Identifying DNA Methylation Patterns as Novel Urinary Biomarkers for Kidney Function

By

RANYA HASSO, HBSc

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

for the degree

Master of Science

McMaster University

Master of Science 2014   McMaster University

(Medical Sciences – Cancer and Genetics)   Hamilton, Ontario

TITLE: Identifying DNA Methylation Patterns as Novel Urinary Biomarkers for Kidney Function

AUTHOR: Ranya Hasso, HBSc

SUPERVISOR: Dr. Guillaume Paré

NUMBER OF PAGES: 82

ABSTRACT

Chronic kidney disease (CKD) is a major public health concern, characterized by an irreversible reduction in renal function. Currently, creatinine-based GFR estimation is predominantly used clinically to characterize CKD. However, this method is known to be an insensitive test for early losses of kidney function. Since patient prognosis relies heavily on slowing further decline of kidney function, uncovering novel biomarkers for kidney function, in conjunction with eGFR, will help improve patient outcome. Epigenetic-based biomarkers have been identified in numerous cancers, as DNA methylation changes alter cellular function. Thus, the objective of this study is to determine novel DNA methylation patterns reflecting altered kidney function. Five healthy participants that have undergone a nephrectomy have donated urine samples before and after their surgery, and global DNA methylation changes were analyzed through the 450K HumanMethylation microarray. Site- and region-level analyses were conducted to determine significant differentially methylated probes post-nephrectomy. The differential associations observed post-nephrectomy are statistically significant in both the site-level and regional analyses. Nineteen significant candidate probes have been systematically selected for validation, based on involvement in kidney function and consistent direction of methylation. Pyrosequencing assays have also been successfully designed and tested with control DNA, however replication of the microarray findings in participant DNA was unsuccessful. The inability to validate these candidate probes may be attributed to many influencing factors, and with this in mind, uncovering novel methylation patterns is still a promising biomarker for evaluating kidney function.

ACKNOWLEDGMENTS

Table of Contents

LIST OF FIGURES

Chapter 5: External Validation of Candidate Probes

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Chronic Kidney Disease

Chronic kidney disease (CKD) is an economic burden on both the national and

worldwide scale, as renal replacement therapy was estimated to cost the Canadian

healthcare system an annual $60,000 per patient in 2007[1–3]. This disease is characterized

by an irreversible and gradual reduction in renal function and a decrease in the number of

functional nephrons[4]. According to the 2012 Clinical Practice Guideline for the

Evaluation and Management of Chronic Kidney Disease issued by the Kidney Disease:

Improving Global Outcome (KDIGO) work group, CKD is defined as abnormalities of

kidney structure or function impacting health[5]. These guidelines characterize CKD by a

sustained glomerular filtration rate (GFR) < 60 mL/min/1.73 m^2 and persistence of

albuminuria[6,7]. A classification system of relative CKD risk, based on GFR and

albuminuria categories, is used clinically to predict prognosis of CKD (**Figure 1.1**). The

1999-2004 NHANES (National Health and Nutrition Examination Survey) suggests a

substantial 13.1% of the US adult population is affected by CKD with a GFR of >90-15

mL/min/1.73 m^2, as compared to 10% in the 1988-1994 survey[8].

Figure 1.1| Prognosis of CKD by GFR and Albuminuria Categories

CKD is established by presence of kidney damage, assessed by the level of albuminuria, and the level of renal filtration, through GFR estimation. The colour blocks indicate the risk category of kidney disease. Green indicates low risk (if no other markers of kidney disease is present, then no CKD), yellow indicates moderate risk, orange indicates high risk, red indicates very high risk. This figure is taken from the KDIGO Work Group (2013).

The prevalence of this disease is rapidly increasing, and the most prevalent outcome of CKD is kidney failure requiring dialysis or transplantation. In addition, individuals with diabetes mellitus, dyslipidemia, hypertension and advanced age are at a greater risk for CKD development and progression[9,10] End-stage renal disease (ESRD) is a final outcome of CKD, and has nearly doubled in the elderly population within the last 25 years[11]. In addition, about 1.3 million to 2.9 million Canadians are estimated to have CKD, and roughly 10% of the national population is affected by stage 1 or 2 of the disease. Since CKD is asymptomatic within these early stages, individuals that do not have health conditions that would prompt screening for kidney dysfunction, such as diabetes or hypertension, may be easily missed based on current practices[9]. Hence, a larger proportion of people are likely affected by CKD than current estimations. Given

the tremendous implications of CKD on patient outcome and healthcare economy, it is important to advance current CKD knowledge and explore other means of detection to optimize disease management.

1.2 Chronic Kidney Disease and Cardiovascular Disease

In 2008, the Heart and Stroke Foundation estimated that 29% of deaths in Canada are a result of cardiovascular disease, which costs the Canadian healthcare system $20.9 billion per year[12]. Cardiovascular disease (CVD) is commonly associated with CKD, and in 2003, the American Heart Association defined CKD as the highest risk factor for subsequent CVD [4]. Patients with CKD are at an increased risk of developing cardiovascular-related outcomes, such as cardiovascular mortality, stroke, and congestive heart failure. In addition, the proportion of CKD-affected individuals that will develop CVD-related events greatly increases over 10 years compared with an absence of CKD, as indicated in the survival analysis conducted by Weiner, *et al.*[13] (**Figure 1.2**). Thus, CKD is shown to independently increase adverse cardiovascular-related outcomes in individuals without CVD[13,14]. However, CKD and CVD are now emerging as a part of a dynamic relationship, and not just separate diseases. It is previously known that both CVD- and CKD-related outcomes can be a result of traditional risk factors, such as diabetes mellitus, hypertension, and family history[15]. While more recently, CKD-related outcomes, such as inflammation, oxidative stress, aggressive vascular calcification, prothrombotic milieu, and anemia, are perceived as novel risk factors for further development of CVD (**Figure 1.3**). Given this dynamic and injurious relationship, uncovering novel markers for kidney function, as a preventative measure for CKD, can also improve cardiovascular-related events.
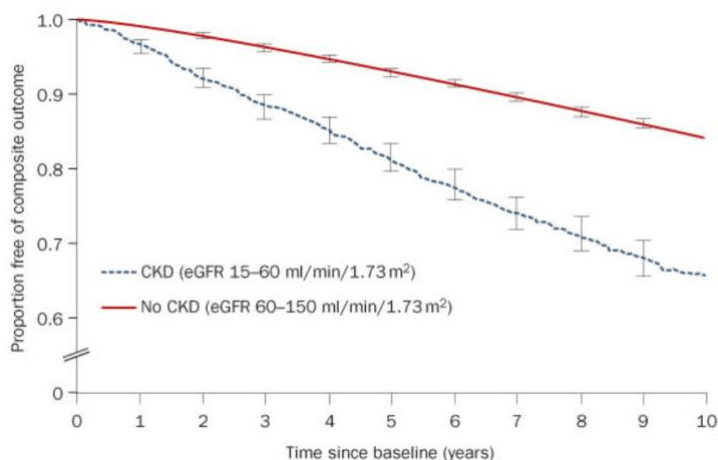
Figure 1.2 | Survival Estimate (Kaplan-Meier) following Cardiovascular Outcomes and Presence of CKD

Presence of CKD defined as eGFR<60 ml/min/1.73m^2 and the composite outcomes refer to the cardiovascular-related events; including myocardial infarction, fatal coronary heart disease, stroke, and all-cause mortality. Figure taken from van der Zee, S. et al. Nature Reviews Cardiology. 6, 580–9 (2009)[15].
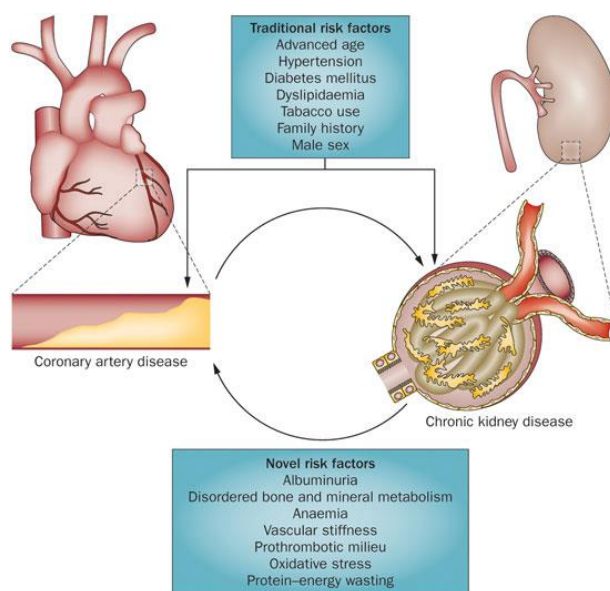
Figure 1.3 | Relationship between CKD and Coronary Heart Disease

Traditional risk factors, such as age, sex and hypertension can affect the progression of both coronary heart disease and chronic kidney disease. However, outcomes of chronic kidney disease, such as anemia, oxidative stress and albuminuria can increase this risk of further developing coronary heart disease. Figure taken from van der Zee, S. et al. Nature Reviews Cardiology. 6, 580–9 (2009).

1. 3 Chronic Kidney Disease Pathogenesis and Diagnosis

Initial nephron injury marks the beginning of CKD pathogenesis and hinders early diagnosis, through the compensatory responses of the remaining nephrons[16]. Three main phases of CKD pathogenesis have been described: 1) nephron injury initially occurs, 2) then stimulating repair mechanisms, which may be misdirected and maladapted to impede nephron function[16,17]. 3) The remnant functional nephrons compensate for the lost nephrons through changes in activity, which may cause further damage[16,17]. CKD pathogenesis ultimately leads to a vicious cycle of nephron death, which can be detected through alterations in GFR. Conversely, repair mechanisms following initial nephron injury leads to functional overload of the existing nephrons, and thus GFR elevation[17]. Therefore, GFR estimation is a misleading marker for early CKD, and novel biomarkers reflecting initial repair mechanisms or early compensation can be a more precise diagnostic method. Since CKD is treatment involves reducing progression of kidney damage[8], developing a marker for earlier detection of CKD may improve both patient outcome and may help relieve the global burden of this disease.

Though GFR estimation is an important tool in CKD diagnosis and prognosis, these approaches are not always well-suited for the assessment of initial decreases in GFR and early stages of CKD. According to the National Kidney Foundation, CKD is diagnosed by a GFR of less than 60 mL/min/1.73m^2 [18,19], and is expected to decrease uniformly in response to the decrease in functional nephrons. Injecting inulin, an exogenous substance, accurately measures GFR as it is present in plasma at stable concentrations, is freely filtered, and not secreted nor reabsorbed[20]. Although inulin is the

5

"gold-standard" in GFR measurement, it is highly invasive and rarely used clinically. Creatinine, on the other hand, is a metabolite of creatine metabolism in muscle mass, and is an endogenous marker currently used to assess GFR[18]. Creatinine is filtered through the glomerulus and is not reabsorbed, but is secreted at the proximal tubules. Many creatinine-based GFR predicting equations exist, however the MDRD equation is most frequently used in clinical settings[21]. This equation incorporates serum creatinine levels, while also adjusting for age, sex and race. While the MDRD is a viable and critical means of estimating GFR, weaknesses of the equation still exist and hinder the ability to detect early losses of kidney function[11,20,22].

Creatinine clearance usually mirrors GFR; however, this method is an insensitive approach for early losses of kidney function, and thus limits treatment options[8,18]. Since creatinine is secreted at a constant rate in the tubules, a slightly lowered GFR during early stages of kidney failure causes a 30% higher GFR estimation[20]. A comparison of serum creatinine levels and estimated GFR via an exogenous marker, as in **Figure 1.4**, shows a GFR between 15-29 mL/min/1.73m^2 is represented well, as creatinine increases dramatically in both males and females[23]. Since creatinine is also secreted within the proximal tubules, CKD stages 1-3 (characterized by GFR >30 mL/min/1.73m^2) is not well-represented within the serum creatinine levels of both sexes, as minimal alterations in creatinine levels is established. Unfortunately, the available treatment options are often limited at the time CKD is diagnosed through this GFR measurement tool. Therefore, uncovering novel urinary biomarkers for early losses of kidney function, in addition to eGFR, can help improve diagnostics and patient outcome.
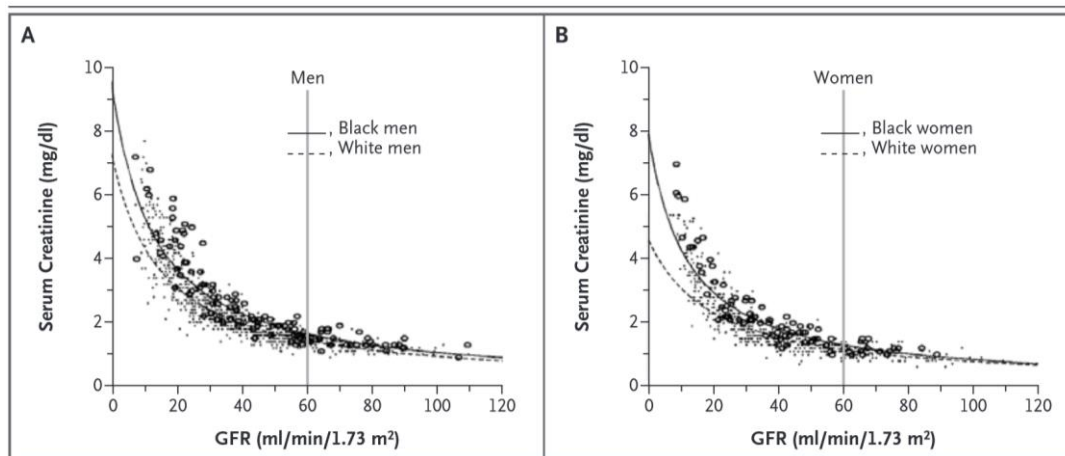
<u>**Figure 1.4**</u> | Relationship of Serum Creatinine Levels versus Measured GFR in Males and Females

Each point on the graph represents the creatinine concentration and measured GFR of an individual within the Modification of Diet and Renal Disease (MDRD) Study. GFR was measured via renal clearance of $_{125}$I-iothalamate, while serum creatinine was measured through the use of a kinetic alkaline picrate assay. The data is separated into panels by gender, while ethnicity is shown, via straight and dashed lines, for each panel. Figure taken from Stevens, L. A. et al. N Engl J Med. 354, 2473-83 (2006).

1.4 Urinary DNA

DNA is found in urine and exists as two basic types. The first category is transrenal DNA, derived from apoptotic cells in the blood releasing highly fragmented DNA sequences small enough to pass the kidney barrier[24]. Botezatu et al. studied urinary DNA of females who have received transfusions from male blood[25]. This group has identified male-specific sequences in the urinary DNA of these females, even 10-14 days after the initial transfusion. These sequences have thus traveled from circulation through the kidney barrier and were found to be short, averaging about 150 bp[25]. The second category of urinary DNA is derived from apoptotic and exfoliated cells of the kidney, bladder and urinary tract, characterized by larger fragments than transrenal DNA ($\geq$1

kb)[24]. Since cancerous cells are known to have distinct genetic variations, DNA and

RNA-based biomarkers have been discovered for a variety of urological cancers[26]. Thus,

differences with larger urinary DNA of renal cells, versus transrenal DNA, can potentially

be used to uncover novel biomarkers after a loss of kidney function.

1.5 DNA Methylation

DNA methylation is a structural, or epigenetic, modification of the genome in

which methyl groups are added to a cytosine paired with guanine, by the means of a

phosphate, in a CpG dinucleotide[27]. DNA methyltransferases convert these cytosine

residues to 5-methylcytosine (5mC), and thus methylating the CpG site[28]. The role of

DNA methylation is to maintain DNA in an inactive state, as the methyl group of 5mC

blocks transcription factors from binding, thus interfering with gene expression[28]. Classes

of methylated DNA-binding proteins exist that also repress gene transcription by

interfering with transcription factors[28]. CpG sites are predominantly found in CG-rich

repeats, spanning about 1 kb, which are referred to as CpG islands and are capable of

altering transcription[29]. These CpG islands are often associated with the transcription start

sites of genes, but are also found within the gene body and promoter regions, and can

directly repress gene expression[29]. As a result, the silencing property of these epigenetic

modifications allows for specific functioning and behavior of cells within a tissue as only

required genes are transcribed, rather than the entire genome[30].

Abnormalities in DNA methylation patterns have the capacity to directly silence

genes thereby causing cell dysfunction, thus DNA methylation has been the major focus

of epigenetic studies[28]. DNA methylation changes have been recently shown as novel

biomarkers for many diseases, especially with respect to carcinogenesis[31,32]. Several studies have newly described methylation patterns specific to bladder cancer, through urinary DNA, for both diagnostic and prognostic purposes[33,34,35]. In particular, a 2007 study profiled the methylation status of 59 tumor-associated genes in both tumor biopsies and urinary DNA of 64 bladder cancer patients and 23 age-matched controls. Methylation-specific PCR was used to analyze changes in methylation patterns in these genes. Cancer-specific hypermethylation was reported in 15 genes within both sample types ($P < 0.05$)[36]. Thus, methylation profiling was described as a sensitive method of bladder cancer detection. For that reason, novel epigenetic biomarkers can also be identified in urinary DNA from renal cells, which has the potential be an important diagnostic and prognostic measure of kidney disease.

1.6 DNA Methylation Analysis

An abundance of DNA methylation analysis technologies have been made available in recent years, such as microarray-based methods and sequencing-based applications[37]. Both technologies are key aspects in the experimental design of this study and will be discussed further.

1.61 Methylation Microarray

Microarray technology allows for hybridization of nucleic acid sample to a larger number of probes attached on the chip and has many applications, including DNA methylation analysis[38]. A global-level of DNA methylation can be examined through the Illumina Infinium HD Assay for methylation. This chip uses two different probe chemistries to profile over 450, 000 CpG loci per sample[39]. DNA samples are initially

treated with sodium bisulfite, which converts any unmethylated cytosine base pairs within the DNA to uracil, while methylated a cytosine is protected and remain unchanged. The bisulfite-converted DNA samples are then prepared for whole-genome amplification (**Figure 1.4 A**) and undergoes fragmentation, through controlled enzymatic processes. These fragments are dispensed and incubated to hybridize onto the methylation chip. Hybridized fragments of DNA then anneal to the locus-specific 50mers that are linked to one of over 450 000 bead types.

As previously mentioned, two bead chemistries are employed within the Illumina 450K BeadChip; the Infinium I and II assays. Infinium I assays contain two bead types corresponding to each CpG site, consisting of one bead type corresponding to methylated sites and another for unmethylated sites. Infinium II assays contain one bead type and the single base-extension step, after hybridization, determines the methylation state. Any unhybridized DNA fragments are washed away, and the chip then undergoes a single base-pair extension and staining steps. Detectable labels are then incorporated on the BeadChip through the single-base extensions of the oligos and determined methylation levels of CpG sites (**Figure 1.5 B**). Finally, the BeadChip is read by the iScan system (**Figure 1.5 C**), and high-resolution images of the fluorophore-emitted light is recorded as a dataset of intensity and β-values for all profiled CpG loci. These β-values are generated Illumina system generates, based on the detected intensity levels of the probes, and represent the level of methylation of a CpG loci[40].
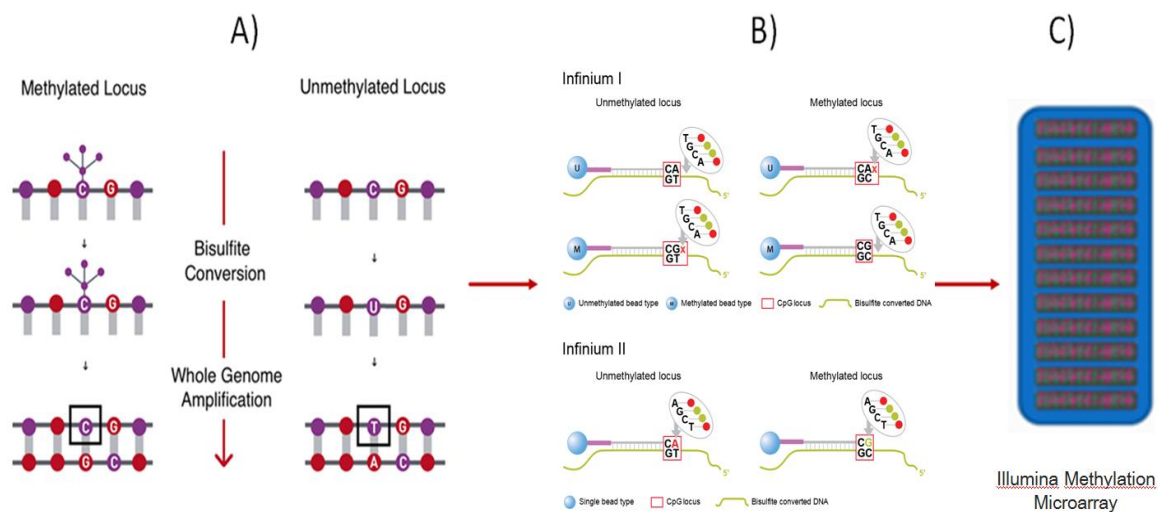
Figure 1.5 | Illumina 450K BeadChip Experimental Process

DNA sample is initially bisulfite-converted to convert any unmethylated cytosine base pair, while protecting methylated ones, and this converted DNA undergoes whole genome amplification. The DNA is then hybridized onto the chip and the two different bead types on the chip's surface allow for single-base pair extension. The chip is then scanned by the iScan system, and the single-base extensions are illuminated, thus giving off a readable methylation signal. Figure adapted from Illumina, Inc. (2014).

Site- and region-level analyses are two statistical methods to interpret data from methylation microarrays, and can be used in conjunction to shed light on epigenetic changes in health and disease. A site-level analysis is used to pinpoint individual probes within the methylation data that are differentially significant between two groups. A region-level analysis can identify entire genomic regions that are associated with the outcome, which is valuable as gene transcription is regulated by heavily methylated regions within the genome. The Illumina 450K BeadChip allows for the analysis of data based on a regional collection of probes, as well as the more traditional site-level analysis. Within the microarray, each methylation probe is assigned an associated gene name as well as regional information, depending on its position. There are eleven region types that

are illustrated by the schematic representation of the probe coverage of both the gene

regions (**Figure 1.6 A**) and CpG Island regions (**Figure 1.6 B**).
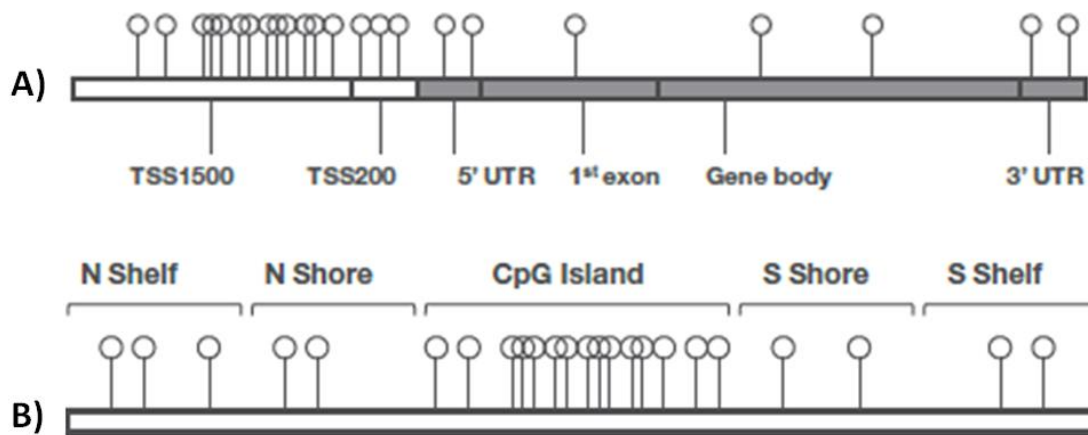
<u>**Figure 1.6** | Genomic Regions within Illumina 450K BeadChip</u>

The following figure is a schematic representation of the eleven types of genomic regions classified within the Illumina 450K BeadChip. A) shows the six regions within the gene region; 1500 bp from the Transcription Start Site (TSS1500), 200 bp from the Transcription Start Site (TSS200), 5' Untranslated Region (5UTR), 1st Exon region, the Gene Body, and 3' Untranslated Region (3UTR). B) shows the five types of promoter regions; CpG Island, North Shore (NShore, 2kb upstream of Island), North Shelf (NShelf, -2-4 kb from Island), South Shore (SShore, 2 kb downstream from Island), South Shelf (SShelf, 2-4 from Island). This figure is taken from the Infinium Human Methylation 450K BeadChip Datasheet (Illumina).

1.62 Methylation-Specific Pyrosequencing

Pyrosequencing is another tool to analyze DNA methylation, however this method

interrogates a small number of methylation sites at a very high level of precision[41].

Pyrosequencing is a sequencing-by-synthesis technique that monitors the quantity of

incorporated nucleotides in real-time through light signals[42]. These signals are released

by enzymatic conversions of the released phosphate during base-pair addition. Bisulfite-

treated DNA is used in this pyrosequencing application, and the degree of methylation is

measured by the proportion of C's and T's within the measured sequence, thus representing the proportion of methylated and unmethylated cytosine base-pairs[42]. Pyrosequencing was conducted in this study through the PyroMark Q24 Assay for methylation, and is used to validate the methylation statuses of specific probes, discovered in the methylation microarray analysis.

1.7 Rationale and Objectives

CKD is characterized by an irreversible reduction in renal function and is largely asymptomatic at early stages. Estimating GFR through creatinine clearance is an important means of diagnosis. However, eGFR overestimates kidney function during early CKD, leading to an underdiagnoses of the disease. Thus, uncovering novel methylation patterns as biomarkers of kidney function would enhance the current means of CKD detection, and can provide further insight to kidney injury pathways.

We hypothesize that urinary DNA, from apoptotic and exfoliated kidney cells, can be used to uncover novel methylation patterns as biomarkers of renal function for CKD detection. Methylation patterns of kidney function have been previously studied in urinary DNA, and hypermethylation of tumor suppressor genes, specific to kidney cancer, were identified between both tumor biopsies and urinary DNA[33,34]. In particular, a 2003 study identified six normally unmethylated tumor-suppressor genes in both tissue types[34]. In particular, hypermethylation of the *VHL* gene was observed exclusively in clear-cell renal carcinoma patients, which is appropriate as this type of kidney cancer is prevalent in

a large subset of Von Hipple Lindau patients[35]. Thus, methylation profiling of urinary DNA is described as a promising means of kidney disease detection.

Urinary DNA from nephrectomy patients, before and after their surgery, has been used as a model to study methylation patterns of kidney function. Individuals that have undergone a nephrectomy have lost half of their functional nephrons, and have therefore lost half of their normal renal function. However, the nephrons within the existing kidney are able to compensate for this loss, and the kidney donor is able to lead a normal life. This compensatory response in nephrectomy patients is similar to the compensatory mechanisms involved in preliminary phases of CKD pathogenesis. Thus, a nephrectomy model can be used to derive novel methylation patterns as markers for kidney injury and may further elucidate these mechanisms. Since characterization of methylation patterns in urinary DNA post-nephrectomy is completely novel, the outcome of these findings may lead to uncovering epigenetic biomarkers of early CKD, which would have tremendous diagnostic and prognostic implications.

Many benefits, as well as caveats, exist for the use of kidney donors as a model for kidney function. The analysis of genetic material from nephrectomy patients evaluates kidney function almost exclusively, as these healthy patients are absent of CKD-related complications, and thus assessing kidney function directly. Evaluating kidney function in CKD patients is difficult as other complications can act as confounding variables, and any genetic alterations may be associated with underlying diabetes, hypertension, or dialysis[43]. However, this nephrectomy model assesses kidney function from a different

perspective, as an acute response to loss of kidney function is observed in these patients, as opposed to chronic injury in a CKD model[16,17].

Overall, the main objective of this study is to determine novel urinary DNA methylation patterns representative of a loss of kidney mass post-nephrectomy, as a model for the loss of kidney functioning in CKD patients. This experimental design is completely novel and methylation patterns for kidney function is not yet established in urine, and could evaluate renal function and help shed light on kidney injury mechanisms.

CHAPTER 2: DIFFERENTIAL ANALYSIS OF METHYLTION MICROARRAY DATA

2.1 Materials and Methods

2.1.1 Urine Sample Collection

In this study, urine and blood samples have been collected from healthy kidney donors at St. Joseph's Healthcare undergoing a nephrectomy. Urine samples were collected from a total of twelve healthy kidney donors. Since the Illumina methylation microarray requires a minimum input of 500 ng of DNA, only five participants met this criteria and were included in the study. Also, DNA from two of these participants were concentrated and pooled due to small amounts of DNA in both participants. In these pooled samples, equal amount of DNA were used from both participants.

To ensure adequate volumes of sample were collected to continue the study, amendments to the sample collection protocol were made after the first microarray was conducted. Previously, urine samples were collected at various time points: 3 months prior to the operation (Pre-Op), the day of the surgery, one day after surgery, two days after, three days after, and three months after surgery (3 Month). The study collection manual had been modified to include the collection of only Pre-Op and 3M donations. DNA quantity was found to be highly variable in urine, thus further alterations were made to the collection protocol to include 400 mL of sample, as compared to 30 mL. All participant samples collected were stored at -80 °C until required for DNA extraction.

2.1.2 DNA Extraction and Quantification

The Norgen Urine DNA Isolation Maxi Kit (Slurry Format) was used to isolate DNA from frozen urine samples. DNA extraction from urine was initially tested on control samples that were stored at room temperature, -20 °C, and -80 °C, to examine the effect of storage conditions on DNA quantity. Urine stored at -20 and -80 °C showed similar quantity of isolated DNA, while urine stored at room temperature showed increased DNA quantity due to bacterial growth. Participant DNA was then isolated from about 25 mL of urine, as outlined in the Norgen protocol, and DNA extracts were stored at -20°C until later use.

The Quant-iT Picogreen Assay by Invitrogen was used to quantify the amount of DNA within each extracted DNA sample. In this assay, the absorbance of the fluorescently stained dsDNA is measured by the fluorescence microplate reader against the linear detection range of the standard fluorometer. This quantification step is important as the downstream methylation microarray requires 500-1000 ng of DNA per sample.

2.1.3 Analysis of DNA Size

The size of DNA from patient urine samples was be determined by the Agilent High Sensitivity DNA assay. Simple preprocessing steps, outlined in the Agilent protocol, were followed for the input of 1 µL of DNA sample into the assay. This step is important to ensure the quality of the DNA, and that DNA samples of 1 kb or higher are only included in the microarray.

2.1.4 DNA Concentration

According to the Illumina Protocol, DNA samples are to contain 500-1000 ng of

DNA within a volume of 45 µL. Concentration of DNA samples to the needed volume

was required as an additional preprocessing step. The speed-vac method for DNA

concentration was used to decrease sample volume by evaporating water within the

sample.

2.1.5 Methylation Microarray

DNA methylation analysis was performed using the Illumina Infinium HD Assay for

methylation, using one or two bead types to profile over 450 000 CpG loci per sample

[16]. Twelve DNA samples were profiled in the first microarray (**Table 1**, **Appendix**),

performed in August 2011. Within these twelve DNA samples, one Pre-Op and one 3M

sample were pooled together from participants 06 and 07. Pooling equal amounts of these

two participant's DNA samples was performed to achieve sufficient amount of DNA

required for the microarray. This chip also includes one positive control of known high

quality DNA, as well as a duplicated set as an indication of assay quality. In total, five

sets of extracted pre- and post-nephrectomy DNA samples were used, as well as a Day-3

sample from Patient-08. The second microarray was performed in December of 2012

(**Table 1**, **Appendix**), which includes one Pre-Op and 3M sample set, their duplicates and

one Pre-Op sample.

2.1.6 Statistical Analysis

Initial quality control of the raw intensity and β-values data has been conducted

through GenomeStudio (Illumina) for all samples. The number of detected CpGs (p >

0.01), sample clustering, as well as bisulfite conversion and other sample dependent and

independent controls available in GenomeStudio were assessed for quality assurance.

Raw intensity signals and average β-values of all Pre-Op and 3M samples have been

exported from GenomeStudio and read by Illumina Methylation Analyzer (IMA), an

R/Bioconductor package[44]. Loci with missing β-value, which describes the level of

methylation, and median detection P-value >0.05 have been filtered and peak correction

is performed, as it is shown to improve Infinium II data quality[45]. Normalization of the

raw methylation data was not conducted as the commonly-used quantile normalization

has been shown to be an insufficient strategy to eliminate technical variation for

methylation microarray data[46]. Normalization methods for this type of data is still an area

of active investigation[46].

Differential methylation analysis between the two sample groups has been conducted

through R/Bioconductor to identify differentially methylated probes after a loss of a

kidney[47]. Two types of differential methylation analyses have been employed in this

project; a site-level and regional analysis. The site-level analysis treats each methylation

probe as independent variables, and the pooled Student's t-test was used for inference of

differentially methylated probes in the post-nephrectomy group[47]. A mentioned

previously, Illumina groups probes of the microarray within categories of regions, such as

Island, Gene Body, and 5' Untranslated regions. A differential analysis was conducted for

each of the eleven region-types. Initially, loci was collected within a specific region and derive the median index of overall region methylation value[44,47]. The index of overall methylation of all the genomic regions within the region-type were grouped, and a t-test was also implemented to identify differentially methylated regions.

2.2 RESULTS

2.2.1 Determining the Size Urinary DNA

Agilent High Sensitivity DNA assays were conducted to identify the size of DNA extracted from the collected urine samples, and a comparative assay was constructed via the Agilent Bioanalyzer. The comparative gel electrophoresis (**Figure 2.1A**) demonstrates the extracted participant DNA were comprised of larger sized fragments, as these DNA samples take a longer time to travel through the gel, as compared to the assay's ladder. The Bioanalyzer-generated electropherograms (**Figure 2.1B**) also demonstrated the size of the DNA extractions. This assay also includes gel markers in each sample for quality control at 35 bp (45 sec) and 7000 bp (115 sec) as shown in all samples of **Figure 2.1A** in green and purple. This experiment is very important to the overall study as larger-sized DNA is required to test methylation status of renal cells, versus trans-renal DNA. Since this experiment showed both DNA elutions contain higher-sized DNA, the two elutions were used in both microarrays performed.
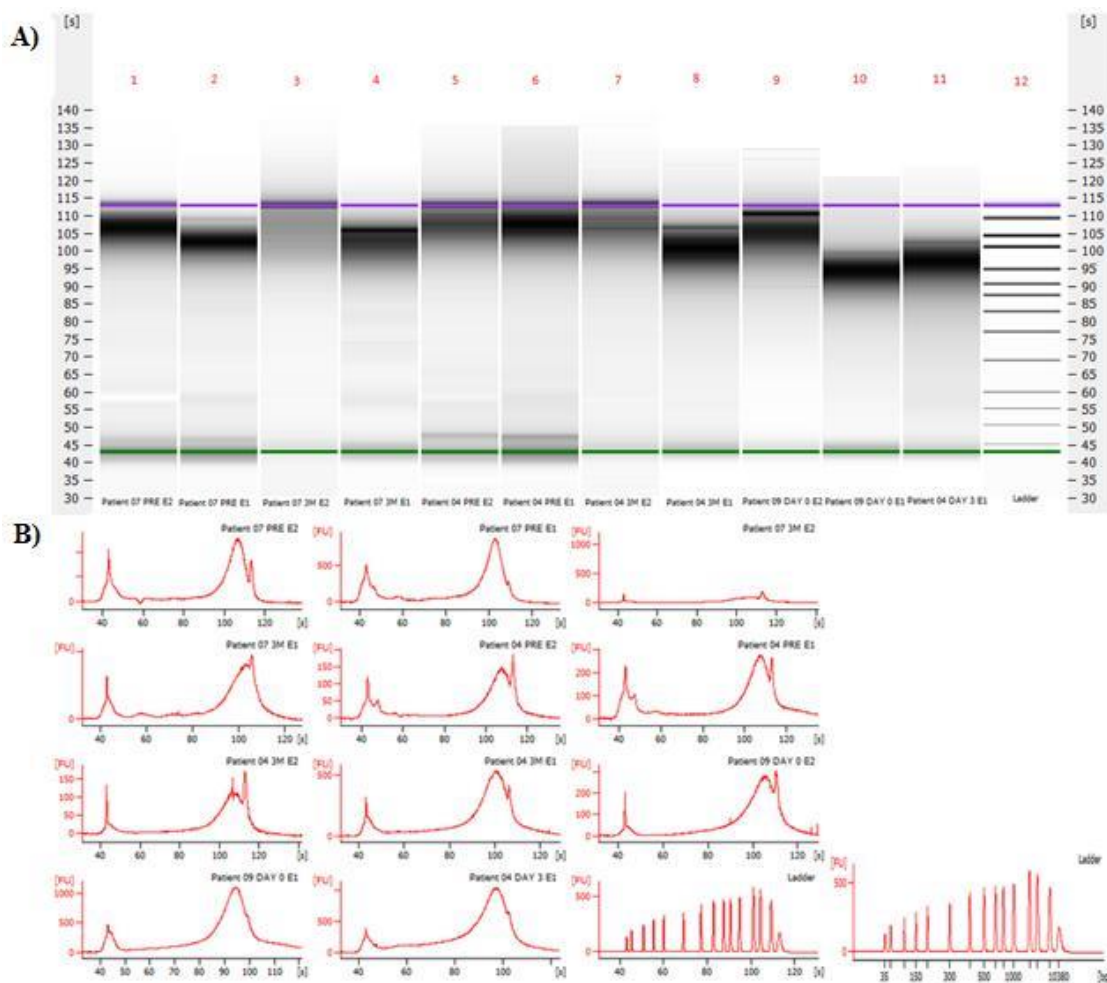
Figure 2.1 | High Sensitivity DNA Assay Gel Electrophoresis and Electropherograms

Comparative assay of first and second elution DNA extracts from patient samples. **A)** Gel electrophoresis 1) 07 Pre E2, 2) 07 Pre E1, 3) 07 3M E2, 4) 07 3M E1, 5) 04 PreE2, 6) 04 Pre E1, 7) 04 3M E2, 8) 04 3M E1, 9) 09 Day 0 E2, 10) 09 Day 0 E1, 11) 04 Day 3 E1, 12) Ladder. **B)** Associated electropherogram traces per well. Two ladders are shown, the first in seconds and the second in basepairs to give a numerical value for the size.

2.2.2 QUALITY CONTROL

Microarray performance is examined by directly testing the average β-Value for each detected CpG site against that of the duplicate, and generating an R^2 value as a function of similarity. A scatter plot (**Figure 1 A - Appendix**) of the average beta value,

representing the level of methylation, of each detected CpG loci for sample 09 Day-0

versus its duplicate. The two are very closely related as the vast majority of the beta

values per CpG site lie close to the trend line and an R^2 value of 0.9868 is observed. This

R^2 value is high and suggests the measurement of intensity and beta values of all CpG

sites is of high quality. After the second methylation microarray was performed, a

scatterplot was also generated of the average β-value of the post-nephrectomy (3 Month)

DNA sample of participant 03 against its duplicate (**Figure 1 B - Appendix**), with an R^2

value of 0.9757. These values were thought to be slightly lower than expected, however

an R^2 above 0.97 also indicates very similar β-values per CpG site (Illumina technical

support, personal communication).

2.2.3 SITE-LEVEL ANALYSIS

Initially, the raw intensity and β-values were exported from GenomeStudio to the

R/Bioconductor package IMA, which automates the pipeline for analysis of 450K

methylation data, for both the site-level and regional analysis. This package allows for

exploratory analysis through a variety of preprocessing and differential testing options.

Differentially methylated positions (DMPs) are single genomic positions having the most

different methylation statuses between sample groups, and are examined by a Quantile-

Quantile plot. A Q-Q plot is used in this analysis to assess the number and magnitude of

observed differential associations between methylation status of every probe before and

after nephrectomy, compared to the association statistics expected under the null

hypothesis of no association[48]. The observed differential associations are shown through

the $-\log_{10}$ value of the calculated P-values, derived from linear models, and are ranked in

order from smallest to largest on the y-axis. These values are plotted against the

theoretical distribution that would be expected under the null hypothesis, on the x-axis. This method for visualizing significant differential associations between two groups is an important tool in both the site-level and regional analysis.

A site-level data analysis was performed on all Pre-op and 3M samples of both methylation chips and fifteen candidate probes were selected for further verification. The samples included in this data set are Patients 08, 06+07 Pool, 04, and 03 and statistical testing generates differential associations displayed in the Q-Q plot in **Figure 2.2**. The outcome of the microarray dataset indicates the highest $-\log_{10}$ P-values deviate above the confidence interval, signifying the P-values of these differentially methylated positions are lower than expected and are statistically significant.
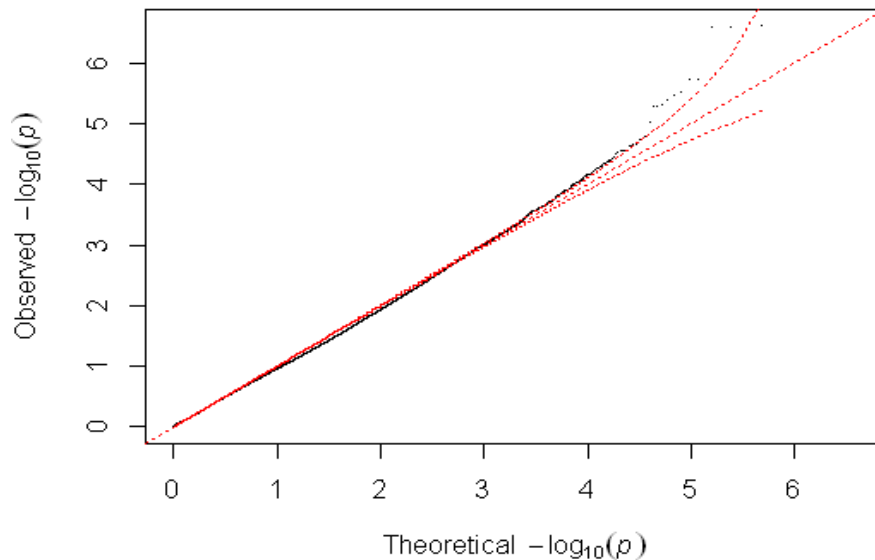
Figure 2.2| Q-Q Plot of Association between Pre- and Post-Nephrectomy Groups from Both Microarrays
The QQ plot of the association between the pre-nephrectomy group (08 Pre, 07+06 Pool Pre, 03 Pre, 04 Pre) and the post-nephrectomy group (08 3M, 04 3M, 07+06 Pool 3M, and 03 3M). Each dot represents the –log of the P value for each probe within the methylation analysis (y-axis), as compared to the –log of the P value under null hypothesis of no association (x-axis).

2.2.4 REGION-LEVEL ANALYSIS

In addition to the site-level analysis, a differential analysis was conducted on each of the eleven genomic regions to determine entire regions which show significant changes in the post-nephrectomy group. These analyses were conducted using the same pre-processed methylation data as the site-level analysis, and generally show some genomic regions contain statistically-significant differentially methylated regions. As previously stated, Illumina assigns the vast majority of methylation probes of the 450K BeadChip to an associated gene. Additionally, the probe's regional information is assigned depending on the probe's position within one of the eleven gene or promoter genomic region categories, such as the shore, shelf or island regions. The QQ Plots of the eleven region-types are provided in **Figure 2 (Appendix)** and demonstrate statistical significance of the individual and specific regions within a single region type. The observed P-values of the individual clusters of probes within the exon 1, north shore, island region, gene body, transcription start site 1500, and transcription start site 200 region-types is equivalent or greater than the expected P-values, and are therefore not considered to be statistically significant. Conversely, select regions within the north shelf, south shelf and 5' untranslated region-types are statistically significant and are differentially associated with a nephrectomy. These statistically significant regions associated with a loss of a kidney are listed in **Table 2.1**, as well as the associated gene and P-value of each region.

Table 2.1 | Statistically Significant Regions within Regional Analysis

The following table lists the statistically significant regions determined during the regional analysis. These individual regions are within the 5'UTR, North Shelf and South Shelf region-types. The gene name the region is associated with is listed, as well as the region-type and P-Value. Genomic regions that are not associated with a gene are classified as NA

Genomic Region	Region-Type	Associated Gene	P-Value
chrX: 37891792-37892347	UTR5	SYTL5	8.640×10^{-06}
chr9:103791944-103792173	UTR5	LPPR1	1.766×10^{-05}
chr22:45182016-45182223	UTR5	ARHGAP8	2.3274×10^{-05}
chr2:219575448-219576080	NSHELF	NA	4.985×10^{-06}
chr7:75677041-75677767	NSHELF	STYXL1	1.681×10^{-05}
chr17:6915824-6916201	NSHELF	ALOX12	2.898×10^{-05}
chr13:25620937-25621554	NSHELF	NA	4.264×10^{-05}
chr7:100253782-100254150	SSHELF	NA	1.874×10^{-06}
chr6:29795553-29796594	SSHELF	HLA-G	2.942×10^{-05}
chr6:108169396-108169790	SSHELF	NA	8.725×10^{-05}

2.3 DISCUSSION

We hypothesize that urinary DNA can be used as a means of novel biomarkers of kidney function through nephron injury, and the overall objective of this study was to identify specific methylation patterns after a nephrectomy. After preprocessing the raw data from the microarrays, differential analyses were conducted to identify individual

probes and genomic regions that are significantly associated with after a kidney is

donated. Overall, the difference in methylation statuses post-nephrectomy of 100

individual probes and ten genomic regions were statistically significant. It is expected that

only a small fraction of the microarray's 480 000 interrogated probes will be differentially

associated with a kidney loss, as only genes involved in renal functioning should be

affected. Also, few confounding factors exist within the experimental design, as both the

pre-nephrectomy and post-nephrectomy samples are from the same tissue type of the

same participant, thus a smaller difference in methylation changes is expected. Thus, the

differential associations observed in both analyses are statistically significant and reflect

the cellular changes that occurred in the remaining kidney of these participants.

CHAPTER 3: BIOLOGICAL RELEVANCE OF CANDIDATE PROBES

3.1 Materials and Methods

3.1.1 Literature Review

A literature review was conducted, through PubMed, to establish that the associated genes of the 100 most differentially associated probes and the ten genomic regions are involved in kidney function. Probes and regions that are not associated with a gene, classified as 'NA', were excluded from the literature search. Scientific articles were found in PubMed by searching the gene name and "renal" or "kidney." Candidate probes and regions were chosen if their associated gene is directly related to kidney function, versus an indirect association though diabetes, hypertension, or creatinine production.

3.2 RESULTS

3.2.1 Site-level Analysis

Scholarly articles were initially found for 27 of the 100 probes of the site-level analysis, however, eleven of these were excluded as these genes were only indirectly associated with kidney function. The eleven excluded genes were directly associated with diabetes, hypertension, inflammatory changes, and creatinine production versus excretion, which then alters kidney function. **Table 3.1** summarizes the sixteen biomarker candidate probes whose genes show biological relevance to kidney function.

Table 3.1| Summary of Differentially Methylated Probes in Site-level Analysis with Genes Related to Kidney Function

The following table lists the probes within the top 100 site-level methylation analysis, their associated gene, chromosome, position (in base pairs), P-Value and the gene's kidney-related function.

Probe	Associated Gene	Chromosome	Position (bp)	Kidney-related Function	P-Value
cg11108676	NTRK3	15	88801004	Cell signaling and may be essential for podocyte foot processes[49]	2.27x10[-05]
cg08552446	PRKG1	10	52898733	Mediator of the nitric oxide/cGMP signaling pathway and involved in glomerular albumin permeability[50]	3.84 x10[-05]
cg00928580	CLU	8	27473612	Chaperone involved in cell death and indicator for early renal injury[51]	5.12 x10[-05]
cg15981475	TIMP2	17	76921948	Maintenance of tissue homeostasis and plays a role in renal injury[5253]	6.47 x10[-05]
cg06081750	WNK1	12	1018307	Important regulator of sodium and potassium homeostasis in nephron[54]	7.48 x10[-05]
cg20197825	PRKDC	8	48744176	DNA double stranded break repair and recombination and mutation increases sensitivity to nephropathy[55]	7.48 x10[-05]
cg23630593	CUX1	7	101517752	Involved in gene expression and cell cycle progression and implicated in polycystic kidney disease[56,57]	9.55 x10[-05]
cg07654864	STY17	16	19183628	Calcium-independent and PKA-dependent processes, and phosphorylated by ADH in proximal tubule cell line[58]	9.79 x10[-05]
cg11806084	SLC22A7	6	43265339	Encodes OAT2 transporter involved in sodium-independent transport of organic anions in basolateral membrane[59]	9.83 x10[-05]
cg16539675	RNASEN	5	31532179	Involved in initial steps of microRNA biogenesis and important in podocyte functioning[60]	9.98 x10[-05]

cg06971503	INF2	14	105170815	Severs actin filaments and defects result in proteinuria and ESRD[61]	1.09 x10^{-04}
cg26571478	NEBL	10	21436607	Down-regulated in podocyte foot processes in podocyte injury[62]	1.32 x10^{-04}
cg05299847	CBS	21	44477612	Regulator of hydrogen sulfide and involved in kidney fibrosis[63]	1.42 x10^{-04}
cg16423624	ALDH2	12	112220997	Responsible for alcohol metabolism and activation may prevent renal IR injury[64,65]	1.44 x10^{-04}
cg00746810	ATG7	3	11564388	Involved in ATP-binding and involved in renal injury and proximal basal and cisplatin-induced autophagy in kidney[66]	1.80 x10^{-04}
cg11263235	CORO1C	12	109085523	Involved in a variety of cellular processes, and up-regulated in kidney cells during PAR1 and PAR2 activation[67]	6.08 x10^{-05}

3.2.2 Region-level Analysis

The biological relevance of the significantly-associated candidate probes within the regional analysis was also determined by selecting articles that show a direct relationship of the candidate probe's gene with renal functioning. The ARHGAP8-PRR5 genomic region within the 5' Untranslated region (UTR5) is the only region that meets three of these criteria. The ARHGAP8-PRR5 region is statistically significant as its P-Value of 2.327×10^{-5} is less than that predicted of the null hypothesis. This gene is a rare read-through sequence, which conjoins the ARHGAP8 and PRR5 neighboring genes[68]. The ARHGAP8 gene encodes a member of the RHOGAP family, and is typically involved in cell signaling and cytoskeletal processes, and PRR5 encodes proline rich 5, renal[69]. Both encoded proteins are predominant in kidney tissue, and PRR5 has been shown to inhibit an organic cation transporter involved in renal excretion[68].

3.3 DISCUSSION

In this study, we hypothesize that urinary DNA can be used to identify epigenetic changes in kidney function post-nephrectomy, and selecting biologically relevant candidate probes is an essential criteria for this hypothesis. Of the 100 statistically significant candidate probes from the site level analysis, scholarly articles were obtained that indicate sixteen probes are involved in kidney function. The genes of these candidate probes are related to renal function through various ways, such as podocyte foot processes, renal injury, and nephropathy. The ARHGAP8-PRR5 UTR5 region, consisting of seven probes, was the only genomic region that was related to renal function through transporter-mediated renal excretion. Overall, 23 candidate probes were both statistically significant post-nephrectomy and biologically related to kidney function, signifying an enrichment of kidney related genes within the analysis. Identifying biologically-relevant genes is key in uncovering potential biomarkers for kidney function, and confirms the differential analyses and experimental design were conducted successfully.

CHAPTER 4: INTERNAL VALIDATION OF CANDIDATE PROBES

4.1 Materials and Methods

4.1.1 Analysis of Raw Beta-Values

Internal validation of the previously selected candidate probes was conducted by analyzing the raw, unprocessed methylation data of each probe in R. For each of the four patients included in the analysis, the average beta-value, corresponding to the level of methylation at the site, was extracted from the raw data. The difference between the raw average beta post-nephrectomy and pre-nephrectomy was determined for all four patient samples, and visualized through generated barplots. Candidate probes with discordance between participants in direction or level of methylation failed the internal validation criteria, and were filtered from the study.

4.2 RESULTS

4.2.1 Site-level Analysis

Internal validation is the third criteria for determining candidate probes for kidney function biomarkers, and was established by examining the difference in raw methylation values before and after surgery. Sixteen candidate probes were shown to be both significantly associated post-nephrectomy and are biologically relevant to kidney function. Fifteen of these probes show consistency in the level and direction of methylation, as shown in **Figure 4.1**. However, the probe within the CORO1C gene (**Figure 4.1 P**) is hyper-methylated in two participant samples and hypomethylated in the other participants.

Figure 4.1 | Change in Raw Beta-Values Before and After Nephrectomy of Probes with Kidney-Related Functions in Site-level Analysis

The following figure illustrates the change in the raw beta-value, representing level of methylation. Each bar shows the difference in methylation (along the y-axis), and the four bars represent the four participants in this study. A barplot was made for all fifteen probes chosen for validation; A) NTRK3, B) PRKG1, C) CLU, D) TIMP2, E) WNK1, F) PRKDC, G) CUX1, H) STY17, I) SLC22A7, J) RNASEN, K) INF2, L) NEBL, M) CBS, N) ALDH2 O) ATG7, P) CORO1C

M) CBS	**N) ALDH2**	**O) ATG7**

P) CORO1C
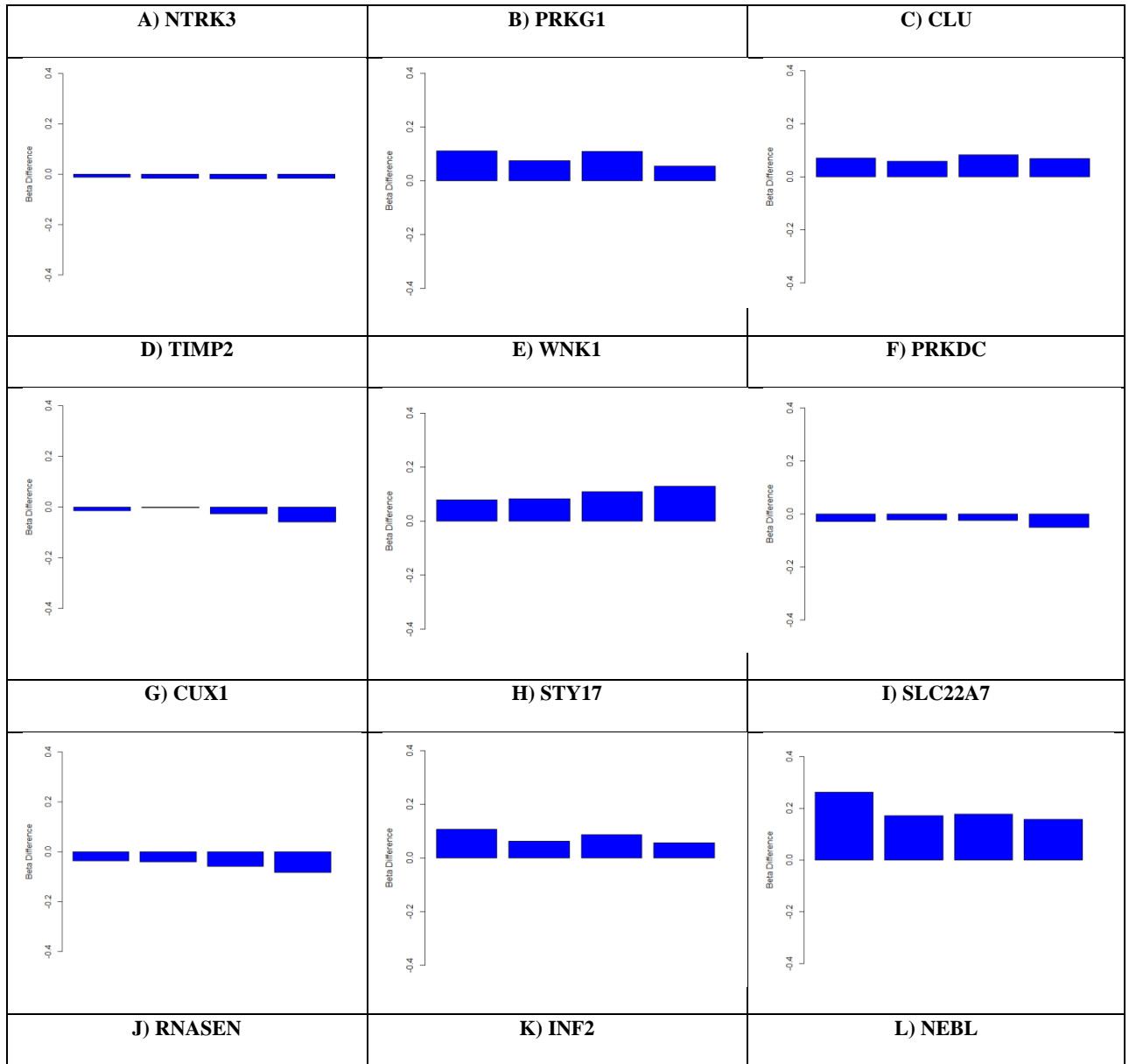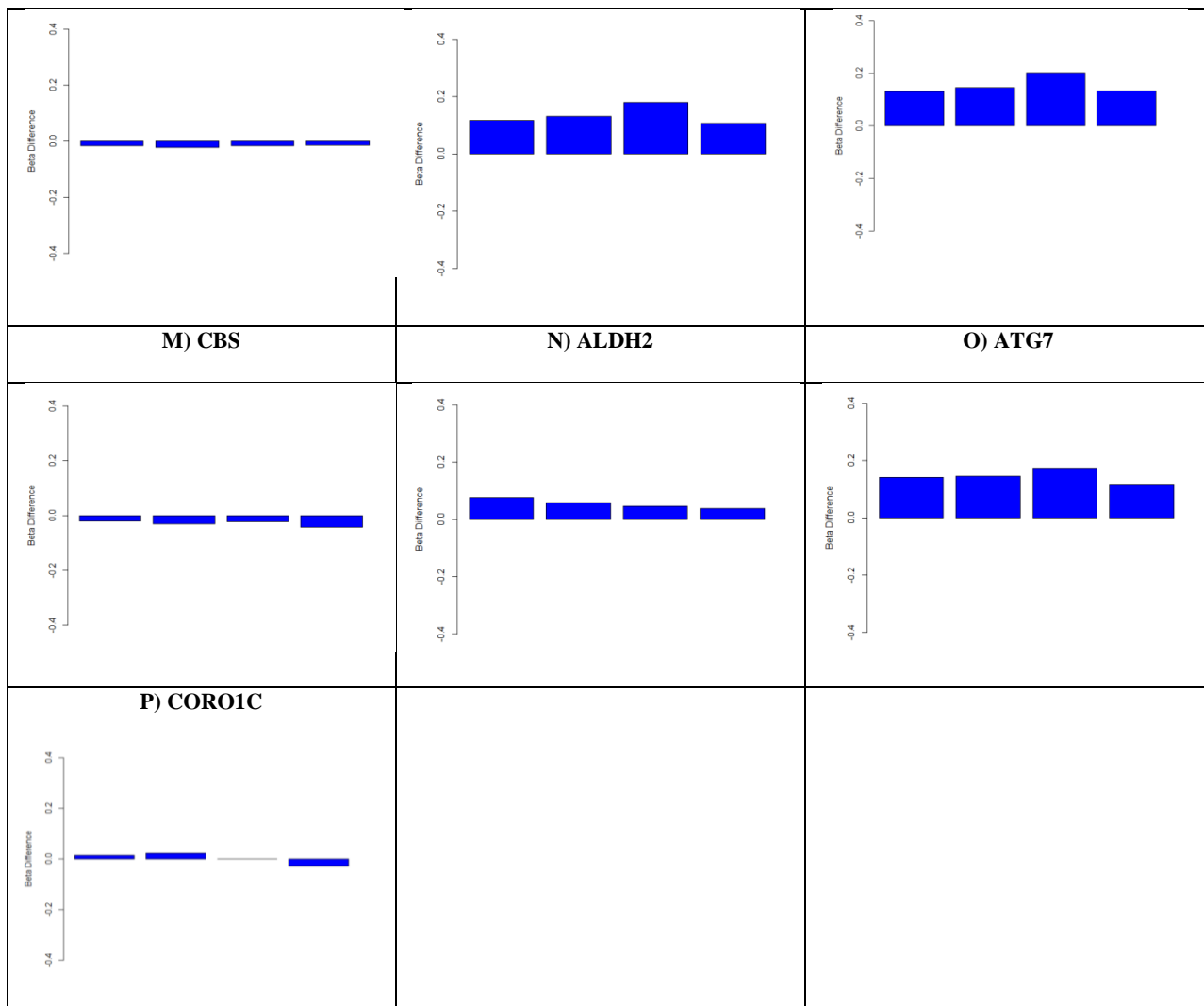
4.2.2 Region-level Analysis

Within the regional analysis, the raw methylation values of the seven probes within the ARHGAP8-PRR5 region have been extracted for each of the four participants of the study. **Figure 4.2** includes the internal validation barplots of the seven probes within the ARHGAP8-PRR5 gene region. Four of these methylation sites have shown consistent hyper- or hypo-methylation within all four participant samples.

Figure 4.2 | Change in Raw Beta-Values Before and After Nephrectomy of Probes with Kidney-Related Functions in Regional-Analysis

The following figure illustrates the change in the raw beta-value, representing level of methylation. Each bar shows the difference in methylation (along the y-axis), and the four bars represent the four participants in this study. A barplot was made for the seven probes within the AHRGAP8-PRR5 genomic region for internal validation in the regional analysis.

Gene Name	Probe	Region	Raw Beta Values
ARHGAP8-PRR5	cg13363124	ISLAND/UTR5	
ARHGAP8	cg13452317	UTR5	
ARHGAP8	cg20686461	UTR5	

ARHGAP8-PRR5	cg26523649	ISLAND/UTR5	
ARHGAP8	cg13757640	UTR5	
ARHGAP8	cg26523649	UTR5	
ARHGAP8	cg27407546	UTR5	

**4.3 DISCUSSION**

Internal validation is the third criteria for determining candidate probes for kidney function biomarkers, and was established through the difference in raw methylation values before and after surgery. Each candidate probe must be consistently hypo- or hyper-methylated in all four participants included in the analysis. Also, the change in the raw methylation value of the four participants must be consistent within all participants of the study. This procedure is an important means eliminating discordant candidate probes, as well as probes with an overall methylation value driven by one outlying participant methylation value.

Within the site-level analysis, the probe within the CORO1C gene did not meet internal validation standard, and has been eliminated as a potential candidate probe for this study. In the region-level analysis, four of the seven candidate probes have been internally validated in the microarray data of all four patient samples. Thus, DNA methylation changes in fifteen probes, from the site-level analysis, and four probes, from the region-level analysis, have been validated in microarray data. Since these nineteen methylation sites met the internal validation criteria, they were considered potential biomarkers of kidney function.

CHAPTER 5: EXTERNAL VALIDATION OF CANDIDATE PROBES

5.1 Materials and Methods

5.1.1 Primer Design

Nineteen probes were found to be differentially associated post-nephrectomy and involved in in kidney function, and these probes will be externally validated through pyrosequencing technology. Primers for these candidate probes have been designed through the PyroMark Assay Design Software (Qiagen). Two of the candidate probes within the AHRGAP8 gene are in densely methylated genomic regions. Thus, the positioning of these probes did allow for functional primers to be designed, as these specific primers would likely anneal to areas with CpG sites. Overall, primers for seventeen of the nineteen candidate probes were designed. However, the eight candidate probes with the lowest P-value were amplified and pyrosequenced, due to limited supply of DNA. Information these eight primers are provided in **Table 2 (Appendix)**, including forward, reverse and sequencing primer sequences and PCR product length (**Figure 5.1**).
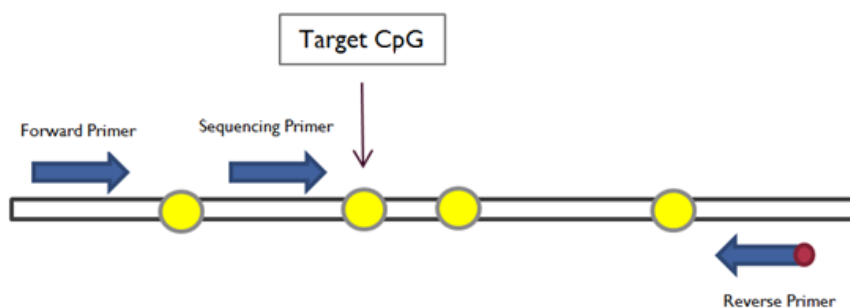
Figure 5.1 | Representation of the Primers Required for the Methylation Application of Pyrosequencing

Three primers are required for the methylation application of pyrosequening; a forward primer, reverse primer, and sequencing primer. These primers must not overlap with any CpG sites (yellow circles) for adequate amplification. Either the forward or reverse primer must be biotinylated (red circle), then HPLC purified. The orientation of the sequencing primer should be within 1-2 base-pairs away from the target CpG site.

5.1.2 Amplification of Candidate Probes

Polymerase Chain Reaction (PCR) was initially conducted on bisulfite-treated control DNA to test the amplification of the candidate probes. These PCR trials were conducted by testing increasing $MgCl_2$ final concentrations and decreasing annealing temperatures per amplicon. Two different control DNAs were used in these trials, as well as a duplicate to ensure amplification. PCR products were analyzed through 1.5% Agarose gels. These trials led to PCR conditions for each candidate probe (**Table 3 – Appendix**) that generates strong and specific amplification. After amplification, these PCR products were stored at -20°C, until required for pyrosequencing testing. Following this pyrosequencing assay experimentation, participant DNA was then amplified according to the determined conditions per amplicon.

5.1.3 Pyrosequencing Approach to Determine Methylation Status of Specific Amplicons

As mentioned previously, pyrosequencing is a sequencing-by-synthesis technique that monitors the quantity of incorporated nucleotides within a strand of DNA. The PyroMark Q24 instrument (Qiagen) has been used to validate the methylation status of eight of the nineteen candidate probes, using two high-quality control DNAs from blood samples. A trial PyroMark Q24 Plate has been analyzed, which contained the eight assays of one control DNA and eight duplicate sets of a second control DNA (**Figure 3 - Appendix**). After analysis from the initial experimental plate, three PyroMark Q24 plates with PCR products from participant DNA were also run and analyzed.

5.2 RESULTS

Pyrosequencing assays have been designed and tested on control DNA for eight candidate probes as biomarkers for kidney function, and six of these assays were successful. The generated pyrograms for the eight candidate probes are shown in **Table 5.1**, along with the software-determined quality of the run and methylation status of each probe. The peaks of the pyrograms are the observed light signals from the base-pair additions, and the thin grey boxes behind each peak is the expected light signals, based on the predicted sequence of that strand of DNA. The PyroMark Q24 software incorporates independent controls within the assay to assess the quality of the run, such as blank dispensations, or nucleotide additions, and bisulfite treatment controls. Added nucleotide quantities assess the quality of the run, as this base pair is not a part of the target sequence, and should yield a blank signal. Bisulfite treatment controls take advantage of cytosine base pairs that are not followed by a guanine, as these nucleotides should always be converted to thymine base pairs after bisulfite conversion. The PyroMark Q24 software adds C nucleotides just before the expected T nucleotide within the target sequence. No signal at the added C nucleotide should be observed to confirm a successful bisulfite treatment has occurred. Signal-to-noise ratios, and algorithms for sum and pattern deviation in variable positions are other independent controls within the PyroMark Q24 analysis that help determine run quality.

5.2.1 PyroMark Q24 Trial Results using Test DNA

The interrogated candidate probes within the six assays of the pyrosequencing run are all of passing quality, however, the probes within the TIMP2 assay have an uncertain

run quality and the candidate probes of the NTRK3 assay were unsuccessful. This failed

quality assessment of the position within the NTRK3 assay was due to an overall lowered

peak height, as noted in the NTRK3 pyrogram in **Table 5.1**. A lowered peak height

indicates a low amount of nucleotides incorporated at that position during base-pair

addition synthesis. The candidate probe within the TIMP2 assay was found to have an

uncertain quality due to a high peak deviation at dispensation 16 within the TIMP2

pyrogram. Overall, six of the eight candidate probes of interest were successfully

pyrosequenced, using control DNA, and a passing quality and bisulfite conversion was

established.

Table 5.1 | Pyrograms of Eight Pyrosequenced Assays

Pyrograms for the eight candidate probes have been generated through the PyroMark Q24 instrument and software. The quality and methylation status, in percentage, of each candidate probe is shown for one of the two control DNAs incorporated in this trial.

Assay	Pyrogram	Quality	Meth (%)
NTRK3		Failed	100

PRKG1		Passed	93
CLU		Passed	95
TIMP2		Check	4

WNK1		Passed	84
PRKDC		Passed	95
CUX1		Passed	87

SYT17		Passed	89
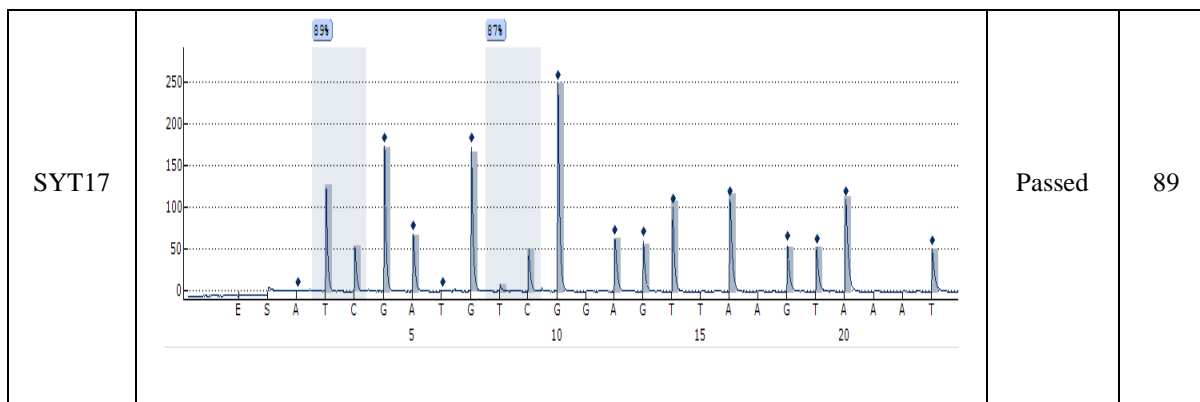
5.2.2 PyroMark Q24 Results using Participant DNA

After the successful completion of PyroMark Assays for six candidate sites, these sites were validated in the five sets of pre- and post-nephrectomy DNA samples from the methylation microarray analysis. Unfortunately, two positions did not show specific and strong amplification in three DNA sample sets, and as a result, these PCR products were not pyrosequenced. The remaining 48 PCR products were pyrosequenced in three PyroMark Q24 plates, each of which included a positive and negative control. The methylation values and quality status of each amplified DNA sample is shown in **Table 4** (**Appendix**). The majority of pyrosequenced amplicons were of passing quality, while five assays obtained a 'Check' quality status, due to slightly lower than expected light signals. These prompts suggest a reassessment of the site, and may still be acceptable. Individual analysis of each of these five assays determined they were all of passing quality. Overall, 75% of the methylation sites were successfully pyrosequenced, while the remaining 25% obtained a failing quality status. These amplicons failed the quality

control measures mostly due to low signal-to-noise ratios and improper bisulfite conversion.

As previously stated, the purpose of the pyrosequencing experiments is to confirm the findings of the methylation microarray through precise methylation level measurements of the candidate probes. In total, six methylation sites were interrogated in five participant pre- and post-nephrectomy DNA samples. No statistically significant difference can be detected in the averaged methylation level of all six candidate probes after a nephrectomy (**Table 5.2**). Furthermore, **Figure 5.2** through to **Figure 5.7** compare methylation statuses per probe in both the microarray and pyrosequencing techniques. These figures illustrate that the pyrosequencing-based methylation values for all six probes do not show consistency in direction of methylation, as compared to the microarray values.

Table 5.2 | Methylation Changes of the Six Candidate Probes

A group comparison was conducted for the altered methylation statuses of the six candidate probes of interest of the pre- and post-nephrectomy DNA samples.

Associated Gene	Level of Methylation (%)		P-value
	Pre-Nephrectomy Sample	Post-Nephrectomy Sample	
PRKG1	95.4 ±3.7	95 ±1.7	0.83
CLU	91.2 ±2.9	91.4 ±3.3	0.945
WNK1	75.8 ±8.8	78.0 ±12.5	0.734
PRKDC	83.6 ±8.4	84.6 ±3.4	0.779
CUX1	68.0 ±2	75.5 ±0.5	0.205
STY17	87.0 ±2	88.5 ±3.5	0.83

Figure 5.2 | Comparison of the Microarray- and Pyrosequencing-Derived Methylation Pattern of the PRKG1 Candidate Probe

The difference in methylation levels of the PRKG1 candidate probe is shown in five DNA sample sets. The beta-values of the candidate probes, which represent the methylation levels, have been extracted from the Illumina 450K methylation microarray data (blue bars). The difference in methylation levels post-nephrectomy for this amplicon are generated from the PyroMark Q24 Assay (orange bars).
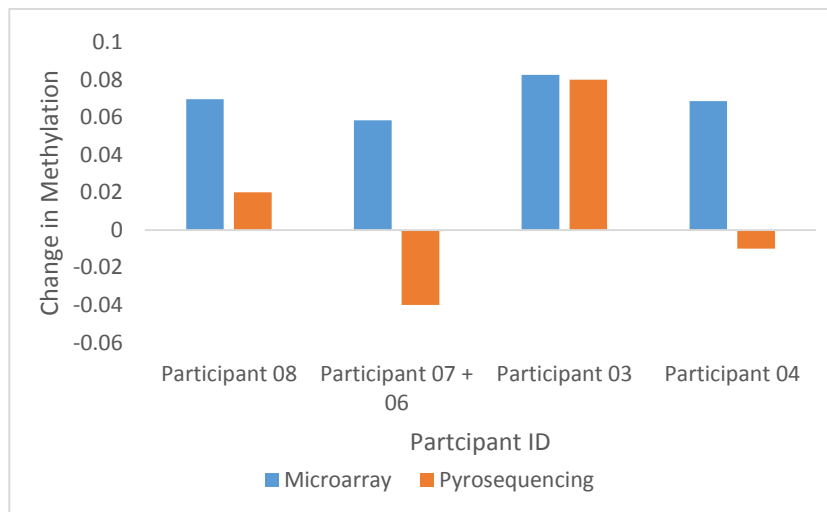
Figure 5.3 | Comparison of the Microarray- and Pyrosequencing-Derived Methylation Pattern of the CLU Candidate Probe

The difference in methylation levels of the CLU candidate probe is shown in five DNA sample sets. The beta-values of the candidate probes, which represent the methylation levels, have been extracted from the Illumina 450K methylation microarray data (blue bars). The difference in methylation levels post-nephrectomy for this amplicon are generated from the PyroMark Q24 Assay (orange bars).
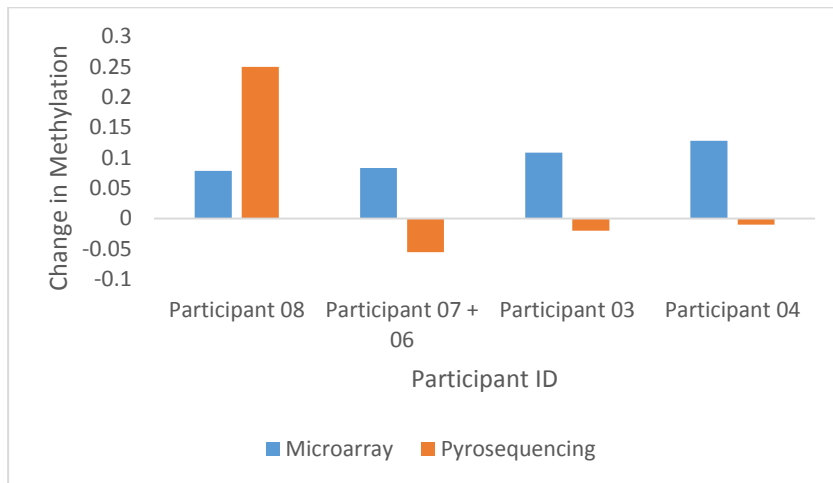
Figure 5.4 | Comparison of the Microarray- and Pyrosequencing-Derived Methylation Pattern of the WNK1 Candidate Probe

The difference in methylation levels of the WNK1 candidate probe is shown in five DNA sample sets. The beta-values of the candidate probes, which represent the methylation levels, have been extracted from the Illumina 450K methylation microarray data (blue bars). The difference in methylation levels post-nephrectomy for this amplicon are generated from the PyroMark Q24 Assay (orange bars).
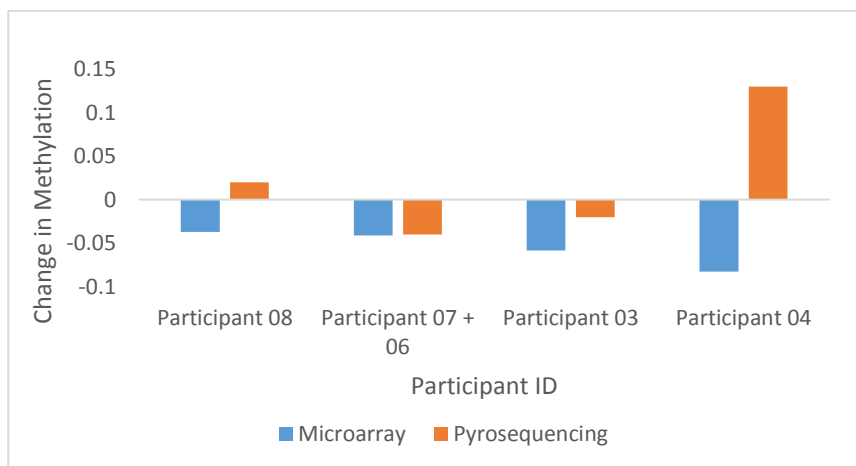
Figure 5.5 | Comparison of the Microarray- and Pyrosequencing-Derived Methylation Pattern of the PRKDC Candidate Probe

The difference in methylation levels of the PRKDC candidate probe is shown in five DNA sample sets. The beta-values of the candidate probes, which represent the methylation levels, have been extracted from the Illumina 450K methylation microarray data (blue bars). The difference in methylation levels post-nephrectomy for this amplicon are generated from the PyroMark Q24 Assay (orange bars).
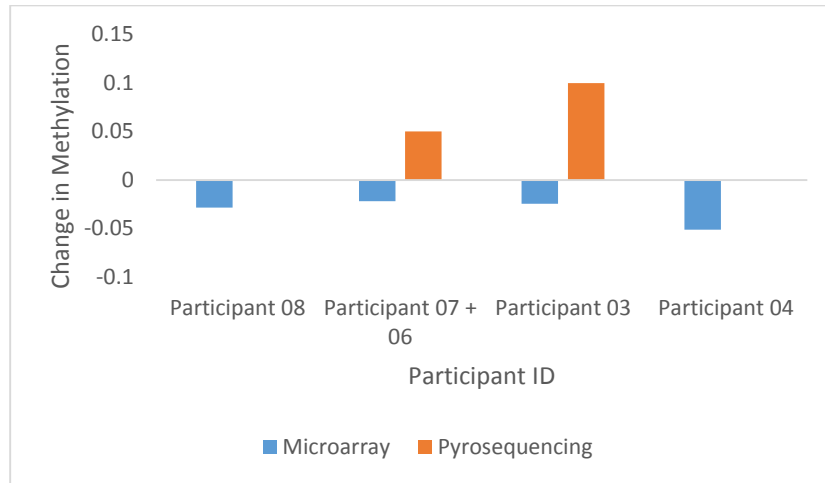
Figure 5.6 | Comparison of the Microarray- and Pyrosequencing-Derived Methylation Pattern of the CUX1 Candidate Probe

The difference in methylation levels of the CUX1 candidate probe is shown in five DNA sample sets. The beta-values of the candidate probes, which represent the methylation levels, have been extracted from the Illumina 450K methylation microarray data (blue bars). The difference in methylation levels post-nephrectomy for this amplicon are generated from the PyroMark Q24 Assay (orange bars). Pyrosequencing could not be conducted on participant samples 08, 07 and 04, and could not be included below.
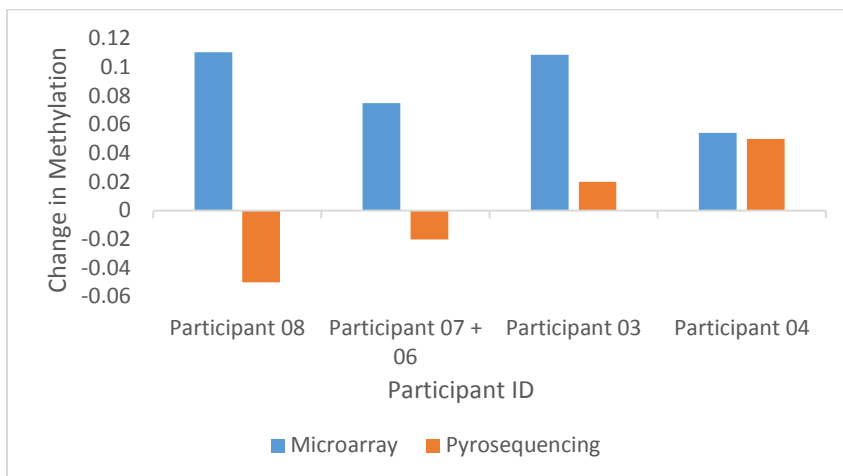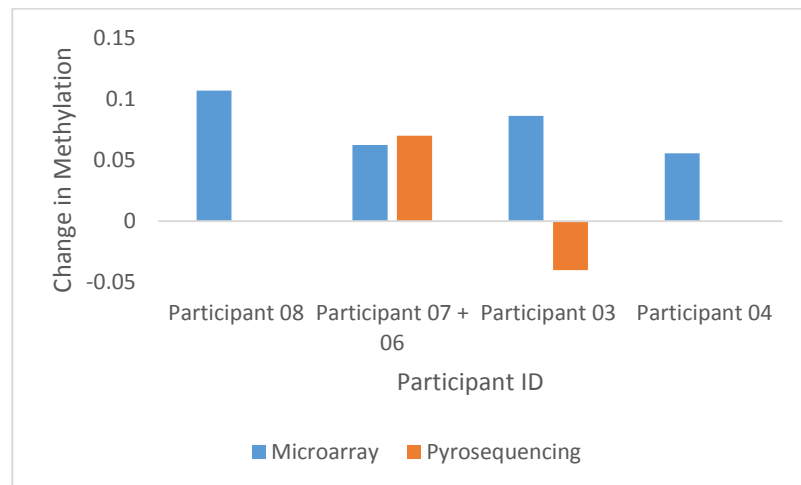

Figure 5.7 | Comparison of the Microarray- and Pyrosequencing-Derived Methylation Pattern of the SYT17 Candidate Probe

The difference in methylation levels of the SYT17 candidate probe is shown in five DNA sample sets. The beta-values of the candidate probes, which represent the methylation levels, have been extracted from the Illumina 450K methylation microarray data (blue bars). The difference in methylation levels post-nephrectomy for this amplicon are generated from the PyroMark Q24 Assay (orange bars). Pyrosequencing could not be conducted on participant samples 08, 07 and 04, and could not be included below.

Another means of testing the validity of this external validation criteria was conducted, and an overall poor concordance between the microarray and pyrosequencing data is evident. The methylation values per candidate probe for each participant's microarray value was plotted against the same participant's pyrosequencing value, in the pre- and post-nephrectomy groups separately. The goal of this analysis was to evaluate both the association with pre and post nephrectomy status and the correlation between chip and microarray methylation measurements. **Figure 5.8** through to **Figure 5.13** show the pre- and post-nephrectomy methylation levels for each of the six candidate probes, with respect to both applications. It is clear that the candidate probes are highly methylated in both applications. There is significant clustering in the pre- and post-nephrectomy groups in these candidate probes. However, clustering is not observed along the y-axis, which does not indicate a significant difference in methylation of both applications, rather only of the microarray technique.
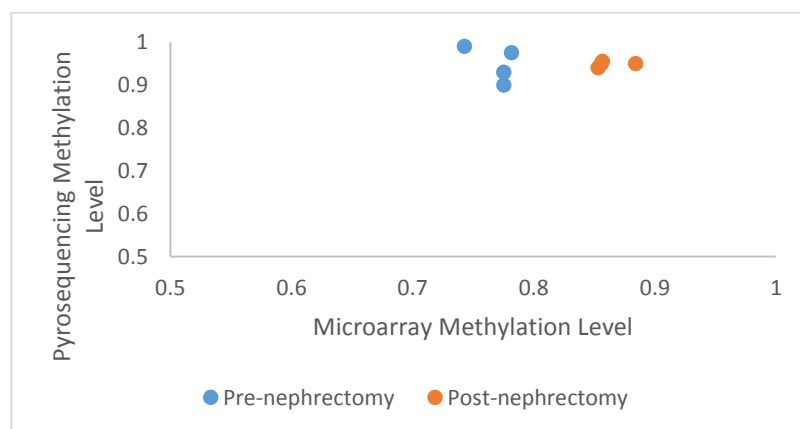
Figure 5.8 | Pre- and Post-nephrectomy Methylation Level Plot for PRKG1 Candidate Probe

The PRKG1 probe methylation values (0 being completely nonmethylated and 1 is completely methylated) for all Pre- and post-nephrectomy samples have been plotted, where each dot represents an individual participant. The microarray values for each participant is plotted on the x-axis, while the pyrosequencing values per participant are plotted on the y-axis. All pre-nephrectomy samples are shown in blue, while post- samples are shown in red.
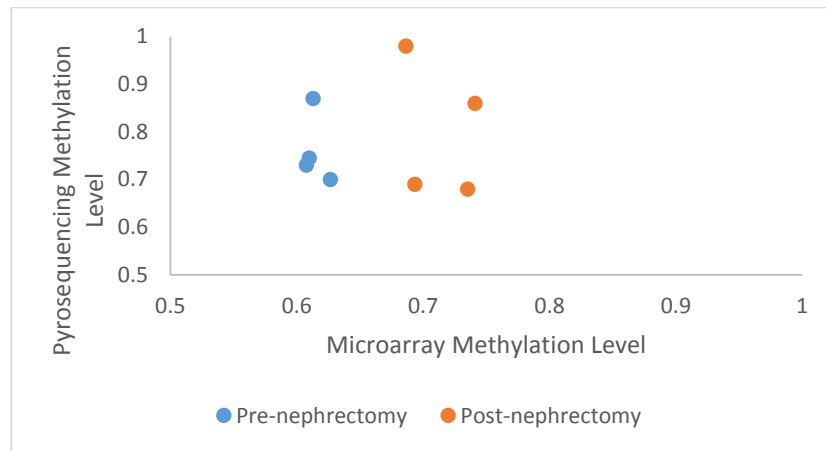
Figure 5.9 | Pre- and Post-nephrectomy Methylation Level Plot for WNK1 Candidate Probe

The WNK1 probe methylation values for all Pre- and post-nephrectomy samples have been plotted, where each dot represents an individual participant. The microarray values for each participant is plotted on the x-axis, while the pyrosequencing values per participant are plotted on the y-axis. All pre-nephrectomy samples are shown in blue, while post- samples are shown in red.
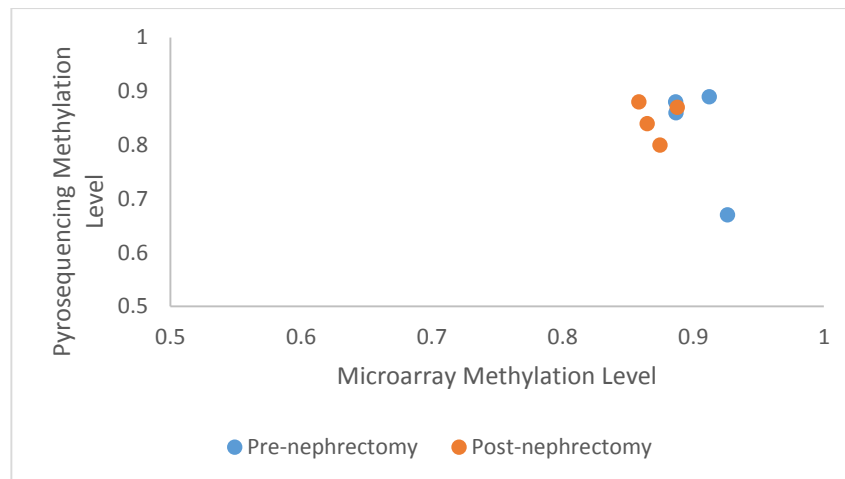
Figure 5.10 | Pre- and Post-nephrectomy Methylation Level Plot for PRKDC Candidate Probe

The PRKDC probe methylation values for all Pre- and post-nephrectomy samples have been plotted, where each dot represents an individual participant. The microarray values for each participant is plotted on the x-axis, while the pyrosequencing values per participant are plotted on the y-axis. All pre-nephrectomy samples are shown in blue, while post- samples shown in red.
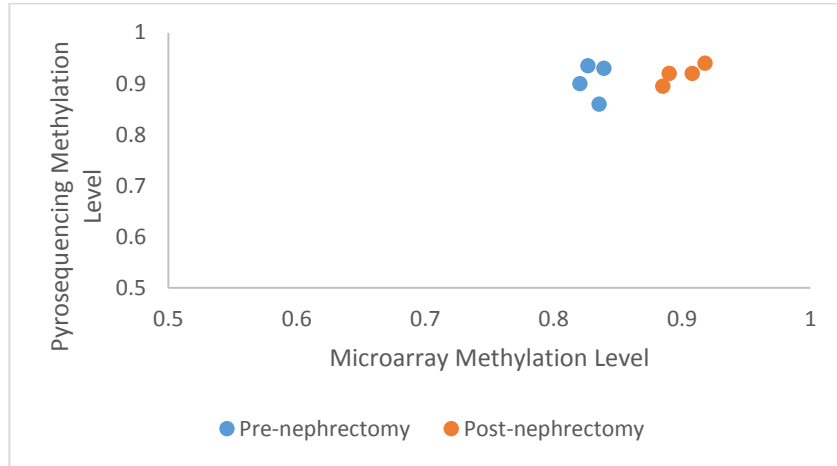
Figure 5.11 | Pre- and Post-nephrectomy Methylation Level Plot in CLU Candidate Probe

The CLU probe methylation values for all Pre- and post-nephrectomy samples have been plotted, where each dot represents an individual participant. The microarray values for each participant is plotted on the x-axis, while the pyrosequencing values per participant are plotted on the y-axis. All pre-nephrectomy samples are shown in blue, while post- samples shown in red.
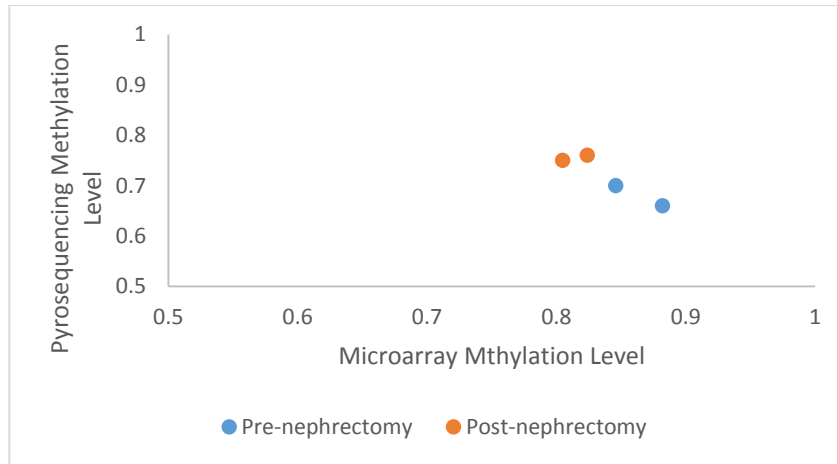
Figure 5.12 | Pre- and Post-nephrectomy Methylation Level Plot in CUX1 Candidate Probe

The CUX1 probe methylation values for all Pre- and post-nephrectomy samples have been plotted, where each dot represents an individual participant. The microarray values for each participant is plotted on the x-axis, while the pyrosequencing values per participant are plotted on the y-axis. All pre-nephrectomy samples are shown in blue, while post- samples shown in red.
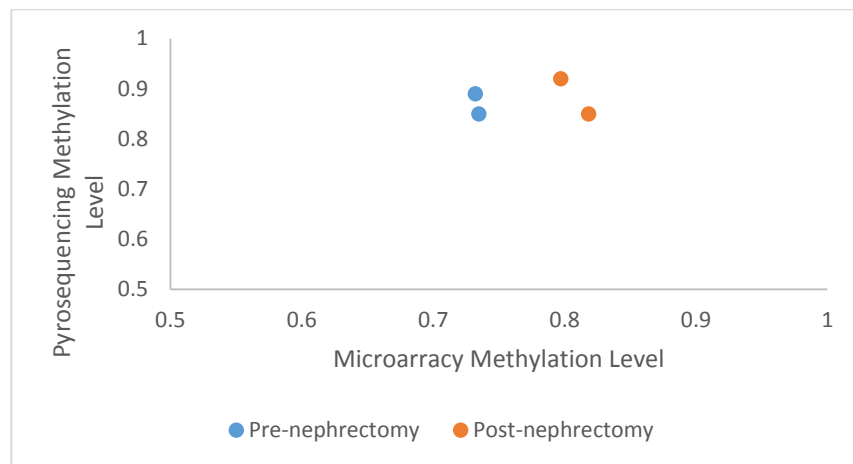
Figure 5.13 | Pre- and Post-nephrectomy Methylation Level Plot in SYT17 Candidate Probe

The SYT17 probe methylation values for all Pre- and post-nephrectomy samples have been plotted, where each dot represents an individual participant. The microarray values for each participant is plotted on the x-axis, while the pyrosequencing values per participant are plotted on the y-axis. All pre-nephrectomy samples are shown in blue, while post- samples shown in red.

5.3 DISCUSSION

External validation of the methylation microarray data is an important aspect in uncovering potential methylation patterns as biomarkers of kidney function. The PyroMark Q24 technology was chosen to externally validate the chosen candidate probes, due to its unprecedented accuracy in detecting the methylation status of defined targets. Primers have been designed for seventeen probes. However, pyrosequencing assays for the eight most statistically significant candidate probes have been tested for quality assurance. Participant pre- and post-nephrectomy DNA samples were then externally validated in the six successfully interrogated sites of interest in control DNA.

Overall, six of the eight designed PyroMark assays are of passing quality control and were reproducible, as these assays were found to have very similar peak heights and methylation levels between the two different DNA samples, as well as the duplicates.

51

Unfortunately, the assay developed to interrogate a candidate probe within the NTRK3 gene was unsuccessful. Although a methylation status of the candidate probe was obtained in this assay, the quality of this targeted position is unsatisfactory due to a low peak height and a very strong signal at the independent control. A low peak height indicates an overall low amount of nucleotides are incorporated at that position during synthesis of the complementary strand of DNA within the sample. The most common causes of this low signal is an unspecific or small amount of PCR product. Other common causes of low peaks may be due to steps involving the actual pyrosequencing process, such as a loss of streptavidin beads during vacuum prep steps, incorrect instrument methods and reagents used during the process. However, it is unlikely this issue is regarding the pyrosequencing protocol, as all other assays of the PyroMark plate achieved a passing quality. Also, visualization of the PCR product, through an agarose gel (**Figure 4 - Appendix**), indicates there is a sufficient amount of specific PCR product.

Another issue regarding the NTRK3 assay is the appearance of a high signal at the added thymine base pair in the sixth nucleotide sequence, possibly due to a mutation within the target sequence. However, no known mutations were found in this genomic area, when comparing to genomic databases, such as the Ensembl and 1000 Genome SNP databases. Also, the bisulfite control in this assay was unsuccessful, which is likely due to the bisulfite control being placed too far into the sequence for the PyroMark Q24 to obtain an accurate read. Overall, the failing quality of all three NTRK3 assays within the pyrosequencing run introduces doubt to the methylation statuses of these targets, and as a result, this candidate probe will not be validated in participant urinary DNA.

Another concern regarding the pyrosequencing trial is the assay's inability to accurately interrogate the TIMP2 candidate probe. This decreased precision is a result of the two stretches of multiple identical nucleotides, called homopolymers, within the DNA sequence. As a result, both the second and sixteenth nucleotide sequence of the assay likely have shorter than expected peaks, which is due to unreliable quantification for a run of five or more nucleotides. Since the methylation site of interest contains a nearby homopolymer, the quantification will not be reliable due to variable quality ratings in future runs. Thus, the candidate probe within the TIMP2 gene will not be externally validated with participant DNA.

External validation of the candidate probes was conducted in participant DNA to replicate the findings of the methylation microarray, which is the final criteria to uncover novel genetic biomarkers for kidney function. Although, the results of the pyrosequencing analysis alone do not show statistically significant methylation changes after the removal of a kidney in all six probes (**Table 5.2**), a variety of technical factors may have contributed to this outcome. Firstly, DNA was extracted from urine samples for the microarray and pyrosequencing experiments separately, which can introduce a level of technical bias. The urinary DNA isolation kit was also updated slightly since the first DNA was needed for the microarray. Specifically, a step involving proteinase K and pronase digestion of proteins within the urine was removed. Although the participant samples are from healthy individuals without proteinuria, this step may impact lysis of exfoliated cells. This change may have also contributed to the inconsistencies between the microarray and pyrosequencing results, as both measures are very sensitive. Secondly, the

CUX1 and SYT17 targets could not be amplified in three participant sample sets, leading to an incomplete validation experiment. Finally, a quarter of the PyroMark Q24 assays analyzed failed quality control due to multiple factors, such as unconverted bisulfite targets and low signals. Failed Pyrosequencing assays and extraction method differences are the most likely reasons that no association can be made in the post-nephrectomy methylation levels.

The overall objective of the external validation step was to verify the candidates from the microarray analysis. Regrettably, this aim could not be achieved as methylation level changes were not confirmed with the pyrosequencing data. An overall poor concordance between the microarray and pyrosequencing methods was evident (**Figures 5.8** to **Figure 5.13**), as the participant methylation values were situated far from the y = x line, or line of equality. In addition, the lack of apparent differences between the pre- and post-nephrectomy samples with the pyrosequencing data, along the y-axis, demonstrates the lack of replication of the microarray results. Many contributing factors within the pyrosequencing experiments are suspected to have caused this lack of replication, such as the failed pyrosequencing assays, inability to amplify candidate probes in all participant samples, and technical variance. To conclude, replication of the candidate probes between the two methods could not be achieved, as a clear change in methylation post-nephrectomy is not evident, which is likely due to the discussed technical factors. Therefore, the candidate probes from the methylation microarray do not reflect changes in kidney function at this time.

CHAPTER 6: SUMMARY

As previously discussed, CKD is a major public health concern, characterized by an irreversible reduction in renal functioning. Patients affected by CKD are at a much greater risk of developing CVD. Creatinine-based eGFR is the current means of CKD diagnosis, however, this method is especially problematic for early detection. Thus, uncovering novel DNA methylation changes as early markers can complement GFR estimation to slow the progression of the disease and maximize patent outcome.

We hypothesized that urine contains DNA from apoptotic and exfoliated cells from the kidney, which can be used to uncover novel biomarkers to better detect CKD. Previous studies indicate urinary DNA contains kidney-specific methylation, as tumor-suppressor genes in the DNA of both renal tumor biopsy and urine contain the same methylation changes. This study demonstrates that methylation analysis to identify changes in kidney functioning, through apoptotic and exfoliated cells from the kidney, is possible. Thus, the overall objective of this study was to determine novel urinary DNA methylation patterns representative of a loss of kidney mass post-nephrectomy, as a model for the loss of kidney functioning in CKD patients.

In this pilot study, patients that have undergone a nephrectomy at St. Joseph's Healthcare in Hamilton have donated urine samples before and after their surgery. DNA from the urine samples has been extracted, and a genome-wide methylation analysis has been conducted through the 450K HumanMethylation microarray (Illumina) and the IMA package (R/Bioconductor). The outcome of the differential analyses was encouraging, as

the differential associations observed between the pre- and post-nephrectomy groups are statistically significant in both the site-level and regional analysis. Through these analyses, nineteen probes have been systematically selected for external validation, based on statistical significance, biological relevance to kidney function, and internal validation of the raw methylation data per candidate probe.

External validation, through methylation-specific pyrosequencing, has been conducted through the PyroMark Q24 technology. Eight candidate probes have been initially tested with control DNA, and six of these assays were successful. Finally, external validation these six candidate probes was conducted in pre- and post-nephrectomy DNA samples. The methylation status of the candidate probes from the pryosequencing steps did not show clear differences post-nephrectomy, and thus did not replicate the findings from the microarrays. With this in mind, many contributing factors exist that could have led to these results, such as technical variance and pyrosequencing assays of failing quality controls. To conclude, nineteen microarray-determined candidate probes were found to be significantly associated post-nephrectomy, biologically relevant to kidney function and consistent in level and direction of methylation. While the hypothesis we originally postulated holds true within the microarray data, these findings could not be replicated through pyrosequencing. Finally, future experimental and clinical studies should aim to explore epigenetic biomarkers of kidney function beyond the nephrectomy model, as genetic markers of CKD is poorly characterized in urinary DNA.

REFERENCES

1.   Levin, A. *et al.* Guidelines for the management of chronic kidney disease. *CMAJ* **179,** 1154–62 (2008).

2.   Piwko, C. *et al.* The STARRT trial: a cost comparison of optimal vs sub-optimal initiation of dialysis in Canada. *J. Med. Econ.* **15,** 96–104 (2012).

3.   Manns, B. J., Mendelssohn, D. C. & Taub, K. J. The economics of end-stage renal disease care in Canada: incentives and impact on delivery of care. *Int. J. Health Care Finance Econ.* **7,** 149–69 (2007).

4.   Zhang, Q.-L. & Rothenbacher, D. Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC Public Health* **8,** 117 (2008).

5.   Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int. Suppl.* **3,** 1–150 (2013).

6.   Abboud, H. & Henrich, W. L. Stage IV Chronic Kidney Disease. (2013).

7.   Zoccali, C., Santoro, A. & Plebani, M. Age, stage and biomarkers for the definition of CKD: a construction in progress. *Clin. Chem. Lab. Med.* **51,** 1919–23 (2013).

8.   Castro, A. F. & Coresh, J. CKD surveillance using laboratory data from the population-based National Health and Nutrition Examination Survey (NHANES). *Am. J. Kidney Dis.* **53,** S46–55 (2009).

9.   Arora, P. *et al.* Prevalence estimates of chronic kidney disease in Canada: results of a nationally representative survey. *CMAJ* **185,** E417–23 (2013).

10.   Echouffo-Tcheugui, J. B. & Kengne, A. P. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Med.* **9,** e1001344 (2012).

11.   Nitta, K., Okada, K., Yanai, M. & Takahashi, S. Aging and Chronic Kidney Disease. *Kidney Blood Press. Res.* **38,** 109–120 (2014).

12.   Sarnak, M. J. *et al.* Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention. *Circulation* **108,** 2154–69 (2003).

13. Weiner, D. E. Chronic Kidney Disease as a Risk Factor for Cardiovascular Disease and All-Cause Mortality: A Pooled Analysis of Community-Based Studies. *J. Am. Soc. Nephrol.* **15,** 1307–1315 (2004).

14. Sarnak, M. J. *et al.* Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention. *Circulation* **108,** 2154–69 (2003).

15. Van der Zee, S., Baber, U., Elmariah, S., Winston, J. & Fuster, V. Cardiovascular risk factors in patients with chronic kidney disease. *Nat. Rev. Cardiol.* **6,** 580–9 (2009).

16. Schnaper, H. W. Remnant nephron physiology and the progression of chronic kidney disease. *Pediatr. Nephrol.* **29,** 193–202 (2014).

17. Schnaper, H. W. *et al.* A conceptual framework for the molecular pathogenesis of progressive kidney disease. *Pediatr. Nephrol.* **25,** 2223–30 (2010).

18. Peralta, C. a *et al.* Detection of chronic kidney disease with creatinine, cystatin C, and urine albumin-to-creatinine ratio and association with progression to end-stage renal disease and mortality. *JAMA* **305,** 1545–52 (2011).

19. Sud, M. *et al.* CKD Stage at Nephrology Referral and Factors Influencing the Risks of ESRD and Death. *Am. J. Kidney Dis.* 1–9 (2014). doi:10.1053/j.ajkd.2013.12.008

20. Frank, M. *et al.* Estimation of glomerular filtration rate in hospitalised patients: are we overestimating renal function? *Swiss Med. Wkly.* **142,** w13708 (2012).

21. Levey, A. S., Inker, L. a & Coresh, J. GFR Estimation: From Physiology to Public Health. *Am. J. Kidney Dis.* **63,** 820–834 (2014).

22. Koppe, L., Klich, A., Dubourg, L., Ecochard, R. & Hadj-Aissa, A. Performance of creatinine-based equations compared in older patients. *J. Nephrol.* **26,** 716–23 (2013).

23. Stevens, L. a, Coresh, J., Greene, T. & Levey, A. S. Assessing kidney function-- measured and estimated glomerular filtration rate. *N. Engl. J. Med.* **354,** 2473–83 (2006).

24. Tsui, N. B. Y. *et al.* High resolution size analysis of fetal DNA in the urine of pregnant women by paired-end massively parallel sequencing. *PLoS One* **7,** e48319 (2012).

25. Botezatu, I. *et al.* Genetic analysis of DNA excreted in urine: a new approach for detecting specific genomic DNA sequences from cells dying in an organism. *Clin. Chem.* **46,** 1078–84 (2000).

26. Dietrich, D. *et al.* Nucleic acid-based tissue biomarkers of urologic malignancies. *Crit. Rev. Clin. Lab. Sci.* **8363,** 1–27 (2014).

27. Jones, P. a & Liang, G. Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* **10,** 805–11 (2009).

28. Virani, S., Virani, S., Colacino, J. a, Kim, J. H. & Rozek, L. S. Cancer epigenetics: a brief review. *ILAR J.* **53,** 359–69 (2012).

29. Jones, P. a. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–92 (2012).

30. Petronis, A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* **465,** 721–7 (2010).

31. Kalari, S. & Pfeifer, G. P. Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. *Adv. Genet.* **70,** 277–308 (2010).

32. Dawson, M. a & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150,** 12–27 (2012).

33. Cairns, P. Detection of promoter hypermethylation of tumor suppressor genes in urine from kidney cancer patients. *Ann. N. Y. Acad. Sci.* **1022,** 40–3 (2004).

34. Battagli, C. *et al.* Promoter Hypermethylation of Tumor Suppressor Genes in Urine from Kidney Cancer Patients Promoter Hypermethylation of Tumor Suppressor Genes in Urine from Kidney Cancer Patients. 8695–8699 (2003).

35. Kim, W. Y. & Kaelin, W. G. Role of VHL gene mutation in human cancer. *J. Clin. Oncol.* **22,** 4991–5004 (2004).

36. Yu, J. *et al.* A novel set of DNA methylation markers in urine sediments for sensitive/specific detection of bladder cancer. *Clin. Cancer Res.* **13,** 7296–304 (2007).

37. Laird, P. W. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* **11,** 191–203 (2010).

38. Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics Off. J. DNA Methylation Soc.* **6,** 692–702 (2011).

39. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98,** 288–95 (2011).

40. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11,** 587 (2010).

41. Shames, D. S., Minna, J. D. & Gazdar, A. F. Methods for detecting DNA methylation in tumors: from bench to bedside. *Cancer Lett.* **251,** 187–98 (2007).

42. Tost, J. & Gut, I. G. DNA methylation analysis by pyrosequencing. *Nat. Protoc.* **2,** 2265–75 (2007).

43. Rodríguez-Ribera, L. *et al.* Time in hemodialysis modulates the levels of genetic damage in hemodialysis patients. *Environ. Mol. Mutagen.* **55,** 363–8 (2014).

44. Wang, D. *et al.* IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. *Bioinformatics* **28,** 729–30 (2012).

45. Dedeurwaerder, S. *et al.* Evaluation of the Infinium Methylation 450K technology. *Epigenomics* **3,** 771–84 (2011).

46. Aryee, M. J. *et al.* Accurate genome-scale percentage DNA methylation estimates from microarray data. *Biostatistics* **12,** 197–210 (2011).

47. Siegmund, K. D. Statistical approaches for the analysis of DNA methylation microarray data. *Hum. Genet.* **129,** 585–95 (2011).

48. Pearson, T. a & Manolio, T. a. How to interpret a genome-wide association study. *JAMA* **299,** 1335–44 (2008).

49. Lefevre, G. M., Patel, S. R., Kim, D., Tessarollo, L. & Dressler, G. R. Altering a histone H3K4 methylation pathway in glomerular podocytes promotes a chronic disease phenotype. *PLoS Genet.* **6,** e1001142 (2010).

50. Piwkowska, A., Rogacka, D., Kasztan, M., Angielski, S. & Jankowski, M. Insulin increases glomerular filtration barrier permeability through dimerization of protein kinase G type Iα subunits. *Biochim. Biophys. Acta* **1832,** 791–804 (2013).

51. Zhou, W. *et al.* Loss of clusterin expression worsens renal ischemia-reperfusion injury. *Am. J. Physiol. Renal Physiol.* **298,** F568–78 (2010).

52.   Wang, Z. *et al.* TIMP2 and TIMP3 have divergent roles in early renal tubulointerstitial injury. *Kidney Int.* **85,** 82–93 (2014).

53.   Ahmed, A. K., El Nahas, A. M. & Johnson, T. S. Changes in matrix metalloproteinases and their inhibitors in kidney transplant recipients. *Exp. Clin. Transplant.* **10,** 332–43 (2012).

54.   Cheng, C.-J., Baum, M. & Huang, C.-L. Kidney-specific WNK1 regulates sodium reabsorption and potassium secretion in mouse cortical collecting duct. *Am. J. Physiol. Renal Physiol.* **304,** F397–402 (2013).

55.   Papeta, N. *et al.* Prkdc participates in mitochondrial genome maintenance and prevents Adriamycin-induced nephropathy in mice. *J. Clin. Invest.* **120,** 4055–64 (2010).

56.   Hulea, L. & Nepveu, A. CUX1 transcription factors: from biochemical activities and cell-based assays to mouse models and human diseases. *Gene* **497,** 18–26 (2012).

57.   Smyth, L. J., McKay, G. J., Maxwell, A. P. & McKnight, A. J. DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. *Epigenetics* **9,** (2013).

58.   Chin, H. *et al.* Protein kinase A-dependent phosphorylation of B/K protein. *Exp. Mol. Med.* **38,** 144–52 (2006).

59.   Sato, M. *et al.* Renal secretion of uric acid by organic anion transporter 2 (OAT2/SLC22A7) in human. *Biol. Pharm. Bull.* **33,** 498–503 (2010).

60.   Zhdanova, O. *et al.* The inducible deletion of Drosha and microRNAs in mature podocytes results in a collapsing glomerulopathy. *Kidney Int.* **80,** 719–30 (2011).

61.   Barua, M. *et al.* Mutations in the INF2 gene account for a significant proportion of familial but not sporadic focal and segmental glomerulosclerosis. *Kidney Int.* **83,** 316–22 (2013).

62.   Miao, J. *et al.* Newly identified cytoskeletal components are associated with dynamic changes of podocyte foot processes. *Nephrol. Dial. Transplant* **24,** 3297–305 (2009).

63.   Jung, K.-J. *et al.* Involvement of hydrogen sulfide and homocysteine transsulfuration pathway in the progression of kidney fibrosis after ureteral obstruction. *Biochim. Biophys. Acta* **1832,** 1989–97 (2013).

64.   Gunness, P., Aleksa, K., Bend, J. & Koren, G. Acyclovir-induced nephrotoxicity: the role of the acyclovir aldehyde metabolite. *Transl. Res.* **158,** 290–301 (2011).

65.   Yuan, Q. *et al.* Preconditioning with physiological levels of ethanol protect kidney against ischemia/reperfusion injury by modulating oxidative stress. *PLoS One* **6,** e25811 (2011).

66.   Jiang, M. *et al.* Autophagy in proximal tubules protects against acute kidney injury. *Kidney Int.* **82,** 1271–83 (2012).

67.   Suen, J. Y., Gardiner, B., Grimmond, S. & Fairlie, D. P. Profiling gene expression induced by protease-activated receptor 2 (PAR2) activation in human kidney cells. *PLoS One* **5,** e13809 (2010).

68.   Shan, Z., Haaf, T. & Popescu, N. C. Identification and characterization of a gene encoding a putative mouse Rho GTPase activating protein gene 8, Arhgap8. *Gene* **303,** 55–61 (2003).

69.   Johnstone, C. N. *et al.* PRR5 encodes a conserved proline-rich protein predominant in kidney: analysis of genomic organization, expression, and mutation status in breast and colorectal carcinomas. *Genomics* **85,** 338–51 (2005).

APPENDIX

Table 1| Patient Samples and DNA Quantity Input into Initial Methylation Microarray

The following table identifies each the DNA samples used for the methylation microarray 1 (in 2011) and microarray 2 (2012) and the quantity of DNA (ng) of each sample. E1 and E2 refer to Elution 1 and Elution 2 of the DNA extraction process. These elutions were pooled for most samples, except sample 09 Day-3, to increase the total quantity of DNA. Equal amounts of DNA were pooled from samples 07 and 06, in both the Pre and 3M samples.. Equal quantities of DNA were used from the patient-09 Day 0 and Day 3 set as duplicates.

Methylation Microarray	Sample	Quantity of DNA (ng)	Methylation Microarray	Sample	Quantity of DNA (ng)
1	08 PRE (E1 + E2)	658	1	DUPLICATE: 09 DAY 0 (E1 + E2)	312
1	08 3M (E1 + E2)	797	1	DUPLICATE: 09 DAY 3 E1	2834
1	04 DAY 0 (E1 + E2)	651	1	DUPLICATE: 09 DAY 3 E1	2834
1	04 3M (E1 + E2)	581	1	POSITIVE CONTROL	800
1	09 DAY 0 (E1 + E2)	312	2	03 PRE (E1 + E2)	307
1	09 DAY 3 E1	2834	2	03 3M (E1 + E2)	1204
1	07 PRE (E1 + E2) + 06 PRE (E1 + E2)	1066	2	DUPLICATE: 03 PRE (E1 + E2)	307
1	07 3M (E1 + E2) + 06 3M (E1 + E2)	674	2	DUPLICATE: 03 3M (E1 + E2)	1204
1	08 DAY 3 (E1 + E2)	377	2	04 PRE (E1 + E2)	193
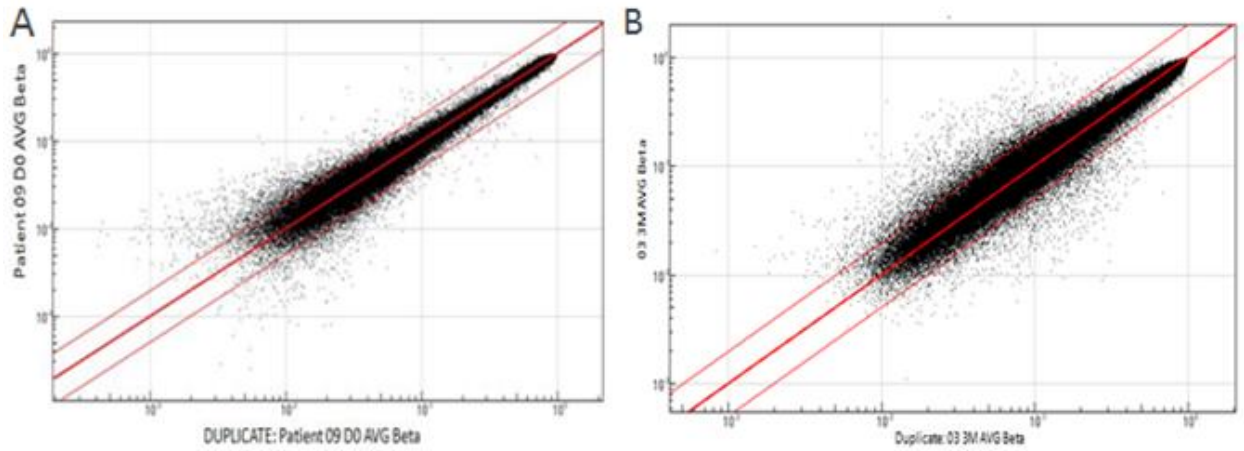
Figure 1 | Scatter Plots of Duplicate Samples in First and Second Methylation Microarrays

A) The following scatter plot demonstrates the average beta value of each detected CpG of Patient 09 Day-0 against its duplicate for the first microarray and **B)** demonstrates the average beta value of each detected CpG of Patient 03 3M against its duplicate within the same chip in the second microarray

Figure 2 | Q-Q Plots of Regional Analysis

The outcome of the statistical analysis of the eleven regions is illustrated in this figure. A Q-Q plot of the A) Exon1, B) Gene Body, C) Island, D) North Shore, E) North Shelf, F) South Shore, G) South Shelf, H) TSS1500, I) 3UTR, J) TSS200, and K) 5UTR are shown.

A) Exon 1 Region

B) Gene Body Region

C) Island Region

D) North Shore Region

E) North Shelf Region

F) South Shore Region

G) South Shelf Region

H) Transcription Start Site 1500 Region

I) 3' Untranslated Region

J) Transcription Start Site 200 Region

K) 5' Untranslated Region
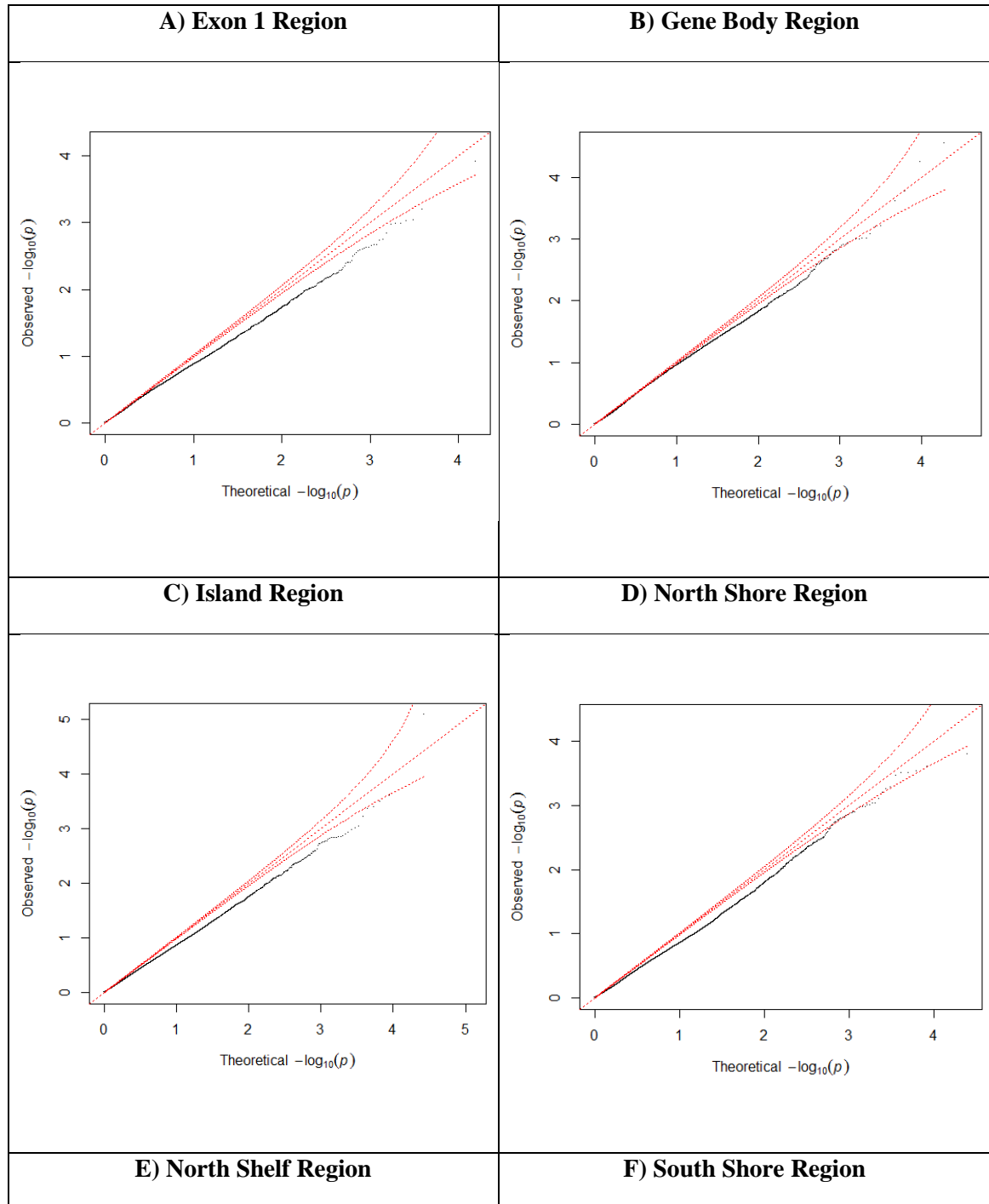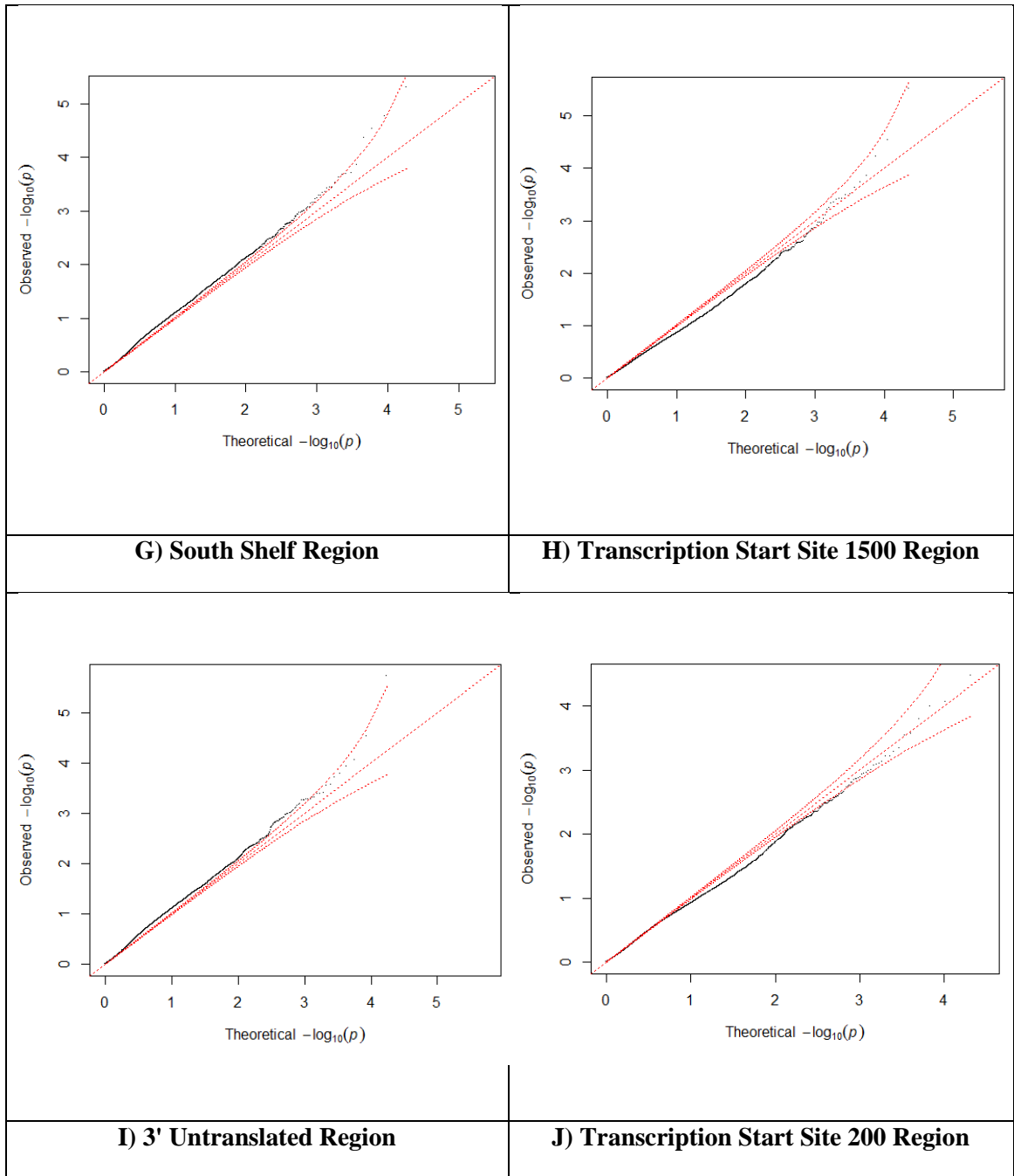
<u>**Table 2**</u> | Designed Primer Sequence Information for Pyrosequencing Validation

The following table lists the primer information, such as type, sequence and PCR product length for the forward and reverse primer for eight of the nineteen probes chosen for validation

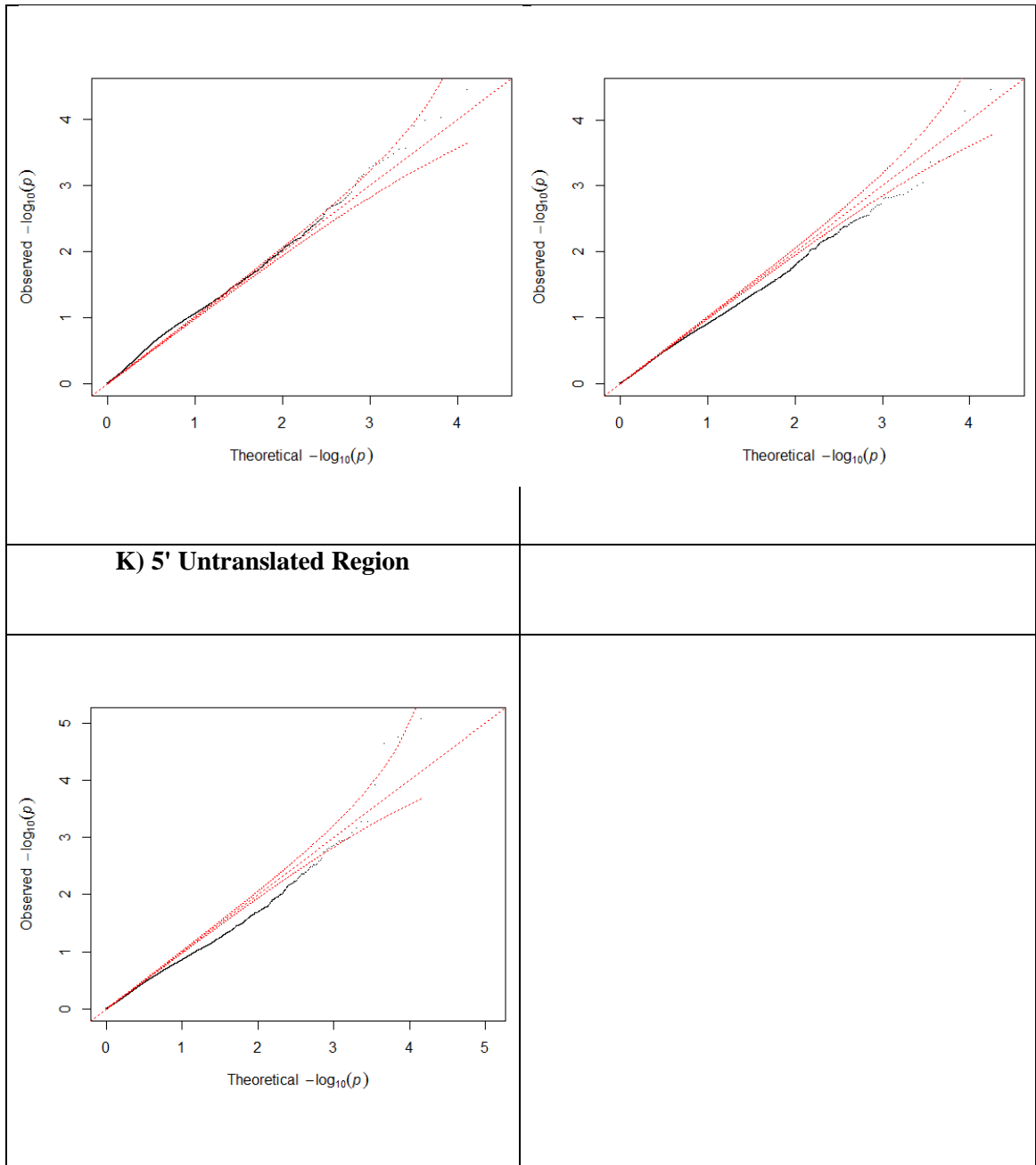| Probe ID	Gene Name	Primer Type	Sequence	PCR product length
cg11108676	NTRK3	Forward	5'GGATTTTAGAAAGATTTTAATGAAGAGTAG'3	183
		Reverse	Biotin-5'AAACATAAATTACTCTAAACAACCAAATA'3	
		Sequencing	5'ATTTTTGAGTAAGAAGAAT'3	
cg08552446	PRKG1	Forward	5'TTTTTAAAGTGTTGGGATTATAGGTATGA'3	83
		Reverse	Biotin-5'ACACCAAAAATAAATCTCATCACCTATTCA'3	
		Sequencing	5'AAGTGTTGGGATTATAGGTATGAGT'3	
cg00928580	CLU	Forward	5'TTGTTTTTGGGTTTTAGTTTTTTTTATTTG'3	112
		Reverse	Biotin-5'AATAACCTTCTAATACTTAACTTAACTCT'3	
		Sequencing	5'TTGGGTTTTAGTTTTTTTTATTTGG'3	
cg15981475	TIMP2	Forward	Biotin-5'TAAGGGGTGTTTTTTGGAGTAG'3	106
		Reverse	5'CCTCCATTTAAAAAAAAAATCCTATCAAT'3	
		Sequencing	5'AAACCACCCCCCCACAAAAAC'3	
cg06081750	WNK1	Forward	Biotin-5'GGGTTAGGAAAGTTTTGTTTATAAGGA'3	107
		Reverse	5'TTCAAAACATATACCAAAAACAAAAACTCC'3	
		Sequencing	5'CCAAAAACAAAAACTCCACCCAC'3	
cg20197825	PRKDC	Forward	5'TTGAGGTAGGTTTGGGATTAG'3	177
		Reverse	Biotin-5'CCAATATCACCTACAAACCAAATTAC'3	
		Sequencing	5'ATTTTTATTAGAGGAGAGAATTGGG'3	
cg23630593	CUX1	Forward	5'AGGTGATTGGAAGGTAGATAGTA'3	189
		Reverse	Biotin-5'ATTCCCAACCATACCTAACACCAAAAC'3	
		Sequencing	5'GGTAGTTGAGGAGTAGGA'3	
cg07654864	STY17	Forward	5'GTGGGTGGTGATAGGAAATATTTTA'3	138
		Reverse	Biotin-5'CCAACCCAAAAAAAACATAATTAACTAACAT'3	
		Sequencing	5'GGAGTAGAATTGGAAGGTAGGGTT'3	

Table 3 | PCR Conditions for NTRK3, PRKG1, TIMP2, WNK1, PRKDC, STY17, CLU and CUX1 Position Amplification

The following table lists the volume and concentrations of the PCR reagents and melting temperatures (°C) used to amplify the NTRK3, PRKG1, TIMP2, WNK1, PRKDC, STY17, CLU and CUX1 Positions. All of these amplicons, except PRKG1, were programmed to heat to 95 °C for 15 minutes, then 50 cycles of 95 °C for 30 seconds, annealing temperature for 30 sec and 72 °C for 10 sec. The samples are then exposed to 72°C or 4 minutes and stored at 4 °C. The ideal amplification cycle of PRKG1 95 °C for 15 minutes, then 50 cycles of 95 °C for 30 seconds, annealing temperature for 30 sec and 72 °C for 30 sec, then 72°C or 4 minutes after the 50 cycles and stored at 4 °C.

Primer Gene Name	NTRK3	PRKG1	TIMP2	WNK1	PRKDC	STY17	CLU	CUX1
Annealing temperature (^{0}C)	55	57.6	57.2	63.8	57.2	65	53.8	53.8
10X PCR Buffer	5 µl	5 µl	5 µl	5 µl	5 µl	5 µl	5 µl	5 µl
MgCl2 Reagent (25 mM)	4 µl	8 µl	4 µl	8 µl	4 µl	6 µl	10 µl	8 µl
dNTP Reagent (1 mM)	4 µl	4 µl	4 µl	4 µl	4 µl	4 µl	4 µl	4 µl
Forward Primer (4uM)	3.3 µl	3 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl
Reverse Primer (4uM)	3.3 µl	3 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl
AmpliTaq Gold DNA Polymerase	0.4 µl	0.5 µl	0.3 µl	0.4 µl	0.3 µl	0.4 µl	0.4 µl	0.4 µl
Bisulphite-converted DNA (20 ng/µL)	2 µl	1 µl	2 µl	2 µl	2 µl	2 µl	2 µl	2 µl
High Purity Water	28 µl	21.5 µl	29.7 µl	25.6 µl	29.7 µl	27.6 µl	23.6 µl	25.6 µl
Final Volume	50 µl	50 µl	50 µl	50 µl	50 µl	50 µl	50 µl	50 µl

A1	A2	A3	A4	A5	A6	A7	A8
NTRK3 DNA 2	PRKG1 DNA2	CLU DNA 2	TIMP2 DNA 2	WNK1 DNA 2	PRKDC DNA 2	CUX 1 DNA 2	SYT17 DNA 2
B1	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**
NTRK3 DNA 5 DUPLICATE 1	PRKG1 DNA 5 (1)	CLU DNA 5 (1)	TIMP2 DNA 5 (1)	WNK1 DNA 5 (1)	PRKDC DNA 5 (1)	CUX1 DNA 5 (1)	SYT17 DNA 5 (1)
C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**
NTRK3 DNA 5 DUPLICATE 2	PRKG1 DNA 5 (2)	CLU DNA 5 (2)	TIMP2 DNA 5 (2)	WNK1 DNA 5 (2)	PRKDC DNA 5(2)	CUX1 DNA 5 (2)	SYT17 DNA 5 (2)

Figure 3 | Trial PyroMark Q24 Plate Layout

The following table is a schematic representation of the PyroMarkQ24 plate setup, which includes a total of eight targets that are being tested o two different DNAs and two duplicate pairs of DNA

Table 4 | Methylation Level and Quality Status of Pyrosequenced Participant DNA

The methylation statuses, in percentage, are shown for all 48 amplified participant DNA, along with a quality status of Passed, Check or Failed.

Participant DNA Sample	Candidate Probe	Methylation Level (%)	Quality
03 Pre	PRKG1	93	Passed
03 3M	PRKG1	95	Passed
06 Pre	PRKG1	95	Passed
06 3M	PRKG1	93	Passed
04 Pre	PRKG1	90	Failed
04 3M	PRKG1	95	Failed
07 Pre	PRKG1	100	Failed
07 3m	PRKG1	98	Failed
08 Pre	PRKG1	99	Failed
08 3M	PRKG1	94	Failed
03 Pre	WNK1	70	Passed
03 3M	WNK1	68	Passed
06 Pre	WNK1	64	Passed
06 3M	WNK1	64	Passed
04 Pre	WNK1	87	Passed
04 3M	WNK1	86	Passed
07 Pre	WNK1	85	Passed

07 3m	WNK1	74	Passed
08 Pre	WNK1	73	Failed
08 3M	WNK1	98	Failed
03 Pre	PRKDC	89	Check
03 3M	PRKDC	87	Passed
06 Pre	PRKDC	88	Passed
06 3M	PRKDC	87	Passed
04 Pre	PRKDC	67	Passed
04 3M	PRKDC	80	Passed
07 Pre	PRKDC	88	Passed
07 3m	PRKDC	81	Passed
08 Pre	PRKDC	86	Failed
08 3M	PRKDC	88	Check
03 Pre	CLU	86	Passed
03 3M	CLU	94	Passed
06 Pre	CLU	94	Passed
06 3M	CLU	85	Passed
04 Pre	CLU	93	Passed
04 3M	CLU	92	Passed
07 Pre	CLU	93	Passed
07 3m	CLU	94	Passed
08 Pre	CLU	90	Check
08 3M	CLU	92	Passed
03 Pre	CUX1	66	Failed
03 3M	CUX1	76	Check
06 Pre	CUX1	70	Check
06 3M	CUX1	75	Failed
03 Pre	SYT17	89	Passed
03 3M	SYT17	85	Passed
06 Pre	SYT17	85	Failed
06 3M	SYT17	92	Passed

Figure 4 | Gel Electrophoresis of NTRK3 Amplicon

The 1.5% agarose gel shows some of the PCR products for the pyrosequencing experiment. The amplification of the NTRK3 target is around 200 bp, in both DNA 2 and DNA 5 (lanes 2 and 3).