GENETICS OF POST-NEPHRECTOMY RESIDUAL RENAL CAPACITY IN LIVING KIDNEY DONORS

By
Esther Tanumihardja

GRADUATION REPORT

Submitted to
Hanze University of Applied Science Groningen

in partial fulfillment of the requirements
for the degree of

Fulltime Honours Master Sensor System Engineering

2016

ABSTRACT

GENETICS OF POST-NEPHRECTOMY RESIDUAL RENAL CAPACITY IN LIVING KIDNEY DONORS

By

Esther Tanumihardja

Human kidneys have been shown to have a residual capacity, a compensation mechanism the kidney undergoes when the other kidney is lost or has lost its function. This thesis is aimed to find genetic variants associated with this capacity. This was done by correlating pre- and post-donation kidney function with genome-wide genotype data in living kidney donors. This measure is interpreted as the donor's ability to adapt to single-kidney situation, a measure relevant for both the donor and recipient in kidney transplant situation. Significant association can increase knowledge of normal renal physiology and more specifically in the situation of unilateral nephrectomy.

Three loci were found to be significantly associated with the residual renal capacity. These loci are in the regions of gene *PAX2*, *PAX3*, and *HLA-DRB6*, which are logical functional candidates. Given the small sample size used in the study, independent studies should be conducted to replicate and validate these findings.

DECLARATION

I hereby certify that this report constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

I declare that the report describes original work that has not previously been presented for the award of any other degree of any institution.

Signed,

Esther Tanumihardja

ACKNOWLEDGEMENTS

First and foremost, I would like to express my greatest gratitude to Martin de Borst and Ilja Nolte for the opportunity of this research project, as well as for their valuable guidance, supervision, patience, and inspirations during the course of this project. My grateful thanks are extended to Ronald van Elburg for his feedback, guidance, and motivation. I am also very thankful for the help, supervision, and feedback provided by Peter van der Most during this project. I also wish to thank Harold Snieder, Mark Eijgelsheim, and Gerjan Navis for their expert input, ideas, and help in this project.

Contents

List of Figures

List of Tables

List of Definitions and Abbreviations

(Uni-)Nephrectomy	Surgical removal of the kidney, uni-nephrectomy refers to removal of one kidney
$\Delta GFR_{measured}$	Notation of residual renal capacity based on GFR measurements before and after nephrectomy
Allele	A variant form of a gene; usually used to express the pinned difference between different phenotypes/traits
BMI	Body mass index
bp	Basepairs; DNA unit of length, based on the number of nucleotide base-pairings
CD-CV	Common disease-common variants
Chromosome	Packaged and organized structure of the DNA of a living organism
CKD	Chronic kidney diseases
CVA	Cerebrovascular accidents
CVD	Cardiovascular diseases
EAF	Effect-allele frequency
eGFR	Estimated glomerular filtration rate
ERPF	Effective renal plasma flow; the volume of blood delivered to the kidneys per unit time
FF	Filtration fraction; defined as the ratio between GFR and ERPF
Gene	DNA that encodes a functional RNA or protein product
Genome	Organism's complete set of DNA, contains all of the information needed to build and maintain that organism
Genotype	An individual's collection of genes
GFR	Glomerular filtration rate; the total amount of glomerular filtrate formed in both kidneys per unit time
GWAS	Genome-wide association study
HWE	Hardy-Weinberg equilibrium
LD	Linkage disequilibrium
MAF	Minor allele frequency
MDRD	Modification of Diet in Renal Disease
Phenotype	An individual's observable traits, such as height, eye colour, and blood type
SNP	Single nucleotide polymorphism

Chapter 1
Rationale

Millions of people worldwide are afflicted with chronic diseases or congenital conditions, for which the only treatment is organ transplantation. Chronic kidney disease (CKD) is one of the most prevalent chronic diseases with increasing incidence and high morbidity [1]. It is defined as progressive loss in kidney function over a period of months or years, due to irreversible pathological changes. CKD has high prevalence of adverse outcomes, including total loss of kidney function, cardiovascular diseases (CVD), and premature death [2]. In the increasingly aging population, CKD acts as a significant barrier to healthy aging.

Hypertension and diabetes mellitus are the most common causes of CKD in the Western world; both underlying conditions can, for an important part, be attributed to lifestyle-related factors. However, given the disease's heritability estimated at 36-75% [3], a genetic component is also present. Over the years, single-gene studies have pinned some single-gene's attributed to kidney function/diseases [4]. Conversely, as genes may act in interaction with environment or other genes, such studies provided limited information for CKD or even for general kidney function [5]. Thus currently, there is still incomplete knowledge regarding which factors contribute to renal physiology and function.

The kidney is the most transplanted organ worldwide. Kidney and liver (though the latter far less common) transplantations are unique in that they can be donated from living donors. Evidently, living-donor kidney transplantation proved to have a higher success rate, as well as higher long-term (graft and patient) survival rate [6]. In addition, living-donor transplanted kidneys have lower risks of complications and better early function [7]. This is due to several factors. Firstly, living-donor transplantation surgery can be planned ahead of time, resulting in much shorter organ ischemia time. Secondly, the surgeon and donor can be placed in optimal conditions for the surgery. Consequently, the number of living kidney donations has been growing over the last decades. Currently, at UMCG, living kidney transplantations are performed equally often as post-mortal kidney transplantations. In 2015, 178 kidney transplantations were performed, 90 of which from living donors [8].

Living donors have to endure single-kidney situation after donation. Though it can be minimal, single-kidney situation increases the relative risk of developing end-stage renal disease [9]. Studies on the last 60 years of living (kidney) donations have proven its safety. Most of approved donors, having complied to a number of acceptance criteria, experience only minimal immediate and long-term risks and ultimately have an unaffected survival rate (compared to healthy population) [7]. A careful screening and post-donation monitoring of residual kidney function hold key to this, making them important values for the living donation practice.

Implications of donor's post-nephrectomy conditions can be predicted pre-donation [10]. Currently, potential living kidney donors are screened based on a number of clinical conditions. The most pivotal parameter is the donor's glomerular filtration rate (GFR) [11], used as the most reliable measure of overall kidney function. Donor age and BMI have also been identified as independent determinants of long-term post-donation GFR [12].

It has been shown that after nephrectomy, the donor's remaining kidney undergoes hyperfiltration as an adaptive reaction [13], to compensate the function of the donated kidney. This adaptive capacity is termed as *kidney's residual/reserve capacity*. Some centres also factor this adaptive capacity to the donor long-term post-nephrectomy prognosis. At the Groningen Transplant Center (GTC), the donor kidney residual capacity is probed experimentally

[14]. Maximum kidney function is simulated by stimulation with low-dose dopamine. GFR levels under regular conditions and GFR following dopamine stimulation are evaluated in precise manner by means of iothalamate clearance [14]. The differential GFR level can be used to predict donor's residual kidney function. This experimental method has the disadvantages of a prolonged and laborious (the procedure takes up to 8 hours), and relatively costly procedure. In addition to their GFR level, potential donors are also screened based on their other clinical conditions, for example the presence of diabetes mellitus, uncontrollable hypertension, etc.

Based on the presence of a genetic component of kidney functions [3], it would be preferred to use pertinent genetic variants as predictors of donor's post-nephrectomy residual kidney function. Such genotypic predictors would be relevant to both donor and recipient. Genes could potentially contribute greatly to the prediction of residual kidney function, alongside of donor's age, BMI, or perhaps in spite of other clinical conditions. E.g. despite of sufficient measured GFR level, an individual might have genotypes that increase his/her risk of developing chronic renal failure, making him an unsuitable donor candidate. Or the other way around, despite insufficient/borderline GFR level or hypertension or diabetes mellitus, an individual might have genotypes for excellent/sufficient renal residual capacity, allowing him to adapt well to unilateral renal situation. Hence, significant genetic associations will increase the current knowledge of renal physiology in the situation of elective unilateral nephrectomy.

In order to achieve the stated objectives, the project is focused in answering the following research question.

> *Can genetic variants be identified that are associated with reported residual renal capacity in single-kidney situation in living kidney donors?*

Sub-research questions addressed in this project include:

- *Which characteristics can be associated with the post-nephrectomy residual renal capacity?*
- *Can known associations with eGFR be replicated in our cohort of living kidney donors?*
- *How much variance in post-nephrectomy residual renal capacity can be explained by genes?*

Chapter 2
Background

Overview

This chapter summarizes the background information regarding the nature of human genetics; especially its functions and relations to the human traits/diseases.

Genes and the Central Dogma

Gregor Mendel was the first who postulated the notion of inheritance back in the mid-nineteenth century. Through experiments, Mendel showed that organism's traits were results of a heritable biological material. Mendel also revealed that these traits are often unrelated, thus can be inherited independently (summed up in Mendel's Laws of Segregation and Independent Assortment), by independent inheritability units which he named *genes*. His experiments also concluded that these genes are passed on with patterned randomness to successive generations. [15]

In 1902, the physical molecular location of this inheritance factor was narrowed down to a cell's chromosome, as specified in the Boveri and Sutton chromosome theory [16], [17]. It was not after 20 years later, however, that the Griffith's experiment confirmed the deoxyribonucleic acid (DNA) as the exact molecule in chromosomes where this inheritance information was stored [18]. In 1958, Watson and Crick modelled the structure of DNA in detail as a double-stranded helix, which is still accepted today [19].

DNA is built of monomers called *nucleotides* [19]. Nucleotides are composed of a common phosphate and sugar group, with a differentiating nitrogenous base, as illustrated in Figure 2.1. The bases are adenine (abbreviated A), guanine (G), cytosine (C), and thymine (T). The different bases form a sequence which codes for the inheritance information. They hold functions in encoding, transmitting, and expressing genetic information responsible for the development and functioning of living organisms. These functions come together in a flow named as the *central dogma of molecular biology* [20], as charted in Figure 2.1Figure 2.1. (A) Structure of nucleotide and its base-pairing *[22]* (B) Illustration of central dogma. DNA segments are transcribed into strands of messenger RNA (a similarly structured nucleic acids with one base differing from DNA, i.e. T is converted to uracil (U)). The messenger RNA (mRNA) strands are then translated into protein in ribosome. [19]–[21] Proteins are the chief functional units within cells (as enzymes, for signalling/binding, and structures), making up the individual's characteristics/phenotypes. Therefore, an organism's DNA can be regarded as the blueprint of their biological makeup.

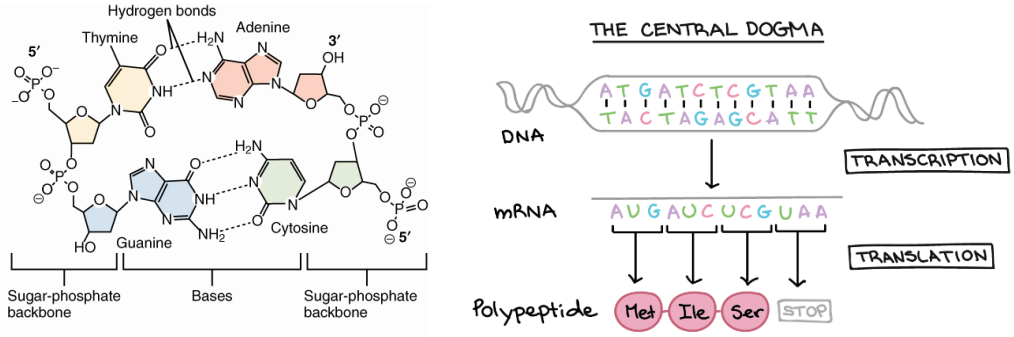

Figure 2.1. (A) Structure of nucleotide and its base-pairing [22] *(B) Illustration of central dogma of molecular biology showing how DNA codes for mRNA then protein* [23]

The Human Genome and its Variation

Genome is the term used to refer to the complete set of DNA sequence in an organism. The human genome consists of 3.3 x 10^9 nucleotides in its length. The complete genome is currently approximated to contain around 25,000 genes, which encompass a mere 1.5% of the genome (these are then called the coding sequence). The rest, however, has more elusive functions (if any at all). Even within genes there are different classes of noncoding DNA currently identified: regions that code for non-protein-coding RNAs (e.g. tRNA, rRNA, miRNA), pseudogenes, introns, untranslated regions of mRNA, regulatory regions, repetitive regions, and transposable elements. Though they do not directly code for protein, a big part of these regions are involved in one way or another in the expression of the protein-coding gene; either pre-, during, or post-transcription [24], [25]. This makes these areas also interesting to study.

Humans have two homologous strands of DNA, making human a *diploid* organism. During cell division, the DNA strands pack themselves around a protein into 23 pairs of thread-like structures, named the *chromosomes*. The pairs of chromosome are *homologous* to each other, meaning that they carry the same genes in the same order, making each location (or *locus*) comparable between two chromosomes of the same pair. However, the bases of same locus can differ, since one chromosome is inherited from the paternal parent and the other maternal. [25] These differences/variant forms are termed as *alleles*. Often, one type of allele is more present in a given population, making it the *major allele*, and the alternative the *minor allele*.

These allele variations can come from multiple forces. Most likely, it involves errors during DNA replication, DNA recombination, the insertion of viral genome components, and/or the insertion and excision of mobile transposon elements. The variation within a population can point to the divergence of the population from more ancestral populations. [26]

There are different forms of variations observed across the genome, namely the single nucleotide polymorphisms (SNPs), indels (insertions and deletions), duplications, copy number variants, and inversions. The most common form of genome variation is SNP, which is a variation at a single base position. Common SNPs, for which the minor allele has a frequency of at least 5%, occur approximately once per 1,000 base pairs, amounting to 88% of the variations currently known in the human genome [27]. SNPs with low frequent or rare alleles occur at an even higher rate and are often population-specific. SNP variations can occur both in coding as well as in non-coding regions. Either variants have been previously associated with traits and/or diseases, suggesting that the regulation of genome is much more complex than previously believed. The cumulative effect of these variations most likely plays a significant role in health and disease.

Mendelian versus Complex Diseases

Human traits can either be monogenic or polygenic. Monogenic traits are those influenced by a single gene. They are often dichotomous (or qualitative) and follow the Mendelian inheritance within families. Genes involved in Mendelian diseases can well be studied by linking the inheritance of genetic variants with presence of the disease among the family members and these so-called *linkage studies* have contributed greatly to the identification of Mendelian disease genes with large effects [28].

However, many human traits (including kidney function [29]) are complex. They are common in the population and believed to be caused by many genes and their interaction with each other or with environment. Because of the complex nature of common diseases, they are often studied through underlying continuous or quantitative characteristics, so-called *endophenotypes.* Endophenotypes are observable characteristics that can be robustly and reliably measured, are thought to be strongly genetic in origin, and are thought to underlie and contribute to the common disease vulnerability but are not to be part of the disorder itself. An example of an endophenotype is eGFR underlying CKD. Only in the early twentieth century quantitative traits were shown to be still bound to Mendel's laws

of inheritance even though such traits were influenced by multiple loci and polygenic in nature [30]. Nevertheless the study of complex diseases or traits and even of the endophenotypes within families has been proven to be difficult for a number of reasons: (1) for most common traits/diseases, many genes are involved, (2) the genes also interact with each other and with environmental factors, (3) a gene variant is also common implying that it might be introduced into a family multiple times, and that (4) the presence of gene variant brings low risk of the manifestation of disease or trait. This all implies that the inheritance of the trait/disease does not follow a straightforward Mendelian inheritance model. [30] As a result researchers turned to *genetic association studies* in which genetic variants are investigated in large numbers of unrelated individuals and their frequencies correlated with the trait, disease, or endophenotype.

Chapter 3
Situational and Theoretical Analysis

Overview

This chapter reviews the physiology of the kidney and introduces the trait investigated in this project. Relevant previous (genetic) studies in the matter were also reviewed in relation to the project's aim, ultimately leading to the hypothesis.

Kidney: its Functions and Measures

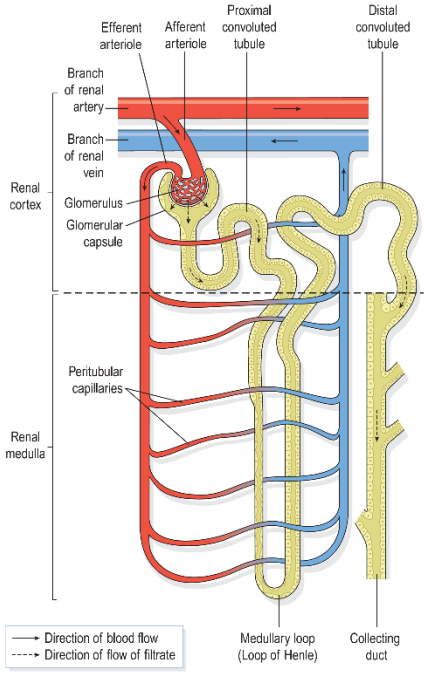

Figure 3.1. Anatomy of a nephron: the smallest functional unit of the kidney and associated blood vessels [31]

The kidney is an organ found in vertebrates, which holds vital roles in the organism's regulatory system. Its primary functions are (1) to regulate volume and composition of extra-cellular fluid in order to maintain homeostasis (e.g. concentration of electrolyte, osmolality, acidity), (2) to excrete metabolic waste products (e.g. urea, uric acid, creatinine), end products of haemoglobin degradation, and foreign chemicals (e.g. drugs, food additives, etc.), and (3) to produce a number of circulating factors (e.g. erythropoietin for red blood cell production, renin for regulation of blood pressure, etc.). These functions are achieved by a number of consecutive processes taking place at microscopic level. The smallest functional unit of the kidney is termed as nephron, illustrated in figure 3.1. Each human kidney contains on average 1.3 million nephrons. With each nephron filtering a small amount of blood, the two human kidneys collectively receive up to 1200 mL of blood per minute. This amounts up to 25% of the cardiac output. [31], [32]

The kidney homeostasis involved different processes/mechanisms – namely filtration, reabsorption, and secretion – taking place at different sites of the nephron. Filtration occurs through the semipermeable walls of the glomerulus and glomerular capsule (also known as Bowman's capsule). The afferent arteriole supplies the glomerulus with blood. The glomerular capillary wall is charged and fenestrated with pores, characteristics that determine their permeability and selectivity. From the glomerular capillaries, unfiltered blood flows to the efferent arteriole, whose diameter is smaller than the afferent arteriole. This increases the pressure inside the glomerular capillaries, which drives filtration across the fenestrated walls. [31], [32]

Glomerular capillaries are invaginated into the dilated, blind end of the nephron, called the *glomerular capsule*. Substances that pass through the glomerular capillaries as well as the specialized epithelium of the capsule (which also has specific permeability and selectivity) are called *glomerular filtrate.* The makeup of glomerular filtrate is the same as blood-makeup, excluding blood cells and proteins. The filtrate then continues into the lumen of the tubule. Throughout the tubules (mostly in the proximal tubule), useful substances (i.e. sodium, glucose, and some water) are reabsorbed and added back to the capillaries in a regulated manner. Further, in the distal tubule, many organic acids

or bases (e.g. end products of metabolism or foreign substances) are avidly secreted by the tubules, forming urine. Urine is then transported into the renal pelvis and then ureter as it leaves the kidney into the bladder, where it is collected in between micturition. [31], [32]

As mentioned, kidney function is most often expressed as the glomerular filtration rate (GFR). It is defined as the total amount of glomerular filtrate formed in both kidneys per minute. Such rate is directly modulated by a number of factors, namely the hydraulic conductivity of glomeruli, glomerular surface area available for filtration, and the net filtration pressure (NFP). GFR can be modelled as the product of these factors, as shown in the equations below.

$$K_f = glomerular\ surface\ area\ \times hydraulic\ conductivity \quad \textit{(eq. 1)}$$

$$NFP = (P_{GC} - P_{BS}) - (\Pi_{GC} - \Pi_{BS}) \quad \textit{(eq. 2)}$$

$$GFR = K_f\ \times NFP \quad \textit{(eq. 3)}$$

Filtration holds a significant (if not, the most significant) and determinant role in kidney function. Moreover, most nephropathies (i.e. diabetic nephropathy, hypertensive nephropathy, etc.) primarily affect the filtration site and process. As the GFR captures this function as a whole, it is an accurate parameter of kidney function both in health and disease.

As physical measurements of glomerular surface area and hydraulic conductivity (often put together as a K_f or filtration coefficient, shown in equation 1) are practically unattainable, GFR can only be inferred experimentally through another measure, namely the renal clearance rate. Generally, clearance rate can be defined as the amount of plasma that is cleared of the specific solute per unit time. For solute that is freely filtered and is neither reabsorbed nor secreted by the tubules, its clearance rate amounts rather accurately to the glomerular filtrate produced per unit time. Such substrate is inulin or the radioactive substance iothalamate, which are not normally found in humans. Thus after injecting inulin/iothalamate into the blood, the rate of its excretion can be calculated and be used to directly approximate the GFR.

In collected urine, the concentration of certain solute (e.g. inulin/iothalamate) can be quantified. Relating to the urine flow, the mass of solute can be calculated. This mass represents the same amount of solute filtered at the glomerulus, as nothing is added or absorbed in the tubules. Dividing this mass by the solute concentration in plasma (known concentration from injection/infusion) calculates back the volume of plasma filtered at the glomeruli (also given in eq. 4).

$$GFR\ (mL/min) = \frac{solute\ concentration\ in\ urine\ (mL/min) \times urine\ flow\ (mg/mL)}{solute\ concentration\ in\ plasma\ (mg/mL)} \quad \textit{(eq. 4)}$$

Another solute used as renal function biomarker is creatinine. It is a solute that occurs naturally in the human's body as a breakdown product of phosphocreatine in the muscle. This rate of breakdown is usually rather constant within some time frame, as a function of muscle mass. Therefore concentration of creatinine in the plasma can be inferred with some confidence according to one's BMI, allowing calculation of the (endogenous) creatinine clearance. However, this rate is sensitive to a number of factors, e.g. hyperglycaemia, protein urea, some medication, and even the time of day (it is highest in afternoon) [33]. Moreover, it tends to overestimates GFR since a small amount of creatinine is also being secreted by the tubules during urine formation.

An even less direct biomarker, the serum creatinine, has also been used as practical solution for assessment of kidney diseases. As said, creatinine is found circulating in blood plasma, to be filtered out by the kidneys. Should renal function be impaired, creatinine excretion is reduced, leading to an elevated level of creatinine in serum. Based on this mechanism, different formulas have been proposed to (rather roughly) estimate GFR value (often termed *eGFR*)

based on serum creatinine level, corrected for different age, gender, and race groups. A widely used formula (the Modification of Diet in Renal Disease, abbreviated MDRD formula) is shown in eq. 5.

$$eGFR\ (mL/min/1.73\ m^2) = 32788 \times {S_{cr}}^{-1.154}\ (\mu mol/L) \times (age)^{-0.203} \times (0.742\ if\ female) \times (1.212\ if\ African\ American) \quad \textit{(eq. 5)}$$

Although very commonly used, MDRD eGFR formula has also been reported to underestimate GFR value in healthy adult individuals (i.e. individuals with eGFR > 60 mL/min/1.73 m^2) [34].

Residual Renal Capacity

As mentioned, in this project we are interested in post-nephrectomy reserve/residual renal capacity. The presence of residual renal capacity has been known and studied for decades [35], [36]. Several human (and animal-model) studies showed that uni-nephrectomy causes the remaining nephrons to go into hyperfiltration to compensate for the lost kidney [37], [36], [38]. Animal model studies demonstrated that this compensatory process can be harmful to rat's remaining kidney [39], but this effect was not observed in humans, as the living kidney donors' long-term renal function and survival rate did not differ from that in non-donating individuals [6], [40]–[42].

Despite its confirmed existence, there currently is still no full consensus on the purpose or mechanism of the residual renal capacity. The compensatory hyperfiltration has been observed to occur immediately after uni-nephrectomy [39]. This suggests that such reaction is an adaptive mechanism so the (resting) kidney could adapt to stress [43]. Among hyperfiltrating diabetes patients, however, it was reported that this capacity was diminished, as well as in patients with extreme hyperfiltration [44].

Experimentally, the (resting) kidneys can be stimulated to undergo hyperfiltration in a number of ways. Many reports showed that GFR increases after an animal-protein-rich meal, as well as by infusion of amino acids [45], [35]. Infusion of (low-dose) dopamine has also been proven to stimulate the kidneys, and thus increases the GFR values [46]. These are currently the commonly used clinical protocols in probing into residual renal capacity, by comparing resting GFR with stimulated GFR (by either, or both, stimulant(s)). Dopamine infusion tests whether there is an increase in renal perfusion, and amino acids infusion tests for an increase of glomerular capillary pressure. A study [36] showed that in following uni-nephrectomy, dopamine-stimulated residual renal capacity was decrease, whereas amino acids-stimulated residual renal capacity remained unchanged. This suggests that increased blood flow may underlie mechanism of the adaptive post-nephrectomy hyperfiltration.

Other studies observed the phenomenon of compensatory hyperfiltration by looking into the change in kidney volume after uni-nephrectomy, revealing an increased kidney volume after some time [13], [37], [47], [48]. A study [47] compared the adaptive response in acquired and congenital single-kidney situation. Long term observations pointed out that individuals in an acquired single-kidney situation are more susceptible to hypertension and renal impairment than individuals in a congenital single-kidney situation, hinting at different underlying adaptation mechanisms. Congenital single-kidney individuals achieved hyperfiltration through kidney hyperplasia, or proliferation of kidney cells, during nephrogenesis, which results in a higher number of nephrons (with normal anatomy). Acquired single-kidney individuals, however, cannot undergo such process due to the low mitotic activity of mature nephrons [37]. Instead, hyperfiltration adaptation in such case is shown to be achieved by hypertrophy, or enlarged cell size, as a result of elongation and increased convolution of the tubuli.

In conclusion, these studies showed the two mechanisms underlying adaptive hyperfiltration in living kidney donors: hyperperfusion and hypertrophy of remaining nephrons [13].

Genetic Association Studies

Genetic association studies strive to map genetic variants that are associated with a disease or trait, following the idea that individuals with common trait share common, specific genetic variants, i.e. the "common disease, common variant"-model (CD-CV model).

The introduction of DNA sequencing method [49] has allowed an exciting boost of genetic association studies. The development kicked off the world's biggest biomedical race, to sequence the entire human genome (also commonly known as the Human Genome Project) [50], [51]. The technological development extended beyond this race. This permitted for genome data to be generated at unprecedented rate and resolution. With the amount of genomic data available, a project (the HapMap project) could even map the common patterns of human genetic variation [52], [53]. Such project discovered that only ~0.1% of the entire human genome is polymorphic (minor allele frequency (MAF) > 5%). Furthermore, it was found that some subsets of (nearby) SNPs in the same chromosome are inherited together (the phenomenon is commonly termed as *linkage disequilibrium*). A series of alleles at correlated SNPs on the same chromosome is called *haploid genotype*, or *haplotype* in short. Haplotypes (within defined population) reduce genetic variability, and thus allowed for further cost reduction in genotyping. By only genotyping a fraction of carefully-selected 'tag' SNPs, an individual's full genome sequence can be imputed based on their haplotype blocks. Samples of haplotypes of the same ethnicity are freely available for research [54]. Affordable high-throughput genome-wide genotypic arrays have been developed in the last 5 years. The array usually contains 300,000-1,000,000 tagSNPs, which have been well-selected to provide reasonable coverage of the human genome.

This availability of genome-wide genotyping also shifted the practice of genetic association studies. Previously, association studies were performed between traits/phenotypes and 'candidate genes', which were selected based on the understanding of the pathogenesis. The study then tests whether at SNPs in that gene the allele and a trait co-occur above chance level. The hypothesis-drive gene candidacy limited such studies to the known pathways involved in the trait and prohibited the discovery of new ones. Over the years, it was found that the outcomes/findings of such studies have low chance of replication in subsequent independent studies. [3], [30]

Genome-wide genotyping opened up the way for genome-wide association studies (GWAS), GWAS offered a hypothesis-free approach, by interrogating every gene/allele for the genetic association. Many GWASs for many diseases and traits have been undertaken and many new genes and pathways have been identified [55]. While it is a promising, powerful tool in probing into complex traits/diseases, there are some limitations and precautions to be taken into account. Each interrogation of allele/SNP against the trait serves a single statistical test. Therefore, the downside of a GWAS is that this requires stringent statistical significance criteria as a result from multiple testing correction in order to avoid false positives. For the same reason, a large sample size is required to achieve adequate statistical power. Quality control, population, and study setup should also be done properly. Mixing populations/analysing a poorly stratified population can lead to false-positive association due to the differences in the frequencies of common variants across subpopulations, termed *population stratification*. Lastly, the study might point to intergenic SNPs/loci. Often times, such finding is then annotated to the gene closest to the loci. It is noteworthy that this annotation does not imply a mechanistic/causal relationship between gene and phenotype, as it merely points to the genomic region of the variant. Further investigation would then be required to map the causal variant more accurately, as well as to study the functional pathway that can lead to definitive identification [3], [30].

Genetic Factors on Kidney Function

In order to increase power of the finding a gene that is associated with residual renal capacity the candidate gene approach is also followed. In the past five years, a lot of effort has also been put in GWASs of kidney function [56]–[60]. The CKDGen consortium has so far successfully identified up to 53 loci associated with kidney function. Nonetheless, these loci could only explain 2% of the variance of human kidney function (out of the estimated heritability of 36-75%) [3], suggesting that there are still many genetic variants to be uncovered.

These studies all used single eGFR or serum creatinine measures as the outcome of interest. In the setting of chronic kidney disease, the kidney function declines over time. Therefore observation in time would reveal novel loci pointing to the phenomenon. Only recently the first study was reported on genome-wide association with kidney function over time [61]. The *UMOD* gene, encoding the Tamm-Horsfall protein, was significantly associated with kidney function decline. Furthermore, there has not been a genetic study reported that looked specifically into residual renal capacity.

So far, most variants discovered by GWAS for complex traits/diseases bring about small effects on its own, leaving a lot of its heritability (portion of phenotypic variance that is attributable to genetic factors) unexplained. As mentioned, complex traits are influenced by an interplay of different genes. In such case, multiple genetic variants can be investigated by an aggregated SNPs risk score. The effect of multiple SNPs are studied together in an association study, by means of a *genetic risk scores* parameter. [62]–[64]

Hypothesis

From the given information, it could be concluded that kidney function has genetic underpinnings, most of which is still unexplained. An interesting phenotype is the kidney's residual capacity. Even currently, its physiology has not been fully explained. Its genetic determinants have also not yet been studied. Studying the underlying genetic determinant(s) of the residual renal capacity can potentially offer new insights into the renal physiology, as well as clinically relevant information in the transplantation setting (both for potential donors and recipients). Therefore the hypothesis of this project is that one or more genes that will be studied, are statistically significantly associated with residual renal capacity. GWAS offers the most promising way to study such complex trait. With a well-defined phenotypic trait, this method can efficiently probe into the genetic determinant of the human residual renal capacity. However as such a study might be limited by the sample size additionally only the 53 genes that were previously identified by CKDGen, were studied for association with residual renal capacity.

Chapter 4
Conceptual Model

Overview

So far, the motivation and aim of the project have been made clear. Relevant advances in the field have also been presented, along with the hypothesis statement. This chapter explores in more detail the most important factors of the study. Choices for each factors are discussed separately, leading to the overall study plan.

Study Population

Choosing a suitable study population is important for association studies. Use of a non-random (e.g. presence of subpopulation with different ancestry) group may yield false-positive results, where associations point to some underlying structure of the population instead of the trait in question. Ideally, a large homogeneous population should be used. Groups with random mating behaviour, little migration, and no mutations are best to yield least random fluctuations in allele frequencies. [65]

Due to the (unique) phenotype of the study, the potential population is limited to living kidney donors with kidney function data observed over time. In this project 280 individuals were enrolled that visited the University Medical Center Groningen (UMCG), the Netherlands, for a uni-nephrectomy during the period of 1993 until 2007. Their hospital records were reviewed and data on renal function (GFR, ERPF, FF, GFRdopa, ERPFdopa, FFdopa, serum creatinine) as well as on possible confounders (age at the time of surgery, sex, hypertension, (micro)albuminuria, myocardial infarction or revascularization, cerebrovascular accident (CVA), smoking status, blood pressure (systolic, diastolic), sodium intake, BMI, and cholesterol (total, LDL, HDL, triglycerides)) were extracted. SPSS was used for data collection.

To achieve a suitable population, only kidney donors that fulfil the following criteria are included in the study.

- Kidney donors with phenotypic data on renal function (i.e. (e)GFR, EPRF, FF both at base level and dopamine-stimulated) for:
    - 3-4 months pre-donation
    - 3-4 months post-donation
- Free of diabetes mellitus
  It is stated for the data available (since diabetic individuals are not eligible kidney donors), as well as because nephropathy is very common in diabetic individuals, thus altering their kidney function. Inclusion of diabetic individual's data might introduce a subpopulation bias, leading the association to lead to the different (diabetic nephropathy-specific) function/pathway.
- Free of hypertension or well-controlled hypertension before donation (<140/90 mmHg with ≤2 anti-hypertensive drugs)
  For the same reason as the previous criterion, unmanageable hypertensive individuals are not eligible for kidney donors, and the potential bias of hypertensive nephropathy pathway.
- eGFR > 60 mL/min/1.73 m^2 before donation
  It is relevant in the case that eGFR values are used as the phenotypic definition in the association study, due to the non-linearity of the MDRD GFR estimation formula.

Phenotypic Model

Residual renal capacity, as explained, is a complex quantitative trait whose expression is not yet clearly/consensually described thus far. From the available data of the study population, a number of options for residual renal capacity expression were considered. Generally, iothalamate measurement of GFR is preferred over eGFR, for its higher independency and precision. Creatinine-based eGFR has also been reported to be affected in hyperfiltration (negative correlation between error of eGFR and the FF) [66]. In clinical setting, living donors' (residual) renal capacity are estimated by infusion of dopamine around 3-4 months before nephrectomy, these data are available for most of the living donors in the population. As well, a follow-up iothalamate measurement (mostly only during rest state/unstimulated) is usually done around 3-4 months after nephrectomy. Based on the available data, the expressions of residual renal capacity given in Table 4.1 were considered.

Measured residual renal capacity:	
Absolute change in GFR	→ $GFR_{after} - (GFR_{before}/2)$
Relative change in GFR	→ $GFR_{after}/(GFR_{before}/2)$
Estimated residual renal capacity (dopamine):	
Absolute change in GFR pre-donation	→ $(GFR_{dopa}\text{-}GFR)/2$
Relative change in GFR pre-donation	→ $(GFR_{dopa})/2GFR$

Table 4.1. Different formulated expressions of residual renal capacity.

Chapter 3 also looked into the anatomical/histological adaptation in single-kidney situation. It was revealed that the number of glomeruli/nephrons do not change in single-kidney situations with mature glomeruli. Instead, the remaining nephrons undergo hypertrophy which increases its function. Therefore the elevated kidney function should be a function of/depend on the number of remaining nephrons. This can then be best captured in an expression relative to the original/pre-nephrectomy renal capacity.

As can be seen in Table 4.1, it is assumed that nephrectomy should leave donors with half of their original renal mass (the division by two in GFR_{before} term). Though this is a practical approach in such situation, it might not be the most accurate assumption. Individuals often have different kidney mass between their left and right kidneys. Should it be anatomically feasible (uncomplicated by vascular system, etc.), clinicians generally suggest/opt for nephrectomy of the smaller kidney to minimize the risk for the donor. This might create a potential bias in our expression of residual renal capacity, since nephrectomy of the smaller kidneys may exaggerate residual kidney capacity. Thus, normalizing the measured residual renal capacity to its pre-nephrectomy function should also give the upper hand in minimizing this potential bias. In conclusion, measured residual renal capacity in relative expression was chosen as the most suitable primary outcome. This is also noted as $\Delta GFR_{measured}$ in the rest of the document.

Genetic Model

In designing genetics study of a complex trait/disease/endophenotype, the frequency of an allele within the population and the corresponding risk incurred by the allele are important considerations. Following the CD-CV model, a common variant should have small effect size relative to that found for rare disorders. Over the past five years, the CD-CV model has been tested for many common diseases/traits. The abundant results (as catalogued in [55]) has proven the applicability of the model. Spectrum given in Figure 4.1 maps the effect size and allele frequency space of disease/trait. Based on the model, the association study is aimed to discover effects in the right lower space. Although not the entire genetic component of any common trait can be attributed to common alleles only, this area of the spectrum is currently very promising to explore with the currently available genotyping technology/data.

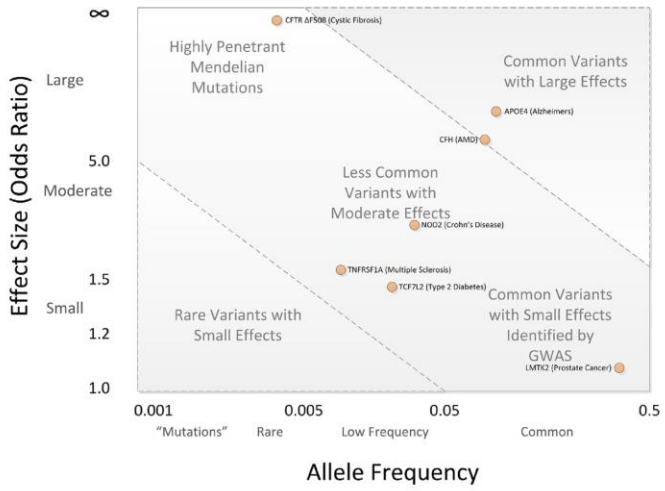

Figure 4.1. Spectrum of endophenotype allele effects. Disease associations can be conceptualized in dimensions of allele frequency and effect size. The discovered genetic associations lie within the dashed lines. [67]

Different models can be assumed about the genetic effect of each allele to the endophenotype, the dominant, recessive, multiplicative, and additive models are explored below. For illustration, consider two alleles for a loci, A and a. The dominant model assumes that having one or more copies of the risk allele (A) increases risk compared to a (i.e. individuals with AA or Aa have increased risk). The recessive model assumes that two copies of the risk allele (A) are required to alter the risk. Thus individuals with AA genotypes should be compared to individuals with Aa and aa genotypes. The multiplicative model assumes that if one copy of risk allele (A) brings k risk, there is a k^2 risk for having two copies of the A allele. The additive model assumes each copy of risk allele (A) increases the risk linearly. So if Aa genotype has k risk, there is a 2k risk for AA.

Commonly, GWAS examines additive models only, it has reasonable power to detect both additive and dominant effects [68]. This additive model, however, may be underpowered to detect some recessive effects. Previous studies have assumed also additive models for other kidney function GWAS [60], [61]. Thus for this study, similarly, an additive model was also used. GWAS examines each SNP independently to the phenotype. Association is done simply through ANOVA (similar to linear regression with categorical predictor variable, the allele). Several assumptions of such models are (1) the (quantitative) trait is normally distributed, (2) the groups are homogenous, and (3) the groups are independent.

Chapter 5
Research Approach

Overview

In answering the research questions, a series of steps were done in the study. This chapter recaps and details these steps/processes separately.

Phenotypic data

Table 5.1 summarizes the data (and their expressions) relevant to the outcome measures and covariates. Clinical data were extracted from UMCG Poliplus patient database system.

Data before nephrectomy		**Data after nephrectomy**	
(Type of data)	*(Values/units)*	*(Type of data)*	*(Values/units)*
Age at nephrectomy	years	Date of GFR measurement	DD.MM.YYYY
Sex (%F)	(F/M)	GFR, ERPF, FF	mL/min, mL/min, -
Race	(Caucasian/Other)	GFRdopa, ERPFdopa, FFdopa	mL/min, mL/min, -
Date of GFR measurement	DD.MM.YYYY	Serum creatinine (24h-urinary creatinine excretion)	mL/min
Date of nephrectomy	DD.MM.YYYY	Medication use	-
GFR, ERPF, FF	mL/min, mL/min, -	Hypertension	(yes/no)
GFRdopa, ERPFdopa, FFdopa	mL/min, mL/min, -	Use of anti-hypertensive(s)	(yes/no)
Serum creatinine (24h-urinary creatinine excretion)	mg/mL	Use of ACE inhibitor	(yes/no)
Medication use	-	(micro)Albuminuria (24h-urinary albumin excretion)	mg/mL
Hypertension	(yes/no)	Blood pressure: systolic, diastolic	mmHg
Use of anti-hypertensive(s)	(yes/no)	Sodium intake (24h-urinary Na excretion)	mmol/24hrs
Use of ACE inhibitor	(yes/no)	BMI: height and body weight	kg/m^2
(micro)Albuminuria (24h-urinary albumin excretion)	mg/mL	Cholesterol: total, LDL, HDL, triglycerides	mmol/L
Myocardial infarction/revascularization	(yes/no)		
Cerebrovascular accident (CVA)	(yes/no)		
Smoking habit	(never/former/ current)		
Blood pressure: systolic, diastolic	mmHg		
Sodium intake (24h-urinary Na excretion)	mmol/24hrs		
BMI: height and body weight	kg/m^2		
Cholesterol: total, LDL, HDL, triglycerides	mmol/L		

Table 5.1. Summary of collected data and their expressions.

Data Cleaning

The gathered phenotypic data were first summarized and visualized in order to understand the population better. Most important characteristics data were plotted in box plots, revealing outlying points. Observed outlying entries were then checked back to the database/patient correspondence file. Suspicious entries were checked manually and filtered out. Analyses were carried out under the assumption that data were normally distributed. In order to check this, data were also plotted in histograms. Logarithmic transformation (base 10) was then applied to skewed distributions to achieve the desired distribution.

Genome-wide data

All individuals in this study participated in the REGaTTA (REnal GeneTics TrAnsplantation) cohort (UMCG, The Netherlands) of which a part was genome-wide genotyped using a customized genome-wide genotyping array the 'TxArray' [69]. Thorough quality control of these data were done elsewhere (as part of the iGeneTrain consortium [70]) and clean data were provided. Among other things non-Caucasian individuals were removed to avoid association coming from underlying population substructure. Genotyped data were next imputed based on the most relevant data available to the population: 1000 Genomes (1000G) [71] and genomes of the Netherlands (GoNL) [72] projects. Resulting genome-wide data were provided in assembly build of *Genome Reference Consortium GRCh37*. Finally, a total of 151 individuals were available with (some expression of) endophenotype and genotype data for this project.

Power Analysis

Power analysis is often practiced in designing a genetic-association study, typically in determining the required sample size. However, in this study (and its time frame), the available population size is rather fixed. Therefore power analysis was done instead to see beforehand what kind of effect size could be detected with the available population.

The power of a genetic association study is the probability of detecting an effect of a particular size. This power depends of a number of factors, i.e. the effect size itself, sample size, frequency of the effect allele, and the required level of statistical significance (type I error rate). The relationships of these factors is illustrated in Figure 5.1 below. As can be seen, to achieve the same significance level and power, the more common the (effect) allele would require smaller sample size. As well, the higher the effect size would also require smaller sample size [73]. Of course, these factors (effect allele frequency and effect size) would be unknown prior to the experiment. Thus such analysis would have to be run under some assumptions.

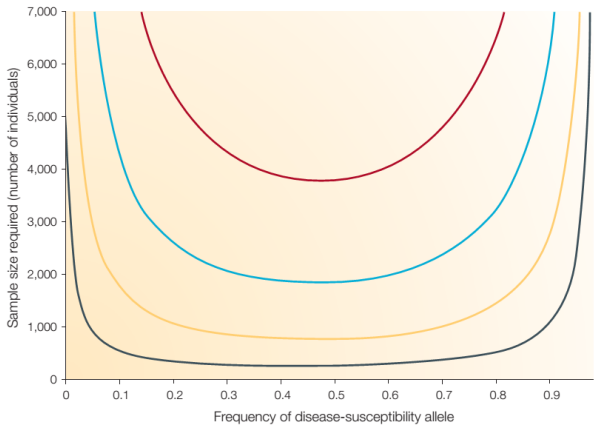

Figure 5.1. Effects of allele frequency on sample-size requirements. The sample size required in an association study to detect variants with odds ratio of 1.2 (red), 1.3 (blue), 1.5 (yellow), and 2 (black) are shown. Numbers shown are for a statistical power of 80% at a significance level of p-value < 10^{-6}. [73]

The traditional computational approach to estimate such power was to simulate hundreds of replicate samples under different settings of the factors. The fraction of simulations could successfully detect an effect provides an estimate of the power. Lately, closed-form power expressions have been generally accepted as a more effective and rapid approach to estimate the power of a study. Such closed-form analysis was done for this study using the program QUANTO [74]. Power analysis was done for the GWAS, looking into gene-only influence modelled in additive mode of inheritance, for the desired power of 80% (generally used as 'acceptable' power level), and allele frequency of 2%. Allele frequency of 2% would mean 2-3 individuals in the population would have this allele, which is a sound assumption for minimum allele frequency based on consultancy with the expert.

Genome-wide Association Study

There were two different methods of interest to analyse the outcome, the hypothesis-free approach of GWAS or the more hypothesis-driven candidate gene or SNP study. While GWAS seemed to be the more promising and unbiased approach, the available population size might limit its success. Nonetheless, we firstly attempted a GWAS analysis.

Each SNP was interrogated against the formulated residual renal capacity measure, corrected for the potential covariates and therefore about $10x10^6$ tests will be performed. These tests were done using the frequentist association test, part of the SNPTEST v2.5.2 [75]. Computation of this test was performed on the Lisa Cluster, Amsterdam.

As a means of quality control, the population was tested for *Hardy-Weinberg equilibrium* (HWE). The HWE models for populations in which the allele and genotype frequencies will remain constant from generation to generation in the absence of other evolutionary influences. Such equilibrium can be tested by determining behaviour of two alleles at a single genetic locus. Testing fitness each locus of the population to the HWE model can assess underlying population structure and reveal genotyping quality. [65], [76] HWE testing was also done to each SNP by the SNPTEST v2.5.2 program.

GWAS results were filtered, SNPs with HWE p-value < 0.001, or MAF < 2% were excluded. This process was done in R. Further, the results were processed with QCGWAS R-package v1.0-8 [77], where the result quality was further checked/cleaned. The program inspected for missing and invalid data, alleles were harmonized across results (by matching it to reference, i.e. the HapMap, and removing unmatched alleles), and quality control plots were generated. Lastly, biological information on specific loci (especially for significant association results) were taken from Ensembl (release 84) [78] and GTExPortal [79]. To control for false-positive results only SNPs with a *p*-value < $5x10^{-8}$ were used as the genome-wide significance threshold.

Candidate SNPs Association Study

As mentioned before, a GWAS analysis on kidney function has previously been done in a large cohort with 133,413 individuals revealing 53 associated loci significantly associated to eGFR [60]. Thus, we also conducted a more prudent, hypothesis-driven association study in which the association of these 53 reported loci with residual renal capacity was tested. The advantages of such an approach were that a less stringent statistical significance threshold was required (*p*-value < 0.05/53 = 0.00094) and this would enable confirmation/validation/replication of the previous work.

Genetic Risk Score Study

Different associated loci can also be aggregated per individual based on certain model to generate a person's genetic risk score (GRS) based on multi-locus analysis. Based on the assumed additive model, the GRS was defined as the sum of the risk allele dosages of the involved SNPs weighted by their corresponding effect size. This analysis was performed in PLINK v1.07 [80], [81]. The GRS was then be associated with each person's measured residual renal capacity to evaluate the combined effect of the discovered/previously known kidney function SNPs.

Ethics Statement

The protocol and study were approved by the Institutional Review Board of the University Medical Center Groningen. Written informed consent was given by all participants (all of whom are adults). According to Dutch law general consent for organ donation and transplantation includes consent for research projects. Conduct of study was also in accordance to the principles of the Declaration of Helsinki. All genetic and phenotypic (clinical) data were anonymized during collection, thus prior to analyses. Processing/analyses were also only done to aggregated data, no information about specified individual was processed, thus nor reported.

Chapter 6
Population and Phenotypic Data

Overview

This chapter presents and briefly discusses the gathered phenotypic data and outcomes to be later used in the association studies. The population data are presented, as well as the chosen primary phenotypic outcome. Choices/selection of covariates are also presented and discussed.

Population Baseline Characteristics

The basic characteristics of the study population are presented in Table 6.1, with more elaborate characteristics given in Appendix B.

Living Donor (1993-2007) Characteristics

	N	Value
Gender	282	
Female		156 (44.7%)
Male		126 (55.3%)
Donor age at time of nephrectomy	282	49 ± 10
Before nephrectomy (93 ± 83 days)		
BMI (kg/m^2)	221	26.2 ± 4.1
Diastolic blood pressure (mmHg)	256	80 ± 9
Systolic blood pressure (mmHg)	256	126 ± 26
ERPF (mL/min)	226	433 ± 89
GFR (mL/min)	249	116 ± 21
FF (mL/min)	228	0.27 ± 0.03
ERPFdopa (mL/min)	201	556 ± 120
GFRdopa (mL/min)	227	134 ± 23
FFdopa (mL/min)	203	0.24 ± 0.03
MDRD eGFR (mL/min)	272	81 ± 15
Serum creatinine (umol/L)	272	82 ± 13
Total cholesterol (mmol/L)	257	5.41 ± 1.01
After nephrectomy (98 ± 94 days)		
Diastolic blood pressure (mmHg)	206	80 ± 9
Systolic blood pressure (mmHg)	206	129 ± 14
ERPF (mL/min)	175	284 ± 52
GFR (mL/min)	207	73 ± 13
FF (mL/min)	176	0.26 ± 0.03
ERPFdopa (mL/min)	84	344 ± 71
GFRdopa (mL/min)	101	84 ± 16

FFdopa (mL/min)	84	0.25 ± 0.04
MDRD eGFR (mL/min)	269	56 ± 11
Serum creatinine (umol/L)	269	114 ± 21
Total cholesterol (mmol/L)	139	5.56 ± 1.05

Table 6.1. Summary of population characteristics. Continuous values are given in mean ± SD.

Table 6.1 shows that the population was rather well-balanced in gender, aged from 21 to 73 with the mean age of 49.53 years. In addition the population has about ideal mean blood pressure, but mean total cholesterol level and BMI value were a little bit higher than what are deemed as ideal. Appendix B showed that there were practically no donors with a history of myocardial infarctions or cerebrovascular accidents within the population. All in all, this suggests that the gathered population was an exceptionally healthy population. Consequently, comparability of our study population to the general population might then be challenged. To address this, the CKDGen genes were also studied for pre-nephrectomy eGFR (calculated based on serum creatinine, using MDRD formula) as a comparative study. CKDGen's study was a population study, taking a large size of the general population. Comparative study with these results would allow us to assess whether results from the general population can be replicated in the living kidney donors' population.

It is also noteworthy that several patients/donors tended to have missing data. Thus different analyses using different phenotypic outcomes (or covariates) were performed on different subsets of the total population discussed here.

Phenotypic Outcome

Based on the gathered data, correlations between the measured and estimated measurements were visualized in Figure 6.1. Some positive correlation was observed between the measured and estimated measurements. Though the correlations were of significance (at 95% confidence level), they were not all that strong. This suggested that there were different phenomena that underlie the different parameters. Considering the reported stimulation mechanism of dopamine infusion, it might be the case that the protocol does not capture hypertrophy of the nephrons (which also occurs over longer time), but only the hyperperfusion. This might also be the cause of the regression line's positive y-intercept. Thus, although it can offer good prediction of donor's adaptive capacity, estimated measurements by dopamine would not fully capture the residual renal capacity in single-kidney situation. Seeing also how the measured parameter taps directly into the single-kidney situation, the study used measured residual renal capacity as its primary endophenotype.

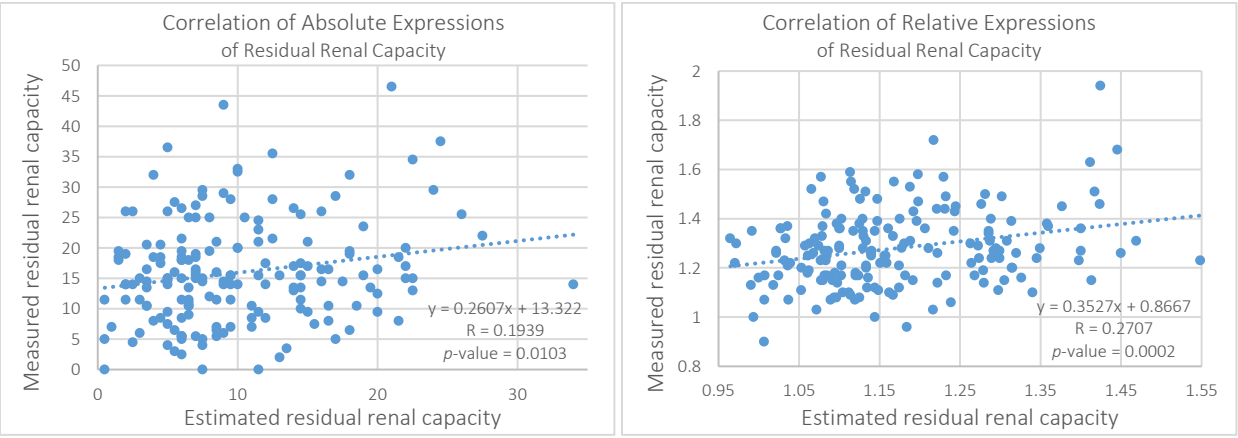

Figure 6.1. (A) Correlation of the measured and estimated absolute expressions of residual renal capacity. (B) Correlation of the measured and estimated relative expressions of residual renal capacity.

Ultimately the measured change of GFR in relative expression ($\Delta GFR_{measured}$) was used in the genetic analyses as a measure for residual renal capacity. In addition, analysis of eGFR was performed for the comparative study with CKDGen results. Logarithmic (base 10) transformation was done to the $\Delta GFR_{measured}$ data. The resulting distribution is shown in Figure 6.2. The observed three outlying points were checked back to the correspondence files. Since no peculiarity was observed (i.e. time of follow up, no special surgery/clinical remarks, etc.) and values were consistent across measurements/examinations, the points were assumed as individuals with exceptional $\Delta GFR_{measured}$.

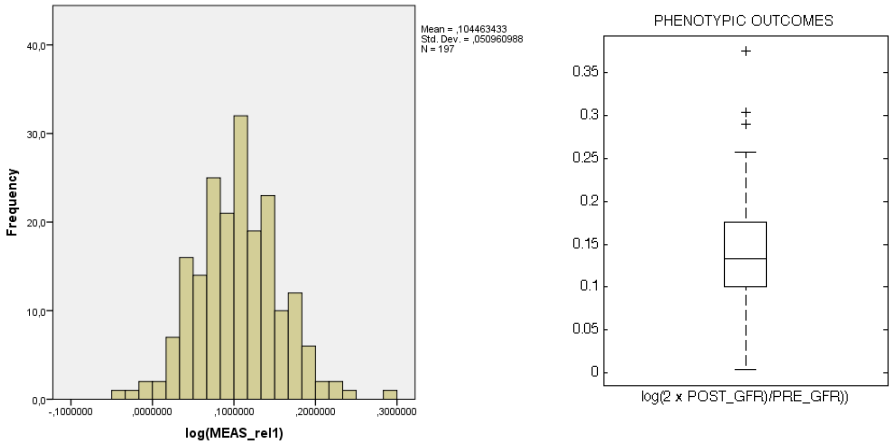

Figure 6.2. (A) Histogram of log transformed $\Delta GFR_{measured}$ (B) Box plot of log transformed $\Delta GFR_{measured}$

eGFR data were also log-transformed, resulting in distribution shown in Figure 6.3. There was one outlying value observed in the cleaned dataset. This measurement was also retained and regarded as an exceptional pre-nephrectomy eGFR value. It might be noteworthy that this measurement was not one of the outlying values of the $\Delta GFR_{measured}$ data.

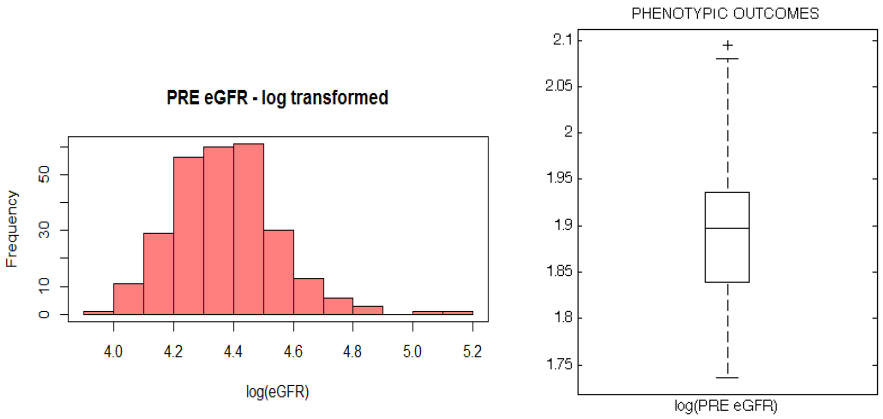

Figure 6.3. (A) Histogram of log transformed pre-nephrectomy eGFR data (mean = 1.89, std = 0.0715) (B) Box plot of log transformed pre-nephrectomy eGFR.

Residual Renal Capacity Covariates

In order to observe confounding effects of the chosen phenotypic measure, the phenotypic data were looked further into especially in respect to the said phenotypic measure. Appendix B also shows bar graphs of different phenotypic data in for different quartiles of $\log(\Delta GFR_{measured})$. These graphs showed that age had a significant influence (or confounding effect) on $\Delta GFR_{measured}$; while BMI, blood cholesterol level, and diastolic blood pressure did not. Systolic blood pressure seemed to have confounding effect to some extent on $\Delta GFR_{measured}$, which is rather expected since hypertension and kidney function are closely related. However, hypertension was already tightly screened for in this population, leaving rather negligible variance explained by systolic blood pressure in the data.

Figure 6.4 shows the correlation between age and $\Delta GFR_{measured}$ better. As presented, age has a significant negative correlation to $\Delta GFR_{measured}$, as older donors tended to have lower $\Delta GFR_{measured}$. Thus age was included as a covariate in the association studies.

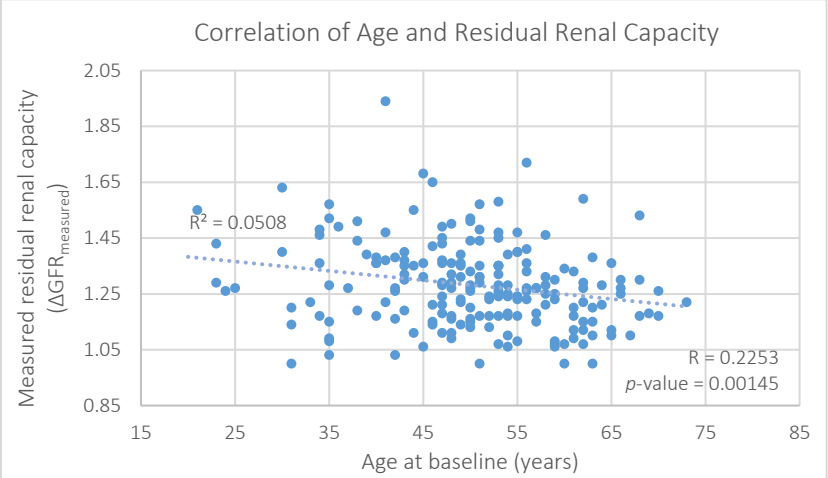

Figure 6.4. Correlation between age and measured residual renal capacity (in relative expression). The observed negative correlation was proven significant.

Other (reviewed) association studies regarding renal capacity have only used eGFR as the phenotypic measure. eGFR is calculated based on gender. Thus to achieve comparable outcome, gender was also used a covariate in the association analyses.

Chapter 7
Genome-wide Association Study of Residual Kidney Capacity

Overview

This chapter presents the results of the GWAS of residual kidney capacity.

Association with $\Delta GFR_{measured}$

Data were available for 103 individuals (thus both with phenotypic and genotypic data). Power analysis was done to determine the effect size that could be detected. To achieve at least 80% of statistical power for an additive model, with allele frequency of (at least) 0.02, and the calculated phenotypic measure (mean of 0.1045 and standard deviation of 0.0510), a SNP effect size of at least 0.15 would be necessary.

Figure 7.1 presents the Manhattan plot of the GWAS results. All of the *p*-values of the interrogated SNPs are plotted (in –logarithm expression) against each SNP's location (grouped per chromosome). The Bonferroni-corrected *p*-value criteria of 5E-08 is also given with the dotted red line. As it can be seen, three SNPs were found to be significantly associated with the $\Delta GFR_{measured}$, located on chromosome two, six, and ten.

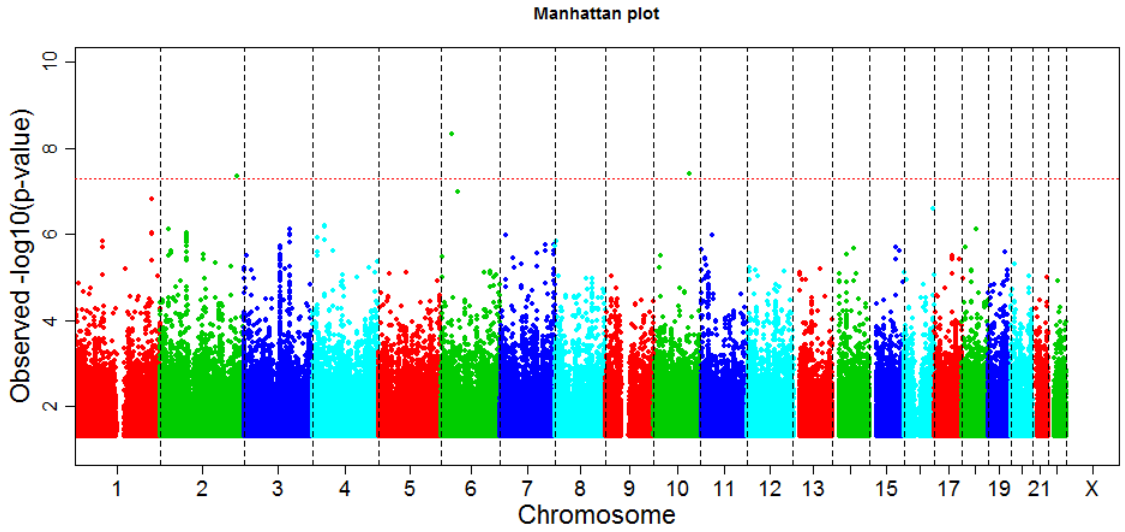

Figure 7.1. Manhattan plot of GWAS results of the residual renal capacity. Each point represents a tested SNP, displayed by chromosomal position (x-axis). Y-axis shows –log(p-value) for each SNP. Significantly associated SNPs are shown on chromosome two, six, and nine.

Figure 7.2 below shows the QQ plots of the *p*-values from the GWAS, based on different allele frequency and imputation quality filtering levels. The majority of SNPs were found along the diagonal (that is, the null line of no association), implying that there was no evidence of population stratification. Furthermore, filtering for allele frequency (alleles with MAF<2% were filtered out) and imputation quality (loci with HWE *p*-value < 0.0001 were also excluded) did not change the QQ plots implying that rare and badly imputed alleles were filtered sufficiently.

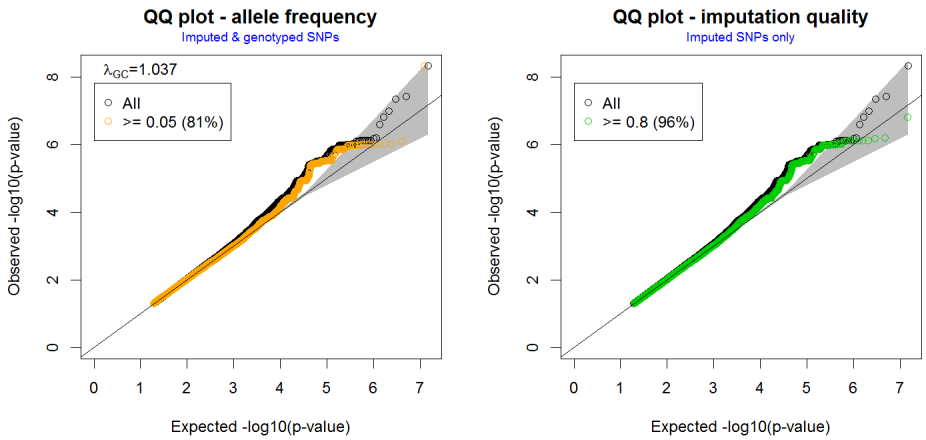

Figure 7.2. Quality control plots. The QQ plots suggest that rare alleles and poorly imputed alleles were filtered sufficiently.

The following, Table 7.1, shows more details of the three loci found to be genome-wide significantly associated with delta eGFR (*p*-value < 5E-08). Given their location on different chromosomes, these loci were not in LD and thus independent of each other. Sign of the beta/effect size of these loci suggests the direction of the effect, where positive numbers would mean that every copy of the risk/effect allele increases residual renal capacity by the magnitude of effect. With the smallest (effect-) allele frequency of 0.02, the power analysis presented earlier holds. Given that all beta (effect size) values were bigger than 0.15, the study did achieve sufficient statistical power (>80%) to reject the null hypothesis.

SNP ID	Chr	Position (bp)	Effect/Non-effect allele (EAF)	*P* value	Beta
rs530595485	6	32,525,768	C/T (0.09)	4.65E-09	0.336
rs80158280	10	102,585,914	T/G (0.04)	3.78E-08	0.411
rs16864916	2	224,406,554	A/G (0.02)	4.53E-08	0.646

Table 7.1. Three loci found to be most associated with measured residual renal capacity. Abbreviations: Chr, chromosome number; bp, basepairs; EAF, effect allele frequency

Chapter 8
CKDGen Loci Association Studies

Overview

This chapter presents the results of the more hypothesis-driven association studies of the loci previously associated with kidney function by CKDGen. Firstly, the loci were associated with the residual renal capacity endophenotype. As comparative study, association of loci with living donor's pre-nephrectomy eGFR was also done and presented in this chapter.

Association with $\Delta GFR_{measured}$

As mentioned, association studies with less markers would cut down the multiple testing correction for a more lenient significance threshold. For the 53 loci published by CKDGen [60], a significance threshold of 0.05/53 = 9.43E-04 was sufficient to avoid false positives.

Firstly, an analysis using the study original endophenotype was done. Results can be found in Table 8.1, sorted by the *p*-values of the 53 loci (in their association with $\Delta GFR_{measured}$). Lowest *p*-value of the studied loci valued at 5.38E-02 (locus rs12136063), so no significant association was observed amongst these loci with the endophenotype.

Power analysis suggested a minimum effect size of 0.065 to achieve 80% power. The highest effect size among the 53 loci reported in the publication was 0.015. This suggests that this study is underpowered for the necessary effect size. As such, no conclusion could be drawn from these results, as it can also be that none of the CKDGen kidney function loci addresses residual kidney function, as the nature of these two endophenotypes (eGFR and delta eGFR) is quite different.

SNP ID	Chr	Position (bp)	Effect/Non-effect allele (EAF)	*P* value	Beta
rs12136063	1	110,014,170	G/A (0.32)	5.38E-02	0.047
rs2453580	17	19,438,321	C/T (0.35)	8.29E-02	0.044
rs164748	16	89,708,292	G/C (0.32)	9.17E-02	0.046
rs163160	11	2,789,955	G/A (0.22)	9.38E-02	-0.049
rs3750082	7	32,919,927	A/T (0.32)	1.33E-01	0.038
rs4744712	9	71,434,707	A/C (0.37)	1.78E-01	0.033
rs12124078	1	15,869,899	G/A (0.34)	2.00E-01	0.034
rs6088580	20	33,285,053	C/G (0.46)	2.13E-01	-0.028
rs10277115	7	1,285,195	A/T (0.26)	2.50E-01	0.032
rs10774021	12	349,298	C/T (0.37)	2.62E-01	-0.029
rs12460876	19	33,356,891	C/T (0.36)	2.75E-01	0.028
rs6431731	2	15,863,002	C/T (0.04)	2.96E-01	-0.063
rs267734	1	150,951,477	C/T (0.23)	2.97E-01	-0.029
rs3925584	11	30,760,335	C/T (0.39)	3.18E-01	-0.026
rs10109414	8	23,751,151	T/C (0.41)	3.47E-01	-0.022
rs2453533	15	45,641,225	A/C (0.39)	3.66E-01	-0.021
rs848490	7	77,555,005	G/C (0.22)	3.76E-01	0.026

rs7805747	7	151,407,801	A/G (0.28)	3.83E-01	0.025
rs6465825	7	77,416,439	C/T (0.33)	3.83E-01	0.020
rs10513801	3	185,822,353	G/T (0.12)	4.35E-01	-0.031
rs10491967	12	3,368,093	A/G (0.1)	4.54E-01	-0.032
rs347685	3	141,807,137	C/A (0.29)	4.61E-01	0.019
rs4014195	11	65,506,822	G/C (0.36)	4.69E-01	0.017
rs4667594	2	170,008,506	T/A (0.43)	4.82E-01	-0.018
rs2279463	6	160,668,389	G/A (0.08)	4.90E-01	-0.030
rs12917707	16	20,367,690	T/G (0.17)	5.20E-01	-0.021
rs228611	4	103,561,709	A/G (0.42)	5.60E-01	-0.014
rs8091180	18	77,164,243	G/A (0.41)	5.77E-01	-0.014
rs7759001	6	27,341,409	G/A (0.23)	6.04E-01	-0.014
rs9895661	17	59,456,589	C/T (0.15)	6.31E-01	-0.016
rs881858	6	43,806,609	G/A (0.27)	6.58E-01	0.012
rs10794720	10	1,156,165	T/C (0.07)	6.59E-01	0.021
rs13538	2	73,868,328	G/A (0.19)	6.68E-01	0.013
rs6459680	7	156,258,568	G/T (0.22)	7.02E-01	0.012
rs17319721	4	77,368,847	A/G (0.48)	7.30E-01	-0.008
rs10994860	10	52,645,424	T/C (0.16)	7.59E-01	0.010
rs626277	13	72,347,696	C/A (0.41)	7.73E-01	-0.007
rs1106766	12	57,809,456	T/C (0.29)	7.90E-01	-0.007
rs1394125	15	76,158,983	A/G (0.39)	8.06E-01	0.006
rs7956634	12	15,321,194	C/T (0.16)	8.33E-01	0.007
rs1260326	2	27,730,940	T/C (0.37)	8.44E-01	-0.005
rs2712184	2	217,682,779	C/A (0.48)	8.87E-01	0.004
rs2802729	1	243,501,763	A/C (0.41)	8.93E-01	0.003
rs3850625	1	201,016,296	A/G (0.1)	9.04E-01	-0.005
rs11959928	5	39,397,132	A/T (0.43)	9.15E-01	-0.003
rs17216707	20	52,732,362	C/T (0.21)	9.22E-01	0.003
rs491567	15	53,946,593	C/A (0.21)	9.23E-01	0.003
rs2928148	15	41,401,550	G/A (0.45)	9.23E-01	0.002
rs9682041	3	170,091,902	C/T (0.1)	9.30E-01	0.003
rs11666497	19	38,464,262	T/C (0.18)	9.45E-01	0.002
rs7208487	17	37,543,449	G/T (0.16)	9.91E-01	0.001
rs6795744	3	13,906,850	A/G (0.12)	9.96E-01	-0.001

Table 8.1. 53 renal function (eGFR) loci found in CKDGen studies [60]*, associated with* $\Delta GFR_{measured}$ *in the current living donor population. Abbreviations: Chr, chromosome number; bp, basepairs; EAF, effect allele frequency*

Association with Pre-nephrectomy eGFR

To scrutinize further the study population, a similar study was also performed with pre-nephrectomy eGFR as endophenotype in order to replicate the results previously found by CKDGen [60]. The same significance threshold was used. The result of this analysis is presented in Table 8.2. As seen, there was also no significant loci associated above the significance threshold. Power analysis also suggested that our study was also underpowered to detect the necessary effect size (as found in the CKDGen paper).

SNP ID	Chr	Position (bp)	Effect/Non-effect allele (EAF)	*P* value	Beta
rs7805747	7	151,407,801	A/G (0.35)	1.10E-02	-3.411
rs12124078	1	15,869,899	A/G (0.33)	2.10E-02	2.929
rs2453580	17	19,438,321	T/C (0.23)	4.78E-02	2.484
rs8091180	18	77,164,243	G/A (0.1)	6.62E-02	2.195

rs163160	11	2,789,955	C/A (0.44)	7.19E-02	2.756
rs6088580	20	33,285,053	T/C (0.05)	1.10E-01	-1.728
rs1260326	2	27,730,940	A/T (0.42)	1.47E-01	1.880
rs848490	7	77,555,005	A/C (0.47)	1.48E-01	2.038
rs347685	3	141,807,137	C/T (0.36)	1.67E-01	-1.809
rs9895661	17	59,456,589	A/G (0.21)	1.68E-01	2.141
rs2453533	15	45,641,225	G/A (0.12)	2.21E-01	1.436
rs9682041	3	170,091,902	A/C (0.27)	2.23E-01	2.315
rs2279463	6	160,668,389	T/C (0.1)	2.63E-01	2.371
rs267734	1	150,951,477	T/G (0.11)	2.75E-01	-1.526
rs2712184	2	217,682,779	G/A (0.46)	2.79E-01	1.363
rs10994860	10	52,645,424	G/A (0.49)	2.92E-01	-1.714
rs6465825	7	77,416,439	T/A (0.45)	2.95E-01	1.206
rs7208487	17	37,543,449	A/G (0.09)	3.26E-01	1.604
rs10794720	10	1,156,165	A/G (0.23)	4.19E-01	-2.020
rs4667594	2	170,008,506	A/G (0.26)	4.84E-01	-0.889
rs13538	2	73,868,328	T/A (0.26)	4.86E-01	1.000
rs7759001	6	27,341,409	G/A (0.3)	4.96E-01	0.921
rs4014195	11	65,506,822	T/G (0.23)	5.04E-01	0.775
rs1394125	15	76,158,983	T/A (0.31)	5.10E-01	0.781
rs626277	13	72,347,696	T/C (0.33)	5.14E-01	0.793
rs6431731	2	15,863,002	C/G (0.22)	5.34E-01	1.714
rs4744712	9	71,434,707	C/T (0.41)	5.38E-01	0.731
rs11666497	19	38,464,262	C/A (0.39)	5.54E-01	-0.968
rs491567	15	53,946,593	C/T (0.07)	5.72E-01	0.847
rs11959928	5	39,397,132	C/T (0.16)	5.87E-01	-0.695
rs7956634	12	15,321,194	A/G (0.19)	5.95E-01	0.839
rs3925584	11	30,760,335	T/C (0.39)	5.98E-01	-0.636
rs10277115	7	1,285,195	C/G (0.38)	6.07E-01	0.688
rs12136063	1	110,014,170	T/C (0.37)	6.16E-01	0.602
rs10513801	3	185,822,353	G/A (0.13)	6.26E-01	-0.964
rs10774021	12	349,298	T/C (0.18)	6.47E-01	0.586
rs6795744	3	13,906,850	C/T (0.29)	6.60E-01	0.804
rs17319721	4	77,368,847	A/C (0.41)	7.64E-01	-0.352
rs12460876	19	33,356,891	A/G (0.47)	7.76E-01	-0.357
rs2802729	1	243,501,763	C/A (0.38)	7.79E-01	-0.360
rs17216707	20	52,732,362	A/C (0.2)	7.95E-01	-0.389
rs881858	6	43,806,609	G/A (0.38)	8.00E-01	-0.336
rs10491967	12	3,368,093	G/T (0.19)	8.02E-01	-0.492
rs228611	4	103,561,709	C/G (0.33)	8.12E-01	-0.267
rs3850625	1	201,016,296	T/C (0.37)	8.34E-01	0.397
rs12917707	16	20,367,690	T/G (0.16)	9.01E-01	0.202
rs1106766	12	57,809,456	T/C (0.16)	9.03E-01	0.160
rs6459680	7	156,258,568	A/G (0.43)	9.30E-01	0.130
rs2928148	15	41,401,550	T/C (0.38)	9.62E-01	0.056
rs3750082	7	32,919,927	C/T (0.18)	9.62E-01	-0.063
rs164748	16	89,708,292	G/C (0.44)	9.95E-01	0.008
rs10109414	8	23,751,151	T/C (0.23)	9.97E-01	-0.004

Table 8.2. 53 renal function (eGFR) loci found in CKDGen studies [60]*, associated with pre-nephrectomy eGFR values in the current living donor population. Abbreviations: Chr, chromosome number; bp, basepairs; EAF, effect allele frequency*

To test whether lack of association was indeed due to small sample size, the found effect size of each locus was correlated to the published effect size of the same locus. This correlation is shown in Figure 8.1 below. A positive, significant correlation (R = 0.242) is found. This implies that the effect sizes observed in this study resemble the ones of the CKDGen consortium and hence that the lack of significant results per SNP can be attributed to the low study power, due to the small sample size.

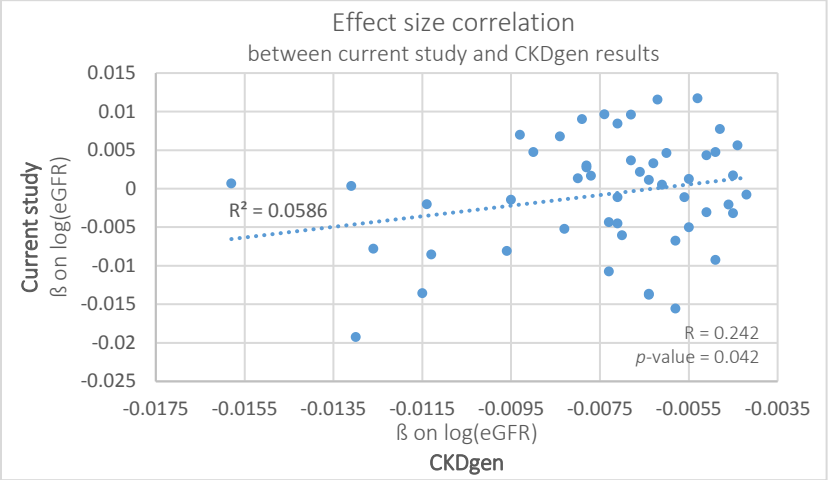

Figure 8.1. Correlation between CKDGen's effect size and current study's effect size (direction of effect is compared to the CKDGen's negative-effect alleles, if the found effect allele was the same, the effect was taken as is, otherwise the direction of effect was inverted). P-value is given for the correlation.

Chapter 9
Genetic Risk Scores

Overview

Last analyses were aimed to see how much phenotypic variance could be explained by the studied loci by means of genetic risk scores. This chapter presents the results of these analyses. The GRS was first associated with the residual renal capacity, then also with pre-nephrectomy eGFR.

Association with $\Delta GFR_{measured}$

The individual's alleles at the 3 loci found to be significantly associated with $\Delta GFR_{measured}$ (Table 7.1) were used to generate his genetic risk score (GRS). This score is weighted by the corresponding effect size found in the CKDGen GWAS study (Chapter 7). The GRS was then associated with the $\Delta GFR_{measured}$, as plotted in Figure 9.1 below. Linear regression revealed that the GRS explained 39.2% of the trait variance.

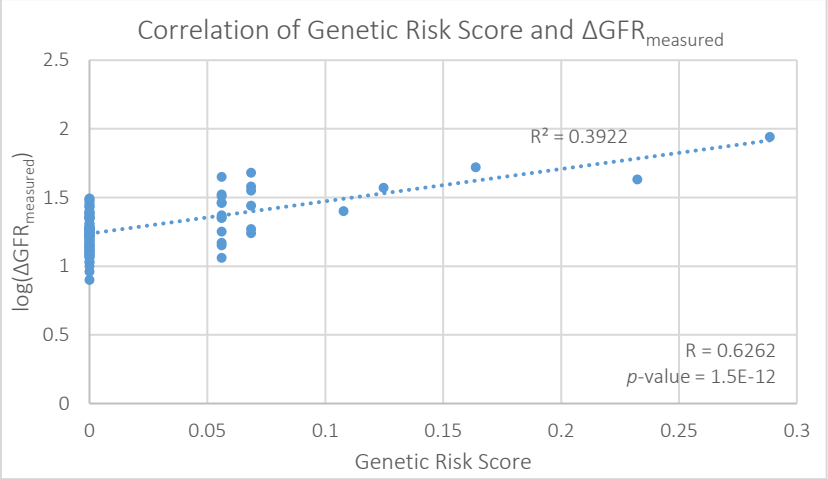

Figure 9.1. Correlation of GRS (based on the three residual renal capacity loci found in GWAS) and $\Delta GFR_{measured}$. GRS was weighted with effect size (and direction) found in the GWAS as well (Table 7.1). P-value is given for the correlation.

Association with Pre-nephrectomy eGFR

Similar protocol was done with the CKDGen 53 kidney function loci. The loci, effect allele, effect size, and direction of effect were taken directly from their latest publication [60]. GRS generated was then associated with pre-nephrectomy eGFR values of the living donor population in the study. Given the two populations were independent from each other (but both of Caucasian ethnicity), effect size was used as given in literature. Similar definition of GRS was used, as the sum of the effect allele dosages weight by the corresponding original effect size.

When all 53 loci were included to generate the GRS, association with trait proved to be very weak ($R^2 \approx 0.7\%$). One potential reason for this was again the small population in the study. It might be the case that the less-common alleles were not captured sufficiently in the population. Thus, loci with MAF less than 15% were filtered out, leaving only loci with effect allele frequency between 15% and 85%. The remaining loci (46 of them) were then used to generate the new GRS, which then associated with pre-nephrectomy GFR measurements, presented in Figure 9.2 below.

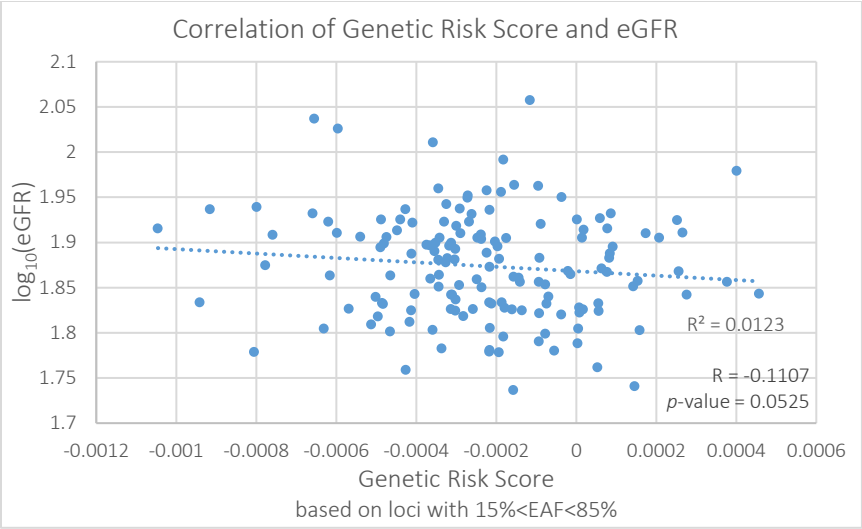

Figure 9.2. Correlation of GRS (based on CKDGen's loci, filtered on MAF>15%) and pre-nephrectomy eGFR. GRS was weighted with effect size (and direction) found in the publication [60]. *P-value was given for the correlation.*

The figure shows that an increase in GRS was associated with a decline in $\log_{10}$(eGFR), albeit not significant. Regression analysis suggested that 1.23% of the variance could be explained by the GRS.

Chapter 10
Discussions, Conclusions, and Recommendations

Lastly, this chapter concludes the findings of the project and answers the formulated central research question. Firstly the central research question is answered, followed by the three sub-research questions stated in Chapter 1. In the end, recommendations for future works are also put forward.

The central research question was '*Can genetic variants be identified that are associated with reported residual renal capacity in single-kidney situation in living kidney donors?*' In this study genetic variants that are associated with residual renal capacity in single-kidney situation in living kidney donors could be identified. Three SNPs were found to be associated beyond the genome-wide significance threshold with the formulated residual renal capacity in a population of (Caucasian) living kidney donors. The three SNPs are rs530595485 (chromosome 6), rs80158280 (chromosome 10), and rs16864916 (chromosome 2). SNP rs80158280 is in an intronic SNP in the *PAX2* gene and is found to be a regulatory variant. A polymorphism of this gene has been evidenced to be associated with congenital anomalies of the kidney and urinary tract (often named *CAKUT*) [82], [83]. This points the SNP's proximity to kidney function and can support real effect of the finding/SNP. Interestingly, SNP rs16864916 is located close to the *PAX3* gene and is also found to be regulatory. The third SNP, rs530595485 is a non-coding transcript exon variant in the *HLA-DRB6* gene in the major histocompatibility complex.

The findings should be replicated in an independent (Caucasian) population. By only validating three loci, multiple testing requirement/correction would be much lenient, allowing use of a much smaller validation population (population of 25 would already yield statistical power of 90%). Due to the unique phenotype used for the residual renal capacity (true measurement of GFR pre- and post-nephrectomy), it might be difficult to have validation population with the same available phenotype and genotype data. The use of other (more commonly available) biomarkers/measurements can be an option in the validation study. However, a thorough study on biomarker's suitability should first be done. For this purpose, further information/elaboration on collected eGFR data in the living donors' population is included in Appendix A.

Next to the central research question, this study also aimed to answer three sub-questions. The first sub-question of '*Which characteristics can be associated with the post-nephrectomy residual renal capacity?*' was mostly addressed in Chapter 6. It was revealed that age, gender, and systolic blood pressure had significant associations with the formulated expression of residual renal capacity. These findings are consistent with previous studies/findings. Body mass index, however, did not prove to have a significant effect on residual renal capacity in the healthy population of living kidney donors. This might be contradictory to the findings in negative correlation between BMI and eGFR in general population [84]. A reason might be the population screening of the living donors, where overweight/obese individuals (BMI>30kg/m^2) are often regarded unfit for donation or are advised to lose weight prior to being accepted as donor. Thus perhaps effects from BMI could not be captured in the available data. Another reason might be that, unlike eGFR, residual renal capacity is not associated/correlated with BMI. This, however, has not yet been studied in the general population before.

Chapter 7-9 addressed in parts the second sub-question, *'Can known associations with eGFR be replicated in our cohort of living kidney donors?*' Known associations with eGFR could not be replicated within the cohort of living kidney donors. Results in Chapter 8 implied that this was due to the underpowered sample size, since a significant association was observed between this study's and the reported effect sizes of kidney function loci. The combined (negative) effect of all known SNPs had a negative correlation between the two studies results. Furthermore, this sub-

question was also aimed to assess the comparability of the living kidney donors' population to the general population. The percentage of variance in eGFR explained by the SNPs combined a GRS found in this study was slightly less than that observed earlier in a general population [85]. In the LifeLines Cohort Study (N = 13,191) association of the weighted GRS composed of 52 of the loci with eGFR (calculated from serum creatinine) revealed that it could explain 2.01% of the eGFR variance. The discrepancy, i.e. the lower variance explained in the current study result compared to LifeLines' results, might be due to exclusion of the less-common alleles as well as to the poorness of eGFR as a renal function biomarker in healthy population.

The last sub-question of *'How much variance in post-nephrectomy residual renal capacity can be explained by genes?'* was addressed in Chapter 9. In the available population, the found genes/genetic variants could explain 39% of the residual renal capacity variance (the rest of variance can be due to epistasis, environment, or just randomness). However, this number is an overestimation for a number of reasons. Firstly, GRS association was done in the same samples/population from which the SNPs were identified. Ideally, GRS association should be done in an independent population to yield unbiased estimates of percentage of explained variance. Furthermore, the population was rather small. Thus this explained variance can only be assumed true for this population and not generalized to other populations. In order to see the true level of variance that can be explained by the 3 loci, validation in an independent population is necessary.

In conclusion, although the sample used in this study was small, three loci were found that were significantly associated with residual renal capacity. Genes in these regions are *PAX2*, *PAX3*, and *HLA-DRB6*, which are logical functional candidates for this phenotype. Nevertheless, independent studies in other samples should be conducted to replicate and validate these three loci.

References

[1] G. Eknoyan, N. Lameire, R. Barsoum, K. U. Eckardt, A. Levin, N. Levin, et al., "The burden of kidney disease: Improving global outcomes," *Kidney Int.*, vol. 66, no. 4, pp. 1310–1314, 2004.

[2] A. S. Levey, K.-U. Eckardt, Y. Tsukamoto, A. Levin, J. Coresh, J. Rossert, et al., "Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO).," *Kidney Int.*, vol. 67, no. 6, pp. 2089–2100, 2005.

[3] C. M. O'Seaghdha and C. S. Fox, "Genome-wide association studies of chronic kidney disease: what have we learned?," *Nat Rev Nephrol*, vol. 8, no. 2, pp. 89–99, Feb. 2012.

[4] F. Hildebrandt, "Genetic kidney diseases," *Lancet*, vol. 375, no. 9722, pp. 1287–1295, 2010.

[5] L. J. Smyth, S. Duffy, a P. Maxwell, and a J. McKnight, "Genetic and epigenetic factors influencing chronic kidney disease.," *Am. J. Physiol. Renal Physiol.*, vol. 307, no. 7, pp. F757–76, 2014.

[6] A. J. Matas, J. M. Smith, M. A. Skeans, B. Thompson, S. K. Gustafson, M. A. Schnitzler, et al., "OPTN/SRTR 2012 Annual Data Report: Kidney," *Am. J. Transplant.*, vol. 14, no. SUPPL. 1, pp. 11–44, 2014.

[7] C. L. Davis, "Living-Donor Kidney Transplantation: A Review of the Current Practices for the Live Donor," *J. Am. Soc. Nephrol.*, vol. 16, no. 7, pp. 2098–2110, 2005.

[8] L. Wang, C. J. Yang, C. D. Medley, S. a. Benner, and W. Tan, "Locked nucleic acid molecular beacons," *J. Am. Chem. Soc.*, vol. 127, no. 45, pp. 15664–15665, 2005.

[9] A. D. Muzaale, A. B. Massie, M.-C. Wang, R. A. Montgomery, M. A. McBride, J. L. Wainright, et al., "Risk of end-stage renal disease following live kidney donation.," *Jama*, vol. 311, no. 6, pp. 579–86, 2014.

[10] J. C. Tan, S. Busque, B. Ho, and B. D. Myers, "Debate: PRO Position. Formal assessment of donor kidney function should be mandatory.," *Am. J. Nephrol.*, vol. 33, no. 3, pp. 198–200; discussion 205, 2011.

[11] P. Delanaye, L. Weekers, B. E. Dubois, E. Cavalier, O. Detry, J. P. Squifflet, et al., "Outcome of the living kidney donor," *Nephrol. Dial. Transplant.*, vol. 27, no. 1, pp. 41–50, 2012.

[12] L. Overleg, N. Lont, and J. J. H. Van Der Heide, "Nederlandse Richtlijn Evaluatie van Potentiële Donoren voor Levende Donor Niertransplantatie," pp. 1–64, 2008.

[13] C. R. Lenihan, S. Busque, G. Derby, K. Blouch, B. D. Myers, and J. C. Tan, "Longitudinal study of living kidney donor glomerular dynamics after nephrectomy," *J. Clin. Invest.*, vol. 125, no. 3, pp. 1311–1318, 2015.

[14] M. Rook, H. S. Hofker, W. J. Van Son, J. J. Homan Van Der Heide, R. J. Ploeg, and G. J. Navis, "Predictive capacity of pre-donation GFR and renal reserve capacity for donor renal function after living kidney donation," *Am. J. Transplant.*, vol. 6, no. 7, pp. 1653–1659, 2006.

[15] G. Mendel, "Versuche über Pflanzenhybriden," *Verhandlungen des naturforschenden Vereines Brunn 4 3*, vol. 44, 1866.

[16] W. S. Sutton, "The Chromosomes in Heredity," *Biol. Bull.*, vol. 4, no. 5, p. 231, 1903.

[17] E. W. Crow and J. F. Crow, "100 Years ago: Walter sutton and the chromosome theory of heredity," *Genetics*,

vol. 160, no. 1, pp. 1–4, 2002.

[18] F. Griffith, "The Significance of Pneumococcal Types," *J. Hyg. (Lond).*, vol. 27, no. 02, p. 113, Jan. 1928.

[19] J. D. Watson and F. H. C. Crick, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid," *Nature*, vol. 171, no. 4356, pp. 737–738, Apr. 1953.

[20] F. Crick, "Central Dogma of Molecular Biology," *Nature*, vol. 227, no. 8, pp. 561–563, 1970.

[21] Lodish H, B. A, and Z. SL, "Structure of Nucleic Acids," in *Molecular Cell Biology*, 4th ed., New York: W. H. Freeman, 2000.

[22] OpenStax College, "Anatomy & Physiology," *Connexions Web site*, 2013. .

[23] Khan Academy, "Intro to gene expression (central dogma)," *Central dogma (DNA to RNA to protein)*, 2016. [Online]. Available: https://www.khanacademy.org/science/biology/gene-expression-central-dogma/central-dogma-transcription/a/intro-to-gene-expression-central-dogma.

[24] D. Latchman, *Gene Regulation*. Taylor & Francis, 2007.

[25] B. Alberts, J. Wilson, and T. Hunt, *Molecular biology of the cell*. New York: Garland Science, 2008.

[26] C. N. Rotimi and L. B. Jorde, "Ancestry and disease in the age of genomic medicine.," *N. Engl. J. Med.*, vol. 363, no. 16, pp. 1551–8, 2010.

[27] S. Levy, G. Sutton, P. C. Ng, L. Feuk, A. L. Halpern, B. P. Walenz, et al., "The diploid genome sequence of an individual human," *PLoS Biol.*, vol. 5, no. 10, pp. 2113–2144, 2007.

[28] C. Vaisse, K. Clement, B. Guy-Grand, and P. Froguel, "A frameshift mutation in human MC4R is associated with a dominant form of obesity.," *Nat. Genet.*, vol. 20, no. 2, pp. 113–114, 1998.

[29] C. M. O'Seaghdha and C. S. Fox, "Genome-wide association studies of chronic kidney disease: what have we learned?," *Nat. Rev. Nephrol.*, vol. 8, no. 2, pp. 89–99, 2011.

[30] I. M. Nolte, J. M. McCaffery, and H. Snieder, "Handbook of Behavioral Medicine: Methods and Applications," A. Steptoe, Ed. New York, NY: Springer New York, 2010, pp. 423–441.

[31] A. Waugh and A. Grant, *Ross & Wilson Anatomy and Physiology in Health and Illness*. Elsevier Health Sciences UK, 2014.

[32] K. E. Barrett, S. M. Barman, S. Boitano, and H. L. Brooks, *Ganong's Review of Medical Physiology (Enhanced EB)*. McGraw-Hill Education, 2009.

[33] M. . J. Alam, F. R. Chowdhury, F. I. Chowdhury, A. Kumarsaha, A. A. Hoque, and M. Mujiburrahman, "Creatinine Clearnce Rate (CCR) in Dignosing Different Stages of Chronic Kidney Disease (CKD) - An Update and Review," *J Med.*, vol. 7, no. 2, pp. 64–68, 2006.

[34] A. D. Rule, T. S. Larson, E. J. Bergstralh, J. M. Slezak, S. J. Jacobsen, and F. G. Cosio, "Using Serum Creatinine To Estimate Glomerular Filtration Rate: Accuracy in Good Health and in Chronic Kidney Disease," *Ann. Intern. Med.*, vol. 141, no. 12, pp. 929–937, 2004.

[35] J. P. Bosch, A. Saccaggi, A. Lauer, C. Ronco, M. Belledonne, and S. Glabman, "Renal functional reserve in humans. Effect of protein intake on glomerular filtration rate," *Am. J. Med.*, vol. 75, no. 6, pp. 943–950, 1983.

[36] P. M. T. E. R. Wee and A. M. T. A. J. M. Donker, "Renal reserve filtration capacity before and after kidney

donation," pp. 393–399, 1990.

[37] T. H. Hostetter, J. L. Olson, H. G. Rennke, M. a Venkatachalam, and B. M. Brenner, "Hyperfiltration in remnant nephrons: a potentially adverse response to renal ablation.," *Am. J. Physiol.*, vol. 241, no. 1, pp. F85–F93, 1981.

[38] S. Barai, S. Gambhir, N. Prasad, R. K. Sharma, and M. Ora, "Functional renal reserve capacity in different stages of chronic kidney disease," *Nephrology*, vol. 15, no. 3, pp. 350–353, 2010.

[39] T. H. Hostetter, T. W. Meyer, H. G. Rennke, and B. M. Brenner, "Chronic effects of dietary protein in the rat with intact and reduced renal mass.," *Kidney Int.*, vol. 30, no. 4, pp. 509–517, 1986.

[40] A. Hartmann, P. Fauchald, L. Westlie, I. B. Brekke, and H. Holdaas, "The risk of living kidney donation," *Nephrol. Dial. Transplant.*, vol. 18, no. 5, pp. 871–873, 2003.

[41] G. S. Rosenblatt, N. Nakamura, and J. M. Barry, "End-Stage Renal Disease After Kidney Donation: A Single-Center Experience," *Transplant. Proc.*, vol. 40, no. 5, pp. 1315–1318, 2008.

[42] A. J. Matas, S. T. Bartlett, A. B. Leichtman, and F. L. Delmonico, "Morbidity and mortality after living kidney donation, 1999-2001: survey of United States transplant centers.," *Am. J. Transplant*, vol. 3, no. 7, pp. 830–834, 2003.

[43] D. M. Thomas, G. a Coles, and J. D. Williams, "What does the renal reserve mean?," *Kidney Int.*, vol. 45, no. 2, pp. 411–6, 1994.

[44] P. M. ter Wee, E. van Ballegooie, J. B. Rosman, S. Meijer, and A. J. M. Donker, "Renal Reserve Filtration Capacity in Patients With Type 1 (Insulin-Dependent) Diabetes Mellitus," *Nephrol. Dial. Transplant.*, vol. 2, no. 6, pp. 504–509, 1987.

[45] M. Williams, J. B. Young, R. M. Rosa, S. Gunn, F. H. Epstein, and L. Landsberg, "Effect of protein ingestion on urinary dopamine excretion. Evidence for the functional importance of renal decarboxylation of circulating 3,4-dihydroxyphenylalanine in man," *J. Clin. Invest.*, vol. 78, no. 6, pp. 1687–1693, 1986.

[46] P. M. ter Wee, J. B. Rosman, S. Van Der Geest, W. J. Sluiter, and A. J. M. Donker, "Renal hemodynamics during separate and combined infusion of amino acids and dopamine.," *Kidney Int.*, vol. 29, no. 4, p. 87O–874, 1986.

[47] P. A. Jaoude, L. Dubourg, J. Bacchetta, J. Berthiller, B. Ranchin, and P. Cochat, "Congenital versus acquired solitary kidney: Is the difference relevant?," *Nephrol. Dial. Transplant.*, vol. 26, no. 7, pp. 2188–2194, 2011.

[48] B. M. Brenner, "Nephron adaptation to renal injury or ablation.," *Am. J. Physiol.*, vol. 249, pp. F324–F337, 1985.

[49] F. Sanger, S. Nicklen, and a R. Coulson, "DNA sequencing with chain-terminating inhibitors.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 74, no. 12, pp. 5463–7, 1977.

[50] M. P. Sawicki, G. Samara, M. Hurwitz, and E. Passaro, "Human Genome Project," *Am. J. Surg.*, vol. 165, no. 2, pp. 258–264, 1993.

[51] E. S. Lander, "Initial impact of the sequencing of the human genome," *Nature*, vol. 470, no. 7333, pp. 187–197, 2011.

[52] T. International and H. Consortium, "The International HapMap Project.," *Nature*, vol. 426, no. 6968, pp. 789–796, 2003.

[53] K. A. Frazer, D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, et al., "A second generation human

haplotype map of over 3.1 million SNPs.," *Nature*, vol. 449, no. 7164, pp. 851–61, 2007.

[54] G. Thorisson and A. Smith, "The international HapMap project web site," *Genome Res.*, vol. 15, pp. 1592–1593, 2005.

[55] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, et al., "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations," *Nucleic Acids Res.*, vol. 42, no. D1, pp. 1001–1006, 2014.

[56] A. Reznichenko, H. Snieder, J. van den Born, M. H. de Borst, J. Damman, M. C. R. F. van Dijk, et al., "CUBN as a novel locus for end-stage renal disease: Insights from renal transplantation," *PLoS One*, vol. 7, no. 5, pp. 1–7, 2012.

[57] C. Pattaro, A. Köttgen, A. Teumer, M. Garnaas, C. A. Böger, C. Fuchsberger, et al., "Genome-wide association and functional follow-up reveals new loci for kidney function," *PLoS Genet.*, vol. 8, no. 3, 2012.

[58] A. Köttgen, C. Pattaro, C. a Böger, C. Fuchsberger, M. Olden, N. L. Glazer, et al., "Multiple loci associated with indices of renal function and chronic kidney disease," *Nat. Genet.*, vol. 42, no. 5, pp. 376–384, 2010.

[59] Y. Okada, X. Sim, M. J. Go, J.-Y. Wu, D. Gu, F. Takeuchi, et al., "Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations," *Nat Genet*, vol. 44, no. 8, pp. 904–909, Aug. 2012.

[60] C. Pattaro, A. Teumer, M. Gorski, A. Y. Chu, M. Li, V. Mijatovic, et al., "Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function.," *Nat. Commun.*, vol. 7, p. 10023, 2016.

[61] M. Gorski, A. Tin, M. Garnaas, G. M. McMahon, A. Y. Chu, B. O. Tayo, et al., "Genome-wide association study of kidney function decline in individuals of European descent.," *Kidney Int.*, pp. 1–13, 2014.

[62] T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, et al., "Finding the missing heritability of complex diseases.," *Nature*, vol. 461, no. 7265, pp. 747–53, 2009.

[63] S. Liu and Y. Song, "Building genetic scores to predict risk of complex diseases in humans: Is it possible?," *Diabetes*, vol. 59, no. 11, pp. 2729–2731, 2010.

[64] T. Johnson, "Efficient Calculation for Multi-SNP Genetic Risk Scores," p. 2012, 2012.

[65] K. Ryckman and S. M. Williams, "Calculation and use of the Hardy-Weinberg model in association studies," *Curr. Protoc. Hum. Genet.*, no. SUPPL. 57, pp. 1–11, 2008.

[66] S. H. S. Huang, A. P. Sharma, A. Yasin, R. M. Lindsay, W. F. Clark, and G. Filler, "Hyperfiltration affects accuracy of creatinine eGFR measurement," *Clin. J. Am. Soc. Nephrol.*, vol. 6, no. 2, pp. 274–280, 2011.

[67] W. S. Bush and J. H. Moore, "Chapter 11: Genome-Wide Association Studies," *PLoS Comput. Biol.*, vol. 8, no. 12, 2012.

[68] C. M. Lewis, "Genetic association studies: design, analysis and interpretation.," *Br. Bioinform*, vol. 3, no. 2, pp. 146–153, 2002.

[69] Y. R. Li, J. van Setten, S. S. Verma, Y. Lu, M. V Holmes, H. Gao, et al., "Concept and design of a genome-wide association genotyping array tailored for transplantation-specific studies.," *Genome Med.*, vol. 7, no. 1, p. 90, 2015.

[70] B. Keating, ..., Y. Li, ..., and F. Asselbergs, "Design and Implementation of the International Genetics and Translational Research in Transplantation Network.," *Transplantation*, vol. 99, no. 11, pp. 2401–12, 2015.

[71] G. A. McVean, D. M. Altshuler (Co-Chair), R. M. Durbin (Co-Chair), G. R. Abecasis, D. R. Bentley, A. Chakravarti, et al., "An integrated map of genetic variation from 1,092 human genomes," *Nature*, vol. 491, no. 7422, pp. 56–65, 2012.

[72] D. I. Boomsma, C. Wijmenga, E. P. Slagboom, M. a Swertz, L. C. Karssen, A. Abdellaoui, et al., "The Genome of the Netherlands: design, and project goals.," *Eur. J. Hum. Genet.*, vol. 22, no. 2, pp. 221–7, 2014.

[73] W. Y. S. Wang, B. J. Barratt, D. G. Clayton, and J. a Todd, "Genome-wide association studies: theoretical and practical concerns.," *Nat. Rev. Genet.*, vol. 6, no. 2, pp. 109–118, 2005.

[74] W. J. Gauderman and J. M. Morrison, "QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies," *http://hydra.usc.edu/gxe*, 2006. [Online]. Available: http://biostats.usc.edu/software.

[75] J. Marchini and B. Howie, "Genotype imputation for genome-wide association studies," *Nat. Rev. Genet.*, vol. 11, no. 7, pp. 499–511, 2010.

[76] D. Balding, "A tutorial on statistical methods for population association studies.," *Nat Rev Genet*, vol. 7, no. 10, pp. 781–791, 2006.

[77] P. J. Van Der Most, A. Vaez, B. P. Prins, M. L. Munoz, H. Snieder, B. Z. Alizadeh, et al., "QCGWAS: A flexible R package for automated quality control of genome-wide association results," *Bioinformatics*, vol. 30, no. 8, pp. 1185–1186, 2014.

[78] P. Flicek, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, et al., "Ensembl 2014," *Nucleic Acids Res.*, vol. 42, no. D1, pp. 749–755, 2014.

[79] J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, et al., "The Genotype-Tissue Expression (GTEx) project.," *Nat. Genet.*, vol. 45, no. 6, pp. 580–5, 2013.

[80] S. Purcell, "PLINK (v1.07)." p. http://pngu.mgh.harvard.edu/purcell/plink/, 2009.

[81] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, et al., "PLINK: A tool set for whole-genome association and population-based linkage analyses," *Am. J. Hum. Genet.*, vol. 81, no. 3, pp. 559–575, 2007.

[82] K. Nakanishi and N. Yoshikawa, "Genetic disorders of human congenital anomalies of the kidney and urinary tract (CAKUT)," *Pediatr. Int.*, vol. 45, no. 5, pp. 610–616, 2003.

[83] D. M. De Miranda, A. C. S. Dos Santos, G. S. Dos Reis, I. S. Freitas, T. G. R. Carvalho, L. A. C. De Marco, et al., "PAX2 polymorphisms and congenital abnormalities of the kidney and urinary tract in a Brazilian pediatric population: Evidence for a role in vesicoureteral reflux," *Mol. Diagnosis Ther.*, vol. 18, no. 4, pp. 451–457, 2014.

[84] R. Kawamoto, K. Kohara, Y. Tabara, T. Miki, N. Ohtsuka, T. Kusunoki, et al., "An association between body mass index and estimated glomerular filtration rate.," *Hypertens. Res.*, vol. 31, no. 8, pp. 1559–64, 2008.

[85] I. M. Nolte, P. J. van der Most, B. Z. Alizadeh, P. I. W. de Bakker, H. M. Boezen, L. Franke, et al., "Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study.," *Eur J Epidemiol*, vol. Submitted, 2016.

Appendix A
eGFR-based Endophenotype

As mentioned in the discussion of Chapter 10, a more commonly available endophenotype data, such as the eGFR, is attractive for validation or other future studies of residual renal capacity. For this reason, more information on the gathered eGFR data is presented below.

Correlation between one time-point GFR and eGFR measurements

GFR and eGFR measurements in one time-point were correlated to each other at two different time-points: before (Figure A.1) and after (Figure A.2) nephrectomy. Both have significant positive correlation ($R \approx 0.45$).

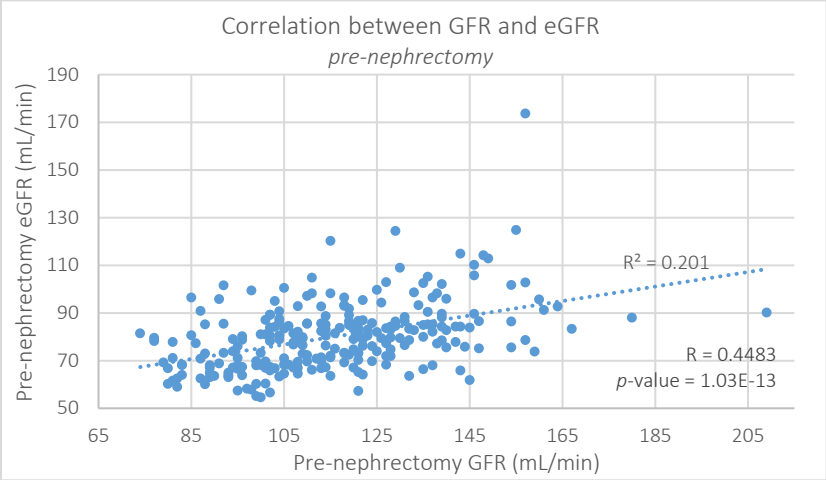

Figure A.1. Correlation between donors' pre-nephrectomy GFR (iothalamate-based measurement, at base level) and eGFR (serum creatinine-based, MDRD formula), measurements were done around 3 months (93 ± 83 days) before nephrectomy.

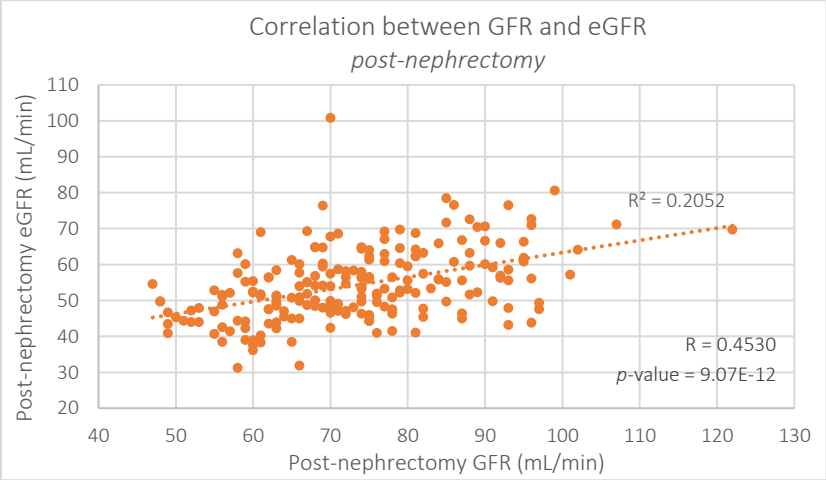

Figure A.2. Correlation between donors' post-nephrectomy GFR (iothalamate-based measurement, at base level) and eGFR (serum creatinine-based, MDRD formula), measurements were done around 3 months (98 ± 94 days) after nephrectomy.

Correlation between GFR- and eGFR-based residual renal capacity

The different expressions and measurements of residual renal capacity are correlated in this section. The different expressions and measurements are presented in Table A.1 below. The correlation plots that followed refer back to this table for the expressions.

Measured residual renal capacity:		
Absolute change in GFR	→ $GFR_{after} - (GFR_{before}/2)$	*(Eq A.1)*
Relative change in GFR	→ $GFR_{after}/(GFR_{before}/2)$	*(Eq A.2)*
Absolute change in eGFR	→ $eGFR_{after} - (eGFR_{before}/2)$	*(Eq A.3)*
Relative change in eGFR	→ $eGFR_{after}/(eGFR_{before}/2)$	*(Eq A.4)*
Estimated residual renal capacity (dopamine):		
Absolute change in GFR pre-donation	→ $(GFR_{dopa}\text{-}GFR)/2$	*(Eq A.5)*
Relative change in GFR pre-donation	→ $(GFR_{dopa})/2GFR$	*(Eq A.6)*

Table A.1. Summary of the different expressions (and measurements) of residual renal capacity.

Figure A.3 shows the correlation between GFR-based measured residual capacity (in absolute expression of Eq A.1) and the comparable eGFR-based expression of Eq A.3. Figure A.4 shows the correlation between GFR-based measured residual capacity (in relative expression of Eq A.2) and the comparable eGFR-based expression of Eq A.4.

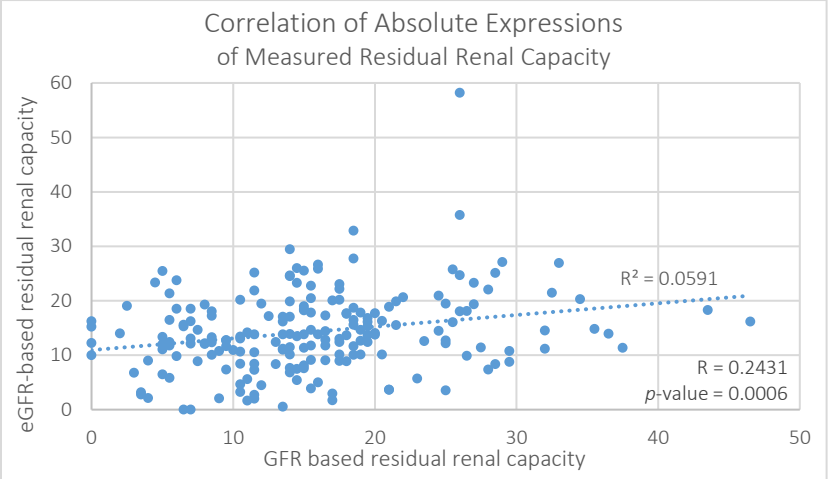

Figure A.3. Correlation of GFR- (Eq A.1) and eGFR-based (Eq A.3) measured residual renal capacity in absolute expression.

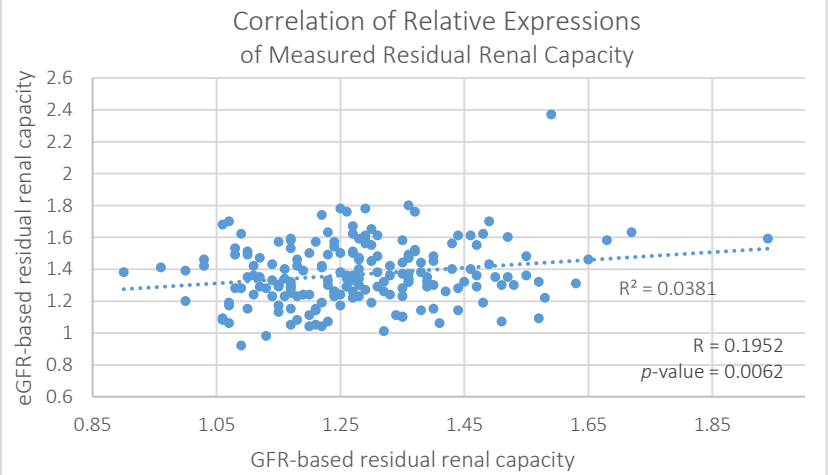

Figure A.4. Correlation of GFR- (Eq A.2) and eGFR-based (Eq A.4) measured residual renal capacity in relative expression.

Figure A.5 shows the correlation between eGFR-based measured residual renal capacity expressed in (Eq A.3) and the estimated residual renal capacity (Eq A.5) both in absolute expressions. Figure A.6 shows the correlation between eGFR-based measured residual renal capacity expressed in (Eq A.4) and the estimated residual renal capacity (Eq A.6) both in relative expressions.

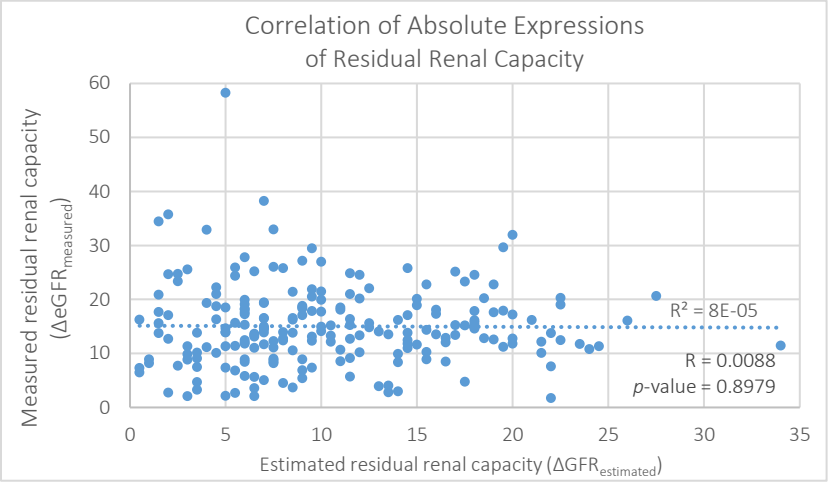

Figure A.5. Correlation of estimated residual renal capacity (Eq A.5) and eGFR-based measured residual renal capacity (Eq A.3) both in absolute expressions.

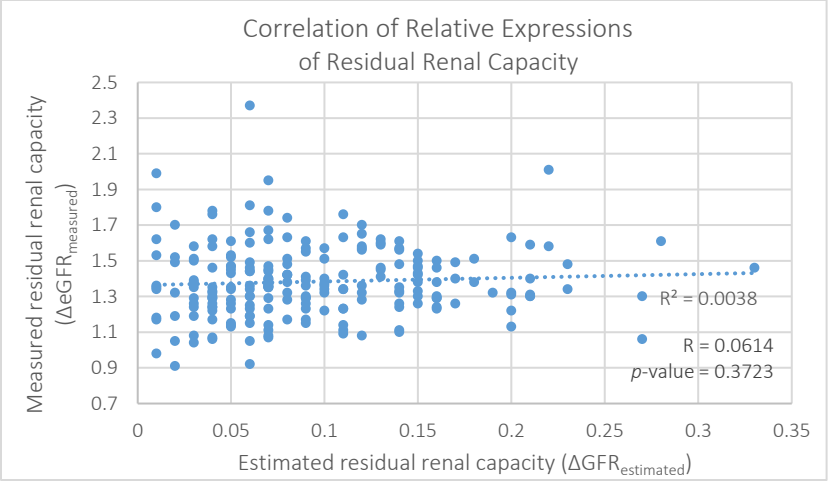

Figure A.6. Correlation of estimated residual renal capacity (Eq A.6) and eGFR-based measured residual renal capacity (Eq A.4) both in relative expressions.

Appendix B
Other Phenotypic Data

Categorical Characteristics

Table B.1 presents the summary of the collected categorical data. It is shown that the very small fraction of the population has notable clinical conditions (e.g. CVA, myocardial infarction, etc.) and also only small fraction (less than 6%) have chronic use of NSAID or ACE inhibitor (which can affect kidney function measurement). Furthermore this supports the healthy population hypothesis.

Living Donor (1993-2007) Characteristics			
		Frequency	Percent (%)
Gender			
	Male	126	44.7
	Female	156	55.3
Use of ACE inhibitor before nephrectomy			
	No	248	87.9
	Yes	10	3.5
	Missing	24	8.5
Chronic use of NSAID before nephrectomy			
	No	243	86.2
	Yes	15	5.3
	Missing	24	8.5
Smoking habit before nephrectomy			
	Never	81	28.7
	Former	49	17.4
	Current	96	34.0
	Missing	56	19.9

Living Donor (1993-2007) Characteristics			
		Frequency	Percent (%)
Alcohol consumption before nephrectomy			
	Never	39	13.8
	Former	1	.4
	Current	128	45.4
	Missing	114	40.4
Hypertension before nephrectomy			
	No	269	95.4
	Yes	13	4.6
CVA before nephrectomy			
	No	281	99.6
	Yes	1	0.4
Myocardial infarction before nephrectomy			
	No	280	99.3
	Yes	2	.7

Table B.1. Summary of categorical characteristics

Other Renal Function Parameters

Figure B.1 below presents the other collected renal function parameters, in box plots. As can be seen, a number of top outliers are shown. However, none of these points showed peculiarity in their data, and consistent information was found across correspondence. Thus these points were left in.

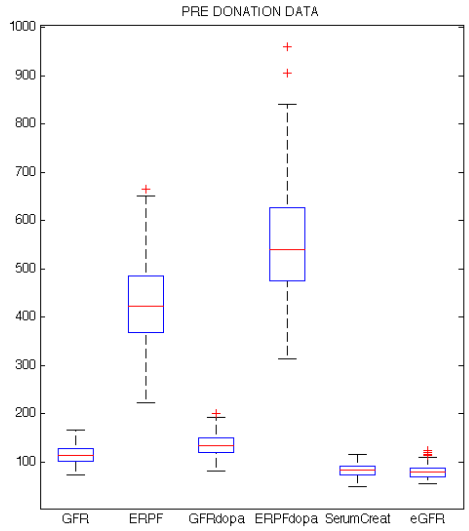

Figure B.1. Pre-donation kidney parameter data, filtered and cleaned. Abbreviations: GFR, glomerular filtration rate at rest; ERPF, estimated renal perfusion rate at rest; GFRdopa, glomerular filtration rate stimulated with low-dose dopamine infusion; ERPFdopa, estimated renal perfusion rate stimulated with low-dose dopamine infusion; SerumCreat, serum creatinine level; eGFR estimated glomerular filtration rate based on MDRD formula.

Stratifications

In analysing the gathered phenotype data, a stratification analysis was done. The population was split into 4 groups (or *quartiles*) based on their calculated $\Delta GFR_{measured}$. The people with top 25% $\Delta GFR_{measured}$ were put into the first quartile (Q1), $\Delta GFR_{measured}$ between 26%-50% into Q2, $\Delta GFR_{measured}$ between 51%-75% into Q3, and the rest into Q4.

Figure B.2 until B.5 plot the different characteristics per each quartile. These plot offer an overview of which characteristics are associated with $\Delta GFR_{measured}$. In Figure B.2, age and systolic blood pressure showed some association with the $\Delta GFR_{measured}$. These two characteristics were then checked for correlation in a scatter plots. Correlation with age is shown in Figure 6.4. Figure B.6 shows the systolic blood pressure correlation, and as a check, correlation with BMI was also checked and plotted in Figure B.7.

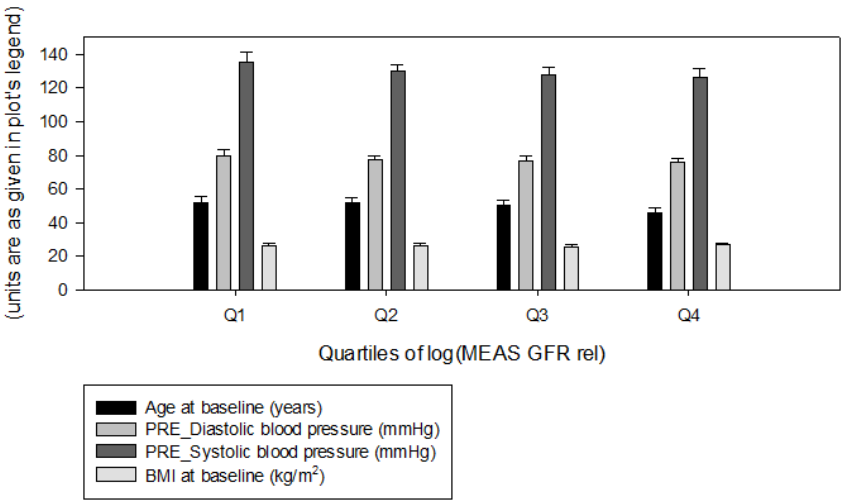

Figure B.2. Bar plots of each quartile's pre-nephrectomy age, diastolic and systolic blood pressure, and BMI. Bar plots present the mean of the characteristic with error bars showing 95% CI.

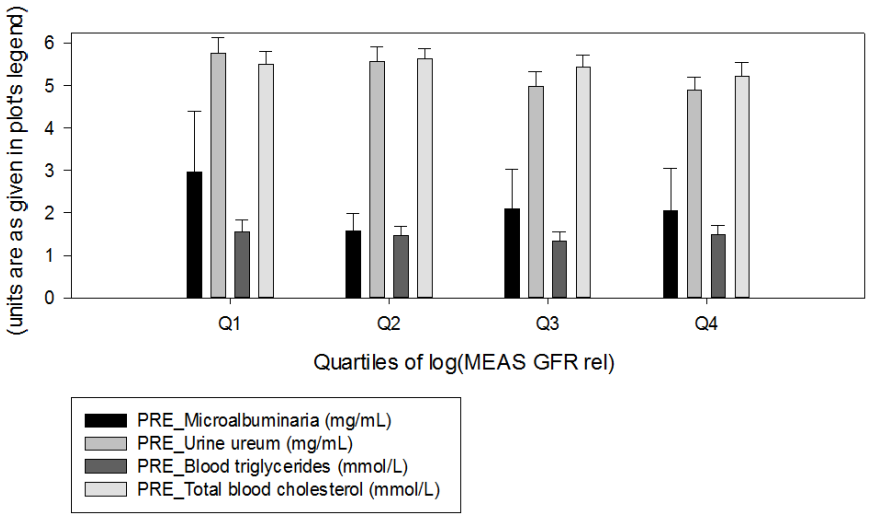

Figure B.3. Bar plots of each quartile's pre-nephrectomy microalbuminaria, urine ureum, blood triglycerides, and total blood cholesterol level. Bar plots present the mean of the characteristic with error bars showing 95% CI.

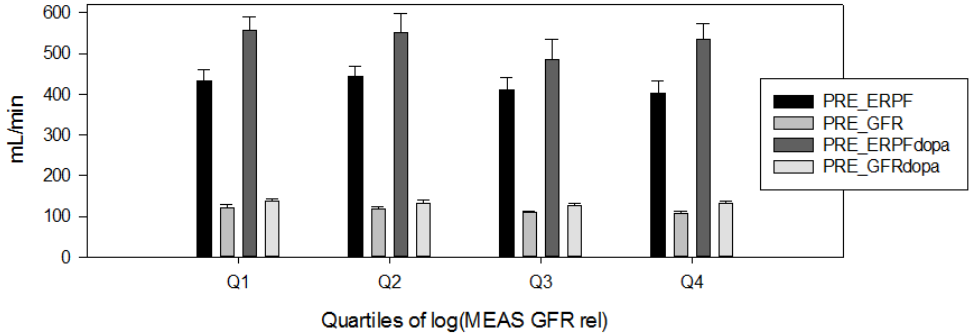

Figure B.4. Bar plots of each quartile's pre-nephrectomy ERPF and GFR at both base and dopamine-stimulated level. Bar plots present the mean of the characteristic with error bars showing 95% CI.

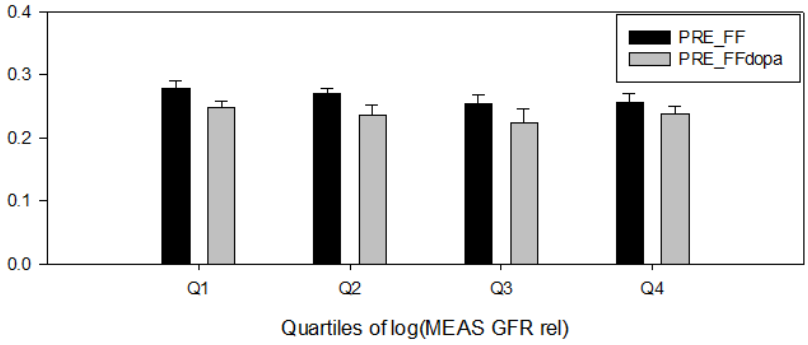

Figure B.5. Bar plots of each quartile's pre-nephrectomy filtration fraction at base and dopamine-stimulated level. Bar plots present the mean of the characteristic with error bars showing 95% CI.

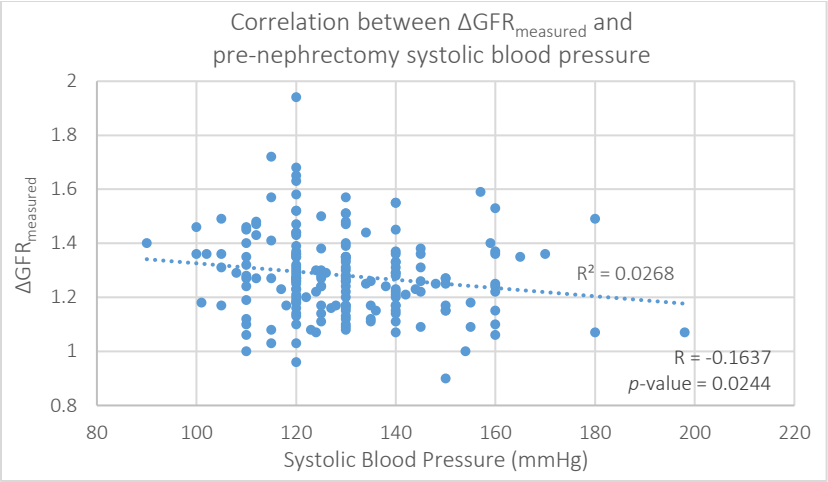

Figure B.6. Correlation between measured residual renal capacity and pre-nephrectomy systolic blood pressure. The two have a significant moderate negative correlation.

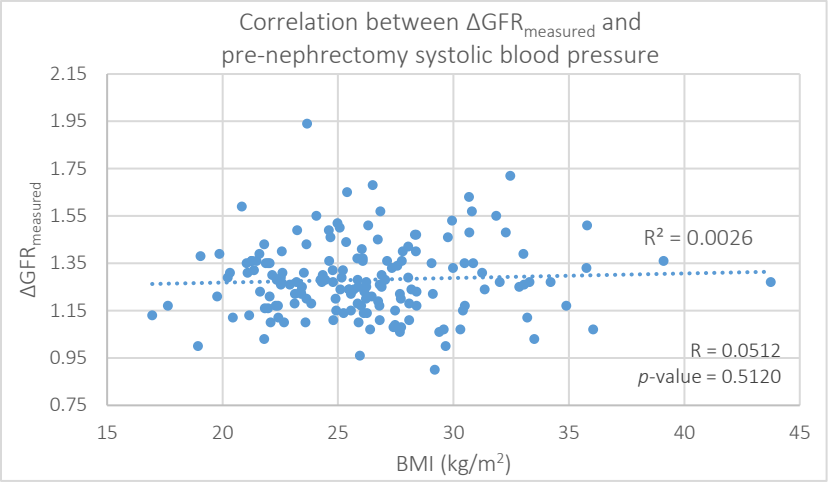

Figure B.7. Correlation between measured residual renal capacity and pre-nephrectomy BMI. The two do not have a significant correlation.

Appendix C
Other GWAS Results

GWAS of $\Delta GFR_{measured}$ with additional covariate of pre-nephrectomy systolic blood pressure

An additional GWAS analysis was done with an added covariate of systolic blood pressure. Systolic blood pressure is not a common covariate used in kidney-function association studies, however, the parameter showed to have a significant negative correlation with the $\Delta GFR_{measured}$. The correlation was not too strong and there was a chance that the collected data might have been biased by the use of anti-hypertensives. Nonetheless, as a check, similar GWAS associating the SNPs with $\Delta GFR_{measured}$ was done with age, gender, and systolic blood pressure covariates.

The plotted results in Figure C.1 shows similar findings with the presented results in the body of the thesis, with three loci found to be significant beyond the genome-wide significance threshold, at chromosome two, six, and ten.

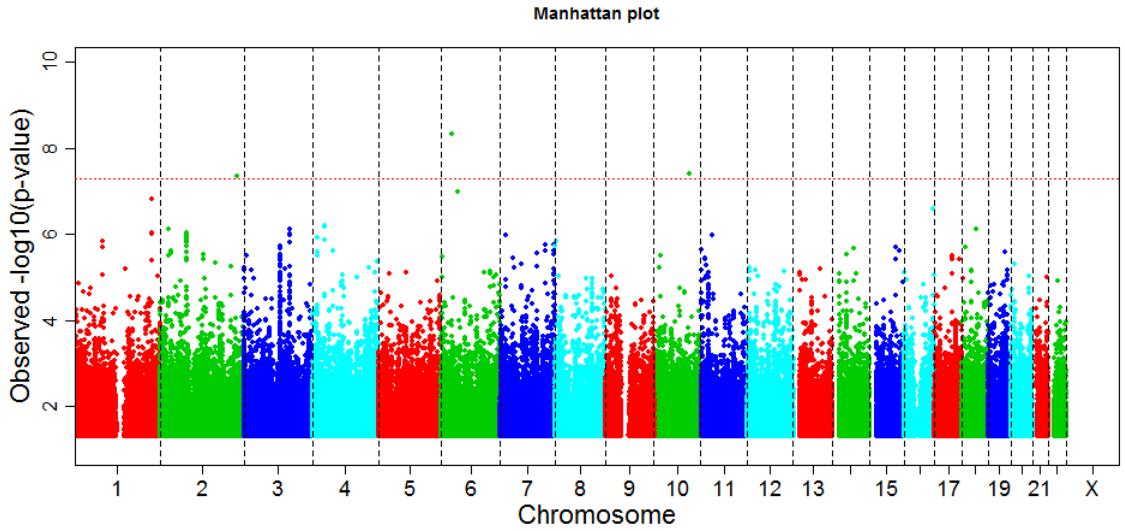

Figure C.1. Manhattan plot of GWAS results of the residual renal capacity, corrected for age, gender, and systolic blood pressure. Each point represents a tested SNP, displayed by chromosomal position (x-axis). Y-axis shows –log(p-value) for each SNP. Significantly associated SNPs are shown on chromosome two, six, and nine.

GWAS with eGFR-based measured residual renal capacity

In the course of the project, a GWAS was also performed with eGFR-based measured residual renal capacity (in relative expression, thus as noted in Eq A.4). Figure C.2 shows the results with rare alleles filtered at MAF >= 2%. The Manhattan plot shows a rather 'noisy' signal. QQ plot of allele frequency shown in Figure C.3 shows an inflation when all data (including rare alleles) are plotted, this implies an insufficient filtering of the rare alleles. Therefore further filtering was done, filtering out rare alleles of MAF < 10%, leaving only 64% of the SNPs. The result is then shown in Figure C.4, showing no significant association at the genome-wide significance threshold.

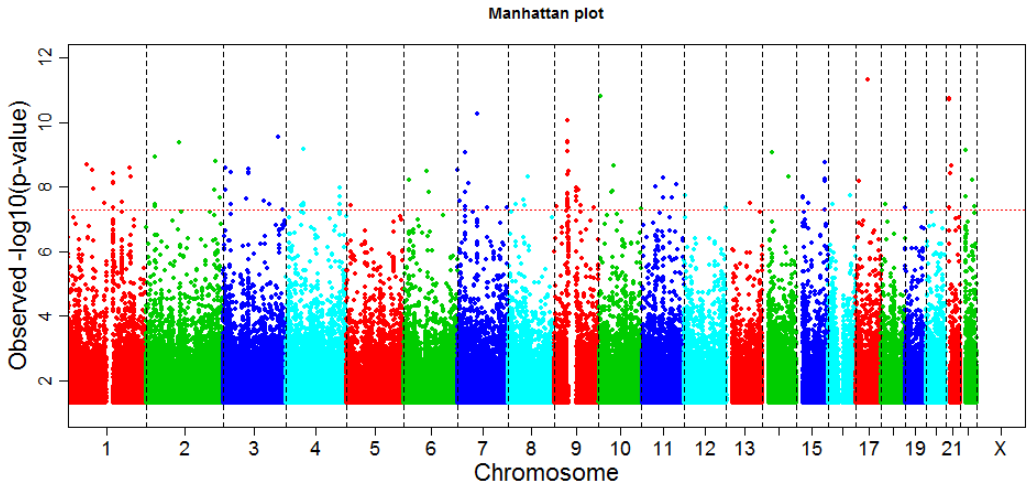

*Figure C.2. Manhattan plot of GWAS results of the residual renal capacity, corrected for age and gender. Filtered at MAF >= 2%. Each point represents a tested SNP, displayed by chromosomal position (x-axis). Y-axis shows –log(*p*-value) for each SNP.*

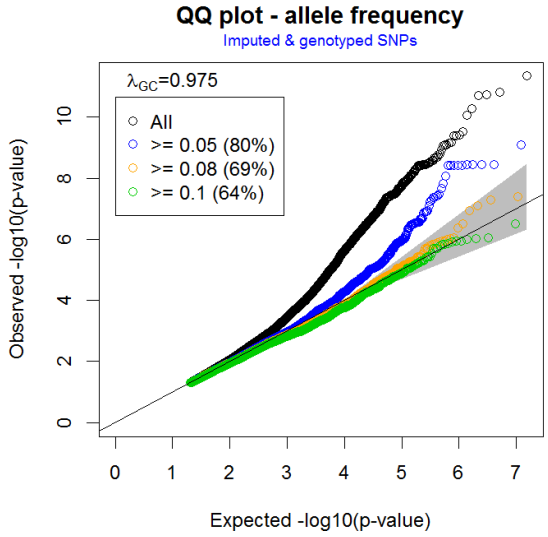

Figure C.3. Quality control plots. The QQ plots suggest that rare alleles need to be filtered at MAF >= 10%.

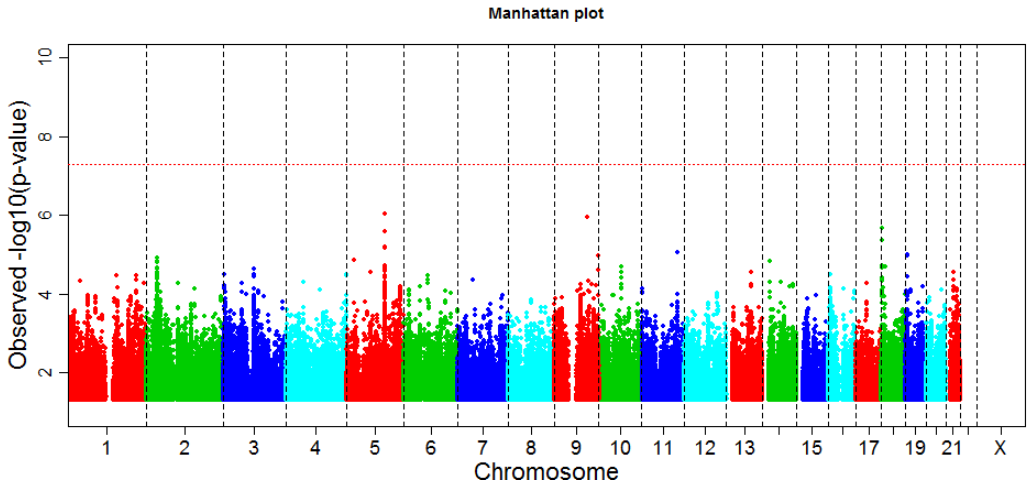

*Figure C.4. Manhattan plot of GWAS results of the residual renal capacity, corrected for age and gender. Filtered at MAF >= 10%. Each point represents a tested SNP, displayed by chromosomal position (x-axis). Y-axis shows –log(*p*-value) for each SNP.*