Genetic Analysis of Renal Diseases (Glomerular and Renal Cystic Diseases*) in Children, Adolescents and Adults in Asia

The DRAGoN 2 Study

Deciphering diversities: **R**enal **A**sian **G**enetics **N**etwork

National University of Singapore (NUS), Singapore
Genome Institute of Singapore (GIS), A*STAR, Singapore

Master Protocol: November 2020, Version 14GC

**This Protocol pertains to BOTH glomerular and renal cystic diseases. For studies involving ONLY glomerular disease or ONLY renal cystic diseases, please refer to other protocols.*

CONTENTS

SUMMARY OF VERSION UPDATES

Changes in version Nov20 v14 from version Aug19 v13

1. Updated the list of Scientists
2. Updated the list of Study Sites and Investigators

Included the variant curation algorithm under "Next generation sequencing (NGS)".

Changes in version Aug19 v13 from version Mar19 v12

1. List of investigators updated (page 7)
2. **Under "Definitions",**
    a. **Definition of nephrotic syndrome: serum albumin level is changed from <25g/L to <30g/L**
    b. **Table for proteinuria is edited to show the preference for urine protein investigations**
    c. **Calcineurin inhibitor resistance changed from 6 weeks of drug dosing to 6 months. Minimum cyclosporin trough changed from 150 ng/ml to 50 ng/ml and minimum tacrolimus trough changed from 5 mg/L to 2mg/L.**
3. Literature review in sections 1a, 3a, 3b is updated
4. Target number of subjects (4a) updated
5. **Inclusion criteria (4b) changed to exclude late-onset steroid-resistance.**
6. Updated on sections 4d (subject withdrawal)
7. Sections 6d, 6f, 6g (study procedures) are edited to make it clearer to reader.
8. Definitions of unmodified derivatives, progeny, modifications are added in section 6l. Elaborations on how we will use residual unused blood samples are added in section 6l.
9. *CRB2* and *SGPL1* are added in the list of genes in 7a.
10. Elaborations on how deidentified data will be handed are added in Section 9a
11. Return of genetic test results and compensation for injury in sections 10 and 13 are updated. Potential benefits for participating in this study are added in section 11.
12. We added how hard copy study documents will be handled 10 years after completion of the study in sections 12a and 16.

Changes in version "Mar19 v12" from version "Jan19 v11"

1. All details on saliva collection had been removed.
2. All "Main Coordinating Site" had been standardized to "Main Coordinating Centre".
3. Under 6d (ii), it had been added that "In addition, if further tests of gene function are necessary, a subsequent or repeat blood sample may be necessary."
4. Information on the phenotype data collection in 6d (ii) had been edited: "The collected data would be a useful reflection of long-term outcomes including progression to end-stage renal failure and death. For this reason, the duration of phenotype follow-up data by site investigators should be for at least 5 years, whenever possible."
5. The subjects' initial cannot be provided to the Main Coordinating Centre. This information had been edited in 6i.
6. The signing of Material Transfer Agreement is added in 6l: "A Material Transfer Agreement will be signed between the Main Coordinating Centre and the Site investigators according to the terms discussed above."
7. Details on the storage of unused residual blood samples and its disposal at National University of Singapore, Laboratory of Paediatrics have been added in 6l.
8. It was added in 7c that "the additional blood samples taken from the affected subjects will be used to isolate white blood cells which can stably proliferate in the laboratory."

9. It was edited in 9a that "The study data in its deidentified form will be stored for a longer period of 30 years."
10. It was added in 11 that "Each subject and their recruited immediate family would be provided with genetic counselling as to make them understand the implication and possibility that might arise in the future (if any discovery is made)."
11. It was added in 16 that "All study data in its deidentified form are to be kept 30 years after the completion of the study."

Changes in version "Jan19 v11" from version "Mar18 v10"

12. Address has been added for study sponsors.
13. The biostatistician was updated to "Rajesh Babu Moorakonda".
14. The target number of subjects to be recruited from the Malaysian sites has been added.
15. The details on subject withdrawal have been updated.
16. Urine collection has been deleted from the 6d (ii).
17. Details on the collection of phenotype data based on patients' clinical visits have been added in 6d (ii).
18. Under 6g (ii), it is added that "If the positive results are deemed worthy of a physician referral (at discretion of the primary physician), they will be referred."
19. Under 9a, it has been added that data will be stored for 20 years.

Changes in version "Mar18 v10" from version "Oct 17 v9GC "

20. The list for the study sites and investigators have been updated for Malaysia and Pakistan.
21. The algorithm for genetic tests has been changed to targeted gene sequencing and/or exome sequencing.

STUDY SPONSORS

National Medical Research Council, Singapore
(NMRC/CSA-INV/0015/2017)
#09-66, Harbourfront Centre
1 Maritime Square (Lobby C)
Singapore 099253

Research Grants:

1. Next Generation Sequencing and Mechanistic Understanding for the Diagnosis and Management of Primary Glomerular Disease in Singapore (NMRC/CSA/0057/2013) (Nov 2013 to Jul 2017)
2. Next Generation Sequencing and Mechanistic Understanding for the Diagnosis and Management of Primary Glomerular Disease in Asia (NMRC/CSA-INV/0015/2017) (Aug 2017 to Aug 2021)

WEBSITE

http://www.scri.edu.sg/crn/deciphering-diversities-renal-asian-genetics-network-dragon/deciphering-diversities-renal-asian-genetics-network-dragon/

CONFLICT OF INTERESTS

There are no conflict of interests for study investigators.

INVESTIGATORS

Main Coordinating Centre

Khoo Teck Puat – National University Children's Medical Institute,
National University Health System, National University of Singapore
Level 12 NUHS Tower Block, 1E Kent Ridge Road, Singapore 119228
Paediatric Nephrology
Tel: +65-6772 4411
Fax: +65-6779 7486

Main Coordinating Centre Investigators

Principal Investigator	Dr Ng Kar Hui (Email: kar_hui_ng@nuhs.edu.sg)
Co-Investigators	Prof Yap Hui Kim
	Dr Isaac Liu

Scientists

Dr Khor Chiea Chuen
Genome Institute of Singapore
60 Biopolis Street, Genome, #02-01
Singapore 138672

Dr Sonia Davila
SingHealth Duke-NUS Institute of Precision Medicine (PRISM)
5 Hospital Drive, National Heart Research Centre, Level 9

Pathologists

A/Prof Tan Puay Hoon
Head and Senior Consultant
Department of Pathology
Singapore General Hospital

Dr Alwin Hwai-Liang Loh
Senior Consultant
Department of Pathology
Singapore General Hospital

Dr Ahmed Syed Salahuddin

Singapore 196609

Consultant
Department of Pathology
Singapore General Hospital

Dr Bruno Reversade
A*STAR Institute of Medical Biology
8A Biomedical Grove, Singapore 138648

Alvin Ng Yu Jin, Poon Kok Siong and Karen ML Tan
Molecular Diagnosis Centre
Department of Laboratory Medicine
National University Hospital

Scientific Collaborator
Dr Lisa Guay-Woodford
Children's National Medical Center
Washington DC, USA

Clinical Geneticist
Dr Beata S Lipska-Zietkiewicz
Clinical Genetics Unit
Medical University of Gdansk
Poland

Research Support Staff and Database Managers
Rajesh Babu Moorakonda
Principal Statistical Analyst
Singapore Clinical Research Institute

Ms Ng Jun Li (jun_li_ng@nuhs.edu.sg)
Ms Liang Ai Wei
Dr U Mya Than (mya_than@nuhs.edu.sg)
Mdm Wee May Lin Vivien

Tel: +65-6601 3307
Fax: +65-6779 7486

Committee members

Sub committee	Members
Glomerular disease	Dr Ng Kar Hui (Singapore) Dr Suprita Kalra (India) Dr Yap Yok Chin (Malaysia) Dr Cho Cho San (Myanmar) Dr Withanage Dona Vindya Nandani Gunasekara (Sri Lanka) Dr Iftikhar Ijaz (Pakistan) Dr Sadaf Asim (Pakistan)

Cystic disease	Dr Isaac Liu (Singapore) Assoc Prof Syed Saimul Huque (Bangladesh) Dr Kiran P. Sathe (India) Dr.Vaishali B. More (India) Dr Yap Yok Chin (Malaysia)
Networking	Assoc Prof Syed Saimul Huque (Bangladesh) Dr Suprita Kalra (India) Dr Sadaf Asim (Pakistan)
Advisors	Professor Franz Schaefer (Germany) Professor Yap Hui Kim (Singapore)

Study Sites and Investigators

Country	**Institution**	**Investigators**
Bangladesh	Bangabandhu Sheikh Mujib Medical University	Assoc Prof Syed Saimul Huque
	Information Center in Bangladesh	Dr Niaz Md Sharif
India	1) Mehta Children's Hospitals, Chennai	Dr Ekambaram Sudha
	2) Sir H.N Reliance Foundation Hospital & Research Center, Mumbai, India	Dr Kiran P Sathe
	3) KLE University's JN Medical College, Belgaum and KLE Dr.Prabhakar Kore Hospital & MRC, Belgaum	Dr Mahantesh V. Patil Dr Preeti Inamder
	4) Armed Force Medical College, Pune	Dr Suprita Kalra
	5) SRCC Children's Hospital S.L.Raheja Fortis Hospital	Dr.Vaishali B. More
Malaysia	1) Hospital Kuala Lumpur, Malaysia	Dr Yap Yok Chin Dr Mirunalini Appadurai Dr Yap Suet Li Dr Tengku Hasnita Tengku Hussain Dr Tham Jia Yi
	2) Hospital Tuanku Jaafar Seremban	Dr Lee Ming Lee Dr Caroline Eng Siew Yin
	3) Hospital Sultan Ismail, Johor	Dr Susan Pee
	4) Hospital Tengku Ampuan Afzan, Kuantan	Dr Selva Kumar Sivapunniam
	5) Hospital Wanital Eanita Dan Kanak-Kanak Sabah, Hospital Likas	Dr Lai Chee Sing
	6) University Malaya Medical Centre, Kuala Lumpur	Dr Karmila Abu Bakar
Myanmar	1) Yangon children Hospital, Yangon	Dr Cho Cho San Dr Khin Le Le Shein

	2) Renal Unit, 300 hundred bedded & 500 bedded children hospital, Mandalay	Dr Chaw Su Khine Dr Kyaw Thuya
	3) Children Ward, Women & Children Hospital, Taungyi	Dr Khin Moh Moh
	4) Yankin Children Hospital, Yangon	Dr Khin Myat Lwin Nyein
Singapore	1) National University Health System National University of Singapore	Dr Ng Kar Hui Professor Yap Hui Kim Dr David Lu Dr Isaac Liu Dr Lau Yew Weng Perry Dr Jimmy Teo
	2) KK Women's & Children's Hospital (KKH)	Dr Ng Yong Hong A/Prof Sing Ming Chao Dr Indra Ganesan Dr Siew Le Chong
Sri Lanka	Lady Ridgeway Hospital For Children, Colombo, Sri Lanka	Dr Withanage Dona Vindya Nandani Gunasekara
Pakistan	National Institute of Child Health, Karachi	Dr Khemchand N Moorani Dr Sadaf Asim Dr Bilqis Naeem
	The Kidney Center, Post-Graduate Training Institute	Dr Khemchand N Moorani Dr. Harnam Hotchandani
	King Edward Medical University	Dr Iftikhar Ijaz
	The Children Hospital and The Institute of Child Health, Multan, Pakistan	Dr Muhammad Imran Dr Khuram Rashid Dr Hashim Raza
Philippines	University of Santo Tomas Hospital	Dr Maria Rosario Cabansag Dr. Angeli Montemor
	National Kidney and Transplant Institute	Dr Ma. Angeles G. Marbella Dr Anne Margaret Canapi
Vietnam	Children's Hospital 1	Dr Nguyen Duc Quang
	Children's Hospital 2	Dr Tan M Nguyen Dr Diem Thuy Dr Trong Thi Nguyen Huynh Dr Dien Duong MD Huynh Ngoc Linh
	National Hospital of Pediatrics	Dr Chị Hương Dr Kien Nguyen

ABBREVIATIONS

ADPKD	Autosomal dominant polycystic kidney disease
ARPKD	Autosomal recessive polycystic kidney disease
CKD	Chronic kidney disease
DMSO	Dimethyl sulfoxide
DPMGN	Diffuse proliferative membranous glomerulonephritis
EDTA	Ethylene-diamine-tetra-acetic acid
eGFR	estimated glomerular filtration rate
ESRD	End Stage Renal Disease
FBS	Fetal bovine serum
FSGS	Focal segmental glomerulosclerosis
GBM	Glomerular basement membrane
GN	Glomerulonephritis
GWAS	Genome-wide Association Studies
hpf	high power field
HRFD	Hepatorenal fibrocystic disease
IgM	Immunoglobulin M
NGS	Next Generation Sequencing
NPHP-RC	Nephronophthisis-related ciliopathy
NSS	Not steady state
PBS	Phosphate-buffered saline
RBC	red blood cells
RPMI	Roswell Park Memorial Institute medium
SDNS	steroid-dependent nephrotic syndrome
SRNS	steroid-resistant nephrotic syndrome

DEFINITIONS

TERM	DEFINITION
Nephrotic syndrome	Edema Nephrotic-range proteinuria Serum albumin <30g/L
Congenital **Infantile**	Onset of nephrotic syndrome in first 3 months of life. Onset of nephrotic syndrome in 4th to 12th month of life.

Proteinuria

TERM	Proteinuria		
	First-line Urine protein: creatinine ratio	**Second-line** Dipstick	**Third-line** 24h urine protein
Normal	< 2 years old <0.06 g/mmol <0.6 mg/mg or g/g >= 2 years old <0.02 g/mmol <0.2 mg/mg or g/g	Neg or trace	<0.3 g/day/1.73m^2 <170 mg/m^2/day <7 mg/m^2/hr
Subnephrotic range	< 2 years old 0.06-0.6 g/mmol 0.6-6.0 mg/mg or g/g >= 2 years old 0.02-0.2 g/mmol 0.2-2.0 mg/mg or g/g	1+	0.3-3.0 g/day/1.73m^2 170-1700 mg/m^2/day 7-70 mg/m^2/hr
Significant	>= 2 years old ≥0.07 and <0.2 g/mmol ≥0.7 and <2.0 mg/mg or g/g		≥1 and <3 g/day/1.73m^2
Nephrotic range	< 2 years old > 0.6 g/mmol > 6.0 mg/mg or g/g >= 2 years old > 0.2 g/mmol > 2.0 mg/mg or g/g	>=2+	> 3.0 g/day/1.73m^2 > 1700 mg/m^2/day > 70 mg/m^2/hr

Relapse / Remission (applies to all immunosuppressive and anti-proteinuric drugs)

	Proteinuria (see definition of Proteinuria)	Dipstick (3 consecutive days)	Albumin (g/L)	Edema
Complete remission	No proteinuria	Neg or trace	>30	None
Partial remission Partial relapse	Proteinuria	≥1+	>30	None
Complete relapse	Nephrotic range	≥2+	<30	Edema

Steroid response **Sensitive** **Resistant**	Complete remission after steroid treatment given alone (not combined with other drugs). No complete remission after 4 weeks of 60mg/m^2/day prednisolone or prednisone or equivalent
Initial resistance	No complete remission after 4 weeks of 60mg/m^2/day prednisolone (or prednisone or equivalent) *started after the first presentation/ manifestation of disease.*
Late resistance	No complete remission after 4 weeks of 60mg/m^2/day prednisolone (or prednisone or equivalent) *after an initial period of steroid-responsiveness*.
Steroid equivalents	Prednisolone 1mg = Hydrocortisone 4mg = Methylprednisolone 0.8mg = Prednisone 1mg
Calcineurin inhibitor resistance	No complete remission after 6 months of cyclosporine A (12-hourly dosing) or tacrolimus (12-hourly dosing) with:

	minimum parent cyclosporine trough (C_0) level of 50ng/ml or ug/L, AND/OR minimum tacrolimus trough level of 2ng/ml or ug/L.
Combined steroid and calcineurin inhibitor resistance	No complete remission after **4 weeks** of 60mg/m^2/day **prednisolone** (or prednisone or equivalent) **combined with** at least **6 months of oral cyclosporine** (achieving trough levels of at least 50ug/L or ng/ml) or **oral tacrolimus** (achieving trough levels of at least 2mg/L).
Onset of disease	First presentation/manifestation of glomerular diseases, and may include incidental finding of microscopic haematuria or proteinuria.

ABSTRACT

Primary glomerular disease, including focal segmental glomerulosclerosis (FSGS), is a common cause of renal failure in adults and children worldwide. Current treatment is empirical and involves immunosuppressive drugs which are toxic, expensive and with only moderate response rates. Proteinuric or familial haematuric diseases are traditionally monogenic diseases, involving as many as 40 known genes. With new sequencing techniques, concomitant variants in different genes are increasingly recognised as important. With the discovery of *MYH9* and *APOL1* as significant risk alleles for FSGS in African Americans, FSGS can also be a polygenic disorder.

There have been evidence that Asians are different genetically. *NPHS1* and *NPHS2* mutations are less prevalent in Asia compared to Europe and the Middle East. The *NPHS2* functional p.R229Q polymorphism is strikingly almost absent in Asians.
In parallel, autosomal recessive polycystic kidney disease (ARPKD) and nephronophthisis-related ciliopathies (NPHP-RC) have in common genetic basis involving proteins that are critical in the structure/function of the primary apical cilium, and have been reclassified as hepatorenal fibrocystic kidney diseases (HRFD).

There has been recent evidence that suggested there may be as yet undiscovered pathogenic genes underlying the HRFD phenotype lying amongst the families in Asia. In addition, a precise genetic diagnosis has important implications for prognostication, as well as family planning and planning for a renal transplant.

Gaps in the current knowledge include undiscovered novel genes, and a generally understudied Asian population. Additionally, tedious conventional sequencing methods are often unable to identify concomitant variants in different genes, and there is no reliable method to predict response to therapy.

We aim to recruit 800 patients with primary glomerular disease and 500 patients with renal cystic disease in Asia and perform the following:

(i) Next Generation Sequencing (NGS) in patients with primary proteinuric glomerular disease which are either familial, onset at less than 3 years old, or resistant to all forms of immunosuppressive therapy, to identify disease-causing variants in known or novel genes.
(ii) Next Generation Sequencing (NGS) in HRFD patients to identify mutations in known and novel ciliopathy related genes.
(iii) Genome-Wide Association Studies (GWAS) in patients with primary sporadic FSGS, to identify genetic variants that may be important in disease susceptibility.
(iv) Preliminary in vitro functional work on identified variants.

Understanding the pathogenesis of these diseases will lead to more directed treatment strategies, which will translate to less side effects and lower healthcare costs. Identification of risk alleles for the FSGS in Asia will allow for estimation of lifetime risks for kidney disease, as well as genetic risk stratification for disease course and response to therapy.

1. BACKGROUND

a) General Introduction

Glomerular disease/Nephrotic syndrome

Primary proteinuric glomerular disease is a common cause of end-stage renal disease in adults and children worldwide. Patients can present with nephrotic syndrome, significant proteinuria, persistent microscopic or recurrent gross haematuria, or chronic renal impairment. Underlying histological aetiologies can include focal segmental glomerulosclerosis (FSGS) and membranous nephropathy.[1,2] Among these, FSGS is the more common histological finding. Classically, FSGS is a histological term describing focal (involving some, but not all, glomeruli) obliteration of capillary lumina by matrix in a segmental pattern (affecting only a portion of the glomerular tuft). FSGS may be primary, caused by genetic defects or immune-related factors such as circulating factors; or secondary caused by other insults, such as infections, drug use and secondary maladaptive responses. Patients with primary FSGS typically present with heavy proteinuria, steroid-dependent (SDNS) or steroid-resistant nephrotic syndrome (SRNS).[1]

Current treatment of primary glomerular diseases is often empirical and involves the use of several concomitant immunosuppressive agents. Response to these highly toxic and expensive drugs is often limited to a small proportion of patients. Indeed, up to 48% of the children with primary nephrotic syndrome will be steroid-dependent or steroid-resistant.[3] Second-line therapy include calcineurin inhibitors such as cyclosporine or tacrolimus. For the refractory cases, mycophenolate mofetil or biological agents such as the anti-CD20 monoclonal antibody, rituximab, may be considered. However, such drugs may not be available to many patients due to costs, and those who do receive such drugs often suffer from chronic morbidities related to the immunosuppressive therapy such as severe infections and malignancy.

There are currently no methods to identify the patients who are likely to respond to immunosuppressive therapy. Clinicians therefore often use a "trial of therapy" strategy in those with no significant renal impairment. Understanding the pathogenesis of these diseases will lead to more directed treatment strategies, which will translate to less side effects and lower healthcare costs.

Up to 30-40% of patients with FSGS progress to ESRD within ten years of onset.[4,5] The poor prognosis despite treatment is due to the lack of understanding of the pathogenesis of these proteinuric glomerular diseases, resulting in the empirical and non-targeted nature of current treatment protocols.

There have been evidence that Asians are different genetically. *NPHS1* and *NPHS2* mutations are less prevalent in Asia compared to Europe and the Middle East. The *NPHS2* functional p.R229Q polymorphism is strikingly almost absent in Asians.

Hepatorenal fibrocystic diseases

Autosomal recessive polycystic kidney disease (ARPKD)[6,7] and associated disorders such as nephronophthisis (MIM 256100), Joubert syndrome (MIM 213300), Bardet–Biedl syndrome (MIM 209900), Meckel–Gruber syndrome (MIM 249000), and oro- facial-digital syndrome I (MIM 311200) have in common the clinical features of variable degrees of renal cystic disease and can be associated with dysgenesis of the portobiliary tract in the form of congenital hepatic fibrosis and/or Caroli disease. Due to their common clinical features, they have been reclassified as hepatorenal fibrocystic diseases (HRFD)[8]. In addition, it has emerged in recent years that they have in common genetic basis involving proteins that are critical in the structure/function of the primary apical cilium. As such, they form a subset of disorders referred to as "ciliopathies".

The classic phenotype-genotype correlation in ARPKD has been that patients with 2 truncating mutations in *trans* do poorer than patients with a mutation. However in recent years it has been found that intronic mutations, by affecting splicing, can result in the ARPKD phenotype[9]. Mutations in *DZIP1L* have been found to cause ARPKD as well[10,11].

Furthermore there is a considerable overlap between ARPKD, its phenocopies and ADPKD. For example, an ultrasound series found that up to 25% of ADPKD cases diagnosed *in utero* were misdiagnosed initially as ARPKD as the features mimicked the latter[12]. In another series, all apparent ARPKD patients without point mutations in *PKHD1* were found to be phenocopies[13]. For this reason, consensus guidelines have recently been released for to aid the differential diagnosis of ARPKD, its phenocopies and ADPKD[14]. However, a definitive diagnosis may be firmly achieved reasonably only with the aid of genetic diagnosis[15,16].

In addition to these clinical diagnostic issues, despite increasing identification of novel genes underlying the ciliopathies[10], pathogenic mechanisms remain poorly understood. It is also likely that not all of the ciliary transition zone/basal body apparatus involved in dysfunction of primary cilia and cystogenesis have been defined. This presents an exciting opportunity to define Asian genes and gene variants that are implicated in HRFD in the regional context. Furthermore, the natural history of ARPKD and other HRFD needs to be better characterized; the impact of recent advances in neonatal care and paediatric dialysis and transplant on patient survival must be systematically evaluated; and biorepositories need to be created which will form the base for ascertainment of novel therapies in new clinical trials.

This proposal outlines the formation of a regional registry in Asia for the HRFD. This consists of a clinical database with longitudinal prospective follow-up, as well as a DNA repository which provides an opportunity for genetic testing for patients and families in Asia who fulfil the inclusion criteria for this study. Novel genes found from this sequencing pipeline will then be tested functionally using in vitro and in vivo models developed by collaborators of the DRAGoN study.

2. HYPOTHESIS AND OBJECTIVES

a) Hypothesis

We hypothesize that genetic susceptibility or disease-causing genetic variants in Asians for primary renal diseases are different from those in other populations. +

b) Primary Aims

DRAGoN aims to establish comprehensive genotype and longitudinal phenotype databases of Asian patients with renal diseases, as well as perform functional studies of novel genetic variants, in order to further our understanding of these diseases in Asia.

Our current specific aims are divided into two parts:

(i) Genome-wide association studies (GWAS): Identify genetic variants in novel or known genes that may be important in disease susceptibility in primary sporadic FSGS in Asia.

(ii) Next Generation Sequencing (NGS): Identify disease-causing sequence variants. We will recruit patients with primary proteinuric glomerular disease which have onset <3 years old, steroid-resistant or have persistent proteinuria that are likely genetic in etiology; as well as patients with hepatorenal fibrocystic disease and nephronophthisis-related ciliopathies.

(iii) Establish a registry or phenotype database of primary glomerular diseases, hepatorenal fibrocystic disease (HRFD) and nephronophthisis-related ciliopathies (NPHP-RC) in Asia. This will form the basis of genotype-phenotype correlations and provide unique insights into the longitudinal disease course of patients with these diseases in Asia.

c) Secondary Aims

(i) Perform preliminary *in vitro* functional work on the genetic variants identified on known or novel genes.

3. RATIONALE AND JUSTIFICATION FOR THE STUDY

a) Rationale for the Study Purpose and Clinical Impact

The current understanding is that proteinuric diseases occur due to insults to either the glomerular basement membranes or podocytes, both of which are components of the glomerular filtration barrier. Podocytes are highly differentiated visceral epithelial cells lining the GBM. The importance of the podocytes in disease is well demonstrated in the identification of hereditary glomerular diseases linked to their structural anomalies. These are mainly single-gene disorders, also known as monogenic diseases, in which a mutation in one gene is sufficient to cause the disease.

Autosomal recessive polycystic kidney disease (ARPKD) occurs with an incidence of 1:20,000[8]. It is due to mutations in the *PKHD1* gene, encoding for fibrocystin (FPC); recently *DZIP1L* has been described as a novel gene causative of ARPKD[10]. FPC is a 4074 amino acid containing a single-pass transmembrane protein, with a large extracellular domain (like PC1) and a short C-terminal tail. FPC has been shown to localize to basal bodies and cilia, and there it interacts with the PC1-PC2 complex. Currently, no approved therapies specifically targeting the cystic condition exist, and all clinical interventions are directed at management of disease symptoms. Consequently, ARPKD presents a significant burden to the patients, their families, and the health-care system. However, advances in neonatal care, ventilation and infant renal replacement therapy have improved outcomes for these patients.

<u>Aim (i): GWAS</u>

There has been increasing recognition that susceptibility to FSGS can have polygenic influences.[17] [18]In polygenic disorders, the pathogenesis involve interactions between the genes and environment, as well as between different gene variants.[19] A genetic change leads to a relative risk of the disease.

Although most associations show polygenic inheritance patterns with limited effect size, this is not the case in the association between *MYH9* and FSGS. Genetic variation at the *MYH9* locus substantially explains the increased burden of FSGS and hypertensive renal failure among African Americans.[20] Yet, not all individuals of African descent with the *MYH9* kidney risk variant will develop clinical kidney disease in their lifetime.[21] Thus, it is likely that additional genes and/or environmental factors interact with the *MYH9* kidney risk variant to trigger glomerular injury. Genetic association between the apolipoprotein L1 (*APOL1*) gene and several severe nondiabetic forms of kidney disease in African Americans also approach Mendelian inheritance patterns and account for a large proportion of glomerulosclerosis in populations of African ancestry.[22] **With the discovery of *MYH9* and *APOL1* as important risk alleles for FSGS and various forms of nephropathy in African Americans,[20,22] FSGS, especially if sporadic, may be considered as a polygenic disorder.[19] Risk alleles in FSGS are largely understudied currently.**

Clinical impact: The exact pathogenesis of FSGS is not clearly understood. The characterization of genetic basis of FSGS has led to the important recognition of podocyte injury to the development of glomerulosclerosis.[23]We need to increase our understanding of the pathomechanism in order to better target treatment strategies. Current treatment of primary glomerular diseases is often empirical and involves the use of several concomitant immunosuppressive agents. Response to these highly toxic and expensive drugs is often limited to a small proportion of patients. There are currently no methods to identify the patients who are likely to respond to immunosuppressive therapy. Clinicians therefore often use a "trial of therapy" strategy. Understanding the pathogenesis of these diseases will lead to more directed treatment strategies, which will translate to less side effects and lower healthcare costs. Identification of risk alleles for the FSGS in Asia will allow for estimation of lifetime risks for kidney disease in Asians, as well as genetic risk stratification for disease course and response to therapy. This can eventually allow for genetic testing as part of personalised medicine.

Aim (ii): NGS Study

Part 1: Glomerular disease

More than 50 genes have been identified in podocytes, and recognised to be contributing to proteinuric diseases.[24-26] Mutations are most commonly found in *NPHS1* (nephrin), *NPHS2* (podocin) and *WT1* for early-onset nephrotic syndrome, and in *ACTN4* and *TRPC6* for late-onset familial or sporadic disease.[24] Targeted gene sequencing of 24 known genes in 36 paediatric patients with SRNS in United Kingdom resulted in identification of definitely or probably pathogenic variants in 86% of those with disease onset less than 2 years old, and only 17-18% in those with disease onset after 2 years old. There are probably still undiscovered genes. Indeed, with the new molecular techniques available, several new genes have been recently discovered, including *ADCK4*[27,28], *ARHGDIA*[29,30], *LMX1B*[31], *COQ6*[32], *TTC21B*[33], anillin[34], *EMP2*[35], *NXF5*[36], *PAX2*[37] and podocalyxin[38], FAT1[39], SGPL1[40,41] and NUP160[42], attesting to ongoing need to study more patients and reanalysing old data.

85% of SRNS cases with onset by 3 months of age and 66% of cases with onset by 1 year of age can be explained by recessive mutations in one of four genes (NPHS1, NPHS2, LAMB2 or WT1) [43]. In a study of 48 individuals with early-onset SRNS in which they were screened for monogenic mutations in 16 recessive and 5 dominant SRNS genes, causative mutations were detected in 16 of the 21 genes in 33% of the individuals [44]. [43]

In several worldwide cohort of individuals with SRNS with onset before 25 years of age, sequencing in genes known at the time to cause monogenic SRNS [45]detected a single-gene cause of SRNS in about 30% of individuals. [45] [46-48]

The fraction of individuals in whom a single gene cause was identified inversely correlated with age at manifestation[49] : onset in the first 3 months of life (69.4%), from 4 to 12 months (49.7%), from 1 to 6 years (25.3%), from 7 to 12 years (17.8%) and from 13 to 18 years (10.8%). [45]As expected, mutations of SRNS genes were found more frequently in consanguineous marriages. Specifically, in non-consanguineous families, the detection rate was ~25% of cases, whereas it was ~50% in consanguineous families.[45]

However, a significant proportion of cases with later-onset SRNS were still molecularly unsolved [43].

Indeed, the phenotype of the patient may not necessarily correlate with the genotype of the patients. A case with congenital nephrotic syndrome was attributed to a homozygous missense mutation in ADCK4, and a de novo missense mutation in TRPC6 was detected in a case with infantile nephrotic syndrome[50].

In addition, familial haematuric diseases are a genetically heterogeneous group of monogenic conditions. The major genes involved in the literature are the collagen IV genes *COL4A3*, *COL4A4*, *COL4A5*, as well as the *FN1* gene and *CFHR5* which causes inherited C3 glomerulonephritis.[51] These conditions have age-dependent penetrance and broad phenotypic heterogeneity which includes progression to chronic renal failure. There are still families with unidentified genetic mutations.[51]

Glomerular disease is increasingly recognised as a complex disease with a multi-hit hypothesis. We are just beginning to realise the significance of genetic epistasis between different genes, and to recognise it as an important "hit".[52] [53] This has been suggested in studies in which simultaneous variants were noted in *WT1* and *NPHS1*.[52,54] Functional studies suggesting interactions between podocin and *TRPC6* have been reported.[55]

Hildebrandt *et al* described the use of a multiplexing method to simultaneously analyse 21 single genes in 48 SRNS individuals, significantly reducing the turnaround time and the cost of mutational analysis compared to the traditional methods.[56] 307 patients with SRNS were screened with targeted NGS of 37 genes and 20% of the paediatrics cases were found to be due to a genetic cause. 45% of the patients with genetic disease had novel pathogenic variants with no previous disease association, including 2 patients that had exonic deletion of *NPHS1* or *NPHS2*. Targeted NGS has provided a genetic diagnosis in SRNS patients, aiding in the stratification and treatment management.[57]

Clinical impact:
This can potentially be enormous.

a) It may allow for proper genetic diagnosis and treatment strategies in local patients. Current evidence-based recommendation is that all steroid-resistant nephrotic patients should undergo genetic testing, and if mutations are found, intensified immunosuppression should not be given.[24,58,59] [50] In most Asian institutions including Singapore, most SRNS patients are not properly managed according to international evidence-based recommendations because genetic testing is not conducted in most patients due to lack of local facilities and high costs and turnover times for genetic tests performed at overseas laboratories. Most patients are given empirical immunosuppressive drugs with significant adverse effects (including malignancy in long term)

and medical costs. With this study, patients can be better stratified and immunosuppressive therapy be given more rationally. The direct impact will be significant reduction in medication costs, and costs incurred from complications of the drugs.

b) Genetic diagnosis can also allow genetic counselling for the patient and family, as well as to plan for renal transplant. Live-related donors with mutations are generally precluded from transplant.[60] Patients with genetic FSGS are less likely to have recurrence post-transplant.

b) Rationale for Study Population

Much of the genetic studies in nephrotic syndrome occurred in North America, Europe and Middle East. Despite the large populations in Asia, published genetic studies from this part of the world are relatively scarce.[61,62]

Similar to animal models in which the genetic susceptibility to FSGS differs among different strains,[17] the genetic susceptibility to renal disease in humans is also related to the ethnicity. Evidence for genetic variation among different ethnicities are:

(i) Increased susceptibility to renal diseases due to *MYH9* or *APOL1* polymorphisms disproportionately affects individuals of African descent compared to other populations worldwide.[22,63] The *MYH9* risk haplotype (E-1) occurs at much higher frequencies in African Americans (more than 60%) than in European Americans (less than 4%).[63]

Mutations in *NPHS1* and *NPHS2* are generally less prevalent in Asia compared to Europe and the Middle East.[61,64,65] *NPHS2* mutations are found in 10-30% of Caucasian patients with sporadic nephrotic syndrome or FSGS and 43-55% of the familial cases,[66-68] while studies in Asia including a preliminary study in Singapore showed a very low prevalence of mutations (0-4%) in sporadic and familial cases.[69-72]

(iii) The *NPHS2* non-synonymous functional polymorphism c.686G>A (p.R229Q, rs61747728) has been shown to be functional,[73] and a recent meta-analysis showed a higher risk of SRNS in homozygous individuals compared to homozygous nonvariant individuals.[74] The allele frequency of this p.R229Q in healthy controls varies with ethnicity, and has been reported as 3-13% in Europeans and 5 to 25% in Africans.[61,74,75] This variant is strikingly almost absent in Asians.[74,76] Founder mutations occurring at higher rates in certain regions of the world have been noted[45]. For example, the R138Q mutation of NPHS2 occurred frequently in Western Europe and the USA.

In studies conducted in Chinese patients[50,72], a genetic etiology was identified in 28.3%-31.7% of the patients. The most common mutated genes were *COQ8B* (formerly known as *ADCK4*) (6.67%), NPHS1 (5.83%), WT1 (5.83%), and NPHS2 (3.33%), and the difference in the frequencies of ADCK4 and NPHS2 mutations between this study and a study on monogenic causes of SRNS in the largest international cohort of 1,783 different families was significant. In Japan, whole exome sequencing was performed on 24 patients with SRNS and/or FSGS and about 20% of the families was found to have causative gene mutations.[77]

Part 2: Hepatorenal fibrocystic disease and nephronophthisis-related ciliopathies.

Identify disease-causing sequence variants in genes known to cause hepatorenal fibrocystic disease and nephronophthisis-related ciliopathies.

Establishing a specific genetic diagnosis has implications for prognostication, family planning as well as transplant planning. Examples exist of novel genes being discovered by NGS: *ANKS6(NPH16)* was discovered using targeted NGS, whereas *WDR19 (NPHP13)* was discovered via whole exome sequencing of patients with Sensenbrenner syndrome. Furthermore, there is the phenomenon of oligogenic inheritance or epistatic manisfestations, that may occur and be discovered only by massively-parallel NGS methods. [78]

Clinical impact:
This can potentially be enormous.
Many countries, especially resource strapped ones, do not currently have resources for genetic testing. Achieving a genetic diagnosis will have implications on prognostication, family planning, as well as planning of transplant (screening the potential donor for carrier status).

Novel genes discovered by NGS, if validated functionally in *in vitro* and *in vivo* models, will help further our understanding of the interplay between ciliary pathobiology and the HRFD/NPHP-RC spectrum of diseases.

Aim (iii): Establish a registry or phenotype database of primary glomerular diseases, hepatorenal fibrocystic disease (HRFD) and nephronophthisis-related ciliopathies (NPHP-RC) in Asia.

There is a current dearth of understanding regarding glomerular diseases, ARPKD and HRFD in Asia. This is an important disease in Asia due to the higher prevalence of consanguinity; additionally, culturally-modified health seeking behaviour and the availability of renal replacement resources may have an impact on outcomes of these patients. This data is important to collect and understand in order to truly benefit the management of these patients in Asia.

A comprehensive phenotype database allows tracking of important clinical parameters so that similar patients may be analysed together, and genetic analysis will be more logical. This database will also allow future comparison with more established glomerular disease databases like the Podonet[47] (Europe) and Neptune[79] (USA). This sharing of clinical data across databases is important in establishing clinically useful genotype-phenotype correlations. It would also maximise the benefits of genetic testing in the routine diagnosis of nephrotic diseases.[80]

Despite the clinical burden of ARPKD and related HRFD, large organized patient cohorts of this rare disease are scarce. Two major registries have been established in the United States[81] (HRFDCC Clinical Database) and Europe (ARegPKD) to chronicle the longitudinal course of disease and complications in these patients, as well as systematic evaluation of potential novel therapies. Currently (2017) in the HRFDCC, 103 patients have been recruited. Findings in this cohort included: hyponatremia in 31.3%, hypertension in 72.3%, and CKD (eGFR <90 ml/min/1.73m^2) in 57.8% and ESRD in 22.9%. In ARegPKD, 400 patients have been recruited (as of 2017) and findings are currently being analysed.

Moreover, genetics alone does not explain the phenotypic variability in these diseases. It is thought that epigenetic and other environmental factors may be important. The longitudinal clinical database will help to capture these features, and in conjunction with other international databases, may be used to answer these questions.

c) Rationale for Study Design

See Study Design.

Both genotype and phenotype databases are important in understanding diseases.

Risk alleles in polygenic diseases are usually identified from GWAS. GWAS investigates the entire genome, and is therefore non-candidate-driven. The GWAS can identify single nucleotide polymorphisms and other variants in DNA which are associated with a disease, but cannot on their own specify which genes are causal. This test is selected because primary FSGS is increasingly recognised to be polygenic disease.

NGS, on the other hand, will allow us to discover novel undiscovered genes, or to identify variants in the less commonly sequenced and implicated known genes. Phenotype does not predict genotype in the way we understand today. The current notion is that out of the 24 genes known to cause SRNS, only about 8 of them are commonly implicated.[58] The other genes are therefore less commonly sequenced and hence implicated less commonly. This was clearly suggested by McCarthy *et al* who performed NGS in 36 patients with familial/sporadic SRNS in United Kingdom and found pathogenic variants in genes that would rarely be considered as highly likely today.[54] A next-generation screening of 27 SRNS causative genes in an international cohort of 1783 unrelated SRNS families revealed a single gene cause in 29.5% of SRNS families that manifested before the age of 25.[82] For example, in a patient with isolated nephropathy, mutations were noted in Coenzyme Q2 homolog (*COQ2*), a gene implicated in mitochondrial disease. In another patient with sporadic SRNS and renal histology that is NOT typical of thin basement membrane or Alport's syndrome, and no sensorineural deafness (hence rendering Alport syndrome unlikely), highly probable pathogenic variants were noted in *COL4A4*, a gene commonly implicated today in Alport syndrome or thin basement disease. In addition, this NGS found pathogenic sequence variants in 70% of the familial cases,[54] in contrast to Santin et al who performed targeted conventional sequencing of eight commonly implicated podocyte genes (*NPHS1*, *NPHS2*, *TRPC6*, *CD2AP*, *PLCE1*, *INF2*, *WT1*, *ACTN4*) in patients with SRNS and found mutations in only 27% of the familial cases.[58] Clearly, there are undiscovered genes, or the other already known genes (out of the 24 genes) currently thought to be unimportant or uncommon are actually quite important!

Our approach to use targeted gene sequencing first, followed by whole exome sequencing is more cost-effective, and allows for more efficient use of time and resources.

4. STUDY POPULATION

a) Target number of subjects

Primary glomerular diseases

Number of subjects to be recruited worldwide: 500 cases and 300 healthy subjects.

Cystic kidney disease

Target is 500 patients for NGS including whole exome sequencing in selected cases. With an estimated mutational rate of 20%, we expect to find disease-causing mutations in about 100 patients.

b) Patients

Inclusion Criteria (PRIMARY GLOMERULAR DISEASES)

1) Congenital or infantile nephrotic syndrome with onset in first year of life* OR
2) Nephrotic syndrome with steroid resistance at first manifestation, with disease onset before 25 years of age* OR

3) Persistent subnephrotic proteinuria with likely genetic etiology**.
See Definitions.

Notes:

- *Includes familial and sporadic (non-familial) cases.
- **Likely genetic etiology refers to
  - Onset <3yrs, OR
  - Significant family history including consanguinity, OR
  - Significant relevant extrarenal malformations (eg hearing loss, lenticonus)
- Recruitment with respect to time course of disease
  - Patients may have presented several years before time of recruitment, and may be recruited at any time point of their disease course, regardless of the degree of renal impairment.
  - Patients who are already in remission, but their initial presentation or previous clinical features satisfy the criteria can still be recruited.
  - There is no limit to the time period between onset and age of the patients at recruitment, as long as the inclusion criteria are satisfied.
  - There is no upper limit to the current age of the patients.

Availability of histological evidence is highly recommended, but not mandatory. See "Histopathological evidence"

- Primary evidence should be as one of the following forms:
  - **Histological slides** (preferred)
  - **Biopsy in paraffin block** (preferred)
  - **Renal biopsy samples in formalin** (to be pre-arranged)
  - **Digital photograph** of light microscopy of histology slides
- Histological evidence may be dated up to 5 years back, as long as it is still clear and undisputable.

Exclusion Criteria (PRIMARY GLOMERULAR DISEASES)

1) Secondary glomerular diseases, such as lupus nephritis, IgA nephropathy, Henoch–Schönlein purpura.
   - We recommend a careful clinical evaluation (history, physical examination, investigations) to ensure secondary causes are not present. These include, *but are not limited to*, the following:

	Symptoms	**Signs**	**Laboratory investigations**	**Histopathology in addition to light microscopy**
Secondary FSGS	Drugs Viruses Obesity History of nephron loss (nephrectomy, congenital anomalies)	Obesity	HIV and Parvovirus in presence of collapsing FSGS	Electron microscopy: podocyte foot process effacement not extensive
Lupus nephritis	SLE features	SLE features	ANA, anti-dsDNA, complements	Immunofluorescence
IgA nephropathy	Recurrent painless synpharyngitic gross haematuria	May have nephritic syndrome	High serum IgA	Immunofluorescence

Henoch–Schönlein purpura	Rash Joint pains Abdominal pain	Rash Joint swelling May have nephritic syndrome	High serum IgA	Immunofluorescence

Inclusion Criteria (CYSTIC KIDNEY DISEASE)

1) ARPKD as defined by
    a. Typical findings on renal imaging (enlarged kidneys with diffuse hyperechogenicity with poor corticomedullary differentiation, or renal cysts) AND
    b. One or more of the following criteria:
        i. Clinical/laboratory signs of hepatic fibrosis
        ii. Hepatic pathology demonstrating ductal plate abnormality
        iii. Absence of renal enlargement and/or multiple cysts in both parents
        iv. Diagnosis of ARPKD in an affected sibling
        v. Family history consistent with autosomal recessive inheritance.
        vi. Parental consanguinity

2) Nephronophthisis-related ciliopathy as defined by
    a. Hyperechogenic kidneys and corticomedullary cysts/poor corticomedullary differentiation
    b. AND one or more of the following extrarenal features
        i. Liver fibrosis
        ii. Ophthalmologic disorders: Blindness, retinitis pigmentosa or retinal dystrophy/degeneration, nystagmus or coloboma
        iii. Structural brain abnormalities
        iv. Polydactyly or skeletal abnormalities
        v. Abnormalities of left-right patterning (also known as heterotaxy) eg dextrocardia, situs inversus

3) Families with suspected nephronophthisis related ciliopathy as defined by two or more family members with hyperechogenic kidneys or cystic kidneys

Exclusion Criteria (CYSTIC KIDNEY DISEASE)

1) Autosomal dominant polycystic kidney disease (ADPKD)
2) Suspected renal dysplasia, with or without urinary tract malformations. Eg cystic dysplastic or multicystic dysplastic disease.
3) Acquired renal cystic disease (associated with long duration of ESRD)

Notes:

1. *Due to resource limitations,* ***priority for genetic sequencing will be given to consanguineous cases, those with more than one family member affected in a recessive manner, and those with extra-renal features suggestive of a ciliopathy****.*
2. *Patients who have already received genetic diagnoses with mutations in PKHD1 gene or other HRFD/NPHP-RC genes are welcome in the registry. DNA received will be banked and genetic sequencing may not be performed.*

3. *The main coordinating centre reserves the right not to run genetic testing for cystic patients who do not fulfil inclusion criteria (eg. Cystic dysplastic kidney disease or ADPKD). Blood received will be banked instead.*

c) Family members

Inclusion criteria

As much as possible, family members should be recruited for both familial and sporadic cases. **These include both parents, ALL affected siblings and as many healthy siblings as possible. The more healthy and affected family members are recruited, the more accurate and comprehensive the genetic analysis.**

For renal cystic disease, we regret that due to limited resources, we may not be able to process the proband's samples for genetic tests if we do not receive at least both parents' samples.

Family members need to be classified into "healthy/unaffected" or "affected".

Healthy/Unaffected

Healthy/Unaffected family members are those with no known renal diseases or kidney dysfunction (eg. Haematuria, proteinuria, raised creatinine).

Affected

Primary Glomerular Diseases

Affected members are those with evidence of glomerular disease (proteinuria, persistent glomerular haematuria, glomerular or GBM abnormalities on light or electron microscopy).

This also includes patients who have chronic kidney disease or end-stage renal disease of unknown cause.

For those family members with known renal diseases, clinical details, such as clinical presentation, histology, treatment response, renal function, will be required.

Renal Cystic Disease

Affected members are those who have had cysts in the kidneys and/or liver found on imaging, or hyperechogenic kidneys on imaging.

Suspected Affected (FOR RENAL CYSTIC DISEASE ONLY)

Patients who have chronic kidney disease or end-stage renal disease with no known cysts are classified as suspected affected members, as other causes (eg glomerulonephritis) may be responsible for the kidney disease. Those with kidney diseases with biopsy evidence of an alternative diagnosis (eg glomerulonephritis) should also be considered as suspected affected as certain genes are known to cause both cystic disease and glomerulonephritis.

For those family members with known renal diseases, clinical details, such as clinical presentation, histology, treatment response, renal function, will be required.

d) Subject Withdrawal

Study subjects and / or their parents or legal guardians have the authority to withdraw from the study and remove all the genetic materials and derivatives for present and future studies, without giving

any reason. Subjects only need to indicate to their local site investigators verbally, via phone, emails or any other forms of communication. The site investigators will inform Main Coordinating Centre investigators via email.

When subjects indicate their wish to withdraw, all unused samples in the Main Coordinating Centre will be destroyed, unless the subjects indicate their wishes for these samples to be shipped back to them.

If the samples have been used for genetic sequencing, the sequencing and phenotype data will still be used for analysis in deidentified form, as per standard ethics protocol in Singapore.

If the sample has not been sent for genetic sequencing upon notification of the withdrawal, all deidentified phenotype data of these withdrawn patients will be deleted.

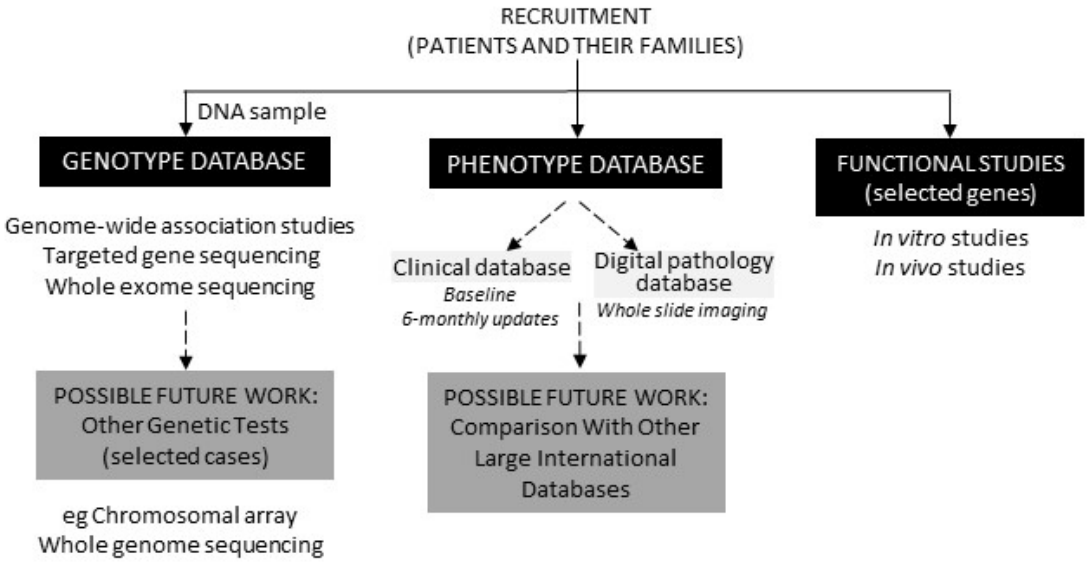
Withdrawn subjects will not be replaced.

e) Subject Termination

Subjects are only expected to have only one study visit for consent taking and provision of samples. There is no need for research-related follow-up visits unless repeated samples are needed in rare cases.

Subject's follow-up data will be entered by site investigators to phenotype database, and this follow-up data will be terminated once subject is no longer on follow-up with the local site (eg transfer, demise, loss to follow-up etc)

5. STUDY DESIGN



6. STUDY PROCEDURES

a) Summary of study procedures

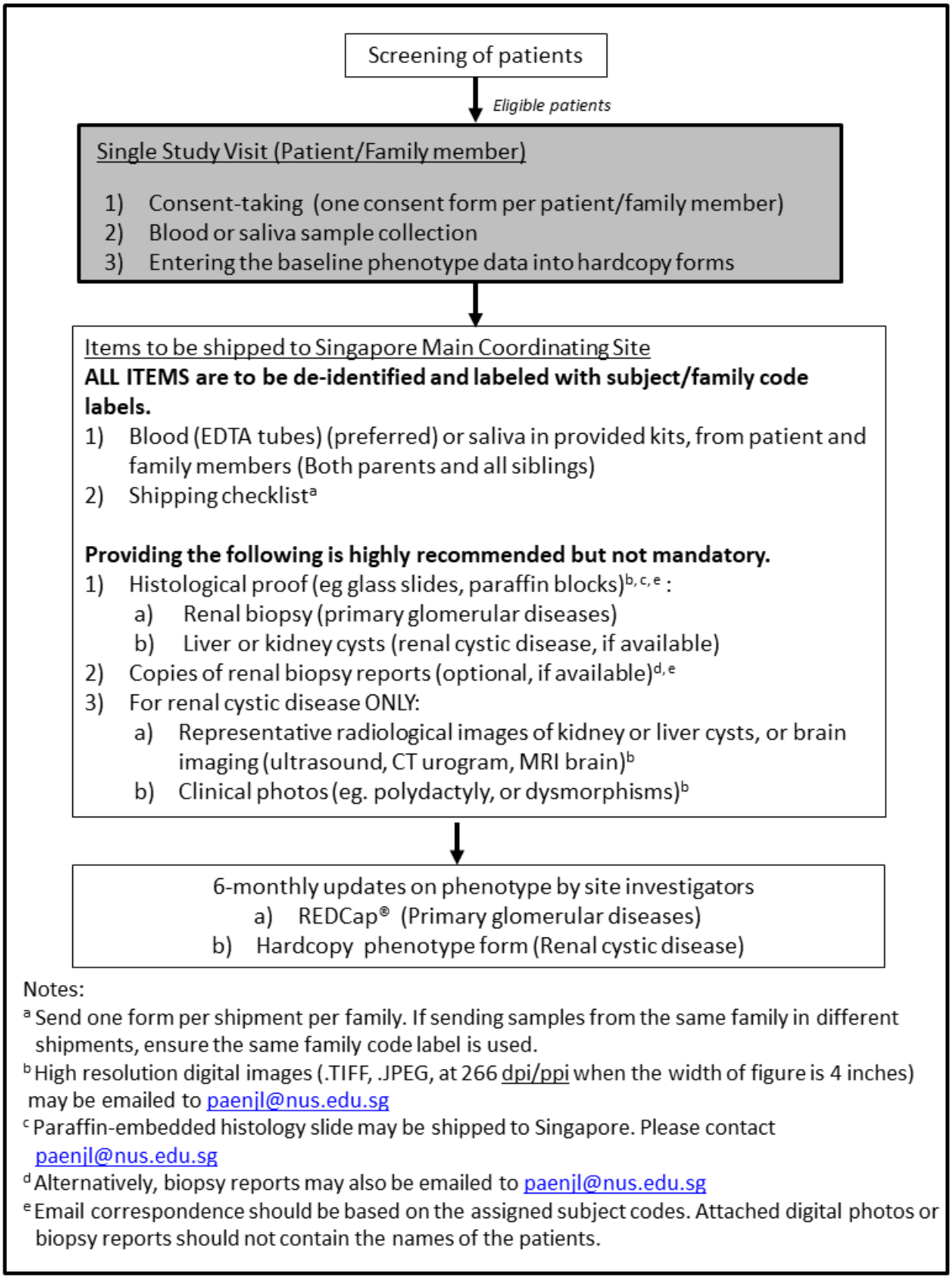
For each recruited affected patient, the participation involves giving a blood sample, with up to 2 repeated samples in case of inadequate sampling. Thereafter, the local study investigators will enter the phenotype data every 6 months for up to 5 years, as long as the patient is on follow-up. The total duration of participation for each affected patient therefore will be up to 5 years.

For each recruited healthy family member, the participation involves giving a urine sample for confirmation of kidney health (urine is for dipstick testing and will not be sent to Main Coordinating Centre) and a blood sample, with up to 2 repeated samples in case of inadequate sampling. Thereafter, there is no more participation required. Local study investigators will only enter phenotype data once, and will NOT enter phenotype data prospectively.

Blood samples of probands will be used for genetic testing, and the family members' blood used only for confirmation of variant pathogenicity, where applicable.

The entire DRAGoN study duration will be up to 20 years. This is to allow for continued recruitment of patients in other participating centres. In addition, patient samples may be used for genetic testing **within this study's aims and scope**, up to 10 years from subjects' recruitment. This long duration is because investigators need repeated testing on samples to confirm findings, and there may be newer techniques of gene testing in the future that may be used to allow better identification of genes.

If there are future plans to use the stored patient samples (stored with consent) **outside of this study's scope**, specific written notice, legal agreements and ethics approval will be obtained before proceeding.



Notes:

[a] Send one form per shipment per family. If sending samples from the same family in different shipments, ensure the same family code label is used.

[b] High resolution digital images (.TIFF, .JPEG, at 266 dpi/ppi when the width of figure is 4 inches) may be emailed to paenjl@nus.edu.sg

[c] Paraffin-embedded histology slide may be shipped to Singapore. Please contact paenjl@nus.edu.sg

[d] Alternatively, biopsy reports may also be emailed to paenjl@nus.edu.sg

[e] Email correspondence should be based on the assigned subject codes. Attached digital photos or biopsy reports should not contain the names of the patients.

b) Site activation

Each site will be activated once the following documents are received by the Main Coordinating Centre.

- Approval Letter from the respective Institution Review Board
- Approved consent forms for affected and unaffected (healthy) subjects

The Main Coordinating Centre will then

(i) Ship the following items to the site:
- Ethylene-diamine-tetra-acetic acid (EDTA) blood tubes
- Urine dipsticks
- Biohazard bags
- Ziplock bags (secondary container)
- Slide mailer boxes (for histology glass slides).
- Subject sticky labels (One family in one sheet of label)
- Blank sticky labels (site investigators to copy subject codes to blank labels).

Note: The other consumable items needed for blood taking (eg syringes, needles, cleaning solutions and gauzes) will not be routinely provided by the Main Coordinating Centre. Site investigators however may contact the Main Coordinating Centre if they require such supplies.

(ii) Provide details of the shipping company account to send samples to Main Coordinating Centre.

(iii) Arrange a web-conferencing session to orientate site investigators on recruitment and phenotype data entry where necessary.

c) Recruitment procedure of patients and family members

Patients may be recruited during hospital or clinic visits for the routine clinical evaluations. Their family members may be recruited when they accompany the patient for the visit or as a separate visit.

Consent and samples from the family members may be taken on different days. Although it is generally recommended that samples from the same family be sent on the same shipment, if this is not possible, then, separate shipment is also possible. It is important that labels with the same family code be used when sending samples from the same family in different shipments (see Subject codes and sample labels).

d) Study visits and procedures for subjects

Only a single study visit is required in most cases. There is no need to schedule visits especially for this study. Blood specimens can be collected during routine clinic visits or hospital admissions whenever possible. If this is not possible, the patients may choose to come to the hospital specially for the study purpose.

During the study visit, consent will be taken, and the specimens collected.

(i) Consent form

Written consent will be obtained from the study subjects and/or their parents/legal guardians by the site investigators according to the institution's requirements. **There is no need to send the signed consent forms to the Main Coordinating Centre.** There are separate consent forms for affected subjects (patients), and healthy unaffected family members. Affected family members should use the consent forms meant for patients. Templates of the Informed Consent Forms can be found in

Appendix IV (Patients and affected family members) and Appendix V (Healthy family members). **These are to be modified according to the institution's requirements.**

(ii) Sample collection

The blood samples need not be taken on the same day as the written consent.

Blood

The amount of blood needed will be 4-5ml (1-2 EDTA tube). It is important that the amount of blood in each EDTA tube do not exceed the maximum fill line shown on the tube (3ml). If more than 3ml of blood is obtained, please use 2 EDTA tubes. For a child less than 2 years old, 3ml of blood may be obtained. Blood will be obtained using 28enipuncture and can be timed, whenever possible, together with 28enipuncture performed for clinical indications. For patients who have received blood transfusion without leukocyte filter, blood sampling should be done three months after the blood transfusion. For these reasons, the blood sample need not be taken on the same day as the written consent.

Repeat collections

Often, a single specimen collection is enough. However, in the rare circumstances in which the specimens are not enough or of a poor quality, a second sample may be needed. In addition, if further tests of gene function are necessary, a subsequent or repeat blood sample may be necessary. The second collection may occur between as early as a week later, or may be much later (up to 12 months later). For subsequent collections, 3ml of blood for less than 2 year old, 4-5ml of blood for more than 3 year old. There should not be more than 3 sample collections in total. The subject may refuse further sample collections.

Note: All recruited subjects only need one research-related visit in which consent is obtained and the blood sample is collected. This can be conducted during a routine clinic / hospital visit. Thereafter, there are no more research-related visits or commitment for the subjects, unless the samples collected are of inadequate quality or quantity (very rare) in which case, they will be requested to provide another sample in another single visit.

Samples from family members

Specimens from the family members of the same family should be sent to the Main Coordinating Center at the same time whenever possible. If this is not possible, they can be sent in batches at different times. Each subject (healthy or affected) with submitted specimens should have a unique Subject Code. Members of the same family should use labels from the same sheet so that they have the same family code within the subject code (See Subject codes and sample labels).

(iii) Sample handling

Blood in EDTA tubes may be kept at 4°C for 3 to 4 weeks before shipping at room temperature. Alternatively, they can be kept at room temperature for a few days before shipping.

The samples should be labelled using the subject code labels and placed inside the biohazard bags provided by the Main Coordinating Center. Specimens from the same family may be put in the same bag. The biohazard bags should be put in another ziplock bag (provided) as a secondary container before shipping.

e) Subject codes and sample labels

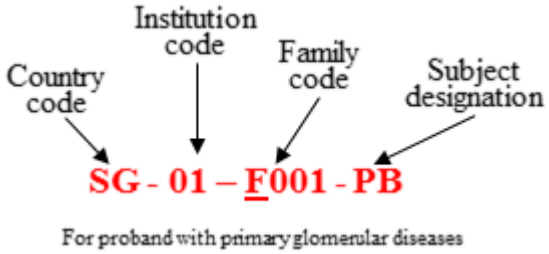
Each subject (healthy or affected) with submitted specimens should have a unique assigned Subject Code.

What do the subject codes mean?

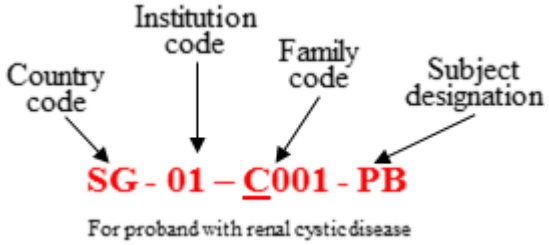
Each subject code is made up of 4 numbers. Country and institution codes will be the same for all subjects recruited from the same institution. Members of the same family will share the same family code. Different families recruited will have different family codes eg F1, F2, F3 or C1, C2, C3 etc.

IMPORTANT: ***Subjects under the cystic disease arm of DRAGoN will have different DRAGoN code running separately from the glomerular arm.***

An example of a subject code for GLOMERULAR DISEASE is SG-01-**F**001-PB:



An example of a subject code for CYSTIC DISEASE is SG-01-**C**001-PB:



The F is replaced by C in the code for cystic diseases. Separate sets of patient labels will be provided for glomerular and cystic arms. Codes for glomerular and cystic arms will run in order of the family code numbers and independently of each other.

The last two digits will be based on the subject designation, based on the table below.

Subject	**Possible relationship to proband**	**Subject designation**
Proband	NA	PB
Father	Father	FR
Mother	Mother	MR
Sibling 1	Sister or brother	S1
Sibling 2	Sister or brother	S2
Sibling 3	Sister or brother	S3
Sibling 4	Sister or brother	S4
Others 1	Grandparent, aunt, cousin, others	T1
Others 2	Grandparent, aunt, cousin, others	T2
Others 3	Grandparent, aunt, cousin, others	T3
Others 4	Grandparent, aunt, cousin, others	T4
Others 5	Grandparent, aunt, cousin, others	T5

How to use subject code labels (Appendix III)?

See Appendix III for a sample of subject labels.

- **ONE sheet of labels will be used for ONE family.** If only the proband is recruited with no family members, then ONLY the proband's label is used. The entire sheet can be kept in case and if the family members can be recruited at a later stage. **USE A NEW SHEET FOR A NEW FAMILY.**
- Use the subject code labels according to the designation of the family member. Eg Proband's father is to use "FR" label. If there are family members not stated (eg step-father, cousin, aunt, grandparent), then use "OTHERS".
- Family members may be healthy or affected. If there are more than one affected in the family, arbitrarily assign one child to be the proband.
- The blank row of labels at the bottom are extra ones to allow site investigators to write in codes if the labels are misplaced, or are damaged or not enough.
- **If there are more family members recruited from the same family and there are not enough assigned labels in the same sheet,** use the blank sheet of labels provided to write in the codes. The country, institution, family codes should be the same in the same family. The subject subject designation will be T6, T7, T8, T9 etc. Using the example above, the subject codes for subsequent family members will be SG-01-F001**-T6,** SG-01-F001**-T7,** SG-01-F001**-T8**, etc.
- There are 5 sticky labels per subject with the same subject code (See Appendix III). These will be used in

	Number of labels
Sample tubes (blood)	1-2
Shipping checklist	1
Primary histological evidence (Slide mailer box etc)	1

If more than 5 sticky labels are needed for one subject, copy the subject code in the blank labels provided.

f) Phenotype data collection

Each centre will observe and record information on subjects who meet criteria after obtaining informed consent. Phenotype data include demographics, age of onset, clinical features at presentation, types of treatment given and the responses to them, renal impairment and transplant remarks, family history.

Primary Glomerular Diseases

REDCap® (https://redcap.scri.edu.sg/), a web-based phenotype database has been set up to allow site investigators to enter baseline and follow-up phenotype data. Instructions on how to use REDCap® database is found in Appendix II. A user-specific password will be given to each user for data entry. All phenotype will have to be de-identified. The nine baseline forms are Demographics, First manifestation, Extrarenal Manifestations, Renal Biopsies, Drug Treatment, Renal Replacement Therapy, Family History and Latest Patient Encounter. For the six-monthly updates on the patient's clinical progress, an email reminder will be sent at 5 months after the baseline data entry. The six update forms are Extrarenal Manifestations, Renal Biopsies, Renal Replacement Therapy, Family History, Latest Patient Encounter and Clinical Course and Medications.

Renal Cystic Disease

At the present time, we will use hardcopy forms for the phenotype data for cystic disease. A baseline phenotype form will have to be shipped with the blood samples. Phenotype data on the proband and the family members will be requested. At yearly follow-up, the follow-up forms can be shipped, emailed or faxed to Ms Ng Jun Li (jun_li_ng@nuhs.edu.sg). *Please also send representative diagnostic images by email at baseline and at yearly follow-up, if available.*

In addition, at least one of the following should be provided, unless there are other supporting features that can demonstrate high possibility of genetic cystic diseases.

1) Representative radiological images of kidney or liver cysts, or brain imaging (ultrasound, CT urogram, MRI brain)
2) Clinical photos (eg. Polydactyly, or dysmorphisms)
3) Histology of liver or kidney cysts (if available)

Local investigators should discuss with the investigators from the Main Coordinating Centre if there are doubts or difficulties in fulfilling the requirements.

Investigators at the Main Coordinating Centre may request for more clinical details if phenotype data is inadequate. Primary histological evidence is highly recommended but not mandatory. **We regret that genetic analysis of the samples will not commence till adequate phenotype data has been received.** Local investigators are welcome to discuss eligibility of cases before formally recruiting the patients.

Site investigators will enter phenotype data into the web-based phenotype database. These phenotype data are obtained from the patients' clinical visits (not part of study requirements). The follow-up data should be entered for as long as the patient is on active follow-up with the local site. The collected data would be a useful reflection of long-term outcomes including progression to end-stage renal failure and death. For this reason, the duration of phenotype follow-up data by site investigators should be for at least 5 years, whenever possible.

g) Phenotype definitions

(i) Definitions for patients

See Definitions.

(ii) Definitions for family members

Family members need to be classified into "healthy/unaffected" or "affected".

Healthy/Unaffected

For all recruited family members not known to have a renal disease, they should have, at the minimum, urine dipstick result in the last 6 months of recruitment. Only if there is no protein or blood detected on the dipstick can they be classified as "healthy or unaffected" family members.

If there is 1+ or trace protein noted on the dipstick, site investigators can choose to repeat on a separate day with an early morning sample. Alternatively, formal urine protein:creatinine ratio or 24h urine total protein may be sent (at subjects' own expenses).

If there is blood noted on the urine dipstick, the dipstick may be repeated, or a formal urine microscopy be sent (at subjects' own expenses). If the positive results are deemed worthy of a physician referral (at discretion of the primary physician), they will be referred.

Affected

For those family members with known renal diseases, clinical details, such as clinical presentation, histology, treatment response, renal function, will be required.

h) Histopathology evidence

- Availability of histological evidence is highly recommended but not mandatory.
- As much as possible, each subject/family is recommended to have at least one type of *primary* histological evidence.

    - *Primary histological evidence* as one of the following types is critical in phenotyping the patients:
        - **Paraffin-embedded histological slides** (preferred)
            - We request the entire set of slides from the biopsy to be sent to us (as many slides as possible). If not possible, a set of 5 slides of consecutive levels will allow pathologist to follow each glomeruli at multiple levels.
            - The slides can be returned to site investigators in a few months. Expedited return may be requested if necessary. PAS staining is preferred. If not present, then trichrome staining is recommended. It is not that ideal if only hematoxylin-eosin stain-slides are present.
        - **Biopsy in paraffin block** (preferred, block and processed histology slides can be returned to site investigator in 4 weeks)
        - **Renal biopsy samples in formalin** (can be pre-arranged with Main Coordinating Centre in Singapore – further instructions will be provided on request)
        - **Digital photographs** of light microscopy of glomeruli (PAS staining (most preferred), trichrome, hematoxylin-eosin (least preferred)) taken at **40x** resolution. Please contact us for specifications of the photographs before sending.

      IMPORTANT: We request evidence of at most 5 years back and of at least 5 glomeruli (slide or biopsy tissue containing at least 5 glomeruli, or photographs of at least 5 glomeruli).

    - *Secondary supporting histological evidence* as one of the following types will further aid in phenotyping the patient (Primary histological evidence should still be provided):
        - Digital photographs of immunofluorescence and electron microscopy (email to Main Coordinating Centre in Singapore if available)
        - Biopsy reports

- Histopathological assessment at Main Coordinating Centre
    - Histological slides, renal biopsy in paraffin block or formalin sent to Singapore will be assessed by pathologists. Whole slide imaging[83-86] will be performed to capture the images which will be stored in a database. A pathologist will validate the histological diagnosis given by the local site investigators. Systematic scoring of the glomeruli may be performed by independent pathologists at a later date.
- De-identification of the histological evidence is important (See "De-identification").

Whole slide imaging

- This will be performed by pathologists in Singapore General Hospital using established protocols[83].
- If facilities are available, scanning of consecutive whole slides may be performed at local sites, de-identified and sent electronically to Singapore. Further technical details will be provided.

i) Subject de-identification

All subjects will subsequently be attributed a unique Subject Code number by the Main Coordinating Centre. The code will be printed on the labels which will be provided by the Main Coordinating Centre. All samples and all documents to be shipped (Shipping Checklist, Phenotype Data Form) should be labelled only with the Subject Code Labels provided by the Main Coordinating Centre. The subjects' full name (including the initials) cannot be written on the code label. Documents to be emailed will be saved with file name using subject code only (eg SG01F001PB). The full name of the patient should be removed.

Investigators at each site will keep subject personal health information and identifiers in a separate log at their site. Only the site investigators have access to this log which links the codes to the subjects' identities. Each centre will maintain its own data and will transfer the de-identified data using the unique subject code number to the Main Coordinating Centre in Singapore. Confidentiality will be strictly maintained. Moreover, the study personnel will be trained to avoid recording any personal information on the REDCap® database by inadvertence.

If it is a hardcopy report, the full names of the subject should be removed (eg using a marker or black pen), and the Subject Code of the subject written on them instead.

Alternatively, if removal of the subject's name and birthdate is not possible, then the document should be emailed to a third party in NUS (ckc@nuhs.edu.sg) who will de-identify the document before sending to the Main Coordinating Centre.

j) Shipping checklist (Appendix I)

This should be shipped with the specimens to the Main Coordinating Centre. One Checklist should be used per family per shipment.
IMPORTANT: Family members from the SAME FAMILY should use labels from the SAME STICKY LABEL SHEET, so that they have the same family code (See Subject codes and sample labels and Appendix III). Site investigators need to take note of this especially if samples from the same family is sent in different shipments.

k) Shipping details

(i) Sample storage at local site

The blood sample should be stored in a 4°C storage (blood) at local site before shipping. Alternatively, they can be kept at room temperature for a few days. Storage for periods longer than this may lead to degradation of the DNA quality and need to repeat sample collection.

(ii) Packing instructions

1. **DNA Samples (Blood)**
   a) One blood sample in one biohazard bag.

b) All DNA samples into big ziplock bag (provided) as secondary container.

2. **Histology evidence**
   a) Ensure glass slides, paraffin blocks, biopsy samples, biopsy reports etc are de-identified. Use marker pens to remove the patient's identity and replace with the subject code.
   b) Place histology glass slides in the slide mailer box provided. **Cover of the slide mailer box needs to be taped down securely. Place 1 subject code label outside the mailer box.** Mailer box should be securely taped to the box or sufficiently cushioned to prevent breakage during shipment.
   Example of a slide mailer box:

   c) Shipment of biopsy samples in paraffin blocks or formalin solutions should be pre-arranged with the Main Coordinating Centre in Singapore.

3. **Forms**
   a) Shipping checklist (one per family) can be put into the box or envelope.
   b) Consent forms SHOULD NOT be sent to Singapore.
   c) Hard-copy phenotype forms (one per family) (ONLY FOR RENAL CYSTIC STUDY)

4. DNA samples (in ziplock bags), histology evidence and forms can all be put into any box or big envelope. Provide some cushioning by wrapping the samples with paper or bubble wraps.

(iii) *Shipping instructions (blood)*

1. FedEx will be the main designated courier company and the shipment will be paid by the Main Coordinating Centre (exceptions in some countries). A FedEx account number will be provided upon site activation.
2. All samples are to be declared as *"Blood. Non-infectious. Non-toxic. Non-hazardous. No commercial value. Strictly for research purposes only."*
3. Ship at room temperature.

(iv) *Frequency of shipping*

Samples are to be shipped every 3 to 4 weeks. If you expect more recruitment in the subsequent 3-4 weeks, then you may store the samples and ship by 4 weeks. Storage for longer than 4 weeks before shipping may lead to poor DNA quality and need to repeat sample collection.

l) Sample storage and future use in Singapore

Site investigators retain ownership of the materials sent, including any materials contained or incorporated in modifications.
Investigators at the Main Investigating Centre retain ownership of:

a) Modifications (except that, the provider retains ownership rights to the material included therein), and
b) Those substances created through the use of the materials or modifications, but which do not contain the original material, progeny or unmodified derivatives.

If either (a) or (b) results from the collaborative efforts of the investigators at providing site and Main Investigating Site, joint ownership may be negotiated.

"**Unmodified Derivatives**" mean substances created by the Main Coordinating Centre which constitute an unmodified functional subunit of the Original Material. Some examples include: Original Material or unmodified portions thereof fixed as tissue sections or in arrays, and unmodified proteins, proteins, RNA or DNA extracted from Original Material.

"**Progeny**" mean unmodified descendant from the Original Material, such as virus from virus, cell from cell, or organism from organism.

"**Modifications**" mean substances created by the Main Coordinating Centre which contain/incorporate the Material but which are not Unmodified Derivatives. For example, genetic modification or manipulation of cells extracted from the Original Material.

A Materials Transfer Agreement is strongly encouraged. Main Coordinating Centre will give written notice to providers if there are any new scope of research, including sample modifications.

The patients and their parents/legal guardians retain their rights to ask the investigators to discard or destroy any unused samples.
Specimens and their derived products obtained in this study **will be used for the purposes of this study**. Subjects and/or their parents/legal guardians need to indicate on the consent form if they agree to **storage of unused residual blood specimens and their additional derived products and use in future studies**.

.

- If they do not agree, the unused residual blood specimens and their derived genetic products will be **used solely for this study** and additional materials will be destroyed as biohazards at National University of Singapore, Laboratory of Paediatrics after completion of this study. Patient samples may be used for genetic testing **within this study's aims and scope**, up to **10 years** from subjects' recruitment. This long duration is because investigators need repeated testing on samples to confirm findings, and there may be newer techniques of gene testing in the future that may be used to allow better identification of genes.

- If they agree to such storage and future use, the unused residual blood specimens and their additional derived products not used in this study may be stored at National University of Singapore, Laboratory of Paediatrics for up to 20 years. It may be used in future research studies involving kidney diseases. These future studies would have to be approved by ethics committees in Singapore and the involved overseas institutions before it can be carried out. However, additional consent from the subjects for these future studies will not be sought. The de-identified samples (and their associated de-identified phenotype data) may be provided to other researchers or other institutions (including overseas) who collaborate with National University of Singapore, Laboratory of Paediatrics. Future work will include laboratory experiments or new techniques of gene sequencing to unravel the causes of kidney diseases. After 20 years, these will be disposed as biohazards at National University of Singapore, Laboratory of Paediatrics.

Site investigators will need to indicate on the web-based phenotype database if consent for storage has been obtained, and Singapore investigators will not proceed with the samples until this field has been entered.

Investigators at the Main Investigating Site may release the de-identified blood samples and their associated de-identified phenotype data to their sub-contractors and collaborators (local and/or overseas), and these will be used solely for the purposes of research in kidney diseases.

7. GENETIC ALGORITHMS AND MOLECULAR METHODS

Samples without adequate primary histological evidence may not be sent for genetic tests due to inadequate phenotype data.

For a start, depending on the phenotype of the subjects, samples will be sent for three possible genetic tests: GWAS, targeted gene sequencing and/or whole exome sequencing.

a) Next generation sequencing (NGS)

Targeted gene sequencing and/or exome sequencing will be performed for all the samples. For targeted gene sequencing, 264 genes that have been reported to be associated with glomerular diseases and renal cystic diseases have been identified (Table 1).[24-38] Customized gene chips will be designed based on the targeted genes. Any definitely or probably pathogenic variants will be validated using capillary sequencing.

Table 1: List of 264 known genes to be included in targeted gene sequencing

Gene	Protein	Gene ID
ACSL4	acyl-CoA synthetase long chain family member 4	2182
ACTN4	Actinin, alpha 4	81
ADAMTS13	ADAM metallopeptidase with thrombospondin type 1 motif, 13	11093
ADCK4	aarF domain containing kinase 4	79934
AGXT	alanine-glyoxylate aminotransferase	189
AHI1	Abelson helper integration site 1	54806
ALG1	ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase	56052
ALMS1	ALMS1, centrosome and basal body associated protein	7840
ANKS6	ankyrin repeat and sterile alpha motif domain containing 6	203286
ANLN	Anillin, actin binding protein	54443
ApoE	Apolipoprotein E	348
APOL1	Apolipoprotein L 1	8542
APRT	adenine phosphoribosyltransferase	353
ARHGAP24	Rho GTPase activating protein 24	83478
ARHGDIA	Rho GDP dissociation inhibitor (GDI) alpha	396
ARL13B	ADP ribosylation factor like GTPase 13B	200894
ARL6	ADP ribosylation factor like GTPase 6	84100
ARMC4	armadillo repeat containing 4	55130
ATP6B1	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B1	525
ATP6N1B	ATPase, H+ transporting, lysosomal V0 subunit a4	50617
AVPR2	arginine vasopressin receptor 2	554
B9D1	B9 domain containing 1	27077

B9D2	B9 domain containing 2	80776
BBS1	Bardet-Biedl syndrome 1	582
BBS2	Bardet-Biedl syndrome 2	583
BBS4	Bardet-Biedl syndrome 4	585
BBS5	Bardet-Biedl syndrome 5	129880
BBS7	Bardet-Biedl syndrome 7	55212
BBS9	Bardet-Biedl syndrome 9	27241
BBS10	Bardet-Biedl syndrome 10	79738
BBS12	Bardet-Biedl syndrome 12	166379
BICC1	BicC family RNA binding protein 1	80114
BMP4	bone morphogenetic protein 4	652
BMP7	bone morphogenetic protein 7	655
BSND	barttin CLCNK-type chloride channel accessory beta subunit	7809
C21orf59	chromosome 21 open reading frame 59	56683
C5orf42	chromosome 5 open reading frame 42	65250
CA2	carbonic anhydrase II	760
CC2D2A	coiled-coil and C2 domain containing 2A	57545
CCDC39	coiled-coil domain containing 39	339829
CCDC40	coiled-coil domain containing 40	55036
CCDC65	coiled-coil domain containing 65	85478
CCDC103	coiled-coil domain containing 103	388389
CCDC114	coiled-coil domain containing 114	93233
CCDC151	coiled-coil domain containing 151	115948
CCNO	cyclin O	10309
CD151	CD151 molecule (Raph blood group)	977
CD2AP	CD2-associated protein	23607
CDC5L	cell division cycle 5-like	988
CEP41	centrosomal protein 41	95681
CEP164	centrosomal protein 164	22897
CEP83	centrosomal protein 83	51134
CEP104	centrosomal protein 104	9731
CEP120	centrosomal protein 120	153241
CEP290	centrosomal protein 290	80184
CFH	Complement factor H	3075
CFHR1	complement factor H-related 1	3078
CFHR3	complement factor H-related 3	10878
CFHR5	Complement factor H-related 5	81494
CLCN5	chloride channel, voltage-sensitive 5	1184
CLCNKB	chloride channel, voltage-sensitive Kb	1188
CLDN16	claudin 16	10686
COQ7	coenzyme Q7, hydroxylase	10229
COQ9	coenzyme Q9	57017
CRB2	crumbs 2, cell polarity complex component/ crumbs family member 2	286204
CSPP1	centrosome and spindle pole associated protein 1	79848

CTLA4	Cytotoxic T-lymphocyte-associated protein 4	1493
COL4A1	Collagen, type IV, alpha 1	1282
COL4A2	Collagen, type IV, alpha 2	1284
COL4A3	Collagen, type IV, alpha 3	1285
COL4A4	Collagen, type IV, alpha 4	1286
COL4A5	Collagen, type IV, alpha 5	1287
COL4A6	Collagen, type IV, alpha 6	1288
COQ2	Coenzyme Q2 4-hydroxybenzoate polyprenyltransferase	27235
COQ6	Coenzyme Q6 monooxygenase	51004
CTNS	cystinosin, lysosomal cystine transporter	1497
CUBN	Cubilin (intrinsic factor-cobalamin receptor)	8029
CYP11B2	cytochrome P450, family 11, subfamily B, polypeptide 2	1585
DCDC2	doublecortin domain containing 2	51473
DDR1	discoidin domain receptor tyrosine kinase 1	780
DGKE	diacylglycerol kinase epsilon	8526
DLX5	distal-less homeobox 5	1749
DLX6	distal-less homeobox 6	1750
DNAAF1	dynein axonemal assembly factor 1	123872
DNAAF2	dynein axonemal assembly factor 2	55172
DNAAF3	dynein axonemal assembly factor 3	352909
DNAAF5	dynein axonemal assembly factor 5	54919
DNAH1	dynein axonemal heavy chain 1	25981
DNAH11	dynein axonemal heavy chain 11	8701
DNAH5	dynein axonemal heavy chain 5	1767
DNAH8	dynein axonemal heavy chain 8	1769
DNAI1	dynein axonemal intermediate chain 1	27019
DNAI2	dynein axonemal intermediate chain 2	64446
DNAL1	dynein axonemal light chain 1	83544
DRC1	dynein regulatory complex subunit 1	92749
DYNC2H1	dynein cytoplasmic 2 heavy chain 1	79659
DNAAF4 (DYX1C1)	dynein axonemal assembly factor 4	161582
DZIP1L	DAZ interacting zinc finger protein 1	199221
E2F3	E2F transcription factor 3	1871
EMP2	Epithelial membrane protein 2	2013
EVC	EvC ciliary complex subunit 1	2121
EVC2	EvC ciliary complex subunit	132884
EYA1	EYA transcriptional coactivator and phosphatase 1	2138
FAN1	FANCD2 and FANCI associated nuclease 1	22909
FAT1	FAT atypical cadherin 1	2195
FN1	Fibronectin 1	2335
FOXC1	forkhead box C1	2296
FOXH1	forkhead box H1	8928
FRAS1	Fraser extracellular matrix complex subunit 1	80144
FREM2	FRAS1 related extracellular matrix protein 2	341640

FXYD2	FXYD domain containing ion transport regulator 2	486
GAS8	growth arrest specific 8	2622
GATA3	GATA binding protein 3	2625
GDF1	growth differentiation factor 1	2657
GLA	galactosidase, alpha	2717
GLIS2	GLIS family zinc finger 2	84662
GPC5	glypican 5	2262
GRHPR	glyoxylate reductase/hydroxypyruvate reductase	9380
GRIP1	glutamate receptor interacting protein 1	23426
HNF1B	HNF1 homeobox B	6928
HNF4A	hepatocyte nuclear factor 4 alpha	3172
IFT122	intraflagellar transport 122	55764
IFT140	intraflagellar transport 140	9742
IFT172	intraflagellar transport 17	26160
IFT80	intraflagellar transport 80	57560
IL-4R	interleukin 4 receptor	3566
IL-6	interleukin 6	3569
IL-13	interleukin 13	3596
INF2	Inverted formin, FH2 and WH2 domain containing	64423
INPP5E	inositol polyphosphate-5-phosphatase E	56623
INVS	inversin	27130
IQCB1	IQ motif containing B1	9657
ITGA3	Integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor)	3675
ITGB4	Integrin, beta 4	3691
KANK1	KN motif and ankyrin repeat domains 1	23189
KANK2	KN motif and ankyrin repeat domains 2	25959
KANK4	KN motif and ankyrin repeat domains 4	163782
KCNJ1	potassium channel, inwardly rectifying subfamily J, member 1	3758
KIAA0586	KIAA0586	9786
KIF7	kinesin family member 7	374654
KIRREL	kin of IRRE like (Drosophila)	55243
LAMA5	laminin subunit alpha 5	3911
LAMB2	Laminin, beta 2	3913
LCORL	ligand dependent nuclear receptor corepressor-like	254251
LMNA	lamin A/C	4000
LMX1B	LIM homeobox transcription factor 1 beta	4010
LRP5	LDL receptor related protein 5	4041
LRRC6	leucine rich repeat containing 6	23639
MAGI2	membrane associated guanylate kinase, WW and PDZ domain containing 2	9863
MCIDAS	multiciliate differentiation and DNA synthesis associated cell cycle protein	345643
MED28	mediator complex subunit 28	80306
MET	MET proto-oncogene, receptor tyrosine kinase	4233
MKKS	McKusick-Kaufman syndrome	8195

*MKS1**	Meckel syndrome, type 1	54903
MRE11A	MRE11 homolog, double strand break repair nuclease	4361
MTTL1	mitochondrially encoded tRNA leucine 1 (UUA/G)	4567
MT-TL2	mitochondrially encoded tRNA leucine 2 (CUN)	4568
MT-TY	mitochondrially encoded tRNA tyrosine	4579
MYD88	myeloid differentiation primary response 88	4615
MYH9	Myosin, heavy chain 9, non-muscle	4627
MYO1E	Myosin IE	4643
MYOG	myogenin (myogenic factor 4)	4656
NEIL1	Nei endonuclease VIII-like 1	79661
NEK1	NIMA related kinase 1	4750
NEK8	NIMA related kinase 8	284086
NME8	NME/NM23 family member 8	51314
NODAL	nodal growth differentiation factor	4838
NOTCH2	notch 2	4853
*NPHP1**	nephronophthisis 1 (juvenile)	4867
*NPHP3**	nephronophthisis 3 (adolescent)	27031
*NPHP4**	nephronophthisis 4	261734
NPHS1	Nephrin	4868
NPHS2	Podocin	7827
NUP107	nucleoporin 107	57122
NUP205	nucleoporin 205	23165
NUP93	nucleoporin 93	9688
NXF5	Nuclear RNA export factor 5	55998
OCRL 1	oculocerebrorenal syndrome of Lowe	4952
OFD1	OFD1, centriole and centriolar satellite protein	8481
SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	5054
PAX2	Paired box 2	5076
PDE6D	phosphodiesterase 6D	5147
PDSS2	Prenyl (decaprenyl) diphosphate synthase, subunit 2	57107
PKD1	polycystic kidney disease 1	5310
*PKD2**	polycystic kidney disease 2	5311
*PKHD1**	polycystic kidney and hepatic disease 1	5314
PMM2	Phosphomannomutase 2	5373
PLCE1	Phospholipase C, epsilon 1	51196
PODXL	Podocalyxin-like	5420
POSTN	periostin	10631
PTPRC (CD45)	protein tyrosine phosphatase, receptor type, C	5788
PTPRO/GLEPP1	Protein tyrosine phosphatase, receptor type, O	5800
RET	ret proto-oncogene	5979
ROBO2	roundabout guidance receptor 2	6092
RPGRIP1L	RPGRIP1 like	23322
RSPH4A	radial spoke head 4 homolog A	345895
RSPH1	radial spoke head 1 homolog	89765

RSPH3	radial spoke head 3 homolog	83861
RSPH4A	radial spoke head 4 homolog	345895
RSPH9	radial spoke head 9 homolog	221421
SALL1	spalt-like transcription factor 1	6299
SCARB2	Scavenger receptor class B, member 2	950
SCNN1A	sodium channel, non voltage gated 1 alpha subunit	6337
SCNN1B	sodium channel, non voltage gated 1 beta subunit	6338
SCNN1G	sodium channel, non voltage gated 1 gamma subunit	6340
SDCCAG8	serologically defined colon cancer antigen 8	10806
SGPL1	sphingosine-1-phosphate lyase	8879
SIX1	SIX homeobox 1	6495
SIX2	SIX homeobox 2	10736
SIX5	SIX homeobox 5	147912
SLC12A1	solute carrier family 12 (sodium/potassium/chloride transporter), member 1	6557
SLC12A3	solute carrier family 12 (sodium/chloride transporter), member 3	6559
SLC3A1	solute carrier family 3 (amino acid transporter heavy chain), member 1	6519
SLC4A1	solute carrier family 4 (anion exchanger), member 1 (Diego blood group)	6521
SLC4A4	solute carrier family 4 (sodium bicarbonate cotransporter), member 4	8671
SLC41A1	solute carrier family 41 member 1	254428
SLC7A9	solute carrier family 7 (amino acid transporter light chain, bo,+ system), member 9	11136
SLC9A3R1	solute carrier family 9, subfamily A (NHE3, cation proton antiporter 3), member 3 regulator 1	9368
SLCA1	solute carrier family 2 (facilitated glucose transporter), member 1	6513
SLIT2	slit guidance ligand 2	9353
SMARCAL1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a-like 1	50485
SPAG1	sperm associated antigen 1	6674
STAT4	signal transducer and activator of transcription 4	6775
STAT6	signal transducer and activator of transcription 6, interleukin-4 induced	6778
SYNPO	Synaptopodin	11346
TCTN1	tectonic family member 1	79600
TCTN2	tectonic family member 2	79867
TCTN3	tectonic family member 3	26123
TLR4	toll-like receptor 4	7099
TLR9	toll-like receptor 9	54106
TMEM138	transmembrane protein 138	51524
TMEM216	transmembrane protein 216	51259
TMEM237	transmembrane protein 237	65062
TMEM67	transmembrane protein 67	91147
TMEM231	transmembrane protein 231	79583
TP63	tumor protein p63	8626
TRIM32	tripartite motif containing 32	22954
TRPC6	Transient receptor potential cation channel, subfamily C, member 6	7225
TSC1	uberous sclerosis 1	7248

TSC2	tuberous sclerosis 2	7249
*TTC21B**	Tetratricopeptide repeat domain 21B	79809
TTC8	tetratricopeptide repeat domain 8	123016
UMOD	uromodulin	7369
UPk3A	uroplakin 3A	7380
VHL	von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase	7428
WDPCP	WD repeat containing planar cell polarity effector	51057
WDR19	WD repeat domain 19	57728
WDR34	WD repeat domain 34	89891
WDR35	WD repeat domain 35	57539
WDR60	WD repeat domain 60	55112
WDR73	WD repeat domain 73	84942
WNK1	WNK lysine deficient protein kinase 1	65125
WNK4	WNK lysine deficient protein kinase 4	65266
WT1	Wilms tumor 1	7490
XDH	xanthine dehydrogenase	7498
XPNPEP3	X-prolyl aminopeptidase 3	63929
XPO5	exportin 5	57510
ZEB1	zinc finger E-box binding homeobox 1	6935
ZIC3	Zic family member 3	7547
ZMPSTE24	Zinc metallopeptidase STE24	10269
ZMYND10	zinc finger MYND-type containing 10	51364
ZNF423	zinc finger protein 423	23090
tRNALeuA3243G	Linked to MELAS	
$tRNA^{Tyr}$ gene	FSGS case study	

Selective **exome sequencing** will be performed by hybridization of fragmented genomic DNA with oligonucleotide probes corresponding to each exonic sequence assembled in one of three platforms (Agilent, Nimblegen, and Illumina) and the captured fragments are then sequenced by next-generation sequencing machines. A multiplexing and pooled sequencing approach will be used which allows sequencing an exome at low cost with high coverage. Samples with known pathogenic variants will be used as positive controls. Any variants will be validated using capillary sequencing.

After exome sequencing, we will curate the variants according to ACMG-based[87] algorithm as described below:

1. The pathogenicity of the identified variants was evaluated based on the American College of Medical Genetics and Genomics guidelines for clinical sequence interpretation.[87]
2. Only variants in coding exons and canonical splice sites were considered (except for intron 9 of *WT1* gene analyzed up to position +6).
3. Variants are considered disease-causing if they are truncating, frameshift, essential splice or previously reported pathogenic mutations
4. Missense variants are considered disease-causing if

- non-synonymous, and
- combined annotation–dependent depletion (CADD) scores of ≥15,[88] and

- deemed deleterious in at least 2 out of 3 in silico algorithms, namely PolyPhen-2,[89], SIFT (Sorting Intolerant from Tolerant), [90,91] and MutationTaster, [92].
- minor allele frequencies, based on exomes in Genome Aggregation Database (gnomAD) v2.1.1 [93](https://gnomad.broadinstitute.org/), GenomeAsia [94](https://genomeasia100k.org/), were below 0.01 for dominant genes or below 0.05 for recessive genes (if available). Where available, the allele frequencies based on the pilot study of SG10K project, a large-scale sequencing study in Singapore, were also checked[95].
- Remaining variants were checked in known databases, namely ClinVar https://www.ncbi.nlm.nih.gov/clinvar/), Varsome (https://varsome.com/),[96]Leiden Open-source Variation Database LOVD v.3.0 (https://www.lovd.nl/)[97]and the Alport syndrome COL4A5 variant database (http://arup.utah.edu/database/ALPORT/ALPORT_welcome.php) [98]where applicable.

Identified variants were classified into 3 groups: Group 1 – (likely) pathogenic variants deemed disease-causing; Group 2 – (likely) pathogenic variants, possibly contributing to disease risk, but were not enough to explain the phenotypes and Group 3 – variants of uncertain significance.

b) Genome wide association studies (GWAS)

This will only be sent for cases with sporadic primary FSGS with adequate primary histological evidence. **All** of the following 4 criteria must be satisfied:

1) Nephrotic syndrome
2) Biopsy-proven primary FSGS (with adequate primary histological evidence)
3) Onset before 25 years old
4) No family history of primary glomerular diseases

(See "Definitions")

GWAS will be performed as previously described.[99] For stage 1, genome-wide genotyping will be performed using the Illumina 610K Quad BeadChip, following the manufacturer's instructions. For stage 2 (replication stage), genotyping was performed using the Sequenom MassArray platform or TaqMan probes (Applied Biosystems). If potentially significant genetic variants are identified, the variants will be validated using capillary sequencing.

Statistical analysis

The patients will be analysed according to ethnic groups/geographic location if necessary. In addition, Genome Institute of Singapore has genetic data of >3000 anonymous local and regional controls in various ethnicities obtained from previous and existing studies. These include countries from collaborating countries like Myanmar, the Philippines, Thailand, Vietnam and Indonesia. Consent for these controls have been obtained from these previous and existing studies. Data from international HapMap will be also used for analysis.

The PLINK software will be used for primary association testing, as well as modeling within a linear regression framework. Individual genotypes will be coded according to the number of copies of the variant allele present: 0 for the wild-type genotype, 1 for heterozygous and 2 for homozygous variants. A trend test incorporated within a linear regression model will be used for primary association testing between genotypes and optic disc area as a quantitative trait, adjusting for age, gender and genetic ancestry.

Manhattan (210 log p-plots) and LD plots will be created using Haploview. Q–Q and regional association plots will be created using the software R [www.r-project.org]. Meta-analysis of results across cohorts will be performed using the inverse-variance method, as previously described for quantitative traits. This method weighs each study according to effective sample size and cohort-specific MAF of the associated variants.

To avoid an otherwise unacceptable number of false positive signals as an artifact of multiple testing, the formal threshold for genome-wide significance, P less than 5.00×10^{-8}, will be considered to be statistically significant. Recruitment target for each site is derived based on estimated number of patients being followed up there.

c) Functional work

If definitely or probably pathogenic variants are identified, and these are either novel variants in known genes or variants in novel genes, *in vitro* functional work may be performed. The type of cells used will depend on the function of the gene involved. Investigators in Singapore Main Coordinating Centre had experience in podocyte cultures, HEK293 cells and tubular epithelial cells, including transfection or lentiviral transduction of genetic variants. If potentially significant variants are found in different genes and interactions between genes are suspected, then co-transfection of these variants may be performed in cells to determine their function. If initial *in vitro* work proves to be promising, further *in vivo* work involving mice, rats or zebrafish may be performed in subsequent research.

Immunofluoresence, immunohistochemistry or in-situ hydridisation can also be performed in kidney (from humans, mice or rats) or in cells to further delineate the function of these genes or variants.

8. SAMPLE SIZE

GWAS: The initial phase of the study will be at the discovery stage, recruitment of 500 cases and 3000 controls will yield power of more than 80% in genes with minimum minor allele frequency (MAF) of 0.40 and minimum allele odds ratio of 1.30 (a less than 1 x 10-4).[99,100] (Note: the 3000 controls are already in the databases of Genome Institute of Singapore and need not be recruited.)

NGS: No sample size determination is necessary since these are relatively rare diseases.

9. DATA HANDLING

a) Data entry and storage

REDCap® (https://redcap.scri.edu.sg/), a web-based phenotype database has been set up to allow site investigators to enter baseline and follow-up phenotype data. All data entered will be de-identified. A user-specific password will be given to each user for data entry. Site investigators will enter the phenotype data after reading through the instructions for the database.

For the renal cystic study, site investigators will fill up hardcopy forms which will later be entered into REDCap by Singapore team.

For whole-slide imaging performed on slides or renal biopsy samples sent to Singapore, patient-specific folders containing digital images and copies of the report will be kept. All data will be stored in password-protected computers in Main Coordinating Centre and in the case of the digital images, it will also be stored in Singapore General Hospital where the digital slide camera and pathologists are located.

All data containing identifiable data will be stored for a period of 20 years, so that comparisons of data with future studies can be made. Only study investigators will have access to the password. All phenotype data are de-identified. The study data in its deidentified form will be stored for a longer period of 30 years. This is because these are precious data on very rare diseases, and accumulation of these data will allow clinicians to better understand the disease over time, using large number of cases.

b) Data quality assurance

The database will be managed by the Main Coordinating Centre investigators and designated research staff, and will be accessed only through passwords. This database will be hosted in a secured server with security features, and backed up in a hard drive.
Data submitted to the Main Coordinating Centre will be monitored for legibility, completeness, and for incongruent information. Only Main Coordinating Centre and designated research staff will have rights to export the data. Exported data will be password protected, and the password will only be known by the study team members. The password will be changed periodically.

The database will incorporate built-in validation audits and flag entries where information is not in accordance with established rules. Site investigators will need to correct data to resolve the errors flagged by database. Research staff in Singapore will validate and audit the data entered.

Site investigators can own their own centre data. They can request for their center data in common formats such as Microsoft Excel, PDF, SAS, Stata, R, or SPSS formats. Centers within the same country can also request for nationwide data provided written approval is obtained from investigators from all centers.

10. RETURN OF GENETIC TEST RESULTS

Return of genetic results of subjects from specific local sites will be returned from Main Coordinating Centre to the site investigators who provided the samples and phenotype data. Genetic results will not be returned to investigators from other sites. If patient/family chooses to know the genetic test results, the results should be communicated to the patient/family by a local genetic counsellor, geneticist or a doctor with adequate genetic counselling skills. Investigators from Main Coordinating Centres may assist site investigators on the interpretation of the results, but will not provide genetic counselling directly to the patients.

For next generation sequencing, changes in the genes not relevant to kidney diseases will not be analysed due to constraints in resources. Hence, analysis of medically actionable non-renal genes will not be conducted in Main Coordinating Centre. However, if the patient/family and the site investigator request, the raw data of genetic sequencing results of specific or all patients may be given to the site investigator, who can then decide to give to the patients and their families. Families may choose to forward the raw data to other parties for analysis.

11. POTENTIAL RISKS AND BENEFITS

Anticipated risks involve only those related to routine venepuncture like pain, bleeding, bruising or swelling at the site of the needle stick and very rarely fainting and infection.

In the process of genetic testing, we may discover (likely) pathogenic gene variants which are transmitted from either or both parents, and parents may feel guilty as a result. Detection of such

genetic variants may also affect the subjects' ability to gain future employment or insurance coverage. Such results may be entered into the subjects' medical records by local doctors. In addition, non-paternity may be discovered in the process. Each subject and their recruited immediate family should be provided with genetic counselling by local doctors as to make them understand the implication and possibility that might arise in the future (if any discovery is made).

Only genes related to kidney conditions will be analysed by Singapore investigators. Genes unrelated to kidney diseases will not be analysed by investigators in Singapore. Subjects may however, request for raw genetic sequencing data to forward to third-party experts for analysis. During this process, they may discover incidental higher genetic predisposition for non-renal related conditions eg cancer or Alzheimer's disease for which preventive therapies may or may not be available. This may cause anxiety in patient and families.

Many subjects may not benefit from participation in this study.

In some cases, this study may potentially identify disease-causing genetic variants. This may help clinicians in deciding the most appropriate immunosuppressive treatment options. If the patients need a kidney transplant in the future, results from this study may also help the doctors to decide on the most suitable live donor, and the most appropriate drug regimens during and after the transplant. In addition, by knowing the gene changes, the doctors can provide genetic counselling to the patients and the family. Family members may also be identified to have early kidney disease as a result of this study. Identification of potential disease-causing variants in (still) healthy family members can also allow early and close surveillance to ensure early diagnosis and interventions, and this may slow down or halt the eventual progression of the disease.

Their participation in this study may add to the medical knowledge on their (child's) condition, and may benefit others as well. This study may eventually identify genetic variants or polymorphisms which may allow doctors to estimate the risk of occurrence of glomerular diseases, as well as the risk of poor outcomes such as renal failure, in Asians. The results of this study can form the basis for future work to allow doctors to better understand the pathogenesis.

The results of this study can form the basis for future work to allow doctors to better understand the pathogenesis of primary proteinuric glomerular disease.

12. ETHICAL CONSIDERATIONS

a) Institutional Review Board (IRB) / Ethics Board and Informed Consent

Each participating centre is to submit to their respective Institutional Review Board (IRB)/Ethics Board for approval of this study. A copy of the approval letter and approved consent form are to be forwarded to the Main Coordinating Site. All regulatory documents are to be kept 10 years after the completion of the study. All study documents in hard copies will be destroyed (shredded) and then discarded 10 years after completion of the study.

b) Confidentiality of Data and Patient Records

To ensure confidentiality of data and patient records, the research data will be stored in the Main Coordinating Centre in a stand-alone computer locked with a password. Only study investigators and designated research assistants will have access to the records. Data collected are the property of National University of Singapore. Individual site investigators also own the rights to their institution's data. The subjects' records, to the extent of the applicable laws and regulations (both

local and overseas) , will not be made publicly available. In the event of any publication regarding this study, the subjects' identities will remain confidential.

13. COMPENSATION FOR INJURY

The risks involved in this study are very minimal and involved only blood collection. Hence, if the subjects are physically injured due to the study procedures in this study, the Main Coordinating Centre in Singapore are not able to pay for the medical expenses for the treatment of that injury. Claims arising from the study may be covered by Medical Indemnity Insurance of the local investigators.

14. FINANCIAL CONSIDERATIONS

This study is sponsored by National Medical Research Council, Singapore, in two grants (NMRC/CSA/0057/2013 and NMRC/CSA-INV/0015/2017).

Subjects and investigators will not receive monetary benefits from participating in this study. The Main Coordinating Centre will provide each site standard consumables (See site activation). Site investigators may contact the Main Coordinating Centre if they require additional supplies.

Subjects will not be expected to pay for any costs incurred in this study. They will have to pay for their standard clinical care. In addition, if the urine dipstick for presumably healthy family members turn out to be abnormal, and if further investigations such as urine microscopy, formal laboratory urine protein quantification or serum creatinine are necessary to further confirm the phenotype of the family member, these cannot be paid for by the Main Coordinating Site, and will have to be borne by the subjects.

The risks involved in this study are very minimal and involved only blood collection. Hence, if the subjects are physically injured due to the study procedures in this study, the Main Coordinating Centre in Singapore are not able to pay for the medical expenses for the treatment of that injury.

The Main Coordinating Centre will pay for the shipping costs involved in shipping items in and out of the sites. Manpower costs involved in identifying and recruiting subjects, sample collections and data collections will not be paid for.

15. PUBLICATIONS

All investigators will be offered co-authorship in all publications resulting from this study. Patients will not be identified in any publications as a result of this study.

If the investigator wants results of all their patients, an email from the site investigator to agree that investigators of the Main Coordinating Centre will be co-authors in the collaborator-generated publications will be requested.

16. RETENTION OF STUDY DOCUMENTS

All study documents in their identifiable form are to be kept 20 years after the completion of the study. All study data in its deidentified form are to be kept 30 years after the completion of the study.

17. REFERENCES

1. Churg, J., Habib, R. & White, R.H. Pathology of the nephrotic syndrome in children: a report for the International Study of Kidney Disease in Children. *Lancet* **760**, 1299-1302 (1970).

2. Glassock, R.J. Diagnosis and natural course of membranous nephropathy. *Semin Nephrol* **23**, 324-332 (2003).
3. Nephrotic syndrome in children: prediction of histopathology from clinical and laboratory characteristics at time of diagnosis. A report of the International Study of Kidney Disease in Children. *Kidney Int* **13**, 159-165 (1978).
4. Kitiyakara, C., Kopp, J.B. & Eggers, P. Trends in the epidemiology of focal segmental glomerulosclerosis. *Semin Nephrol* **23**, 172-182 (2003).
5. U.S. Renal Data System, USRDS 2012 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States. . (National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2012).
6. Guay-Woodford, L.M. Autosomal recessive polycystic kidney disease: the prototype of the hepato-renal fibrocystic diseases. *Journal of pediatric genetics* **3**, 89-101 (2014).
7. Hartung, E.A. & Guay-Woodford, L.M. Autosomal recessive polycystic kidney disease: a hepatorenal fibrocystic disorder with pleiotropic effects. *Pediatrics* **134**, e833-845 (2014).
8. Cramer, M.T. & Guay-Woodford, L.M. Cystic kidney disease: a primer. *Advances in chronic kidney disease* **22**, 297-305 (2015).
9. Chen, J.*, et al.* A rare deep intronic mutation of PKHD1 gene, c.8798-459 C > A, causes autosomal recessive polycystic kidney disease by pseudoexon activation. *Journal of human genetics* (2019).
10. Hartung, E.A. & Guay-Woodford, L.M. Polycystic kidney disease: DZIP1L defines a new functional zip code for autosomal recessive PKD. *Nature reviews. Nephrology* **13**, 519-520 (2017).
11. Lu, H.*, et al.* Mutations in DZIP1L, which encodes a ciliary-transition-zone protein, cause autosomal recessive polycystic kidney disease. *Nature genetics* **49**, 1025-1034 (2017).
12. Garel, J.*, et al.* Prenatal ultrasonography of autosomal dominant polycystic kidney disease mimicking recessive type: case series. *Pediatric radiology* (2019).
13. Szabo, T.*, et al.* Comprehensive genetic testing in children with a clinical diagnosis of ARPKD identifies phenocopies. *Pediatric nephrology (Berlin, Germany)* **33**, 1713-1721 (2018).
14. Gimpel, C.*, et al.* Imaging of Kidney Cysts and Cystic Kidney Diseases in Children: An International Working Group Consensus Statement. *Radiology*, 181243 (2018).
15. Bergmann, C. Early and Severe Polycystic Kidney Disease and Related Ciliopathies: An Emerging Field of Interest. *Nephron* **141**, 50-60 (2019).
16. Bergmann, C. ARPKD and early manifestations of ADPKD: the original polycystic kidney disease and phenocopies. *Pediatric nephrology (Berlin, Germany)* **30**, 15-30 (2015).
17. Fogo, A.B. Causes and pathogenesis of focal segmental glomerulosclerosis. *Nature reviews. Nephrology* (2014).
18. Harita, Y. Application of next-generation sequencing technology to diagnosis and treatment of focal segmental glomerulosclerosis. *Clin Exp Nephrol* **22**, 491-500 (2018).
19. Dryer, S.E. & Reiser, J. TRPC6 channels and their binding partners in podocytes: role in glomerular filtration and pathophysiology. *Am J Physiol Renal Physiol* **299**, F689-701 (2010).
20. Kopp, J.B.*, et al.* MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nat Genet* **40**, 1175-1184 (2008).
21. Kopp, J.B., Winkler, C.A. & Nelson, G.W. MYH9 genetic variants associated with glomerular disease: what is the role for genetic testing? *Semin Nephrol* **30**, 409-417 (2010).
22. Freedman, B.I. APOL1 and nephropathy progression in populations of African ancestry. *Semin Nephrol* **33**, 425-432 (2013).
23. Akchurin, O. & Reidy, K.J. Genetic causes of proteinuria and nephrotic syndrome: impact on podocyte pathobiology. *Pediatric nephrology (Berlin, Germany)* **30**, 221-233 (2015).
24. McCarthy, H.J. & Saleem, M.A. Genetics in clinical practice: nephrotic and proteinuric syndromes. *Nephron Exp Nephrol* **118**, e1-8 (2011).
25. Hildebrandt, F. Genetic kidney diseases. *Lancet* **375**, 1287-1295 (2010).

26. McCarthy, H.J., *et al.* Simultaneous sequencing of 24 genes associated with steroid-resistant nephrotic syndrome. *Clin J Am Soc Nephrol* **8**, 637-648 (2013).
27. Ashraf, S., *et al.* ADCK4 mutations promote steroid-resistant nephrotic syndrome through CoQ10 biosynthesis disruption. *J Clin Invest* **123**, 5179-5189 (2013).
28. Malaga-Dieguez, L. & Susztak, K. ADCK4 "reenergizes" nephrotic syndrome. *J Clin Invest* **123**, 4996-4999 (2013).
29. Gee, H.Y., *et al.* ARHGDIA mutations cause nephrotic syndrome via defective RHO GTPase signaling. *J Clin Invest* **123**, 3243-3253 (2013).
30. Gupta, I.R., *et al.* ARHGDIA: a novel gene implicated in nephrotic syndrome. *J Med Genet* **50**, 330-338 (2013).
31. Boyer, O., *et al.* LMX1B mutations cause hereditary FSGS without extrarenal involvement. *J Am Soc Nephrol* **24**, 1216-1222 (2013).
32. Heeringa, S.F., *et al.* COQ6 mutations in human patients produce nephrotic syndrome with sensorineural deafness. *J Clin Invest* **121**, 2013-2024 (2011).
33. Cong, E.H., *et al.* A Homozygous Missense Mutation in the Ciliary Gene TTC21B Causes Familial FSGS. *J Am Soc Nephrol* (2014).
34. Gbadegesin, R.A., *et al.* Mutations in the gene that encodes the F-actin binding protein anillin cause FSGS. *J Am Soc Nephrol* **25**, 1991-2002 (2014).
35. Gee, H.Y., *et al.* Mutations in EMP2 cause childhood-onset nephrotic syndrome. *Am J Hum Genet* **94**, 884-890 (2014).
36. Esposito, T., *et al.* Unique X-linked familial FSGS with co-segregating heart block disorder is associated with a mutation in the NXF5 gene. *Hum Mol Genet* **22**, 3654-3666 (2013).
37. Barua, M., *et al.* Mutations in PAX2 Associate with Adult-Onset FSGS. *J Am Soc Nephrol* **25**, 1942-1953 (2014).
38. Barua, M., *et al.* Exome sequencing and in vitro studies identified podocalyxin as a candidate gene for focal and segmental glomerulosclerosis. *Kidney Int* **85**, 124-133 (2014).
39. Gee, H.Y., *et al.* FAT1 mutations cause a glomerulotubular nephropathy. *Nature communications* **7**, 10822 (2016).
40. Lovric, S., *et al.* Mutations in sphingosine-1-phosphate lyase cause nephrosis with ichthyosis and adrenal insufficiency. *J Clin Invest* **127**, 912-928 (2017).
41. Prasad, R., *et al.* Sphingosine-1-phosphate lyase mutations cause primary adrenal insufficiency and steroid-resistant nephrotic syndrome. *J Clin Invest* **127**, 942-953 (2017).
42. Zhao, F., *et al.* Mutations in NUP160 Are Implicated in Steroid-Resistant Nephrotic Syndrome. *J Am Soc Nephrol* (2019).
43. Hinkes, B.G., *et al.* Nephrotic syndrome in the first year of life: two thirds of cases are caused by mutations in 4 genes (NPHS1, NPHS2, WT1, and LAMB2). *Pediatrics* **119**, e907-919 (2007).
44. Lovric, S., *et al.* Rapid detection of monogenic causes of childhood-onset steroid-resistant nephrotic syndrome. *Clin J Am Soc Nephrol* **9**, 1109-1116 (2014).
45. Sadowski, C.E., *et al.* A Single-Gene Cause in 29.5% of Cases of Steroid-Resistant Nephrotic Syndrome. *J Am Soc Nephrol* (2014).
46. Giglio, S., *et al.* Heterogeneous genetic alterations in sporadic nephrotic syndrome associate with resistance to immunosuppression. *J Am Soc Nephrol* **26**, 230-236 (2015).
47. Trautmann, A., *et al.* Spectrum of steroid-resistant and congenital nephrotic syndrome in children: the PodoNet registry cohort. *Clin J Am Soc Nephrol* **10**, 592-600 (2015).
48. Lipska, B.S., *et al.* Genetic screening in adolescents with steroid-resistant nephrotic syndrome. *Kidney Int* **84**, 206-213 (2013).
49. Lovric, S., Ashraf, S., Tan, W. & Hildebrandt, F. Genetic testing in steroid-resistant nephrotic syndrome: when and how? *Nephrol Dial Transplant* **31**, 1802-1813 (2016).

50. Wang, Y.*, et al.* Mutation spectrum of genes associated with steroid-resistant nephrotic syndrome in Chinese children. *Gene* **625**, 15-20 (2017).
51. Deltas, C., Pierides, A. & Voskarides, K. Molecular genetics of familial hematuric diseases. *Nephrol Dial Transplant* **28**, 2946-2960 (2013).
52. Lowik, M.*, et al.* Bigenic heterozygosity and the development of steroid-resistant focal segmental glomerulosclerosis. *Nephrol Dial Transplant* **23**, 3146-3151 (2008).
53. Feltran, L.S.*, et al.* Targeted Next-Generation Sequencing in Brazilian Children With Nephrotic Syndrome Submitted to Renal Transplant. *Transplantation* **101**, 2905-2912 (2017).
54. McCarthy, H.J.*, et al.* Simultaneous Sequencing of 24 Genes Associated with Steroid-Resistant Nephrotic Syndrome. *Clin J Am Soc Nephrol* (2013).
55. Anderson, M., Kim, E.Y., Hagmann, H., Benzing, T. & Dryer, S.E. Opposing effects of podocin on the gating of podocyte TRPC6 channels evoked by membrane stretch or diacylglycerol. *Am J Physiol Cell Physiol* **305**, C276-289 (2013).
56. Lovric, S.*, et al.* Rapid detection of monogenic causes of childhood-onset steroid-resistant nephrotic syndrome. *Clinical journal of the American Society of Nephrology : CJASN* **9**, 1109-1116 (2014).
57. Sen, E.S.*, et al.* Clinical genetic testing using a custom-designed steroid-resistant nephrotic syndrome gene panel: analysis and recommendations. *Journal of medical genetics* **54**, 795-804 (2017).
58. Santin, S.*, et al.* Clinical utility of genetic testing in children and adults with steroid-resistant nephrotic syndrome. *Clin J Am Soc Nephrol* **6**, 1139-1148 (2011).
59. Buscher, A.K.*, et al.* Immunosuppression and renal outcome in congenital and pediatric steroid-resistant nephrotic syndrome. *Clin J Am Soc Nephrol* **5**, 2075-2084 (2010).
60. Gbadegesin, R.A., Winn, M.P. & Smoyer, W.E. Genetic testing in nephrotic syndrome--challenges and opportunities. *Nature reviews. Nephrology* **9**, 179-184 (2013).
61. Bouchireb, K.*, et al.* NPHS2 mutations in steroid-resistant nephrotic syndrome: a mutation update and the associated phenotypic spectrum. *Hum Mutat* **35**, 178-186 (2014).
62. Laurin, L.P.*, et al.* Podocyte-associated gene mutation screening in a heterogeneous cohort of patients with sporadic focal segmental glomerulosclerosis. *Nephrol Dial Transplant* (2014).
63. Oleksyk, T.K., Nelson, G.W., An, P., Kopp, J.B. & Winkler, C.A. Worldwide distribution of the MYH9 kidney disease susceptibility alleles and haplotypes: evidence of historical selection in Africa. *PLoS One* **5**, e11474 (2010).
64. Lee, J.H.*, et al.* Genetic basis of congenital and infantile nephrotic syndromes. *Am J Kidney Dis* **58**, 1042-1043 (2011).
65. Sako, M.*, et al.* Analysis of NPHS1, NPHS2, ACTN4, and WT1 in Japanese patients with congenital nephrotic syndrome. *Kidney Int* **67**, 1248-1255 (2005).
66. Caridi, G., Perfumo, F. & Ghiggeri, G.M. NPHS2 (Podocin) mutations in nephrotic syndrome. Clinical spectrum and fine mechanisms. *Pediatr Res* **57**, 54R-61R (2005).
67. Weber, S.*, et al.* NPHS2 mutation analysis shows genetic heterogeneity of steroid-resistant nephrotic syndrome and low post-transplant recurrence. *Kidney Int* **66**, 571-579 (2004).
68. Ruf, R.G.*, et al.* Patients with mutations in NPHS2 (podocin) do not respond to standard steroid treatment of nephrotic syndrome. *J Am Soc Nephrol* **15**, 722-732 (2004).
69. Yu, Z.*, et al.* Mutations in NPHS2 in sporadic steroid-resistant nephrotic syndrome in Chinese children. *Nephrol Dial Transplant* **20**, 902-908 (2005).
70. Maruyama, K.*, et al.* NPHS2 mutations in sporadic steroid-resistant nephrotic syndrome in Japanese children. *Pediatr Nephrol* **18**, 412-416 (2003).
71. Kitamura, A.*, et al.* Genetics and clinical features of 15 Asian families with steroid-resistant nephrotic syndrome. *Nephrol Dial Transplant* **21**, 3133-3138 (2006).
72. Wang, F.*, et al.* Spectrum of mutations in Chinese children with steroid-resistant nephrotic syndrome. *Pediatr Nephrol* **32**, 1181-1192 (2017).

73. Tsukaguchi, H., *et al.* NPHS2 mutations in late-onset focal segmental glomerulosclerosis: R229Q is a common disease-associated allele. *J Clin Invest* **110**, 1659-1666 (2002).
74. Lu, L., *et al.* The p.R229Q variant of the NPHS2 (podocin) gene in focal segmental glomerulosclerosis and steroid-resistant nephrotic syndrome: a meta-analysis. *International urology and nephrology* **46**, 1383-1393 (2014).
75. Santin, S., *et al.* Clinical value of NPHS2 analysis in early- and adult-onset steroid-resistant nephrotic syndrome. *Clin J Am Soc Nephrol* **6**, 344-354 (2011).
76. Kar-Hui Ng, C.-K.H., Chiea-Chuen Khor, Hui-Kim Yap. Genes in FSGS: Diagnostic and Management Strategies in Children. *Curr Pediatr Rep* (2014).
77. Ogino, D., *et al.* Analysis of the genes responsible for steroid-resistant nephrotic syndrome and/or focal segmental glomerulosclerosis in Japanese patients by whole-exome sequencing analysis. *J Hum Genet* **61**, 137-141 (2016).
78. Renkema, K.Y., Stokman, M.F., Giles, R.H. & Knoers, N.V.A.M. Next-generation sequencing for research and diagnostics in kidney disease. *Nature reviews. Nephrology* **10**, 433-444 (2014).
79. Gadegbeku, C.A., *et al.* Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney Int* **83**, 749-756 (2013).
80. Stokman, M.F., *et al.* The expanding phenotypic spectra of kidney diseases: insights from genetic studies. *Nature reviews. Nephrology* **12**, 472-483 (2016).
81. Alzarka, B., Morizono, H., Bollman, J.W., Kim, D. & Guay-Woodford, L.M. Design and Implementation of the Hepatorenal Fibrocystic Disease Core Center Clinical Database: A Centralized Resource for Characterizing Autosomal Recessive Polycystic Kidney Disease and Other Hepatorenal Fibrocystic Diseases. *Frontiers in pediatrics* **5**, 80 (2017).
82. Sadowski, C.E., *et al.* A single-gene cause in 29.5% of cases of steroid-resistant nephrotic syndrome. *J Am Soc Nephrol* **26**, 1279-1289 (2015).
83. Barisoni, L., *et al.* Digital pathology evaluation in the multicenter Nephrotic Syndrome Study Network (NEPTUNE). *Clin J Am Soc Nephrol* **8**, 1449-1459 (2013).
84. Arnold, M.A., *et al.* The College of American Pathologists guidelines for whole slide imaging validation are feasible for pediatric pathology: a pediatric pathology practice experience. *Pediatr Dev Pathol* **18**, 109-116 (2015).
85. Higgins, C. Applications and challenges of digital pathology and whole slide imaging. *Biotech Histochem* **90**, 341-347 (2015).
86. Nast, C.C., *et al.* Morphology in the Digital Age: Integrating High-Resolution Description of Structural Alterations With Phenotypes and Genotypes. *Semin Nephrol* **35**, 266-278 (2015).
87. Richards, S., *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424 (2015).
88. Kircher, M., *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-315 (2014).
89. Adzhubei, I.A., *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249 (2010).
90. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-1081 (2009).
91. Sim, N.L., *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* **40**, W452-457 (2012).
92. Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* **11**, 361-362 (2014).
93. Karczewski, K.J., *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).

94. GenomeAsia, K.C. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* **576**, 106-111 (2019).
95. Wu, D.*, et al.* Large-Scale Whole-Genome Sequencing of Three Diverse Asian Populations in Singapore. *Cell* **179**, 736-749.e715 (2019).
96. Kopanos, C.*, et al.* VarSome: the human genomic variant search engine. *Bioinformatics* **35**, 1978-1980 (2019).
97. Fokkema, I.F.*, et al.* LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* **32**, 557-563 (2011).
98. Crockett, D.K.*, et al.* The Alport syndrome COL4A5 variant database. *Hum Mutat* **31**, E1652-1657 (2010).
99. Vithana, E.N.*, et al.* Genome-wide association analyses identify three new susceptibility loci for primary angle closure glaucoma. *Nat Genet* **44**, 1142-1146 (2012).
100. Purcell, S., Cherny, S.S. & Sham, P.C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149-150 (2003).

18. ACKNOWLEDGMENT

We thank Mr Timothy Chan from the School of Art, Design & Media, Nanyang Technological University, Singapore, for the design of the DRAGoN logo.

19. LIST OF APPENDICES