma Mantovani

Esame finale anno 2017

SUMMARY

ABSTRACT **1**

1. BACKGROUND **2**

1.1 Polycystic Kidney Disease (PKD) as Ciliopathy 2

1.1.1 Autosomal Dominant Polycystic Kidney Disease (ADPKD) 2

1.1.1.1 Genetics of ADPKD 4

1.1.1.2 Complex inheritance pattern 5

1.1.1.3 PKD1 and PKD2 genes 8

1.1.1.4 Polycystin 1 - 2 and PC1/PC2 Complex 9

1.1.1.5 Signalling pathway modified by PC1 and PC2 11

1.1.1.6 Cysts formation 13

1.1.2 Autosomal Recessive Polycystic Kidney disease (ARPKD) 14

1.1.2.1 Genetics of ARPKD 16

1.1.2.2 *PKHD1* gene 17

1.1.2.3 Fibrocystin 1/Polyductin (FC1) 18

1.1.2.4 Polycystin complex and Fibrocistin/Polyductin interaction 20

1.1.3 Ultrasonography in ADPKD and ARPKD diagnosis 21

1.1.4 Genetic testing 22

1.2 From Sanger sequencing to Next-Generation Sequencing 23

1.2.1 Sanger sequencing, the conventional sequencing method 23

1.2.2 Next-Generation Sequencing technology 25

1.2.2.1 NGS platform 26

1.2.2.2 Ion Torrent PGM platform 28

2. PRINCIPAL AIM **30**

3. MATERIALS AND METHODS **31**

3.1 Patients recruitment 31

3.2 DNA extraction routine 31

3.3 Next-Generation Sequencing protocol 32

3.3.1 *PKHD1* and *PKD2* libraries construction protocol 32

3.3.1.1 Assay Design 32

3.3.1.2 Libraries amplifications 32

3.3.1.3 Primers digestion 33
3.3.1.4 Adapter and barcodes ligation 33
3.3.1.5 Libraries purification 33
3.3.1.6 Libraries quantification and pooling 34
3.3.2 *PKD1* libraries construction protocol 34
3.3.2.1 Assay design 34
3.3.2.2 LR-PCR purification, quantification and pooling 37
3.3.2.3 Sonication and end repair 37
3.3.2.4 Barcode and adapters ligation, nick repair and purification 37
3.3.2.5 Size selection of DNA libraries and quantifications 38
3.3.3 Sequencing template preparation: Emulsion PCR 38
3.3.4 Template-positive ISPs enrichment 39
3.3.5 Ion Torrent PGM™ sequencing 40
3.3.6 NGS data analysis 41
3.3.6.1 Variant filtering strategy 41
3.4 Validation of the NGS data by Sanger sequencing 42
3.5 Multiplex Ligation-dependent Probe Amplification: MLPA 43

4. RESULTS AND DISCUSSIONS 45

4.1 Critical assay analysis 45
4.2 NGS protocols: retrospective validation 49
4.3 Pathogenic variants identified by NGS tests 52
4.3.1 Molecular screening of *PKD1* and *PKD2* 52
4.3.1.1 Cases with more than one variant in PKD1 59
4.3.2 Molecular screening of *PKHD1* 62
4.4 Time and cost analysis 68

5. CONCLUSIONS 70

Bibliography 72

ABSTRACT

Polycystic Kidney Disease (PKD) is the most common genetic cause of kidney failure in children and adults and can be inherited as an autosomal dominant trait (ADPKD) or an autosomal recessive trait (ARPKD). ADPKD is the most common form, characterized by a late onset, caused by mutation in two causative gene *PKD1* and *PKD2*; while ARPKD represents the rarest and most severe form, with early onset, caused by mutation in *PKHD1* gene. The large size, the molecular complexity, the lack of mutational hotspot characterizing the causative genes and, in particular, the high homology of *PKD1* with six pseudogenes make the molecular diagnostics challenging, time-consuming and expensive when using conventional sequencing methods.

The aim of this work is to develop and validate a rapid and cost-saving genetic test for the molecular diagnosis of ADPKD and ARPKD based on Next-Generation Sequencing (NGS) using the Ion PGM^{TM} platform. For *PKD2* and *PKHD1* screening the standard protocol for targeted-resequencing based on Ion Ampliseq technology was reliable; however for the *PKD1* gene a strategy based on target-preselection using LR-PCR was applied in order to overcome the pseudogenes issue. The method validation, carried out on a cohort of patients with known molecular defects, showed 100% sensitivity and specificity. The prospective validation phase, carried out on different cohorts of patients with clinical suspect of ADPKD (n=125) or ARPKD (n=28), showed a detection rate of 90.4% and 85.7% respectively. Overall, 154 causative mutations were detected, 84 (54.5%) of which resulted previously undescribed, contributing the widening of the mutational spectrum of PKD.

In conclusion, this NGS-based genetic approach is highly accurate and reliable for mutation analysis, achieves a high sensitivity, a faster turnaround time and lower cost in comparison to conventional Sanger sequencing.

NGS would be an appropriate new standard for clinical genetic testing of PKD.

1. BACKGROUND

1.1 Polycystic Kidney Disease (PKD) as Ciliopathy

Polycystic Kidney Disease (PKD) is the most common genetic cause of kidney failure in children and adults. PKD is characterized by clusters of fluid-filled sacs, called cysts, which may arise in any part of nephron and can interfere with the kidney ability to filter waste products from the blood. The formation of cysts leads to the renal parenchyma destruction and causes renal enlargement, thereby disrupting the normal architecture of the kidney. Cysts may also develop in other organs, particularly in the liver (Wilson 2004).

As these kinds of disorders are caused by molecular defects in genes encoding for particular primary cilium proteins they are classified as "Ciliopathies".

Primary cilium is a hair-like appendage, present in most human cells, extending from the cell surface into the extracellular space, acts as cellular "*antennae*" to sense and to respond to extracellular signaling molecules, thanks to over 400 different proteins contained within its (Badano et al 2006). In particular, in kidney epithelial cells, primary cilium is indispensible for solute exchange, cellular signal transduction, and mechanical sensor of urine flow (Rodat-Despoix et al 2009). Finally, it plays also crucial role during differentiation and organogenesis (Marshall et al 2006; Poureetezadi et al 2016).

The term Polycystic Kidney Disease refers to two different diseases which differ in their mode of genetic inheritance and clinical phenotype: Autosomal Dominant Polycystic Kidney Disease and Autosomal Recessive Polycystic Kidney Disease (Wilson 2004).

1.1.1 Autosomal Dominant Polycystic Kidney Disease (ADPKD)

ADPKD (MIM #173900) is inherited as dominant trait and it is typically diagnosed in adults after the third decade of life with an incidence of 1:400–1,000. The disease is

characterized by the progressive, bilateral development and enlargement of focal cysts that in many cases ultimately result in end-stage renal disease (ESRD) (Torres et al 2007) (Figure 1).

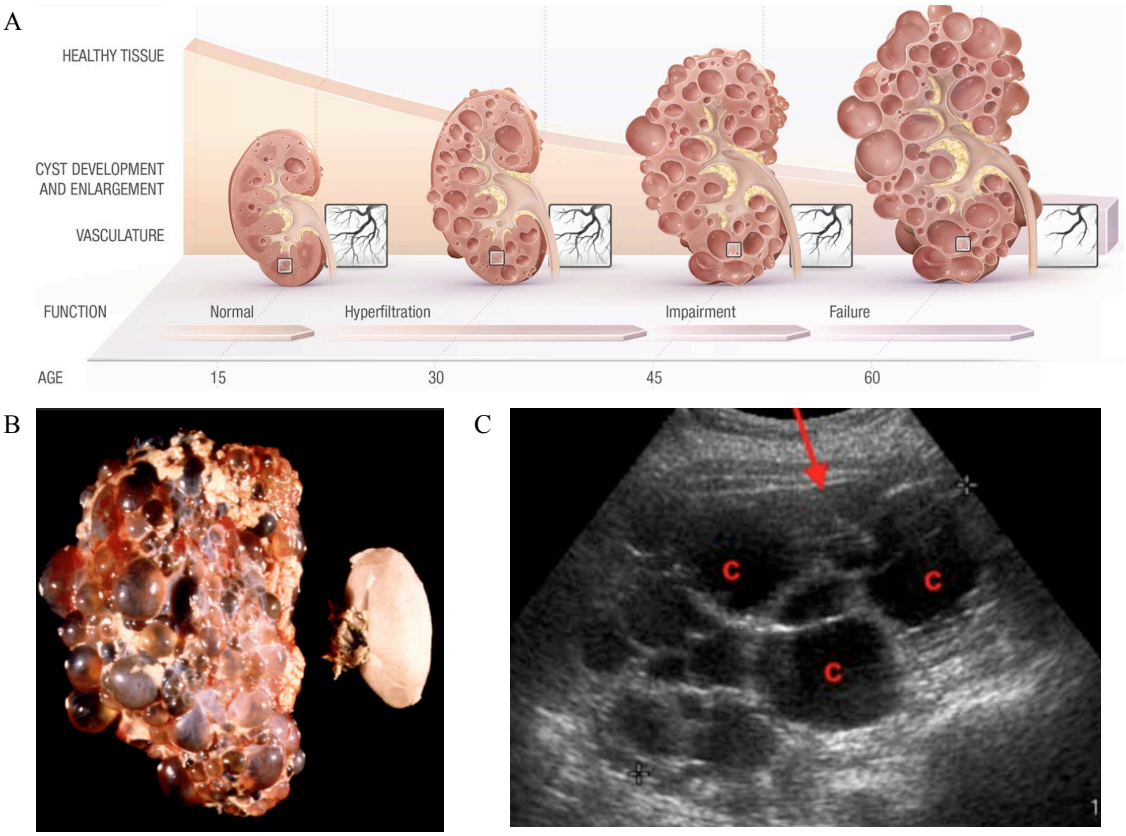


Figure 1: A) Cysts formation and development in an age dependent manner. The cysts development causes the progressive renal failure (PKD foundations 2017); **B)** ADPKD kidney compared with normal kidney (University of Virginia 2015); **C)** ADPKD patient renal ultrasound. The dark areas labeled with "C" represent cysts. Several smaller cysts are can also be seen (University of Virginia 2015).

The relentless development and grow of cysts cause progressive kidney enlargement associated with hypertension, abdominal fullness and pain, sporadic episode of cysts hemorrhage, gross hematuria, nephrolithiasis, cysts infection and a reduced quality of life (Grantham et al 2006; Torres et al 2007; Chapman et al 2012). Despite continuous destruction of renal parenchyma, the surviving glomeruli can maintain the normal renal function for decades by compensatory hyperfiltration. (Grantham et al 2006). Only when the majority of nephrons have been compromised, renal function decline leading to ERSD.

As ADPKD is a systemic disorder other organs are involved with potentially serious complications such as massive hepatomegaly and intracranial aneurysm (ICA) rupture. Additional extrarenal manifestations mainly encompass cysts in other organs (seminal vesicle: 40%; pancreas: 10%; arachnoid membrane: 8%; spinal meningeal:

2%) and connective tissue abnormalities (mitral valve prolapse, abdominal hernia, and diverticular disease (Torres et al 2007) (Table 1).

Table 1: Extrarenal manifestation in ADPKD

Extrarenal Manifestation	Associated	% Affected
Cardiac valve abnormalities	Yes	Mitral valve prolapse 25%
Pericardical effusion	Yes	Up to 35%
Extracranical Aneurysm	Yes, case report (Yeung et al. 2000)	Unknown
Arachonid cysts	Yes	8-12%
Spinal meningeal cysts	Yes	1.7%
Pancreatic cysts	Yes	10%
Diverticular disease	Possibly in association with ERSD	20-50% in ERSD
Abdominal hernias	Yes	Unknown
Seminal vescicle cysts	Yes	40%
Male infertility	Unknown	Unknown
Bronchiectasis	Possibly	37% in one series vs 13% controls
Congenital hepatis fibrosis	Yes, usually affecting only one generation within a family ADPKD	Rare

Table 1: Extrarenal clinical manifestations of ADPKD are listed in table. Abbreviations: ADPKD, autosomal dominant polycystic kidney disease, ERSD, end-stage renal disease (Chapman et al. 2015).

1.1.1.1 Genetics of ADPKD

Autosomal Polycystic Kidney Disease is caused by mutations in two different genes: *PKD1* that accounts for 80-85% of cases and *PKD2* for the remainder. ADPKD has strikingly high phenotypic variability; the severity of the disease can depend on the identity of the affected locus (*PKD1 vs PKD2* mutation), the allelic variant, timing of gene inactivation, mosaicism, and genetic background. Mutations in *PKD2* versus *PKD1* lead too much milder disease, with average ages at ESRD of 79.7 and 58.1 years, respectively. Milder disease is also noted in ADPKD cases associated with non-truncating versus truncating mutations of *PKD1*. However, the genotype-phenotype relationship is not completely understood. The disease is associated with a variety of phenotypes, from newborn infants with massive cystic kidneys to patients whose kidney function persists at adequate levels well into old age (Cornec-Le Gall et al 2014; Rossetti et al 2007).

A high level of allelic heterogeneity is found for both genes. A total of 2,323 different variants is reported for *PKD1* and 278 for *PKD2* in the ADPKD Mutation Database at Mayo Clinic (http://pkdb.mayo.edu), the most complete mutation database for ADPKD. The vast majority of mutations are unique to a single pedigree. For *PKD1*, 1,273 (54.7%) are classified as pathogenic (nonsense (20.7%), large deletion/frameshift/indel/ (44.9%), missense (24.7%), splicing variant (9.7%)). For

PKD2 a larger proportion of mutation is represented by large deletion/frameshift/indel (44.9%) and nonsense (20.7%) whereas one third are missense and splicing variants (33.7%) (Figure 2).

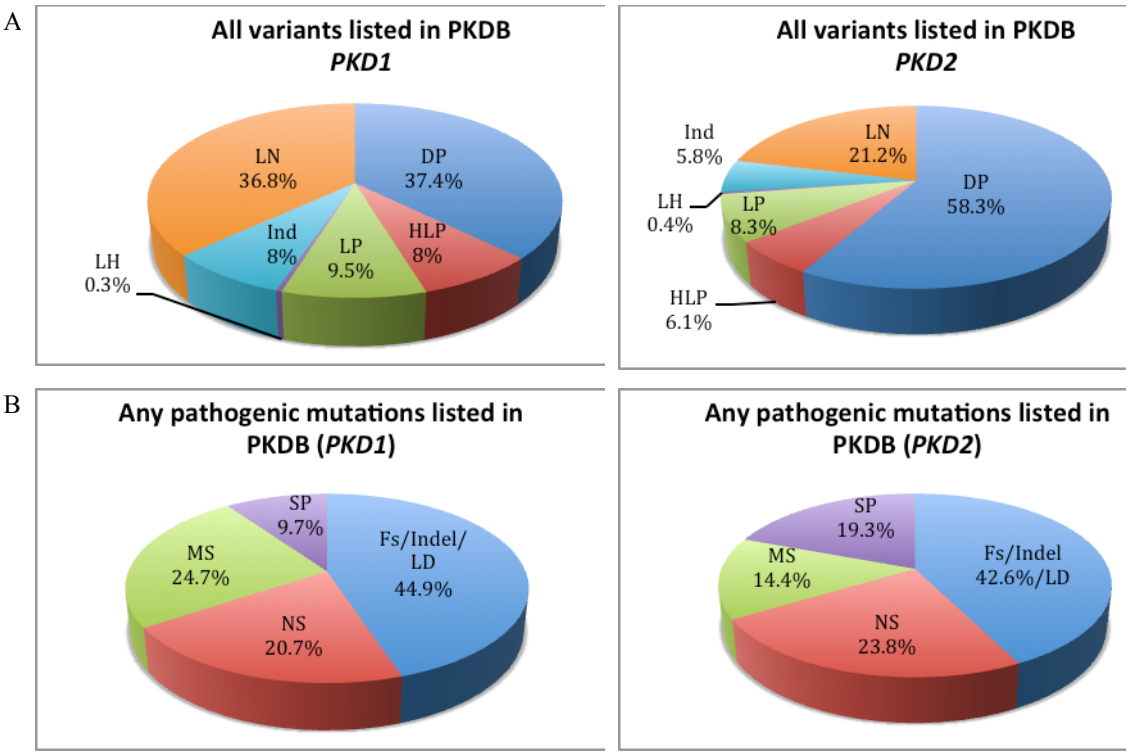


Figure 2: variants reported, up to date (2016), in ADPKD Mutation Database (PKDB) per *PKD1* and *PKD2,* grouped by pathogenicity level (**A**) and type of pathogenic variants (**B**).
A): The variants are grouped according their pathogenicity level (HLP, Higly Likely Pathogenic; DP, Definitely Pathogenic; LP, Likely Pathogenic; LH, Likely Hypomorphic; Ind, Indeterminate; LN, Likely Neutral; **B):** for all pathogenic variants are reported the percentage for type (Fs/Indel/LD, Framshift/Insertion Deletion/Large Deletion); Abbreviations: NS, Nonsense; MS, missense; SP, Splicing variant) (http://pkdb.mayo.edu).

A positive family history is absent in 10–15% of patients with ADPKD because of *de novo* mutations, mosaicism, mild disease from *PKD2* and non-truncating *PKD1* mutations, or because of unavailability of parental medical records.

1.1.1.2 Complex inheritance pattern

Although ADPKD is classically inherited as an autosomal dominant disease resulting from heterozygous mutations in either *PKD1* (#MIM 601313) or *PKD2* (#MIM 173910), recent and major advances in molecular genetic basis underline complex inheritance pattern of ADPKD, leading to the discovery of hypomorphic or incompletely penetrant alleles.

Homozigosity. As findings in both *PKD1* and *PKD2* knockout mouse suggest, homozygosity of *PKD1* or *PKD2* mutations would be embryonically lethal (Wu et al

2002). Consistently, the inheritance of two disease-causing *PKD1* alleles from both consanguineous (Paterson et al 2002) and non-consanguineous (Peces et al 2008) affected parents caused recurrent fetal losses. However, homozygosity involving either *PKD1* or *PKD2* has been reported to be a cause of ADPKD in four unusual families: c.9829C>T (p.Arg3277Cys) and c.9563A>G (p.Asn3188Ser) *PKD1* homozygous mutations were each found in a consanguineous families (Rossetti et al. 2009) and the third homozygous *PKD1* mutation, comprising two *cis*-linked nucleotide changes (c.[3133G>A;4709C>T] p.[Val1045Met;Thr1570Met]) was found in a family with unknown consanguinity (Vujic et al 2010). A common feature of these variants is that they were associated with no or at most suspected PKD in the heterozygous state but caused a definite and often severe disease in the homozygous state. Therefore, each of the three mutations can be reasonably defined as hypomorphic (Vujic et al 2010).

Compound heterozygosity. More compound heterozygotes ADPKD cases are known than homozygotes, and all involve the *PKD1* gene. These observations can be accounted for by (i) the much larger contribution of *PKD1* to the disease compared with *PKD2* and (ii) the high allelic heterogeneity of *PKD1* mutations. The reported compound heterozygotes (Table 2) can be divided into two categories: (i) either of the variants alone was associated with no or at most suspected disease but together caused a severe disease; (ii) one variant was a definite pathogenic mutation while the other was not associated with the disease alone but if coinheritance occurs the disease severity. The variants associated with no disease by themselves, in both categories, may be defined as hypomorphic. However, their roles in disease pathogenesis are context-dependent. Thus, although the variants in first category may be defined as co-causative, those in the second one may be better described as disease-modifying.

Trans-Heterozygousity. To date, the only known digenic inheritance involving *PKD1* and *PKD2* genes was reported in a unique multigenerational family with ADPKD (c.2159delA (p.Asn720fs) in *PKD2* and c.1583A>G (p.Tyr528Cys) in *PKD1*). The heterozygous *PKD1* p. Tyr528Cys carriers were shown to have a uniformly mild renal disease similar to the heterozygous *PKD2* p.Leu736* carriers. The two trans-heterozygotes lead a more severe renal disease in the proband (Pei et al 2012).

Table 2: compound heterozygousity found for *PKD1* gene in different family

Mutations	Genotype-Phenotype description	References
c.[8293C>T];[9313C>T] p.[Arg2765Cys];[Arg3105Trp]	The proband had multiple cysts consistent with PKD at 11 years. Both parents (single heterozygotes) were apparently unaffected.	(Rossetti et al. 2009)
c.[6658C>T];[9829C>T] p.[Arg2220Trp];[Arg3277Cys]	The two compound heterozygous children had in utero onset PKD. Renal ultrasounds of the single heterozygous parents (father at the age of 41 and mother at the age of 34) found no abnormalities.	(Vujic et al. 2010)
c.[3820G>A];[3820G>A;8716G>A] p.[Val1274Met]; [Val1274Met;Gly2906Ser]	The three compound heterozygous children had PKD. Single heterozygous parents (mother, 25 years old; father, 29 years old) had normal renal ultrasound.	(Bergmann et al. 2011)
c.[6472C>T];[9829C>T] p.[Gln2158*];[Arg3277Cys]	The one p.Arg3277Cys heterozygote (#25 years) had a negative renal ultrasound. Coinheritance with the definite pathogenic mutation caused in utero onset ADPKD	(Rossetti et al. 2009)
c.[11457C>A];[8293C>T] (p.[Tyr3819*];[Arg2765Cys])	The one p.Arg2765Cys heterozygote (#25 years) had a negative renal ultrasound. Coinheritance with the definite pathogenic mutation caused in utero onset ADPKD	(Rossetti et al. 2009)
c.[7915_7916ins20];[8293C>T] (p.[Arg2639fs];[Arg2765Cys])	The one p.Arg2765Cys heterozygote (#36 years) had a negative renal ultrasound. Coinheritance with the definite pathogenic mutation caused in utero onset ADPKD. Coinheritance with the definite pathogenic mutation caused in utero onset ADPKD.	(Rossetti et al. 2009)
c.[8259C>G];[6763C>T] (p.[Tyr2753*];[Arg2255Cys])	Three p.Arg2255Cys heterozygotes were associated with no or suspected disease. Coinheritance with the definite pathogenic mutation caused a more severe disease.	(Bergmann et al. 2011)
c.[4051_4052delGG];[8087T>G] (p.[Arg1351fs];[Leu2696Arg])	The one p.Leu2696Arg heterozygote had a negative renal ultrasound. Coinheritance with the definite pathogenic mutation caused a more severe disease.	(Bergmann et al. 2011)
c.[4199delT];[12413G>A] (p.[Leu1400fs];[Arg4138His])	The p.Arg4138His heterozygote had a negative renal ultrasound. Coinheritance with the definite pathogenic mutation caused a more severe disease.	(Bergmann et al. 2011)
c.[3312dupC];[6749C>T] (p.[Val1105fs];[Thr2250Met])	The one p.Thr2250Met heterozygote had a negative renal ultrasound. Coinheritance with the definite pathogenic mutation caused a more severe disease.	(Reiterova et al. 2013)
c.[8362_8363ins34];[5848G>A] (p.[Ser2788fs];[p.Val1950Met])	The one p.Val1950Met heterozygote had a negative renal ultrasound. Coinheritance with the definite pathogenic mutation caused a more severe disease.	(Gilbert et al. 2013)

Table 2: all compound heterozygous cases known are shown. The mutation nomenclature follows the format indicated in the website http://www.hgvs.org/mutnomen/. The DNA mutation numbering system is based on cDNA sequence with a "c." symbol before the number, using $^{+}$1 as the A of the ATG translation initiation codon in the reference sequence. Reference sequences: NM_001009944.2 (*PKD1*) (Cornec-Le Gall et al. 2014).

Triallelic Inheritance. Only one patient was found to carry the *PKD2* c.1445T>G (p.Phe482Cys) homozygous variant and a *de novo* heterozygous *PKD1* splice-site mutation, c.8017-1delAG (Dedoussis et al 2008). The *PKD2* p.Phe482Cys variant is not a common polymorphism and functional studies revealed a lower PC-2 channel function compared with controls in homozygous cells and heterozygous as well. As the patient appeared to have earlier ERSD, the *PKD2* variant was considered to have a modifier role in this particular context (Dedoussis et al 2008).

Coinheritance of mutations in one of the ADPKD genes and an additional PKD-causing gene. A recent report shows the coinheritance of mutations in one of the two

ADPKD causative genes and in *PKHD1* or *HNF1β* gene in four families (Bergmann et al 2011). Mutations in the *HNF1β* gene cause autosomal dominant "renal cysts and diabetes syndrome" (Horikawa et al 1997) and homozygous mutations in the *PKHD1* gene are associated with autosomal recessive polycystic kidney disease (ARPKD) (Mucher et al 1994). In these families the coinheritance of mutations in *PKD1* o *PKD2* genes and in an additional PKD-causing gene seems to be associated with early and severe renal phenotypes.

1.1.1.3 PKD1 and PKD2 genes

PKD1 (16p13.3) is a very large gene consisting of 46 exons span over 52 kb of genomic DNA and it's characterized by GC-rich regions (62,4%), with a CpG/GpG ratio of 0,485 (The American *PKD1* Consortium 1995). The 3'-end of the gene partially overlaps with 3'-end of *TSC2* (#MIM 191092) gene in a "tail-to-tail" orientation. Contiguous deletions removing both of the tails of *PKD1* and *TSC2* genes cause a severe form of infanitile polycystic kidney disease (Sampson et al 1997). Moreover, the *PKD1* region encompasses 5' to 33 exon is replicated in six pseudogenes, expressed as mRNA transcript but not translated, mapping proximally on chromosome 16 (International Polycystic Kidney Disease Consortium 1994). These six pseudogenes differed from the genuine gene by deletion and other rearrangement, but share 98% sequence similarity in exonic sequences (Bogdanova et al 2001).

PKD1 real gene encodes a 14,138 bp mRNA transcript (NM_001009944.2) with a coding sequence of 12,912 bp that is translated into a protein composed of 4,302 amino acids: Polycystin 1 (NP_001009944).

PKD2 (4q21-23) is a less longer and complex gene. It consists of only 15 exons encoding an mRNA transcript of 5,056 bp (NM_0000297) with a coding sequence of 2,907 bp that translated into a protein composed of 968 amino acids: Polycystin 2 (PC2) (NP_000288).

Polycystin 1 and 2 belong to the transient receptor potential (TRP) channel superfamily and are also known as TRPP1 and TRPP2, respectively.

1.1.1.4 Polycystin 1 - 2 and PC1/PC2 Complex

Polycystin-1 (PC1). PC1 is a receptor-like protein, with a molecular mass of 600 kDa (uncleaved and glycosylated).

It is expressed in epithelial cells of the differentiating and mature renal tubules, as well as in a variety of somatic tissues including liver, bone, heart and endocrine glands (Ward et al 1996; Ibraghimov-Beskrovnaya et al 1997). At cellular level, the protein localizes in primary cilium, as well as to the lateral domain of the plasma membrane and in adhesion complexes in polarized epithelial cells (Ibraghimov-Beskrovnaya et al 2000).

Despite the 3D structure remains unresolved because the protein size and complexity, many study focused only on isolated domain to try to understand the protein modular structure. The current view of the protein is shown in Figure 3.

The large extracellular domain comprises a number of domains involved in protein-protein and protein-carbohydrate interactions. The Leucine-Rich repeat (LRR) domain (N-terminus) is followed by sixteen immunoglobin-like domains named PKD repeats, a receptor for egg-jelly (REJ) domain and a GPCR-Autoproteolysis Inducing (GAIN) domain. The first PKD domain is separated from the other 15 by a c-lectin type domain. The PKD domain is reminiscent of proteins with structural and mechanic role, as fibronectin, and may mediate the formation of multicellular structures. The GAIN domain includes G protein-coupled receptor proteolysis site (GPS) motifs that mediates *cis*-autoproteolysis of PC1 to generate a 320 kDa N-terminal product and a 140kDa C-terminal non-covalently bound. The role of auto-proteolytic cleavage is not completely understood yet but several studies suggest that may be important during cellular development (Chapin et al 2010; Yu S 2007).

Functional studies demonstrates the role of the entire extracellular domain in mediating cell adhesion and/or cell junction formation via trans-homophilic interactions of PKD domains and cis-heterophilic interactions with E-cadherin (Streets et al 2009). The PC1 is anchored to the cellular membrane by eleven transmembran domain. In the intracytoplasmatic portions are located the Polycystin-1 Lipoxygenase Alpha-Toxin (PLAT) domain and the C-terminal tail. PLAT domain is considered a signature domain of *PKD1*-like protein and represents a lipid/protein-binding scaffold to integrate cell signaling and PC1 traffiking. The C-terminal tail is the best studied region and contain a coiled-coil domain for PC2 binding and a G-

protein binding and activation sequence. It can bind many phosphatases which regulates PC1 and PC2 phosphorilation and other key signaling proteins or effectors as mTOR, Wnt, JAK/STAT) (Shillingford et al 2006; Lal et al 2008; Bhunia et al 2002).

Polycystin-2 (PC2). PC2 is a high conductance non-selective Ca^{2+} permeable channel with a molecular mass of ~110 kDa. Although a portion of PC2 colocalizes with PC1 to the cilium, the majority of the PC2 is found in intracellular calcium stores where regulates the release of this cation in response to local increase in Ca^{2+} concentrations (Vassilev et al 2001; Koulen et al 2002).

Polycystin 2 consists of six transmembrane domains, which partially share homologies with PC1 TM domain, comprised between the C-N terminal domains (Figure 3).

The calcium-conducting pore is formed by the loop between the fifth and the sixth transmembrane domains (Koulen et al 2002). The C-terminal domain contains two coiled-coil domains (CT1 and CT2) and an EF-hand domain. CT1 is essential for recognition and binding of the PC1 C-terminus to form a PC1/PC2 complex, whereas the CT2 mediates the oligomerization between different PC2s to form oligomeric complexes. The EF-hand domain is a helix-loop-helix structures with a Ca^{2+}-binding site that permits to sense or buffer changes in calcium concentration (Gifford et al 2007). Slightly overlapping with both coiled-coil and EF-hand is one of the two membrane targeting motifs required to maintain the PC2's ER and Golgi localization, the other one is in N-terminal portions and is required to guides and maintain ciliary localization (Cai et al 1999).

PC1/PC2 complex. PC1 and PC2 form a heteromeric molecular complex in cilia, with PC1 serving as a mechanosensor and PC2 functioning as a Ca^{2+} channel. Changes in urinary flow in extracellular space lead to the bending of cilia and PC1-PC2 complex-dependent Ca^{2+} entry into cell resulting in measurable increases in cytosolic calcium concentration. This flow-mediated increase in cell calcium is diminished if either PC1 or PC2 are absent or nonfunctional (Nauli et al 2003).

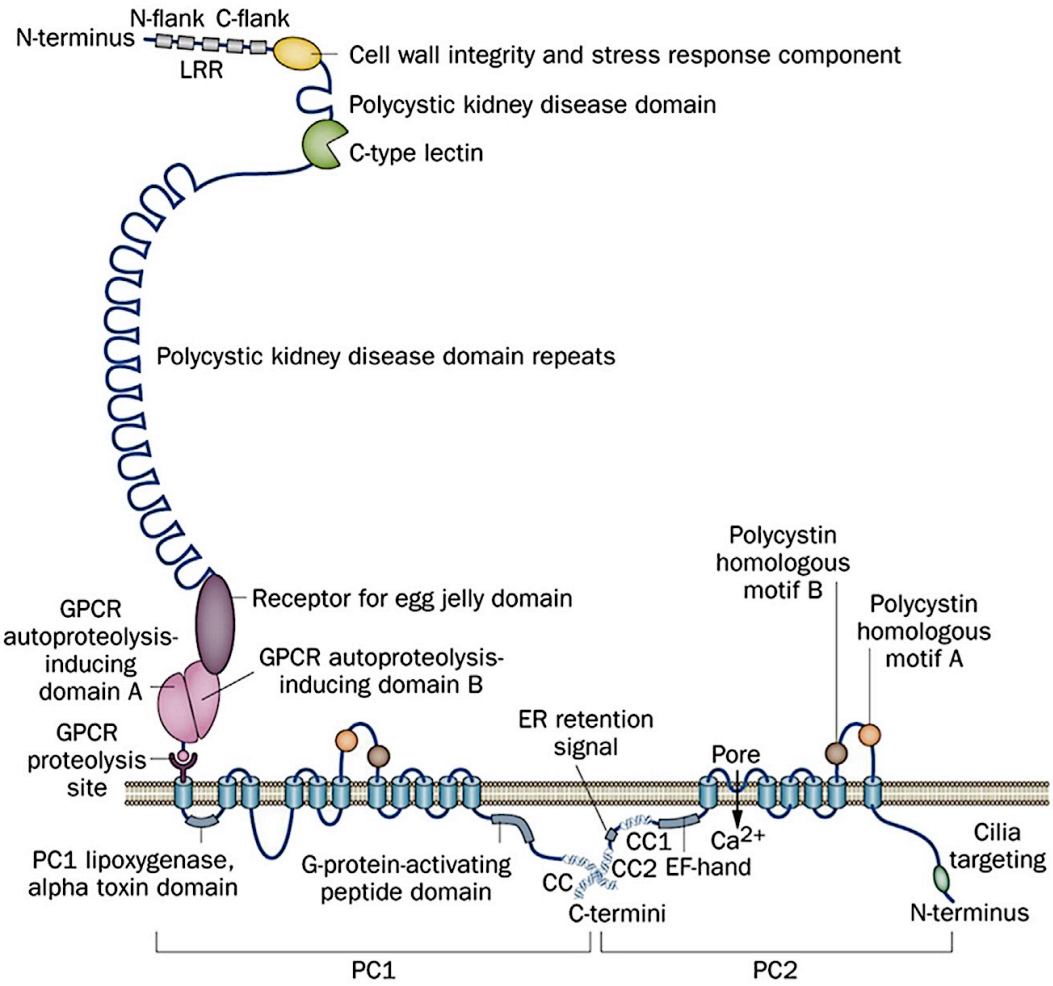


Figure 3: Predicted structures of Polycystin 1 (PC1) and Polycystin 2 (PC2): PC1 is a receptor-like protein with a large ectodomain, 11 transmembrane domains and a cytoplasmic tail. The last 6 transmembrane domains of PC1 are homologous to the transmembrane region of PC2. PC2 is a transient receptor potential–like calcium channel that has an EF-hand motif and an endoplasmic reticulum (ER) retention signal in the carboxy (C) terminus and a proposed cilia targeting sequence in the amino (N) terminus. PC1 and PC2 physically interact through coiled-coil domains in the cytoplasmic tail of PC1 and in the carboxy-terminal tail of PC2 (Ong et al 2015).

1.1.1.5 Signalling pathway modified by PC1 and PC2

Polycystins protein modulates several pathways most of them implied in cellular differentiation, growth, division, apoptosis and planar cell polarity determination. Alteration in these pathways may activate the cystogenesis process. Misregulating PC1 and PC2 expression or *PKD1* cleavage appears to result in aberrant signaling, which may in turn lead to the abnormal cellular growth behaviors that are likely to contribute to ADPKD pathogenesis. The major involved signaling pathways are negative growth regulation, G protein activation, and Wnt pathway modulation (Figure 4).

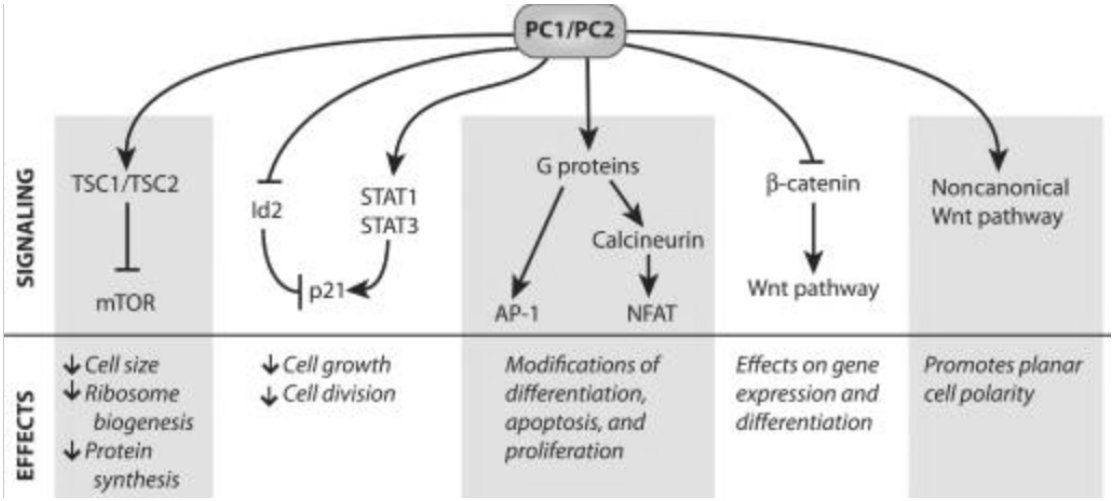


Figure 4: PC1 and PC2 affect multiple signaling pathways. Summary of the effects that PC1 and PC2 exert on signaling pathways. Multiple direct and indirect interactions allow the polycystin proteins to inhibit or stimulate pathways involved in cellular growth and differentiation (Chapin H.C 2010).

G-Protein activation. PC1 functions as a G-protein coupled receptor (GPCR), as contains sequences that are found in GPCRs (Igarashi et al 2002). Downstream signaling through GPCRs can lead to the activation of c-Jun N-terminal kinase (JNK) and activator protein-1 (AP-1) pathways that can regulate cell proliferation, differentiation, cell cycle, apoptosis and inflammation.

Growth regulation. The mechanisms utilized by PC1/PC2 include signaling through the JAK-STAT pathway (Bhunia et al 2002). Signaling through this pathway leads to the activation of STAT1 and STAT3, subsequent upregulation of the cyclin-dependent kinase inhibitor (CKI) p21, and the inhibition of cyclin dependent kinase 2 (CDK2) which ultimately leads to cell cycle arrest at the G0/G1 transition (Li et al 2005). Additional evidence for a direct role of polycystins in cell cycle regulation was shown by the direct interaction of PC2 with the helix-loop-helix (HLH) protein Id2, which regulates cell proliferation and differentiation. Phosphorylation of PC2 by PC1 leads to its interaction with Id2-E47 complex. The interaction of the Id2-E47 complex with PC2 sequesters this complex outside the nucleus. When PC2 is unable to bind to this complex, Id2 is translocated to the nucleus and exerts its dominant negative effects on E47 and other HLH proteins. In this manner, the normal expression of PC1 and PC2 leads to cell cycle arrest by an increase in p21 while a mutation in one of the polycystins leads to the dysregulation of this pathway resulting in increased cell proliferation (Low et al 2006).

An additional significant effect of PC1 involves inhibition of the mTOR(mammalian target of rapamycin) cascade which regulates protein translation, cell proliferation, and cell growth (Shillingford et al 2006). Normally TSC1 and TSC2 (tuberous sclerosis 1 and 2) complex acts as a negative regulator of the mTOR complex (Huang et al 2008) and Shillngfors et al showed that PC1 C-terminal tail interacts with TSC2 stabilizing the functional TSC1– TSC2 complex. In the absence of a functional PC1, as in most cases of ADPKD, the mTOR pathway is activated (Shillingford et al 2006).

Wnt-Pathway. ADPKD cysts and PC1-null cells manifest up-regulation of Wnt signaling activity markers, suggesting that PC1 exerts a negative effect on this system. The Wnt pathways affect growth, differentiation, and establishment of planar cell polarity. In the canonical pathway, the presence of the Wnt ligand induces β-catenin stabilization and nuclear translocation, leading to T cell factor (TCF)–dependent transcriptional activity. Cleaved PC1 CTT inhibits this pathway by directly or indirectly binding to β-catenin, moving with it to the nucleus, and reducing its ability to promote TCF-dependent transcription (Lal et al 2008). PC2 may also regulate the expression of some components of the Wnt pathway. Knocking out PC2 in cultured mouse cells resulted in increased levels of β-catenin protein (Kim et al 2009).

PC1 may also regulate noncanonical Wnt signaling, which is in turn related to the maintenance of planar cell polarity. The cells lining renal tubules generally divide parallel to the tubule's axis, lengthening the tubule rather than expanding its diameter. Tubule-lining cells in models of polycystic kidney disease, however, show a tendency to divide at an angle to the tubule's axis, which could lead to expansion of the tubule diameter. This deviation can occur before cysts appear, suggesting that a loss of this planar cell polarity may be a precursor to cyst formation (Fisher et al 2006; Patel et al 2008).

A disruption in the switch between canonical and non-canonical Wnt signaling could be involved in cystogenesis process (Fisher et al 2006).

1.1.1.6 Cysts formation

Cysts form as small dilations in renal tubules, which then expand to form fluid-filled cavities of different sizes. Factors that are thought to lead to cystogenesis include a germ-line mutation in one of the polycystin gene alleles, a somatic second hit which

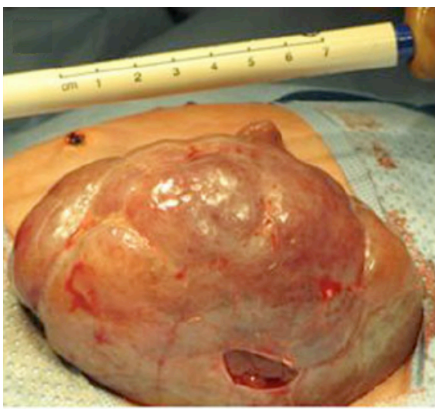
leads to the loss of the normal allele, and a third hit, which can be anything that triggers cell proliferation, leading to the dilation of the tubules. Continued dilation of the tubules through increased cell proliferation, fluid secretion, and separation from the parental tubule will lead to the formation of cysts. Other factors that are involved in cystogenesis and/or cyst progression include defective planar cell polarity, extracellular matrix abnormalities, inflammation, increased apoptosis, modifying genes, and environmental factors (Igarashi et al 2002).

1.1.2 Autosomal Recessive Polycystic Kidney disease (ARPKD)

ARPKD (#MIM 263200) is inherited as recessive trait and belongs to a group of congenital hepatorenal fibrocystic (CHF) syndromes and is a cause of significant renal and liver-related morbidity and mortality in children. It is more rare than ADPKD with an incidence of 1-20.000 live births and the carrier frequency in the general population was estimated at 1:70 (Zerres et al 1998), therefor, the risk of ARPKD in offspring of a proband is approximately 0,7%. Rare milder forms are also known in adolescents and adults.

In infancy, the disease results in significantly enlarged echogenic polycystic kidneys, with pulmonary hypoplasia resulting from oligohydramnios as a major cause of morbidity and mortality. Renal failure is rarely a cause of neonatal demise. On sectioning of the enlarged kidney, fusiform dilatations arranged radially through the renal parenchyma from medulla to cortex are seen, with rounded cystic dilatations in the medulla. Liver involvement is detectable in approximately 45% of infants and is often the major feature in older patients (Harris et al 2004) (Figure 5).

A 

B 

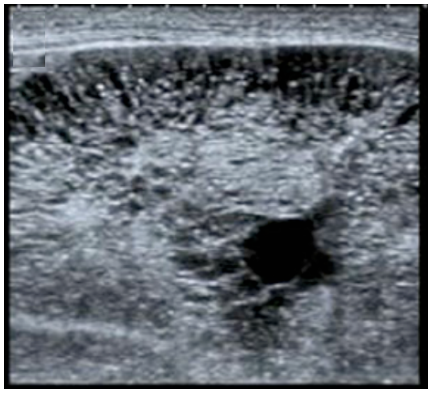
C  D 

Figure 5: Baby with autosomal recessive polycystic kidney disease (ARPKD). **A)** Distended abdomen due to voluminous kidneys that lead to respiratory problems; **B):** nephrectomized kidney of this girl; **C) - D)** Ultra-sound showed symmetrically enlarged echogenic kidneys with fusiform dilations of collecting ducts and distal tubules arranged radially throughout the renal parenchyma from medulla to cortex (Bergmann et al 2011).

In ARPKD patients surviving the neonatal period, have as much as 80% chance of living to more than 15 years of age (Kaplan et al 1989). Advances in neonatal intensive care and renal replacement therapies have improved the survival rates of ARPKD patients, with some of them reaching adulthood. However, life expectancy is severely diminished and a wide range of associated comorbidities often evolve, including systemic hypertension (HTN), endstage renal disease (ESRD), and clinical manifestations of congenital hepatic fibrosis (CHF) (Roy et al 1997). ARPKD is invariably associated with biliary disgenesis and the Caroli's disease is often observed.

The most severely affected fetuses have enlarged echogenic kidneys and display a "potter" oligohydramnios phenotype with pulmonary hypoplasia, a characteristic facies, and contracted limbs with clubfeet (Roy et al 1997). The only signs potentially detectable in utero are enlargement and increased echogenicity of both kidneys. Improved respiratory treatment leads to increase neonatal survival, but death still occurs in the neonatal period in approximately 25%-30% of affected individuals primarily because of respiratory insufficiency (Kaplan et al 1989; Roy et al 1997). Liver involvement is always present and may be the predominant clinical extrarenal feature; in some of these cases, renal ultrasonography may be required to detect clinically silent renal disease. Complications of the hepatic disease include portal hypertension induced by hepatic fibrosis (Alvarez 1981) and acute bacterial cholangitis associated with dilatation of the bile ducts (Kääriäinen et al 1988). Results of liver function tests usually remain within the normal range. Ultrasonography of the liver reveals dilatation of the peripheral intrahepatic ducts and the main bile ducts.

Hepatic cysts may be present and there may also be signs of portal hypertension. Magnetic resonance cholangiography, a non-invasive imaging technique, can also visualize non-obstructive dilations of the intrahepatic bile ducts (Jung et al 1999).

Approximately 50% of affected individuals progress to end stage renal disease (ESRD) within the first decade of life. (Roy et al 1997). Hypertension may occur in up to 80% of children with ARPKD, it is frequently severe and is generally correlated with decrease in renal function. Therapies with an angiotensin converting enzyme inhibitors or ATII receptor inhibitors are generally effective (Guay-Woodford et al, 2003; Kaplan et al 1989). Additional clinical complications include nephrogenic diabetes insipidus, failure to thrive, and hyponatremia (Dell et al 2004; Guay-Woodford et al 2003; Zerres et al 1996).

1.1.2.1 Genetics of ARPKD

ARPKD is caused by mutations affecting *PKHD1* (MIM* 606702) gene. Up to date, the major mutation database that collects and classifies all described *PKHD1* variants, associated to ARPKD clinical phenotype, is Humgen (http://www.humgen.rwth-aachen.de) (Figure 6). Of all mutant alleles found 49% are single nucleotide missense substitutions, about 12% are small deletions, insertions, or duplications, approximately 9% contain nonsense mutations, and a few mutations (30%) affect the canonical splice site sequences.

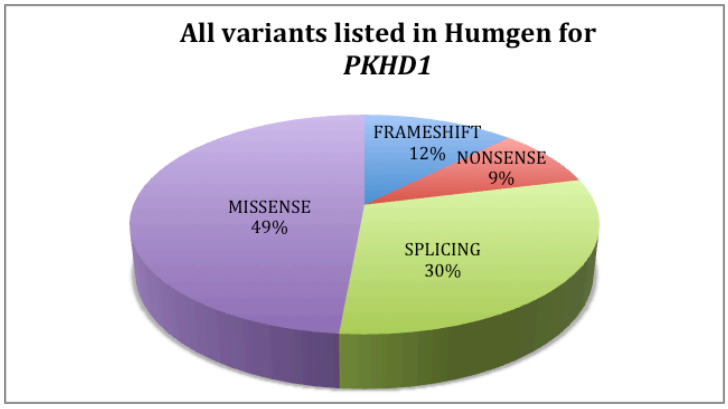


Figure 6: The pie chart reports the percentage (%) of variants, described up to date (2016), in Mutation Database Autosomal Recessive Polycystic Kidney disease (ARPKD/*PKHD1*).

Thus far, gross genomic rearrangements (deletions/ insertions/ inversions), promoter alterations or mutations in alternatively spliced exons not included in the longest ORF, have not yet been encompassed into the analysis in most of the studies.

The most common pattern of mutations in ARPKD patients is two different changes (compound heterozygotes) (Onuchic et al 2002; Ward et al 2002; Bergmann et al 2004; Furu et al 2003; Rossetti et al 2003). Indeed, it is possible that some rare ‘‘polymorphisms’’ may be mutagenic when found in combination with an inactivating mutation. As a wide range of compound heterozygous changes are associated with ARPKD, it has been difficult to determine clear genotype/phenotype correlations. In different studies, affected families were categorized by clinical presentation into ‘‘severe’’ and ‘‘moderate”: the ‘‘severe’’ cohort comprised families in which at least one affected child presented with perinatal disease and neonatal demise, while the ‘‘moderate’’ group included families in which affected patients either survived complications during the first month of life or first became symptomatic beyond the neonatal period. Thanks to this classification it was seen that in the moderate cohort, missense changes were more than two times as frequent as chain-terminating alterations, whereas the compound heterozygosis of Loss of Function (LOF) mutations caused severe phenotype.

1.1.2.2 *PKHD1* gene

PKHD1 (6p12) gene is the only candidate gene for all typical forms of ARPKD. *PKHD1* extends over 469 kb and include a minimum ok 86 exons (Onuchic et al 2002; Ward et al 2002). The longest *PKHD1* transcript includes 67 exons (Figure 7) with an open reading frame (ORF) composed of 66 exons that encode a 4074 amino acid protein, polyductin/fibrocystin. Alternatively spliced transcripts are predicted to fall into two broad groups. The first subset, polyductin-M, is comprised of polypeptides that contain the single TM element but vary with respect to inclusion of the other predicted domains. The second subset, polyductin-S, lacks the TM domain and thus its members may be secreted (Onuchic et al 2002).

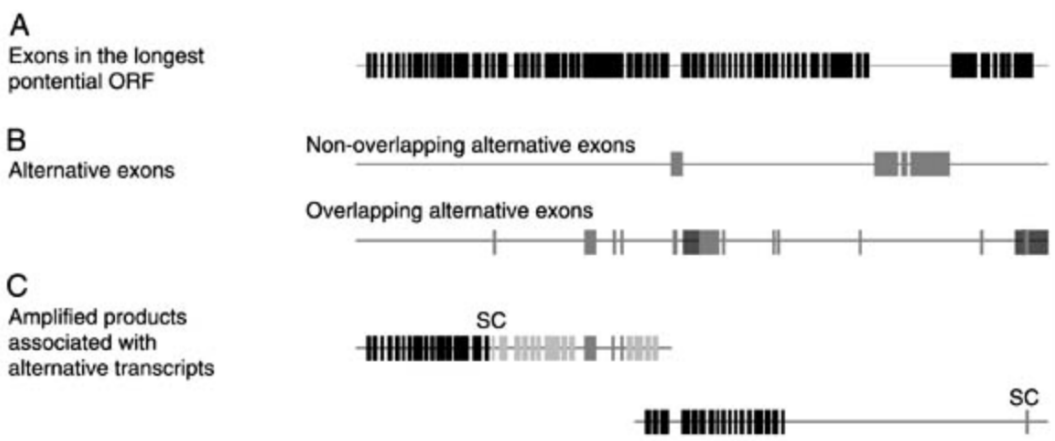


Figure 7: *PKHD1* gene structure. The gene has 86 exons so far identified; 71 are non-overlapping while 15 are exons with alternative splicing boundaries (A and B). The putative longest open reading frame (ORF) transcript comprises 67 exons (A). Preliminary analyses suggest that *PKHD1* encodes a high number of alternative transcripts. Two examples of differential assembling are shown at the bottom of the figure (C). SC: approximate location of the stop codon; dark gray: alternative exons; light gray: noncoding exons in the corresponding transcripts (Menezes et al.2006).

1.1.2.3 Fibrocystin 1/Polyductin (FC1)

The *PKHD1* protein, Fibrocystin 1/Polyductin (FC1), is predicted to have 4074 amino acids (4059 amino acids in the mouse) with a calculated unglycosylated molecular weight of 447 kDa (Ward et al 2002; Onuchic et al 2002). However, the mature protein may be significantly larger as there are multiple potential N-linked glycosylation sites. Fibrocystin has a signal peptide, and structural predictions indicate a large extracellular region (3860 aa) with multiple copies of the TIG domain (an immunoglobulin-like fold), domains that shared homology with TMEM protein of unknown function, several parallele b-helix (PbH), a single transmembrane region and a short cytoplasmic tail (192 aa) (Ward et al 2002) (Figure 8). The role of other TIG containing proteins, such as the hepatocyte growth factor receptor (Met) and the plexins, and the structure of fibrocystin, suggest that it may be a receptor protein (Ward et al 2002). PbH1 repeats are present in virulence factors, adhesins, and toxins in bacterial pathogenesis and known to bind to carbohydrate moieties (Cowen et al 2002). Several secreted forms of fibrocystin may also be generated from alternatively spliced transcripts (Ward et al 2002; Onuchic et al 2002).

The external region also contains 64 potential N-glycosylation sites, suggesting that the protein may be highly glycosylated. In addition, FC1 short carboxylic tail includes putative cAMP/cGMP-dependent protein kinase phosphorilation sites (PKA and PKC) (Ward et al 2002).

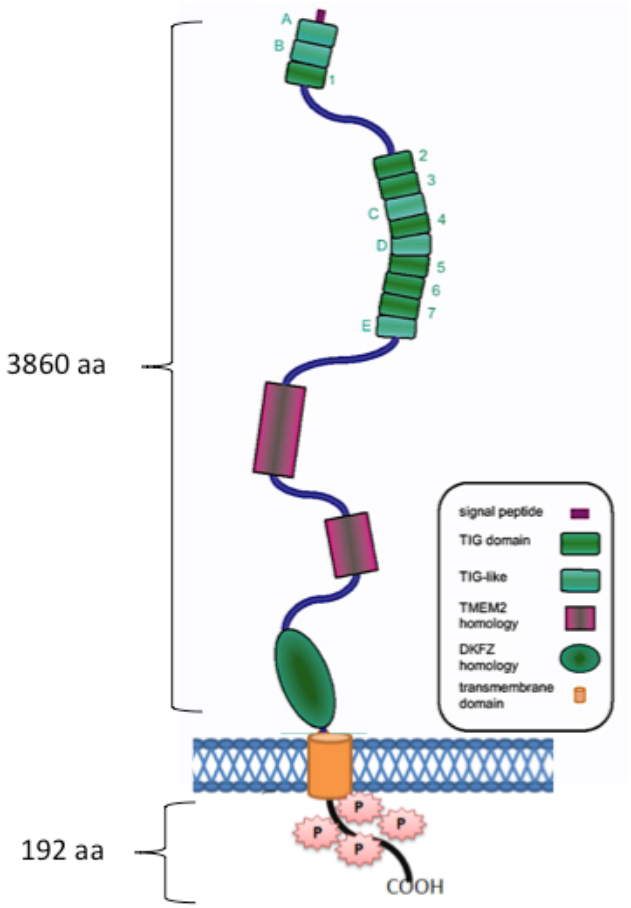


Figure 8: Predicted structure of FC1, a large integral membrane protein, 4074-amino-acid The structure of FC1 as an integral membrane protein with a large extracellular portion and a short intracellular C-terminal suggest that this protein acts as a transducer of extracellular information into the cell by eliciting signal transduction cascades resulting in the modulation of gene transcription (Ward et al. 2002).

The function of FC1 is not known, but the structure and homologies suggest a role as a receptor protein possibly involved in modulating the terminal differentiation of collecting ducts and the biliary system.

Immunohistochemistry analyses revealed staining of cortical and medullary collecting ducts and thick ascending limbs of Henle in kidney and biliary and pancreatic duct epithelia. In human fetus and mouse developing tissues, staining was observed in the branching ureteric bud but not in the metanephric mesenchyme, S-shaped bodies or glomeruli (Menezes et al 2006).

Several studies have shown that at the subcellular level FC1 is expressed in primary apical cilia in kidney cells and cholangiocytes (Menezes et al 2004; Ward et al 2003; Wang et al 2004).

Immunoreactive FC1 localized at the apical domain of polarized epithelial cells, suggesting it may be involved in the tubulogenesis and or maintenance of duct–lumen architecture. FC1 is localized at the apical membrane in collecting duct cells and in the cytoplasm of inner medullary collecting duct (IMCD) cells grown in culture (Menezes et al 2006).

1.1.2.4 Polycystin complex and Fibrocistin/Polyductin interaction

PC2, one of the two ADPKD protein involved in calcium signaling, was seen to participate in an indirect way at the same protein complex with FC1 (Figure 9).
The linker protein involved in the creation of a triplex (PC2–KIF3B–FC1) is the kinesin-2 motor subunit KIF3B. Moreover, several experiments demonstrated that PC2 and FC1 are in the same complex only if KIF3B is present; altering KIF3B level in IMCD cells by over-expression or by siRNA significantly affected complexing between PC2 and FC1 (Wu et al 2006).
Thus, KIF3B represents the first molecular linker between ADPKD and ARPKD proteins, suggesting that the protein complex PC2–KIF3B–FC1 is part of a common molecular pathway implicated in renal cystic diseases. Kinesin-2, as a heterotrimer formed by KIF3A, KIF3B and KAP3, is important for numerous cell functions, in particular, cilium growth and cell cycle (Miki et al 2001; Cole 1999) and has profound pathological implications. In ciliated IMCD cells, PC2 partially co-localized with KIF3B and FC1 in primary cilia suggesting the possibility that the flow sensor is composed of a larger complex containing not only PC1 and PC2, but also KIF3B, FC1 and possibly other proteins (Wu et al 2006).
FC1 is capable of increasing PC2 channel function in the presence, but not in the absence, of KIF3B. Thus, KIF3B likely links PC2 and FPC together not only in a structural complex, but also for the regulation of PC2 channel function. So, FC1 is another partner of PC2, which functionally upregulates its channel function through KIF3B regulations.

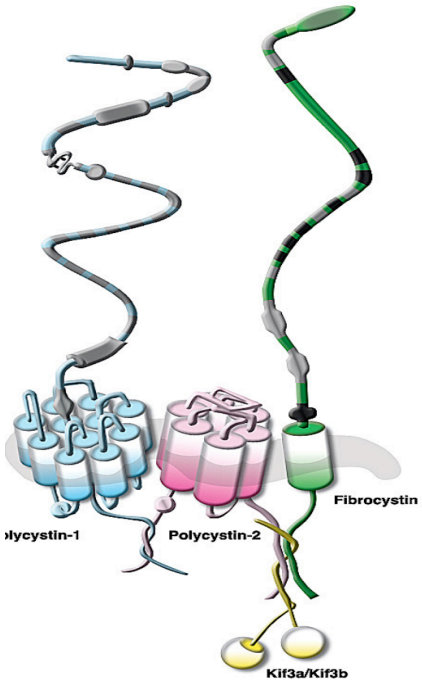


Figure 9: Mechanosensory protein complex. Among other subcellular localizations, it is thought that polycystin-1, polycystin-2 and fibrocystin form a mechanosensory complex protein in the cilium to sense fluid-shear stress. Polycystin-1 and polycystin-2 interact with each other at their COOH termini forming a polycystin complex that interacts with Fibroystin1/Polyductins via Kif as linker protein. (Kolb et al 2008)

It is also possible that the roles that KIF3B plays in the PC2–FC1 complexing and in the modulation of PC2 by FC1 are different from its role as a kinesin-2 motor subunit. (Wu et al 2006).

Unlike the biliary ductal epithelial cells (Masyuk et al 2004), however, successful, but not total, reduction of FC1 in the RNA and protein levels did not result in obvious defects in ciliogenesis, suggesting that the residual FC1 is sufficient for normal ciliogenesis or there might be other specific splicing events of *PKHD1* (Onucich et al 2002). Evidences show that the disruption of either FC1 or the PC1/PC2 complex at the primary cilium results in defects in mechanosensation of fluid flow (Wang et al 2007).

1.1.3 Ultrasonography in ADPKD and ARPKD diagnosis

Ultrasonography (US) is the diagnostic method of choice for assessing ARPKD and ARPKD because it is cost effective, painless, widely available, and does not require radiation or sedation.

Typical imaging findings from patients with ADPKD reveal large kidneys with multiple bilateral cysts. Important factors in diagnosing include family history of

ADPKD, age of patient, and number of kidney cysts. Age-dependent ultrasound criteria for both diagnosis and disease exclusion have been established for patients with a positive family history. If ultrasonography results are equivocal, magnetic resonance imaging (MRI) or computed tomography (CT) may clarify the diagnosis (Pei et al 2015). In the absence of a family history, these imaging-based criteria do not apply. In several situations, multiple factors should be considered, including the age of the patient, the presence of additional associated manifestations, and findings or family history suggestive of other genetic disorders. ADPKD is the most likely diagnosis in the presence of bilaterally enlarged kidneys and innumerable cysts in each kidney, therefore other genetic diseases can be associated with kidney cysts so, when suggestive findings are noted, the differential diagnosis should be broadened (Pei et al 2015).

During the pregnancy it may be possible to detect ARPKD during the usually ultrasound scans routine. The kidneys of a baby with ARPKD may appear unusually large or bright on the scan. Early signs of ARPKD are sometimes visible during the first routine ultrasound scan carried out at week 12 of pregnancy, although the condition isn't usually detected until the second routine scan at around 20 weeks. US may demonstrate echogenic, enlarged, reniform kidneys, oligohydramnios, or an empty urinary bladder in severe cases of ARPKD (Turkbey et al 2009; Sweeney et al 2011). Severely affected fetuses with oligohydramnios may have pulmonary hypoplasia and high mortality due to pulmonary insufficiency, or multiple intrauterine compression anomalies of lethal Potter sequence. The presence of large reniform echogenic kidneys with poor corticomedullary differentiation and oligohydramnios on prenatal ultrasound examination suggests ARPKD, although other diagnoses are possible.

In infant and child the findings on renal imaging are noted as above and renal size may actually decrease with age as fibrosis progresses (Turkbey et al 2009).

1.1.4 Genetic testing

Genetic testing is not always required for diagnosis but may be helpful when imaging results are uncertain, i.e. in all that cases in which clinical phenotype overlaps with other syndromic renal cystic forms. Moreover a molecular diagnosis is required to identify living related kidney donors, or in atypical cases (e.g., early and severe

polycystic kidney disease, kidney failure without significant enlargement of the kidneys, marked discordant disease within family, marked asymmetry in disease severity between kidneys, or very mild PKD). Genetic testing is also useful in the diagnosis of sporadic PKD with no family history. Such testing can be helpful in reproductive counseling as well. This can be followed by multiplex-dependent probe amplification in cases with negative DNA sequencing results (Zerres et al 1998). The large size, molecular complexity, lack of mutational hotspot characterizing the causative genes and, in particular, the high homology with six psudogene for the first 33 exons of *PKD1* gene make the molecular diagnostics challenging.

Currently, the most common method used for molecular diagnosis of ADPKD and ARPKD is direct mutation screening by Sanger sequencing of the *PKD1 PKD2* and *PKHD1* genes respectively (Audrézet et al 2012, Bergmann et al 2004, Bataille et al 2011).

1.2 From Sanger sequencing to Next-Generation Sequencing

1.2.1 Sanger sequencing, the conventional sequencing method

Sanger sequencing, also called the chain-termination method, was developed by Frederick Sanger and colleagues in 1977 (Sanger et al 1977). It is a technique for DNA sequencing based upon the selective incorporation of chain-terminating dideoxynucleotides (ddNTPs) by DNA polymerase during in vitro DNA replication.

Classical Sanger sequencing requires a single-stranded DNA template, a DNA polymerase, a DNA primer, normal deoxynucleosidetriphosphates (dNTPs), and modified nucleotides (ddNTPs) that terminate DNA strand elongation. These ddNTPs lack a 3′-OH group that is required for the formation of a phosphodiester bond between two nucleotides, causing the extension of the DNA strand to stop when a ddNTP is added. The DNA sample is divided into four separate sequencing reactions, containing all four of the standard dNTPs (dATP, dGTP, dCTP, and dTTP), the DNA polymerase, and only one of the four ddNTPs (ddATP, ddGTP, ddCTP, or ddTTP) for each reaction. After rounds of template DNA extension, the formed DNA fragments are separated by size using gel electrophoresis in four parallel lanes representing ddA, ddT, ddC, and ddG terminators. The shortest sequences are moved further down to

the bottom of the gel and the reading can be achieved from the bottom to the top (Figure 10).

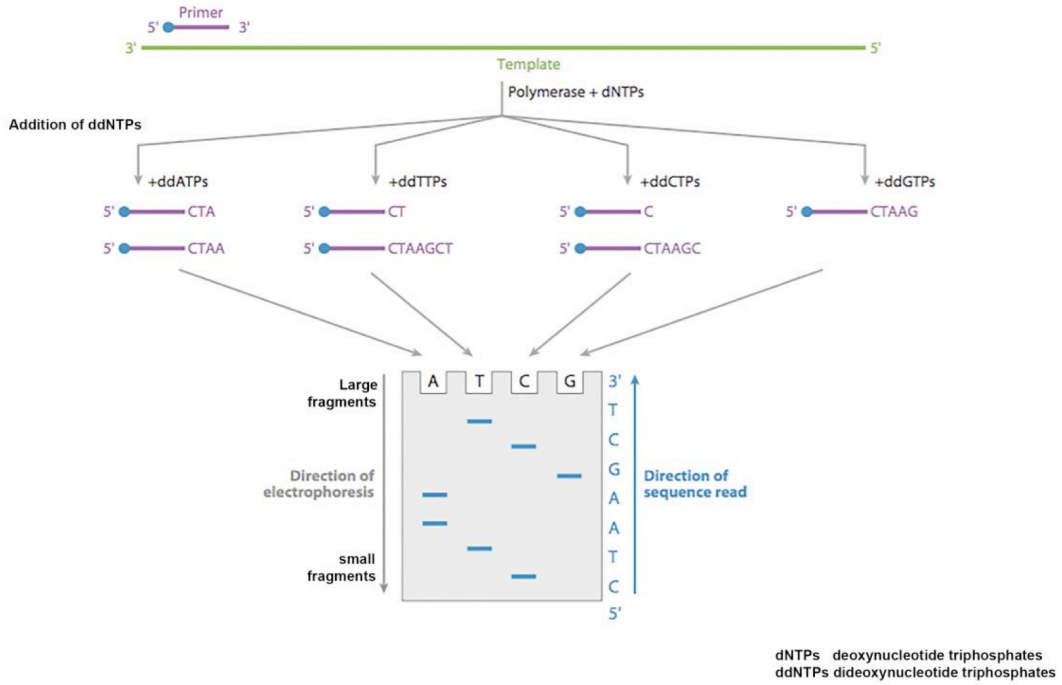


Figure 10: A complementary primer anneals to a 3' single-stranded template to start the extension of the template by adding dNTPs using polymerase enzyme. Equal amounts of the reaction mixture are placed into four tubes, and four ddNTPs are added to each tube, which terminate the sequence synthesis. The contents of the reaction tubes are then subjected to four lanes of polyacrylamide electrophoresis gel and the oligonucleotide sequences are separated according to the size and type of nucleotides. The smallest sequences are moved further down to the bottom of the gel, while the biggest sequences are remained at the top of the gel. The reading can be carried out from the bottom to the top as depicted with the blue arrow. Adapted from (Mardis, 2013).

In 1986, a company named Applied Biosystems began to manufacture automated DNA sequencing machines based on the Sanger method. These machines used fluorescent dyes to tag each nucleotide, allowing the reactions to be run in one column and read by color during the capillary elettrophoresis (Prober et al 1987). In newer dye terminator sequencing, the ddNTPs are labeled with fluorescent dyes to make the fragments readable through a laser light source at unique wavelengths in capillary gels, rather than on traditional gels. Sequencers based on Sanger sequencing produce a read length (the length of a DNA fragment that can be sequenced at one time) of 800–1000 bp. Because at one time only one read can be sequenced in one capillary of the sequencer, the total output of the run is equal to the read length. However, sequencers with multiple capillaries enable one to sequence multiple samples at a time (e.g. 8, 16, 48, or 96, etc.). The automation of 96 or 384 independent capillaries was designed to increase the throughput and facilitate sequencing of more samples (Shendure et al 2008). Furthermore, the read length of sequencing can be achieved by Sanger sequencing up to around 1000 bp and a high

level of accuracy of 99.999% can be obtained. Interestingly, Sanger sequencing was used in sequencing the entire human genome in the Human Genome Project (HGP) (Bentley 2000). The sequencing of five individuals, around 14.8 billion bp, took over nine months (Venter et al 2001). Therefore, the requirement of new platforms to provide rapid sequencing at an affordable price was needed in order to obtain the benefits of the sequencing.

1.2.2 Next-Generation Sequencing technology

Over the past 10-12 years, the advent of Next-Generation sequencing (NGS) techniques has dramatically increased the speed and reduced the costs of sequencing. The growing adoption of NGS technology and the integration of sequencing standards into laboratory practices may make NGS the new gold standard for clinical routine.

In contrast to Sanger sequencing, NGS uses a different principle in its mode of action. In all NGS platforms, following the extraction of the genetic materials, a sequencing library is prepared. Initially, the sequencing library is constructed by shearing the targeted DNA into small fragments, either by using enzymes or physical sonication, and then a ligation step takes place by attaching both ends of the fragmented DNA to adaptors. These adapters are universal sequences, specific to each platform, that can be used to polymerase-amplify the library fragments during specific steps of the process. The fragmented DNA along with the adaptors forms the sequencing library (paired end library). The sequencing library is then fixed to bead particles or solid surfaces, depending on the platform used, in order to prepare the sequencing template for clonal amplification. The clonal amplification either uses emulsion PCR or solid phase amplification depending on the platform used, which is performed in order to enrich and obtain a high number of sequencing templates to ascertain appropriate signal production in the sequencing reaction that is sufficient for a sensitive detection (Goodwin et al 2016).

Depending on the application, various NGS methods are available to suit diverse study designs and objectives, including whole-genome sequencing, exome sequencing, and targeted sequencing. Whole-genome sequencing determines the entire DNA sequence, while exome sequencing analyzes only on the coding portion

of the genome. Targeted sequencing focuses on specific genes of interest and is the most commonly used NGS method in molecular pathology (Goodwin et al 2016).

1.2.2.1 NGS platform

The variety of NGS technology features supports the coexistence of multiple platforms in the marketplace, with some having clear advantages for particular applications over others. Currently, six sequencing platforms are available (454, Illumina, SOLiD, HeliScope, Ion Torrent, PacBio), whereas a couple (StarLight and Nanopore) are in advanced development. The most salient features of the platforms are described below.

- **454** was the first commercial NGS platform acquired by Roche. 454 uses beads that start with a single template molecule, which is amplified via emPCR (emulsion PCR). Millions of beads are loaded onto a picotitre plate designed so that each well can hold only a single bead. All beads are then sequenced in parallel by flowing pyrosequencing reagents across the plate (http://www.454.com).

- **Illumina** uses a solid glass surface to capture individual molecules and bridge PCR to amplify DNA into small clusters of identical molecules. These clusters are then sequenced with a strategy similar to Sanger sequencing, except only dye-labelled terminators are added, the sequence at that position is determined for all clusters, then the dye is cleaved and another round of dye-labelled terminators is added (http://www.illumina.com).

- **SOLiD** uses ligation to determine sequences and until the most recent of Illumina's software and reagents, SOLiD has always had more reads (at lower cost) than Illumina (http://www.appliedbiosystems.com).

- **HeliScope** developed by Helicos, which was the first commercial single-molecule sequencer. Unfortunately, the high cost of the instruments and short read lengths limited adoption of this platform. Helicos no longer sells instruments, but conducts sequencing via a service centre model (http://www.helicosbio.com).

- **Ion Torrent** an instrument system that detects the release of hydrogen ions, as a result of nucleotide incorporation, quantifying changes in pH through a novel coupled silicon detector. In 2010, the first early access instruments were deployed and Ion Torrent was purchased by Life Technologies, but it is still known as Ion Torrent (http://iontorrent.com).

- **PacBio** has developed an instrument that sequences individual DNA molecules in real time. Individual DNA polymoerases are attached to the surface of microscope slides. The sequence of individual DNA strands can be determined because each dNTP has unique fluorescent label, immediately detected prior to being cleaved off during synthesis. Low cost per experiment, fast run times and cool factor generated much enthusiasm for this platform, which first early instruments were deployed in 2010 (http://www.pacificbioscience.com).

StarLight and Nanopore are the upcoming sequencing technologies aiming to longer read length and reduced cost per sample.

- **StarLight** is the more extensively Life Technologies Single Molecule Real-Team Sequencing Technology. The platform uses quantum dots to achieve single-molecule sequencing. DNA is attached to the surface of a microscope slide where sequencing occurs in similar way to PacBio. As the DNA polymerase can be replaced after it has lost activity, the sequencing can continue along the entire length of a template. The peculiar innovation is the ability to perform 3-Dimensional DNA sequencing of ultra-long DNA fragments, wherein DNA-sequence *vs* time *vs* imaging-reagent-space are simultaneously collected. Moreover, completely phased and ordered reads are simultaneously obtained, and the effective ”mate-pairs” for each DNA fragment increase combinatorially with the number of sequencers on each individual DNA fragment. This type of 3-D sequencing information is ideal for quantitating genomic structural variation and for generating de novo scaffolds for shorter read-length sequencing data. Many characteristics of the Starlight technology are known (Karrow 2010), but timing of a commercial

launch, target costs and other details are unknown (http://www.lifetechnologies.com).

- **Nanopore** is an under development method performing 'strand sequencing', a technique where intact DNA polymers pass through a nanopore, being sequenced in real time as the DNA translocates the pore. A nanopore is simply a small hole with an internal diameter of the order of 1 nanometer. Certain porous transmembrane cellular proteins act as nanopores. The theory behind nanopore sequencing is that when a nanopore is immersed in a conducting fluid and a potential (voltage) is applied across it, an electric current due to conduction of ions through the nanopore can be observed. The amount of current is very sensitive to the size and shape of the nanopore. If single nucleotides (bases), strands of DNA or other molecules pass through or near the nanopore, this can create a characteristic change in the magnitude of the current through the nanopore. DNA could be passed through the nanopore for various reasons. For example, electrophoresis might attract the DNA towards the nanopore, and it might eventually pass through it. Alternatively, enzymes attached to the nanopore might guide DNA towards the nanopore. The potential is that a single molecule of DNA can be sequenced directly using a nanopore, without the need for an intervening PCR amplification step or a chemical labelling step or the need for optical instrumentation to identify the chemical label. Nanopore technologies promise no read length associated limitation and the possibility to sequence at 25X depth of coverage the human genome in minutes at a cost of 100 dollars (https://nanoporetech.com/).

1.2.2.2 Ion Torrent PGM platform

Ion PGM was released by Ion Torrent at the end of 2010. PGM uses semiconductor sequencing technology and is the first commercial sequencing machine that does not require fluorescence and camera scanning, resulting in higher speed, lower cost, and smaller instrument size. Currently, it enables 200 bp reads in 2 hours and the sample preparation time is less than 6 hours for 8 samples in parallel.

Library construction includes DNA fragmentation, enzymatic end polishing, and adapter ligation. Amplification of library fragments occurs by a unique approach known as emulsion PCR. An oil–water emulsion is created to partition small reaction vesicles that each ideally contains one sphere, one library molecule and all the reagents needed for amplification. The beads have covalently linked adapter complementary sequences on their surfaces to facilitate amplification on the bead. Although one emPCR reaction can generate billions of templated spheres, some aspects inherent to the emPCR method, in addition to the general biases during PCR amplification, prevent optimal output. In fact, due to the double Poisson distribution behaviour, it is impossible to achieve optimal loading of one library molecule into all individual vesicles. In fact 1/3 of the vesicles will have the one molecule to one vesicle ratio, the remaining 2/3 will be either without a molecule or have more than one. In the final step spheres containing amplified DNA are selected in an enrichment step from empty spheres and the loaded spheres are deposited into the sequencing chip. The Ion torrent chip consists of a flow compartment and solid state pH sensor micro-arrayed wells that are manufactured using processes built on standard CMOS technology. The sequencing reaction take place in every chip microwell which can contain only a single template-positive beads and DNA polymerase enzyme. In every sequencing step, or flow, the chip is washed over with a specific nucleotide. The nucleotide in the flow is incorporated by all consecutive complementary nucleotides 'hanging' at the end of each template. Each incorporation releases an ion, so that the change in pH level indicates whether incorporation occurred and, if so, the number of consecutive bases incorporated (Figure 11) (Flusberg et al 2010).

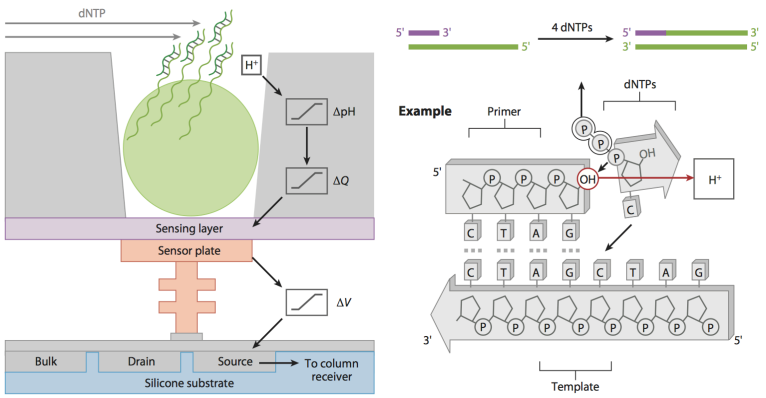


Figure 11: On the right, a schematic structure of the Ion Chip demonstrates the addition of nucleotide causing release of the hydrogen ion H+. This leads to changing of pH of the surrounding solution and the detection then occurs using a sensor underneath each well, working as a solid-state pH meter. Off-chip electronics finally digitise the voltage to digital information and the base calling. On the left is shown the sequencing by synthesis process based on pH sensing. When every nucleotide is incorporated, a hydrogen ion is released and the base is called (Mardis 2013).

2. PRINCIPAL AIM

The goal of this project is to develop and validate a broad, reliable, rapid and cost-saving genetic test for PKD patients based on NGS technology, with a sensitivity comparable with Sanger sequencing.

Up to date, Sanger sequencing represent the most used method in molecular diagnostics, however the large size and the molecular complexity of PKD causative gene, i.e. GC rich regions, high allele heterogeneity and in particular the high homology for the first 33 exons with 6 pseudogenes of *PKD1* gene make this method technically challenging, labor intensive and costly.

This innovative test may be useful for the molecular diagnostics of ADPKD and ARPKD guaranteeing to reach quickly medical reports for all patients and their presyntomatic related, so should have great consequences on disease management and patient treatment.

3. MATERIALS AND METHODS

3.1 Patients recruitment

All the patients submitted to the developed NGS protocols were recruited by Nephrology and Medical Genetics Operative Units of the Sant'Orsola University Hospital in Bologna after obtaining informed consent for genetic study.

- **ADPKD cohort:** 125 unrelated patients with an ADPKD clinical suspect were sequenced by our NGS approach with the purpose of carry out a retrospective and a prospective method validation. This cohort was constituted of 20 patients, previously investigated by other specialized Italian laboratories with Sanger sequencing, and 105 patients with unknown mutations.
- **ARPKD cohort:** 28 unrelated patients recruited were divided in two groups (A and B): Group A consisted of 15 unrelated probands with clinical suspect of ARPKD, including 2 fetuses from terminated pregnancy and 2 severely affected newborns who died shortly after birth; Group B counted 13 unrelated healthy carriers, as parents of affected ARPKD fetuses.

  5 probands (Group A) and 7 healthy carriers (Group B) had been previously analyzed with traditional sequencing by other specialized Italian laboratories. These 12 patients and the remaining 16 subjects were submitted to the NGS test in order to obtain a retrospective and prospective validation of the developed method, respectively.

3.2 DNA extraction routine

Genomic DNA was extracted from 400 µl. of peripheral blood using the semi-automatic Maxwell® 16 instrument and the Maxwell® 16 DNA Purification kit (Promega Corporation, Madison, WI, USA) and were eluted in 300 µl of steril water. For the purpose of this study, when prompted by the machine the protocol selected was "Blood" and the sample type selected was "DNA".

All samples were quantified using Nanodrop 3.0.0 spectrophotometer (Celbio S.P.A. Milano, Italia) and were normalized to 30-50 ng/µl.

3.3 Next-Generation Sequencing protocol

For the libraries constructions step, two different protocols were developed, one was the standard protocol based on Ion Ampliseq Technology used for the sequencing of *PKD2* and *PKHD1*; the second was an alternative protocol, based on target preseletion using LR-PCR, for the specific amplification of *PKD1*.

3.3.1 *PKHD1* and *PKD2* libraries construction protocol

3.3.1.1 Assay Design

The assay design for the mutational screening of *PKD2* and *PKHD1* was carried out by Ion AmpliSeq™ Designer (https://ampliseq.com/browse.action), a free design tool available online, that allow to obtain custom assay designs based on PCR target selection. The tool allowed us to obtain two different primer pools, for each gene, in order to set up ultra-high multiplex PCR for the simultaneous target amplification, i.e the entire coding sequence and flanking exon regions . A total of 122 amplicons for *PKHD1* gene (61 amplicon for each pools) and 39 amplicons for *PKD2* (18 in pool 1 and 19 in pool 2) were obtained for the 99.88% and 99.9% of target coverage respectively.

3.3.1.2 Libraries amplifications

Libraries were carried out using 10ng of DNA for each pool using Ion AmpliSeq^{TM} Library Kit 2.0 (Life Technologies, Carlsb8ad, CA, USA). The reaction mix and the termalcycler conditions are reported below (Table 3).

Reaction mix for sample	Volume (μl)
Ion AmpliSeq™ HiFi Master Mix (5X)	2
Ion AmpliSeq™ Primer Pool 1 (or for Pool 2) (2X)	5
gDNA (10 ng)	1
H_2O Nuclease-free	2
Final volume	10

Temperature T(°C)	Time	Cycles
99 °C	2′	1
99 °C	15″	19 for *PKHD1*/ 21 for *PKD2*
60 °C	4′	
10 °C	Hold	End

Table 3: reaction mix and termalcycler conditions for libraries construction.

3.3.1.3 Primers digestion

Partial digestion of primers was carried out by adding 2 µL of FuPa reagent and the thermocycler was set to the following programme: 50°C for 10 min, 55°C for 10 min, 60°C for 20 min and then held at 10°C for up to 1 hour.

3.3.1.4 Adapter and barcodes ligation

First for every barcoded adapter a final diluition of 1:4 was prepared by mixing Ion P1 adapters and Ion Xpress™ Barcode (X), then the reaction mix reported in table below was added to each samples (Table 4).

Reaction mix for sample	Volume (μl)
Switch solution	2
Barcode adapter mix (1:4)	1
DNA ligase	1
H_2O nuclease-free	1
Final volume	6

Temperature T(°C)	Time
22 °C	30′
72 °C	10′
10 °C	Hold

Table 4: reaction mix for barcoding stage and termalcycler conditions.

3.3.1.5 Libraries purification

A purification step then took place using the Agencourt® AMPure® XP reagent (Beckman Coulter, Brea, CA, USA) by adding 22,5 µL of the beads to each sample (1.8X of the original sample). The mixture was incubated for 5 minutes at room temperature and then placed in a magnetic rack for 2 minutes. As the required sequencing libraries were in the pellet, the supernatant was discarded without touching the beads. Two ethanol washes were performed by adding 100 µL of 70%

ethanol and then the pellet was left to air-dry for up to 5 minutes. Finally the sample was eluted using 25 µL of low Tris-EDTA pH 8.0.

3.3.1.6 Libraries quantification and pooling

For each library a 100-fold dilution was carried out by mixing 2 µL of the sample with 198 µL of nuclease-free water and three 10-fold serial dilutions were prepared of the *E. coli* DH10B Ion Control Library (~68 pM; from the Ion Library Quantitation Kit) at 6.8 pM, 0.68 pM, and 0.068 pM. Each standard, negative control and the samples were analysed in duplicate using Ion Library Quantitation Kit (Life Technologies, Carlsbad, CA, USA). Reaction mix and PCR conditions are shown in tables below (Table 5).

Reaction mix for sample	**Volume (µl)**
2X Ion TaqMan® Master Mix	5
20X Ion TaqMan® Assay	0.5
Library (1:100) or standards DH10B	4,5
Final volume	10

Temperature	**Time**	**Cycles**
50 °C	2′	Hold
95 °C	20″	Hold
95 °C	1″	40
60 °C	20″	
10 °C	End	Hold

Table 5: reaction mix for quantification stage and termalcycler conditions.

Following quantitation the sequencing libraries were pooled together to 9 pM as to ensure they were suitable for template preparation.

3.3.2 *PKD1* libraries construction protocol

3.3.2.1 Assay design

In order to obtain the specific and selective *PKD1* genuine gene amplifications, rare mismatch sequencing between *PKD1* and pseudogenes had been used to develop *PKD1* locus-specific amplicons. Primer set design was carried out using two different tools available online: Primer3 (http://primer3.ut.ee/) was used for primers design, PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to checked the target specificity and Operon Oligo Analyzer (http://eu.idtdna.com/calc/analyzer) was finally used to exclude the self complementarity between primers. The primers set designed is reported in Table 6.

Exons covered	Forward primer (5'-3')	Reverse primer (5'-3')	Kb
1	CGCAGCCTTACCATCCACCT	TCATCGCCCCTTCCTAAGCA	2,3
2-12	CCAGCTCTCTGTCTACTCACCTCCGCATC	CCACGGTTACGTTGTAGTTCACGGTGACG	8,7
13-21	TGGAGGGAGGGACGCCAATC	ACACAGGACAGAACGGCTGAGGCTA	7,9
22-33	CCGCCTCTCCTCTCTCCCCTCCT	ACCTGCTGGATCAGGTCTTC	7,9
34-41	AGGAGGGGGCTCTGAAGCTCACCCTT	AGAGGGTGCGGGTCAGTAGG	5,7
42-46	CCAGGAGCCCACCCTCACTC	CCATTCTGCCTGGCCCTC	2,7

Table 6: Oligonucleotide primers for long-range specific templates from exons 1 to 46 of the *PKD1* gene

LR-PCRs were performed using two different kits: GoTaq® Long PCR Master Mix, (Promega, Madison, WI, USA) and PrimeSTAR GXL DNA Polymerase (Takara Bio USA, Inc). Reaction mixes and termalcycler conditions are reported in the table below.

***LR*-PCR Exon 1 (*TARGET*: 2278bp)**

GoTaq® Long *PCR* Master Mix (Promega, Madison, WI, USA)	
Reaction mix	**µl**
H_2O	6.5
GC rich Buffer (5X)	10
Master Mix (2X)	25
Primer forward 1F (20 pmol/µl)	1.25
Primer reverse 1R (20 pmol/µl)	1.25
gDNA (35-50ng)	6
Final volume	50

Temperature (°C)	Time	Cycles
95	2′	1
95	30″	35
64	30″	
72	3′	
72	10′	1
10	HOLD	END

LR-PCR *Exons 2-12 (TARGET: 8727bp)*

PrimeSTAR GXL DNA Polymerase (Takara)	
Reaction mix	**µl**
H_2O	25
Buffer (5X)	10
dNTPS mixture (2.5 mM each)	5
Primer Forward 2-12F (20 pmol/µl)	1.5
Primer reverse 2-12R (20 pmol/µl)	1.5
Taq (1.25 U/µl)	1
gDNA (35-50ng)	6
Final volume	50

Temperature (°C)	Time	Cycles
98	10″	35
68	8'30"	
10	∞	END

LR-PCR Esons 13-21 (TARGET: 78927bp)

GoTaq® Long PCR Master Mix (Promega, Madison, WI, USA)	
Reaction mix	**µl**
H_2O	6.5
GC rich buffer (5X)	10
Master Mix (2X)	25
Primer forward 13-21F (20 pmol/µl)	1.25
Primer reverse 13-21R (20 pmol/µl)	1.25
gDNA (35-50ng)	6
Final volume	50

Temperature (°C)	Time	Cycles
95	2'	1
95	30"	35
66	30"	
72	8:30"	
72	10'	1
10	∞	HOLD

LR-PCR Esons 22-33 (TARGET: 78927bp)

GoTaq® Long PCR Master Mix (Promega, Madison, WI, USA)	
Reaction mix	**µl**
H_2O	16,5
Master Mix (2X)	25
Primer forward 22-33F (20 pmol/µl)	1.25
Primer reverse 22-33R (20 pmol/µl)	1.25
gDNA (35-50ng)	6
Final volume	50

Temperature (°C)	Time	Cycles
95	2'	1
95	30"	10
72(-0,5)	30"	
73	7:30'	
95	30"	25
64	30"	
73	7:30'	
73	5'	1
10	∞	END

LR-PCR Esons 34-41 (TARGET: 5700bp)

GoTaq® Long PCR Master Mix (Promega, Madison, WI, USA)	
Reaction mix	**µl**
H_2O	16.5
Master Mix (2X)	25
Primer forward 34-41F (20 pmol/µl)	1.25
Primer reverse 34-41R (20 pmol/µl)	1.25
gDNA (35-50ng)	6
Final volume	50

Temperature (°C)	Time	Cycles
95	2'	1
95	30"	35
70	6'	
72	10'	1
10	∞	END

***LR-PCR* Esons 42-46 (*TARGET*: 2763bp)**

GoTaq® Long PCR Master Mix (Promega, Madison, WI, USA)	
Reaction mix	**µl**
H_2O	6.5
GC rich buffer (5X)	10
Master Mix (2X)	25
Primer forward 42-46F (20 pmol/µl)	1.25
Primer reverse 42-46R (20 pmol/µl)	1.25
gDNA (35-50ng)	6
Final volume	50

Temperature (°C)	Time	Cycles
94	2'	1
94	30"	35
62	30"	
72	3'	1
72	10'	
10	∞	END

Table 7: all reaction mix and termalcycler conditions for LR-PCR

3.3.2.2 LR-PCR purification, quantification and pooling

All LR-PCR products were then purified by filtrations over Montage PCR 96 clean up plates (millipore Corp., Billerica, MA, USA) and quantified using Qubit dsDNA BR Assay Kit and Qubit® Fluorometer 2.0 instrument (Invitrogen, Carlsbad, CA, USA). Finally, all LR-PCR, for each patient, were pooled together in equimolar manner (50pM each) in a final volume of 100 µl.

3.3.2.3 Sonication and end repair

In order to obtain uniform genomic libraries constituted of 200bp amplification fragments, all the pool were sonicated using Bioruptor® Pico (Diagenode) for 13 cycles at full amplitude 30 seconds ON/ 30 seconds OFF. The tubes containing the DNA were held in an ice bath during sonication. The sticky-end generated by fragmentation step were then converted in blunt-end using Ion Plus Fragment Library Kit (Life Technologies, Carlsbad, CA, USA). The equimolar amplicons of 100 ng were mixed with 20 µl of 5X end repair buffer and 1 µl of end repair enzyme in a 1.5 ml low binding tube, then the tube was incubated at room temperature for 20 minutes. The fragmented DNA was purified by Agencourt® AMPure® XP reagent by adding 1.8X of the total volume of the fragmented sample as shown in section 3.3.1.5.

3.3.2.4 Barcode and adapters ligation, nick repair and purification

The purified DNA fragments were then transferred ligated to adaptors and nick-repaired following the reaction mix and the termalcycler conditions as indicated in table 8.

Reaction mix	Volume (µl)
DNA	~25
Ligase Buffer (10X)	25
Ion P1 Adapter	2
Ion Xpress™ Barcode X[1]	2
dNTP Mixture	2
Nuclease-free Water	49
DNA Ligase	2
Nick Repair Polymerase	8
Final volume	100

Temperature (°C)	Time	Cycles
25 °C	15′	Hold
72 °C	5′	Hold
4 °C	up to 1h	Hold

Table 8: reaction mix and termalcycler conditions for barcode-adapters ligations and nick repair
[1] refers to the barcode number.

Following the run on the thermocycler, the ligation reactions were purified with 1.4X of Agentcourt® AMPure® reagent for 200base-reads sequencing as discussed above in section 3.3.1.5.

3.3.2.5 Size selection of DNA libraries and quantifications

In order to select only ligated library fragments of ~250 bp in size (~200bp of fragment library + 40bp adapters + 10 bp barcode) the size selection was performed using E-Gel™ iBase™ unit and E-Gel™ Safe Imager™ transilluminator instruments and E-Gel™ SizeSelect™ 2% Agarose precasted Gel (Invitrogen, Carlsbad, CA, USA).

10 µl of 50bp DNA ladder (1 µg/µL, diluited 1:40) and 20 µl of ligated libraries were loaded to the loading well on the top of the E-Gel™ SizeSelect™ 2% Agarose so all the recovery well were filled with 20 µl of Nuclease-free water. When prompt by the machine "SizeSelected™ 2%" program was selected and the run was monitored until the 370bp sized ladder band (350bp) was in the reference line then the run was stopped and the solution within recovery wells (at the bottom of the precasted gel) was collected in 0,2 ml tubes.

Prior to preparing the sequencing template, the size-selected sequencing libraries were quantified using the Ion Library Quantitation Kit (Life Technologies, Carlsbad, CA, USA) as described in section 3.3.1.6 and the different libraries were pooled together at final concentration of 9pM.

3.3.3 Sequencing template preparation: Emulsion PCR

Sequencing template preparation was performed using the Ion PGM Hi-Q OT2 Kit and Ion OneTouch™ 2.0 instrument (Life Technologies, Carlsbad, CA, USA). Reaction mix was prepared in the 2 ml tube containing 800 µl of Ion PGM™ Hi-Q™ Reagent Mix as reported in the table 9.

Reaction mix	Volume (µl)
Nuclease-free water	25
Ion PGM™ HI-Q™ Enzyme mix	50
Diluited library (9pM)	25
Ion PGM™ HI-Q ISPs	100

Table 9: reaction mix used for emulsion PCR

The amplification mixture was added to the sample port of the Ion PGM™ OneTouch Plus Reaction Filter and then 1700 µl of Ion OneTouchTM Reaction Oil was added to the filter sample port, as well. The filled Ion PGM™ OneTouch Plus Reaction Filter was inverted and installed into the three holes on the top stage of Ion OneTouch™ 2.0 instrument and two recovery tubes with 150 µl of Ion Onetouch™ Breaking Solution were inserted in the instrument centrifuge. Finally, the clonal amplification was performed by running the emulsion mixture setting the "Ion PGM Hi-Q OT2 200" protocol.

At the end of the run and final spin, all but 100 µl of the Ion PGM™ OT2 Recovery Solution (surnatant) was removed from each recovery tubes, then 500 µl of Ion One Touch™ Wash Solution was added in each tube and the ISPs (Pellet) were mixed by pipetting and were combined into a new labeled 1.5 ml eppendorf LoBind™ Tube. The tube was centrifuged for 2.5 minutes at 15,500 x g and all but 100 µl of the surnatant was discarded.

The quality of the unenriched template was assessed using Qubit® 2.0 Fluorometer.

3.3.4 Template-positive ISPs enrichment

The template-positive ISPs entichment was carried out using Ion OneTouch™ ES instrument and Ion PGM™ Enrichment Beads Kit (Life Technologies, Carlsbad, CA, USA).

Dynabeads® MyOneTM Streptavidin C1 Beads were created by mixing the tube for 30 seconds to resuspende the beads, immediately 13 µl of Dynabeads® MyOne™ Streptavidin C1 Beads was transferred into a 1,5mL Eppendorf LoBind tube and placed on a magnetic rack for 2 minutes. The surnatant was removed carefully without disturbing the pellet, then 130 µl of MyOne™ Beads Wash Solution was added and beads were mixed by pipetting.

An opened 0.2 mL tube containing 10 µl of Neutralization Solution was insert into the hole at the base of instrument and the 8-well strip (Figure 12) was filled as reported above.

1. 100 µl of entire template-positive ISP sample;
2. 130 µl of Dynabeads® MyOneTM Streptavidin C1 Beads resuspended in MyOne™ Beads Wash Solution:
3. 300 µl of Ion OneTouch™ Wash Solution;
4. 300 µl of Ion OneTouch™ Wash Solution;

5. 300 µl of Ion OneTouch™ Wash Solution;
6. Empty;
7. 300 µl of freshly prepared Melt-Off Solution (125mM NaOH and 0.1% Tween™ 20 detergent);
8. Empty.

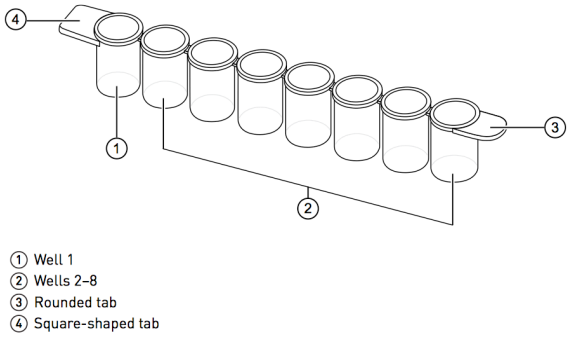


Figure 12: the 8-well strip using for template enrichment

3.3.5 Ion Torrent PGM™ sequencing

Ion PGMTM sequencing platform was cleaned and initialized following the manufacture's instructions.

The template-positive ISPs enriched sample was centrifuged for 2.5 minute at 15,500 x g, then all but 15 µl of surnatant was carefully discarded without disturbing the pellet and 3 µl of sequencing primer was added, finally the sample was incubated in termalcycler following the condition reported in table 10.

Temperature (°C)	Time	Cycles
95 °C	2′	1
37 °C	2′	1

Table 10: termalcycler conditions for annealing primer.

Following the incubation step, 3 µl of Ion PGM™ Hi-Q sequencing polymerase was added and the sample was incubated at room temperature for 5 minutes.

The Ion 318™ v2 Chip was assessed for sequencing. First of all, the chip was tilted to 45 degrees to discard the liquid inside and a pipette tip was firmly inserted into the loading port, then it was placed upside-down in the centrifuge adapter bucket and the bucket was transferred to the MiniFuge with the chip tab pointing in and it was centrifuged for 5 seconds. After liquid removal, the chip was loaded slowing into

loading port with 30 µl of sequencing reaction. After three different step of centrifuged of 30 seconds, alternating the tab chip positions (in and out for each cycle) the reaction sequencing was mixed by tilting the chip to 45 degrees by pipetting the sample in and out of the chip.
At the end of centrifuge cycles, all the liquid was removed from the loading port and the chip was finally positioned in the Ion PGM™ sequencer.

3.3.6 NGS data analysis

Ion PGM™ Sequencer transferred the row sequencing data to the Torrent Server, that converted this one to base calls and the Torrent Suite™ generated a summary sequencing report and different files for analysis including Fastq files, Binary Alignment Map (BAM) in conjunction with Binary Alignment Index (BAI) and Variant Call Format (VCF) files. The sequencing samples were aligned to two reference genome in order to obtain Variant caller Files: *PKD2* and *PKHD1* reads were aligned to human genome (hg19), whereas *PKD1* reads were aligned against an artificial reference genome based on chromosome 16 of hg19, where all the nucleotides outside the *PKD1* locus were masked and replaced with "*Ns*".
The BAM/BAI files, generated following the alignment, were visualized using Integrative Genome Viewer (IGV) Software. IGV was used to assess the depth of coverage of the sequencing reads, zygosity, quality of the sequencing reads and the mapping quality.
All the VCF files were uploaded into Ion Reporter software (https://ionreporter.thermofisher.com/ir/) selecting the Annotation Variant workflow in order to associated to each variant the nucleotide change in mRNA transcript, the aminoacidic change, the exons or IVS and the function. Moreover the Ion Reporter can readily access several prediction programs used to evaluating the pathogenic potential of missense change, providing important information about the pathogenicity of the variants analyzed (see section 3.3.6.1).

3.3.6.1 Variant filtering strategy

All the variant found were classified following the The American College of Medical Genetics and Genomics 2015 (Richards et al 2015) in 5 principal categories:

1. Pathogenic: all the Loss of Function (LOF) mutations as nonsense, frameshift, indel non-frameshift, canonical splicing variants, and the SNP occurring into the ATG starting codon;
2. Likely Pathogenic: all the missense described as pathogenic variants listed in Autosomal Dominant Polycystyc Kidney Disease Mutation Database (PKDB: http://pkdb.mayo.edu) for ADPKD patients, Humgene (http://www.humgen.rwth-aachen.de) for ARPKD patients and in the Leiden Open Variations Database (LOVD3.0: http://www.lovd.nl/3.0/home), and the novel missense variants with a MAF < 0,002 (http://exac.broadinstitute.org), predicted as pathogenic by bioinformatic tools as Polyphen 2.0 (http://genetics.bwh.harvard.edu/pph2/), SIFT BLink (http://sift.jcvi.org/www/SIFT_BLink_submit.html);
3. Variant of Uncertain Significance (VUS) all novel missense rare variants with discordant prediction results.
4. Likely Benign: all frequent missense variants and rare synonims and intronic (near the canonical splicing site) variants predicted as not affects function by Human Splicing finder (http://www.umd.be/HSF3/);
5. Benign: all the deep intronic variants and frequent synonyms variant in non-conserved sequence domains.

3.4 Validation of the NGS data by Sanger sequencing

Sanger sequencing was performed in order to validate all the variants classified as pathogenic, likely pathogenic and VUS.

For *PKD1* validation variants, LR-PCR products were obtained starting to a new DNA extraction as described in section 3.3.2.1, then the obtained amplicons were purified as described in section 3.3.2.2 and the purified template was used to perform nested PCR by specific primers designed to amplified the region containing the variant.

The nested PCR were purified again following the procedure described in section 3.3.2.2 and then sequenced by Sanger technique using ABI Prism 3730 DNA Analyzer (Applied Biosystem, Foster City, CA, USA). The reaction mix and the termalcycler conditions are reported in table 11.

Reaction mix	Volume (µl)
Nuclease-free water	4
Big-Dye™ Buffer (5X)	2
Big-Dye™ Terminator 3.1	1
Primers (3,2 pmol/μl)	1
Purified Template (25 fmol)	2
Final volume	10

Temperature (°C)	Time	Cycles
96 °C	30″	1
96 °C	10′	25
60 °C	3′	
10 °C	4	Hold

Table 11: reaction mix and termalcycler conditions for Sanger sequencing reaction.

For *PKHD1* and *PKD2* variant validations the LR-PCR step was avoided and the specific sequences including the variants were directly amplified and sequenced following the instruction reported above.

3.5 Multiplex Ligation-dependent Probe Amplification: MLPA

In order to identified large genomic rearrangement and copy number variations (CNVs), all the ADPKD patients resulting wild-type to NGS test were submitted to MLPA using commercial kit containing specific probe for *PKD1* and *PKD2* gene (Salsa® MLPA® P351*PKD1* probe mix and Salsa® MLPA® P352 *PKD1-PKD2* probe mix (MRC Holland, Amsterdam, NL)).

The protocols consisted of three different steps:

- **DNA denaturation and probe ibridization:** genomic DNA, freshly extracted, was normalized on a concentration of 20ng/µl then was incubated at 98°C for 10 minutes. In the meanwhile, a probe ligation mix was prepared and added to each samples and the samples were incubated in termalcycler as reported in table 12.

Ibridization ligation mix	Volume (µl)	Temperature (°C)	Time	Cycles
Salsa® MLPA® P351*PKD1* probe mix or Salsa® MLPA® P352 *PKD1-PKD2* probe mix	1.5	95	1’	1
Salsa® MLPA® Buffer	1.5	60	16-20h	HOLD
Final volume	3			

Table 12: ibridization reaction mix and termalcyler conditions

- **Probe Ligation**: the following day all the probes were ligated as reported in table 13.

Ligation mix	Volume (µl)
Nuclease-free H2O	25
Ligase Buffer A	3
Ligase Buffer B	3
Salsa Ligase-65	1
Final volume	32

Temperature (°C)	Time	Cycles
54	15’	1
98	5’	1
20	END	HOLD

Table 13: ligation mix and termalcycler conditions.

- **Probe Amplification**: all the ligated probes were amplified following the reaction mix and termalcycler conditions reported in table 14.

Amplification mix	Volume (µl)
Nuclease-free H2O	7.5
Salsa PCR primer	2
Salsa Polymerase	0.5
Final volume	10

Temperature (°C)	Time	Cycles
95	30’’	34
60	30’’	
72	1’	
72	20’	
10	END	HOLD

Table 14: amplification probe mix and termalcycler conditions.

At the end of the probe amplification 1 µl was used for capillary electrophoresis on ABI Prism 3730 DNA Analyzer (Applied Biosystem , Foster City, CA, USA) and the electopherograms were analyzed by Coffalyzer software (MRC Holland, Amsterdam, NL).

4. RESULTS AND DISCUSSIONS

4.1 Critical assay analysis

The Next-Generation Sequencing technology, characterized by high throughput and convenient benchtop workflows, was used to develop a rapid and cost-effective molecular diagnostic method for polycystic kidney disease.

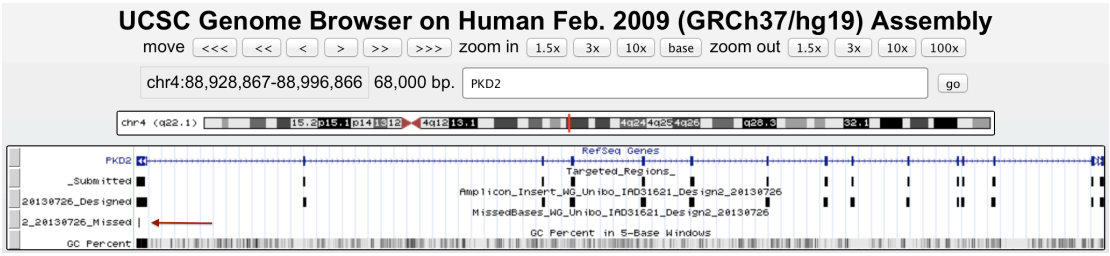
The NGS platform used in this work was the Ion Torrent Personal Genome Machine™ (Life Technologies, Carlsbad, CA, USA), a highly suitable instrument for targeted-resequencing assay for DNA, RNA and small genomes.

For the purpose of this study two different protocols were developed, one was based on Ion Ampliseq™ technology for the target-resequencing of *PKD2* and *PKHD1* genes, and one was based on a target pre-selection strategy for the molecular screening of *PKD1*.

Two different custom panels were obtained for *PKD2* and *PKHD1* using Ion Ampliseq Designer™. The overlapping amplicons generated by the software, divided in two pools for each gene, allowed us to achieve ultra-high multiplexing-PCR for the complete coverage of the entire coding sequences, i.e. exons and their flanking regions.

The custom *PKD2* and *PKHD1* panels, respectively formed by 39 and 122 overlapping amplicons, provided for the simultaneous amplification of the entire target regions for each genes in only two multiplex reactions with 99.9% and 99.88% target coverage respectively (Figure 13).

All the missed regions, 2 bp in exon 2 of *PKD2* and 15bp in exon 65 of *PKHD1,* were sequenced by Sanger sequencing.



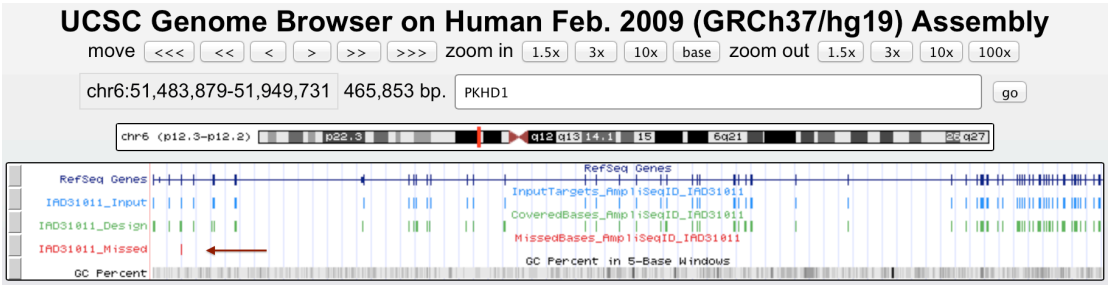


Figure 13: the output of Ion Ampliseq Designer for *PKD2* (**A**) and *PKHD1* (**B**) were reported in figure. The first line (blue) represents the gene reference sequence (hg19); full squares and line indicate exons and intronic regions respectively. The second line (light blue) reports submitted regions and the third (green) and fourth (red) lines represent respectively the covered target and the missed regions (indicated by the red arrow).

The library barcoding allowed the simultaneous sequencing of more patients *per* run. Despite the considerable dimension of *PKHD1* target region an average coverage of ≥ 300X was obtained for the simultaneous sequencing of 12 patients, whereas an average coverage of over 400X was obtained submitting 24 patients in a single run for *PKD2* screening

Although the Ampliseq-based protocol proved to be effective and suitable for the *PKD2* and *PKHD1* sequencing, the Ampliseq technology was not reliable for *PKD1* specific sequencing. The majority of the overlapping amplicons generated by Ampliseq Designer^{TM} caused the unspecific co-amplification of homologus regions, leading to a substantial loss of sensitivity and specificity, especially in the 1-33 exons encompassing region.

To improve *PKD1* sequencing assay and to eliminate the unspecific pseudogene amplification, a LR-PCR-based strategy was developed (Figure 14A). By using computational analysis, a set of six oligonucleotide primer pairs was designed that matched the single-copy gene and rare sequence differences between *PKD1* and the pseudogenes. The unique LR-PCR oligonucleotides were carefully designed to cover approximately 35,1 kb of genomic sequence, including entire coding sequence, exon-intron boundaries and most of the 5'- and 3'-untranslated regions.

The agarose gel electrophoresis with ethidium bromide staining of the LR-PCR products demonstrated that the fragment dimensions corresponded exactly to the expected base pair for *PKD1* (Figure 14A-C).

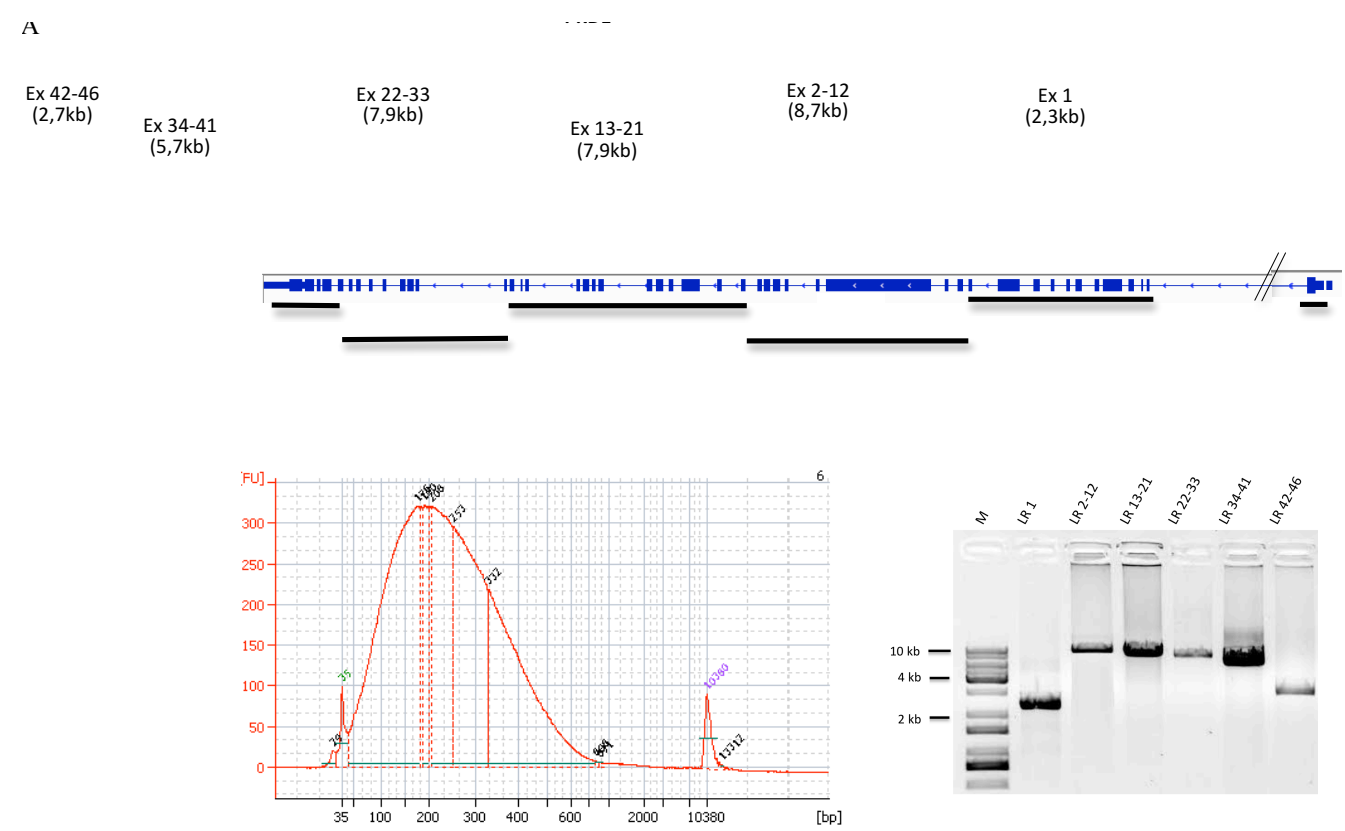


Figure 14: visualization of *PKD1* NGS workflow. *PKD1* was amplified as 6 locus-specific LR-PCR products. A: Map of the *PKD1* gene showing the position of the 6 primer pairs used for LR-PCR amplification of the coding region. B: LR-PCR products from each patients were pooled together in equimolar ratio and sonicated to obtained 200bp genomic library than were assessed for quality using Agilent Bioanalizer Instrument. **C:** Amplification quality was verified using agarose gel electrophoresis.

All LR-PCR products were purified, quantified and pooled in equimolar manner in order to obtain a single pool for each patient that underwent sonication with the purpose of creating a 200 bp fragment genomic library (Figure 14B).

To evaluate the NGS workflow, individually barcoded library 16 patients were submitted to this alternative protocol and the reads were visualized by Integrative Genome Browser software (Figure 15) in order to estimate the target coverage and the reads quality. As it is shown in the figures, despite 100% of target coverage beeing obtained (Figure 15A), a low mapping quality score characterized some reads (Figure 15B), in particular, all the reads which sharing high homology with pseudogenes regions.

It occurred because, using the assembly hg19 as reference genome, the reads sharing high homology with pseudogenes misaligned in pseudogenes loci and were recognized as “errors” by the software. To solve this problem, we created an artificial genome, based on chromosome 16 annotated in hg19, where all the nucleotides outside *PKD1* locus were masked and replaced with “Ns”. Thanks to this artificial genome, a high mapping score was obtained for all the reads and the test sensitivity and specificity improved (Figure 15C). Moreover this procedure allowed maintaining the genomic coordinates of the variants and enabling subsequent variant annotations using standard NGS analysis software.

A

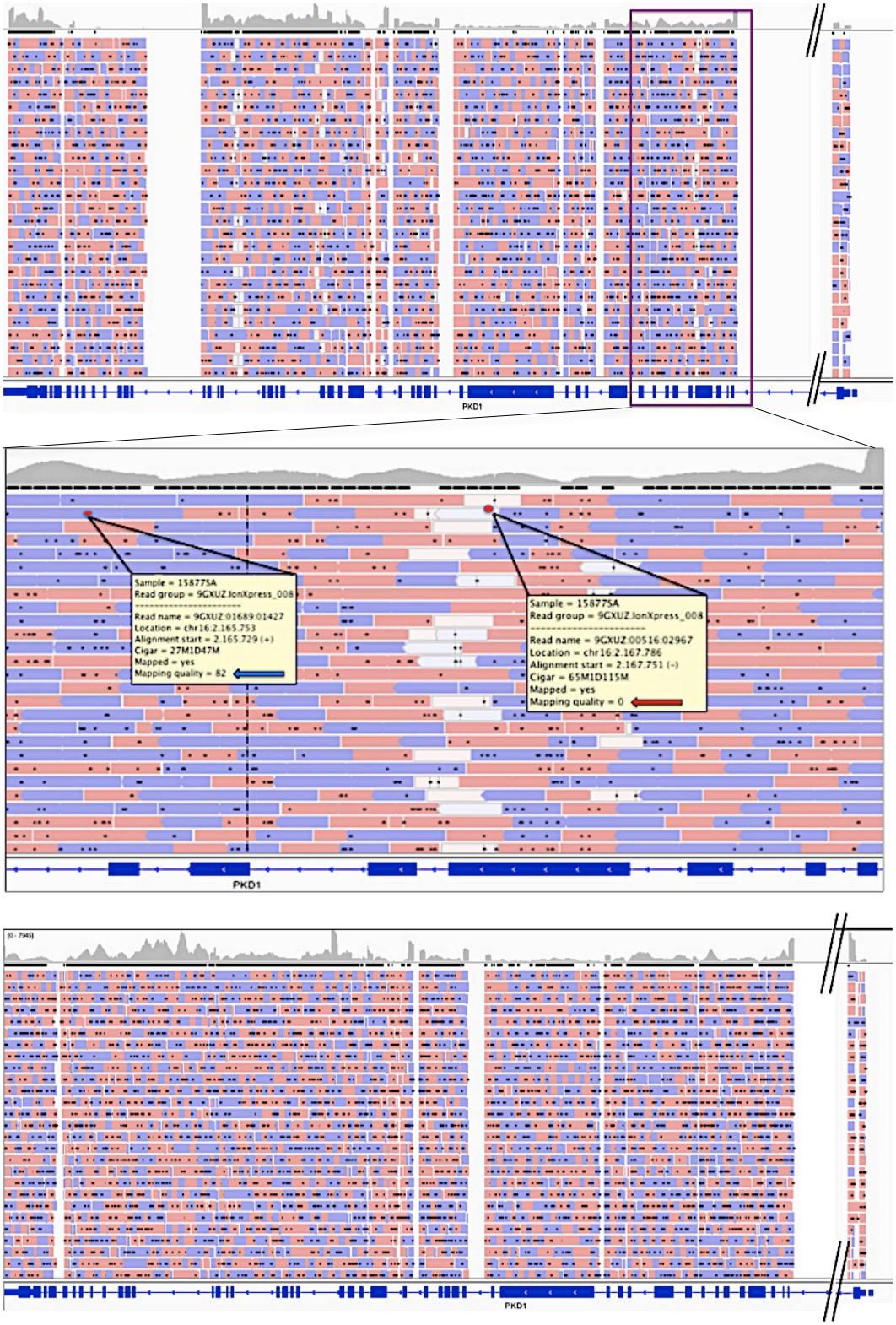


Figure 15: visualization of *PKD1* mapped reads using assembly hg19 (**A** and **B**) and the Chr16 modified genome (**C**) as reference sequences. Forward and reverse reads are represented in pink and blue respectively; all reads characterized by low mapping quality score are indicated in white. **A**): Using hg19 as reference genome several reads had low mapping quality score producing a missed variant calling. **B):** a detail in 2-8 exons encompassing regions is shown and quality score mapping is reported for "good" (blue arrow) and "bad" (red arrow) reads. **C):** the use of artificial Chr16 modified genome allows to obtaining high mapping quality score for all reads increasing sensibility and specificity of *PKD1* screening test.

4.2 NGS protocols: retrospective validation

To determine the accuracy of our NGS protocols, we collected and sequenced DNA from 20 ADPKD and 12 ARPKD previously characterized patients. All these patients had been sequenced by Sanger by other specialized Italian laboratories and were used to conduct a blind analysis in order to compare the two techniques.

- **ADPKD**

In a limited 20-sample validation cohort which included 19 patients with known mutations in *PKD1* and one in *PKD2,* the NGS protocols showed 100% analytic sensitivity in detecting all the pathogenic variations (Table 15).

Table 15: schematic representation of retrospective validation phase

Previously characterized mutation (SANGER)					Confirmed mutations
Gene	**c.DNA**	**Protein**	**EX/IVS**	**n° of unrelaed patients**	**NGS PROTOCOL**
PKD1	c.803dup	p.Phe269Leufs*102	5	1	✓
PKD1	c.1295C>T	p.Ala432Val	6	1	✓
PKD1	c.2098-2_2109del	p.(?)	IVS10	1	✓
PKD1	c.2494dupC	p.Arg832Profs*40	11	1	✓
PKD1	c.2695C>G	p.Leu899Val	11	1	✓
PKD1	c.2986-2_2987del	p.(?)	IVS12/EX13	1	✓
PKD1	c.4429del	p.Leu1479Trpfs*55	15	1	✓
PKD1	c.4494C>G	p.Tyr1498*	15	1	✓
PKD1	c.6040C>T	p.Gln2014*	15	1	✓
PKD1	c.7663G>A	p.Val2555Met	19	1	✓
PKD1	c.8000C>A	p.Ala2667Glu	21	1	✓
PKD1	c.8311G>A	p.Glu2771Lys	23	2	✓
PKD1	c.8428G>T	p.Glu2810*	23	1	✓
PKD1	c.8860G>T	p.Glu2954*	24	1	✓
PKD1	c.9404 C>T	p.Thr3135Met	27	1	✓
PKD1	c.9713-2A>G	c.9713-2A>G	IVS28	1	✓
PKD1	c.11257_11269+3del	p.(?)	EX39/IVS39	1	✓
PKD1	c.11819dup	p.Leu3941Alafs*20	43	1	✓
PKD2	c.540_555dup	p.Arg186Glyfs*32	1	1	✓

Table 15: Are listed all the previously identified mutations with traditional sequencing method and the NGS confirmation. For each mutation are reported: the gene, the nucleotide change in mRNA transcript (*PKD1*: NM_001009944.2 and *PKD2*: NM_000297.3), the aminoacidic change, the exons or IVS (intronic variant sequence number), and the number of times in which the mutations were found in unrelated patients. ✓ indicated the confirmed mutations with our NGS protocol.

The data shown in table 15 demonstrates that both ampliseq-based and the LR-PCR-based NGS protocols were specific and accurate for the correct screening of *PKD1* and *PKD2* gene. Moreover, the identification of the frameshift c.540_555dup, p. Arg186Glyfs*32 in two unrelated patients demonstrates the test robustness.

- **ARPKD**

To test the efficacy of *PKHD1* NGS Ampliseq-based protocol, a limited cohort of 12 previously characterized patients was submitted to our assay.

The cohort included 5 probands with an ARPKD clinical phenotype of and 7 obligated heterozygotes (healthy carriers) for a *PKHD1* pathogenic variant as parents of an affected child. Three of these healthy carriers (9, 11, 12) had received a previously negative molecular diagnosis in *PKHD1* and in two probands (1, 2) only one pathogenic variant had been found.

Table 16: schematic representation of retrospective validation phase

n. of unrelated	Identified mutation with NGS protocol			Previously identified mutation (SANGER)		
	ex	cDNA	Protein	ex	cDNA	Protein
1	36	c.5830G>A	p.Asp1944Asn	36	c.5830G>A	p.Asp1944Asn
	52	**c.8206T>G**	**p.Trp2736Gly**	-	-	-
2	32	c.4870C>T	p.Arg1624Trp	32	c.4870C>T	p.Arg1624Trp
	61	**c.10444C>T**	**p.Arg3482Cys**	-	-	-
3	32	c.4292G>A	p.Cys1431Tyr	32	c.4292G>A	p.Cys1431Tyr
	36	c.5895dup	p.Leu1966Thrfs*4	36	c.5895dup	p.Leu1966Thrfs*4
	30	**c.3407A>G**	**p.Tyr1136Cys**	-	-	-
4	33	c.5336A>G	p.Asn1779Ser	33	c.5336A>G	p.Asn1779Ser
	54	c.8470C>T	p.Gln2824*	54	c.8470C>T	p.Gln2824*
5	32	c.4417C>T	p.Gln1473*	32	c.4417C>T	p.Gln1473*
	32	c.5068T>C	p.Ser1690Pro	32	c.5068T>C	p.Ser1690Pro
6	15	c.1197delC	p.Leu400Cysfs*13	15	c.1197delC	p.Leu400Cysfs*13
7	3	c.107C>T	p.Thr36Met	3	c.107C>T	p.Thr36Met
8	55	c.8581A>G	p.Ser2861Gly	55	c.8581A>G	p.Ser2861Gly
	61	**c.10444C>T**	**p.Arg3482Cys**	-	-	-
9	**52**	**c.8206T>G**	**p.Trp2736Gly**	-	-	-
10	2	c.1A>G	p.Met1Val	2	c.1A>G	p.Met1Val
11	**55**	**c.8606C>A**	**p.Thr2869Lys**	-	-	-
12	-	-	-	-	-	-

Table 16: schematic representation of the retrospective validation phase. All the identified mutations with the NGS protocol and the previously diagnosed mutations with conventional sequencing are listed. For each unrelated patient the mutations, the nucleotide change in mRNA transcript (*PKHD1*: NM_138694), the aminoacidic change, the exons or IVS (intronic variant sequence number) are reported. The mutations written in bold character represents the additional variant not previously discovered with traditional method.

As reported in table 16, the NGS protocol correctly identified all the known mutations. Moreover, it allowed the individuation of additional pathogenic variants in 3 probands (1, 2, 3) consenting the complete molecular diagnosis. The c.8206T>G, p.Trp2736Gly (ExAc MAF=0.00001) and c.10444C>T, p.Arg3482Cys (ExAc MAF=0.00007) variants found respectively in probands 1 and 2 were rare variants described as being associated to the ARPKD phenotype (Bergmann et al 2004; Sharp et al 2005) and were classified as pathogenic mutations in Humgen and LOVD. The c.10444C>T, p.Arg3482Cys was also found to be an additional mutation in the healthy carrier 8 in association with the previously diagnosed variant c.8581A>G, p.Ser2861Gly, a novel variant, predicted as benign by the pathogenicity prediction tools. In this particular case, the molecular diagnosis by the NGS method was found to be more exhaustive, providing additional information on the genetic carrier status that can be advantageous for a future genetic counseling. Unfortunately we can't establish if these two variant were in cis or trans because we were not able to enroll family members.

Moreover, the c.3407A>G p.Tyr1136Cys, identified as additional variant in proband 3, that had not been found by traditional sequencing method was reported in literature as a pathogenic mutation (Loosekoot et al 2005).

In healthy carriers 9 and 11, which were considered wild type for *PKHD1* mutations by the other laboratory, the NGS protocol detected the c.8206T>G, p.Trp2736Gly and c.8606C>A, p.Thr2769Lys respectively. The first one was the same identified in proband 1, whereas the second one was a rare variant (ExAC MAF=0.009332) in a non-conservative aminoacidic change but was found in literature in several ARPKD pedigrees (Rossetti et al 2003), therefore we considered this variant to be as uncertain significance (VUS).

The healthy carrier 12 come back as a wild type for mutation in *PKHD1* in both screening tests. The investigation of large rearrangement, in this case, is recommended.

In conclusion, the NGS method correctly identified all the 12 molecular defects known to be present in the previously characterized samples, showing a sensitivity of 100%. Moreover, it allowed the identification of 4 additional missense variants previously undetected, which were confirmed by Sanger sequencing.

4.3 Pathogenic variants identified by NGS tests.

4.3.1 Molecular screening of *PKD1* and *PKD2*

125 patients, 20 of which with known mutations, were analyzed in this work. All patients were previously sequenced with NGS protocol in order to detect molecular defects in *PKD1* and *PKD2* genes; for all the wild type cases, large genetic rearrangements were investigated with MLPA.

In table 17 and in figure 16 the type and the frequency of pathogenic variants, and the novel numbers for each type, found in *PKD1* and *PKD2* are reported.

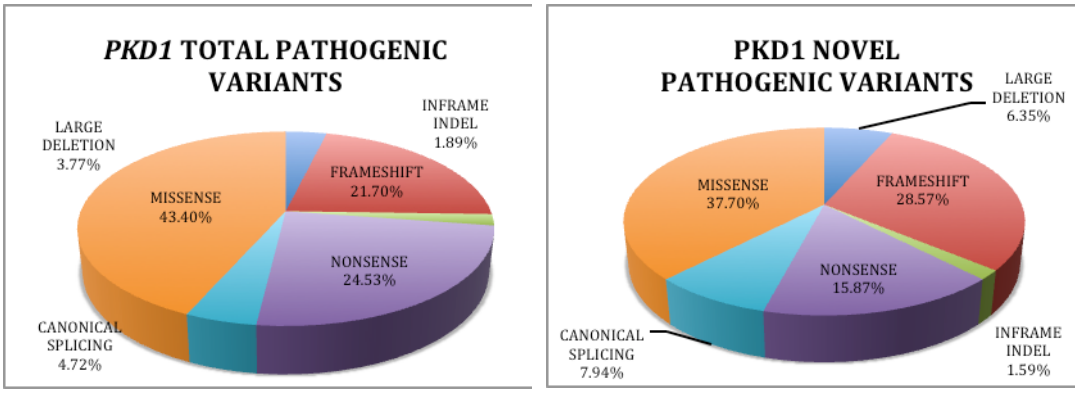
All the identified variants were confirmed by Sanger sequencing to exclude false positives.

Table 17: all identified mutations in PKD1 and PKD2 with NGS method.

TYPE	*PKD1*		*PKD2*	
	TOTAL	NOVEL	TOTAL	NOVEL
LARGE DELETION	4(3.77%)	4(100%)	1(9.09%)	1(100%)
FRAMESHIFT	23(21.70%)	18(78.26%)	2(18.18%)	2(100%)
INDEL NON FRAMESHIFT	2(1.89%)	1(50%)	0 (0%)	0(0%)
NONSENSE	26(24.53%)	10(38.46%)	2(18.18%)	0(0%)
SPLICING	5(4.72%)	5(100%)	2(18.18%)	0(0%)
MISSENSE	46(43.40%)	23(50,00%)	4(36.36%)	2(50%)
TOTALE	**106**	**61** (57.54%)	**11**	**5** (45.45%)

Table 17: all mutations found, in this study, in *PKD1* and *PKD2* sorted by type. The total variant type percentage is calculated by dividing the number of each variant type by the total variants, while the percentage for each type of novel is calculated to each variant type.

A



B

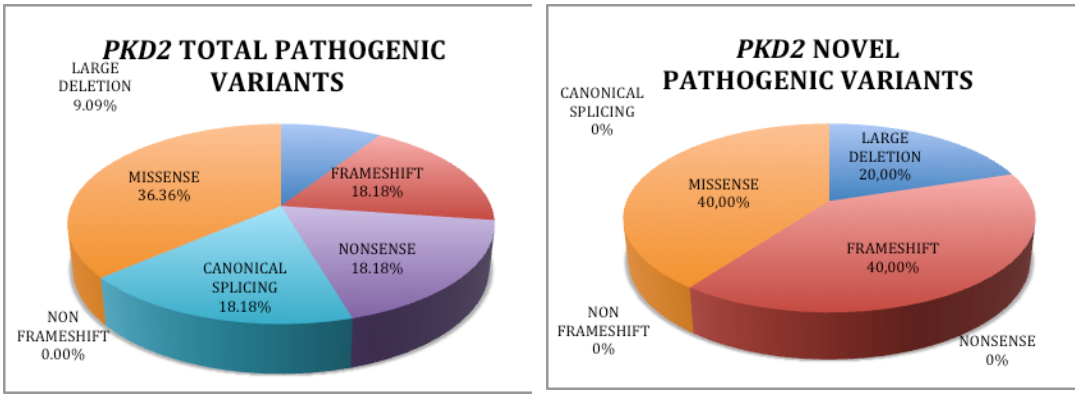


Figure 16: The pie chart **(A)** reports all the mutations found in *PKD1* on 99 ADPKD patients on the left, and the novel variants identified, on the right; **(B)** reports all the mutations found in *PKD2* on 14 ADPKD patients on the left, and the novel variants identified, on the right.

Our investigation allowed us to identify molecular defects in 90.4% (n=113) of the patients, 87.6% (n=99) presented mutations in *PKD1* while 12.4% (n=14) in *PKD2*.

In total 67 loss of function (LOF) mutations were detected, that we considered pathogenic. In particular, 28 were nonsense (26 (24.53%) in *PKD1 vs* 2 (18.18%) in *PKD2*), 25 were frameshift (23 (21.70%) in *PKD1 vs* 2 (18.18%) in *PKD2*), 7 were variants in a canonical slicing site (5(4.72%) in *PKD1 vs* 2 (18.18%) in *PKD2*) and thanks to the MLPA technique 4 large deletions (4 (3.77%) in *PKD1* vs 1 (9.09%) in *PKD2*) were found.

In the table 18-24 all the LOF variants identified are listed.

Table 18: nonsense variants identified in *PKD1* with the NGS developed method within ADPKD cohort.

c.DNA	Protein	Ex	n. of unrel. patients	PKDB/ LOVD	Reference
c.1477C>T	p.Gln493*	7	1	-/-	present study
c.2329C>T	p.Gln777*	11	1	-/-	present study
c.3250C>T	p.Gln1084*	14	1	DP/-	(A)
c.4494C>G	p.Tyr1498*	15	1	-/-	present study
c.4797C>A	p.Tyr1599*	15	1	DP/-	(Audrézet et al. 2012)
c.5482C>T	p.Gln1828*	15	2	DP/-	(Audrézet et al. 2012)
c.5511G>A	p.Trp1837*	15	1	DP/AF	(Rossetti et al. 2002)
c.5884C>T	p.Gln1962*	15	1	-/-	present study
c.6040C>T	p.Gln2014*	15	1	DP/-	(Audrézet et al. 2012)
c.6424C>T	p.Gln2142*	15	1	DP/AF	(Rossetti et al. 2012)
c.7126C>T	p.Gln2376*	17	1	DP/-	(Rossetti et al. 2002)
c.7137C>G	p.Tyr2379*	17	1	DP/-	(Audrézet et al. 2012)
c.7288C>T	p.Arg2430*	18	1	DP/-	(Rossetti et al. 2012)
c.7987C>T	p.Gln2663*	21	1	DP/-	(Audrézet et al. 2012)
c.8428G>T	p.Glu2810*	23	1	DP/-	(Rossetti et al. 2012)
c.8560C>T	p.Gln2854*	23	2	DP/-	(Audrézet et al. 2012)
c.8860G>T	p.Glu2954*	24	1	-/-	present study
c.8972_8973insA	p.Tyr2991*	25	1	DP/-	(Audrézet et al. 2012)
c.9051C>A	p.Tyr3017*	25	1	-/-	present study
c.9760C>T	p.Gln3254*	29	2	-/-	present study
c.9852C>A	p.Cys3284*	29	1	-/-	present study
c.11417G>A	p.Trp3806*	41	1	DP/-	(Stekrova et al. 2009)
c.11424G>A	p.Trp3808*	41	1	-/-	present study
c.11614G>T	p.Glu3872*	42	1	DP/-	(Tan et al. 2009)
c.12168G>A	p.Trp4056*	45	1	-/-	present study
c.12240G>A	p.Trp4080*	45	1	DP/-	(A)

Table 19: nonsense variants identified in *PKD2* with the NGS developed method within ADPKD cohort.

c. DNA	Protein	Ex	n. of unrel. patients	PKDB/ LOVD	Reference
c.916C>T	p.Arg306*	4	1	DP/-	(Audrézet et al. 2012)
c.1960C>T	p.Arg654*	9	2	DP/-	(Audrézet et al. 2012)

Table 20: frameshift variants identified in *PKD1* with the NGS developed method within ADPKD cohort.

c.DNA	Protein	ex	n. of unrel. patients	PKDB/ LOVD	Reference
c.484del	p.Ala162Leufs*128	4	1	-/-	Present Study
c.803dup	p.Phe269Leufs*102	5	1	-/-	Present Study
c.2494dup	p.Arg832Profs*40	11	1	DP/-	(Rossetti et al. 2012)
c.3210_3222del	p.Tyr1071Serfs*29	14	1	-/-	Present Study
c.3236del	p.Asp1079Alafs*25	14	1	-/-	Present Study
c.3398_3399del	p.Val1133Glufs*2	15	1	-/-	Present Study
c.3735_3741del	p.Phe1245Leufs*26	15	1	-/AF	(Carrera et al. 2016)
c.4429del	p.Leu1479Trpfs*55	15	1	-/-	Present Study
c.4631_4640del	p.Val1544Alafs*16	15	1	-/-	Present Study
c.5014_5015del	p.Arg1672fs*98	15	2	DP/AF	(Audrézet et al. 2012)
c.5425del	p.Ala1809Profs*27	15	1	-/-	Present Study
c.5536del	p.Ser1846Alafs*103	15	1	-/-	Present Study
c.6016del	p.Trp2006Glyfs*110	15	1	-/-	Present Study
c.8824del	p.Leu2942Trpfs*52	24	1	-/-	Present Study
c.10700dup	p.Val3568Cysfs*59	36	1	-/-	Present Study
c.10719_10720del	p.Gly3574Valfs*52	36	1	DP/-	(Audrézet et al. 2012)
c.10904del	p.Ala3635Valfs*2	37	1	-/-	Present Study
c.11134del	p.Arg3712Glyfs*114	38	1	DP/AF	(Turco et al. 1997)
c.11635_11645del	p.Ala3879Profs*78	42	1	-/-	Present Study
c.11819dup	p.Leu3941Alafs*20	43	1	-/-	Present Study
c.12011dup	p.Gln4005Alafs*152	44	1	-/-	Present Study
c.12232_12241del	p.Glu4078Thrfs*117	45	1	-/-	Present Study
c.12308del	p.Ala4103Valfs*95	45	1	-/-	Present Study

Table 21: indel non-frameshift variants identified in *PKD1* with the NGS developed method within ADPKD cohort.

c.DNA	Protein	Ex	n. of unrel. patients	PKDB/ LOVD	Reference
c.2059_2064del	p.Leu687_Phe688del	10	1	-/-	Present Study
c.8935_8937del	p.Phe2979del	24	1	HLP/PAF	(Rossetti et al. 2007)

Table 22: frameshift variants identified in *PKD2* with the NGS developed method within ADPKD cohort.

c.DNA	Protein	Ex	n. of unrel. patients	PKDB/ LOVD	Reference
c.540dup	p.Arg186Glyfs*32	1	1	-/-	Present Study
c.1449dup	p.Ile484Tyrfs*42	6	1	-/-	Present Study

Table 23: canonical splicing site variants identified in *PKD1* with the NGS developed method within ADPKD cohort.

c.DNA	Protein	IVS/EX	n. of unrel. patients	PKDB/ LOVD	Reference
c.9713-2A>G	p.(?)	IVS28	1	-/-	present study
c.2098-2_2109del	p.(?)	IVS10	1	-/-	present study
c.2986-2_2987del	p.(?)	IVS12/EX13	1	-/-	present study
c.11257_11269+3del	p.(?)	EX39/IVS39	1	-/-	present study
c.3161+1G>A	p.(?)	IVS13	1	-/-	present study

Table 24: canonical splicing site variants identified in *PKD2* with the NGS developed method within ADPKD cohort.

c. DNA	Protein	IVS/EX	n. of unrel. patients	PKDB/ LOVD	Reference
c.1094+1_1094+4del	p.(?)	IVS4	2	HLP/AF	(Magistroni et al. 2003)
c.1319+1G>A	p.(?)	5	1	DP/AF	(Magistroni et al. 2003)

Table 18-24: tables reported all the Loss of Function mutations found in the ADPKD entire cohort divided grouped by type and gene. For each mutation are reported the gene, the nucleotide change in mRNA transcript (*PKD1*: NM_001009944.2 and *PKD2*: NM_000297.3), the aminoacidic change, the exons or IVS (intronic variant sequence number), the number of times in which the mutations were found in unrelated patients, the classification reported in Autosomal Dominant Polycystic Kidney Mutations Database (PKDB) and in Leiden Open Variation Database (LOVD3.0) and the references. Abbreviations:DP: definitely pathogenic; AF: affected function; HLP: High likely pathogenic; PAF: Probably Affected Function.

All these definitely pathogenic variants were checked into LOVD and PKDB, two free databases, available online, containing a comprehensive coverage of published ADPKD human mutations. As the tables show, 10 (38.46%) nonsense mutations were novel in *PKD1*. In particular, the c.8860G>T, p.Glu2954*, c.9051C>A, p.Tyr3017*, c.2329C>T, p.Gln777* were also found in the proband affected-relatives. An interesting case in our familiar studies involved the c.6040C>T, p.Gln2040*, described in literature as a definitely pathogenic mutation (Audrézet et al 2012). Analyzing two healthy and four clinically affected family members, we confirmed the correct segregation of the mutation within only three of the four affected family members. The child of the proband, in fact, resulted wild type for the nonsense mutation but showed a similar clinical phenotype. Allele drop-out was excluded by using different primer pairs for LR-PCR, and a second investigation performed on the child excluded other pathogenic variants in *PKD1* and *PKD2* so further investigations have been advised.

Moreover, 20 frameshift novel mutations (18 (78.26%) in *PKD1* vs 2 (100%) in *PKD2*) were identified, the c.3210_3222del, p.Tyr1071Serfs*29, c.803dup, p.Phe269Leufs*102, c.11819dup, p.Leu3941Alafs*20, were recurrent in the proband affected-relatives and the c.3236del, p.Asp1079Alafs*25 resulted as a *de novo* in the proband; only 1 (50%) *PKD1* novel indel non frameshift mutations c.2059_2064del, p.Leu687_Phe688del was identified in one of the patients and in his affected brother.

Finally, 5 (100%) novel variants in canonical splicing site were detected in *PKD1* and, among these, the presence of the c.9713-2A>G and the c.11257_11269+3del were confirmed in affected relatives.

In order to classify all the missense variants, we combined results from known pathogenic variants and the current classifications (see section 3.3.6.1).

Table 25: missense variants identified in *PKD1* with the NGS developed method within ADPKD cohort.

c.DNA	Protein	ex	n. of unrel. patients	PKDB/ LOVD	SIFT/ PP2	ExAC	Reference
c.74G>T	p.Gly25Val	1	1	IND/-	T/B	-	(A)
c.194T>A	p.Ile65Asn	1	1	-/UNK	D/B	-	(Carrera et al. 2016)
c.230A>G	p.Asn77Ser	2	1	HLP/-	D/B	-	(A)
c.301A>G	p.Asn101Asp	3	1	HLP/-	D/PD	-	(A)
c.971G>T	p.Arg324Leu	5	1	LN/PNAF	T/PD	0.006	(Rossetti et al. 2007)
c.974A>G	p.Tyr325Cys	5	1	HLP/PAF	D/PD	-	(Audrézet et al. 2012)
c.1261C>T	p.Arg421Cys	6	1	-/AF	D/PD	-	(Carrera et al. 2016)
c.1295C>T	p.Ala432Val	6	1	HLP/-	D/PD	-	(Rossetti et al. 2007)
c.1777G>A	p.Glu593Lys	9	1	-/-	D/PD	-	present study
c.2438G>A	p.Cys813Tyr	11	1	-/-	D/PD	-	present study
c.2695C>G	p.Leu899Val	11	1	-/UNK	D/PD	0.00006	(Carrera et al. 2016)
c.2830C>T	p. Arg944Cys	11	1	LP/-	D/PD	-	(A)
c.3490G>A	p.Gly1164Arg	15	1	-/-	D/PD	-	present study
c.4628C>T	p.Thr1543Met	15	1	-/-	T/PD	0.00005	present study
c.4856C>T	p.Ser1619Phe	15	1	LN/UNK	D/PD	0.0009	(Irazabal et al. 2011)
c.5102A>T	p.Asn1701Ile	15	1	-/-	D/PD	-	present study
c.5830G>A	p.Gly1944Arg	15	1	IND/-	D/PD	0.0002	(A)
c.5999C>G	p.Ser2000Cys	15	1	-/-	D/PD	-	present study
c.6070C>T	p.Arg2024Cys	15	1	-/-	D/PD	0.0003	present study
c.6799G>C	p.Val2267Leu	15	1	-/-	T/PD	0.00002	present study
c.6908C>T	p.Ser2303Leu	15	1	-/-	D/PD	0.00004	present study
c.7166T>C	p.Leu2389Ser	17	1	-/-	D/PD	-	present study
c.7535T>C	p.Leu2512Pro	19	1	-/-	D/PD	-	present study
c.7546C>T	p.Arg2516Cys	19	1	HLP/-	D/PD	-	(Cornec-Le Gall et al. 2013)
c.7622C>T	p.Pro2541Leu	19	1	-/PAF	D/PD	0.0006	(Carrera et al. 2016)
c.7663G>A	p.Val2555Met	19	1	-/-	D/PD	-	present study
c.7979A>G	p.Asp2660Gly	21	1	-/-	D/PD	-	present study
c.8000C>A	p.Ala2667Glu	21	1	-/-	D/PD	-	present study
c.8179G>C	p.Ala2727Pro	23	1	-/-	D/B	-	present study
c.8270C>G	p.Ser2757Cys	23	1	-/-	D/PD	-	present study
c.8293C>T	p.Arg2765Cys	23	2	Lhypo/PAF	D/PD	0.004	(Rossetti et al. 2007)
c.8311G>A	p.Glu2771Lys	23	4	HLP/-	D/PD	-	(Audrézet et al. 2012)
c.8696T>C	p.Val2899Ala	23	1	-/-	D/PD	-	present study
c.9404 C>T	p.Thr3135Met	27	1	LP/PAF	D/PD	-	(Bataille et al. 2011)
c.9583T>G	p.Trp3195Gly	28	1	-/-	D/PD	-	present study
c. 9739C>T	p.Arg3247Cys	29	1	-/-	D/PD	0.00003	present study
c.9829C>T	p.Arg3277Cys	29	1	Lhypo/PAF	D/PD	0.0002	(Vujic et al. 2010)
c.10043G>A	p.Arg3348Gln	30	1	LP/PAF	D/PD	-	(Cornec-Le Gall et al. 2013)
c.11015G>A	p.Arg3672Gln	37	1	LN/UNK	T/PD	0.0009	(A)
c.11258G>A	p.Arg3753Gln	39	1	HLP/-	D/PD	-	(Rossetti et al. 2012)
c.11390A>G	p.Tyr3797Cys	40	1	-/PAF	D/PD	-	(Carrera et al. 2016)
c.11440T>C	p.Tyr3814His	41	1	-/-	D/PD	-	present study
c.12032A>T	p.Gln4011Leu	44	1	-/-	D/PD	-	present study
c.12051G>C	p.Lys4017Asn	44	1	-/-	D/PD	0.0008	present study
c.12444G>C	p.Glu4148Asp	45	1	-/-	D/PD	-	present study
c.12460C>T	p.Arg4154Cys	46	1	LP/PAF	D/PD	0.009	(Perrichot et al. 1999)

Table 26: missense variants identified in *PKD2* with the NGS developed method within ADPKD cohort.

c.DNA	Protein	Ex	n. of unrel. patients	PKDB/ LOVD	SIFT/ PP2	EXAC	REFERENZE
c.974G>A	p.Arg325Gln	4	1	HLP/AF	D/PD	/	(Bataille et al. 2011)
c.1445T>G	p.Phe482Cys	6	2	LN/AF	D/PD	0.002	(Dedoussis et al. 2008)
c.2186T>A	p.Leu729Gln	11	1	-/-	D/PD	0.00002	present study
c.2865G>T	p.Met955Ile	15	1	-/-	T/B	/	present study

Table 25-26: all the missense variants found in the ADPKD entire cohort are reported. For each mutation the gene, the nucleotide change in mRNA transcript (*PKD1*: NM_001009944.2 and *PKD2*: NM_000297.3), the aminoacidic change, the exons or IVS (intronic variant sequence number), the number of times in which the mutations were found in unrelated patients, the classification reported in Autosomal Dominant Polycystic Kidney Mutations Database (PKDB) and in Leiden Open Variation Database (LOVD3.0), the pathogenicity prediction carried out using Polyphen 2.0 (PP2) and SIFT and the minor frequency allele annotated in ExAC are reported. The last column reports the references. Abbreviations: AF: affected function; HLP: High likely pathogenic; PAF: Probably Affected Function; LP: likely pathogenic; PNAF: potentially not affect function; Lhypo: Likely Hypomorphic; IND: indeterminate significance; UNK: unknown significance; LN: likely neutral; PD: Possibly Damaging; D: Damaging: T: tolerated; B: benign.

In our cohort, 25 *novel* missense mutations (23(50%) in *PKD1* vs 2 (50%) in *PKD2*) all predicted as likely pathogenic by bioinformatic tools were also detected (Table 25-26). Among these, the c.8000C>A, p.Ala2667Glu and c.5102A>T, p.Asn1701Ile were also confirmed in all affected relatives and we classified them as likely pathogenic variants.

The c.8293C>T, p.Arg2765Cys, c.8311G>A, p.Glu2771Lys in *PKD1* and the c.1445T>G, p.Phe482Cys in *PKD2* were found in more than one unrelated patients. All these three variants have already been reported in literature as associated with ADPKD. In particular, the *PKD1* variants were all classified as potentially affected functions and the *PKD2* variant resulted as hypomorfic and functional studies revealed a lower PC-2 channel function compared with controls in homozygous cells and heterozygous as well (Dedoussis et al 2008).

4.3.1.1 Cases with more than one variant in PKD1

Although ADPKD is classically inherited as an autosomal dominant disease resulting from heterozygous mutations in either *PKD1* or *PKD2,* recent and major advances in molecular genetic basis underline a complex inheritance pattern of ADPKD, leading to the discovery of hypomorphic or incompletely penetrant alleles. In our study we found 13 unrelated patients with more than one variant in *PKD1* (Table 27). Usually, in cases with more than one variant, familiar study could help to understand the pathogenic role. In our study, unfortunately, the unavailability of sufficient affected family members for some cases did not allow for the in-depth study of the variant role in ADPKD.

.Table 27: Cases with more than 1 variant in *PKD1*.

n°	Gene	ET/ OMO	c.DNA	Protein	EX/ IVS	M/ P	SIFT/ PP2	PKDB/ LOVD	Reference
1	*PKD1*	ET	c.5830G>A	p.Gly1944Arg	15	-	D/PD	IND/-	(A)
		ET	c.11390A>G	p.Tyr3797Cys	40	-	D/PD	-/PAF	(Carrera et al. 2016)
2	*PKD1*	ET	c.6799G>C	p.Val2267Leu	15	-	T/PD	-/-	present study
		ET	c.8311G>A	p.Glu2771Lys	23	-	D/PD	HLP/-	(Audrézet et al. 2012)
3	*PKD1*	ET	c.2098-2_2109del	p.(?)	IVS10	-	-/-	-/-	present study
		ET	c. 9739C>T	p.Arg3247Cys	29	-	D/PD	-/-	present study
4	*PKD1*	ET	c.8270C>G	p.Ser2757Cys	23	-	D/PD	-/-	present study
		ET	c.11614G>T	p.Glu3872*	42	-	-/-	DP/-	(Tan et al. 2009)
5	*PKD1*	ET	c.3490G>A	p.Gly1164Arg	15	M	D/PD	-/-	present study
		ET	c.8293C>T	p.Arg2765Cys	23	P	D/PD	LHypo/PAF	(Rossetti et al. 2012)
		ET	c.10043G>A	p.Arg3348Gln	30	P	D/PD	LP/PAF	(Cornec et al. 2013)
6	*PKD1*	ET	c.3735_3741del	p.Phe1245Leufs*26	15	-	-/-	-/AF	(Carrera et al. 2016)
		ET	c.6070C>T	p.Arg2024Cys	15	-	D/PD	-/-	present study
7	*PKD1*	ET	c.11440T>C	p.Tyr3814His	41	M	D/PD	-/-	present study
		ET	c.12032A>T	p.Gln4011Leu	44	M	D/PD	-/-	present study
8	*PKD1*	ET	c.1477C>T	p.Gln493*	7	-	-/-	-/-	present study
		ET	c.2830C>T	p. Arg944Cys	11	-	D/PD	LP/-	(A)
9	*PKD1*	ET	c.1777G>A	p.Glu593Lys	9	-	D/PD	-/-	present study
		ET	c.6908C>T	p.Ser2303Leu	15	-	D/PD	-/-	present study
10	*PKD1*	ET	c.4797C>A	p.Tyr1599*	15	M	-/-	DP/-	(Audrézet et al. 2012)
		ET	c.5999C>G	p.Ser2000Cys	15	P	D/PD	-/-	present study
11	*PKD1*	ET	c.2695C>G	p.Leu899Val	11	n	D/PD	-/UNK	(Carrera et al. 2016)
		ET	c.11134del	p.Arg3712Glyfs*114	38	P	-/-	DP/AF	(Turco et al. 1997)
12	*PKD1*	ET	c.230A>G	p.Asn77Ser	2	-	D/B	HLP/-	(A)
		ET	c.12460C>T	p.Arg4154Cys	46	-	D/PD	LP/PAF	(Perrichot et al. 1999)
13	*PKD1*	OMO	c.9829C>T	p.Arg3277Cys	29	T	D/PD	LHypo/PAF	(Vujic et al. 2010)

Table 27: table reported ADPKD unrelated cases with more than 1 variant in *PKD1*. For each mutation the inheritance (M=maternal; P=Paternal; n=de novo), the nucleotide change in mRNA transcript (*PKD1*: NM_001009944.2 and *PKD2*: NM_000297.3), the aminoacidic change, the exons or IVS (intronic variant sequence number), the classification reported in Autosomal Dominant Polycystic Kidney Mutations Database (PKDB) and in Leiden Open Variation Database (LOVD3.0), the pathogenicity prediction carried out using Polyphen 2.0 (PP2) and SIFT are reported. The last column reports the references. Abbreviations: HLP: High likely pathogenic; PAF: Probably Affected Function; PNAF: potentially not affect function; Lhypo: Likely Hypomorphic; IND: indeterminate significance; PD: Possibly Damaging; D: Damaging: T: tolerated; B: benign.

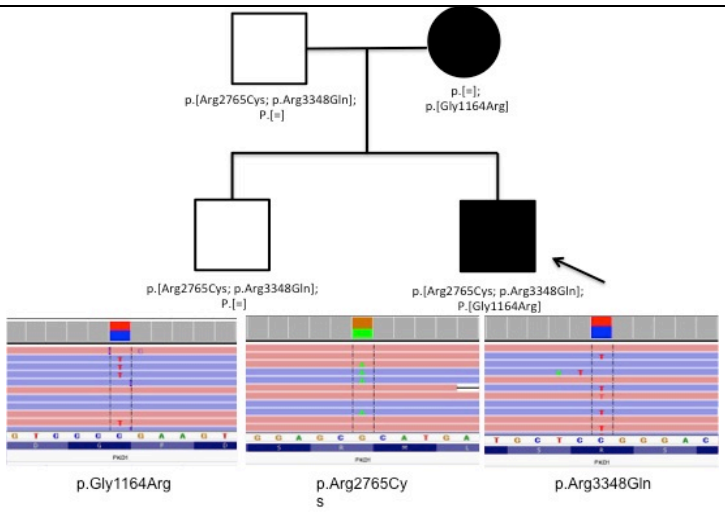
Thanks to the collaboration of family members we were able to identify pathogenic variants in families 5, 7,10 and 11.

The NGS protocol detected in proband 5 (Figure 17) two described causative variants the c.8293C>T, p.Arg2765Cys and c.10043G>A, p.Arg3348Gln and the novel c.3490G>A, p.Gly1164Arg (Table 25). The two causative variants were found in cis in his healthy brother and father, while the novel variant was inherited by his affected mother. This family study allowed us to establish a causative role of the variants found in the present study and to conclude that the compound heterozygous with the other variants can result in a more severe phenotype. The two variants in cis were not sufficient for the phenotype manifestation but can act as disease-modifying. However, the lack of accurate renal ultrasound reports of "healthy" relatives did not allow the exclusion of a mild phenotype.

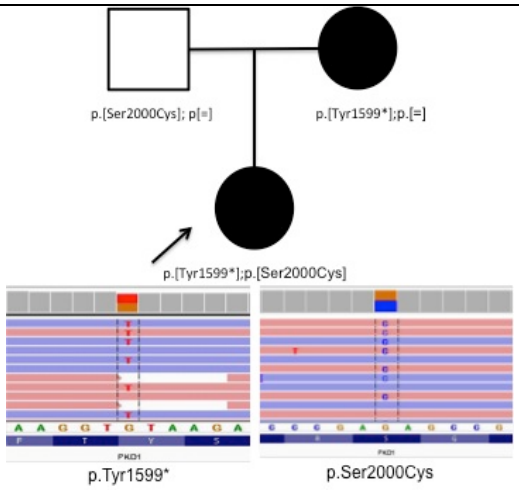
Another disease-modifying variant was the c.5999C>G, p.Ser2000Cys. In family 10 (Figure 17), the proband showed the c.5999C>G, p.Ser2000Cys in compound heterozygous with a truncated mutation inherited by the affected mother. Although the missense variant alone was not sufficient for the phenotype manifestation in the father, the coinheritance occurred as a more severe disease.

In family 11 (Figure 17) the proband inherited the frameshift mutation c.11134del, p.Arg3712Glyfs*114 from his affected father but showed an ARPKD phenotype. The molecular analysis of *PKHD1* excluded molecular defects in this gene but we found an addictional *de novo* causative mutations in *PKD1*, the c.2695C>G, p.Leu899Val. This second variant, in trans with the inactivated-*PKD1* allele caused in the proband an in-utero-onset of ADPKD. The proband in family 7 presented two different novel mutations, both of them predicted as potential pathogenic by bioinformatics tools. These two variants, the c.11440T>C, p.Tyr3814Hys and the c.12032A>T, p.Gln4011Leu resulted inherited both from her affected mother, so the cis configuration did not help us define which of them was the causative one. Finally, an interesting case was represented by the patients 13 that showed an ARPKD phenotype but resulted wild type for the molecular screening of *PKHD1*. This patient presented the homozygosity of an hypomorphic allele c.9829C>T, p.Arg3277Cys in *PKD1*. Rossetti et al. reported that the heterozygous mutation alone may result in mild cystic disease, while the homozygosity can cause an ARPKD-like phenotype (Rossetti et al 2009).

Family 5



Family 10



Family 11

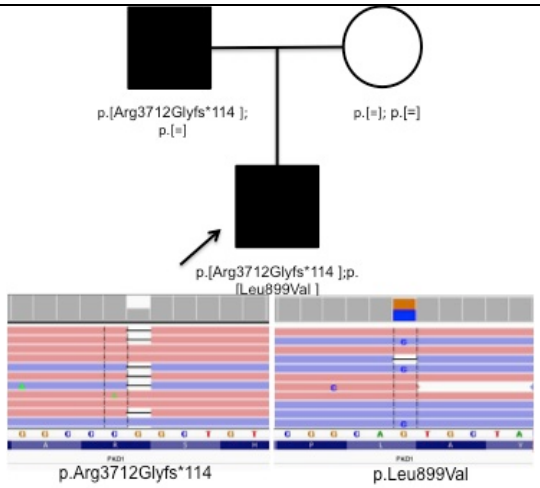


Figure 17: the pedigree of families 5, 10 and 11 are represented. Below each family tree the mutations visualized by IGV are reported. Probands are indicated by the arrow.

4.3.2 Molecular screening of *PKHD1*

The mutational screening of *PKHD1* gene was performed on a cohort of 28 unrelated patients divided in two groups (A and B): Group A consisted of 15 unrelated probands with clinical suspect of ARPKD, including 2 fetuses from terminated pregnancy and 2 severely affected newborns who died shortly after birth; Group B counted 13 unrelated healthy carriers, as parents of affected ARPKD fetuses. All the fetuses from terminated pregnancies presented oligo/anhydramnios and renal pathology including dilated tubules and/or collecting ducts, changed kidney size and bilaterally enlarged hyperechogenic kidneys.

In table 28 and in the pie charts in figure 18 are reported the total and the novel variants found in this study divided for type. The frequency of the total variant is calculated on the total number and the frequency of the novel is calculated as the ratio between the novel and the total for each type.

Table 28: all identified mutations in *PKHD1* with the NGS method.

TYPE	*PKHD1*	
	TOTAL	NOVEL
NONSENSE	7(18.91%)	4(57.14%)
FRAMESHIFT	4(10.81%)	2(50%)
INDEL NON FRAMESHIFT	1(2.70%)	1(100%)
CANONICAL SPLICING	1(2.70%)	1(100%)
MISSENSE	24(64.86%)	10(41.66%)
TOTAL	37	18(48.64%)

Table 28: all mutations found, in this study, in *PKHD1* sorted by type. The total variant type percentage is calculated by dividing the number of each variant type by the total variants, while the percentage for each type of novel is calculated to each variant type.

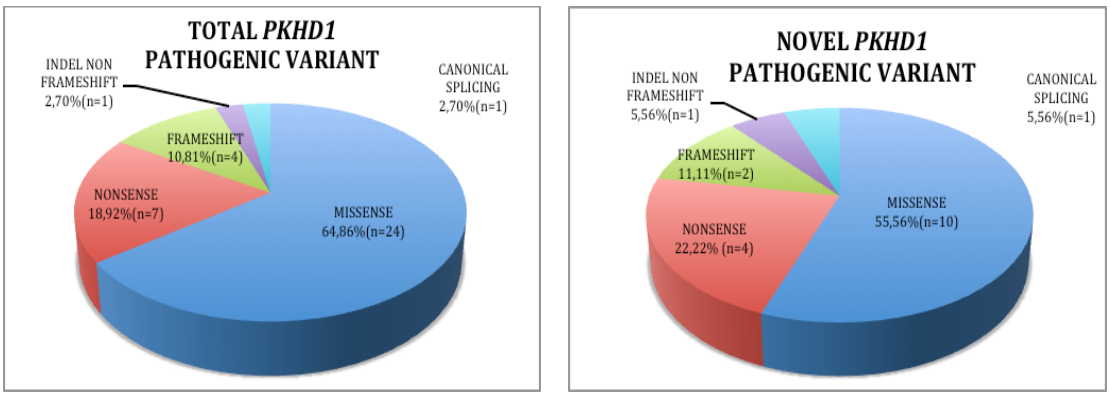


Figure 18: The pie charts reports all the mutations found in *PKHD1* in ARPKD cohort on the left, and the novel variants identified, on the right.

In total, 37 different variants, almost evenly distributed throughout the entire gene transcript, were detected. 13 were loss of function (18.91% (n=7) nonsense, 10,81% (n=4) frameshift and 2.70% (n=1) indel non frameshift and 2.70% (n=1) variant in canonical splicing site) and 64.86% (n=24) were missense mutations.

All the variants were checked against the mutation databases for ARPKD and LOVD 3.0 and for each variant the minor frequency allele was obtained by ExAC browser. The tables below reports the mutation divided by type in detail. All the missense were submitted to predictional pathogenicity tools SIFT and POLYPHEN 2.0 in order to achieve a preliminary pathogenic classifications.

Table 29: nonsense variants identified in *PKHD1* with the NGS developed method within ARPKD cohort.

c.DNA	Protein	ex	n° of unrel patients	Humgen/ LOVD	EXAC	Reference
c.370C>T	p.Arg124*	5	1	P/AF	0.00002	(Bergmann et al. 2005)
c.4417C>T	p.Gln1473*	32	1	-/-	/	present study
c.6472G>T	p.Glu2158*	39	1	-/-	/	present study
c.7328T>G	p.Leu2443*	46	1	-/-	/	present study
c.8470C>T	p.Gln2824*	54	1	-/-	/	present study
c.10174C>T	p.Gln3392*	61	1	P/AF	0.000009	(Loosekoot et al. 2005)
c.11314C>T	p.Arg3772*	63	1	P/AF	/	(Bergmann et al. 2004)

Table 30: frameshift variants identified in *PKHD1* with the NGS developed method within ARPKD cohort.

c.DNA	Protein	ex	n° of unrel patients	Humgen/ LOVD	EXAC	Reference
c.1197delC	p.Leu400Cysfs*13	15	1	-/-	0.000008	present study
c.5895dup	p.Leu1966Thrfs*4	36	2	P/AF	/	(Gunay-Aygun et al. 2010)
c.6390dup	p.Thr2131Hisfs*17	39	1	-/-	/	present study
c.10856del	p.Lys3619Serfs*7	61	1	P/AF	/	(Rossetti et al. 2003)

Table 31: indel non-frameshift variants identified in *PKHD1* with the NGS developed method within ARPKD cohort.

c.DNA	Protein	ex	n° of unrel patients	Humgen/LOVD	EXAC	Reference
c.9708_9737del	p.Arg3237_Pro3246del	58	1	-/-	/	present study

Table 32: canonical splicing site variants identified in *PKHD1* with the NGS developed method within ARPKD cohort.

cDNA	Protein	IVS	n° of unrel patients	Humgen/LOVD	EXAC	Reference
c.8555-2A>C	p.(?)	54	1	-/-	-	present study

Table 33: Missense variants identified in *PKHD1* with the NGS developed method within ARPKD cohort.

c.DNA	Protein	ex	n° of unrel pat	Humgen /LOVD	SIFT/ PP2	EXAC	Reference
c.1A>G	p.Met1Val	2	2	-/-	-/-	-	present study
c.107C>T	p.Thr36Met	3	1	PD/-	D/PD	0.0005	(Loosekoot et al. 2005)
c.664A>G	p.Ile222Val	9	1	PD/NAF	T/B	0.0001	(Loosekoot et al. 2005)
c.1736C>T	p.Thr579Met	19	2	POL/-	D/P	0.02	(Loosekoot et al. 2005)
c.2171C>G	p.Pro724Arg	22	1	PD/NAF	D/P	-	(Gunay-Aygun et al. 2010)
c.2702A>C	p.Asn901Thr	25	1	PD/-	D/PD	0.000008	unpublish data
c.2929A>T	p.Asn977Tyr	27	1	-/-	T/D	0.000008	present study
c.3407A>G	p.Tyr1136Cys	30	1	PD/UNK	T/PD	0.008	(Loosekoot et al. 2005)
c.3686G>C	p.Trp1229Ser	32	1	nd/-	T/B	0.003	(Suzuki et al. 2006)
c.4292G>A	p.Cys1431Tyr	32	1	-/-	D/PD	0.00003	present study
c.4519C>G	p.Gln1507Glu	32	1	-/-	T/B	0.00003	present study
c.4870C>T	p.Arg1624Trp	32	1	PD/AF	T/D	0.0002	(Onuchic et al. 2002)
c.5068T>C	p.Ser1690Pro	32	1	-/-	T/D	-	present study
c.5336A>G	p.Asn1779Ser	33	1	-/-	T/PD	-	present study
c.5830G>A	p.Asp1944Asn	36	1	-/-	D/P	0.00002	present study
c.8206T>G	p.Trp2736Gly	52	2	nd/AF	D/B	0.00002	(Denamur et al. 2010)
c.8518C>G	p.Arg2840Gly	54	1	-/-	T/B	0.00003	present study
c.8521A>G	p.Met2841Val	54	1	PD/NAF	T/B	0.003	(Loosekoot et al. 2005)
c.8581A>G	p.Ser2861Gly	55	1	PD/NAF	T/B	0.003	(Gunay-Aygun et al. 2010)
c.8606C>A	p.Thr2869Lys	55	1	POL/AF	D/PD	0.009	(Rossetti et al. 2003)
c.9163G>T	p.Gly3055Cys	58	1	-/-	D/D	-	present study
c.9189C>A	p.Asn3063Lys	58	1	-/-	D/D	-	present study
c.9370C>T	p.His3124Tyr	58	1	PD/AF	D/D	-	(Furu et al. 2003)
c.10444C>T	p.Arg3482Cys	61	2	PD/AF	D/D	0.00006	(Bergmann et al. 2005)

Table 29-33: reported all the missense variants found in the ADPKD entire cohort. For each mutation the gene, the nucleotide change in mRNA transcript (*PKHD1*: NM_138694), the aminoacidic change, the exons or IVS (intronic variant sequence number), the number of times in which the mutations were found in unrelated patients, the classification reported in Autosomal Recessive Polycystic kidney disease database (Humegene) and in Leiden Open Variation Database (LOVD3.0), the pathogenicity prediction carried out using Polyphen 2.0 (PP2) and SIFT and the minor frequency allele reports in ExAC are reported. The last column reports the references. Abbreviations: AF: affected function; NAF: Not Affected Function; nd: classification not available; UNK: unknown significance; POL: Polymorphism; PD: Possibly Damaging; D: Damaging: T: tolerated; B: benign.

Overall, 18 (48.64%) novel variants were identified in this study, 44.4% (n=8) were classified as pathogenic because predicted to caused a non-functional proteic product (4 nonsense, 2 frameshift, 1 indel non frameshift and 1 canonical splice site variant).

The majority of mutant alleles were unique for each family, however 5 variants, the frameshift mutation c.5895dup, p.Leu1966Thrfs*4, and the 4 missense variants c.1A>T p.Met1Val; c.1736C>T, p.Thr579Met; c.8206T>G, p.Trp2736Gly c.10444C>T, p.Arg3482Cys appeared at least 2 families.

In Group A (Table 34), 14 of the 15 probands presented at least two pathogenic or probably pathogenic mutations in compound heterozygous status, one of these

patients (#5) presented 3 variants, however we could not establish if they were in cis or in trans configuration.

40% (n=6) of patients with two causal mutations harbored a combination of a missense and chain-terminating mutation while the combination of two missense or two truncating mutations was accounted for the 33.33% (n=5) and the 6,67% (n=1), respectively.

Group B (Table 35) counted 12 potentially healthy carriers in 6 non-consanguineous marriage, with at least one child affected by ARPKD or with termination of pregnancies for oligo/anhydramnios and fetus renal pathologies, and one with a previously molecular diagnosis of healthy carriers. In all but 3 cases, at least one causative variant was identified. 3 cases reported a truncating or frameshift mutations, 2 showed a nucleotide change in the initiation codon and, finally, 5 patients presented at least one missense variant. All the missense variants were described in either one of the mutational databases mentioned above.

In the mother of family 16, 2 missense variants were identified, the c.10444C>T p.Arg3482Cys was predicted as damaging, whereas the c.8581A>G p.Ser2861Gly was predicted as tolerated by prediction tools. Both parents of this family were healthy obligate carriers of an ARPKD defect because their fetuses in the two terminated pregnancies presented ARPKD features. The presence of c.10444C>T, p.Arg3482Cys and the paternal c.1A>G, p.Val1Met were confirmed in both fetal DNAs extracted from conserved tissues sample.

The overall detection rate amounts to 93.33% in Group A and to 77% in Group B and the overall detection rate for the entire patient cohort is 85.7%.

For all cases considered wild type for molecular defects in *PKHD1*, the investigation of large genomic rearrangement is recommended.

Table 34: Summary of mutations and clinical phenotype found in Group A.

# unrel patient	Ex/ IVS	c.DNA	Protein	M/P	age of diagnosis/ Death	Info	Reference
1	32	c.4870C>T	p.Arg1624Trp	M	perinatal	renal echog↑/ brohter ARPKD (†21dd)/mother healty ARPKD carrier	(Onuchic et al. 2002)
	61	c.10444C>T	p.Arg3482Cys	P			(Bergmann et al. 2005)
2	54	c.8518C>G	p.Arg2840Gly	P	infantil (4y) /† (34y)	renal and hepatic echog↑/HSM/HTN/no familiarity for ARPKD	present study
	61	c.10856del	p.Lys3619Serfs*7	M			(Rossetti et al. 2003)
3	30	c.3407A>G	p.Tyr1136Cys	-	Childhood (15y)	bilatera renal cysts/poor cortico-medullary differentiation	(Loosekoot et al. 2005)
	32	c.4292G>A	p.Cys1431Tyr	-			present study
	36	c.5895dup	p.Leu1966Thrfs*4	-			(Gunay-Aygun et al. 2010)
4	33	c.5336A>G	p.Asn1779Ser	P	perinatal/ †(2dd)	OH/renal and hepatic echog↑	present study
	54	c.8470C>T	p.Gln2824*	M			present study
5	25	c.2702A>C	p.Asn901Thr	P	Childhood (11y)	bilatera renal cysts/poor cortico-medullary differentiation/dialysis/brother with ARPKD	unpublish data
	58	c.9189C>A	p.Asn3063Lys	M			present study
6	36	c.5830G>A	p.Asp1944Asn	M	infantil (2y)	renal and liver echog↑ bilatera renal cysts/poor cortico-medullary differentiationdialysis/brother with ARPKD	present study
	52	c.8206T>G	p.Trp2736Gly	P			(Denamur et al. 2010)
7	32	c.4417C>T	p.Gln1473*	P	perinatal (4m)	bilatera renal cysts/poor cortico-medullary differentiation/renal and hepatic echog↑	present study
	32	c.5068T>C	p.Ser1690Pro	M			present study
8	27	c.2929A>T	p.Asn977Tyr	P	perinatal (7m)	bilatera renal cysts/poor cortico-medullary differentiation/renal and hepatic echog↑/KTX	present study
	58	c.9370C>T	p.His3124Tyr	M			(Furu et al. 2003)
9	30	c.3385G>A	p.Gly1129Arg	M	infantil (1y)	bilatera renal cysts/poor cortico-medullary differentiation/renal and hepatic echog↑	present study
	58	c.9708_9737del	p.Arg3237_Pro3246del	P			present study
10	IVS 54	c.8555-2A>C	p.(?)	M	CVS/TOP	OH	present study
	58	c.9163G>T	p.Gly3055Cys	P			present study
11	22	c.2171C>G	p.Pro724Arg	M	infantil (6y)	renal and hepatic cysts/brother with ARPKD/healty consanguineous parents	(Gunay-Aygun et al. 2010)
	39	c.6472G>T	p.Glu2158*	P			present study
12	32	c.3686G>C	p.Trp1229Ser	M	perinatal/ † 28gw	OH/HSM/renal and liver echog↑	(Suzuki et al. 2006)
	54	c.8521A>G	p.Met2841Val	P			(Sharp et al 2005)
13	39	c.6390dup	p.Thr2131Hisfs*17	M	CVS/TOP	OH	present study
	46	c.7328T>G	p.Leu2443*	P			present study
14	9	c.664A>G	p.Ile222Val	M	perinatal (7m)	bilatera renal cysts/poor cortico-medullary differentiation/renal and hepatic echog↑	(Ward et al. 2002)
	36	c.5895dup	p.Leu1966Thrfs*4	P			(Ward et al. 2002)
15	5	c.370C>T	p.Arg124*	-	Adolescent (11y)	bilateral renal cortico-medullary cysts	(Bergmann et al. 2005)

Table 34: all mutations found in unrelated patients of ARPKD cohort (GroupA). For each patients the nucleotide change mutations in mRNA transcript (*PKHD1*: NM_138694), the aminoacidic change, the exons or IVS (intronic variant sequence number), the segregational study results, the patient phenotype and clinical informations are reported. The last column reports the references. Abbreviations: M:maternal; P:paternal; † deceased; y:years; dd: days; m:month; gw:gestational week; CVS/TOP: prenatal diagnosis after chorionic villus sampling (CVS) based on haplotyping with subsequent termination of pregnancy (TOP), echog↑: increased echogenicity, HSM:hepatosplenomegaly; HTN: ipertension; KTX:kidney transplantions; OH:oligohydramnios;

Table 35: Summary of mutations and clinical phenotype found in Group B.

#family/ M/F	info	c.DNA	Protein	ex	HUMGEN/ LOVD	SIFT/ PP2	Reference
16.F	2 TOP	c.107C>T	p.Thr36Met	3	PD/-	D/PD	(Bergmann et al. 2005)
16.M		c.8581A>G	p.Ser2861Gly	55	PD/NAF	T/B	(Furu et al. 2003)
		c.10444C>T	p.Arg3482Cys	61	PD/AF	D/PD	(Bergmann et al. 2005)
17.M	2 TOP/ 1child ARPKD (†6dd)	c.8206T>G	p.Trp2736Gly	52	nd/AF	D/B	(Denamur et al. 2010)
17.F		c.1A>G	p.Met1Val	2	-	D/PD	present study
18.M	1 TOP/OH	c.11314C>T	p.Arg3772*	63	PD/AF	-/-	(Bergmann et al. 2004)
18.F		c.10174C>T	p.Gln3392*	61	PD/AF	-/-	(Loosekoot et al. 2005)
19.M	healthy carrier	c.1197delC	p.Leu400Cysfs*13	15	-	-/-	present study
20.F	1 child with ARPDK	c.1736C>T	p.Thr579Met	19	POL/-	D/PD	(Loosekoot et al. 2005)
20.M		-	-	-	-	-	-
21.M	1 TOP/ OH	c.8606C>A	p.Thr2869Lys	55	POL/-	D/PD	(Bergmann et al. 2005)
21.F		-	-	-	-	-	-
22.M	1 TOP/ OH	c.1A>G	p.Met1Val	2	-	D/PD	present study
22.F		-	-	-	-	-	-

Table 35: all mutations found in healthy carriers of ARPKD cohort (GroupB). For each subject the number of family, the clinical family informations, the nucleotide change mutations in mRNA transcript (*PKHD1*: NM_138694), the aminoacidic change, the exons or IVS (intronic variant sequence number), the classification reported in Autosomal Recessive Polycystic kidney disease database (Humegene) and in Leiden Open Variation Database (LOVD3.0), the pathogenicity prediction carried out using Polyphen 2.0 (PP2) and SIFT are reported. The last column reports the references. Abbreviations: M:mother; F:father ; TOP: prenatal diagnosis termination of pregnancy; OH:oligohydramnios; AF: affected function; NAF: Not Affected Function; nd: classification not available; POL: Polymorphism; PD: Possibly Damaging; D: Damaging; T: tolerated; B: benign.

4.4 Time and cost analysis

The NGS-based developed protocols turned out to be both faster and cost-saving compared to the traditional Sanger sequencing while allowing us to obtain rapid medical reports for PKD patients in just a few weeks.

In this section we will compare the conventional Sanger sequencing and the NGS method analyzing reagent costs and their respective labor time to reach a complete molecular diagnosis of ADPKD or ARPKD patients (Table 36).

According to our experience, by mixing barcoded-libraries obtained from the two differential developed protocols, we were able to obtain the simultaneous screening of *PKD1* and *PKD2* analyzing 16 patients *per* run. Using a 318v2 chip we reached an average coverage of ≥ 300X for all the target regions in all the 16 submitted patients, which was sufficient to ensure a complete genotyping of the two ADPKD causative

genes. As the *PKHD1* transcript gene is larger compared to *PKD1* and *PKD2* transcripts we decided to mixing barcoded-libraries of maximum 12 patients in order to reach an average coverage ≥ 300X for all the targets, as well.

As table 36 shows, up to 16 ADPKD patients can be genotyped with our NGS protocols at reagent costs of € 250 *per* patient compared with €1250 *per* patient using the current Sanger sequencing method. The same cost reduction is applicable for *PKHD1* NGS protocol.

Although the hands-on time required for setting up the LR-PCR reactions (in the case of *PKD1*), preparing the library, and sequencing still requires approximately 1 week, data analysis is considerably faster with NGS than with the Sanger method and can be completed in only a few hours without the need for expensive computer clusters and making it attractive to the standard diagnostic laboratory.

Table 36: comparison of sequencing reagents costs and time of labor for Sanger sequencing and NGS

	ADPKD (*PKD1/PKD2*)		**ARPKD (*PKHD1*)**	
	Sanger Sequencing	NGS	Sanger Sequencing	NGS
Patients per run	1	16	1	12
Costs (per patients)	1.250 €	250 €	1.100 €	194 €
Labor-time (per patients)	≥ 15dd	8-10dd	≥ 15dd	6-8dd

Table 36: comparison of sequencing reagents costs and time of labor necessary to obtain a complete genotyping for ADPKD and ARPKD causative genes with NGS protocols and traditional Sanger sequencing. Abbreviations: dd: days.

5. CONCLUSIONS

The NGS differential protocols and the data analysis pipeline developed in this study for the molecular screening of PKD-causative genes showed an enhanced sensitivity compared to conventional diagnostic techniques.

In particular, the association between the alternative libraries construction protocol developed for *PKD1*, and the our artificial reference genome created for variant calling process, allowed us to obtain substantial improvement of sensibility and specificity for *PKD1* screening and allowing us to overcome the pseudogene issues.

The NGS assays, preformed on the previously genotyped cohorts that had been received a molecular diagnosis for ADPKD and ARPKD by Sanger sequencing, allowed for the correct detection of all the mutations previously identified. For ADPKD patients, the NGS test identified all 20 mutations (19 in *PKD1* and 1 in *PKD2*), including a repetitive variant in two unrelated patients demonstrating repeatability and reproducibility. In ARPKD validation cohort the individuations of additional variants in six patients suggested a greater sensitivity compared to conventional sequencing.

In the total cohort (n=125) of patients with suspicions of ADPKD phenotype, the NGS method in association with MLPA assay identified molecular defects in 90.4% (n=113) of the patients, 87.6% (n=99) presented mutations in *PKD1* while 12.4% (n=14) in *PKD2*. In 5 of the 12 patients found to be mutation-negative by NGS, the MLPA assay identified large genomic rearrangement. The combination of these two methods achieved a complete and accurate genetic test suitable for ADPKD diagnostic practice. Among all the pathogenic variants found, 58.11% (61 in *PKD1* vs 5 in *PKD2*) resulted previously undescribed: 41 were loss of function that we classified as pathogenic and 25 missense variants that we classified as likely pathogenic or VUS.

The failure to find pathogenic mutations in the remaining 9.6% (n=12) of ADPKD families may be accounted for by the following, not mutually exclusive, reasons. First, the intronic regions, promoters and 5′- and 3′-untranslated regions of both the

PKD1 and *PKD2* loci have never been systematically analyzed in literature, therefore no functional studies were available for non-coding variants in these particular regions. Second, coding variants currently classified as nonpathogenic, including silent mutations and some missense mutations, may be of pathogenic relevance through modulation of mRNA splicing and/or stability; this possibility has not been systematically tested to date. Third, some "ADPKD" cases may have been wrongly diagnosed. Finally, the existence of a third ADPKD locus cannot be excluded.
In the ARPKD cohort, the molecular screening correctly identified two mutations in compound heterozygous status in 14 patients in Group A, whereas in the healthy carriers cohort (Group B) at least one causative mutation was identified in all but three cases, demonstrating an overall detection rate of 85,7%. In those cases where less than two mutations were found on the NGS in probands cohort and in the healthy carriers resulting wild type for the mutational screening the MLPA assay for the investigation of large genomic rearrangement is recommended.

In conclusion, the NGS-based genetic analysis for ADPKD and ARPKD is a highly accurate and reliable approach for mutation analysis achieving high sensitivity and improved intronic coverage, in particular for *PKD1* screening, with a faster turnaround time and lower cost in comparison to Sanger sequencing.
NGS would be an appropriate new standard for clinical genetic testing of polycystic kidney disease.

Bibliography

Alvarez, F., Bernard, O. and Brunelle, F. (1981) 'Congenital hepatic fibrosis in children', *J Pediatr*, vol. 99, p. 370.

Arnould, T., Kim, E. and Tsiokas, L. (1998) 'The polycystic kidney disease 1 gene product mediates protein kinase C alpha-dependent and c-Jun N-terminal kinase-dependent activation of the transcription factor AP-1', *J Biol Chem.*, vol. 273(11), pp. 6013-8.

Audrézet, M., Cornec-Le Gall, E., Chen, J., Redon, S. and al (2012) 'Autosomal dominant polycystic kidney disease: comprehensive mutation analysis of PKD1 and PKD2 in 700 unrelated patients.', *Human Mutation*, vol. 33, pp. 1239-50.

Badano, J., Mitsuma, N., Beales, P. and Katsanis, N. (2006) 'The ciliopathies: an emerging class of human genetic disorders', *Annu Rev Genomics Hum Genet*, vol. 7, pp. 125-48.

Bae, K., Tao, C., Zhu, F., Bost, J., Chapman, A., Grantham, J. and al (2009) 'MRI-based kidney volume measurements in ADPKD: reliability and effect of gadolinium enhancement.', *Clin J Am Soc Nephrol*, vol. 4, pp. 719-25.

Bajwa, Z., Sial, K., Malik, A. and Steinman, T. (2004) 'Pain patterns in patients with polycystic kidney disease. ', *Kidney Int.*, vol. 66, pp. 1561-9.

Bataille, S., Berland, Y., Fontes, M. and al (2011) 'High resolution melt analysis for mutation screening in PKD1 and PKD2', *BMC Nephrol*, vol. 12, p. 57.

Bateman, A. and Sandford, R. (1999) 'The PLAT domain: a new piece in the PKD1 puzzle.', *Curr Biol*, vol. 9, pp. 588-90.

Bentley, D. (2000) 'The Human Genome Project—An Overview', *Med Res Rev*, vol. 20, pp. 189-96.

Bentley, D., Balasubramanian, S., Swerdlow, H. and al (2008) 'Accurate whole human genome sequencing using reversible terminator chemistry', *Nature*, vol. 456, pp. 53-9.

Benzing, T., Simons, M. and Walz, G. (2007) 'Wnt signaling in polycystic kidney disease', *J Am Soc Nephrol*, vol. 18(5), pp. 1389-98.

Bergmann, C., Senderek, J., Kupper, F. and al (2004) 'PKHD1 mutations in Autosomal-Recessive Polycystic Kidney Disease (ARPKD)', *Hum Mutat*, vol. 23(5), pp. 453-63.

Bergmann, C., Senderek, J., Windelen, E. and al (2005) 'Clinical consequences of PKHD1 mutations in 164 patients with Autosomal-Recessive Polycystic Kidney Disease - ARPKD', *Kid Int*, vol. 67(3), pp. 829-48.

Bergmann, C., von Bothmer, J., Ortiz-Bruchle, N. and al. (2011) 'Mutations in multiple PKD genes may explain early and severe polycystic kidney disease', vol. 22, pp. 2047-56.

Bhunia, A., Piontek, K., Boletta, A. and al (2002) 'PKD1 Induces p21(waf1) and Regulation of the Cell Cycle via Direct Activation of the JAK-STAT Signaling Pathway in a Process Requiring PKD2', *Cell*, vol. 109, pp. 157-68.

Bogdanova, N., Markoff, A., Gerke, V., McCluskeyc, M. and al (2001) '. Homologues to the first gene for autosomal dominant polycystic kidney disease are pseudogenes.', *Genomics*, vol. 74, pp. 333-41.

Bogdanova, N., Markoff, A. and Horsta, J. (2002) 'Autosomal Dominant Polycystic Kidney Disease Clinical and Genetic Aspects ', *Kidney Blood Press Res* , vol. 25, pp. 265-83.

Bonon, A., Mangolini, A., Pinton, P., Del Senno, L. and Aguiari, G. (2013) 'slow cell growth in autosomal dominant polycystic kidney disease cells ', *Biochem Biophys Res Commun* , vol. 441, pp. 668-74.

Bycroft, M., Bateman, A., Clarke, J., Hamill, S. and al (1999) 'The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease.', *EMBO J.*, vol. 18(2), pp. 297-305.

Cai, Y., Maeda, Y., Cedzich, A., Torres, V. and al (1999) 'Identification and characterization of polycystin-2, the PKD2 gene product.', *J Biol Chem*, vol. 274, pp. 28557-65.

Carrera, P., Calzavara, S., Magistroni, R. and al (2016) 'Deciphering variability of PKD1 and PKD2 in an italian cohort of 643 patients with autosomal dominant polycystic kidney disease (ADPKD)', *Sci Rep*, vol. 6, pp. 1-13.

Casuscelli, J., Schmidt, S., DeGray, B. and al (2009) 'Analysis of the cytoplasmic interaction between polycystin-1 and polycystin-2', *Am J Physiol Renal Physiol*, vol. 297(5), pp. 1310-5.

Chapin, H. and Caplan, M. (2010) 'The cell biology of polycystic kidney disease', *J Cell Bio*, vol. 191(4), pp. 701–710.

Chapin, H., Rajendran, V. and Caplan, M. (2010) 'Polycystin-1 surface localization is stimulated by polycystin-2 and cleavage at the G protein-coupled receptor proteolytic site', *Mol Biol Cell*, vol. 21, pp. 4338-48.

Chapman, A. (2003) 'Cystic disease in women: clinical characteristics and medical management', *Adv Ren Replace Ther* , vol. 10, pp. 24-30.

Chapman, A., Bost, J., Torres, V. and al (2012) 'Kidney volume and functional outcomes in autosomal dominant polycystic kidney disease', *Clin J Am Soc Nephrol*, vol. 7, pp. 479-86.

Chapman, A., Devuyst, O., Eckardt, U., Ganevoortsevoort, R. and al (2015) 'Autosomal-Dominant Polycystic Kidney Disease (ADPKD): executive summary from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference', *Kidney Int*, vol. 88(1), pp. 17-7.

Chauvet, V., Tian, X., Husson, H., Grimm, D. and al (2004) 'Mechanical stimuli induce cleavage and nuclear translocation of the polycystin-1 C terminus.', *J Clin Invest*, vol. 114(10), pp. 1433-43.

Cole, D. (1999) 'Kinesin-II, the heteromeric kinesin ', *Cell Mol Life Sci*, vol. 56, pp. 217-26.

Consugar, M., Wong, W., Lundquist, P., Rossetti, S. and al (2008) 'Characterization of large rearrangements in autosomal dominant polycystic kidney disease and the PKD1/TSC2 contiguous gene syndrome.', *Kidney Int.*, vol. 74, pp. 1468-79.

Cornec-Le Gall, E., Audrézet, M., Chen, J., Hourmant, M. and al (2013) 'Type of PKD1 mutation influences renal outcome in ADPKD', *J Am Soc Nephrol*, vol. 24, pp. 1006-13.

Cowen, L., Bradley, P., Menke, M., King, J. and Berger, B. (2002) 'Predicting the beta-helix fold from protein sequence data ', *J Comput Biol* , vol. 9, pp. 261-76.

Dedoussis, G., Lou, Y., Starremans, P. and al (2008) 'Co-inheritance of PKD1 mutation and homozygous PKD2 variant: a potential modifier in autosomal dominant polycystic kidney disease', *Eur J Clean Invest*, vol. 38, pp. 180-90.

Denamur, E., Delezoide, A., Albert, C. and al (2010) 'Genotype-phenotype correlations in fetuses and neonates with autosomal recessive polycystic kidney disease', *Kid Int*, vol. 77(4), pp. 350-8.

Dere, R., Wilson, P., Sandford, R. and al (2010) 'Carboxy terminal tail of polycystin-1 regulates localization of TSC2 to repress mTOR', *PLOSE ONE*, vol. 5(2), p. 9239.

Dohm, J., Lottaz, C., Borodina, T. and al (2008) 'Substantial biases in ultra-short read data sets from high-throughput DNA sequencing', *Nucleic Acid Res*, vol. 36, p. 105.

Ecder, T. and Schrier, R. (2009) '. Cardiovascular abnormalities in autosomal-dominant polycystic kidney disease.', *Nat Rev Nephrol*, vol. 5, pp. 221-8.

Erlich, Y., Mitra, P., De La Bastide, M. and al (2008) 'Alta-Cyclic: a self- optimizing base caller for next-generation sequencing', *Nat Methods*, vol. 5, pp. 679-82.

European Polycystic Kidney Disease Consortium (1994) 'The polycystic kidney disease gene 1 encodes a 14 kb transcript and lies within a duplicated region on chromosome 16.', *Cell*, vol. 77, pp. 881-94.

Fisher, E., Legue, E., Doyen, A., Nato, F. and al (2006) 'Defective planar cell polarity in polycystic kidney disease', *Nat Gen*, vol. 38(1), pp. 21-3.

Flusberg, B., Webster, D., Lee, J. and al (2010) 'Direct detection of DNA methylation during single-molecule, real-time sequencing', *Nat Meth*, vol. 7, pp. 461-65.

Furu, L., Onuchic, L., Gharavi, A. and al (2003) 'Milder presentation of recessive polycystic kidney diseases requires presence of aminoacid substitution mutations', *J Am Soc Nephrol*, vol. 14(8), pp. 2004-14.

Gifford, J., Walsh, M. and Vogel, H. (2007) 'Structures and metal-ion-binding properties of the Ca2+binding helix-loop-helix EF-hand motifs', *Biochem J.*, vol. 405(2), pp. 199-221.

Gilbert, R., Sukhtanka, P., Lachlan, K. and Fowler, D. (2013) 'Bilineal inheritance of PKD1 abnormalities mimicking autosomal recessive polycystic disease', *Pediatr Nephrol* , vol. 28, pp. 2217-20.

Glazer, A. and Mathies, R. (1997) 'Energy- transfer fluorescent reagent for DNA analyses ', *Curr Opin Biotechnol*, vol. 8(1), pp. 94-102.

Gonzalez-Perrett, S., Kim, K., Ibarra, C., Damiano, A. and al (2001) 'Polycystin-2, the protein mutated in autosomal dominant polycystic kidney disease (ADPKD), is a Ca2+-permeable nonselective cation channel.', *Proc Natl Acad Sci USA*, vol. 98(3), pp. 1182-7.

Goodwin, S., McPherson, J. and McCombie, W. (2016) 'Coming of age: ten years of next-generation sequencing technologies', *Nat Rev Genet*, vol. 17(6), pp. 333-51.

Grantham, J. (1997) 'Polycystic kidney disease: huge kidneys, huge problems, huge progress ', *Trans Am Clin Climatol Assoc.*, vol. 108, pp. 165-70.

Grantham, J., Chapman, A. and Torres, V. (2006) 'Volume progression in autosomal dominant polycystic kidney disease: the major factor determining clinical outcomes', *Clin J Am Soc Nephrol* , vol. 1, pp. 148-57.

Grantham, J., Torres, V., Chapman, A., Guay-Woodford, L. and al (2006) 'Volume Progression in Polycystic Kidney Disease', *N Eng J Med*, vol. 354, pp. 2122-130.

Guay-Woodford, L. (2003) 'Murine models of polycystic kidney disease: molecular and therapeutic insight', *Am J of Physiol*, vol. 285, pp. 1034-49.

Gunay-Aygun, M., Tuchman, M., Font-Montgomery, E. and al (2010) 'PKHD1 sequence variation in 78 children and audults with Autosomal Recessive Kidney Disease and congenital hepatic fibrosis', *Mol Genet Metab*, vol. 99(2), pp. 160-73.

Hadimeri, H., Lamm, C. and Nyberg, G. (1998) 'Coronary aneurysms in patients with autosomal dominant polycystic kidney disease.', *J Am Soc Nephrol* , vol. 9, pp. 837-41.

Halvorson, C., Bremmer, M. and Jacobs, S. (2010) 'Polycystic kidney disease: inheritance, pathophysiology, prognosis, and treatment', *Int J Nephrol Renovasc Dis* , vol. 3, pp. 69-83.

Hanaoka, K., Qian, F., Boletta, A., Bhunia, A. and al (2000) 'Co-assembly of polycystin-1 and -2 produces unique cation-permeable currents', *Nature*, vol. 48(6815), pp. 990-4.

Harris, P. (1997) 'The TSC2/PKD1 contiguous gene syndrome', *Contrib Nephrol* , vol. 122, pp. 76-82.

Harris, P. and Rossetti, S. (2004) 'Molecular genetics of autosomal recessive polycystic kidney disease', *Mol Genet Metab* , vol. 81, pp. 75-85.

Harris, P. and Torres, V. (2009) 'Polycystic kidney disease', *Annu Rev Med*, vol. 60, pp. 321-37.

Hateboer, N., v Dijk, M., Bogdanova, N., Coto, E., Saggar-Malik, A. and al (1999) 'Comparison of phenotypes of polycystic kidney disease types 1 and 2. European PKD1-PKD2 Study Group', *Lancet*, vol. 353, pp. 103-7.

Higashihara, E., Horie, S., Muto, S., Mochizuki, T., Nischio, S. and Nutahara, K. (2012) 'Renal disease progression in autosomal dominant polycystic kidney disease. ', *Clin Exp Nephrol* , vol. 16, pp. 622-28.

Hildebrandt, F., Benzing, T. and Katsanis, N. (2011) 'Ciliopathies', *N Engl j Med*, vol. 364, pp. 1533-43.

Hogan, M., Manganelli, L., Woollard, J., Masyuk, A. and al (2009) 'Characterization of PKD protein-positive exosome-like vescicles', *J Am Soc Nephrol*, vol. 20(2), pp. 278-88.

Horikawa, Y., Iwasaki, N., Hara, M., Furuta, H., Hinokio, Y., Cockburn, B., Lindner, T., Yamagata, K., Ogata, M., Tomonaga, O., Kuroki, H., Kasahara, T., Iwamoto, Y. and Bell, G. (1997) 'Mutation in hepatocyte nuclear factor-1-beta gene (TCF2) associated with MODY', *Nature Genet*, vol. 17, pp. 384-5.

Horsley, V. and Pavlath, G. (2002) 'NFAT: ubiquitous regulator of cell differentiation and adaptation', *Cell Biol*, vol. 156(5), pp. 771-4.

Huang, J. and Manning, B. (2008) 'The TSC1-TSC2 complex: a molecular switchboard controlling cell growth', *Biochem J* , vol. 412(2), pp. 179-9.

Huang, E., Picotal, M., McCune, T., Melancon, J., Montgomery, R. and al (2009) 'DNA testing for live kidney donors at risk for autosomal dominant polycystic kidney disease.', *Transplantation*, vol. 87, pp. 133-37.

Huan, Y. and Van Adelsberg, J. (1999) 'Polycystin-1, the PKD1 gene product, is in a complex containing E-cadherin and the catenins. ', *Clin Invest* , vol. 104(10), pp. 1459-68.

Hughes, J., Ward, C.J., Peral, B., Aspinwall, R. and al (1995) 'The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains', *Nature Genetics*, vol. 10, pp. 151-60.

Huse, S., Huber, J., Morrison, H. and al (2007) 'Accuracy and quality of massively parallel DNA pyrosequencing', *Genome Biol*, vol. 8, p. 143.

Ibraghimov-Beskrovnaya, O. and Bukanov, N. (2008) 'Polycystic kidney disease: from molecular discoveries to targeted therapeutic strategies', *Cell Mol Life Sci*, vol. 65, pp. 605-19.

Ibraghimov-Beskrovnaya, O., Bukanov, N., Donohue, L., Dackowski, W., Klinger, K. and Landes (2000) 'Strong homophilic interactions of the Ig-like domains of polycystin-1, the protein product of an autosomal dominant polycystic kidney disease gene, PKD1', *Hum Mol Gen*, vol. 9, pp. 1641-49.

Ibraghimov-Beskrovnaya, O., Dackowski, W., Foggensteiner, L., Coleman, N., Thiru, S., Petry, L., Burn, T., Connors, T., Van Raay, T. and Bradley, J. (1997) 'Polycystin:

in vitro synthesis, in vivo tissue expression, and subcellular localization identifies a large membrane-associated protein ', *Proc Natl Acad Sci* , vol. 94, pp. 6397-402.

Igarashi, P. and Somlo, S. (2002) 'Genetics and pathogenesis of polycystic kidney disease', *J Am Soc Nephrol.*, vol. 13(9), pp. 2384-98.

International Polycystic Kidney Disease Consortium (1994) 'The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16', *Cell*, vol. 77, pp. 881-94.

Irazabal, M., Huston, J.3., Kubly, V., Rossetti, S. and al (2011) 'Extended follow-up of unruptured intracranial aneurysms detected by presymptomatic screening in patients with autosomal dominant polycystic kidney disease', *Clin J Am Soc Nephrol*, vol. 6(6), pp. 1274-85.

Itty, C., Farshid, A. and Talaulikar, G. (2009) 'Spontaneous coronary artery dissection in a woman with polycystic kidney disease. ', *Am J Kidney Dis* , vol. 53, pp. 518-21.

Jung, G., Benz-Bohm, G., Kugel, H. and al, e. (1999) 'MR cholangiography in children with autosomal recessive polycystic kidney disease', *Pediatr Radiol*, vol. 29, p. 463.

Kääriäinen, H., Koskimies, O. and Norio, R. (1988) 'Dominant and recessive polycystic kidney disease in children: Evaluation of clinical features and laboratory data', *Pediatr Nephrol*, vol. 2, p. 296.

Kaplan, B., Kaplan, P., Rosenbeg, H. and al (1989) 'Polycystic kidney disease in childhood', *J Pediatr*, vol. 115, p. 867.

Kim, I., Ding, T., Fu, Y., Li, C., Cui, L. and al (2009) 'Conditional mutation of Pkd2 causes cystogenesis and upregulates beta-catenin', *J Am Soc Nephrol*, vol. 20(12), pp. 2556-69.

King, B., Torres, V., Brummer, M., Chapman, A., Bae, K., Glockner, J. and al (2003) 'Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP): Magnetic resonance measurements of renal blood flow as a marker of disease severity in autosomal-dominant polycystic kidney disease.', *Kidney Int*, vol. 64, pp. 2214-21.

Kolb, R. and SM, N. (2008) 'Ciliary dysfunction in polycystic kidney disease: an emerging model with polarizing potential', *Front in Bioscience*, vol. 13, pp. 4451-66.

Koulen, P., Cai, Y., Geng, L., Maeda, Y., Nishimura, S., Witzgall, R. and al (2002) 'Polycystin-2 is an intracellular calcium release channel', *Nat Cell Biol*, vol. 4(3), pp. 191-7.

Lal, M., Son, g.X., Pluznick, J. and al (2008) 'Polycystin-1 C-terminal tail associates with beta-catenin and inhibits canonical Wnt signaling', *Hum Mol Genet*, vol. 17, pp. 3105-17.

Lantinga-van Leeuwen, I., Dauwerse, J., Baelde, H. and al (2004) 'Lowering of Pkd1 expression is sufficient to cause polycystic kidney disease', *Hum Mol Genet*, vol. 13(24), pp. 3069-77.

Lawson, C., Doulton, T. and MacGregor, G. (2006) 'Autosomal dominant polycystic kidney disease: role of the renin-angiotensin system in raised blood pressure in progression of renal and cardiovascular disease', *J Renin Angiotensin Aldosterone Syst*, vol. 7, pp. 139-45.

Le, N., Van Der Wal, A., Van Der Bent, P. and al (2005) 'Increased activity of activator protein-1 transcription factor components ATF2, c-Jun, and c-Fos in human and mouse autosomal dominant polycystic kidney disease', *J Am Soc Nephrol*, vol. 16(9), pp. 2724-31.

Liang, G., Yang, J., Wang, Z., Li, Q. and al (2008) 'Polycystin-2 down-regulates cell proliferation via promoting PERK-dependent phosphorylation of eIF2alpha.', *Hum Mol Genet*, vol. 17(20), pp. 3254-62.

Li, X., Luo, Y., Starremans, P. and al (2005) 'Polycystin-1 and polycystin-2 regulate the cell cycle through the helix-loop-helix inhibitor Id2', *J Nat Cell Biol*, vol. 7(12), pp. 1202-12.

Loosekoot, M., Haarloo, C., Ruivenkamp, C. and al (2005) 'Analysis of missense variants in the PKHD1 gene in patiens with Autosomal-Recessive Polycystic Kidney Disease (ARPKD)', *Hum Genet*, vol. 118(2), pp. 185-206.

Losekoot, M., Ruivenkamp, C., Tholens, A., Grimbergen, J., Vijfhuizen, L., Vermeer, S., Dijkman, H., Cornelissen, E., Bongers, E. and Peters, D. (2012) 'Neonatal onset autosomal dominant polycystic kidney disease (ADPKD) in a patient homozygous for a PKD2 missense mutation due to uniparental disomy', *J Med Genet*, vol. 49, pp. 37-40.

Low, S.H., Vasanth, S., Larson, C.H., Mukherjee, S., Sharma, N., Kinter, M.T., Kane, M.E., Obara, T. and Weimbs, T. (2006) 'Polycystin-1, STAT6, and P100 function in a pathway that transduces ciliary mechanosensation and is activated in polycystic kidney disease', *Dev. Cell.*, vol. 10, pp. 57-69.

Lu, W., Peissel, B., Babakhanlou, H. and al (1997) 'Perinatal lethality with kidney and pancreas defects in mice with a targetted Pkd1 mutation', *J Nat Genet*, pp. 179-81.

Magistroni, R., He, N., Wang, K. and al (2003) 'Genotype-renal function correlation in type 2 autosomal dominant polycystic kidney disease', *J Am Soc Nephrol*, vol. 14(5), pp. 1164-74.

Mardis, E. (2013) 'Next-Generation Sequencing Platforms', *Ann Rev of Anal Chem*, vol. 6, pp. 287-303.

Margulies, M., Egholm, M., Altman, W. and al (2005) 'Genome sequencing in microfabricated high-density picolitre reactors', *Nature*, vol. 437, pp. 376-80.

Marshall, W. and Nonaka, S. (2006) 'Cilia: tuning in to the cell's antenna', *Curr Biol*, vol. 16(15), pp. 604-14.

McNeill, H. (2009) 'Planar cell polarity and the kidney', *J Am Soc Nephrol*, vol. 20(10), pp. 2104-11.

Menezes, L. and Onuchic, L. (2006) 'Molecular and cellular pathogenesis of autosomal recessive polycystic kidney disease', *Braz J Med Biol Res*, vol. 39, pp. 1537-48.

Merriman, B., Rothberg, J. and al (2012) 'Progress in ion torrent semiconductor chip based sequencing', *Electrophoresis*, vol. 33(23), pp. 3397-417.

Miki, H., Setou, M., Kaneshiro, K. and Hirokawa, N. (2001) 'All kinesin superfamily protein, KIF, genes in mouse and human', *Proc Natl Acad Sci* , vol. 39, pp. 7004-11.

Mochizuki, T., Wu, G., Hayashi, T., Xenophontos, S. and al (1996) 'PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein', *Science*, vol. 272, pp. 1339-42.

Mucher, G., Wirth, B. and Zerres, K. (1994) 'Refining the map and defining flanking markers of the gene for autosomal recessive polycystic kidney disease on chromosome 6p21.1-p12', *J Hum Genet*, vol. 55, pp. 1281-84.

Nauli, S.M., Alenghat, F.J., Luo, Y., Williams, E., Vassilev, P., Li, X. and Elia, A.E.H. (2003) 'Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells', *Nat. Genet*, vol. 33, pp. 129-37.

Nims, N., Vassmer, D. and Maser, R. (2003) 'Transmembrane domain analysis of polycystin-1, the product of the polycystic kidney disease-1 (PKD1gene: evidence for 11 membrane-spanning domains', *Biochemistry*, vol. 42(44), pp. 13035-48.

Ong, A. and Harris, P. (2015) 'A polycystin-centric view of cyst formation and disease: the polycystins revisited', *Kidney Int*, vol. 88(4), pp. 699-710.

Onuchic, L., Furu, L., Nagasawa, Y. and al (2002) 'PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-trascription factor domains and parallele Beta-Helix 1 repeats', *Ann J Hum Genet*, vol. 70(5), pp. 1305-17.

Parnell, S., Magenheimer, B., Maser, R. and al (1998) 'The polycystic kidney disease-1 protein, polycystin-1, binds and activates heterotrimeric G-proteins in vitro', *Biochem Biophys Res Commun*, vol. 251(2), pp. 625-31.

Patel, A. and Honoré, E. (2010) 'Polycystins and renovascular mechanosensory transduction', *Nat Rev Nephrol*, vol. 6(9), pp. 530-8.

Patel, V., Li, L., Cobo-Stark, P., Shao, X. and al (2008) 'Acute kidney injury and aberrant planar cell polarity induce cyst formation in mice lacking renal cilia', *Hum Mol Genet*, vol. 17(11), pp. 1578-90.

Paterson, A., Wang, K., Lupea, D., St George-Hyslop, P. and Pei, Y. (2002) 'Recurrent fetal loss associated with bilineal inheritance of type 1 autosomal dominant polycystic kidney disease', *Am J Kidney Dis*, vol. 40, pp. 16-20.

Pearson, B., Gaskin, D., Segers, R. and al (2007) 'The complete genome sequence of Campylobacter jejuni strain 81116 (NCTC11828)', *J Bacteriol*, vol. 189, pp. 8402-3.

Peces, R., Peces, C., Coto, E. and Selgas, R. (2008) 'Bilineal inheritance of type 1 autosomal dominant polycystic kindney disease (ADPKD) and recurrent fetal loss', *NTD plus*, vol. 5, pp. 289-91.

Pei, Y. (2001) 'A "two-hit" model of cystogenesis in autosomal dominant polycystic disease?', *Trend Mol Med*, vol. 7(4), pp. 151-6.

Pei, Y., Hwang, Y., Conklin, J. and al (2015) 'Imaging-based diagnosis of autosomal dominant polycystic kidney disease', *J Am Soc Nephrol*, vol. 26(3), pp. 746-53.

Pei, Y., Lan, Z., Wang, K., Garcia-Gonzalez, M., He, N., Dicks, E., Parfrey, P., Germino, G. and Watnick, T. (2012) 'A missense mutation in PKD1 attenuates the severity of renal disease', *Kidney Int* , vol. 81, pp. 412-17.

Pei, Y., Obaji, J., Dupuis, A., Paterson, A., Magistroni, R., Dicks, E. and al (2009) 'Uni ed criteria for ultrasonographic diagnosis of ADPKD', *J Am Soc Nephrol*, vol. 20, pp. 205-12.

Perrichot, R., Mercier, B., Simon, P., Whebe, B., Cledes, J. and Ferec, C. (1999) 'DGGE screening of PKD1 gene reveals novel mutations in a large cohort of 146 unrelated patients', *Hum Genet*, vol. 105(3), pp. 231-9.

Peters, D., Van de Wal, A., Spruit, L., Breuning, M. and al (1999) 'Cellular localization and tissue distribution of polycystin-1', *J Pathol*, vol. 188(4), pp. 439-46.

PKD foundations (2017) *Polycystic Kydney Disesase*, [Online], Available: HYPERLINK "https://pkdcure.org/what-is-pkd/adpkd/what-are-cysts/" https://pkdcure.org/what-is-pkd/adpkd/what-are-cysts/ [14 March 2017].

Poureetezadi, S. and Wingert, R. (2016) 'Little fish, big catch: zebrafish as a model for kidney disease', *Kidney Int*, vol. 89, pp. 1204-10.

Praetorius, H.A. and Spring, K.R. (2001) 'Bending the MDCK cell primary cilium increases intracellular calcium', *J. Membr. Biol*, vol. 184, pp. 71-79.

Presenti Gritti, A. and Boletta, A. (2015) 'Le Malattie del Ciglio-Rene Policistico', *GIN*, vol. 6.

Pritchard, L., Sloane-Stanley, J., Sharpe, J. and al (2000) 'A human PKD1 transgene generates functional polycystin-1 in mice and is associated with a cystic phenotype', *Human Mol Genet*, vol. 9(18), pp. 2617-27.

Prober, J., Trainor, G., Dam, R., Hobbs, F. and al (1987) 'A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides', *Science*, vol. 238, pp. 336-41.

Qian, F., Germino, F., Cai, Y., Zhang, X. and al (1997) 'PKD1 interacts with PKD2 through a probable coiled-coil domain', *Nat Genet*, vol. 16(2), pp. 179-83.

Quian, F., Boletta, A., Bhunia, A., Xu, H., Liu, L. and al (2002) 'Cleavage of polycystin-1 requires the receptor for egg jelly domain and is disrupted by human autosomal-dominant polycystic kidney disease 1-associated mutations', *Proc Natl Acad Sci USA*, vol. 99(26), pp. 16981-6.

Rantman, S. and Nauli, S. (2010) 'Hypertension in autosomal dominant polycystic kidney disease: a clinical and basic science perspective', *Int J Nephrol Urol*, vol. 2, pp. 294-308.

Rechsteiner, M. and Rogers, S. (1996) 'PEST sequences and regulation by proteolysis', *Biochem Sci.*, vol. 21(7), pp. 267-71.

Reiterova, J., Stekrova, J., Merta, M., Kotlas, J., Elisakova, V., Lnenicka, P., Korabecna, M., Kohoutova, M. and Tesar, V. (2013) 'Autosomal dominant polycystic kidney disease in a family with mosaicism and hypomorphic allele ', *BMC Nephrol* , vol. 14, p. 19.

Reynolds, D., Hayashi, T., Cai, Y., Veldhuisen, B. and al (1999) 'Aberrant splicing in the PKD2 gene as a cause of polycystic kidney disease', *J Am Soc Nephrol*, vol. 10(11), pp. 2342-51.

Richards, S., Aziz, N., Bale, S. and al (2015) 'Standards and Guidelines for the enterpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the association for molecular pathology', *Gen in Med*, vol. 17, pp. 405-23.

Riella, C., Czarnecki, P. and Steinman, T. (2014) 'Therapeutic advances in the treatment of polycystic kidney disease.', *Nephron Clin Pract*, vol. 128, pp. 297-02.

Rizzo, J. and Buck, M. (2012) ' Key principles and clinical applications of "next-generation" DNA sequencing cing. Cancer Prev Res (Phila) 5: 887-900. ', *Cancer Prev Res* , vol. 5, pp. 887-900.

Rodat-Despoix, L. and Delmas, P. (2009) 'Ciliar functions in the nephron', *Eur J Physiol*, vol. 458, p. 179.

Romão, E., Moysés Neto, M., Teixeira, S., Muglia, V., Vieira Neto, O. and Dantas, M. (2006) 'Renal and extrarenal manifestation of autosomal dominant polycystic kidney disease', *Braz J Med Biol Res* , vol. 39, pp. 533-38.

Rossetti, S., Burton, S., Strmecki, L., Pond, G., San Millan, J. and al, e. (2002) '. The position of the polycystic kidney disease 1 (PKD1) gene mutation correlates with the severity of renal disease', *J Am Soc Nephrol*, vol. 13, pp. 1230-37.

Rossetti, S., Chauveau, D., Kubly, V., Slezak, J., Saggar-Malik, A. and al (2003) 'Association of mutation position in polycystic kidney disease 1 (PKD1) gene and development of a vascular phenotype', *Lancet*, vol. 361, pp. 2196-201.

Rossetti, S., Consugar, M., Chapman, A., Torres, V., Guay-Woodford, L. and al (2007) 'Comprehensive molecular diagnostics in autosomal dominant polycystic kidney disease', *JASN*, vol. 18, pp. 2143-60.

Rossetti, S. and Harris, P. (2007) 'Genotype-phenotype correlations in autosomal dominant and autosomal recessive polycystic kidney disease', *J Am Soc Nephrol*, vol. 18(5), pp. 1374-80.

Rossetti, S. and Harris, P.C. (2007) 'Genotype-phenotype correlations in autosomal dominant and autosomal recessive polycystic kidney disease ', *J. Am. Soc. Nephrol*, vol. 18, pp. 1374-80.

Rossetti, S., Hope, K., Sikkink, R. and al (2012) 'Identification of gene mutations in autosomal dominant polycystic kidney disease through target resequencing', *J Am Soc Nephrol*, vol. 23(5), pp. 915-33.

Rossetti, S., Kubly, V., Consugar, M., Hopp, K., Roy, S., Horsley, S., Chauveau, D., Rees, L., Barratt, T., van't Hoff, W., Niaudet, P., Torres, V. and al (2009) 'Incompletely penetrant PKD1 alleles suggest a role for gene dosage in cyst initiation in polycystic kidney disease', *Kid Int*, vol. 75, pp. 848-55.

Rossetti, S., Strmecki, L., Gamble, V., Burton, S. and al (2001) 'Mutation analysis of the entire PKD1 gene: genetic and diagnostic implications.', *Am J Hum Genet*, vol. 68, pp. 46-63.

Rossetti, S., Torra, R., Coto, E. and al (2003) 'A completed mutation screen of PKHD1 in Autosomal-Recessive Polycystic Kidney Disease (ARPKD) predigrees', *Kid Int*, vol. 64(2), pp. 391-403.

Rowe, I., Chiaravalli, M., Mannella, V., Ulisse, G. and al (2013) 'Defective glucose metabolism in polycystic kidney disease identifies a new therapeutic strategy', *Nature Medicine*, vol. 19, pp. 488-93.

Roy, S., Dillon, M., Trompeter, R. and TM, B. (1997) 'Autosomal recessive polycystic kidney disease: long-term outcome of neonatal survivors. ', *Pediatr Nephrol*, vol. 11, pp. 302-6.

Sampson, J., Maheshwa, M., Aspinwall, R., Thompson, P. and al (1997) 'Cystic disease in tuberous sclerosis: role of the polycystic kidney disease 1 gene.', *Am J Hum Genet*, vol. 61, pp. 843-51.

Sanger, F., Nicklen, S. and Coulson, A. (1977) 'DNA sequencing with chain-terminating inhibitors.', *Proceedings of the National Academy of Sciences USA*, vol. 74, pp. 5463-67.

Schouten, J., McElgunn, C., Waaijer, R. and al (2002) 'Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification', *Nucleic Acids Res*, vol. 30(12), p. 57.

Seeman, T., Dusek, J., Vondràk, K., Blàhovà, K. and al (2004) 'Renal concentrating capacity is linked to blood pressure in children with autosomal dominant polycystic kidney disease.', *Physiol Res.*, vol. 53, pp. 629-34.

Sharp, A., Messiaen, L.P.G. and al (2005) 'Comprehensive genomic analysis of PKHD1 mutations in ARPKD cohorts', *J Med Genet*, vol. 42(4), pp. 336-49.

Shaulian, E. and Karin, M. (2002) 'AP-1 as a regulator of cell life and death', *Nat Cell Biol* , vol. 4(5), pp. 131-6.

Shendure, J., Ji and H (2008) 'Next-generation DNA sequencing.', *Nat Biotech*, vol. 26, pp. 1135-45.

Shillingford, J., Murcia, N., Larson, C., Low, S. and al (2006) 'The mTOR pathway is regulated by polycystin-1, and its inhibition reverses renal cystogenesis in polycystic kidney disease', *Proc Natl Acad Sci USA*, vol. 103(14), pp. 5466-71.

Singla, V. and Reiter, J.F. (2006) 'The primary cilium as the cell's antenna: signaling at a sensory organelle ', *Science*, vol. 313, pp. 629-33.

Somlo, S., Torres, V. and Caplan, M. (2008) 'Autosomal dominant polycystic kidney disease and inherited cystic diseases', *In The Kidney* , vol. 2, pp. 2283-314.

Srivastava, A. and Patel, N. (2014) 'Autosomal dominant polycystic kidney disease', *Am Fam Physician* , vol. 90, pp. 303-7.

Stekrova, J., Reiterova, J., Svobodova, S. and al (2009) 'New mutation in the PKD1 gene in czech population with autosomal dominant polycystic kidney disease', *BMC Med Genet*, vol. 10, p. 78.

Strausberg, R., Levy, S. and Rogers, Y. (2008) 'Emerging DNA sequencing technologies for human genomic medicine', *Drug Discov Today*, vol. 13, pp. 569-77.

Streets, A., Wagner, B. and Harris, P. (2009) 'Homophilic and heterophilic polycystin 1 interactions regulate E-cadherin recruitment and junction assembly in MDCK cells', *J Cell Sci*, vol. 122, pp. 1410-17.

Suzuki, T., Delgado-Escueta, A., Alonso, M. and al (2006) 'Mutation analyses of genes on 6p12-p11 in patients with juvenile myoclonic epilepsy', *Neurosci Lett*, vol. 405(1-2), pp. 126-31.

Sweeney, W.J., Gunay-Aygun, M., Patil, A. and Avner, E. (2011) 'Diagnosis and management of childhood polycystic kidney disease.', *Pediatr Nephrol*, vol. 26, pp. 675-92.

Takamitsu, S. and Bell, P. (2015) 'Molecular Pathways and Therapies in Autosomal-Dominant Polycystic Kidney Disease', *Physiology*, vol. 30, pp. 195-207.

Takiar, V. and Caplan, M. (2011) 'Polycystic kidney disease: pathogenesis and potential therapies.', *Biochim Biophys Acta*, vol. 1812, pp. 1337-43.

Tan, Y., Blumenfeld, J., Angel, R. and al (2009) 'Novel methods for genomic analysis of PKD1 and PKD2 mutations in Autosomal Dominant Polycystic Kidney Diesase', *Hum Mutat*, vol. 30(2), pp. 264-73.

Tan, Y., Michaeel, A., Genyan, L. and al (2014) '. Molecular Diagnosis of autosomal dominant polycystic kidney disease using next-generation sequencing.', *J Mol Diag*, vol. 16, pp. 216-28.

The American PKD1 Consortium (1995) 'Analysis of the genomic sequence for the autosomal dominant polycystic kidney disease (PKD1) gene predicts the presence of a leucine-rich repeat', *Hum Mol Genet*, vol. 4, pp. 575-82.

Torres, V., Chapman, A.B., Devuyst, O., Gansevoort, R.T. and al (2012) 'TEMPO 3:4 trial investigators. Tolvaptan in patients with autosomal dominant polycystic kidney disease.', *N Engl J Med*, vol. 367, pp. 2407-18.

Torres, V., Harris, P. and Pirson, Y. (2007) 'Autosomal Dominant Polycystic Kidney Disease', *Lancet*, vol. 369, pp. 1287-301.

Tsiokas, L., Arnould, T., Zhu, C.K.E. and al (1999) 'Specific association of the gene product of PKD2 with the TRPC1 channel.', *Proc Natl Acad Sci USA*, vol. 96(7), pp. 3934-9.

Tsiokas, L., Kim, E., Arnould, T. and al (1997) 'Homo- and heterodimeric interactions between the gene products of PKD1 and PKD2', *Proc Natl Acad Sci USA*, vol. 94(13), pp. 6965-70.

Turco, A., Bresin, E., Rossetti, S., Englisch, S., Pignatti, P. and al (1997) 'Molecular genetic investigations in autosomal dominant polycystic kidney disease. Gene Mutation detection, linkage analysis, and preliminary ACE gene I/D polymorphism association studies: an update', *Contrib Nephrol*, vol. 122, pp. 53-7.

Turkbey, B., Ocak, I., Daryanani, K., Font-Montgomery, E. and al (2009) 'Autosomal recessive polycystic kidney disease and congenital hepatic fibrosis (ARPKD/CHF)', *Pediatr Radiol*, vol. 39, pp. 100-11.

University of Virginia (2015) 'University of Virginia', Available: HYPERLINK "https://www.med-ed.virginia.edu/courses/rad/gu/kidneys/Images/adpckdus.jpg" https://www.med-ed.virginia.edu/courses/rad/gu/kidneys/Images/adpckdus.jpg [28 January 2015].

Vassilev, P., Guo, L., Chen, X., Segal, Y. and al (2001) 'Polycystin-2 is a novel cation channel implicated in defective intracellular Ca(2+) homeostasis in polycystic kidney disease', *Biochem Biophys Res Commun*, vol. 282(1), pp. 341-50.

Venter, J., Adams, M., Myers, E., Li, P., Mural, R. and al (2001) 'The sequence of the human genome', vol. 291, pp. 1304-51.

Vujic, M., Heyer, C., Ars, E. and al (2010) 'Incompletely penetrant PKD1 alleles mimic the renal manifestations of ADPKD', *J Am Soc Nephrol*, vol. 21, pp. 1097-102.

Wang, S., Zhang, J., Nauli, S.M., Li, X., Starremans, P.G., Luo, Y., Roberts, K.A. and Zhou, J. (2007) 'Fibrocystin/polyductin, found in the same protein complex with polycystin-2, regulates calcium responses in kidney epithelia', *Mol. Cell. Biol.*, vol. 27, pp. 3241-52.

Ward, C., Hogan, M., S, R. and al (2002) 'The gene mutated in Autosomal Recessive Polycystic Kidney disease encodes a large-receptor like protein', *Nat Genet*, vol. 30(3), pp. 259-69.

Ward, C.T.H., Ong, A., Comley, M., Biddolph, S., Chetty, R., Ratcliffe, P., Gattner, K. and Harris, P. (1996) 'Polycystin, the polycystic kidney disease 1 protein, is expressed by epithelial cells in fetal, adult, and polycystic kidney', *Proc Natl Acad Sci*, vol. 93, pp. 1524-28.

Wheeler, D., Srinivasan, M., Egholm, Y., Shen, L. and al (2008) 'The complete genome of an individual by massively parallel DNA sequencing', *Nature*, vol. 452, pp. 872-6.

Wilson, P. (2004) 'Polycystic Kidney Disease', *N Engl J Med*, vol. 350, pp. 151-64.

Wu, Y., Dai, X., Li, Q., Chen, C., Mai, W., Hussain, Z. and al, e. (2006) 'Kinesin-2 mediates physical and functional interactions between polycystin-2 and fibrocystin. ', *Hum Mol Genet* , vol. 15(22), pp. 3280-92.

Wu, G., Tian, X., Nishimura, S., Markowitz, G., D'Agati, V., Park, J. and al (2002) 'Trans-heterozygous Pkd1 and Pkd2 mutations modify expression of polycystic kidney disease', *Hum Mol Genet*, vol. 11, pp. 1845-54.

Xu, C., Rossetti, S., Jiang, L., Harris, P. and al (2007) 'Human ADPKD primary cyst epithelial cells with a novel, single codon deletion in the PKD1 gene exhibit defective ciliary polycystin localization and loss of flow-induced Ca2+ signaling', *Am J Physiol Renal Physiol.*, vol. 292(3), pp. 930-45.

Yeung, B., Danielpour, M., Matsumura, J., Ailawadi, G., Batjer, H. and Yao, J. (2000) 'Incidental asymptomatic cerebral aneurysms in patients with extracranial cerebrovascular disease: is this a case against carotid endarterectomy without arteriography?', *Cardiovasc Surg* , vol. 8(7), pp. 513-8.

Yoder, B.K. (2007) 'Role of primary cilia in the pathogenesis of polycystic kidney disease ', *J. Am. Soc. Nephrol* , vol. 18, pp. 1381-88.

Yoder, B., Hou, X. and Guay-Woodford, L. (2002) 'The polycystic kidney disease proteins, polycystin-1, polycystin-2, polaris, and cystin, are co-localized in renal cilia.', *J Am Soc Nephrol* , vol. 13(10), pp. 2508-16.

Yu, S., Hackmann, K. and Gao, J. (2007) 'Essential role of cleavage of Polycystin-1 at G protein-coupled receptor proteolytic site for kidney tubular structure', *Prot Natl Acad Sci*, vol. 104, pp. 18688-93.

Zerres, K., Rudnick, S. and al (1998) 'Autosomal Recessive Kidney Disease', *J Mol Med*, vol. 76, pp. 303-9.