UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova
Dipartimento di Medicina

SCUOLA DI DOTTORATO DI RICERCA IN: SCIENZE MEDICHE, CLINICHE E SPERIMENTALI
INDIRIZZO: FISIOPATOLOGIA CLINICA E SCIENZE NEFROLOGICHE
CICLO: XV

GENOMIC APPROACH TO IDIOPATHIC CALCIUM NEPHROLITHIASIS: SEARCHING FOR SUSCEPTIBILITY GENES IN A THREE GENERATION FAMILY

Direttore della Scuola : Ch.mo Prof. Gaetano Thiene
Coordinatore d'indirizzo: Ch.mo prof. Aldo Baritussio
Supervisore : Dott.ssa Franca Anglani

Dottorando : Alessandra Familiari

TABLE OF CONTENTS

ABSTRACT

La nefrolitiasi calcica idiopatica (NCI) è una malattia multifattoriale, la cui patogenesi dipende dall'interazione di fattori ambientali, metabolici, anatomici e genetici.
L'importanza dei fattori ereditari è emersa da diversi studi su famiglie e gemelli affetti da NCI. Il modello di ereditarietà più probabile sembrerebbe, come predetto da programmi bioinformatica, ad eredità monogenica co-dominante o poligenica.
La maggior parte degli studi a gene candidato o di associazione condotti fin ora ha prodotto risultati negativi o marginali, lasciando pertanto le basi genetiche della NCI una questione ancora aperta.
Tuttavia, le nuove tecnologie di analisi genetica sono in continua evoluzione e potrebbero contribuire ad approfondire le nostre conoscenze sulla patogenesi della NCI.
Le tecnologie *whole-genome scanning,* quali *Next-Generation Sequencing* (NGS) and CGH (*Comparative Genomic Hybridization*) array, rappresentano gli approcci di genetica molecolare più recenti, e permettono di tracciare una mappa completa della variabilità presente nel genoma umano, in particolare SNPs (*Single Nucleotide Polymorphisms*) e varianti strutturali. Negli ultimi anni l'importanza di queste ultime, maggiormente rappresentate dalle *Copy-Number Variants* (CNVs), è stata infatti dimostrata in numerosi studi che ne hanno individuato un ruolo critico per lo sviluppo di malattie genetiche complesse.

Scopo di questo studio è stata la ricerca di alcune delle maggiori varianti di suscettibilità alla NCI, applicando l'enorme potenziale delle tecnologie *whole genome scanning* ad un caso studio molto peculiare, sia per manifestazioni cliniche che per il *bagkground* genetico.
Il probando da circa 30 anni è affetto da una forma severa di NCI, con ipercalciuria intermittente e tubulopatia fosfaturica, espellendo fino ad oggi circa 300 calcoli, sia spontaneamente che mediante litotripsia (33 interventi). Egli appartiene ad una famiglia con consanguineità, in cui la nefrolitiasi viene trasmessa in modo apparentemente dominante, con maschi affetti più severamente e meno frequentemente rispetto alle femmine.
Da un'analisi CGH array, al fine di individuare la presenza di eventuali CNV, è stato possibile identificare una duplicazione di una regione di 308 Kb del cromosoma X (Xq22.2

CNV) del probando. Questa CNV è risultata particolarmente interessante poiché fin ora mai identificata e condivisa da altri tre familiari (eredità materna).

Inoltre, confrontando la duplicazione Xq22.2 con un database di 14375 individui sono state identificate solo 5 duplicazioni " overlappanti", confermandone pertanto l'estrema rarità. In particolare, tutte queste 5 duplicazioni mostravano una regione minima comune, corrispondente al gene NUP62CL, suggerendone una possibile significatività. In uno studio di espressione da noi effettuato si è inoltre osservata una significativa e marcata down-regolazione di NUP62CL. Pertanto, pur non avendone avuto prova definitiva, è sembrato che la CNV Xq22.2 giocasse un ruolo importante nella patogenesi della NCI nella nostra famiglia. Tuttavia, data la sua presenza anche in un membro non evidentemente affetto e considerata anche la severità del fenotipo NCI di alcuni membri del ramo paterno, è stato ipotizzato che la presenza aggiuntiva di altri determinanti genetici potesse essere responsabile del fenotipo estremo riscontrato nel probando. In quest'ottica, è stato sequenziato l'intero esoma del probando (*Whole Exome Sequencing*) al fine di individuare varianti puntiformi rare che potessero agire da ulteriori varianti di suscettibilità alla NCI nella famiglia caso studio.

La fase di analisi bioinformatica e validazione molecolare ha portato a delineare un quadro di eredità poligenica, con l'individuazione di quattro possibili geni di suscettibilità: XDH, ATP6V1B1, HNF1B, NUP62CL. Il gene XDH, fattore causativo della xanthinuria di tipo I, è stato evidenziato per la presenza di due varianti rare con un predetto effetto dannoso sulla struttura proteica. Il gene ATP6V1B1, di cui è già noto il ruolo nell'acidosi tubulare distale, presentava una variante missenso che sembrava render conto del fenotipo intermedio ipocitraturia osservato nella famiglia. E' stata individuata, inoltre, una variante estremamente rara nel gene HNF1B, un fattore causativo di nefropatia cistica e, concordemente, il probando recentemente aveva mostrato la presenza di cisti pre-caliceali. Infine, soprendentemente, è stato possibile evidenziare ancora una volta il gene NUP62CL per la presenza di due varianti omozigoti, che agiscono attivando dei siti criptici di splicing e quindi con un probabile effetto di produzione di un mRNA aberrante. Questo gene non è stato mai associato alla NCI ma, possedendo un ruolo noto nel trasporto dell'ossalato, potrebbe essere responsabile di alcuni dei meccanismi che sottendono alla patogenesi della NCI.

ABSTRACT

Idiopathic calcium nephrolithiasis (ICN) is a multifactorial disease with a pathogenesis depending upon the interplay of environmental, metabolic, anatomical and genetic factors. The importance of hereditary factors in ICN has emerged from a number of studies on families and twins affected by ICN. Computer programs predicted the best inheritance fit with a model of single gene co-dominant model/polygenic model.

Most candidate gene or association studies have produced, up to date, negative or only marginal results, leaving the genetic basis of ICN still an open question. But, genetic analysis techniques are rapidly evolving and promise to improve our knowledge of the genetic basis of nephrolithiasis and allied disorders.

Whole-genome scanning techniques, like *Next-Generation Sequencing* (NGS) and CGH (*Comparative Genomic Hybridization*) array, are the most recent technical approaches of molecular genetics that allow to trace out a complete draw of variation present in the human genome, namely SNPs (*Single Nucleotide Polymorphisms*) and heterogeneous structural variants. In the last years, indeed, several studies have recognized the critical role of structural genetic variants, mostly represented by *Copy-Number Variants* (CNVs), in modulating gene expression and complex diseases phenotype.

Aim of this project was to identify some of the major susceptibility genes involved in idiopathic calcium nephrolithiasis applying the potential of whole-genome scanning technologies to a case study very peculiar, both for clinical manifestations both for its genetic background.

The proband, since 30 years, suffers of a recurrent and severe form of ICN, with intermittent hypercalciuria and phosphaturic tubulopathy, expelling, up to date, about 300 calculi both spontaneously and by lithotripsy (33 treatments). Strikingly, he belongs to a family with consanguinity in which nephrolithiasis is present in several members and is transmitted in an apparently dominant fashion, with males being more severely affected than females.

CGH arrays has been performed in the proband and in eight members of the family in order to detect the presence of CNVs in the genome. A duplication of 308 kb region of the

Xq22.2 chromosome has been detected in the proband. This CNV resulted particularly interesting because it was a novel (not known) variant and it was shared by other three family members (maternal inheritance).

Moreover, matching the Xq22.2 duplication over a database of 14375 individuals only 5 overlapping duplications were found, confirming its extreme rarity. Strikingly, all these 5 duplication shared a common region that corresponded to NUP62CL gene, suggesting a likely importance of this genetic trait. Furthermore, performing an expression study, we observed a significant and marked down-regulation of NUP62CL.

Thus, even if we could not have a definitive proof of principle, this Xq22.2 CNV seemed to have an important role for ICN pathogenesis in our family. Nonetheless, given its presence in a member of the family that was not evidently affected and given the severity of the phenotype of some members of the paternal branch, we postulated that additional genetic determinants were also responsible of the severe phenotype encountered in the proband. In this light, we performed whole exome sequencing of the proband to identify rare single nucleotide variants that may have a role as susceptibility variants for ICN in the case-study family.

By bioinformatic analyses and molecular validation, we could draw a polygenic model of heredity, with the identification of four possible susceptibility genes: XDH, ATP6V1B1, HNF1B, NUP62CL. The XDH gene, normally responsible of xanthinuria type I, was highlighted for the presence of two very rare variants with a damaging predicted effect on protein. Next, we discovered a missense homozygous variant in ATP6V1B1 gene, whose role is already known in distal tubular acidosis, that seemed to account for the ICN intermediate phenotype hypocitraturia observed in the family. In addition, we found an extremely rare heterozygous variant in HNF1B gene, a causative factor in cystic nephropathy, and the proband, indeed, recently showed pre-calyceal cysts. Strikingly, we again evidenced NUP62CL gene for the presence of two homozygous variants that operate activating criptic splicing sites, thus eventually generating an aberrant mRNA. This gene has never been associated to ICN but, having a known role in oxalate transport, could be responsible of some of the biological mechanisms that are on the basis of ICN pathogenesis.

1. INTRODUCTION

1.1. NEPHROLITHIASIS

Nephrolithiasis is a common disease, typically occurring between 30 and 60 years of age and it represents the most chronic kidney condition after hypertension. In the majority of patients the symptoms and consequences are not life threatening, but stones in the urinary tract are a major cause of morbidity, hospitalisation and days lost from work (Saigal et al., 2005). In Italy for example, the number of patients receiving hospital treatment between 1988 and 1993 increased from 60 to 80,000 per year; among these, 14% (12,000 patients/year) required surgical treatment or urological manoeuvres and the number of extracorporeal shock wave lithotripsy sessions was approximately 50,000 per year (Gambaro et al., 2004).

Initiation and growth of stones is determined by supersaturation of urine with respect to a solute (calcium, oxalate, uric acid and cystine), that leads to a phase change in which dissolved salts condense into solids thus forming the stone.

Nephrolithiasis and the metabolic characteristics of urinary stones are heterogeneous: there are many types of renal stones, but the most common are calcium-containing stones making up 75% of all stones. Calcium oxalate (CaOx) is the predominant component of most stones (Worchester and Coe, 2008), often admixed with some calcium phosphate (CaP), which may form the initial nidus of the stone.

For CaOx stones, the most important determinants of urinary supersaturation are the total daily calcium excretion and urine volume, that is, the urine calcium concentration. In addition to solute concentration, urine pH constitutes a critically important determinant of solubility for CaP, uric acid and cystine (Coe et al., 2007).

However, urinary CaOx supersaturation is found frequently in normal subjects and this occurs, most likely, because of the presence of crystallization inhibitors in urine (citrate, pyrpophopshate, as well as at least a dozen proteins - like osteopontin - and glycosaminoglycans), which can impede the nucleation, growth, and aggregation of crystals

in vitro and have been shown experimentally to interfere with crystal attachment to renal epithelial cells (Kumar and Lieske, 2006).

The initial presentation of nephrolithiasis occurs often with renal colic - severe pain caused by stone passage- triggered by movement of a stone from the renal pelvis into the ureter, which leads to ureteral spasm and possibly obstruction. A radiograph of the kidneys, ureters and bladder often can visualize calcium-containing stones in the kidney or ureter, including struvite stones, but uric acid, purine and cystine stones are often visualized poorly. Stones smaller than 5 mm in diameter usually can pass spontaneously, but about 50% of stones larger than 5 mm require urologic intervention for removal, and those larger than10 mm are very unlikely to pass unaided.

Therapeutic attention has been directed more to the intervention on already formed stones, than to the prevention of stone formation, very often because the nature of stones is not threatened by any drug. There are several options of surgical treatment for the 10% to 20% of stones that fail to pass spontaneously (Auge and Preminger, 2002), and the appropriate chose depends on the size, the location, and the type of stone, or the presence of anatomic abnormalities or infections. Nevertheless every urologic treatment is made by invasive extra- and intra-corporeal manoeuvres. *Extracorporeal shock wave lithotripsy* (*ESWL*) uses sound waves to fragment stones into small pieces that can be passed easily, but is effective for most stones smaller than 2 cm and phosphate and cystine stones appears to be resistant to fragmentation. Larger stones can be removed by percutaneous lithotripsy, through a small flank incision that allows direct visualization, stone disruption and removal of fragments.

It is clear that all these urologic measures have no effects over the incidence and worsening of nephrolithiasis, in some cases being even cause of new colic episodes due to the permanence of stone fragments in the urinary tract (Fine et al, 1995), thus representing no ideal treatments of nephrolithiasis.

Therefore, it appears more and more necessary to understand deeply basic metabolic-physiologic mechanisms that are causative factors of nephrolithiasis pathogenesis and eventually develop new preventive and targeted therapies.

1.1.1. Renal pathology in stone formers

Retention of crystals within the kidney is necessary for stone formation and it occurs with several patterns of deposition in kidney of stone formers, and these patterns are associated with specific stone types. Patients who have idiopathic CaOx stones have white deposits on their papillae, called "Randall's plaque" (Evan AP, Lingeman JE, Coe FL et al., 2003). Biopsies of these areas reveals interstitial deposits of CaP, in the form of biologic apatite, which starts in the basement membrane of the thin loops of Henle and contain layers of protein matrix. Deposits can extend down to the tip off the papilla, and if the overlying urothelium is denuded, the exposed plaque can become an attachment site for stones (Evan P, Coe FL, Lingeman JE et al., 2007). Stones seem to start as deposits of amorphous CaP overlying the exposed plaque, interspersed with urinary proteins. With time, more layers of protein and mineral are deposited, and the mineral phase becomes predominantly CaOx (Fig. 1).

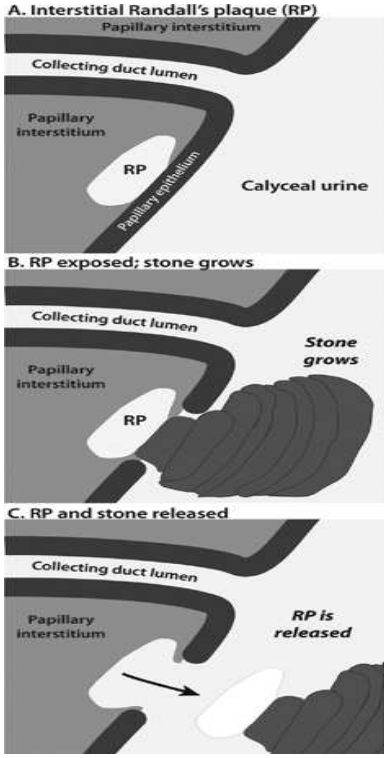


Figure 1. Concept of stone growing on Randall's plaque (RP).
A) The plaque appears in the interstitial tissue within the renal papilla, with no crystals present in any tubular lumens. The plaque is composed of calcium phosphate in the mineral form of apatite. **B)** Papillary epithelium is lost, and the plaque is uncovered so that its surface is bathed with urine in the renal calyx. The resulting calcium oxalate stone may grow only very slowly, and can remain small for an extended period of time, as the plaque keeps the stone from flowing out with the urine, and the insolubility of the calcium oxalate makes the stone quite stable with time. The layers in the schematic represent hypothetical growth periods for the stone, periods which could be separated by weeks or even months. **C)** Some stones that are formed on Randall's plaques are released from the papilla in which they formed, by a process that is not known

By contrast, in patients with stone constitution mainly of CaP (apatite or brushite) stones are not attached to plaque. Instead, many collecting ducts are filled with crystal deposits made of apatite that fill the tubule lumen and may protrude from the mouths of the ducts of Bellini. Overall, most stone formers studied so far have crystal deposits in the medullary collecting ducts, except for those who have idiopathic CaOx stones, that have no intratubular deposits but instead have abundant deposits of apatite in the papillary interstitium.

Strikingly, if the manifestation of urinary stones is due to multiple different underlying pathologies, then studying patients grouped together simply under the category of "stone former" will certainly result in confusion. Even specifying patients under the term "calcium stone former" can lump together pathologies that appear to be completely separated in primary cause of the disease.

In conclusion, it appears necessary to integrate pathology mechanisms in the general picture of every stone former patient, thus enabling to define a clearer phenotype that results essential to further basic biology and genetic studies.

1.1.2. Risk factors

Environmental factors, especially diet, play an important role in expression of the tendency of stone formation. Evidence of environmental influence includes the different distribution of the disease in wealthier as opposed to poorer areas, its constant increase in industrialized countries during the last century (an increase only temporarily interrupted by World Wars I and II), and the recently reported increasing prevalence of nephrolithiasis in younger women (Robertson et al., 1999). The reasons for increasing prevalence are not clear but, for instance, mass index and waist circumference seem to play a role, especially in women (Taylor et al., 2005).

However, also inherited factors play a role in stone formation. In fact, under the same environmental conditions, only a few individuals tend to develop stone disease (incidence

estimated at ~1%/y; prevalence, ≥10%), and these subjects frequently have family members affected by the disease.

Nephrolithiasis, indeed, occurs more frequently in individuals with a positive family history of renal stones. Large-scale epidemiological surveys have shown a familial aggregation for kidney stones, with a greater risk for first-degree relatives (McGeown, 1960; Resnick et al., 1968; Robertson et al., 1983).

For example, as regards patients with idiopathic calcium nephrolithiasis, 16 to 50% of have a positive family history of nephrolithiasis (Mehes and Szelid, 1980; Coe et al., 1979). In the United States, a family history of nephrolithiasis was reported in 16% and 17% of stone patients by Resnick et al (1968) and Curhan et al. (1997), respectively. A much greater prevalence has been observed in Italy. Trinchieri et al. (1988) observed a familial aggregation in 37% of cases, and we reported that as many as 38% of non-recurrent and 50% of recurrent stone formers have a positive family history of nephrolithiasis (Gambaro et al., 1996). In 71 Canadian families from a genetically and geographically homogeneous population, hypercalciuria segregates with stone formation (Tessier et al., 2001).

The importance of hereditary factors in idiopathic calcium nephrolithiasis has emerged from a number of studies of families containing members affected by renal stone disease and from twin studies. Computer programs predicted the best inheritance fits with a model of single gene co-dominant model/polygenic model. Both models gave a heritability scores of 58%. As a whole, genes might determine over 50% of the urinary calcium excretion rate.

The role of inheritance is clearest in the monogenic forms of nephrolithiasis such as cystinuria, Dent's disease, and primary hyperoxaluria (Coe et al., 2005). There is a clear familial tendency in idiopathic stone formation as well (Goldfarb et al., 2005), although the genes involved are currently unknown.

In the last decade important additions have been made to our knowledge by Bushinsky's animal model (Bushinsky, 1999), the important theoretical contribution from Goodman et al. (Goodman et al., 1995) on idiopathic calcium nephrolithiasis as a polygenic disorder, and the studies of Canadian stone-former sib-ships by the Bonnardeaux group (Scott et al., 1998).

1.1.3. Idiopathic Calcium Nephrolithiasis

Calcium stones are the most common in *idiopathic calcium nephrolithiasis* (ICN), and they are associated with a number of metabolic derangements, the most common is hypercalciuria. Elevate urine calcium excretion, indeed, is the most common abnormality found in both adults and children who form kidney stones: 30% to 60% of adult and paediatric stone formers have hypercalciuria. Many monogenic diseases are associated with hypercalciuria, but the majority of cases in stone formers are caused by either idiopathic *hypercalciuria* (*IH*) or *primary hyperparathyroidism* (*PHPT*).

The term "idiopathic hypercalciuria" is generally applied to cases in which serum calcium is normal and other causes of increased calcium excretion, such as vitamin D excess, renal tubular acidosis, granulomatous disease such as sarcoid, steroid use and hyperthyroidism have been excluded.

IH can involve normally calcium handling by gut, kidney and bone. Patients with IH very often have elevated serum 1,25.dihydroxy vitamin D levels and increased intestinal calcium adsorption (Coe et al., 2004) . Their kidneys exhibit a decreased ability to reabsorb filtered calcium. A decreased renal reabsorption of phosphate may also occur, and serum phosphate levels are low in patients with IH.

For idiopathic hypercalciuric who have recurrent stones, thiazide diuretics, which can lower urine calcium, are the treatment of choice (Borghi et al., 1993).

Low urinary citrate excretion may occur in a large fraction of stone formers as a consequence of acidosis or potassium depletion or in case of an idiopathic disorder; it is frequently associated with other metabolic disorders that increase stone risk. Citrate can inhibit stone formation for its ability to chelate calcium, forming a soluble complex that prevents calcium binding with oxalate or phosphate. In addition, citrate can act on the surface of preformed CaOX or CaP crystals as a growth inhibitor.

Citrate has also been used as a treatment for idiopathic calcium stone formers, especially in those who have low urine citrate levels (Barcelo et al., 1993).

Hyperoxaluria is a quite common metabolic condition among stone formers and, apart from genetic primary hyperoxaluria, may be caused by increased oxalate absorption triggered by a low calcium diet (Asplin, 2002).

Patients who forms CaOx stones may have also elevetaed uric acid excretion, and hyperuricosuria decreases solubility of CaOx and promotes stones. Patients who have hyperuricosuric calcium stones differ from patients who have gout and uric acid stones because they have higher urine pH and a higher urine uric acid level as well.

Most calcium stones are composed predominantly of CaOx with small amounts of admixted CaP. Normally, stones containing more than 50% CaP are uncommon and form when urinary CaP supersaturation is persistently elevated. The major determinants of CaP supersaturation are alkaline urine (pH>6.3) combined with hypercalciuria (Parks et al., 2004). This condition is seen in patients who have distal renal tubular acidosis, whether genetic or acquired, but most patients who have CaP stones do not have metabolic acidosis, and the cause of their persistently alkaline pH is unclear.

1.1.4. Genetics of idiopathic calcium nephrolithiasis

During the past years it appeared more and more clearly that genetic factors are important determinants of the individual's risk of developing kidney stones. As many as 40% of hypercalciuric stone formers have at least one first-degree relative with nephrolithiasis (Pak, 1979). Men form stones three times more than women, and men with kidney stones are three time more likely to have a parent or sibling with a history of kidney stones (Curhan et al., 1997). This indicates that, although as previously discussed shared dietary or environmental factors may contribute, the principal determinants for these associations are probably genetic in origin.

Nevertheless, the genetics of ICN is very often complicated by a number of issues. Primarily, some patients may inherit a predisposition to an abnormality, but the abnormality itself may not be manifest at all times. This can therefore make it very difficult

to assign phenotype properly when conducting family studies and it obligates to make observations under standardized conditions. A second complicating issue is that many patients form stones in the setting of several metabolic abnormalities. Thus, for example, a particular patient with calcium stones may have hypercalciuria, hyperoxaluria, and hypocitraturia, whereas other members of the family may share some but not all of the same risk factors.

A third complicating issue is that, with the exception of rare monogenic diseases, some risk factors for stones may reflect not a single gene but rather the interacting effects of multiple genes.

Idiopathic hypercalciuria, indeed, occurs 40% or more in patients that have a family history of nephrolithiasis. Eventhough several studies have presented families in which hypercalciuria appears to be inherited as an autosomal trait (Coe et al., 1979; Mehes et al., 1980), there are several reasons to conclude that hypercalciuria is not a monogenic but rather a quantitative trait.

In the nineties, studies of gene mapping and positional cloning have identified a candidate gene for hypercalciuria, CLC-5 chloride channel, that is mutated in X-linked nephrolithiasis or Dent's disease, characterized generally by LMW (*low molecular weight*) proteinuria, low level of parathyroid hormone (PTH), and hypercalciuria, and very often by nephrocalcinosis and nephrolithiasis.

CLC-Kb, another channel of chloride channel family and ROMK (KCNJ1), a potassium voltage gated channel, have been already identified as responsible of Bartter syndrome, another hypercalciuric condition (Simon DB et al., 1997).

From a number of candidate gene, association and linkage studies, it has been demonstrated that polymorphisms in a variable number of candidate genes could be associated with higher calcium excretion and/or nephrolithiasis, among them VDR and CaSR genes (Pearce et al., 1996; Scott et al., 1999) (Table1).

Nevertheless, with the expanding scenario of novel molecular genetic techniques, it is reasonable to expect exciting advances in our understanding of the genetics of ICN.

Renal stone disease[a]	Mode of inheritance[b]	Gene[c]	Human chromosomal location (mouse)
Associated with hypercalciuria			
	A-d	*SAC*	1q23.3-q24
	A-d	*VDR*	12q12-q14
	A-d	?	9q33.2-q34.2
ADHH	A-d	*CASR*	3q21.1
Hypercalcemia with hypercalciuria	A-d	*CASR*	3q21.1
Bartter syndromes			
Type I	A-r	*SLC12A1/ NKCC2*	15q15-q21.1
Type II	A-r	*KCNJ1 /ROMK*	11q24
Type III	A-r	*CLCNKB*	1q36
Type IV	A-r	*BSND*	1q31
Type V	A-d	*CASR*	3q21.1
Type VI	X-r	*CLCN5*	Xp11.22
Dent's disease	X-r	*CLCN5*	Xp11.22
Lowe syndrome	X-r	*OCRL1*	Xq25
HHRH	A-r	*NPT2c/SLC34A3*	9q34
Nephrolithiasis, osteoporosis and hypophosphatemia	A-d	*NPT2a/SLC34A1*	5q35
Familial hypomagnesemia with hypercalciuria and nephrocalcinosis	A-r	*PCLN1/CLDN16*	3q28
Familial hypomagnesemia with hypercalciuria and nephrocalcinosis with ocular abnormalities	A-r	*CLDN19*	1p34.2
dRTA	A-d	*SLC4A1/kAE1*	17q21.31
dRTA with sensorineural deafness	A-r	*ATP6B1/ ATP6V1B1*	2p13
dRTA with preserved hearing	A-r	*ATP6N1B/ ATP6V0A4*	7q34
Not associated with hypercalciuria			
Primary hyperoxaluria type 1	A-r	*AGXT*	2q37.3
Primary hyperoxaluria type 2	A-r	*GRHPR*	9p13.2
APRT deficiency	A-r	*APRT*	16q24.3
Cystinuria type A	A-r	*SLC3A1*	2p16.3
Cystinuria type B	A-r	*SLC7A9*	19q13.1
Wilson's disease	A-r	*ATP7B*	13q14.3

Table 1. Genetic defects related to monogenic forms renal stone disease associated or not associated to hypercalciuria (Image by *Stechman et al., 2009*).

1.2 MOLECULAR GENETICS: COMPLEXITY OF HUMAN GENOME AND COMPLEX GENETIC DISEASES.

The approach to molecular genetic studies of complex diseases evolved considerably during the recent years. In particular the candidate gene approach, restricted normally to an analysis of a few single-nucleotide polymorphisms, has been supplanted by the unbiased approach of genome-wide studies that exploit the enormous potential of *whole-genome*

scanning techniques. To better understand the extreme complexity of human genetics and complex genetic diseases and the fine relationship that exists between them, it is firstly necessary to clarify some new genetic concepts. Whole genome scanning techniques, indeed, revealed extreme variability in humans and allowed to discover new forms of genetic variation

1.2.1 Single-Nucleotide Polymorphims

A *single-nucleotide polymorphism* (*SNP*) is a DNA sequence variation occurring when a single nucleotide — A, T, C or G — in the genome differs between members of a biological species or paired chromosomes in a human.

Humans are genetically diverse and they differ in approximately 0.1% of their genomes. The single-nucleotide polymorphism database (dbSNP) lists more than 37 million variants among humans. With the exception of identical twins, no two humans have identical genomes. Every genome contains approximately 4 million of *DNA sequence variants* (*DSVs*) that affect half of the genes in each genome collectively, and many are private (Pennisi E, 2010; Wheeler et al., 2008).

Among the approximately 3.5 million SNPs in each genome, approximately 10,000 SNPs are *nonsynonymous SNPs* (*nsSNPs),* of which approximately two thirds are predicted by *in silico* analysis to impart potentially damaging effects. In addition, each genome contains approximately 50 to 100 variants that have been associated with inherited disorders and approximately 30 *de novo* variants, whose presence is indicative of continuous introduction of new variants to the genetic pool (Durbin et al., 2010).

1.2.2 Structural variations

Most DSVs in the genome are SNPs, but a big portion of variants is constituted by *structural variations* (*SVs*). They affect more nucleotides in the genome, simply because of their sizes and the first identified were small deletions, insertions and duplications (<1kb).
Genome-scanning array technologies and comparative DNA-sequence analyses have begun to reveal DNA variations that involve segments that are smaller than those recognized microscopically (e.g. aneuploidies, chromosomal rearrangements, heteromorphisms, fragile sites) but larger than those that are readily detected by conventional sequence analysis (e.g. SNPs, micro- and mini-satellites) . These variants, which range from ~1 kb to 3 Mb in size, are normally refererred as *submicroscopic structural variants*.
In the last decade rearrangements of larger segments of DNA (>1 kb) were discovered through *CGH* (*Comparative Genomic Hybridization*) *array* technology, with the contemporary understanding that they might increase or decrease the copy number of the genes from the natural two copies (Mills RE et al., 2011; Eichler EE et al., 2010). Such SVs are referred to as *Copy Number Variants* (*CNVs*).

1.2.2.1 Copy-Number Variants: meaning and identification techniques

The DNA encodes around 23,500 genes. It was generally thought that genes were almost always present in two copies in a genome. However, recent discoveries have revealed that large segments of DNA, ranging in size from thousands to millions of DNA bases, can vary in copy-number. In particular, a Copy- Number Variants represents a segment of DNA that is 1 kb or larger and is present at variable copy number in comparison with a reference genome. Classes of CNVs include insertions, deletions and duplications. This definition also includes *large-scale copy number variants*, which are variants that involve segments of DNA ≥50 kb (*LCV*).

Such copy number variations can encompass genes leading to dosage imbalances. For example, genes that were thought to always occur in two copies per genome have now been found to sometimes be present in one, three, or more than three copies. In a few rare instances the genes are missing altogether (Fig. 2). Thus, the new findings seems to indicate that our DNA is less than 99.9% identical, as was previously thought.

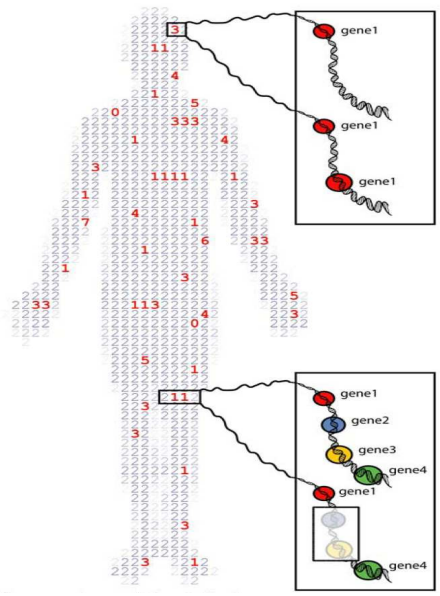


Figure 2. Copy number variation in the human genome. The 30,000 genes are usually present in two copies. A new map of the genome has been unveiled that catalogues DNA and genes variable in copy number (those numbers others that 2 are highlighted in red). Duplication of a gene (top) and deletion of two genes (bottom) are depicted. (Image by http://www.sickkids.ca/mediaroom/custom/genomevariation06.asp)

Studies over the past six years have resulted in increasing recognition of the critical role of structural genetic variation (most of which appear to be in the form of copy number variation) in modulating gene expression and disease phenotype. CNVs, indeed, are now known to be a prevalent form of common genetic variation and represent a substantial proportion of total genetic variability in human populations.

Moreover, association studies have already demonstrated the importance of CNVs as disease-susceptibility variants, with specific CNVs found to confer differential risk to HIV infections, autoimmune disease, and asthma (Gonzalez et al., 2005; Fanciulli M et al., 2007; Brasch-Andersen C. et al., 2004). Recently, genome-wide surveys have demonstrated that rare CNVs altering genes in neurodevelopmental pathways are implicated in several neuropsychological disorders, like autism and schizophrenia (Sebat J,

et al. 2007; Walsh T, et al., 2008). It is therefore becoming increasingly clear that genetic studies of complex diseases must pay closer attention to the contribution of CNVs.

Several distinguishing features of CNVs support their role in disease pathogenesis. First, though less abundant than SNPs, it has been suggested that CNVs account for more nucleotide variation than do SNPs because of their size. By spanning thousands of bases, CNVs often encompass (and can sometimes disrupt) functional DNA sequences. Second, there appears to be an enrichment of currently-known CNVs toward "environmental sensor" genes – i.e. genes that are not necessarily critical for early embryonic development, but rather help us to perceive and interact successfully with our ever-changing environment (Sebat J et al., Science 2004). This includes enrichment for olfactory receptors, immune and inflammatory response genes, cell signaling and cell adhesion molecules, structural proteins, and ion channels. Third, a recent comparison of the relative impact of SNPs and CNVs on gene expression noted that a substantial proportion (~18%) of gene expression variability was attributable to known CNVs greater than ~40 kb in size (Stranger BE et al., 2007). Lastly, like other forms of genetic variation, both purifying and adaptive natural selective pressures appear to have influenced the frequency distribution of selective CNVs, suggesting their functional significance (Redon R et al., 2006).

Currently, the main approaches for identifying unbalanced structural variants are array-based analyses and quantitative primarily PCR-based assays. *Array-based comparative genome hybridization* (*array-CGH*) approaches (Pinkel D et al., 1998) provide the most robust methods for carrying out genome-wide scans to find novel CNVs. These approaches use labeled fragments from a genome of interest, which are competitively hybridized with a second differentially labeled genome to arrays that are spotted with cloned DNA fragments, revealing copy-number differences between the two genomes (Fig. 3).

Nevertheless, the most robust assays for screening targeted regions of the genome are mainly PCR-based, whose the best established is *real-time quantitative PCR* (Bieche I et al., 1998). However, although most protocols for this method work well for detecting individual deletions and duplications, they are generally not suitable for multiplexing.

Structural variants can also be identified *in silico* by comparing DNA sequences from different sources. In the simplest approach, two assemblies from unique human DNA

sources are aligned to detect differences (Tuzun et al, 2005). One advantage of this method is that all types of variant, including balanced variants, can be detected. In addition, there is no limit to the resolution, and the variants that are identified can be defined at the nucleotide level (Feuk et al., 2005)

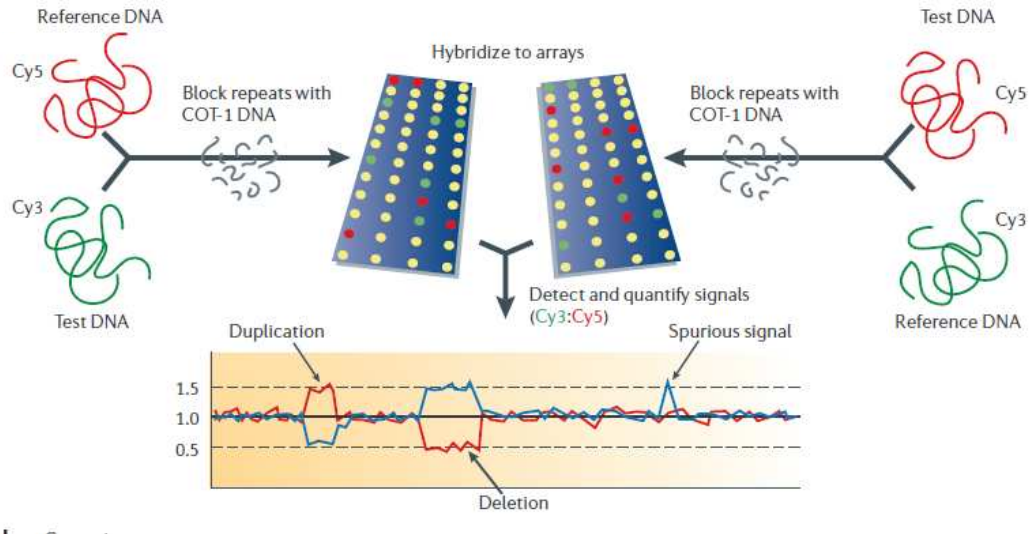


Figure 3. Array-CGH for the identification of copy-number variants. In array based comparative genome hybridization (array-CGH), reference and test DNA samples are differentially labeled with fluorescent tags (Cy5 and Cy3, respectively), and are then hybridized to genomic arrays after repetitive-element binding is blocked using COT-1 DNA. The array can be spotted with one of several DNA sources, including BAC clones, PCR fragments or oligonucleotides. After hybridization, the fluorescence ratio (Cy3:Cy5) is determined, which reveals copy-number differences between the two DNA samples. Typically, array-CGH is carried out using a 'dye-swap' method, in which the initial labelling of the reference and test DNA samples is reversed for a second hybridization (indicated by the left and right sides of the panel). This detects spurious signals for which the reciprocal ratio is not observed. (Image by Feuk et al., 2006)

The current map of structural variation in the human genome is far from being complete. While several databases exist to catalog this newly-appreciated form of human genetic variation (notably the *Database of Genomic Variants* - http://projects.tcag.ca/variation/ and the *Human Structural Variation* database - http://humanparalogy.gs.washington.edu/structuralvariation/), quality control is lacking,

and studies have differed in technological approaches, precise boundary definition of CNVs, DNA quality, and even discrepancies in terminology (Scherer SW et al., 2007). Nevertheless, emerging technologies are more sensitive for detection of CNVs and provide more precise definition of boundaries (Perry et al., 2008). Undoubtedly, as a clearer map of human structural genetic variation emerges, we will begin to more comprehensively include this type of genetic variation in genome-wide studies that attempt to elucidate the role of CNVs in human disease.

1.2.3 Molecular genetics of complex diseases

The plethora of DSVs in the genome and the multilayer regulation of gene expression and function are indicative of the intricacy of the determinants of the complex diseases and phenotypes. The clinical phenotypes are presumed to result from the additive effects and interactions among multiple causative alleles with various genomics and environmental factors. In a complex phenotype, the effect sizes of the involved alleles are expected to vary and to follow a gradient that ranges from minimal or indiscernible to large and significant effect size (Fig. 4). Only a few alleles are expected to impart large effect sizes and, hence, could be detected by the commonly used approaches to genetic studies of complex phenotypes. Many are expected to exert modest effects that per se might not be discernible based on the usual phenotyping and genetic approaches. A complex trait results from confluence of various genetic and non genetic determinants. Genetic factors are major determinants of the complex phenotype and this notion is supported by heritability of the complex trait (Barabasi et al., 2011). By influencing gene expression, protein structure, and function, DSVs could impact various interacting networks that together influence susceptibility to a complex phenotype and account for the genetic component of a complex trait etiology (Barabasi et al., 2010).

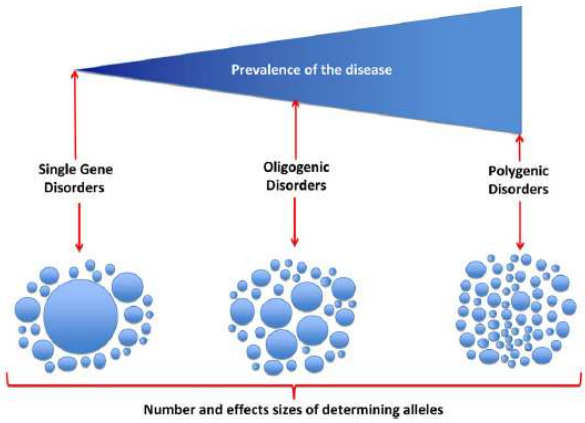


Figure 4. Gradients of disease prevalence and the effect sizes of the causative alleles. The prevalence of disease, number of determinant DSVs and their effect sizes are shown. Single-gene disorders are caused by rare variants with large effect sizes. In addition to the main causal variant, which typically exhibits a Mendelian pattern of inheritance, several other non-Mendelian variants contribute to expression of the phenotype. On the opposite end of the spectrum are the common complex traits, which are caused, partly, by the cumulative effects of a large number of DSVs, each imparting a modest effect size. In oligogenetic phenotypes, several alleles with moderate size effects and a large number of alleles with small effect sizes contribute to the phenotype (Image by *Marian AJ, 2012*).

The contributing DSVs individually might be neither necessary nor sufficient to alter the susceptibility to a complex phenotype. This is in contrast to single-gene disorders, where the presence of the mutation typically indicates the expression of disease phenotype, even if with variable penetrance and expressivity. Likewise, the effect size of each contributing variant is typically small and often negligible. In general, the effect sizes of DSVs on phenotypes are expected to follow a gradient, being the largest for those phenotypes that are influenced directly by genes, such as messenger RNAs (mRNAs) and proteins. These phenotypes are referred to as *proximal phenotypes* or *endophenotypes*. In contrast, the effect sizes of DSVs are expected to be smaller for phenotypes that are not influenced directly by DSVs, such as clinical outcomes. Such phenotypes are referred to as *distant phenotypes* (Barabasi et al., 2011). The gradient of effect sizes in part relates to the number of competing genetic and non genetic factors that contribute to the phenotype. For the proximal phenotypes, such as mRNA and protein levels, fewer determining factors are expected to contribute to the phenotype, thus each determinant might have a considerable effect size. In contrast, for the distant phenotype, such as mortality, many competing

genetic and non genetic factors contribute to the phenotype and dilute the effect size of an individual DSV (Kathiresan et al., 2008; Debette et al., 2011).

1.2.4 Genetic approaches to complex diseases

The full spectrum of allele frequency in a population is expected to follow a gradient ranging from private to extremely common alleles. Conventionally, however, the variants are categorized into 3 classes based on their *minor allele frequencies* (*MAFs*) in the population. Common and rare variants are those that have population MAFs of >5% and <1%, respectively. Variants that have population MAFs from 1% to 5% are considered uncommon or infrequent (Marian AJ, 2009). As observed for the genetic causes of Mendelian diseases, DSVs with large effects are expected to be rare. However, the converse is not the rule as most rare variants could have weak or no discernible clinical or biologic effects. Likewise, most common variants are expected to exert minimal or clinically indiscernible effect sizes.

Genetic studies of complex diseases might be instigated by a priori knowledge of potential involvement of a gene in the pathogenesis of the phenotype, which is referred to as the **candidate gene approach** (Marian AJ et al., 2011). An alternative is an **unbiased approach**, taking in consideration many genes and variants and typically the entire genome, as in *GWAS* (*Genome Wide Association Studies*) or *NGS* (*Next Generation Sequencing*), to identify the associated alleles. Genetic studies of complex phenotypes are designed typically on either **common disease–common variant** (**CD-CV**) or **rare variant–common disease** (**RV-CD**) hypotheses (Risch et Merikangas, 1996).

The former surmises that complex phenotypes results of cumulative effects of a large number of common variants, each exerting a modest effect. In contrast, the latter posits that multiple rare variants with large effect sizes are the main determinants of heritability of the complex phenotypes. Given that the anticipated effect sizes of the alleles in the genome are

likely to be a continuum, one might expect a combination of rare, uncommon, and common alleles to contribute to heritability of the complex diseases. In a given population, however, common alleles, despite having modest effects, might have a greater attributable fraction because of their number, even though each rare allele might have a larger effect size (Pritchard and Cox, 2002).

1.3 NEXT-GENERATION SEQUENCING: ORIGIN, RESOURCES AND PITFALLS

DNA sequencing techniques have revolutionized our understanding of human biology over the last forty years. They originated in the early 70s due to the work of Walter Gilbert and Frederick Sanger (Maxam and Gilbert, 1977; Sanger and Coulson, 1975). Continuous technological improvements in DNA sequencing instrumentation ever since has created an environment in which the *Human Genome Project* (*HGP*) (Watson JD, 1990) could be finally realized in the year 2001 after a decade of work. The HGP provided the initial draft of mankind's DNA sequence by 23 collaborating laboratories using *Sanger sequencing* of mapped regions as well as *shotgun sequencing* techniques in a process that had a cost ~$3 billion. The HGP was expected to provide mankind with a dramatic advance in our understanding of human health and generate a revolution in personalized healthcare approaches.

Today, *Next Generation Sequencing* (*NGS*) techniques represent the next phase in the evolution of DNA sequencing technology at dramatically reduced cost and time compared to traditional Sanger sequencing (Zhou X et al., 2010).

Several NGS technologies have been developed using diverse approaches since 2001, each with its own distinctive strengths and weaknesses. The major commercial entities which came into existence after the success of the HGP include *454 sequencing* (http://www.my454.com/), *Solexa/Illumina* (http://www.illumina.com), *SOLiD* (http://www.appliedbiosystems.com), and *Polonator* (http://www.polonator.org/).

The mapping software then attempts to “map” the individual sequence NGS “reads” onto a reference genome sequence available from online genome databases. This process is known as *reference mapping.* Alternatively, when the reference genome is unknown *apriori,* the individual “*read*” fragments are linked to each other by overlapping the common sequences at the ends of each read to form a longer, much complete version of the genome under study. This is known as *denovo mapping* of DNA sequence. For the purposes of clinical sequencing, reference mapping is performed most of the time due to our pre-existing knowledge of the human reference genome sequence (Mardis ER, 2008).

The first step of NGS data processing is the alignment of reads obtained followed by assembly of the genome of the patient sample. Once the alignment process is completed, downstream bioinformatics analysis is performed to detect the clinically relevant structural genomic alterations. Different software programs are designed to detect different kinds of genetic variants: SNPs are the most reliably detected variants in the genome and the most abundant; other genomic variants such as *indels* are equally detected routinely. NGS technology also can identify structural variations, like CNVs, in the genome, although routine alignment tools are not suited to perform such analysis since they cannot identify more than a few nucleotide mismatches. Thus, specialized software for analyzing indels from *paired-end reads* are being developed which enables identification of structural variants by identifying the flanking end regions of the NGS read data (Medved et al., 2010.

NGS technology supports both *whole genome* and *whole exome sequencing (WGS* and *WES*), this last indicating a technique to selectively capture and sequence the coding regions of all annotated protein-coding genes. The genome comprises approximately 180,000 exons that reside in approximately 23,500 genes, therefore, collectively, the coding regions encompass about 1% of the genome (Lander et al., 2001). WES requires generating a genomic DNA library, capture and enrichment of all exons, and sequencing using a NGS platform.

Despite the high accuracy of base calling by NGS platforms, because of the enormity of the data output, the error rate could represent a significant challenge for an accurate identification of the variants, particularly for heterozygous alleles or rare variants in a small fraction of DNA templates (Ansorge WJ et al.,2009). An important aspect is the *coverage rate* of each allele.

In DNA sequencing by the NGS platforms, multiple fragments of DNA are sequenced simultaneously, and the outputs are aligned to the reference genome. Thus, it is essential that both strands of a diploid genome are represented adequately in the sequence read out. In addition, multiple reads of an allele increases the signal to-noise ratio. Therefore, in low-coverage sequencing, which is more practical because of size of the sequence output, less costly, and computationally less laborious, accurate determination of the genotype is less certain. In contrast, a higher coverage at each nucleotide increases the confidence in accurate allele calling. The coverage rate, however, is often non-homogenous and certain genomic regions or exons might not be covered at sufficient depth to provide for accurate allele calling (Nielsen et al., 2011).

Inadequateness of the sequence reads, in part, might reflect inadequate coverage by the capture probes. Nevertheless, the gaps could lead to inadequate detection of the variants as well as miscalling because of a poor signal-to-noise ratio. Different depth of coverage might be necessary according to intended application of the NGS data.

For instance, in NGS studies in families to detect a causative allele in an autosomal dominant Mendelian disorder, a higher depth of coverage might prove essential to detect robustly or exclude the presence of a heterozygous mutation. In addition, the relatively small number of family members makes high-depth coverage feasible. In contrast, in allelic association studies of complex phenotypes, wherein normally sequencing of many cases and controls is required, low-coverage is practical and probably a more powerful approach than covering a smaller number of individuals at a greater depth.

1.3.1 Whole Exome Sequencing: an unbiased approach to study complex genetic diseases

Evidently, NGS technology high sequencing accuracy has become of relevant importance in the identification of genetic variants that may cause complex inherited diseases.

The discovery of genetic traits that are at the basis of complex diseases, indeed, has been greatly simplified by NGS, apart from the significantly reduced time frame from years to

weeks, in the effective identification of rare variants present in the whole genome that could exert large effect sizes on the phenotype (*RV-CD hypothesis).*

WES it is probably the most commonly used direct DNA sequencing approach today to identify the genetic causes of rare Mendelian and common non-Mendelian disorders, because of practical reasons including cost, data storage, and bioinformatics analyses.

The rationale for *whole exome sequencing* is based on the notion that variants located in exons and affect protein sequence are more likely to be pathogenic than those located in introns or intergene regions. Thus, the approach posits primarily that infrequent and rare nonsense, frame shift, and nonsynonymous DSVs are likely to play major etiologic roles in susceptibility to complex phenotypes.

The whole-exome sequencing approach has been applied successfully to identify of the causal mutations for rare Mendelian disorders, such as Freeman-Sheldon syndrome, congenital chloride-loosing enteropathy, Kabuki syndromes, systemic hypertension caused by hyperaldosteronism, and others (Ng SB et al.,2010; Choi M et al., 2011).

Exome sequencing, likewise, revealed the complexity of discerning the pathogenic alleles based on sequence data alone, as many apparently pathogenic variants might be found in clinically unaffected individuals (Klassen T et al., 2011).

Efforts are ongoing to apply the whole-exome sequencing approach to delineate the genetic causes of common forms of Mendelian diseases, particularly those with an autosomal dominant pattern of inheritance, and complex traits. In addition to limitations imposed by family size and structure, which could restrict discerning co-segregation of the variants with the phenotype, several other limitations render the approach challenging. The enormous genetic diversity of the humans and the presence of many variants in each genome in conjunction with incomplete penetrance of the causative variants pose significant difficulties in establishing a clear genotype-phenotype co-segregation.

But, if on one side identification of a vast number of rare variants in coding and regulatory regions constitutes a successful starting-point to dissect complex diseases, on the other side the necessity to effectively analyze the huge amount of WES data and to relate them to complex phenotypes is still an open field.

Interpretation of exome information derived from NGS is still a prickly question. The initial approach to establish causality connection is to focus on protein-changing variants,

inferring biologic and functional significance by *PolyPhen-2* (*National Institutes of Health, Bethesda, Md*) and *SIFT* (*J. Craig Venter Institute, Md*) (Adzhubei IA et al., 2010; Kumar P et al., 2009) especially those that are rare (low MAF values). Usually 150-1500 private variants in the patient are identified as potentially disease causing. Subsequently, there are six strategies that could be applied to prioritize these variants (Oetting WS, 2012). The choice of the strategy requires knowledge of the inheritance of the disorder, genetic heterogeneity, and availability of family members. The *linkage strategy* prioritizes variants that segregate with the disease or lie within a region that segregate with the disease. Whereas, in the *homozygosity strategy* known consanguinity can help identifying mutations associated with a recessive disease. The *double-hit strategy* is instead mainly used for recessive diseases, by selecting the variants that are homozygous or compound heterozygous, and in some cases a single exome can be sufficient to identify the gene associated with the disease. In the *overlap strategy* a well-defined phenotype is the starting point and by sequencing multiple unrelated patients with the same phenotype. The *de novo strategy*, in which a patient and his parents are sequenced, allows the identification of *de novo* mutations. Finally, the *candidate strategy*, by using biological information at the variant and gene level, can help in finding disease-causing mutations.

Anyway, there are some potential problems that can limit success: a lack of sequence coverage, resulting in the mutation not being targeted; misalignments of reads or miscalling of variants; misinterpretation of the variants; and the over cited clinical and/or genetic heterogeneity.

Notwithstanding these limitations, there is an estimated success rate of 60% and WES is expected to offer and open a full spectrum of genetic determinants of the phenotype in Mendelian and complex diseases (Oetting WS, 2012).

2. AIM OF THE STUDY

Few monogenic diseases associated with hypercalciuria, renal calcification, and calcium nephrolithiasis have been identified so far, and the few genetics investigations by genome wide or candidate gene association studies of ICN have produced contradictory results; thus, the issue of the genetics of ICN is still awaiting a solution.

Discovering the genes involved in ICN hopefully will lead to breakthroughs in pharmacological targets for treating or preventing these conditions, perhaps to tools for diagnosing the risk of developing them, and, finally, but no less importantly, to basic science discoveries and a better understanding of renal/intestinal/bone physiology relevant to ICN.

As emerged clearly in the last decade, the approach to molecular genetics has evolved considerably, pointing out whole-genome scanning technologies as priority tools to investigate genetic basis of complex genetic diseases.

The choice of the best strategy in a successful genetic analysis requires knowledge of the inheritance of the disorder, genetic heterogeneity, and availability of family members.

Thus, we proposed to study a case-study family in which ICN seems to be mostly an hereditary trait and clinical and familiar characteristics are so peculiar to make it a candidate ideally suited to be analyzed with this type of approaches.

The proband, indeed, suffers of a severe form of ICN and, strikingly, he belongs to a family with consanguinity in which ICN is present in several members, from both maternal and paternal branch (see family tree), and is transmitted in an apparently dominant fashion, with males being more severely affected than females.

MV, 50 years old, since 19 formed calcium oxalate and calcium phosphate renal stones, expelling up to date about 300 calculi both spontaneously and by lithotripsy (he underwent 33 treatments). The metabolic phenotype revealed intermittent hypercalciuria, hypocitraturia and phosphaturic tubulopathy. Pre-calyceal renal cysts were also observed. He was resistant to all pharmacological treatments so far adopted.

MV belongs to a family with consanguinity (his parents are second cousins and both are affected by nephrolithiasis). In this family nephrolithiasis is present in several members and

is transmitted in an apparently dominant fashion, with males being more severely affected than females.

Specific aim of this project was to exploit the potential of CGH array and SOLiD™ technology, in combination with new developed bioinformatic tools, in order to identify some of the major susceptibility genes involved in ICN.
Although the ultimate goal of this project was to find answers to medical genetic questions, it gave us the fundamental opportunity to elaborate efficient NGS data analysis pipelines.
NGS technologies, indeed, are still on their infancy as regards appropriate data analysis and filtering. Thus, in this context, our role of collaborating unit with specific bioinformatic unit (CRIBI, Padova) was critical to evaluate the output of bioinformatic analysis of NGS data in order to: identify problems related to the interface bioinformatic tool/ medical-biological questions; and establish valid criteria of inclusion or exclusion of genetic variants for the subsequent biological analyses. This could definitively provide our medical and genetic expertise to help in the development of a more user-friendly bioinformatic tool (QueryOR) and NGS data analysis guidelines.

3. MATERIALS AND METHODS

3.1 CASE-STUDY

MV, 50 years old, since 19 formed calcium oxalate and calcium phosphate renal stones, expelling up to date about 300 calculi both spontaneously and by lithotripsy (he underwent 34 treatments). He has been followed for thirty years by prof. D'Angelo (Nephrology Division, University of Padova).

The metabolic phenotype revealed:

- alternatively absolute (>300 mg/die) or marginal (UCa/Cr >0,14) idiopathic hypercalciuria;
- persistent hypocitraturia (≤0,95 mmol/die);
- phosphaturic tubulopathy: TRP<80% and TmPO4/GFR < 2.2 mg/dl, with salt depletion;
- absence of thyroid and parathyroid metabolism alterations;
- very recently, bilateral pre-calyceal renal calcifications were also observed;
- resistance to all pharmacological treatments so far adopted (thiazide diuretics, assumption of crystallization inhibitors as citrate and magnesium, abundant hydrotherapy).

Interestingly, MV belongs to a family with consanguinity (his parents are second cousins, both affected by nephrolithiasis). In this family ICN is present in several members (Fig. 5) and is transmitted in an apparently dominant fashion, with males being more severely affected than females.

The most striking clinical characteristics of family members are reported below.

- Father: bilateral symptomatic recurrent nephrolithiasis that led to monolateral nephrectomy.
- Paternal grand-father: recurrent nephrolithiasis, similarly to the father.
- Mother: nephrolithiasis diagnosis with constant "renella" emission, hypocitraturia and hypomagnesuria;

- Maternal (2/2) and paternal (3/3) aunts affected by recurrent nephrolithiasis, with one of the maternal branch certainly presenting hypocitraturia.
- Nephew (maternal branch): presenting a diagnosis of bilateral microlithiasis, constant " renella" emission, hypophosphoremia and sporadic hypocitraturia.

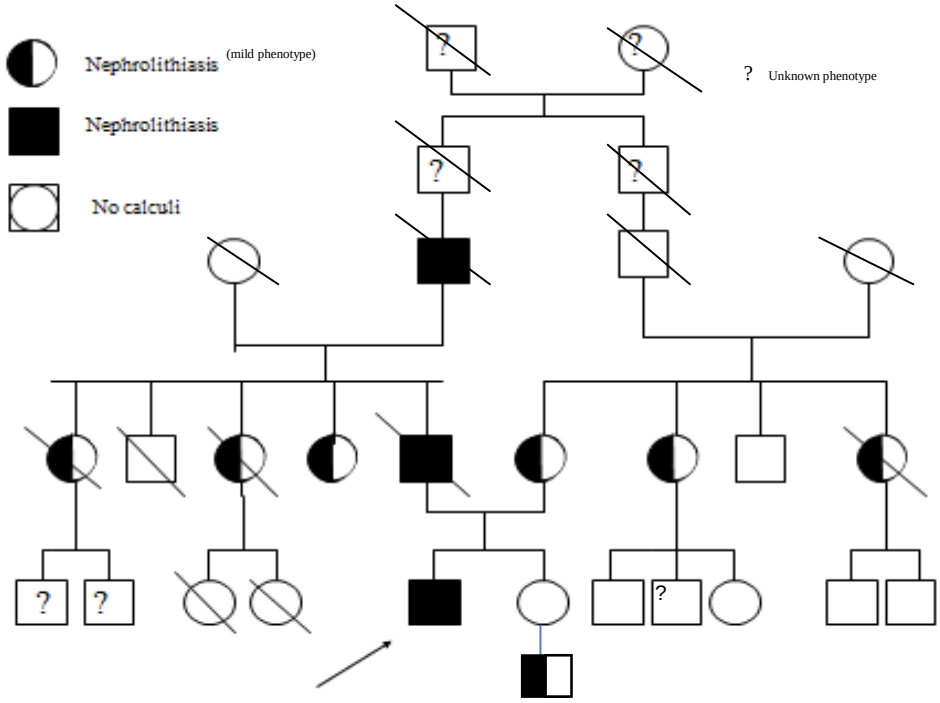


Figure 5. Genealogical tree of the case-study family.

From both clinical and metabolic points of view, the ICN encountered in this family does not present any of the characteristics of the monogenic forms. Nevertheless, the proband and affected and not affected family members have been analyzed for the presence of mutations /variations in candidate genes known to have a role in nephrolithiasis such as CLCN5 (in the suspicion of Dent's disease) and CaSR genes, but any relevant findings have been found.

3.2 GENOMIC STUDIES

3.2.1 DNA extraction and purification, PCR amplification

Peripheral blood leucocytes DNA was extracted by using *QIAamp® DNA Blood Mini Kit* (*QIAGEN*, CA, USA), following the indicated protocol.

DNA was lately quantified with *NanoDrop ND-1000* (*Celbio*, Italy) and its purity was checked basing on A_{260}/A_{280} ratio (between 1,8 and 2). It was subsequently used for needed applications or stored at -20°C.

PCR amplification was performed on *myCycler* thermal cycler (*Bio-Rad*, Italy) using a DNA working concentration of 25 ng in a reaction volume of 25 µl. Reaction mix constituents were: 1mM or 1.5 mM MgCl2 (*Sigma*, MO, USA), 0.4 mM forward and reverse primers (*Eurofins MWG Operon*, Germany), 0.2 mM dNTPs (*Boehringer*, Germany), 0.04 U/ml JumpStart Taq DNA Polimerase (*Sigma*, MO, USA), 50mM KCl 1X buffer (*Sigma*, MO, USA), and 10 mM Tris-HCl (pH 8.3).

PCR program used is detailed below:

Step 1. 95 °C, 5 min (Denaturation)

Step 2. repeated per 40 cycles (Amplification)

94 °C, 45 sec

T °C, 45sec (annealing T variable basing on primer sets)

72 °C, 1 min

Step 3. 72 °C, 7 min (Final extension)

If following sequencing applications were needed amplified DNA was further purified to remove residuals of all PCR reaction components (nucleotides, salts, polymerase,etc.) with MinElute® PCR Purification Kit (*QIAGEN*, CA, USA).

3.2.2 Primer design and optimization

The DNA sequence for a gene or for a genomic region was retrieved from the *UCSC genome browser* (http://genome.ucsc.edu).

Primer pairs for the region of interest were designed according to stringent parameters to ensure successful assays and convenient experiment design, by using *Primer3Plus software* (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). Subsequently, the generated primers were subjected to *in silico* validation to avoid single nucleotide polymorphisms (SNPs) and copy number polymorphisms at the annealing sites. These were excluded at the annealing sites using the corresponding tracks from the UCSC browser (*in silico PCR* tool) to open the Genome Browser at the position of the amplicon. The *BLAST* program from the NCBI browser (http://www.ncbi.nlm.nih.gov/BLAST/) was used for *in silico* specificity analysis.

Anyway, after thorough *in silico* quality control, an extensive empirical validation of the primer pairs was performed. First, amplification efficiencies are calculated based upon the generation of standard curves using genomic DNA (gDNA) dilution series. Subsequently, melting curve analysis, acrilamide gel electrophoresis or microchip electrophoresis (2100 BioAnalyzer, Agilent) were used to check the specificity of the PCR reactions.

3.2.3 CGH array analysis, CNVs detection and characterization

CGH arrays has been performed in the proband and in eight members of the family in order to detect the presence of Copy Number Variants (CNVs) in the genome, in collaboration with Dott. Zavan (Laboratory of Tissue Engineering, University of Padova).
Slide array 4x44K and 4x180K Agilent (Agilent Technologies) were used, ensuring a resolution of about 80-100 Kb and 13 kb, respectively. Map position was in accord to genomic assembly of March 2006. Analysis was carried out with Feature Extraction and

CGH-Analytics software. Evaluation of results based on Genome Assembly 2006 (hg18), *UCSC Genome Browser* and *DGV* (*Database of Genomic Variants*).
DGV (https://decipher.sanger.ac.uk/) is a database that provides a comprehensive summary of structural variations (larger than 50bp) in the human genome of healthy control samples.
Gene database at NCBI (http://www.ncbi.nlm.nih.gov/gene) was lately used to clarify CNV genetic content.

3.2.3.1 Real time CNV-profiling

For copy-number determination, a qPCR assay was developed. In particular it has been used to dose particular genomic fragments of Xq22.2 region, allowing a correct quantification of the presence of one, two or more copies. This assay was set up on available members of the case study family and lately

Dedicated design of the qPCR primers and extensive *in silico* validation were the starting points, as detailed above.

Experiment design was the subsequent but indispensible (but often overlooked) step in the workflow of accurate real-time PCR based genomic quantification. Reference assays were included in the screening to accurately measure and correct for variations in the total amount of input DNA. These assays amplified a piece of genomic DNA that is known not to be affected (i.e. not registered as a known copy number polymorphism). Most autosomal inherited genes with an essential function, not related to the studied phenotype, can be used as a reference sequence. We have used an assay amplifying GPR15 genomic DNA as references for normalization of the qPCR data since 2005 (RTprimerDB #1022, http://www.rtprimerdb.org) (Lefever et al., 2009) At least two types of control samples should be included in every qPCR-based copy number analysis. “No template controls” were included to detect the presence of contaminating DNA. Specific for qPCR-based copy number analysis, was the inclusion of reference samples with a known copy number.

During calculations and result interpretation, these control samples were used as a reference point (or calibrator) for the determination of the copy numbers. The inclusion of multiple reference samples result in more accurate results. We used two samples, one with a normal and the other with a known CNV. The latter CNV sample with known Xq22.2 duplication served as reference point and as positive control for the detection of CNVs.
A list of primers sequences used in this study is showed below in Table 2.

Table 2. Primer sequences used in CNV-profiling assays.

Primer name	Forward sequence	Reverse sequence
gGPR15	GGTCCCTGGTGGCCTTAATT	TTGCTGGTAATGGGCACACA
gNUP62CL	TTAGTGTAGTAGCAACTCCTG	CTCAATCAATGTATGGTCCC
gRBM41/I	TCCTCAATTGCAGTGGTCTG	TGATACATGTTACCAGGGGCA
gRBM41/II	GAAGAGTGATGAGCATGTCCT	CATACATTCCTCAATGGAGACAG
gCXorf41	TGGAATGTACCAGTGCCAAA	TCTTTTTCATGCAGTTTCAGAAG

Real-time quantitative polymerase chain reaction (*qRT-PCR*) was performed to quantify genomic DNA levels, using *SYBR Green* technology. The principle consists in the quantitative relationship that exists between the amount of starting material (target sequence) and the amount of PCR product at any given cycle. When amplifying the DNA, an increasing amount of double-stranded DNA is created, which binds the *Sybr® Green* (fluorescent) probe resulting in an increase of fluorescence. By plotting the increase in fluorescence versus cycle number it is possible to analyze the PCR kinetics in real-time (Pfaffl, 2001). For each fragment to analyze, a reaction mix was made (15 µl final volume), prepared as follows: 7,5 µl of *iQ SYBR Green Mastermix* (*BioRad*, Italy); 4,55 µl of milliQ water; 0,225 µl of 20 µM Forward Primer (300 nM); 0,225 µl of 20 µM Forward Primer (100 nM); 0,5 µl of 5 µM Reverse Primer (300 nM). An overview of the primers used is given below in Table 2. The reaction mix was loaded, 12,5 µl/well, in 96-well plates

(*BioRad*, Italy) and 2,5 µl of the DNA of interest was added to each well. The plate was subsequently loaded on *iCycler* thermal cycler (*BioRad*, Italy).

To ensure quality control on the precision and accuracy of the obtained qPCR data, provide better accuracy, we used 3 technical PCR replicates and a 3-points standard curve) with a dilution factor of 10 (starting from 10 ng/µl concentration).

PCR reactions were performed on the *iCycler* (*BioRad*, Italy).

The PCR program used is detailed below:

Step 1: 50°C, 2 min (Incubation)
Step 2: 95°C, 2 min
Step 3: 95°C, 15 sec (Denaturation) (Step 3, repeated per 40 cycles)
60°C, 45 sec (Extension)
Step 4: 95°C, 15 sec (Dissociation)
Step 5: 60°C, 20 sec (Melting curve)

Step 6: 95°C, 15 sec

The level of each genomic fragment was always compared to GPR15(*G protein-coupled receptor 15*), the housekeeping gene used as reference assay, level calculating the ΔCt sample value as follows: $Ct_{\text{genomic fragment}} - Ct_{\text{GPR15}}$. The ΔΔCt was calculated subtracting ΔCt calculated for the calibrator ($\Delta Ct_{\text{sample}} - \Delta Ct_{\text{calibrator}}$). Finally, the ultimate calculation to obtain genomic DNA level of each fragment was 2^(-ΔΔCt).

3.2.3.2 CNVs burden in case and "control" cohorts

In collaboration with Dott. Sanna Cherchi (Division of Nephrology, Columbia University, New York), CNVs burden in control cohorts was examined.

The “control” group consisted of 14,375 anonymized adults and children: these are not to be considered as properly defined control in our ICN study because they were collected for allowing another kind of genetic studies, on renal developmental abnormalities, and therefore we cannot exclude that they were eventually affected by ICN.

Two cohorts consisted of white European affected individuals recruited from pediatric centers in Italy, Poland, Macedonia, Croatia, and the Czech Republic. All cases were unrelated. Inclusion criteria included the presence of a primary renal-parenchyma defect—such as renal agenesis, a congenital solitary kidney or renal hypodysplasia (finding of a small or cystic kidney for age). Another cohort consisted of 134 multiethnic North American individuals (63% white, 23% African American, and 10% admixed) diagnosed with renal hypodysplasia. Additionally, six cohorts of European (80,4%), Asian (13,4%), and African American ancestry (6,1%), were examined: they were genotyped on high-density *Illumina* platforms as cases or controls for genetic studies of complex traits not related to any developmental phenotypes.

The CNV-profiling assay was applied to screen a case cohort of 85 ICN patients (collected by prof. Gambaro G., Catholic University of Rome; and Dott. Fabris A., University of Verona).

3.2.4 Whole Exome Sequencing analysis, data filtering and validation

Whole Exome Sequencing of the proband was performed (in collaboration with prof. Valle G., CRIBI, Padova) with *SOLiD 5500xl* sequencer, at an average 15X coverage.

3.2.4.1 Exome sequencing data analysis: criteria for identifying single-nucleotide and copy-number variants

To analyze and filter single-nucleotide variant findings, we used *QueryOR* platform, a web-based query platform that aggregates several functional annotation of variants and genes (developed by Bioinformatic group, CRIBI, prof. Valle G., Padova), for giving valuable prioritization criteria and reduce the search space of analysis of exome sequencing data. The prioritization strategy consist in a ranking system that sorts results based on the satisfied criteria, in contrast with other developed tools that applies filters to reduce the results.

We generally applied two criteria of search to identify the most promising *variant call set*: a **gene-centered** search, starting from known notions about genes and focusing attention on those that harbored promising variants; a **variant-centered** search, starting from characteristic intrinsic to the variants per se like type (sense, missense, nonsense), homozygosis and genomic position.

For missense variants, we generally applied a subsequent prioritization strategy by ranking variants on the basis of query coverage, low *MAF* (*Minor Allele Frequency*) values (< 0,001) and a *possibly-probably damaging/deleterious* prediction (*PolyPhen/SIFT* prediction tools), as could be inferred from QueryOR reports.

To investigate the possible pathogenic significance of sense variants, we used *ASSEDA* software (*Automated Splice Site and Exon Definition Analysis*, http://splice.uwo.ca/) that allowed to study the effect of a single nucleotide variant in modificating canonical splicing, with the prediction of activation an strenght of criptic sites of splicing.

To investigate the presence of CNVs we filtered exome sequencing data with bioinformatic software developed by CRIBI Genomics that allow the detection of small structural variations.

3.2.4.2 Molecular validation and segregation study

Because NGS technique is not exempt from errors and a low percentage of false positive calls always exist, variants that passed the *in silico* prioritization strategy were submitted to molecular validation by Sanger sequencing. This validation step was performed on proband's DNA. Furthermore, once the variant was confirmed, we extended the validation procedure to the other members of the family to perform a segregation analysis.

Primer were designed, as detailed above, to sequence a genomic fragment that contained the single-nucleotide variant to be confirmed; primer sequences are listed below (Table 3). PCR reaction and purification were performed as previously described.

Table 3. Primer sequences used for single-nucleotide variant validation

Primer name	**Forward sequence**	**Reverse sequence**
gNCX1/rs148215685	AGCACAAAGAGGCAGGATGT	CATCGCTGCCATCTACCAC
gNUP62CL/rs1285590	CTGTCACAGCAGCAGGAACT	ACATCTCCACATGCTCCTCA
gNUP62CL/rs1298577	CTGTCACAGCAGCAGGAACT	ACATCTCCACATGCTCCTCA
gATP6V1B1/rs11681642	CCAGCTGGACCTGAAGTCTC	GGCCTGCTGTCTATCTCCAT
gATP6V1B1/not known	CTAACACTCCCTCCCGCTCT	CGGGAATAGAACTCGTCGAT
gFGF23/not known	TAATTCACTTCAACACCCCCA	TCGGGAGCTCCTGTGAACAG
gHNF1B/not known	GCAATTACTCCATGATTATGCTACTT	ATGTGAATTATTTGATTAAAATCTGA
gXDH/ rs45624433	CTGAGCCTCACCTGTCCAAT	TCTGCTTGGGAACGTACTCT
gXDH/ rs17011368	CTGCTTCGGAAAACCCCTTC	CAGCTGAGGAAATGGAGGAA
gUSP29/ rs9973206	AGAACTTTTTCCCGAAATGGA	CCGGGAATTTCTCTCTTGAA
gKCNS1/not known	AAAGGCGAAGACGCACAGCTCGT	AGGGAGGAGCACTGAAACCT

Sequencing reaction of PCR products was set up by using *BigDye® Terminator v1.1 Cycle Sequencing Kit* (*Applied Biosystems*, CA-USA), with a purified amplified DNA concentration of 2 ng/ µl. For each PCR product to be sequence, a reaction mix was made (10 µl final volume), prepared as follows: 2 µl Terminator Ready Reaction Mix; 0,5 µl Forward or Reverse primer 3.2 µM; 5,5 µl milliQ water.

Sequencing program is detailed below:

Step 1. 96 °C, 1 min

Step 2. repeated per 25 cycles

94 °C, 10 sec

50 °C, 5 sec

60 °C, 4 min

Sequence products were lately stored at -20°C or immediately processed to purification.

Tagged terminators not incorporated in the reaction were subsequently removed through *CENTRISEPT Spin Columns* (Princeton Separations, Applied Biosystems, USA).

Elution product was later loaded on sequencing plate and read on *3130 Genetic Analyzer* (*Applied Biosystems*).

Sequences were displayed with *Chromas 1.45* tool and compared to genomic sequence by performing *BLAST (bl2seq)* (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi), or using *UCSC Genome Browser* to investigate the presence/absence of the detected variant.

3.3 TRANSCRIPTIONAL STUDIES

3.3.1 RNA exctraction, quantification and cDNA synthesis

RNeasy Micro Kit *(Qiagen, cat no. 74004)* was used to isolate total RNA from leucocyte cells, following the spin column protocol. RNA was finally eluted in RNAse free water and stored at –80°C.

RNA concentration was measured with Nanodrop ND-100 spectrophotometer (Celbio, Italy) and purity was checked by the A_{260}/A_{280} ratio (between 1,8 and 2).

RNA integrity was valued by capillary electrophoresis on chip with Agilent 2100 Bioanalyzer (Agilent Technology), by observing and quantifying integrity of 28S and 18S rRNAs.

We routinely performed two-step RT-PCR reactions, in which the first step is cDNA

synthesis. This protocol allowed us to make cDNA by using MuLV DNA polymerase starting from a single-strand RNA template.

RNA was retro-transcribed from a starting quantity of 100 ng in a reaction volume of 20 µl. Reaction mix was prepared as follows: 4 µl MgCl2 25 mM (Sigma); 2µl dNTPs 10 µM (Roche); 1µl random examers 50 µM (Applied Biosystem); 1µl RNAse inhibitor 20 U/ µl (Applied Biosystem); 1µl MuLV inverse transcriptase 50 U/ µl (Applied Biosystem), 2µl Buffer 10X (Sigma) constituted by KCl 50 mM e Tris-HCl 10 mM (pH8.3) and milliQ water to reach the final volume.

Reaction was performed on iCycler thermal cycler by applying the following protocol:

Step 1. T amb, 10 min (Primer annealing)

Step 2. 42 °C, 30 min (Primer extension)

Step 3. 65 °C, 5 min (Denaturation)

Step 4. 4°C, 5 min (Cooling)

3.3.2 Quantification of gene expression: Real Time PCR

Real-time quantitative polymerase chain reaction (qRT-PCR) was performed to analyze gene expression levels, using *iQ SYBR Green SuperMix (BioRad,* Italy) as detailed above.

For each fragment to analyze, a reaction mix was made (15 µl final volume), prepared as follows: 12,5 µl of iQ SYBR Green Mastermix; 4,55 µl of milliQ water; 0,225 µl of 20 µM Forward Primer (300 nM) and 0,225 µl of 20 µM Reverse Primer (300 nM) (MWG Operon, Germany). An overview of the primers is given below in Table 4.

Table 4. Primer sequences used for expression study.

Primer name	Forward sequence	Reverse sequence
NCX1	TGAATTCAAGAGTACTGTGGAC	TGCATCACGTAATCGAAACAG
MORC4	ACCTACCTTCACAAATAAGCAG	ATTACTCCTACACCTTCTCCA
NUP62CL	ACTGTTAGTGTAGTAGCAACTCC	CATTCTCAATCAATGTATGGTCCC
CLDN2	TGTCTTCTAGATGCCTTCTTGAG	CAGACCTCTCAGTAGAAGCG
PRPS1	TAAGAAAGATAAGAGCCGGGC	CTCTGCATACAAATTGTCTACTGG

The reaction mix was loaded, 14 μl/well, in 96-well plates (Bior) and 1 μl of the DNA of interest was added to each well.

PCR reactions were performed on the *iCycler* (*BioRad*, Italy). The PCR program used is detailed below:

Step 1: 50°C, 2 min (Incubation)

Step 2: 95°C, 2 min

Step 3: 95°C, 15 sec (Denaturation) (Step 3, repeated per 40 cycles)

60°C, 45 sec (Extension)

Step 4: 95°C, 15 sec (Dissociation)

Step 5: 60°C, 20 sec (Melting curve)

Step 6: 95°C, 15 sec

The expression level of each gene was always compared to GAPDH (*Glyceraldehyde 3-phosphate dehydrogenase*) (forward: GAAGGTGAAGGTCGGAG; reverse: TGGCAACAATATCCACTTTACCA) housekeeping gene calculating the ΔCt sample value as follows: $Ct_{gene} - Ct_{GPR15}$. The ΔΔCt was calculated subtracting ΔCt calculated for the calibrator ($\Delta Ct_{sample} - \Delta Ct_{calibrator}$). Finally, the ultimate calculation to obtain the relative gene expression was 2^(-ΔΔCt).

4. RESULTS AND DISCUSSION

4.1 CGH ARRAY ANALYSIS: DISCOVERY OF A NOVEL CNV

To investigate the presence of CNVs in the case study family, we performed CGH array analysis on all available family members.

We found the presence of a duplication of 308,398 Kb of chromosome X q22.2 region in the proband and in other three family member (Fig. 6). This duplication having START at 106,126,808 and STOP at 106,435,205 chrX positions (basing on hg18 Assembly), respectively.

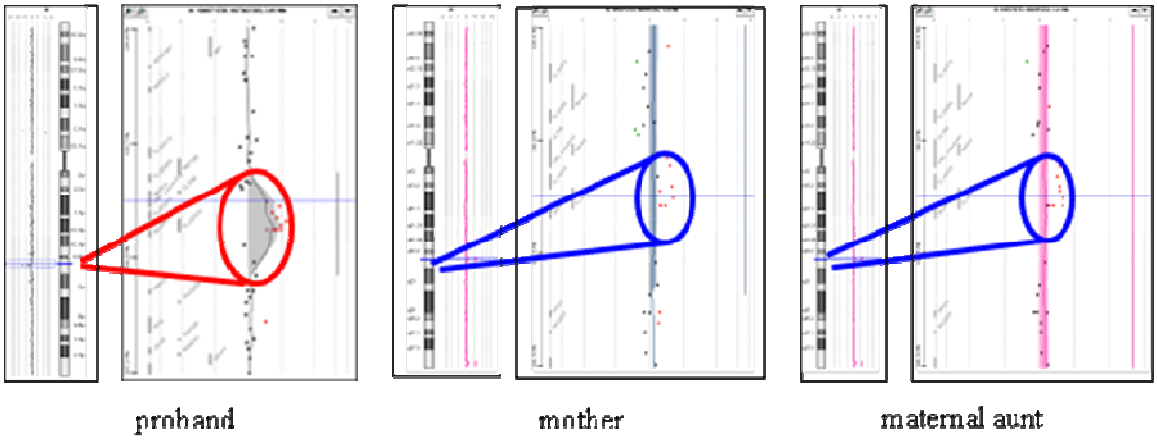


Figure 6. Feature extraction software visualization of CGH results: the Xq22.2 duplication of 308,397 Kb is evidenced by the probes (circled red dots) that overlie in the +1 axis; results are shown for three over four family members that harboured the Xq22.2 duplication.

As can be inferred by the genealogic tree in Fig. 7, three over four family members that present Xq22.2 duplication are certainly affected. Thus, this seemed a first hint indicating the striking importance of this finding. Furthermore, by consulting the available public database we found that this was a novel CNV, because it resulted absent in the DGV (Database of Genomic Variants).

Strikingly, this last finding seemed to indicate that we encountered certainly a very rare copy-number variant.

The CNV has been inherited from the affected mother and was also present in the maternal affected aunt. However this variant cannot be shared by the paternal branch since the progenitors of maternal and paternal branch are two brothers (see family tree, Fig. 7). Thus, we hypothesized that additional genetic determinants have been inherited from the paternal branch. Alternatively, the Xq22.2 duplication might be the result of an X/autosome balanced translocation occurred in one of the common progenitors and segregating differently in the descendents of maternal and paternal branches.

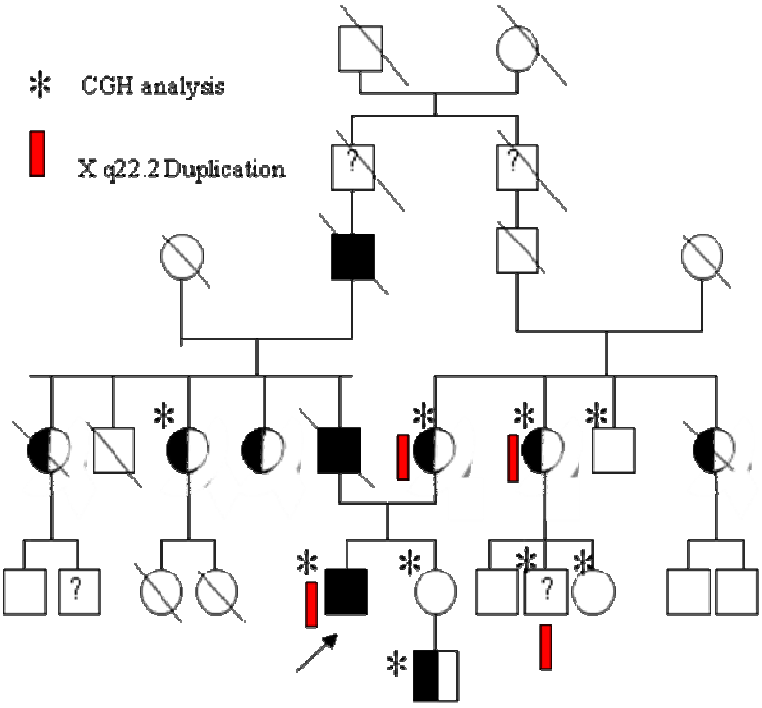
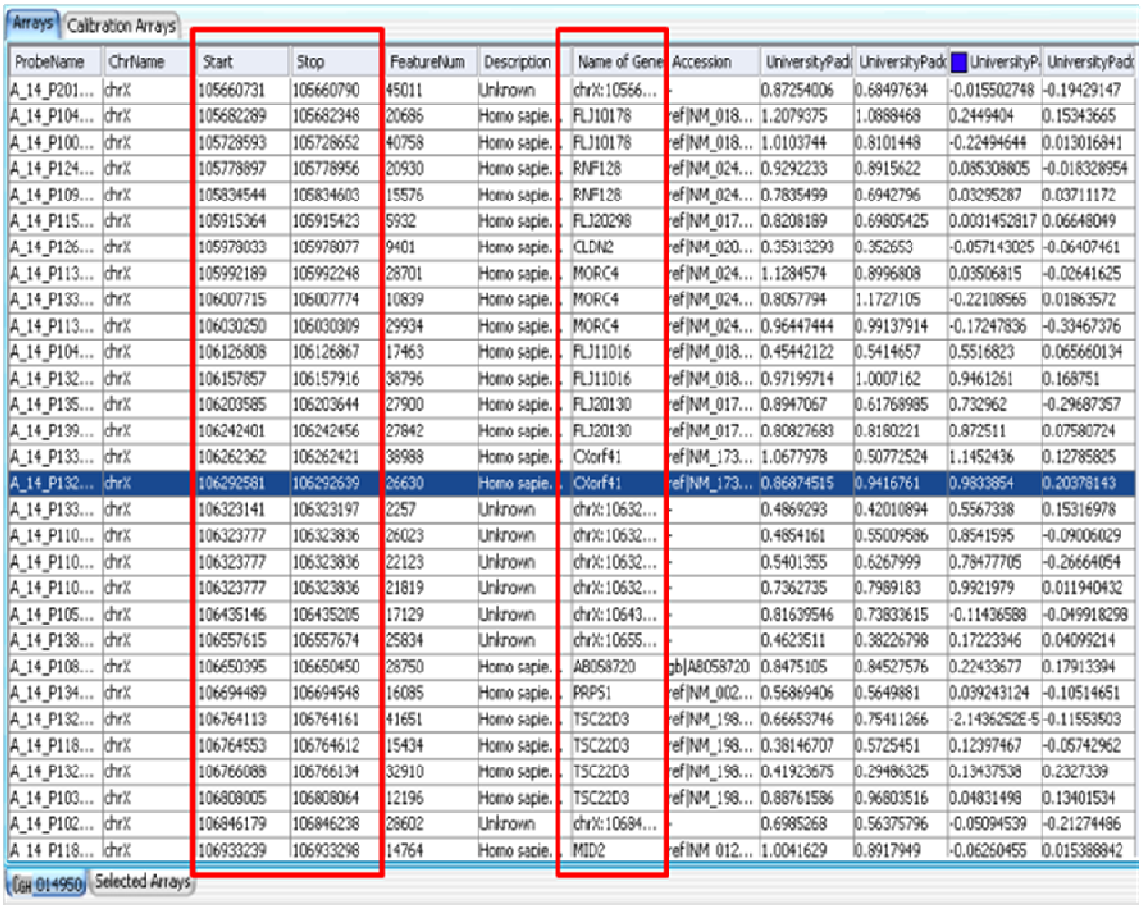


Figure 7. CGH array analysis pin the case-study family. It was performed in nine family members (as indicated by the asterisk) and the Xq22.2 duplication was detected in the four members tagged with the red bar.

Considering the intriguing finding, we started to deeply investigate the duplicated Xq22.2 region by analyzing its genetic content. By extracting information from the Feature Extraction Software, we could infer which genes were identified by the probes that resulted with a double signal (Fig. 8) and by comparing it with *Gene* database we could clarify that three genes were included in the duplicated region: RBM41, NUP62CL and CXorf41. This was subsequently confirmed by UCSC Genome browser analysis (Fig. 9).
RBM41 gene encodes for the *RNA binding motif protein 41*; NUP62CL (*nucleoporin 62 C-terminal like)* is an homolog gene of nucleoporin 62, NUP62; CXorf41 or PIHD3 (*PIH1 domain containing 3*) is a gene implicated in a positive regulation of human rRNAs.



ProbeName	ChrName	Start	Stop	FeatureNum	Description	Name of Gene	Accession	UniversityPadi	UniversityPadi	UniversityP	UniversityPadi
A_14_P201...	chrX	105660731	105660790	45011	Unknown	chrX:10566...	-	0.87254006	0.68497634	-0.015502748	-0.19429147
A_14_P104...	chrX	105682289	105682348	20686	Homo sapie...	FLJ10178	ref|NM_018...	1.2079375	1.0888468	0.2449404	0.15343665
A_14_P100...	chrX	105728593	105728652	40758	Homo sapie...	FLJ10178	ref|NM_018...	1.0103744	0.8101448	-0.22494644	0.013016841
A_14_P124...	chrX	105778897	105778956	20930	Homo sapie...	RNF128	ref|NM_024...	0.9292233	0.8915622	0.085308805	-0.018328954
A_14_P109...	chrX	105834544	105834603	15576	Homo sapie...	RNF128	ref|NM_024...	0.7835499	0.6942796	0.03295287	0.03711172
A_14_P115...	chrX	105915364	105915423	5932	Homo sapie...	FLJ20298	ref|NM_017...	0.8208189	0.69805425	0.0031452817	0.06648049
A_14_P126...	chrX	105978033	105978077	9401	Homo sapie...	CLDN2	ref|NM_020...	0.35313293	0.352653	-0.057143025	-0.06407461
A_14_P113...	chrX	105992189	105992248	28701	Homo sapie...	MORC4	ref|NM_024...	1.1284574	0.8996808	0.03506815	-0.02641625
A_14_P133...	chrX	106007715	106007774	10839	Homo sapie...	MORC4	ref|NM_024...	0.8057794	1.1727105	-0.22108565	0.01863572
A_14_P113...	chrX	106030250	106030309	29934	Homo sapie...	MORC4	ref|NM_024...	0.96447444	0.99137914	-0.17247836	-0.33467376
A_14_P104...	chrX	106126808	106126867	17463	Homo sapie...	FLJ11016	ref|NM_018...	0.45442122	0.5414657	0.5516823	0.065660134
A_14_P132...	chrX	106157857	106157916	38796	Homo sapie...	FLJ11016	ref|NM_018...	0.97199714	1.0007162	0.9461261	0.168751
A_14_P135...	chrX	106203585	106203644	27900	Homo sapie...	FLJ20130	ref|NM_017...	0.8947067	0.61768985	0.732962	-0.29687357
A_14_P139...	chrX	106242401	106242456	27842	Homo sapie...	FLJ20130	ref|NM_017...	0.80827683	0.8180221	0.872511	0.07580724
A_14_P133...	chrX	106262362	106262421	38988	Homo sapie...	CXorf41	ref|NM_173...	1.0677978	0.50772524	1.1452436	0.12785825
A_14_P132...	chrX	106292581	106292639	26630	Homo sapie...	CXorf41	ref|NM_173...	0.86874515	0.9416761	0.9833854	0.20378143
A_14_P133...	chrX	106323141	106323197	2257	Unknown	chrX:10632...	-	0.4869293	0.42010894	0.5567338	0.15316978
A_14_P110...	chrX	106323777	106323836	26023	Unknown	chrX:10632...	-	0.4854161	0.55009586	0.8541595	-0.09006029
A_14_P110...	chrX	106323777	106323836	22123	Unknown	chrX:10632...	-	0.5401355	0.6267999	0.78477705	-0.26664054
A_14_P110...	chrX	106323777	106323836	21819	Unknown	chrX:10632...	-	0.7362735	0.7989183	0.9921979	0.011940432
A_14_P105...	chrX	106435146	106435205	17129	Unknown	chrX:10643...	-	0.81639546	0.73833615	-0.11436588	-0.049918298
A_14_P138...	chrX	106557615	106557674	25834	Unknown	chrX:10655...	-	0.4623511	0.38226798	0.17223346	0.04099214
A_14_P108...	chrX	106650395	106650454	28750	Homo sapie...	AB058720	gb|AB058720	0.8475105	0.84527576	0.22433677	0.17913394
A_14_P134...	chrX	106694489	106694548	16085	Homo sapie...	PRPS1	ref|NM_002...	0.56869406	0.5649881	0.039243124	-0.10514651
A_14_P132...	chrX	106764113	106764161	41651	Homo sapie...	TSC22D3	ref|NM_198...	0.66653746	0.75411266	-2.1436252E-5	-0.11553503
A_14_P118...	chrX	106764553	106764612	15434	Homo sapie...	TSC22D3	ref|NM_198...	0.38146707	0.5725451	0.12397467	-0.05742962
A_14_P132...	chrX	106766088	106766134	32910	Homo sapie...	TSC22D3	ref|NM_198...	0.41923675	0.29486325	0.13437538	0.2327339
A_14_P103...	chrX	106808005	106808064	12196	Homo sapie...	TSC22D3	ref|NM_198...	0.88761586	0.96803516	0.04831498	0.13401534
A_14_P102...	chrX	106846179	106846238	28602	Unknown	chrX:10684...	-	0.6985268	0.56375796	-0.05094539	-0.21274486
A_14_P118...	chrX	106933239	106933298	14764	Homo sapie...	MID2	ref|NM_012...	1.0041629	0.8917949	-0.06260455	0.015388842

Figure 8. Feature extraction software frame showing in columns the probes used to perform the CGH analysis in the Xq22.2 region, indicating for each one its name, start and stop positions. The name of the corresponding gene is also shown. Data shown are relative to Human Mar. 2006 (NCBI36/hg18) Assembly.

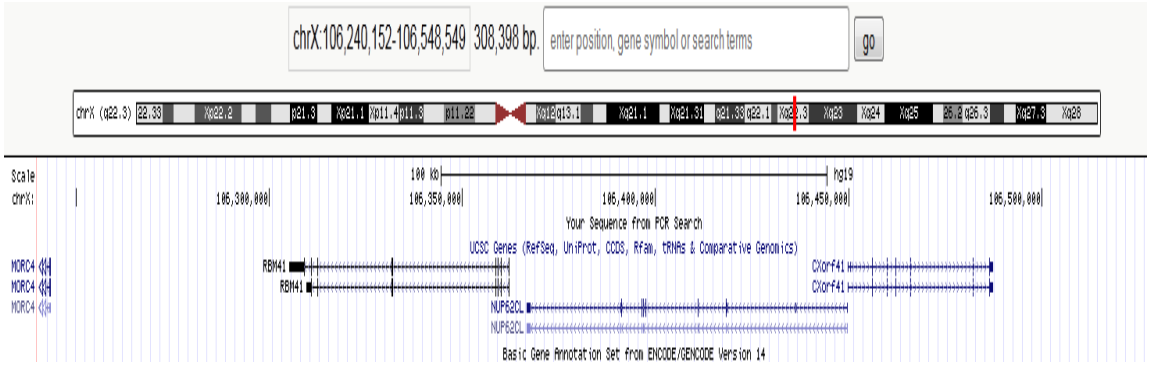


Figure 9. UCSC GENOME Browser frame illustrating the Xq22.2 duplication of 308,398 Kb, with the new genomic coordinates (chrX: 106,240,152-106,548,549) updated to Human Feb. 2009(GRCh37/hg19) Assembly.

4.2 Xq22.2 CNV IN-DEPTH ANALYSIS: TRANSCRIPTIONAL STUDY

Most copy number variants exist in healthy individuals; however, these variants are hypothesized to cause diseases through several mechanisms. First, copy number variants can directly influence gene dosage through insertions or deletions, which can result in altered gene expression and potentially cause genetic diseases.

Gene dosage describes the number of copies of a gene in a cell, and gene expression can be influenced by higher and lower gene dosages (Feuk et al., 2006). For example, deletions can result in a lower gene dosage or copy number than what is normally expressed by removing a gene entirely (Fig. 10a). Deletions can also result in the unmasking of a recessive allele that would normally not be expressed (Fig. 10b). Structural variants that overlap a gene can reduce or prevent the expression of the gene through inversions, deletions, or translocations (Fig. 10b). Variants can also affect a gene's expression indirectly by interacting with regulatory elements. For instance, if a regulatory element is deleted, a dosage-sensitive gene might have lower or higher expression than normal (Fig.10c). Sometimes, the combination of two or more copy number variants can produce a complex disease, whereas individually the changes produce no effect (Fig. 10d).

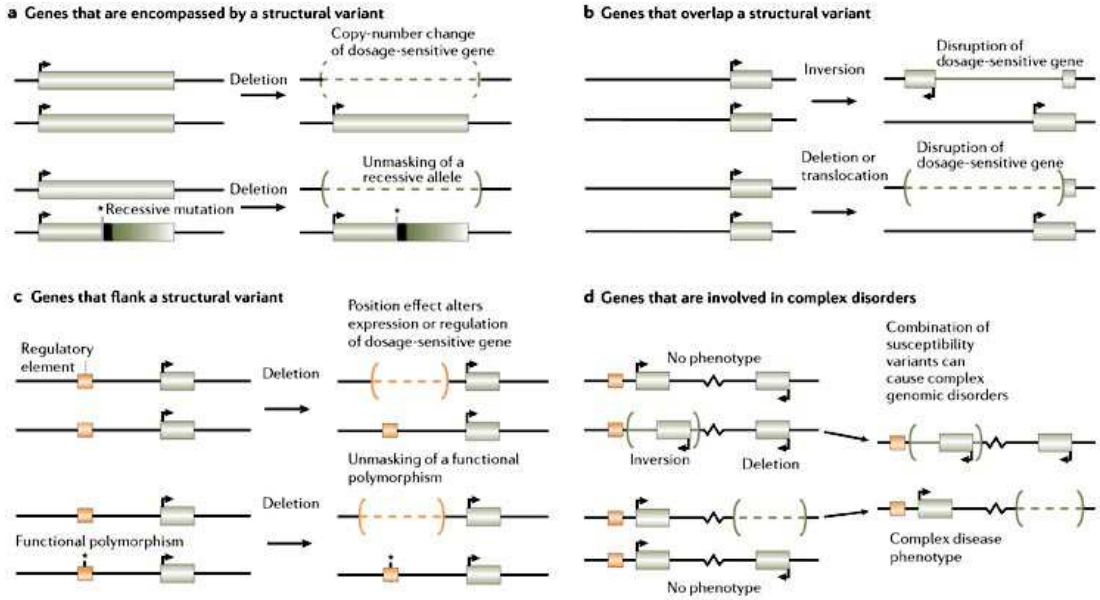


Figure 10. Influence of structural variants on phenotype (Image by *Feuk et al., 2006*).

Some variants are flanked by homologous repeats, which can make genes within the copy number variant susceptible to nonallelic homologous recombination and can predispose individuals or their descendants to a disease (Freeman et al., 2006). Additionally, complex diseases might occur when copy number variants are combined with other genetic and environmental factors (Feuk et al., 2006).

We firstly decided to examine Xq22.2 duplication effect at gene expression level and, by taking all the over cited basics in considerations, we planned a well comprehensive expression study.

In brief, we considered both genes strictly included in the duplicated region, both genes situated slightly upstream or downstream, and genes that could be influenced by a long-range position effect (Kleinjan and van Heyningen, 2005).

Thus, we set up our expression study taking in consideration five candidate genes. The nucleoporin NUP62CL, that is the only gene strictly included in the Xq22.2 duplication region. MORC4, that encodes for a protein containing an ATPase domain and a Zinc-finger domain crucial for protein-.protein and protein-DNA interactions. CLDN2, that encodes for a claudin of epithelial and endothelial tight junctions responsible of small ions and solutes transport. PRPS1 (*phosphoribosylpyrophosphate synthetase 1*), whose overexpression has

been found linked to a genetic disorder characterized by hyperuricemia and hyperuricosuria (De Brouwer et al., 2008). Finally, we also considered a gene not contained in the X chromosome, NCX1 (sodium/calcium exchanger member 1), because it could be influenced by a long-range positional effect. Upstream the duplicated region, indeed, it is located FXYD6P3 gene, belonging to a family of ion transport regulators that in some tissues regulate the NCX1 exchanger (Cheung et al., 2013). This gene, in the kidney, is responsible of calcium extrusion from the basolateral membrane, and for this reason it represents a good candidate for ICN, even if it has never been analyzed in human nephrolithiasis.

Examining gene expression in leucocytes of control and mutated members (harboring Xq22.2 duplication), we found a significant increase in the expression of NCX1 in the mutated family members (Fig. 11). Thus, considering its role in the kidney, our results point it out as possible susceptibility gene for ICN.

We either detected a decrease in expression of CLDN2 and NUP62CL, even though it could not be assumed as significant (Fig. 11).

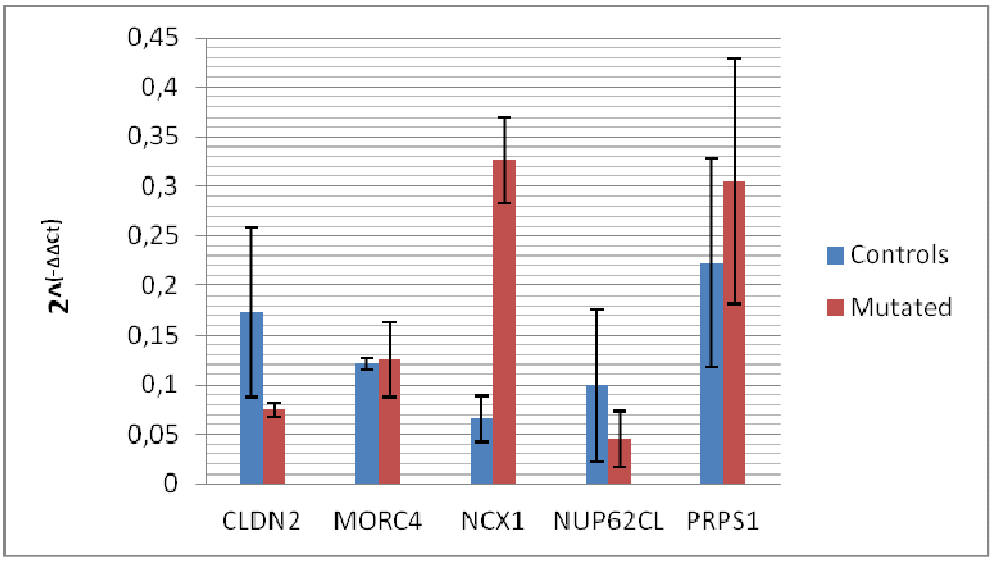


Figure 11. Expression levels of the five candidate genes are shown in the graph. Results are referred to leucocytes samples.

Subsequently, we could analyze expression levels of the candidate genes on fibroblast samples taken from a bladder biopsy of our proband and we compared them to control fibroblasts.

Interestingly, NUP62CL gene showed a significant and huge decrease in the expression in the proband sample, compared to controls (Fig.12).

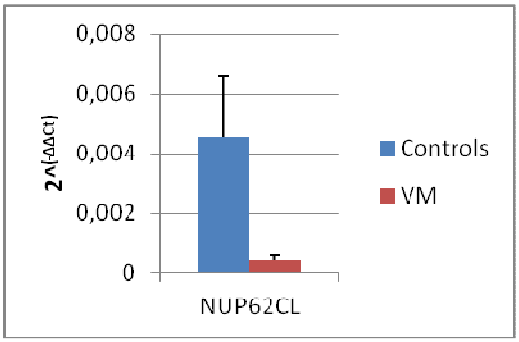


Figure 12. Expression level of NUP62CL gene in fibroblast samples.

This last finding appeared particularly appealing in our ICN susceptibility genes search, NUP62CL being the homolog of NUP62, also known as *nuclear pore complex oxalate binding protein 62.* This protein, indeed, mediates oxalate transport from the cytoplasm into the nucleus and its expression has been found increased in experimental hyperoxaluria (Sivakamasundari et al., 2004), a common metabolic condition in ICN patients (Asplin et al., 2002). In recent years, our proband did not show the intermediate phenotype hyperoxaluria but, at his nephropathy exordium, he formed calcium oxalate stones. Furthermore, the over cited studies, dealing with an *in vitro* mimicked hyperoxaluria, prove a clear modulation of NUP62 by oxalate. Therefore, the down-regulation that we observed well fits in this scenario.

4.3 GENOMIC STUDIES

4.3.1 Characterization of Xq22.2 duplication and "case-control" study

Real-time PCR represents, indeed, the golden standard for a quick and reliable confirmation of CGH findings (D'haene et al., 2011).

Thus, we firstly served of our case study family to set up a CNV-profiling assay, to detect and confirm the presence of one, two or three copies of the Xq22.2 region in object (see materials and methods). We initially chose a genomic fragment encompassing a region contained in NUP62CL gene (chrX: 106,397,338-106,397,480, hg19), that was the gene lying in the center of the duplicated region.

The chosen genomic fragment resulted effectively capable to discriminate the different genomic profiles in the case-study family: dealing with a chrX region, indeed, we detected a single dose for normal males, a double dose for mutated males and normal females and a triple dose for mutated females (Fig. 13).

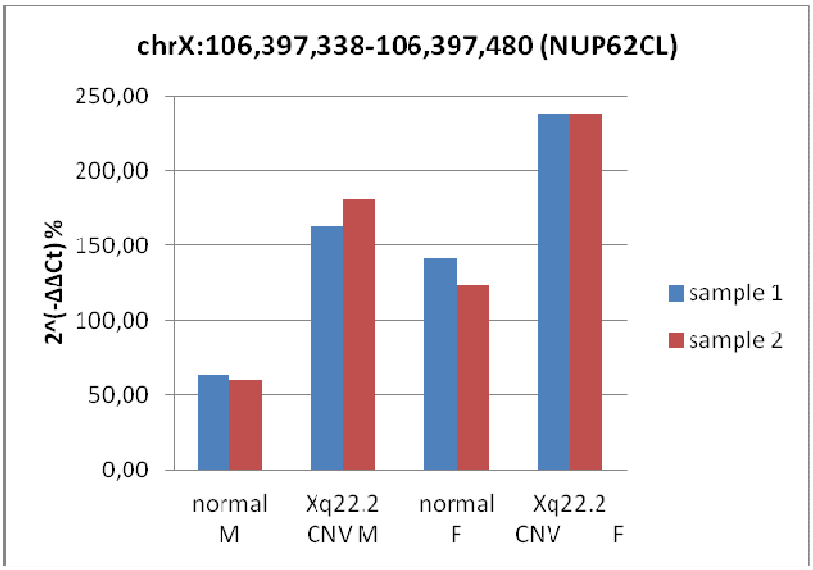


Figure 13. CNV-profiling assay set up. The NUP62CL genomic fragment resulted quantitatively reliable in the discrimination of the different genomic profiles. Samples 1 and 2 stand for two family members each one for each category (normal males, Xq22.2 CNV males, normal females and Xq22.2 CNV females).

Lately, we tested three more genomic fragments to confirm and better characterize the Xq22.2 duplication. We chose two fragments in the proximity of START and STOP regions, RBM41/I (chrX:106,312,496-106,312,608) and CXorf41 (chrX:486,428-106,486,544), respectively; and another fragment lying more centrally in the duplicated region, but upstream the first one tested (NUP62CL), RBM41/II (chrX:106,359,873-106,359,981). By observing the underlying graph (Fig. 14a), it appears clear that the fragment proximal to START region, RBM41/I, actually did not seem internal to the duplication because both normal and mutated (Xq22.2 CNV) males presented a single dose (near to 100%).

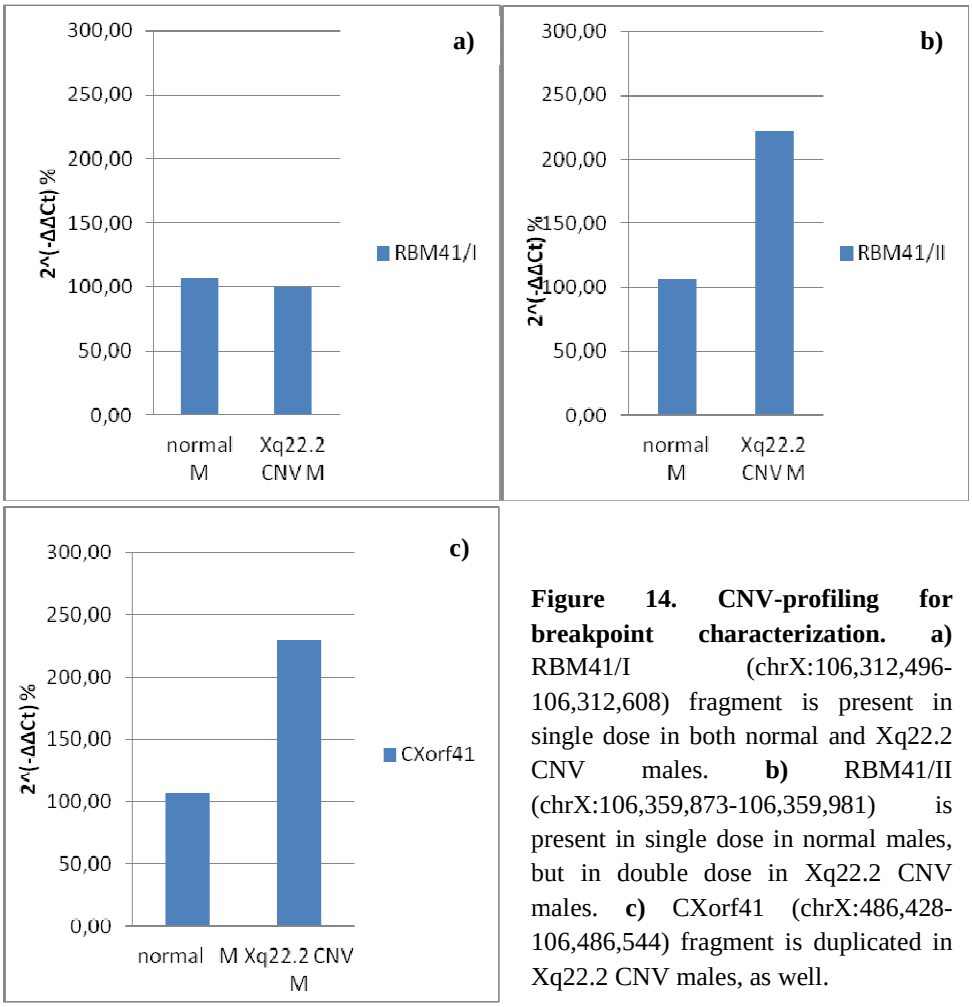


Figure 14. CNV-profiling for breakpoint characterization. a) RBM41/I (chrX:106,312,496-106,312,608) fragment is present in single dose in both normal and Xq22.2 CNV males. **b)** RBM41/II (chrX:106,359,873-106,359,981) is present in single dose in normal males, but in double dose in Xq22.2 CNV males. **c)** CXorf41 (chrX:486,428-106,486,544) fragment is duplicated in Xq22.2 CNV males, as well.

Whereas, both the more internal fragment RBM41/II and the fragment proximal to the END region, CXorf41, resulted effectively duplicated in mutated males, presenting a double dose compared to the single one of normal males (Fig. 14b and c).
These results confirmed that the males of the case-study family in which CGH array analysis detected the Xq22.2 duplication effectively presented a double dose of genomic fragments lying in the central and END regions, contrarily to the normal condition of hemizigosis. However, they even indicated that the START breakpoint of the duplication, as detected by CGH array analysis, needed to be refined more precisely because in this case the presence of the duplication was not confirmed. Furthermore, taken all in consideration, these results proved once more the reliability and necessity of a quantitative assay to validate CGH findings.

To deeply investigate the frequency and pathogenic significance of the novel Xq22.2 duplication we performed a collaboration study, with Dott. Sanna Cherchi (Columbia University), in which we screened a cohort of 14375 individuals (more than 10000 caucasians), previously characterized for the presence of any CNVs.
Eminently, we found 16 individuals harboring CNVs overlapping our Xq22.2 CNV region, of which only 5 were duplications (Fig. 15).
Notably, the genetic content common to all the five duplication overlapping our Xq22.2 duplication is represented by NUP62CL gene.
This result, due to the size of the control cohort considered, confirmed the extreme rarity of Xq22.2 duplication that we found in the case-study family and highlighted a possible significance of NUP62CL gene.
Nonetheless, dealing with a cohort of individuals coming from a different pursued study, these individuals were not selected for having no nephrolithiasis condition, thereby could not be considered as pure controls. We cannot exclude, indeed, that, ICN being a common disease, some of those 16 individuals were eventually affected by ICN. This would have, in case, strengthened more and more the hypothesis of Xq22.2 duplication association to ICN.

Chr	Start (hg18)	End(hg18)	N_SNPs	Size	CN_State	ID	Start_SNP	End_SNP	Conf_Sc
ChrX	106108146	106348183	40	240.038	3	4419829362_R01C01	rs12012022	rs6523935	38,568
ChrX	106125974	106348183	36	222.210	3	4419829620_R02C01	rs17326228	rs6523935	37,516
ChrX	106125974	106430726	124	304.753	3	HYPIMA_C001776	rs17326228	rs6523944	55,355
ChrX	106150815	106623476	58	472.662	3	4331122363_R01C01	rs1285740	rs6622233	40,994
ChrX	106253758	106354408	54	100.651	1	HYPVHP C00C181	rs1285582	rs1936015	33,909
ChrX	106253758	106354408	54	100.651	1	HYPSSA GC0844	rs1285582	rs1936015	26,621
ChrX	106256522	106349061	50	92.540	1	HYPEPE C001419	rs6622156	rs1954366	27,14
ChrX	106264934	106354408	49	89.475	1	HYPVHP C001407	rs1285566	rs1936015	25,729
ChrX	106264934	106354408	49	89.475	1	HYPVHP C001703	rs1285566	rs1936015	38,664
ChrX	106264934	106354408	49	89.475	1	HYPVHP C001790	rs1285566	rs1936015	50,084
ChrX	106282845	106354408	43	71.564	1	HYPVHP C001642	rs2275789	rs1936015	43,723
ChrX	106282845	106354408	43	71.564	1	HYPVHP C001691	rs2275789	rs1936015	33,751
ChrX	106302307	106354408	32	52,102	1	HYPVHP C000898	rs1285574	rs1936015	47,527
ChrX	106302307	106356085	34	53.779	1	HYPVHP C001627	rs1285574	rs6622185	25,317
ChrX	106320296	106516549	66	196.254	3	HYPHCS_C000062	rs7884600	rs6523953	229,51
ChrX	106416698	106516549	15	99.852	0	4485703879_C000062	rs2107066	rs6523953	32,309

Figure 15. List of the 16 CNVs overlapping Xq22.2 CNV detected region. CN_State describes the type of CNV: 0= homozygous deletion, 1= heterozygous deletion, and 3= heterozygous duplication (highlighted in gray).

To better clarify the possible association of the detected duplication with the pathogenesis of ICN, we applied the CNV-profiling assay to screen easily a case cohort of 85 ICN patients for the presence of Xq22.2 duplication. In this case, CNV profiling did not evidenced any duplication in sample individuals.

Thus, this result did not actually prove the effective association of the Xq22.2 duplication, *per se*, with ICN. Even though, it is still possible that the case cohort screened was too small to be realistically representative of an ICN population, also considering that we deal with a very heterogeneous disease both for cause and manifestations (Taylor et al., 2005; Worchester and Coe, 2008).

4.3.2 Whole exome sequencing: chrX CNV detection and validation

As mentioned before, although the primary necessity and achieved goal of WES is to obtain the sequence of an entire exome, subsequently it emerged clearly as it can serve as valuable instrument to afford CNV detection, as well as its breakpoints characterization (Kidd et al., 2008).

In collaboration with prof. Valle (CRIBI, Padova), we performed WES of the proband and subsequently we applied some public and home-made bioinformatic tools to detect the presence of CNVs in the chrX of the proband.

Normally, analyzing WES data, we consider a duplication that unique region in the exome that shows a coverage unusually over the mean coverage, whereas for putative deletions we should observe a signal depletion, that stands for no coverage at all.

In this chrX circle map (Fig. 16) four different traces are plotted to indentify CNVs and discriminate them from false positives. All these traces were obtained measuring the mean coverage signal of a 100,000 base pairs window, with progressive steps of 10,000 bases (overlap region of 90,000 and resolution of 10 Kb).

Considering the CNV trace (Fig. 16), we could observe 4 blue-colored regions from positions 105 to 120 (corresponding to chrX coordinates) that seem to be duplications.

Nonetheless, only the duplications lying nearby positions 106 and 110 are to be considered realistic (Fig. 16). This is true because those are the only two that present a significant signal in both BEST-HIT and UNIQ traces, representing multiple and unique alignments in the exome respectively.

Taken all these results in consideration, we could firstly confirm the Xq22.2 duplication, beforehand identified by CGH array, and gain the previously unknown information about its occurrence in tandem on chrX. This first result identified this duplication with START and STOP at chrX:106,305,001-106,625,001, with 10 Kb resolution (Fig. 17), therefore it seemed shifted of 55,849 base pairs in respect to the precedent identification.

Unexpectedly, we discovered another duplication lying 70 Kb downstream the first one, spanning from positions chrX:110,695,001–110,845,001, for a overall size of 150 Kb (Fig. 17).

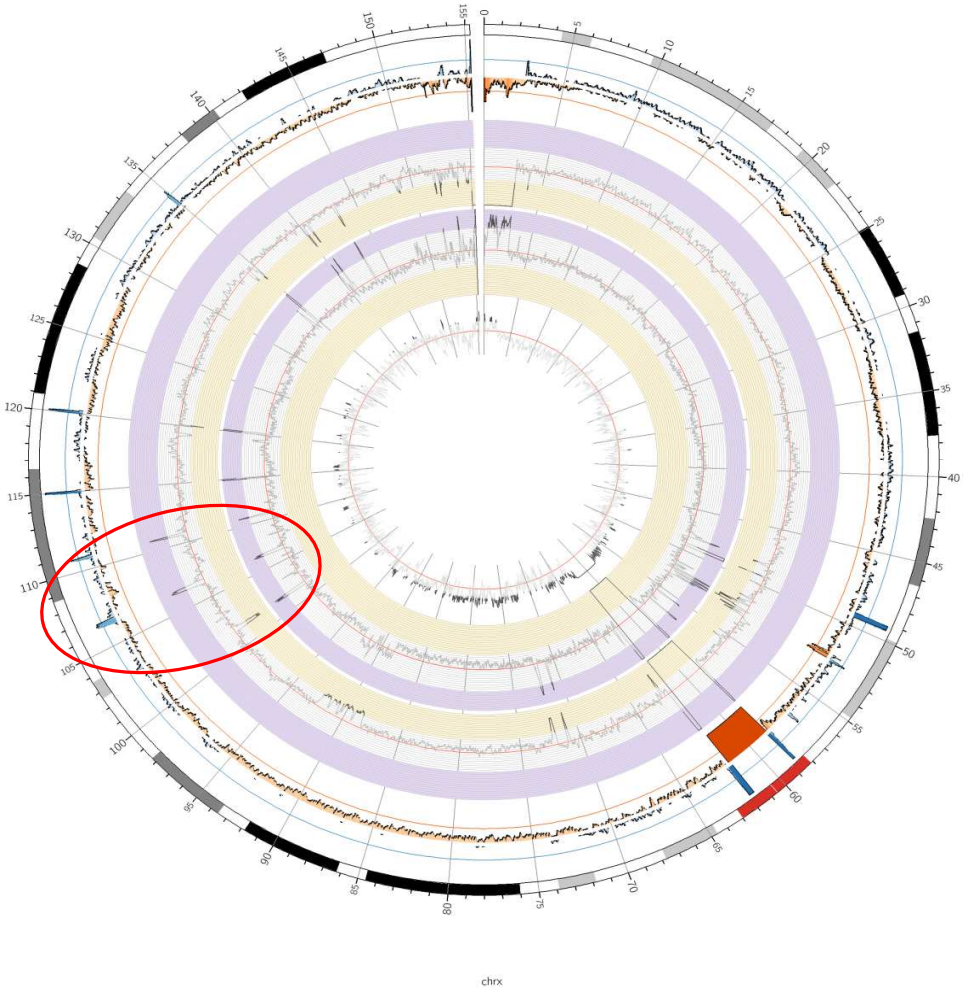


Figure 16. Circle chrX map. WES bioinformatic traces relative to chrX plotted in a circle shape. Starting from the middle of the circle we encounter, in order, 4 traces and for each one the red line indicates the mean signal. The MASK trace plots signal given from the fraction of N bases present in the 100,000 window is measured; peaks that exceed 0,8 value are evidenced (those windows with at least 80% of bases equal to N). The BEST-HIT trace plots the signal given by the "mate pairs" *best-hit* (multiple alignments in the genome); the violet band indicate those value in which the signal exceed the mean signal more than a selected significant threshold, and for the yellow one the *vice versa* is true. The UNIQ trace plots coverage signal given by the "mate pairs" *uniq* (reads that align with a unique position in the genome); the violet and yellow band have the same meaning of *best-hit* trace. The CNV trace normalizes signal arising from *best-hit* + *uniq* trace; regions evidenced in blue should be duplication, whereas orange regions represents deletions.

This result seemed to recall the theory of second-site variants that could exert a synergistic effect (Girirajan et al., 2012). This second duplication spans a region where it is present only one gene whose function it still unknown.

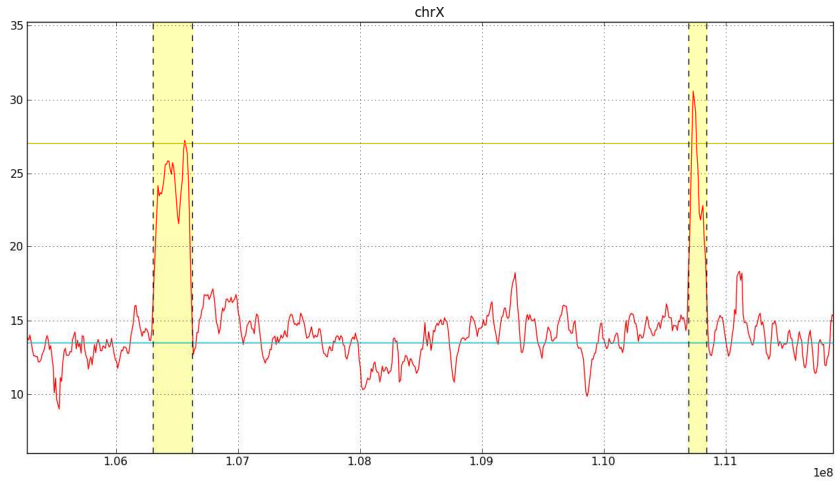


Figure 17. WES coverage in Xq22.2 region. The two yellow highlighted regions present a double coverage in respect to the mean coverage of the whole exome, thus indicating that those fragment sequences are present in double quantity.

Next, we deeply investigated this new finding through CNV-profiling assay and we could confirm the second detected duplication (Fig. 18).

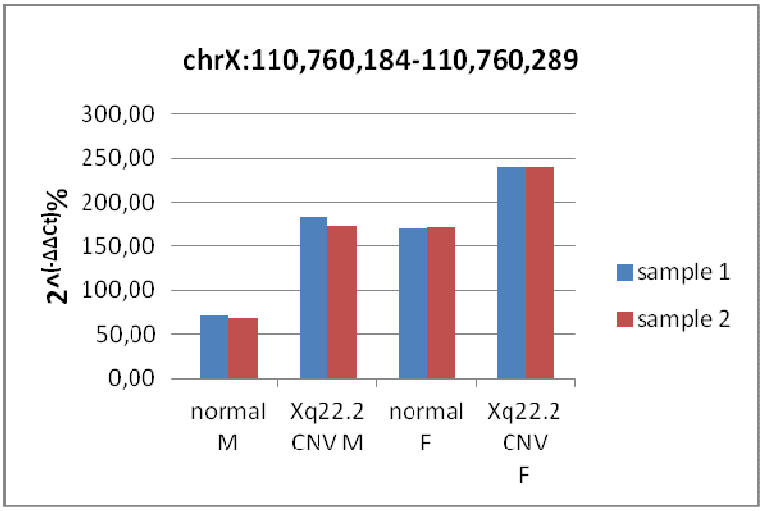


Figure 18. CNV profiling assay applied to the second duplication. Samples 1 and 2 stand for two family members each one for each category (normal males, Xq22.2 CNV males, normal females and Xq22.2 CNV females).

4.3.3 Whole exome sequencing: single nucleotide variant detection and validation

We previously formulated hypothesis about the presence of additional genetic determinants, beyond the chrX duplication inherited from the maternal branch, that could have been inherited from the paternal branch. To answer this question proband's exome was sequenced and investigated for the presence of single nucleotide variants that could act as major susceptibility variants and contribute to the pathogenesis of ICN in this family.
Data analysis took advantage of *QueryOR* bioinformatic platform, developed by prof. Valle (CRIBI, Padova), that allowed a good filtering of all the variants identified (Fig. 19).

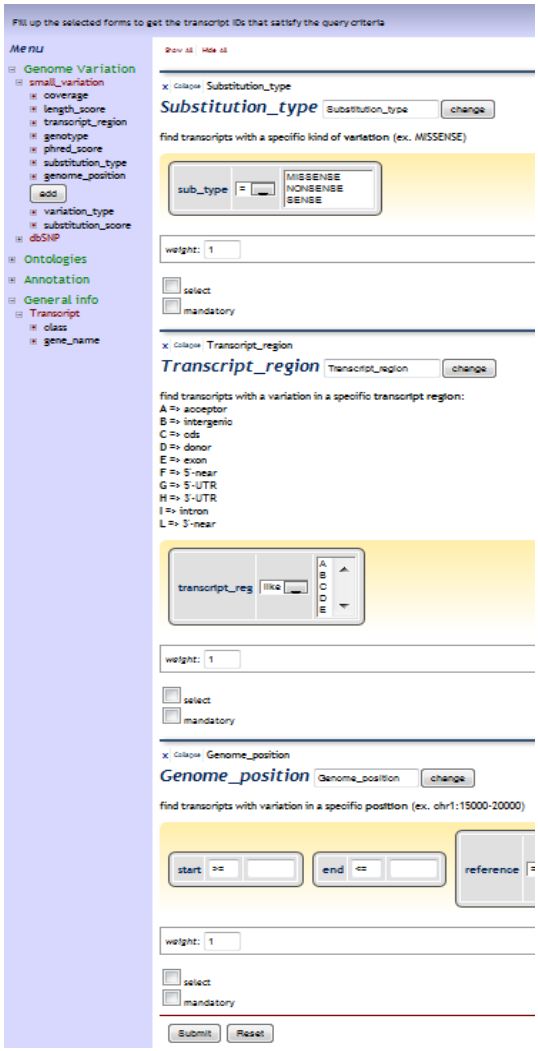


Figure 19. QueryOR example window. QueryOr allows the search starting from a known gene (General info – gene name) or from the type of variant. In particular, for example, we can choose the substitution of interest (sense, missense,nonsense), a particular genome position, the presence or absence in dbSNP database, the kind of aminoacid change (conservative, moderately conservative, radical) and other more useful characteristics that can facilitate the search.

Initially we applied a gene-centered search, that usually starts from the knowledge of genes and their function, by selecting 82 genes from the literature for their known relationship with nephrolithiasis or other different renal pathologies. In this way we could find out 22 sense and 42 missense variants.

As regards sense variants and conservative missense variants, they have been normally believed less impacting on human phenotype, but lately their role in disease pathogenesis has been more and more acknowledged for the presence of ESE (*Exonic Splicing Enhancers*) and ESS (*Exonic Splicing Silencers*) that, if modified by the presence of a sequence variant, can act altering the correct splicing and producing an aberrant mRNA (Sterne-Weiler et al., 2011). For these reasons we focused our attention on two sense variants lying in NCX1 and NUP62CL genes, because they leapt out from our previous investigations, and passed them to the next phase of validation through Sanger sequencing.

The NCX1 sense variant (rs148215685) was confirmed in heterozygosis in the proband and, studying its segregation in the family we could identify the presence in heterozygosis in an affected paternal aunt (Fig. 20). Furthermore, NUP62CL sense variant (rs1285590), that was also confirmed in homozygosis in the proband and in the other male of the family (proband's cousin) that presented Xq22.2 duplication (Fig. 20). This explain why, contrarily to its presence on chromosome X, this variant results effectively in an homozygous condition. We found that NUP62CL harbored also a missense variant (rs1298577) that, even though it presented a *benign/tolerated* prediction on protein level, because of its presence in the neighborhood of the first sense variant investigated (they both presented a MAF value of 0,44), was analyzed for a possible effect on splicing. We used ASSEDA software (*Automated Splice Site and Exon Definition Analysis,* see materials and methods) and it predicted the activation of two criptic acceptor splicing sites of notable strenght. Thus, the potential effect on mRNA splicing need to be clarified.

For missense variants, basing on our prioritization strategies (see materials and methods), we restricted the *variant call set* to six variants lying in four genes: ATP6V1B1, FGF-23, HNF1B, XDH. Thus, all these variants presented a *possibly or probably damaging/ deleterious* prediction (*PolyPhen/SIFT* tools) at the protein level, but not always a low MAF because, even though it represented a criterion of prioritization, it has been observed

that sometimes it can be not actually discriminating (Cassa et al., 2013). Next, the five variants were analyzed in the subsequent phase of molecular validation.

The ATP6V1B1 gene encodes for an ATPase with a known role in the pathogenesis of distal renal tubular acidosis (*dRTA*), that frequently leads to calcium nephrolithiasis. This gene presented two missense variants: one, not known, revealed to be a false positive; whereas the other one (rs11681642), that presented a low coverage and quite high MAF(=0,39), was confirmed in homozygosis. Next, we studied the segregation in the family and we found it in homozygosis in another member of the family (nephew) with US diagnosis of microlithiasis (Fig. 20). In the proband dRTA diagnosis has not been ever postulated, but hypocitraturia, the intermediate phenotype that normally accompanies tubular acidosis (Escobar et al.,2013), was always encountered; the nephew that shared this variant also presented hypocitraturia. In other respect, the variant the we identified and confirmed, even with the worst prediction on protein, appears quite frequent in the normal population indicating that in some way its potential damaging effect is naturally encompassed by other biological mechanisms. This does not change the fact that it could have marginal effects on phenotypes like, for example, hypocitraturia.

The FGF-23 variant, presenting a low coverage, revealed to be a false positive.

Mutations of HNF1B gene are responsible in heterozygosis in the adult of cystic nephropathy sometimes associated with hyperuricemia or hypomagnesemia and hypokalemia (Heidet et al., 2010; Faguer et al., 2011). The variant that we identified (rs144425230) was known, but so rare that it even not presented a quantifiable MAF, thus appearing quite intriguing. This, in spite of his low coverage, was confirmed in heterozygosis in the proband. Lately, by studying its segregation in the family, we could find it present in heterozygosis in an affected paternal aunt (Fig. 20). However, the proband did not present hyperuricemia and hypomagnesemia; but, recently, pre-calyceal renal cysts were observed.

Mutations in XDH gene are normally responsible of xanthinuria, an autosomic recessive disease that leads to xanthine stone formation (Gok et al., 2003). We identified two variants, rs45624433 and rs17011368 with MAF=0,0005 and MAF=0,017 respectively, that were confirmed. The missense variant rs45624433, that had a very bad prediction (*probably damaging/deleterious*) was confirmed in heterozygosis in the proband (Fig. 20).

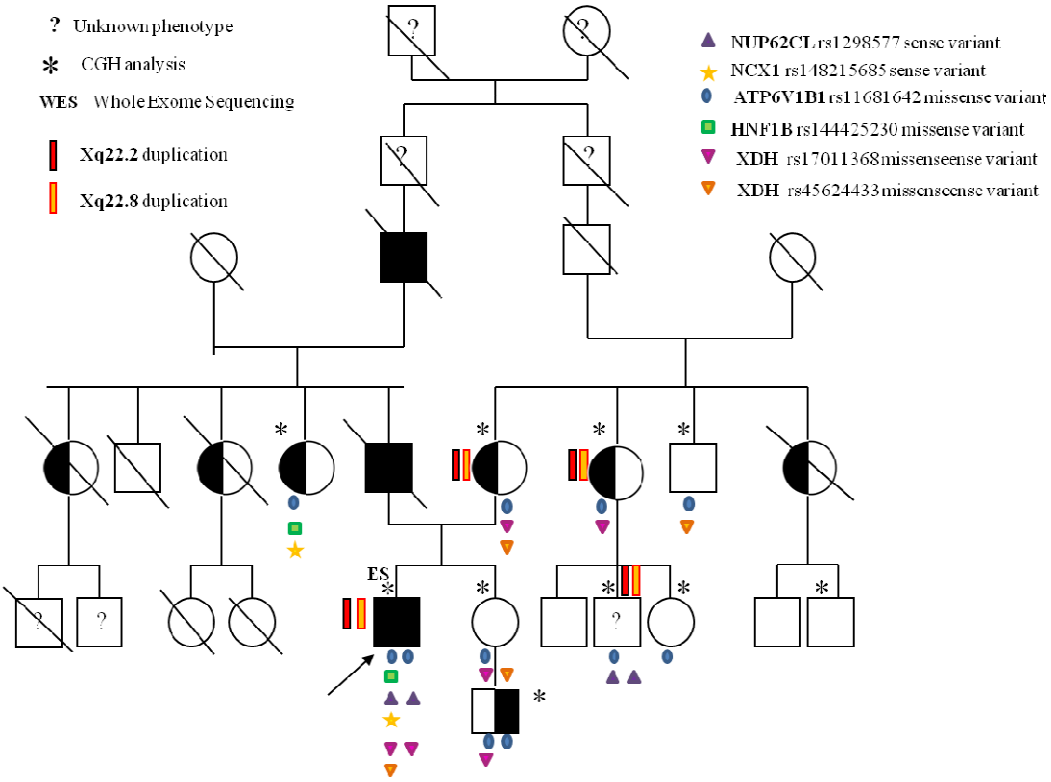


Figure 20. Genealogic tree where copy-number and single-nucleotide variants identified in the proband are illustrated. Their segregation in the case-study family is also shown.

A segregation study showed that it probably came from maternal branch, because it was also present in the mother, the sister and in a maternal uncle (Fig. 20).

Whereas, missense variant rs17011368, that had a milder prediction (*bening/deleterious*), was confirmed in homozygosis, although it was reported by QueryOR software in the heterozygous state (Fig. 20). By analyzing its segregation in the family, we found that it was also inherited from the maternal branch because it appeared present in heterozygosis in the mother, the maternal aunt, the sister and his son (proband's nephew) (Fig. 20). The inheritance from the paternal branch was presumed because of the homozygous mutation state in the proband. However, in the proband neither xanthine in stone composition nor

hypouricemia, intermediate phenotype typical of xanthinuria, were present, but metabolic analysis to assess xanthinuria phenotype should be performed.

Next, we moved to a variant centered search, that is normally more dispersive because it does not take origin from a previous knowledge on genes, thus appearing more like "looking for a needle in a haystack".

Our starting points were: the identification of nonsense variants, because they were normally believed as the most damaging; and not known homozygous missense variants, for both of their novelty and their presumed major impacting effect because of their homozygosis.

As regards nonsense variants, we could identify 25 genes harboring, of these, only seven presented not known or very rare (low MAF values) variants: USP12 and USP29, two ubiquitine peptidase; ALK, a tyrosine kinase associated with cancer; SLC5A9, a sodium-glucose transporter; and FAM108B, ANKR36, ANKR36C, whose function is not well known. As primary focus we pointed the attention to the two ubiquitine peptidase, but USP29 variant revealed to be a false positive while USP12 is still in phase of validation.

In the analysis of missense variant we could evidence 35 genes but, after the usual prioritization strategy (see materials and methods), only 4 genes deserved further investigations: KCNS1, RP1L1,NADK, SHKBP1. We initially passed KCNS1 to the next phase of validation because of its possibly significant role: this gene encodes, indeed, for a potassium channel that belongs to the known family KCN; to this family belongs the most known KCNJ1, with a recognized role in neonatal or late onset Bartter syndrome, characterized by hypercalciuria, nephrocalcinosis and kidney stones (Brochard et al., 2009; Sharma and Linshaw, 2011). In spite of its appealing significance, we could not confirm by Sanger sequencing the two variants identified (both having low coverage).

5. CONCLUSIONS

ICN is a complex disease in which environmental, metabolic and genetic factors play important and sometimes indiscernible roles, from different perspectives.

The importance of hereditary factors in ICN has emerged from a number of studies on families and twins affected by ICN. Computer programs predicted the best inheritance fit with a model of single gene co-dominant model/polygenic model.

Most candidate gene or association studies have produced, up to date, negative or only marginal results, leaving the genetic basis of ICN still an open question. But, genetic analysis techniques are rapidly evolving and promise to improve our knowledge of the genetic basis of nephrolithiasis and allied disorders.

Molecular genetics techniques have evolved considerably in the last decade allowing great strides forward in the study of complex diseases: the discovery of new genetic variants as previously not possible, such as *Copy-Number Variants* (CNV) thanks to *Comparative Genomic Hybridization* (CGH) technology; and the possibility to sequence entire genomes or exomes with a very reduced timeframe and cost, that from the start of a new sequencing era with the completion of Human Genome project in 2001 led to the advent of the revolutionary *Next Generation Sequencing* (NGS) technologies.

In this study we could take advantage of a case-study remarkably peculiar that permitted us to approach ICN genetic basis with these two innovative technologies. This case study was extremely peculiar both for its heterogeneous clinical and metabolic manifestations, ranging from the exceptionally severe phenotype encountered in the proband to very mild phenotypes presented by other family members, both for its clear hereditary component, even being a consanguineous family.

Nevertheless, although dealing with a family in which hereditary factors should play a pivotal role, mimicking a mendelian disorder, we could not find any causative mutation of the ICN phenotype recurring in the family. None of the detected variants – either the private X-linked CNV, either the damaging missense or sense mutations in ICN significant genes - segregated in the family with ICN phenotype. We should remark, indeed, that, despite of their consanguineous origin, family members of this family show very heterogeneous ICN phenotypes. Thus, this heterogeneity can most likely be the

manifestation of an oligogenic or polygenic inheritance. This is clearly manifested in the proband who harbored, in addition to the CNV with its long range effect on the up-regulation of NCX1 gene and the NUP62CL rs1298577 variant in homozygosity, also the ATP6V1B1 rs11681642 and XDH rs17011368 in the homozygous state and rs45624433 in the heterozygous state, as well as HNF1B rs144425230 in the heterozygous state. All these predicted damaging variants in significant genes for the pathogenesis of ICN can act in synergy determining the worst ICN phenotype of the proband and conditioning each other regarding their influence on intermediate phenotype expression.

None of the variants we identified have been previously found associated with the mendelian nephropathies that usually lead to nephrolithiasis, nor they have been previously considered as risk factors for nephrolithiasis, thus they might be considered in future case-control association studies as candidate susceptibility variants for idiopathic calcium nephrolithiasis.

However, these technologies need to be critically evaluated in their ability to answer specific medical and/or biological questions such as posed by our study.

From this point of view, we could definitively prove, that molecular validation of variants identified by WES is always necessary for two reasons: the first, of more immediate comprehension, is that it helps in the identification of false positives due to low coverage regions; the second, less evident but not less important, is that it avoids also false negative results because it can prove the reality of some variants that would have normally been considered not believable on the basis of their coverage.

Discovering the genes involved in ICN hopefully will lead to breakthroughs in pharmacological targets for treating or preventing these conditions, perhaps to tools for diagnosing the risk of developing them, and, finally, but no less importantly, to basic science discoveries and a better understanding of renal/intestinal/bone physiology relevant to ICN.

6. LIST OF ABBREVIATIONS

ALK: anaplastic lymphoma kinase

ANKRD: ankyrin repeat domain

ANKRD36C : ankyrin repeat domain 36 C

ATP6V1B1: ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B1

CaR: calcium sensing receptor

CD-CV: common disease-common variant

CGH: comparative genomic hybridization

CLC-5: chloride channel, voltage-sensitive 5

CLDN2: claudin-2

CNV: copy-number variants

CXorf41: chromosome X open readin frame 41

DGV: database of genomic variants

DSV: DNA sequence variants

ESE: exonic spkicing enhancer

ESS: exonic splicing silencer

ESWL: extracorporeal shock-wave lithotripsy

FAM108B: family with sequence similarity 104, member B

FGF-23: fibroblast grow factor-23

GPR15: G-protein coupled receptor 15

GWAS: genome wide association study

HGP: human genome project

HNF1B: hepatocyte nuclear grow factor

ICN: idiopathic calcium nephrolithiasis

IH: idiopathic hypercalciuria

KCNJ1: potassium inwardly-rectifying channel, subfamily J, member 1

KCNS1: potassium voltage-gated channel, delayed-rectifier, subfamily S, member 1

LCV: large copy-number variants

LMW: low molecular weight

MAF: minor allele frequency

MORC4: MORC family CW-type zinc finger 4

NADK: NAD kinase

NCX1: Na^{+}/Ca^{2+} exchanger 1 (SLC8A1)

NGS: next-generation sequencing

NUP62CL: nucleoporin 62 C-terminal-like

PHPT: primary hyperparathyroidism

PRPS1: phosphoribosyl pyrophosphate synthetase 1

PTH: parathyroid hormone

qPCR: quantitative PCR (real-time PCR)

RBM41: RNA binding motif 41

RP1L1: retinis pigmentosa 1- like 1

RV-CD: rare variant-common disease

SHKB1: SH3KBP1 binding protein 1

SLC5A9: solute carrier family 5 (sodium/glucose cotransporter), member 9

SNP: single-nucleotide polymorphisms

SV: structural variation

USP12: ubiquitine specific peptidase 12

USP29: ubiquitine specific peptidase 29

VDR: vitamin D receptor

WES: whole exome sequencing

WGS: whole genome sequencing

XDH: xanthine dehydrogenase

7. BIBLIOGRAPHY

Ansorge WJ. Next-generation DNA sequencing techniques. N Biotechnol 2009;25:195–203.

Asplin JR. Hyperoxaluric calcium nephrolithiasis. Endocrinol Metab Clin North Am 2002; 31(4):927-49.

Auge BK, Preminger GM. Surgical management of urolithiasis. Endocrinol Metab Clin North Am 2002; 31:1065-82.

Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12:56–68.

Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12(1):56-68.

Barcelo P, Wuhl O, Servitge E et al. Randomized double-blind study of potassium citrate in idiopathic hypocitraturic calcium nephrolithiasis. J Urol 1993;150(6):1761-4.

Bièche I, Olivi M, Champème MH et al. Novel approach to quantitative polymerase chain reaction using real-time detection: application to the detection of gene amplification in breast cancer. Int J Cancer 1998 23;78(5):661-6.

Borghi L, Meschi T, Guerra A et al. Randomized prospective study of a nonthiazide diuretic, indapamide, in preventing calcium stone recurrences. J Cardiovasc Pharmacol. 1993;22 Suppl 6:S78-86.

Brasch-Andersen C, et al. Possible gene dosage effect of glutathione-S-transferases on atopic asthma: using real-time PCR for quantification of GSTM1 and GSTT1 gene copy numbers. Hum Mutat 2004;24:208–14.

Brochard K, Boyer O, Blanchard A et al. Phenotype-genotype correlation in antenatal and neonatal variants of Bartter syndrome. Nephrol Dial Transplant 2009;24(5):1455-64.

Bushinsky DA. Genetic hypercalciuric stone-forming rats. Curr Opin Nephrol Hypertens 1999; 8:479-488.

Cassa CA, Tong MY, Jordan DM. Large numbers of genetic variants considered to be pathogenic are common in asymptomatic individuals. Hum Mutat 2013;34(9):1216-20.

Cheung JY, Zhang XQ, Song J et al. Coordinated regulation of cardiac Na(+)/Ca (2+) exchanger and Na(+)-K (+)-ATPase by phospholemman (FXYD1). Adv Exp Med Biol 2013;961:175-90.

Choi M, Scholl UI, Yue P, et al. K1 channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. Science 2011;331:768–72.

Coe FL, Evan A, Worchester E. kidney stone disease. J Clin Invest 2005.

Coe FL, Parks JH, Asplin JR. Nephrolithiasis. In: Brenner BM, Rector FC Jr. editors. The kidney. 7th edition. Philadelphia: Elsevier; 2004. pp 1819-66.

Coe FL, Parks JH, Evan AP et al. Pathogenesis and treatment of nephrolithiasis. In: Alpern RJ, Herbert SC,editors. Seldin and Giebisch's the kidney.4th edition. Burlington (MA): Academic Press; 2007. pp 1945-77.

Coe FL, Parks JH, Moore ES. Familial idiopathic hypercalciuria. N Engl J Med 1979; 300:337-340.

Coe FL, Parks JH, Moore ES. Familial idiopathic hypercalciuria. N Engl J Med 1979; 300:337-340.

Curhan GC, Willett WC, Rimm EB et al. Familial history and risk of kidney stones. J Am Soc Nephrol 1997; 8:1568-1573.

De Brouwer APM, Duley JA, Christodoulou J. Phosphoribosylpyrophosphate Synthetase Superactivity. 2008 Sep 23 [updated 2013 Aug 08]. In: Pagon RA, Adam MP, Bird TD, Dolan CR, Fong CT, Stephens K, editors.

Debette S, Visvikis-Siest S, Chen MH et al. Identification of cis- and trans-acting genetic variants explaining up to half the variation in circulating vascular endothelial growth factor levels. Circ Res 2011;109(5):554-63.

D'haene B, Vandesompele J, Hellemans J. Accurate and objective copy number profiling using real-time quantitative PCR. Methods 2011;53(3):325.

Durbin RM, Abecasis GR, Altshuler DL, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.

Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 2010;11:446–50.

Escobar L, Mejía N, Gil H et al. Distal renal tubular acidosis: a hereditary disease with an inadequate urinary H^+ excretion. Nefrologia 2013;33(3):289-96.

Evan AP, Coe FL, Lingeman JE et al. Mechanisms of formation of human calcium oxalate renal stones on Randall's plaque. Anat Rec (Hoboken)2007; 290:1315-23.

Evan AP, Lingeman JE, Coe FL et al. Randall's plaque of patients with nephrolithiasis begins in basement membranes of thin loops of Henle. J Clin Invest 2003; 111(5):607-16.

Faguer S, Decramer S, Chassaing N et al. Diagnosis, management, and prognosis of HNF1B nephropathy in adulthood. Kidney Int 2011;80(7):768-76.

Fanciulli M, , Norsworthy PJ, Petretto E et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet 2007; 39:721–3.

Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006;7(2):85-97.

Feuk L, MacDonald JR, Tang T et al. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS Genet 2005;1(4):e56.

Fine JK, Pak CYC, Preminger G. Effect of medical management and residual fragments on recurrent stone formation following shock wave lithotripsy. J Urol 1995;153: 27-33.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C. Copy number variation: new insights in genome diversity. Genome Res 2006 Aug;16(8):949-61.

Gambaro G, Marchini F, Piccoli A et al. The abnormal red-cell oxalate transport is a risk factor for idiopathic calcium nephrolitiasis: a prospesctive study. J Am Soc Nephrol 1996; 7:608-612.

Gambaro G, Vezzoli G, Casari G et al. Genetics of hypercalciuria and calcium nephrolithiasis: from the rare monogenic to the common polygenic forms. Am J Kidney Dis. 2004; 44:963-86

Girirajan S, Rosenfeld JA, Coe BP et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med 2012;367(14):1321-31.

Gok F, Ichida K, Topaloglu R. Mutational analysis of the xanthine dehydrogenase gene in a Turkish family with autosomal recessive classical xanthinuria. Nephrol Dial Transplant 2003;18(11):2278-83.

Goldfarb DS, Fisher ME, Keich Y et al. A twin study of genetic and dietary influences on nephrolithiasis: a report from the Vietnam Era Twin (VET) Registry. Kidney Int 2005; 67(3): 1053-61.

Gonzalez E, Kulkarni H, Bolivar H et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 2005; 307:1434–40.

Goodman HO, Holmes RP, Assimos DG. Genetic factors in calcium oxalate stone disease. J Urol 1995; 153:301-307.

Heidet L, Decramer S, Pawtowski A et al. Spectrum of HNF1B mutations in a large cohort of patients who harbor renal diseases. Clin J Am Soc Nephrol 2010;5(6):1079-90.

Kathiresan S, Melander O, Guiducci C et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet 2008;40(2):189-97.

Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. Nature 2008;453(7191):56-64.

Klassen T, Davis C, Goldman A, et al. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. Cell 2011;145:1036–48.

Kumar V, Lieske JC. Protein regulation of intrarenal crystallization.Curr Opin Nephrol Hypertens 2006; 15:374-80.

Lander ES, Linton LM, Birren B et al., International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860-921.

Lefever S, Vandesompele J, Speleman Fet al. RTPrimerDB: the portal for real-time PCR primers and probes. Nucleic Acids Res 2009;37(Databaseissue):D942-5.
Mardis ER. Next-Generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387–402.

Marian AJ, Belmont J. Strategic approaches to unraveling genetic causes of cardiovascular diseases. Circ Res 2011;108(10):1252-69.

Marian AJ. Nature's genetic gradients and the clinical phenotype. Circ Cardiovasc Genet 2009;2(6):537-9.

Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A 1977;74:560–4.

McGeown MG. Hereditary in renal stone disease. Clin Sci 1960; 19:465-471.

Medvedev P, Fiume M, Dzamba M et al. Detecting copy number variation with mated short reads. Genome Res 2010;20(11):1613-22.

Mehes K, Szelid Z. Autosomal dominant inheritance of hypercalciuria. Eur J Pediatr 1980; 133:239-242.

Mehes K, Szelid Z. Autosomal dominant inheritance of hypercalciuria. Eur J Pediatr 1980; 133:239-252, 1980.

Mills RE, Walter K, Stewart C, et al. Mapping copy number variation by population-scale genome sequencing. Nature 2011;470:59–65.

Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 2010;42:790–3.

Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2010;42: 30–5.

Nielsen R, Paul JS, Albrechtsen A et al. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 2011;12(6):443-51.

Oetting WS. Exome and genome analysis as a tool for disease identification and treatment: the 2011 Human Genome Variation Society scientific meeting. Hum Mutat 2012;33(3):586-90.

Pak CYC. Physiological basis for absorptive and renal hypercalciuria. Am J Physiol 1979;237:F415-F423.

Pearce SH, Williamson C, Kifor O et al. A familial syndrome of hypocalcemia with hypercalciuria due to mutations in the calcium-sensing receptor. N Engl J Med 1996;335:115-1122.

Pennisi E. Genomics. 1000 Genomes Project gives new map of genetic diversity. Science. 2010;330(6004):574-5.

Perry GH, Ben-Dor A, Tsalenko A et al. The fine-scale and complex architecture of human copy-number variation. Am J Hum Genet 2008;82(3):685-95.

Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 2001;29(9):e45.

Pinkel D, Segraves R, Sudar D et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998;20(2):207-11.

Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant or not? Hum Mol Genet 2002;11:2417–23.

Redon R, ishikawa S, Fitch KR et al. Global variation in copy number in the human genome. Nature 2006;444:444–54.

Resnick M, Pringen DB, Goodman HO. Genetic predisposition to formation of calcium oxalate renal calculi. N Engl J Med 1968; 278:1313-1318.

Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7.

Robertson WG, Longhorn SE, Whitfield HN et al. The changing pattern of the age at onset of urinary stone disease in the UK, in Borghi L, Meschi T, Briganti A, Schianchi T, Novarini A (eds): Proceedings of the 8th European Symposium on Urolithiasis. Cosenza, Italy, Editoriale 1999. pp165-168.

Robertson WG, Peacock M, Baker M et al. Studies on the prevalence and epidemiology of urinary stone disease in men in Leeds. Br J Urol 1983; 55:595-598.

Saigal CS, Joyce G, Timilsina AR. Urologic Diseases in America project. Direct and indirect costs of nephrolithiasis in an employed population: opportunity for disease management. Kidney Int 2005; 68:1808-14(A).

Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 1975;94:441–8

Scherer SW, Lee C, Birney E et al. Challenges and standards in integrating surveys of structural variation. Nat Genet 2007;39(7 Suppl):S7-15.

Scott P, Ouimet D, Proulxy et al. The α-hydroxylase locus is not linked to calcium stone formation or calciuric phenotypes in French-Canadian families. J Am Soc Nephrol 1998; 9:425-432.

Scott P, Ouimet D, Valiquette L, et al. Suggestive evidence for a susceptipility gene near the vitamin D receptor locus in idiopathic calcium stone formation. J Am Soc Nephrol 1999. 10:1007-1013.

Sebat J, Lakshmi B, Troge J et al. Large-scale copy number polymorphism in the human genome. Science 2004;305:525–8.

Sharma A, Linshaw MA. A novel compound heterozygous ROMK mutation presentingas late onset Bartter syndrome associated with nephrocalcinosis and elevated 1,25(OH)(2) vitamin D levels. Clin Exp Nephrol 2011;15(4):572-6.

Simon DB, Bindra RS, Mansfield TA et al. Mutations in the chloride channel gene, CLCNKB, cause Bartter's syndrome type III. Nat Genet 1997 ;17(2):171-8.
Sterne-Weiler T, Howard J, Mort M et al. Loss of exon identity is a common mechanism of human inherited disease. Genome Res 2011;21(10):1563-71.

Sivakamasundari P, Varalakshmi P, Kannabiran M. Nuclear pore complex oxalate binding protein p62: its expression on oxalate exposure to VERO cells. J Cell Biochem 2004;93(6):1099-106.

Stranger BE, Forrest MS, Dunning M et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007;315:848–53.

Taylor EN, Stampfer MJ, Curhan GC. Obesity, weight gain, and the risk of kidney stones. JAMA 2005; 293 (4):455-62.

Tessier J, Petrucci M, Trouve ML et al. A family-based study of metabolic phenotypes in calcium urolitiasis. Kidney Int 2001; 60:1141-1147.

Trinchieri A, Mandressi A, Luongo P et al. Familial aggregation of renal calcium stone disease. J Urol 1988; 139:478-481.

Tuzun E, Sharp AJ, Bailey JA et al. Fine-scale structural variation of the human genome. Nat Genet 2005;37(7):727-32.

Watson JD. The human genome project: Past, present, and future. Science 1990;248:44–9.

Wheeler DA, Srinivasan M, Egholm M et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008;452(7189):872-6.

Worcester EM, Coe FL. Nephrolithiasis. Prim Care. 2008 Jun; 35(2):369-91, VII.

Worchester EM, Gillen DL, Evan AP, et al. Evidence that postprandial reduction of calcium reabsorption mediates hypercaliuria of patients with calcium nephrolithiasis. Am J Physiol Renal Physiol 2007; 292(1):F66-75.

Zhou X, Ren L, Li Y et al. The next-generation sequencing technology: A technology review and future perspective. Sci China Life Sci 2010;53:44–57.