New methods clear the dust off old biopsies

RNA sequencing of FFPE kidney biopsies

Øystein Solberg Eikrem

Thesis for the Degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2019

UNIVERSITY OF BERGEN

New methods clear the dust off old biopsies

RNA sequencing of FFPE kidney biopsies

Øystein Solberg Eikrem

Thesis for the Degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defence: 07.06.2019

Year: 2019

Title: New methods clear the dust off old biopsies

Name: Øystein Solberg Eikrem

Print: Skipnes Kommunikasjon / University of Bergen

Contents

List of abbreviations

CA9 Carbonic anhydrase 9

ccRCC clear cell renal cell carcinoma

CKD Chronic kidney disease

DNA Deoxyribonucleic acid

DV200 Percentage of nucleotides being 200 nucleotides or longer

ELISA Enzyme linked immunosorbent assay

FDA Food and Drug Administration

FFPE Formalin-fixed, paraffin-embedded

IHC Immunohistochemistry

IPA Ingenuity pathway analysis

LCM Laser-capture microdissection

MDS Multiple dimensional scaling

NPTX2 Neuronal pentraxin 2

NGS Next generation sequencing

NKBR Norwegian Kidney Biopsy Registry

NRR Norwegian Renal Registry

NTNU Norwegian University of Science and Technology

PCA Principal component analysis

PNCK Pregnancy up-regulated nonubiquitous Calmodulin kinase

RIN RNA integrity number

RNA Ribonucleic acid

RRT Renal replacement therapy

TNM Tumor Node Metastasis staging system

UMOD Uromodulin

Scientific environment

This work was carried out within the Renal Research Group, Department of Clinical Medicine (Klinisk Institutt 1), Faculty of Medicine, University of Bergen, Bergen, Norway.

Collaboration partners were the:

Department of Urology, Haukeland University Hospital, Bergen, Norway.

Department of Pathology, Haukeland University Hospital, Bergen, Norway.

Department of Medicine, Haukeland University Hospital, Bergen, Norway.

Research funding was secured through the University of Bergen, the Regional Health Authorities of Western Norway and from international competitive research grants from Sanofi Genzyme and Shire.

Acknowledgements

Looking back at my years as a PhD student, I am very grateful for the opportunity to immerse myself in scientific work, do research, go to conferences, and to meet gifted colleagues. I am really impressed by many of my fellow students and colleagues. Their dedication and skill are extraordinary. Many of them deserve my warm thanks for their contribution and support. I cannot possibly mention them all, but some persons deserve to be named.

At first, I want to express my gratitude to the best main supervisor I could possibly have. It is still a little bit odd that someone of Professor Hans-Peter Marti merits should come to rainy little Bergen, Norway. He has always been available and very helpful. His door has always been open. He excused himself for the late reply of revisions when he replied to emails sent during night that was answered the next/"same" day. I am impressed by his working capacity and by his devotion to his field of study. His high demands for scientific work have been challenging, and his curiosity has been a great source of inspiration. Co-supervisor Cand. Scient. Trude Skogstrand PhD trained me well in the lab skills needed to perform the work in this thesis. Her instructions, support and encouragement have been very much appreciated. She really deserves warm thanks for her guidance, her interest and supportive comments! I have passed on the detailed and careful precision needed when handling RNA.

I owe a great thanks to esteemed researchers and clinicians in the Renal Research Group. Professor Emeritus Einar Svarstad is a great mentor and has meant a lot to me and was the colleague that recruited me to the nephrology field. Also, several other great researchers in our group like Professor Bjørn Egil Vikse, Assoc. Professor Sabine Leh, Consultant Nephrological pediatrician Camilla Tøndel and Professor Rune Bjørneklett have paved the way and set the standard!

I would like to express gratitude to some of the co-authors Andreas Scherer PhD, Vidar Beisvåg PhD and MSc Arnar Flatberg.

I am very grateful for the highly skilled performance of MSc Sten-Even Erlandsen in the NTNU sequencing facility in Trondheim. Many thanks also to the local genomics core facility at Haukeland with MSc Rita Holdhus as manager.

I am very grateful for the skilled and dedicated help of supreme technicians from the hospital's kidney pathology lab; Bendik Nordanger, Brynhild Johanna Haugen, Nina Holmelid and Tina Dahl. Also from the University technicians I have received a lot of help. Thank you very much Dagny Ann Sandnes and Gry Hilde Nilsen. Without the highly skilled help of Bendik and Sabine to help me in the development of the LCM sectioning protocol the LCM work would have been much harder to achieve.

The Department of Clinical Medicine, Medical Faculty, University of Bergen has been very helpful in providing me with the best possible working environments and administrative help for making all of this possible. Thanks a lot to Jorunn Skei, Nils Erik Gilhus and Kjell Morten Myhr and co-workers for your kind help.

I owe a great thanks to all of the members of the research group. Lea Landolt, Rannveig Skrunes, Thomas Knoop, Ingegjerd Sekse, Rolf Christiansen. The cooperation with Lea Landolt has been excellent and very helpful ever since she also joined the lab.

I am very grateful for being part of a group where so many new and aspiring researchers are educated. The medical students and more recent PhD fellows I have had the pleasure of tutoring throughout the course of my PhD-fellowship have really kept me at my toes! I would like to thank Philipp Strauss, Even Koch, Bjørnar Lillefosse, Sigrid Nakken, Hassan Elsaid, August Hoel, Ole Petter Nordbø, Magnus Farstad, Magnus Granly, Tedd Walther and Tonje Myklebust.

Many thanks to the industrial sponsors Sanofi Genzyme and Shire for your financial support making it possible to set the methods developed in this thesis into motion with many groundbreaking projects to come in the future.

Thanks to Silje Solberg at Dermatology for, after hearing about my project in one of the joint research courses, basically handed me the title of my thesis.

Thanks to all my office colleagues situated in a luxurious suite on top of the laboratory building containing no more than 12 people in one room for making every day a little easier.

I am privileged to have a loving and supportive family. My parents have always encouraged me to study, set an aim and work hard to achieve it. Their understanding and support have been most important. Without their repeated help with understanding the importance of studying hard I might not have been able to study Medicine at all. My mother as well as my sister have taken their PhD's and discussions with them have been very useful.

Most of all I want to thank my wife for her important contributions to my PhD and life in general. She has been managing director and chairman of the board in the family and done much more than her fair share when it comes to all that has to be done in a family of four. I love you and both of our adorable children Nora and Emma!

Abstract

Background and aims: Formalin-fixed, paraffin-embedded (FFPE) tissues are an underused resource for molecular analyses. We wanted to exploit renal biopsies also on the mRNA level to elucidate pathophysiological mechanisms and to ultimately define novel therapies of kidney diseases. The work in this thesis aimed to assess the technical feasibility of RNA sequencing per se and the quality of the respective mRNA data derived from extracted RNA of whole FFPE tissue sections. In the first paper the main aim was to test whether lower quality, partially degraded RNA obtained from archival formalin-fixed and paraffin-embedded (FFPE) renal tissues could serve as appropriate source of material for RNA sequencing. This was approached by testing transcriptome sequencing of RNA from concurrently harvested FFPE and fresh stored kidney biopsies. In the second paper we aimed to validate and expand the first analysis by investigating a second cohort of FFPE kidney biopsies from local ccRCC patients. The secondary aim of this thesis was to assess the technical feasibility and the quality of mRNA data obtained from LCM renal tissues. Further, the aim of the third paper was to evaluate the most appropriate method to extract RNA from FFPE renal tissues and to compare yield and quality of extracted RNA between the different methods with the target of conducting RNA sequencing, especially from LCM glomerular cross-sections.

Methods: Kidney biopsies from resected tissues belonging to patients undergoing nephrectomy were obtained with a 16g core biopsy needle. In paper I, tumor samples and adjacent normal tissue specimens were FFPE or RNAlater® stored. In paper II, only FFPE kidney biopsies were used. In the third paper, FFPE biopsies from rat and human tissues were utilized. In all papers RNA sequencing libraries were built with the newly released Illumina's TruSeq® Access library preparation kit (recently re-named RNA exome kit). Comparative analyses were done using voom/Limma package in R.

Main results: In the first paper we demonstrated that the FFPE and RNAlater® datasets gave comparable numbers of detected genes, differentially expressed transcripts and affected pathways. The average expression and the differentially

expressed genes had very high correlation between the FFPE and RNAlater® stored samples. In paper I and II the detected genes relevant for ccRCC were in accordance with the current literature. The number of detected transcripts in the "discovery/paper I" and "confirmation/paper II" data set gave 8957 and 11,047 detected transcripts, respectively. These data sets shared 1193 of differentially expressed genes. The average expression and the differentially expressed transcripts in both data sets correlated, with R^2 of 0,95 and R^2 of 0,94, respectively.

In the third paper, several kits were eligible for RNA extraction from FFPE tissues from both whole kidney biopsy sections and from LCM samples.

Conclusions: Gene expression data obtained from FFPE kidney biopsies are comparable to data obtained from freshly stored material, thus expanding the utility of archival tissue specimens. Next-generation sequencing expands the clinical application of tissue analyses from FFPE biopsies and gives results well in line with the current literature. RNA can be extracted from archival renal biopsies in sufficient quality and quantity from a single human kidney biopsy section and from around 100 LCM glomerular cross-sections to enable successful RNA library preparation and sequencing using commercially available RNA extraction kits.

List of Publications

Paper I

Transcriptome Sequencing (RNAseq) Enables Utilization of Formalin-Fixed, Paraffin-Embedded Biopsies with Clear Cell Renal Cell Carcinoma for Exploration of Disease Biology and Biomarker Development.
Eikrem O, Beisland C, Hjelle K, Flatberg A, Scherer A, Landolt L, Skogstrand T, Leh S, Beisvag V, Marti HP (2016)
PLoS ONE 11(2): e0149743.
https://doi.org/10.1371/journal.pone.0149743

Paper II

Development and confirmation of potential gene classifiers of human clear cell renal cell carcinoma using next-generation RNA sequencing
Eikrem O, Strauss P, Beisland C, Scherer A, Landolt L, Flatberg A, Leh S, Beisvag V, Skogstrand T, Hjelle K, Shresta A, Marti HP (2016)
Scandinavian Journal of Urology, 50:6, 452-462,
DOI: 10.1080/21681805.2016.1238007

Paper III

RNA extraction for RNA sequencing of archival renal tissues
Landolt L, Marti HP, Beisland C, Flatberg A, Eikrem O (2016)
Scandinavian Journal of Clinical and Laboratory Investigation, 76:5, 426-434,
DOI: 10.1080/00365513.2016.1177660

Reprints of the papers were made with permission from the publishers.

Related papers (not included in the Thesis presentation)

Clear Cell Renal Cell Carcinoma is linked to Epithelial-to-Mesenchymal Transition and to Fibrosis. Lea Landolt, Øystein Eikrem, Philipp Strauss, Andreas Scherer, David H. Lovett,Christian Beisland, Kenneth Finne, Tarig Osman, Mohammad M. Ibrahim, Gro Gausdal,Lavina Ahmed, James B. Lorens, Jean Paul Thiery, Tuan Zea Tan, Miroslav Sekulic & Hans-Peter Marti. *Physiol Rep. 2017;5(11).*

Expanding the utilization of formalin-fixed, paraffin-embedded archives: feasibility of miR-seq for disease exploration and biomarker development from biopsies with clear cell renal cell carcinoma. Philipp Strauss, Hans-Peter Marti, Christian Beisland, Andreas Scherer, Sabine Leh, Arnar Flatberg, Even Koch, Vidar Beisvag, Lea Landolt, Trude Skogstrand, Øystein Eikrem. *Int J Mol Sci. 2018;19(3).*

Fine Needle Aspirates of Kidneys: a promising tool for RNA sequencing in renal allografts. Oystein Eikrem; Tedd C. Walther; Arnar Flatberg; Vidar Beisvag; Philipp Strauss; Magnus Farstad; Christian Beisland; Even Koch; Thomas F. Müller; Hans-Peter Marti. *BMC Nephrol. 2018;19(1):221.*

1. Introduction

1.1 Archival formalin-fixed paraffin-embedded biopsies

For several decades tissue biopsies have been fixed in formalin and embedded in paraffin for long time storage in the archives of the pathology departments and tissue biobanks throughout the world. It is well appreciated that this way of fixing and preserving tissues for downstream analyses is far from optimal. Previously, full transcriptome RNA sequencing from FFPE materials has been considered impossible because of RNA degradation. Liquid nitrogen snap-frozen and further fresh-frozen in minus 80 °C is the gold standard tissue storage method for subsequent sequencing of extracted RNA or DNA. The overall performance and acquired information from an experiment is mainly reliant on the quality of the sample itself, which tissue preservation method used, what kind of nucleotide extraction protocol was applied, and of the chosen sequencing setup. RNA derived from FFPE tissue blocks are more degraded and yields less RNA once extracted. The RNA from FFPE blocks can also be molecularly modified based on cross-linkage of nucleotides to formalin and proteins. Also, other molecular changes like addition of mono-methylol (–CH2OH) especially to the adenin bases can occur (1, 2). Potential differences in the formalin fixation duration time and fixation method as well as the age of the archival tissue samples further add to the variation of RNA quality. The first time DNA was extracted and Sanger sequenced from FFPE tissues was in 1985 in a study of oncogenes (3). After this, many other studies have demonstrated the feasibility of DNA extraction from tissue specimens up to 40 years old (4, 5). Already in 1988 RNA was isolated from tissues by Rupp and Locker (6). They acknowledged the fact that RNA was more difficult to handle because of its fast degradation compared to DNA, yet not impossible as RNA has been isolated from tissues as old as 20 years (7). In the context of getting best possible RNA quality, many factors contribute. Warm ischemia time, autolysis, time before the sample is put in formalin and the time it takes for the formalin to thoroughly fixate the sample. Formalin penetrates the tissues during fixation at a rate of 1mm per hour, which implies that the size of the

specimen is indirectly associated with the yield and quality of extracted RNA (1). Lastly, other issues regarding RNA quality can come from contamination from ubiquitous RNAses that can derive from hardware, skin and surroundings. The automated process of embedding into paraffin together with other samples is not completely sterile, nor is it RNAse-free either (8, 9). Tissue preservation and light-microscopical investigation is the main purpose of the FFPE tissue preservation method. While there are many drawbacks of FFPE tissues for molecular analyses, the distinctive advantage is that the material designated for RNA sequencing can be concurrently investigated by light microscopy. Thus, allowing for excellent specificity for what entity to be further examined with molecular methods. Although of greatest potential, FFPE tissues are an underused resource for molecular analyses. New methods hold great promise and the Illumina TruSeq RNA Access Kit® (recently changed name to RNA Exome kit) released in 2014 is designed to overcome these challenges for RNA sequencing applications by isolating mRNA through a sequence-specific capture protocol resulting in reduced ribosomal RNA and enriched exonic RNA sequences.

1.2 Historical aspects and background of the molecular biology field

Already in 1871, Friedrich Miescher from Basel published the first evidence of a substance that fundamentally differed from proteins. Because of its occurrence within the cells nuclei; he termed the novel substance “nuclein” (10). Almost a century later, work also from others eventually led to the landmark paper, “the structure of DNA”, of much more known Watson and Crick in 1953 (11). Later, the invention of DNA sequencing by Frederick Sanger in the 1970s represented a paradigm shift in the era of modern medicine (12). With this technique it was possible to determine the sequence of nucleotides much faster and more accurate than with previous methods. This method was improved in the decades to follow with the addition of fluorescently stained nucleotides rather than radioactively labelled gels. In 2004, after 14 years of intense work, the Human Genome Project on Sanger sequencing of the entire compiled human genome was published (13). Altogether, paralleled by a revolution

in computer science, this led to several new projects, and finally to the introduction of the contemporary commercial sequencing platforms, which we know as next generation sequencing (NGS) (14-17). With Sanger's DNA chain termination method it was possible to sequence one gene at a time. Contrarily, with NGS it is achievable to sequence the whole epigenome, genome, exome or transcriptome within a few days of work.

Concurrent with the development of DNA sequencing technologies, the evolution of methods to study gene expression took place. Quantitative real-time PCR, micro-array technologies and RNA sequencing are powerful tools for characterization of mRNA abundances. The foundation of this field of research was pioneered by Kjell Kleppe and co-workers who first described the method of replicating a short DNA template with primers in vitro (18). However, the invention of polymerase chain reaction (PCR) is generally accredited to the Nobel Prize in Chemistry awardee, Kary Mullis (19). To date RNA sequencing has largely replaced microarray technologies for elucidating gene expression patterns. Compared with other approaches, like microarray, NGS offers the possibility of detecting novel transcripts that are not based on a priori assumptions. This has led, and will continue to lead to the discovery of potential biomarkers or targets for novel therapies (14, 20).

1.3 Norwegian Kidney Biopsy Registry and Norwegian Renal Registry

The Norwegian Kidney Biopsy Registry (NKBR) was founded in 1988. It holds information about clinical and histopathological parameters of non-neoplastic kidney biopies at the time of the biopsy. Until 2013, there was a designated nephropathologist in Bergen that reanalyzed all biopsies with a second opinion for all biopsies registered. In the last five years however, the diagnosis from the respective pathology departments have been communicated to the registry without a second opinion examination. In 2012 NKBR became a national quality registry. Since 2013 the registry is building a digital slide archive of all registered non-neoplastic kidney

biopsies in Norway. To date the registry contains information of more than 14000 non-neoplastic kidney biopsies. In 2016 the NKBR fused with the Norwegian ESRD and transplant registry and became part of The Norwegian Renal Registry (NRR). The NNR now consists of two sections: Section for dialysis and transplantation and Section for kidney biopsy.

The NRR based in Oslo has been operative since 1980 and contains information about all patients from the time they reach chronic kidney disease (CKD) stage 5, or from all patients at the time they start renal replacement therapy (RRT). Data are reported annually on a standardized form with information about etiology, comorbidities, laboratory findings and treatment and treatment response. In the 2017 annual report from the NRR, the reported number of patients starting RRT was 579 (21).

In Norway all habitants have their own unique 11-digit personal identification number and this number is used in all health-related registries. This makes it possible to link data from the NRR to other registries like the Norwegian National Cause of Death Registry, or the Norwegian Population Registry. Several high-impact publications have arisen from the NRR (22-25).

Altogether the NKBR holds a unique opportunity for research based on the long time period of biopsy inclusion and for the high number of included biopsies. Previous works from the NRR/NKBR have been more clinically-, epidemiologically-, and transplant- oriented (22-25). The work in this thesis expands on the recent developments in the tissue-based research already performed on biopsies from the NKBR. Professor Bjørn Egil Vikse and Kenneth Finne PhD, recently established methods for studying the proteome with *proteomics* from these biopsies (26-30). With my PhD thesis work we aim to establish methods to study the transcriptome with *transcriptomics*.

Figure 1 shows the principle workflows of non-neoplastic kidney biopsies at pathology departments in Norway.

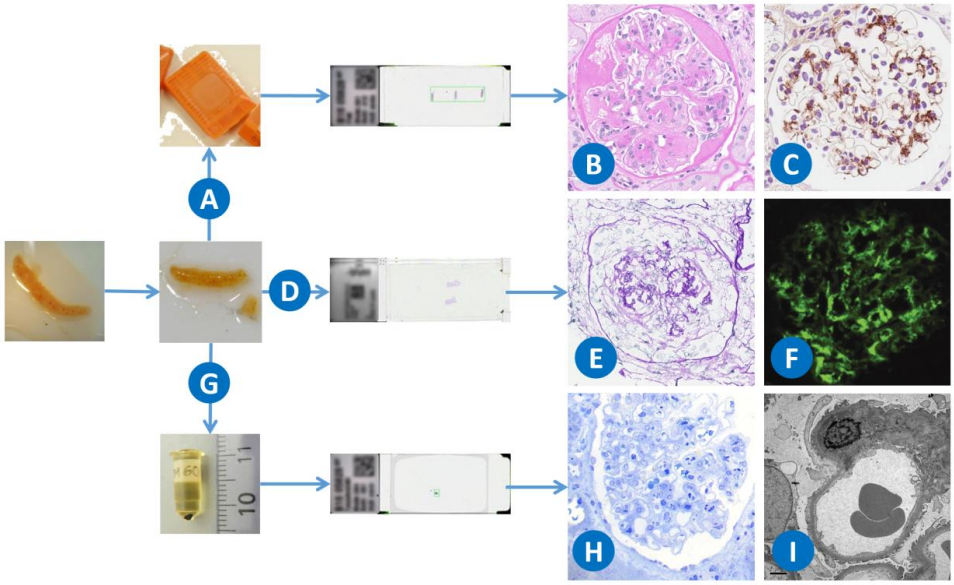

Figure 1 Overview of principle processing steps of kidney biopsies in respective pathology departments.

A) The largest part of the core gets fixed in formalin and embedded in paraffin. (B) shows a typical example of a section from a paraffin blokk with the standard periodic acid Schiff stain (PAS). Sections from paraffin embedded and formalin fixed material can also be used for immunohistochemistry (C).

D) Some of the pathology departments also freeze a part for immunofluorescence. Morphology from frozen sections shows reduced detail (E), but is well suited for immunofluorescense (F) (Reproduced from Koivuviita, N. with permission (31).)

A smaller part of the biopsy is fixed in a special medium and embedded into an epoxy resin. As fixation medium McDowell solution is used, a mixture of formalin and glutaraldehyde. The semithin section from this material is shown in (H, toluidine blue stain) and the ultrathin sectionin (I).

Sections can be made from all types of processed material and can be digitalized for permanent storage (B, C, E, F, H, I).

Reproduced from S. Leh with permission (32).

1.4 Next generation sequencing

In the mid-2000s the Human Genome Project was completed (13). Approximately ten years after its completion the first of many high-throughput sequencing platforms were established. These various platforms are all known as next generation sequencing (NGS), massively parallel sequencing or high-throughput sequencing (33). The key advantage of the NGS technology is the massive parallelization of millions of reactions simultaneously (33). Thousands of clonally amplified DNA fragments in a defined area ensures a strong base-calling signal (33). Continuous advancements of the NGS technology over the last decade have led to an additional 100-1000 fold increased capacity since its release (34). The release of NGS heralded a 50 000 fold drop in the cost of a human genome since the Human Genome Project (35). Prices have continued to drop and in the late 2015 the 1000 $ barrier for a whole human genome was breached (33).

1.5 Clear cell Renal Cell Carcinoma

We chose clear cell renal cell carcinoma primarily because of tissue availability to set up the RNAseq method in FFPE tissues. However, there is concurrently a lack of stringent diagnostic and prognostic blood-based panels in clear cell renal cell carcinoma (ccRCC) and novel therapies for advanced stages are urgently needed. In addition, we received extensive local expertise and collaborative help from our Deptartment of Urology. Therefore, ccRCC was a good model disease to develop RNAseq from FFPE tissues, as described in more detail below. Due to our later success, we have continued the ccRCC research until this date.

Specifically, ccRCC is the most frequent primary renal neoplasm with both increasing incidence and considerable morbidity and mortality (36, 37). Renal cell cancer ranks among the ten most frequent cancers in women and men (38). In 2017 there were approximately 64 000 new cases and 14 400 deaths attributed to kidney

cancer in the United States (39). In Norway 869 patients have been diagnosed with renal cell cancer during 2017 causing a total of 245 deaths (40).

ccRCC only has a favorable prognosis if it is diagnosed once the disease is still localized (41, 42). At this stage it is curable with early surgical intervention alone. However, up to 20% with initially localized disease develop metastases after five years (41). Only half of the patients with locally advanced disease are alive after five years of follow-up (43). Long term survival rate of metastatic disease is extremely poor (44). Even small tumors (1–2 cm) have metastatic potential (45, 46). Currently, no established biomarker for renal cell carcinoma is in use in clinical practice, despite intensive efforts (47, 48). Therefore, elucidation of the molecular mechanisms of this disease is important. We need to unravel prognostic and predictive markers as well as potential novel drug targets.

1.6 Biomarkers

To test the deliverables from the ccRCC RNAseq data we investigated the diagnostic properties of some of the known biomarkers in this field. Some of these biomarkers might also be linked to prognostic or predictive markers as well as potential novel drug targets. The "*Biomarkers Definitions Working Group*" put the following definition into words: "A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, patho-genic processes, or pharmacologic responses to a therapeutic intervention" (49). Biomarkers can be utilized as a diagnostic tool, as a tool for staging of disease, as an indicator of prognosis and as a tool with the capacity of predicting treatment response following an intervention. Yet, the bench to bedside translation takes a long time and there are only a very limited number of biomarkers that have been incorporated into clinical practice (50, 51). This is also the case for renal cancer, where there are no FDA-approved biomarkers for renal cancer (52). Still, both vast resources and intense efforts are invested in this field of research and with better bioinformatics approaches, careful selection of study

inclusion and external validation of potential markers, better markers will come in the future (53).

1.6.1 Predictive biomarkers

Predictive biomarkers hold the capacity to anticipate the response to a therapeutic intervention. The addition of a monoclonal antibody targeting HER2 in patients with HER2 overexpressed breast cancer serves as a classical example of a marker that can predict treatment response (54).

1.6.2 Prognostic biomarkers

Prognostic biomarkers have the capacity to forecast the natural disease course. Prognostic biomarkers are unable to predict the response to a specific therapy, yet they might be helpful in guiding treatment so that high-risk patients will get aggressive treatment and reduce overtreatment in low-risk patients (55). There are several examples of prognostic biomarkers like CA 125 levels in ovarian cancer and PSA levels in prostate cancer (56, 57).

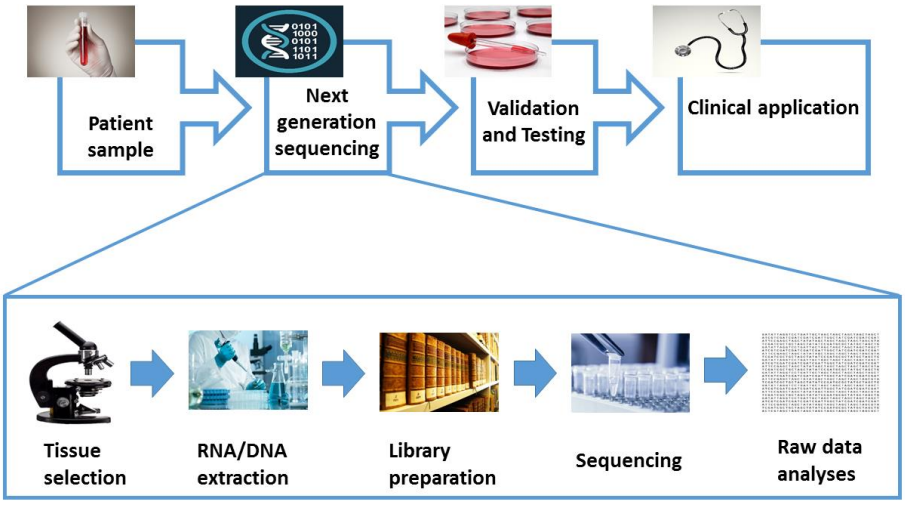

Figure 2. The upper section displays the application of next generation sequencing (NGS). The lower section illustrates the general workflow of NGS. Reproduced from EMJ Reviews with permission (58).

Although there are no established blood sample prognostic biomarkers in clinical use in renal cancer, there are several good clinical and histopathological scoring algorithms that perform well (42, 59-63). The TNM staging system gives prognostic information based on the anatomical characteristics and distribution of the disease (64). Lower stage disease has much more favorable prognosis than higher stage, at least compared to nodal positive and systemic metastatic disease.

1.7 Laser-capture microdissection

Laser-capture microdissection (LCM) is a method that can be used to cut out specific cells or tissue compartments under direct visualization for further molecular analyses (65). Previous attempts to capture specific cells or tissues were performed by protection of areas of interest by covering with pigments and deactivating all unprotected DNA through application of short wave ultraviolet light (66). Instead of studying the gene expression from a whole kidney biopsy section, where differences in specific nephron compartments might go under the radar, with LCM you can study precisely and specifically the desired tissue area (e.g. glomeruli). Since FFPE tissues are so easily available, development of a method to perform RNA extraction from FFPE LCM tissues will be highly valuable. This will have the capacity of enabling downstream molecular analyses independent of fresh tissues and therefore unlocking pathological archival tissues for these kinds of molecular analyses.

2. Hypothesis and aims of the thesis

2.1 Hypothesis

Our prime hypothesis was that RNAseq of whole tissue sections and microdissected nephron compartments from archival FFPE kidney biopsies is feasible. Furthermore, we also hypothesized that RNAseq of FFPE tissues leads to the delivery of quality data allowing in-depth data mining for biomarker and drug target programs.

2.2 Rationale

Our rationale for performing this project is the high potential of FFPE kidney biopsies for the detection of pathophysiological mechanisms and ultimately of the definition of novel therapies of kidney diseases using omics-based technologies. Due to very recent technical advances in the field *(most notably RNA Exome kit, Illumina),* the feasibility of RNAseq suddenly appeared to be very likely. Furthermore, for our studies we can exploit the local ccRCC biobank and later the NKBR. We are in a strong position to both test this and make high-impact, sub-sequent studies based on the NKBR that can be connected with a spectrum of outcome registries, e.g. comprising long follow-up periods.

2.3 Aims

2.3.1 Main aims

The overall primary aim of this thesis was to assess the technical feasibility of RNA sequencing per se and the quality of the respective mRNA data derived from extracted RNA of whole tissue FFPE sections.

In the first paper, the main aim was to test whether lower quality, partially degraded RNA obtained from archival formalin-fixed and paraffin-embedded renal tissues could serve as appropriate source of material for RNA sequencing. This was approached by testing transcriptome sequencing of RNA from concurrently harvested

FFPE and freshly stored kidney biopsies (RNAlater®-fixed prior to -80°C). In the second paper we aimed to validate and expand the first analysis by investigating a second cohort of FFPE kidney biopsies from local ccRCC patients. These additional samples served two purposes: to provide evidence for the reproducibility of RNAseq data from FFPE samples, and to serve as a validation set for biomarker development.

2.3.2 Secondary aims

The secondary aim of this thesis was to assess the technical feasibility and the quality of mRNA data obtained from LCM renal glomeruli. Accordingly, the aim of the third paper was to evaluate the most appropriate method to extract RNA from FFPE renal glomeruli tissues and to compare yield and quality of extracted RNA between the different methods. Furthermore, we wanted to demonstrate if the extracted glomerular RNA is of sufficient quality and quantity for potential subsequent RNAseq.

3. Materials and methods

3.1 Subjects

3.1.1 Paper I

Patients (n=16) from Haukeland University Hospital diagnosed with ccRCC which underwent partial (n = 10) or full (n = 6) nephrectomy were included chronologically. I have harvested perioperative biopsies directly after the specimens have been surgically removed. Patients had a mean age of 58.2±6.8 years (3 females and 13 males). Patients had pT tumor stages T1a (n = 10), T2a or b (n = 2) and T3a or b (n = 4).

3.1.2 Paper II

Patients (n=12) from Haukeland University Hospital with ccRCC undergoing full (n=8) or partial (n=4) nephrectomy were included. Similarly, perioperative biopsies were harvested directly after the specimens were operated out. Patients had a mean age of 56.9 ± 6.8 years (seven females and five males). Patients had pT tumor stages T1a or b (n=7), T2a or b (n=2) and T3a or b (n=3).

3.1.3 Paper III

Biopsies with non-tumorous, 'normal' tissue adjacent to the tumor from two human patients (one male and one female) undergoing nephrectomy due to ccRCC were obtained in the operating room in June 2015.

Healthy, normal FFPE renal tissue was collected from two male Wistar Hannover rats from Taconic (Ry, Denmark), used in a prior study (67). The FFPE tissue blocks were approximately 3.5 years old.

3.2 Ethical permissions

The regional ethics committee of Western Norway/ Institutional review board (IRB) has approved all of our studies (REC West no. 78/05). All participants provided written informed consent before enrollment. The studies were performed in compliance with the Declaration of Helsinki (2002) and Good Clinical Practice guidelines. Concerning paper III; the experiments were performed also under the approval of the Norwegian State Board for Biological Experiments with living animals.

3.3 Kidney biopsies

All human biopsies were obtained by me with a 16G core biopsy needle for all studies. Biopsies were collected perioperatively just after the specimen was operated out. The 16 G core biopsies were put directly into formalin or RNAlater®(Qiagen, Germay). The formalin biopsies were embedded in paraffin the next day. The RNAlater® stored biopsies were transported to the -80°C freezer and stored upon usage. For paper I; a total of four biopsies from each patient were stored and utilized. Each individual pair of histologically-confirmed clear cell renal cell carcinoma (ccRCC) and adjacent non-tumorous ("normal") tissue were either stored as FFPE tissue or fresh-frozen in an RNA-stabilizing agent (RNAlater®). For paper II and III, only the formalin-fixed biopsies were utilized. For paper III rat whole kidney sections were also used.

3.4 RNA extraction

The FFPE tissues were cut into 5 µm and 10 µm sections for rat and human tissue, respectively. Kidney biopsy sections were cut by Bendik Nordanger on an ordinary rotary microtome (Leica RM 2155, Leica Microsystems, Wetzlar, Germany). All sections were directly inserted into RNAse-free Eppendorf tubes prior to

deparaffinization and RNA extraction (Eppendorf, Hamburg-Eppendorf, Germany). I extracted Total RNA with the miRNeasy FFPE kit (Qiagen) for the formalin-fixed samples (paper I and II) and with the miRNeasy micro kit (Qiagen) for the fresh-frozen samples (paper I). In paper III, numerous different FFPE RNA extraction kits were utilized, including the above mentioned. First, seven different RNA extraction kits were tested on rat tissue. Later, the four kits with the highest amount of RNA extracted from rat whole kidney sections were used to extract RNA from human renal tissues from both whole sections and laser-microdissected tissues.

3.5 RNA concentration and quality

Total RNA concentration was measured using Qubit RNA HS Assay Kit on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and with the NanoDrop spectrophotometer (NanoDrop ND-1000, NanoDrop Technologies, Wilmington, NC, USA). The NanoDrop also reports the absorbance ratios at 260/280 nm and 260/230 nm which reflects the purity of the nucleic acids measured.
RNA quality was assessed using Agilent RNA 6000 Nano Kit on a 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA) and the percentages of RNA fragments larger than 200 nucleotides were calculated. Illumina's guidelines state that samples with DV200 below 30 % are too degraded for further processing. Samples with DV200 of 30-50 % are of low quality and samples with 50-70 % are of medium quality. Lastly, samples with DV200 of over 70% are of high quality. The Bioanalyzer also computes the RNA integrity numbers (RIN). A RIN of above 7 is usually considered to be of high quality when considering fresh-frozen samples.

3.6 cDNA library preparation and sequencing performed at the Norwegian Genomics Consortium, NTNU

cDNA libraries were prepared utilizing the TruSeq RNA Access library kit (Illumina, Inc., San Diego, CA, USA) according to the manufacturer`s instructions. This was performed at the Norwegian Genomics Consortium (Oslo, Trondheim and Bergen) at NTNU by Vidar Beisvåg, Arnar Flatberg and Sten Even Erlandsen based on their previous experience with FFPE tissues. Libraries were validated using the Agilent DNA 1000 kit on a 2100 Bioanalyzer instrument. Specific exome capture probes were used for the first hybridization step. With 200 ng of each DNA library, a 4-plex pool of libraries was made. Probes hybridized to the target regions were captured with streptavidin coated magnetic beads. Afterwards a second round of hybridization and capture were performed to ensure high specificity of the capture regions. Thereafter a thorough wash procedure to remove non-specific binding from the beads was carried out. Finally, AMPure XP beads were used to clean up the libraries prior to PCR amplification. The libraries were quantitated by qPCR using the KAPA Library Quantification Kit—Illumina/ABI Prism1 (Kapa Biosystems, Inc., Wilmington, MA, USA) and validated using Agilent High Sensitivity DNA Kit on a Bioanalyzer. Library were normalization to 22 pM and single read sequencing was performed for 50 cycles on a HiSeq2500 instrument (Illumina, Inc. San Diego, CA, USA). Base calling was done on the HiSeq instrument by RTA 1.17.21.3. FASTQ files were generated using CASAVA 1.8.2 (Illumina, Inc. San Diego, CA, USA).

3.7 Statistics and NGS Data Processing (In collaboration with A. Flatberg and A. Scherer)

Sequencing reads were aligned to the Human genome assembly GRCh38 guided by Tophat and Bowtie. Genes with more than 15 counts per million (cpm) in more than 8 samples per dataset were set as an empirical expression filter. Differentially expressed genes were defined as Benjamini-Hochberg adjusted p-value <0.05 with an absolute fold change of >2 using the voom/Limma R-package (R Bioconductor v 3.4;

available online: www.bioconductor.org). Pathway analysis was performed with Ingenuity Pathway Analysis (Qiagen, USA; version 24718999). The Ingenuity Knowledge Base information was used as reference set. Canonical pathways were sorted by smallest Benjamini-Hochberg-adjusted p-values.

3.8 Histology and Immunohistochemistry

Immunohistochemistry was performed on 4 µm thick FFPE sections from the tumor and adjacent non-tumorous tissue. For positive controls, tissues with known positive reactivity were used, for negative controls the primary antibody was omitted. Stained slides were scanned with the Aperio ScanScope® XT system (Leica Biosystems Imaging, Wetzlar, Germany) at ×40 objective magnification and viewed in ImageScope 12 (Leica Biosystems Imaging, Wetzlar, Germany).

3.9 ELISA for CA9 Serum Levels (Paper I)

CA9 serum concentrations of 38 patients were measured using the Quantikine Human Carbonic Anhydrase IX Immunoassay (R&D Systems, Minneapolis, USA, catalogue number DCA900) according to instructions of the manufacturer, but with an overnight incubation at 4°C after having added the serum. Results were assessed with the Kruskal-Wallis and Dunn's test.

3.10 LaserCapture Microdissection (LCM) of glomerular cross-sections

A new protocol for Laser Capture Microdissecion (LCM) from tissue block to samples ready for RNA extraction was designed. Tissue sections (5-10 µm thick) were mounted on nuclease-free Membrane Slides NF 1.0 PEN (Zeiss, Oberkochen,

Germany). Following deparaffinization with xylene and standard alcohol series, staining was performed with Hematoxylin and Eosin (HE) according to a shortened procedure under RNAse-free conditions. Please confer appendix for complete overview of the tissue sectioning and preparation protocol.
The glomeruli were microdissected using a PALM Laser-Microbeam System (PALMVR Robo software V 2.2.2, P.A.L.M, Bernried, Germany).
Microdissected glomeruli were captured into the lid of RNAse-free Safe-Lock Eppendorf tubes (Eppendorf, Hamburg-Eppendorf, Germany). Several different volumes within the lid was tested before it was decided that 65µl of the lysis buffer from the High Pure RNA extraction kit worked optimally for capture of the material without risking the droplet in the lid to fall out. The remaining 35 µl of lysis buffer was added upon RNA extraction.

4. Summary of main results

4.1 Results paper I

Each of the 16 patients donated four core biopsies, which included two with ccRCC and two from adjacent non-affected "normal" tissue. Each individual pair of ccRCC and "normal" tissue was stored either as FFPE *or* put in RNAlater® and fresh-frozen. This paired design facilitated the evaluation of the impact of storage condition on expression profiles using RNAseq.

The mean RNA integrity number (RIN) and mean DV200 values (95% CI) were 5.7 (5.10–6.30) and 61% (58–64) for RNAlater® samples and 2.53 (2.33–2.73) and 75% (72–79) for FFPE samples, respectively.

We detected a similar number of genes, which passed the expression filter in the FFPE (n=9164) and the RNAlater® (n= 9205) dataset. Around 94% of these genes (n = 8893) were common to both datasets and the correlation of the logarithmic fold change was $R^2 = 0.93$, and correlation of the average expression $R^2 = 0.97$ (As shown in Fig. S1, paper I). These two datasets shared 1106 differentially expressed genes, which correlated with an R^2 of 0,96 (Fig.2, paper I). In a multidimensional scaling (MDS) plot (Fig.1, paper I) it was clear that the samples segregated by diagnosis, rather than by storage condition (FFPE or RNAlater®).

The genes with the highest fold change in both formalin-fixed paraffin-embedded and RNAlater® dataset were Uromodulin (UMOD, -183.2 fold change in tumor versus normal in FFPE dataset, -158.7 fold change in tumor versus normal in RNAlater® dataset), neuronal pentraxin-2 (NPTX2, 140.9 fold change in tumor versus normal in FFPE data set and 220 fold change in tumor versus normal in the RNAlater® dataset) and carbonic anhydrase 9 (CA9, 121.2 fold change in tumor versus normal in FFPE and 304 fold change in tumor versus normal in RNAlater® dataset). Immunohisto-chemistry was used to confirm protein presence of UMOD, NPTX2 and CA9.

Later, pathway analysis revealed TGFB1 as an important expression regulator of 237 genes (17% of differentially expressed genes) in the datasets. Thus, epithelial-to-mesenchymal transition was linked to clear cell renal cancer as markers such as

Vimentin (Vim), Endothelin 1 (EDN1) and Fibronectin (FN1) were up-regulated, whereas epithelial markers such as E-Cadherin (CDH1), epithelial-cell adhesion molecule (EPCAM) and inhibitors of epithelial-to-mesenchymal transition like Grainyhead-like 2 (GRHL2) were down-regulated.
RNA sequencing results from formalin-fixed paraffin-embedded biopsies could also be used for tumor classifier analysis. Using the example of the known clear cell renal cancer biomarker CA9, our samples could be classified in tumorous and normal tissues with a sensitivity and specificity of 93.8%.

4.2 Results paper II

The main aim of this paper was to study the reproducibility of the previously published gene expression analysis from FFPE ccRCC and normal biopsies.
This paper included 12 adult patients and each of the patients donated two core FFPE biopsies, one from tumor and one from adjacent non-affected "normal" tissue.
The mean DV200 value for the samples in this study was 54% (95% CI of 48–61%).
In this investigation 11047 mRNAs passed the expression filter. Around 98% of the 8957 detected genes in the discovery data set (paper I) were common to both FFPE the discovery and confirmation (paper II) data set. The correlation of the average expression was R^2=0,96 and the correlation of the logarithmic fold change was R^2= 0.89. The discovery data detected 1367 differentially expressed genes. The confirmation data set had 2176. These two data sets shared 1193 differentially expressed genes. The correlation of the average expression of these 1193 genes was R^2=0,95. The log2-fold changes of these differentially expressed genes correlated by R^2 =0,94. In a principal component analysis (PCA) comparing the common differentially regulated genes in paper I vs paper II, the samples segregated by biological condition and not by group affiliation; discovery or confirmation set. Comparing the 20 most up- or downregulated genes the results from paper I and paper II were highly concordant.

Again IPA revealed TGFB1 as an important regulator of gene expression in the confirmation data set. TGFB1 itself was overrepresented 3.1-fold (2.8-fold in the FFPE discovery data set). Also, high concordance with regards to the detected pathways were found between paper I and paper II.
The classification with the K nearest neighbor algorithm was used as the training set and the confirmation data set as the test set. In this scenario, the 24 confirmation samples were stratified with 100% accuracy into tumor or normal samples. When the data sets were swapped and the confirmation data were used as the training set to stratify the discovery samples, 30 out of 32 samples were assigned correctly. Two samples in the discovery study that were misclassified, had either admixture of tumor tissue in a normal sample, or a tumor sample with some adjacent tissue that had been judged to be normal. The KNN algorithm with leave-one-out internal cross-validation and unsupervised feature selection showed that TNFAIP6 classified almost all samples correctly in both the confirmation and discovery set. TNFAIP6 was overrepresented in the tumor samples compared to normal samples in both mRNA and also in the immunohistochemical stainings.

4.3 Results paper III

At first, seven different FFPE RNA extraction kits were tested on kidney rat tissue with both quantitative and qualitative analyses. All kits extracted sufficient amounts of RNA, above the required minimum of 30–100 ng RNA for RNA sequencing with the Illumin RNA Access library preparation kit, from a single rat whole-kidney section. From LCM tissues, we found that around 100 LCM glomerular cross-sections were sufficient for RNA sequencing according to the requirements of the Illumina Access library preparation protocol. The four kits that gave the best results regarding RNA yield and RNA quality were selected to be tested also on human tissues. These four kits were the High Pure kit, miRNeasy, RNeasy and the ExpressArt kit. Both whole kidney sections and laser microdissected samples were tested. All of these four kits yielded enough RNA from a single human kidney biopsy section measured by both NanoDrop and Qubit to enable RNA sequencing.

The High Pure FFPE kit and the ExpressArt kit extracted RNA of high quality from LCM glomerular cross-sections and human kidney biopsy sections.
Further on, we tested RNA extracted from six LCM human glomerular cross-sections samples using the High Pure, ExpressArt and the miRNeasy FFPE kit for the library preparation according to the TruSeq® RNA Access Library Preparation Kit protocol. The Access libraries were sequenced on an Illumina NS500 flowcell with 75 basepair single read. Both library preparation and sequencing were evaluated as successful for all of the samples, based on the quality control of the libraries and the sequencing reaction. The High Pure kit had an average amount of reads per sample of 24.1M. This was twice the amount of the two other kits which had 12.3M for the miRNeasy kit and 10.6M for the ExpressArt kit. The number of reads mapped to the genome and the transcriptome were in average 20.8M for the High Pure kit, 10.1M for the miRNeasy kit and 9.1M for the ExpressArt kit. Percentages of duplicate reads were similar in all samples.

5. Discussion

5.1 Methodological considerations

To adequately test the new methods of cDNA library preparation for FFPE archival biopsies, we needed a gold standard to compare with. We therefore chose RNAlater®-fixation and storage in -80°C as the comparator. RNAlater® is considered a good RNA stabiliser and studies show that RNA yields and gene expression results with RNAlater® are comparable to those obtained using fresh-frozen tissues (68, 69). In addition, in paper I of this thesis we collected all samples prospectively in a pairwise fashion of which each individual pair of tumor and normal samples were both FFPE-stored and RNAlater® fixed and frozen. By the time we planned these studies, there were no in-depth report yet comparing matched RNAlater® and FFPE stored samples for RNA sequencing. Another study have demonstrated success with the use of RNA sequencing in FFPE compared to fresh-frozen material from a ribosomal depletion cDNA library protocol (70). Although this method works well, also on FFPE tissues, it requires approximately four times as much sequencing effort compared to mRNA sequencing to achieve the same amount of gene detection (71). When it is sufficient to study the coding regions opposed to intergenic or non-coding regions, the RNA exome kit (Illumina) (formerly TruSeq RNA access kit) provides a highly reliable and cost-effective method. Others have also investigated the effect of storage time in up to 10 year old biopsies in FFPE and the feasibility in mRNA expression experiments. Both microarrays and RNAseq investigations have been demonstrated (72-74). We have also evaluated some of our own kidney biopsies from the NKBR that are up to 30 years old. Based on the RNA quantity and quality measured by DV200 in some of our still unpublished data, their suitability for RNAseq have been demonstrated. The use of a capture-specific protocol for the coding regions is further supported by a high impact publication in Genome Research, where accurate estimates of RNA abundance, uniform transcript coverage and broad dynamic range were found investigating FFPE and flash frozen cancer tissues (75). But for genome-

wide detection of novel transcripts, whole exome enrichment of RNA could be necessary in addition (76).

In this work, there are of course also some limitations. We tested only the same library preparation kit in all comparisons. We could have used a more standard poly-A capture kit (e.g. the TruSeq Stranded mRNA kit from Illumina) for the fresh frozen samples. This could have rendered even more sequencing reads in this group. The number of samples could have been much higher, although at a higher cost. The power calculations performed in paper I, did however claim that the number of included samples were sufficient to achieve a power of 0,85 with an alpha of 0,05. Another limitation was the use of RNAlater® fixation and storage at -80°C. I believe that we might have had an even better gold-standard comparator with the use of snap-freezing instantaneously in liquid nitrogen before long time storage in -80°C.
Also, the issue of warm ischemia time and the impact of RNA degradation prior to the tissues being biopsied were not systematically assessed in this work. Still, the perioperatively collected biopsies in these studies are more controlled than biopsies in the archives of the pathology departments based on the varying delay before the samples are put in formalin.
Further methodological considerations are also discussed in each of the following paragraphs related to each of the papers.

5.2 Discussion of the main results

5.2.1 Discussion paper I

These proof-of-concept studies have shown that it is possible to sequence RNA from previously considered useless RNA. Still, the RNA quality and its implications together with the verification of biological findings will be discussed here. The quality of the RNA samples was determined with the Agilent RNA integrity number (RIN) and the DV200 number (77). RIN is widely accepted as a good RNA quality

measure for gene expression analysis (77). When it comes to FFPE samples, RIN is not a sensitive measure of RNA quality nor a reliable predictor of successful library preparation (78). This is probably because most FFPE samples have a RIN-value of only 1-3 out of a scale from 1-10. Thus, RIN is not a sensitive measure for distinguishing between "poor" or "good" RNA quality between different FFPE samples. Therefore, previous investigations have used mean RNA fragment size as a determinant of RNA quality when working with RNA obtained from FFPE tissues (79-81). It has been demonstrated that high-quality libraries can be prepared from low-quality FFPE samples with a DV200 value as low as 30% (79).

The DV200 numbers achieved in our three studies were therefore of adequate quality to be tested with the newly relased RNA Access library preaparation kit. Although the FFPE biopsies' quality was of sufficient quality, we had to verify that we could attain similar biological results from the FFPE and the RNAlater dataset. In the first paper, we achieved high similarity of the two datasets indicating that archival FFPE-samples can be used in coming studies. We had 94% overlap of the transcripts that passed passing the expression filter in the FFPE and RNAlater® sample groups, 80% of differentially regulated genes were in common, and 75% of the differentially affected pathways were present in both datasets. We could have got even higher numbers of similarity, but the differences in gene expression can probably be explained by the cell-composition of the respective biopsies. There is a well-described intra-tumor heterogeneity in renal cancer (82). The remainder of the difference could be explained by difference in the RNA quality between the FFPE and the RNAlater® dataset.

Importantly, beyond the numerical values of the similarities, biologically relevant information well in line with the literature between normal and tumor biopsies have been found. Three of the highest differentially expressed genes in tumor vs. normal in both data sets were upregulation of CA9 and NPTX2 and downregulation of UMOD. CA9 has been extensively investigated for its capacities as a diagnostic biomarker in ccRCC (83-87). It has a very high diagnostic accuracy in solid tumors (83). Also in preoperative biopsies <4 cm, a recent publication has demonstrated that, 25/25 tumor biopsies were CA9 positive on RT-PCR and 31/34 on immunohistochemistry (IHC)

(84). This has also been confirmed with IHC positivity from all of our stained cancer samples. CA9 is also in some settings a good predictive biomarker of outcome (87, 88). Following anti-VEGF therapy, increasing levels of CA9 after treatment are associated with a better prognosis (87). Several studies have also evaluated other biomarkers for predictive measures, especially in metastatic RCC, although some positive results have been demonstrated, further validation is needed (89-92). We have also demonstrated good concordance with microarray gene expression studies of ccRCC, where upregulation of one of our top regulated molecules were found; NPTX2 is also well in line with the current literature (93). We found that 17 of the 20 genes with the biggest absolute fold-change in the microarray meta-analysis also were differentially regulated in the NGS datasets. One limitation and uncertainty in this comparison is still the large discrepancy in the fold changes detected in the microarray studies (Table 4, paper I), and from the fact that all genes in the Table 4 were differentially expressed in just two or three of five of these microarray studies. It has been discussed earlier that NGS has a wider dynamic range, giving more accuracy of the abundance of reads either lowly or highly expressed. Microarrays can reach a certain threshold of the highly expressed genes (94). Regardless, our data more or less verify most of the gene expression changes found in microarray studies of ccRCC.

To further strengthen the evidence of the similarities between the RNAlater® stored and the FFPE stored biospies, we carefully reviewed an important and specific signaling pathway (Figure 4, paper I) and reported the fold changes of the different datasets for each of the different molecules of the VEGF/NOTCH/DLL4 signaling cascades (95, 96). There is a striking similarity in the fold changes of the different signaling molecules important for some of the molecules in this pathway in the FFPE vs. the RNAlater® datasets.

TGFB1 was the most significant gene *regulator* in our study (Figure 5, paper I). By the time the first paper of the thesis was written, targeted therapy against TGFB1 was still not in clinical use. There are now a growing body of evidence supporting the use of TGFβ intervention in phase I-III clinical trials(97). There are many important mechanisms by which TGFβ can play a role TGFβ has been involved in

angiogenesis, cell proliferation, metastases dissemination, epithelial-to-mesenchymal transition, immune infiltration and drug resistance (98). Several recent high-impact papers also point to the immune evasion mechanisms as being very important for the role of targeting TGFβ in cancer (97, 99-101).

5.2.2 Discussion paper II

In the second paper of the thesis, we expand on the number of included renal cancer tissue samples. This aids the biomarker development to be able to confirm initial exploratory data in a second cohort of patients. There are several methods by which one can evaluate biomarkers in data sets with a low sample number. Internal cross-validation, iteratively leaving one sample out and predicting its endpoint based on the other samples is one example. It is also possible to split data sets in two and use one half as a training set and the other half for validation. The problem with this method is that it relies on samples being unbiased. In a clinical setting there are many sources for variation including differences in patient populations, recruitment differences, change in clinical practice as well as batch effects in the handling of tissue samples over the course of time. The way samples are included in this thesis (paper I and paper II) reflect the way samples are being handled in a real life clinical setting. We used the findings in paper I as training or exploratory data and the findings in paper II as a second cohort of patients for validation. Then, to reflect the Microarray Quality Control-II (MAQC-II) project analysis strategy we swapped the discovery and confirmation data set for a more complete evaluation of the diagnostic biomarker performance of some of our best classifiers (102). The classification accuracy of the discovery data set (paper I) is not as good as it could have been because of a tissue contamination of two samples. We had a normal sample classified as tumor and this specimen contained admixture of tumor tissue detected at a second evaluation of the light microscopy. We also had a tumor sample with some admixture of normal tissue. The classifiers tested revealed that these samples were grouped incorrectly and pointed to experimental inconsistencies. But at the same time this strengthens the belief in the diagnostic capacities of the biomarkers tested. In both data sets CA9 expression as well as a clinical microarray data set from another publication clearly

categorize ccRCC biopsies from normal renal biopsies (93). The second paper of this thesis also shows that a more novel potential biomarker for ccRCC, TNFAIP6, has very high diagnostic qualities. TNFAIP6 is a hyaluronan-binding secreted protein that drives epithelial–mesenchymal transition (EMT), which is an important factor in renal cancer pathophysiological mechanisms (103, 104). In paper I we also pointed to TGFB1 and DLL4 further supporting the EMT involvement in ccRCC. These findings were confirmed in the second paper. TNFAIP6, also very highly upregulated in both papers, is thought to be an important regulator of von Hippel-Lindau signaling (105). In addition to these molecules, most up-regulated genes were common in the validation study and the first discovery study. When comparing the results to a recent publication with a high number of subjects, our findings are well in line with the literature (48). Some genes, however are not detected in a microarray study like the one performed by Schrodter et al (48). This might be because the panel did not include an interesting gene like PNCK. We found this conspicuously interesting because PNCK was first linked to ccRCC carcinogenesis in 2010 and overexpression is linked to poor prognosis (106, 107). Also, a 2015 publication in breast cancer mentions PNCK as a novel calmodulin kinase, important for epidermal growth factor receptor stability and function, and as a marker for Trastuzumab resistance and a novel therapeutic target (108). Thus, PNCK could be a diagnostic, prognostic and predictive biomarker in ccRCC. The fact that this result was found in a series of RNA sequencing experiments from FFPE biopsies and not found in a large series of microarray patient series demonstrates the clinical biomarker development potential utilizing FFPE biopsies. Markers like CA9, TNFAIP6 or even PNCK could be used in larger series to evaluate the prognostic biomarker potential.

5.2.3 Discussion paper III

In the first and second paper in this thesis we used a well-known RNA extraction kit, namely the miRNeasy FFPE RNA extraction kit by Qiagen. In the third paper of this thesis we expanded on the technicalities of RNA extraction testing which kit gave the best yield and quality of RNA as well as developing a protocol for performing LCM and subsequent RNA extraction and RNAseq from these minutes amount of tissue. At

first we needed to establish which method was the most appropriate in quantifying low concentrations of RNA. In the Illumina TruSeq® RNA Access (now; RNA exome) Library Preparation Kit suggest the use of NanoDrop as a reference for RNA input requirements (78).

Spectrophotometry has been in use for decades to quantify the amount of nucleotides and proteins. NanoDrop® is one of the commercial spectrophotometric UV absorbance analysis platforms (109). UV spectrophotometry are not as specific in distinguishing RNA fragments from other nucleotides, phenols or proteins as compared to other methods (110, 111). Contaminants can generate false high results and altered 260/280 and 230/260 ratios. For general lab work spectrophotometric methods are reproducible and convenient, also no addition of reagents is needed. With the Qubit measurement system, a fluorescent dye selectively stains RNA and this is more accurate and specific than spectrophotometric methods (110, 111). In the third paper of this thesis, we got consistently higher RNA measurements in the NanoDrop compared to the Qubit method. It is imperative to have accurate information about RNA of low amounts, especially for our FFPE LCM tissues, also considering the low average 260/280 and 260/230 ratios of these samples. Repeated Qubit measurements should be performed when more precise numbers regarding the amount of RNA is needed. Nevertheless, the trends were similar in differentiating between the different kits in terms of RNA yield for both rat- and human tissue. The High Pure, miRNeasy, RNeasy and ExpressArt kit gave the highest amounts by both NanoDrop and Qubit measurements.

It is natural that the RNA yield from extracted LCM glomerular cross-sections is lower because of less tissue input quantity, but there is also evidence that supports the loss of RNA during Hematoxylin-staining (112). Shortened staining protocols as well as the addition of RNA-inhibitors into the staining solutions can decrease RNA degradation (112, 113). In the tissue sectioning protocol we developed for the preparation of slides for LCM we made sure all steps were carried out with precautions to prevent contaminations. All solutions used were RNAse free and we used RNAse free slides and cleaned all surfaces with RNAse-away. We also

shortened down all steps for the staining process to an absolute minimum. The highest risk for contamination took place when the tissues were harvested originally for the sole purpose of light microscopy. There can be cross-over from other samples and the paraffin itself is not “molecular grade/PCR clean”.

Together with the issue of getting the correct measurements in terms of RNA quantity. What is probably even more imperative is the RNA quality. We found the BioAnalyzer smear analyses and DV200 calculations very helpful for this task. There exist however different methods RNA quality assessments. The 28s/18s ribosomal ratio from an agarose gel electrophoresis can be used. The more up to date RNA Integrity number (RIN), calculated by the BioAnalyzer software, depicts the quality of extracted RNA by an algorithm, which takes the same 28s/18s ratio into consideration (114). The RIN number from 1 to 10 gives high quality information about fresh frozen samples and can differentiate what samples to use and not for a microarray experiment or a standard poly-A capture RNAseq experiment. When assessing FFPE tissues, all samples are usually between 1 (lowest) and 3. Since the FFPE samples are in the far lowest end of the scale it is appreciable that this is a poor predictor of cDNA yield from an RNA source (78). Regardless of low or even unmeasureable RIN values, RNA can be input for library preparation and RNA sequencing considering the requirements of the Illumina TruSeq® RNA Access Library Preparation Kit. According to the kit guidelines, the DV200 value of the extracted RNA determines the necessary quantity sufficient to obtain RNA suitable for RNA sequencing (78). One drawback for the DV200 evaluation is that one has to consider the size selection of RNA species for the RNA extraction kit applied. For instance the miRNeasy FFPE kit gave lower DV200 values than the other kits. This does not come as a surprise because it also extracts miRNA. The BioAnalyzer smears clearly shows an “early peak” representing the pool of miRNA. This gets calculated into the RNA fragments below 200 nucleotides. RNA extracted from LCM human glomerular cross-sections using the High Pure kit, ExpressArt and miRNeasy kit were successfully sequenced using the Illumina TruSeq® RNA Access Library

Preparation Kit. Despite lower DV200 values the miRNeasy kit gave acceptable number of sequencing reads. The High Pure kit had the most reads and we concluded this is therefore probably better than the other kits for mRNA seq from FFPE tissues.

6. Conclusions

We have demonstrated the technical feasibility of RNA sequencing from FFPE tissue per se and shown that the quality of the respective mRNA data derived from extracted RNA of whole FFPE tissue sections are sufficient and comparable to matched RNAlater® data. We have also provided evidence for the reproducibility of RNAseq data from two separate FFPE sample cohorts with a test and validation set for biomarker development. Lastly, we have also demonstrated the technical feasibility of RNA sequencing obtained from LCM renal tissues.

We used the proof of concept data to explore and to confirm published biological findings, and findings which may be worth following up in larger ccRCC patient cohorts, leading to possible novel therapeutic strategies, e.g. based on TGFB1-regulated genes, the NOTCH signaling cascade, and EMT.

Our studies open the door to transcriptome analyses of the archival, FFPE stored tissues from patients with ccRCC and supports CA9 as a potential marker for ccRCC. This work enables researchers to investigate archival tissue blocks from retrospectively defined clinical patient cohorts with long follow-up time and already available endpoint data, on a molecular level that previously has been considered very hard to achieve. Classifier models consisting of features such as gene expression data in combination with a decision algorithm are powerful tools to support diagnostic and prognostic evaluation of patient data. Further, RNA can be extracted from both a single human kidney biopsy section and from around 100 LCM glomerular cross-sections from human archival renal biopsies to successfully perform RNA sequencing using commercially available kits.

7. Future perspectives

We would like to use the novel methods established in this thesis on other projects targeting neoplastic and non-neoplasic kidney diseases. In the renal cancer field we would like to target our local archival tissue biobanks and address important clincal issues related to patient stratification based on correct estimation of prognosis where we have the necessary follow-up data already at hand. In this respect, an ongoing project is a cohort of low risk patients that still progress to metastatic disease. From 443 patients we have characterized 8 patients with initially low risk of disease progression that still progress to metastatic disease. Together with these eight primary tumors, we have 16 stable matched controls, and a total of 10 metastatic lesions. Full mRNAseq and MiR-seq will be performed from tissue biopsies with accompanying MiR-seq from serum samples at baseline. In another project in final revision at AJP-RP, we are completing a "cross-omics" approach fusing data from Mir-seq, mRNAseq and proteomics in ccRCC.

In general nephrology, we are already on our way with several ground-breaking projects including LCM from four different tissue compartments of three serial biopsies per patient in Fabry disease presented as a poster at the annual meeting of the American Society of Nephrology (ASN) 2018 (115).
In primary membranous nephropathy we have also performed LCM on the glomerular compartment to better characterize the gene expression pattern in PLA2R positive-, PLA2R negative- cases and controls presented as an oral abstract at the ASN 2018 (116). With state-of-the-art bioinformatics' we are displaying a transcriptomic landscape linking Drugbank®-listed compounds to relevant up- or down-regulated mRNA products/proteins within relevant clusters of gene sets (117).

8. References

1. von Ahlfen S, Missel A, Bendrat K, Schlumpberger M. Determinants of RNA quality from FFPE samples. PLoS One. 2007;2(12):e1261.
2. Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. Nucleic Acids Res. 1999;27(22):4436-43.
3. Goelz SE, Hamilton SR, Vogelstein B. Purification of DNA from formaldehyde fixed and paraffin embedded human tissue. Biochem Biophys Res Commun. 1985;130(1):118-26.
4. Shibata D, Martin WJ, Arnheim N. Analysis of DNA sequences in forty-year-old paraffin-embedded thin-tissue sections: a bridge between molecular biology and classical histology. Cancer Res. 1988;48(16):4564-6.
5. Iwamoto KS, Mizuno T, Ito T, Akiyama M, Takeichi N, Mabuchi K, et al. Feasibility of using decades-old archival tissues in molecular oncology/epidemiology. Am J Pathol. 1996;149(2):399-406.
6. Rupp GM, Locker J. Purification and analysis of RNA from paraffin-embedded tissues. Biotechniques. 1988;6(1):56-60.
7. Coombs NJ, Gough AC, Primrose JN. Optimisation of DNA and RNA extraction from archival formalin-fixed tissue. Nucleic Acids Res. 1999;27(16):e12.
8. Mies C. Molecular biological analysis of paraffin-embedded tissues. Hum Pathol. 1994;25(6):555-60.
9. Lewis F, Maughan NJ, Smith V, Hillan K, Quirke P. Unlocking the archive--gene expression in paraffin-embedded tissue. J Pathol. 2001;195(1):66-71.
10. Dahm R. Friedrich Miescher and the discovery of DNA. Dev Biol. 2005;278(2):274-88.
11. Watson JD, Crick FH. The structure of DNA. Cold Spring Harb Symp Quant Biol. 1953;18:123-31.
12. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-7.
13. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.
14. Buermans HP, den Dunnen JT. Next generation sequencing technology: Advances and applications. Biochim Biophys Acta. 2014;1842(10):1932-41.
15. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell. 2013;155(1):27-38.
16. Schloss JA. How to get genomes at one ten-thousandth the cost. Nat Biotechnol. 2008;26(10):1113-5.
17. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418-26.
18. Kleppe K, Ohtsuka E, Kleppe R, Molineux I, Khorana HG. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. J Mol Biol. 1971;56(2):341-61.

19. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 1985;230(4732):1350-4.
20. Renkema KY, Stokman MF, Giles RH, Knoers NV. Next-generation sequencing for research and diagnostics in kidney disease. Nat Rev Nephrol. 2014;10(8):433-44.
21. Reisaeter AV. Annual Report of the Norwegian Renal Registry. Oslo, Norway: Renal Unit, OUS Rikshospitalet, Oslo, Norway. 2017:p.20.
22. Vikse BE, Irgens LM, Leivestad T, Skjaerven R, Iversen BM. Preeclampsia and the risk of end-stage renal disease. N Engl J Med. 2008;359(8):800-9.
23. Skrunes R, Svarstad E, Reisaeter AV, Vikse BE. Familial clustering of ESRD in the Norwegian population. Clin J Am Soc Nephrol. 2014;9(10):1692-700.
24. Chesnaye NC, Schaefer F, Bonthuis M, Holman R, Baiko S, Baskin E, et al. Mortality risk disparities in children receiving chronic renal replacement therapy for the treatment of end-stage renal disease across Europe: an ESPN-ERA/EDTA registry analysis. Lancet. 2017;389(10084):2128-37.
25. Eide IA, Halden TAS, Hartmann A, Dahle DO, Asberg A, Jenssen T. Associations Between Posttransplantation Diabetes Mellitus and Renal Graft Survival. Transplantation. 2017;101(6):1282-9.
26. Vethe H, Finne K, Skogstrand T, Vaudel M, Vikse BE, Hultstrom M, et al. Distinct protein signature of hypertension-induced damage in the renal proteome of the two-kidney, one-clip rat model. J Hypertens. 2015;33(1):126-35.
27. Paunas TIF, Finne K, Leh S, Marti HP, Mollnes TE, Berven F, et al. Glomerular abundance of complement proteins characterized by proteomic analysis of laser-captured microdissected glomeruli associates with progressive disease in IgA nephropathy. Clinical proteomics. 2017;14:30.
28. Ovrehus MA, Zurbig P, Vikse BE, Hallan SI. Urinary proteomics in chronic kidney disease: diagnosis and risk of progression beyond albuminuria. Clinical proteomics. 2015;12(1):21.
29. Finne K, Vethe H, Skogstrand T, Leh S, Dahl TD, Tenstad O, et al. Proteomic analysis of formalin-fixed paraffin-embedded glomeruli suggests depletion of glomerular filtration barrier proteins in two-kidney, one-clip hypertensive rats. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2014;29(12):2217-27.
30. Finne K, Marti HP, Leh S, Skogstrand T, Vethe H, Tenstad O, et al. Proteomic Analysis of Minimally Damaged Renal Tubular Tissue from Two-Kidney-One-Clip Hypertensive Rats Demonstrates Extensive Changes Compared to Tissue from Controls. Nephron. 2016;132(1):70-80.
31. Koivuviita N, Tertti R, Heiro M, Metsarinne K. A case report: a patient with IgA nephropathy and coeliac disease. Complete clinical remission following gluten-free diet. NDT Plus. 2009;2(2):161-3.
32. Skrunes R, Leh S. Annual Report of the Norwegian Renal Registry. Section for Kidney Biopsy. Medical Department, Haukeland University Hospital, Bergen, Norway. 2016.
33. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333-51.

34. Kircher M, Kelso J. High-throughput DNA sequencing--concepts and limitations. Bioessays. 2010;32(6):524-36.
35. Wetterstrand KA. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). National Human Genome Research Institute. 2016 [updated 06.07.2016]. Available from: https://www.genome.gov/sequencingcosts/.
36. Eisengart LJ, MacVicar GR, Yang XJ. Predictors of response to targeted therapy in renal cell carcinoma. Arch Pathol Lab Med. 2012;136(5):490-5.
37. Ljungberg B, Campbell SC, Choi HY, Jacqmin D, Lee JE, Weikert S, et al. The epidemiology of renal cell carcinoma. Eur Urol. 2011;60(4):615-21.
38. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7-30.
39. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin. 2017;67(1):7-30.
40. Larsen IK, Møller, B., Johannesen, T.B., Robsahm, T.E., Grimsrud, T.K., Larønningen, S., Jakobsen, E., Ursin, G. Cancer Registry of Norway. Cancer in Norway 2017 - Cancer incidence, mortality, survival and prevalence in Norway. 2018.
41. Dabestani S, Beisland C, Stewart GD, Bensalah K, Gudmundsson E, Lam TB, et al. Long-term Outcomes of Follow-up for Initially Localised Clear Cell Renal Cell Carcinoma: RECUR Database Analysis. Eur Urol Focus. 2018.
42. Leibovich BC, Blute ML, Cheville JC, Lohse CM, Frank I, Kwon ED, et al. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. Cancer. 2003;97(7):1663-71.
43. Escudier B, Porta C, Schmidinger M, Rioux-Leclercq N, Bex A, Khoo V, et al. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2016;27(suppl 5):v58-v68.
44. Choueiri TK, Motzer RJ. Systemic Therapy for Metastatic Renal-Cell Carcinoma. N Engl J Med. 2017;376(4):354-66.
45. Guethmundsson E, Hellborg H, Lundstam S, Erikson S, Ljungberg B, Swedish Kidney Cancer Quality Register G. Metastatic potential in renal cell carcinomas </=7 cm: Swedish Kidney Cancer Quality Register data. Eur Urol. 2011;60(5):975-82.
46. Thorstenson A, Harmenberg U, Lindblad P, Holmstrom B, Lundstam S, Ljungberg B. Cancer Characteristics and Current Treatments of Patients with Renal Cell Carcinoma in Sweden. Biomed Res Int. 2015;2015:456040.
47. Escudier B, Porta C, Schmidinger M, Algaba F, Patard JJ, Khoo V, et al. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2014;25 Suppl 3:iii49-56.
48. Schrodter S, Braun M, Syring I, Klumper N, Deng M, Schmidt D, et al. Identification of the dopamine transporter SLC6A3 as a biomarker for patients with renal cell carcinoma. Mol Cancer. 2016;15:10.
49. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89-95.

50. Duffy MJ, Sturgeon CM, Soletormos G, Barak V, Molina R, Hayes DF, et al. Validation of new cancer biomarkers: a position statement from the European group on tumor markers. Clin Chem. 2015;61(6):809-20.
51. Mordente A, Meucci E, Martorana GE, Silvestrini A. Cancer Biomarkers Discovery and Validation: State of the Art, Problems and Future Perspectives. Adv Exp Med Biol. 2015;867:9-26.
52. Parkinson DR, McCormack RT, Keating SM, Gutman SI, Hamilton SR, Mansfield EA, et al. Evidence of clinical utility: an unmet need in molecular diagnostics for patients with cancer. Clin Cancer Res. 2014;20(6):1428-44.
53. Jardillier R, Chatelain F, Guyon L. Bioinformatics Methods to Select Prognostic Biomarker Genes from Large Scale Datasets: A Review. Biotechnol J. 2018;13(12):e1800103.
54. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med. 2001;344(11):783-92.
55. Oldenhuis CN, Oosting SF, Gietema JA, de Vries EG. Prognostic versus predictive value of biomarkers in oncology. Eur J Cancer. 2008;44(7):946-53.
56. Huang J, Hu W, Sood AK. Prognostic biomarkers in ovarian cancer. Cancer Biomark. 2010;8(4-5):231-51.
57. Nalejska E, Maczynska E, Lewandowska MA. Prognostic and predictive biomarkers: tools in personalized oncology. Mol Diagn Ther. 2014;18(3):273-84.
58. Landolt L, Strauss P, Marti HP, Eikrem O. Next Generation Sequencing: A Tool for This Generation of Nephrologists. EMJ 2016;1[2]:50-57.
59. Fuhrman SA, Lasky LC, Limas C. Prognostic significance of morphologic parameters in renal cell carcinoma. Am J Surg Pathol. 1982;6(7):655-63.
60. Greene FL, Sobin LH. A worldwide approach to the TNM staging system: collaborative efforts of the AJCC and UICC. J Surg Oncol. 2009;99(5):269-72.
61. Ljungberg B, Cowan NC, Hanbury DC, Hora M, Kuczyk MA, Merseburger AS, et al. EAU guidelines on renal cell carcinoma: the 2010 update. Eur Urol. 2010;58(3):398-406.
62. Delahunt B, Cheville JC, Martignoni G, Humphrey PA, Magi-Galluzzi C, McKenney J, et al. The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters. Am J Surg Pathol. 2013;37(10):1490-504.
63. Dagher J, Delahunt B, Rioux-Leclercq N, Egevad L, Srigley JR, Coughlin G, et al. Clear cell renal cell carcinoma: validation of World Health Organization/International Society of Urological Pathology grading. Histopathology. 2017;71(6):918-25.
64. Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Hora M, et al. EAU guidelines on renal cell carcinoma: 2014 update. Eur Urol. 2015;67(5):913-24.
65. Espina V, Wulfkuhle JD, Calvert VS, VanMeter A, Zhou W, Coukos G, et al. Laser-capture microdissection. Nat Protoc. 2006;1(2):586-603.
66. Shibata D, Hawes D, Li ZH, Hernandez AM, Spruck CH, Nichols PW. Specific genetic analysis of microscopic tissue after selective ultraviolet radiation fractionation and the polymerase chain reaction. Am J Pathol. 1992;141(3):539-43.

67. Skogstrand T, Leh S, Paliege A, Reed RK, Vikse BE, Bachmann S, et al. Arterial damage precedes the development of interstitial damage in the nonclipped kidney of two-kidney, one-clip hypertensive rats. J Hypertens. 2013;31(1):152-9.
68. Medeiros F, Rigl CT, Anderson GG, Becker SH, Halling KC. Tissue handling for genome-wide expression analysis: a review of the issues, evidence, and opportunities. Arch Pathol Lab Med. 2007;131(12):1805-16.
69. Weber DG, Casjens S, Rozynek P, Lehnert M, Zilch-Schoneweis S, Bryk O, et al. Assessment of mRNA and microRNA Stabilization in Peripheral Human Blood for Multicenter Studies and Biobanks. Biomark Insights. 2010;5:95-102.
70. Li P, Conley A, Zhang H, Kim HL. Whole-Transcriptome profiling of formalin-fixed, paraffin-embedded renal cell carcinoma by RNA-seq. BMC Genomics. 2014;15:1087.
71. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics. 2014;15:419.
72. Webster AF, Zumbo P, Fostel J, Gandara J, Hester SD, Recio L, et al. Mining the Archives: A Cross-Platform Analysis of Gene Expression Profiles in Archival Formalin-Fixed Paraffin-Embedded Tissues. Toxicol Sci. 2015;148(2):460-72.
73. Hedegaard J, Thorsen K, Lund MK, Hein AM, Hamilton-Dutoit SJ, Vang S, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. PLoS One. 2014;9(5):e98187.
74. Ribeiro-Silva A, Zhang H, Jeffrey SS. RNA extraction from ten year old formalin-fixed paraffin-embedded breast cancer samples: a comparison of column purification and magnetic bead-based technologies. BMC Mol Biol. 2007;8:118.
75. Cieslik M, Chugh R, Wu YM, Wu M, Brennan C, Lonigro R, et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. Genome Res. 2015;25(9):1372-81.
76. Halvardson J, Zaghlool A, Feuk L. Exome RNA sequencing reveals rare and novel alternative transcripts. Nucleic Acids Res. 2013;41(1):e6.
77. Roberts L, Bowers J, Sensinger K, Lisowski A, Getts R, Anderson MG. Identification of methods for use of formalin-fixed, paraffin-embedded tissue samples in RNA expression profiling. Genomics. 2009;94(5):341-8.
78. Evaluating RNA Quality from FFPE Samples [updated Pub. No. 470-2014-001 Current as of 26 October 2016. Available from: https://emea.illumina.com/content/dam/illumina-marketing/documents/products/technotes/evaluating-rna-quality-from-ffpe-samples-technical-note-470-2014-001.pdf.
79. Huang W, Goldfischer M, Babayeva S, Mao Y, Volyanskyy K, Dimitrova N, et al. Identification of a novel PARP14-TFE3 gene fusion from 10-year-old FFPE tissue by RNA-seq. Genes Chromosomes Cancer. 2015;54(8):500-5.
80. Walther C, Hofvander J, Nilsson J, Magnusson L, Domanski HA, Gisselsson D, et al. Gene fusion detection in formalin-fixed paraffin-embedded benign fibrous histiocytomas using fluorescence in situ hybridization and RNA sequencing. Lab Invest. 2015;95(9):1071-6.

81. Puls F, Hofvander J, Magnusson L, Nilsson J, Haywood E, Sumathi VP, et al. FN1-EGF gene fusions are recurrent in calcifying aponeurotic fibroma. J Pathol. 2016;238(4):502-7.
82. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883-92.
83. Tostain J, Li G, Gentil-Perret A, Gigante M. Carbonic anhydrase 9 in clear cell renal cell carcinoma: a marker for diagnosis, prognosis and treatment. Eur J Cancer. 2010;46(18):3141-8.
84. Li G, Feng G, Zhao A, Peoc'h M, Cottier M, Mottet N. CA9 as a biomarker in preoperative biopsy of small solid renal masses for diagnosis of clear cell renal cell carcinoma. Biomarkers. 2017;22(2):123-6.
85. Gimenez-Bachs JM, Salinas-Sanchez AS, Serrano-Oviedo L, Nam-Cha SH, Rubio-Del Campo A, Sanchez-Prieto R. Carbonic anhydrase IX as a specific biomarker for clear cell renal cell carcinoma: comparative study of Western blot and immunohistochemistry and implications for diagnosis. Scand J Urol Nephrol. 2012;46(5):358-64.
86. Oosterwijk E. Carbonic anhydrase expression in kidney and renal cancer: implications for diagnosis and treatment. Subcell Biochem. 2014;75:181-98.
87. Stewart GD, O'Mahony FC, Laird A, Rashid S, Martin SA, Eory L, et al. Carbonic anhydrase 9 expression increases with vascular endothelial growth factor-targeted therapy and is predictive of outcome in metastatic clear cell renal cancer. Eur Urol. 2014;66(5):956-63.
88. Stillebroer AB, Mulders PF, Boerman OC, Oyen WJ, Oosterwijk E. Carbonic anhydrase IX in renal cell carcinoma: implications for prognosis, diagnosis, and therapy. Eur Urol. 2010;58(1):75-83.
89. Xu J, Pham CG, Albanese SK, Dong Y, Oyama T, Lee CH, et al. Mechanistically distinct cancer-associated mTOR activation clusters predict sensitivity to rapamycin. J Clin Invest. 2016;126(9):3526-40.
90. Voss MH, Chen D, Marker M, Hakimi AA, Lee CH, Hsieh JJ, et al. Circulating biomarkers and outcome from a randomised phase II trial of sunitinib vs everolimus for patients with metastatic renal cell carcinoma. Br J Cancer. 2016;114(6):642-9.
91. Voss MH, Hakimi AA, Pham CG, Brannon AR, Chen YB, Cunha LF, et al. Tumor genetic analyses of patients with metastatic renal cell carcinoma and extended benefit from mTOR inhibitor therapy. Clin Cancer Res. 2014;20(7):1955-64.
92. Funakoshi T, Lee CH, Hsieh JJ. A systematic review of predictive and prognostic biomarkers for VEGF-targeted therapy in renal cell carcinoma. Cancer Treat Rev. 2014;40(4):533-47.
93. von Roemeling CA, Radisky DC, Marlow LA, Cooper SJ, Grebe SK, Anastasiadis PZ, et al. Neuronal pentraxin 2 supports clear cell renal cell carcinoma by activating the AMPA-selective glutamate receptor-4. Cancer Res. 2014;74(17):4796-810.
94. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One. 2014;9(1):e78644.

95. Iacovelli R, Sternberg CN, Porta C, Verzoni E, de Braud F, Escudier B, et al. Inhibition of the VEGF/VEGFR pathway improves survival in advanced kidney cancer: a systematic review and meta-analysis. Curr Drug Targets. 2015;16(2):164-70.
96. Lai Y, Zhao Z, Zeng T, Liang X, Chen D, Duan X, et al. Crosstalk between VEGFR and other receptor tyrosine kinases for TKI therapy of metastatic renal cell carcinoma. Cancer Cell Int. 2018;18:31.
97. de Gramont A, Faivre S, Raymond E. Novel TGF-beta inhibitors ready for prime time in onco-immunology. Oncoimmunology. 2017;6(1):e1257453.
98. Neuzillet C, Tijeras-Raballand A, Cohen R, Cros J, Faivre S, Raymond E, et al. Targeting the TGFbeta pathway for cancer therapy. Pharmacol Ther. 2015;147:22-31.
99. Chakravarthy A, Khan L, Bensler NP, Bose P, De Carvalho DD. TGF-beta-associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. Nat Commun. 2018;9(1):4692.
100. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018;554(7693):544-8.
101. Tauriello DVF, Palomo-Ponce S, Stork D, Berenguer-Llergo A, Badia-Ramentol J, Iglesias M, et al. TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. Nature. 2018;554(7693):538-43.
102. Shi L, Campbell G, Jones WD, Campagne F, Wen Z, Walker SJ, et al. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. Nat Biotechnol. 2010;28(8):827-38.
103. Bommaya G, Meran S, Krupa A, Phillips AO, Steadman R. Tumour necrosis factor-stimulated gene (TSG)-6 controls epithelial-mesenchymal transition of proximal tubular epithelial cells. Int J Biochem Cell Biol. 2011;43(12):1739-46.
104. Mikami S, Mizuno R, Kosaka T, Saya H, Oya M, Okada Y. Expression of TNF-alpha and CD44 is implicated in poor prognosis, cancer cell invasion, metastasis and resistance to the sunitinib treatment in clear cell renal cell carcinomas. Int J Cancer. 2015;136(7):1504-14.
105. Chi JT, Wang Z, Nuyten DS, Rodriguez EH, Schaner ME, Salim A, et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. PLoS Med. 2006;3(3):e47.
106. Zhou L, Chen J, Li Z, Li X, Hu X, Huang Y, et al. Integrated profiling of microRNAs and mRNAs: microRNAs located on Xq27.3 associate with clear cell renal cell carcinoma. PLoS One. 2010;5(12):e15224.
107. Wu S, Lv Z, Wang Y, Sun L, Jiang Z, Xu C, et al. Increased expression of pregnancy up-regulated non-ubiquitous calmodulin kinase is associated with poor prognosis in clear cell renal cell carcinoma. PLoS One. 2013;8(4):e59936.
108. Deb TB, Zuo AH, Barndt RJ, Sengupta S, Jankovic R, Johnson MD. Pnck overexpression in HER-2 gene-amplified breast cancer causes Trastuzumab resistance through a paradoxical PTEN-mediated process. Breast Cancer Res Treat. 2015;150(2):347-61.

109. NanoDrop, 260/280 and 260/230 ratios: NanoDropVR ND-1000 and ND-8000 8-Sample Spectrophotometers. Techn Supp Bull T009 Nanodrop Technol2007 [
110. O'Neill M. J Phys: Conf Ser 307 012047. 2011.
111. Thermo Fisher Scientific Accurate and sensitive quantitation of nucleic acids, even at low concentrations. Comparison of the qubit quantitation platform with spectrophotometry. Bio Probes 2010;62:20–2.2010 31.01.2019. Available from: https://www.thermofisher.com/no/en/home/references/newsletters-and-journals/bioprobes-journal-of-cell-biology-applications/bioprobes-issues-2010/bioprobes-62/comparison-of-qubit-quantitation-platform-with-spectrophotometry.html.
112. Wang H, Owens JD, Shih JH, Li MC, Bonner RF, Mushinski JF. Histological staining methods preparatory to laser capture microdissection significantly affect the integrity of the cellular RNA. BMC Genomics. 2006;7:97.
113. Yee JY, Limenta LM, Rogers K, Rogers SM, Tay VS, Lee EJ. Ensuring good quality RNA for quantitative real-time PCR isolated from renal proximal tubular cells using laser capture microdissection. BMC Res Notes. 2014;7:62.
114. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006;7:3.
115. Eikrem O, Strauss P, Sekulic M, Tøndel C, Flatberg A, Skrunes R, et al. Fabry nephropathy: Transcriptome sequencing of microdissected renal compartments from archival kidney biopsies at baseline, and after 5 & 10 years of enzyme replacement therapy. ASN-Kidney Week; 2018-10-25 - 2018-10-28.
116. Eikrem O, Koch E, Knoop T, Vikse B, Schordan E, Debiec H, et al. Primary Membranous Nephropathy: Glomerular RNA Sequencing from Archival Kidney Biopsies. ASN-Kidney Week 2018-10-25 - 2018-10-28.
117. Delaleu N, Nguyen CQ, Tekle KM, Jonsson R, Peck AB. Transcriptional landscapes of emerging autoimmunity: transient aberrations in the targeted tissue's extracellular milieu precede immune responses in Sjogren's syndrome. Arthritis Res Ther. 2013;15(5):R174.

9. Appendix

9.1 Tissue sectioning protocol for LCM slides

	All steps are carried out with precautions to prevent contamination with keratin. Gloves and a cap are used during handling the blocks and the glass slides. It is important that the time the sections are in room temperature is kept to a minimum. Careful handling of the slides in the deparaffinization and staining procedure to avoid that the sections might fall off the membrane.	
Preparation	Nuclease-free PEN (polyethylene naphthalate) slides are used. Slides are irradiated by UV light for 30 min. to make the membrane more hydrophilic and so achieve a better adherence. This is done in 5327A (lab. for molecular pathology) under the fume hood.	
Sectioning	The microtome should be cleaned before start as well as the tools for handling the sections with RNAse away solution by simply wiping with a paper towel soaked in RNAse away. Afterwards wash with sterile gauze soaked in autoclaved distilled water.	
	Use a new cutting knife and move the knife over for every sample to avoid sample cross-over.	
	Wipe the water bath container once with a paper-towel with RNAse away solution. Rinse once with autoclaved distilled water. Fill the water bath with autoclaved distilled water. If needed Øystein can supply the autoclaved distilled water.	
	Use only autoclaved distilled water throughout. Should any of the reagents used need dilution in water, autoclaved distilled water should also be used then (exception PAS and Sirius red).	
	Discard the first section	
	Section thickness: 10 µm for PEN slides, otherwise **2**-3 µm.	
	After expanding in the water bath sections are mounted on PEN membrane slides	
	Make sections for the PEN slides and parallel sections on normal slides for microscopy as shown below in the table with sectioning numbers	
Drying	Dry for 20 min at 60 ºC in a drying oven, take out and look carefully at all slides. Pinch out any big water bubbles with a sterile gauze from the **outside** of the membrane. Do not swipe across the membrane and risk damaging the membrane or slide. Dry for another 10 minutes in the 60 ºC drying oven. Total 60 ºC drying time 30 mins. Let air dry for some minutes.	
Deparaffinization	Freshly prepared molecular grade Xylene and ethanol solutions and autoclaved containers.	
	Xylene	2 x 2 minutes each
	Ethanol absolute	1 minute
	Ethanol 96%	1 minute
	Ethanol 70%	1 minute
	Autoclaved distilled water	1 minute
Staining	Under the fume hood. Clean jars (autoclaved) and clean solutions. Filter the staining solutions.	
HE	for microdissection (PEN-slides)	
	1 min Shandon hematoxylin solution (The “strong one”=powder solution - mixed with autoclaved water)	
	5 min total rinsing in autoclaved distilled water. 2,5 minutes in one container, and 2,5 minutes in a new container. Use caution not to loosen the tissue from the slides.	
	30 sec Eosin	
	Increasing ethanol series	
	Let air-dry for approximately 15-20 mins before storage in freezer.	
Storage	Immediately for -20 ºC storage.	

For PAS and Sirius red stained sections, water does not need to be destilled and autoclaved	
PAS	for intermittent control sections for accurate diagnostics and tissue selection (PAS staining is performed together with routine nephrology samples)
	10 min perjodic acid 1 %
	10 min rinsing in running tap water (lukewarm)
	10 – 15 min Schiff's reagent
	10 min rinsing in running tap water (lukewarm)
	5 min hematoxylin
	Rinsing in running tap water
	Increasing ethanol series
	Xylene
	Coverslip
	Scan at 40x and notify Sabine, glass slides in renal biopsy archive
Sirius red	procedure including a nuclear stain has to be developed, for example:
	8 min Weigert's hematoxylin
	10 min rinsing in running tap water
	60 min Sirius red
	Increasing ethanol series
	Xylene
	Coverslip
	Scan

PLOS | ONE

RESEARCH ARTICLE

Transcriptome Sequencing (RNAseq) Enables Utilization of Formalin-Fixed, Paraffin-Embedded Biopsies with Clear Cell Renal Cell Carcinoma for Exploration of Disease Biology and Biomarker Development

Oystein Eikrem[1], Christian Beisland[2], Karin Hjelle[2], Arnar Flatberg[3], Andreas Scherer[4], Lea Landolt[1], Trude Skogstrand[1], Sabine Leh[1,5], Vidar Beisvag[3], Hans-Peter Marti[1,6]*

1 Department of Clinical Medicine, Nephrology, University of Bergen, Bergen, Norway, **2** Department of Clinical Medicine, Urology, University of Bergen, Bergen, Norway, **3** Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway, **4** Spheromics, Kontiolahti, Finland, **5** Department of Pathology, Haukeland University Hospital, Bergen, Norway, **6** Department of Medicine, Nephrology, Haukeland University Hospital, Bergen, Norway

* hans-peter.marti@uib.no

OPEN ACCESS

Citation: Eikrem O, Beisland C, Hjelle K, Flatberg A, Scherer A, Landolt L, et al. (2016) Transcriptome Sequencing (RNAseq) Enables Utilization of Formalin-Fixed, Paraffin-Embedded Biopsies with Clear Cell Renal Cell Carcinoma for Exploration of Disease Biology and Biomarker Development. PLoS ONE 11(2): e0149743. doi:10.1371/journal.pone.0149743

Editor: Christos Chatziantoniou, Institut National de la Santé et de la Recherche Médicale, FRANCE

Received: September 25, 2015

Accepted: February 4, 2016

Published: February 22, 2016

Data Availability Statement: Data underlying the study are available in the repository Gene Expression Omnibus, (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76207) under accession number GSE76207.

Funding: The funder, namely the University of Bergen, provided support in the form of the salary for author Oystein Eikrem (O.E.) as PhD student scholarship but it did not have any additional role in the study design, data collection and analysis,

Abstract

Formalin-fixed, paraffin-embedded (FFPE) tissues are an underused resource for molecular analyses. This proof of concept study aimed to compare RNAseq results from FFPE biopsies with the corresponding RNAlater® (Qiagen, Germany) stored samples from clear cell renal cell carcinoma (ccRCC) patients to investigate feasibility of RNAseq in archival tissue. From each of 16 patients undergoing partial or full nephrectomy, four core biopsies, such as two specimens with ccRCC and two specimens of adjacent normal tissue, were obtained with a 16g needle. One normal and one ccRCC tissue specimen per patient was stored either in FFPE or RNAlater®. RNA sequencing libraries were generated applying the new Illumina TruSeq® Access library preparation protocol. Comparative analysis was done using voom/Limma R-package. The analysis of the FFPE and RNAlater® datasets yielded similar numbers of detected genes, differentially expressed transcripts and affected pathways. The FFPE and RNAlater datasets shared 80% (n = 1106) differentially expressed genes. The average expression and the log2 fold changes of these transcripts correlated with $R^2 = 0.97$, and $R^2 = 0.96$, respectively. Among transcripts with the highest fold changes in both datasets were carbonic anhydrase 9 (*CA9)*, neuronal pentraxin-2 (*NPTX2)* and uromodulin (*UMOD)* that were confirmed by immunohistochemistry. IPA revealed the presence of gene signatures of cancer and nephrotoxicity, renal damage and immune response. To simulate the feasibility of clinical biomarker studies with FFPE samples, a classifier model was developed for the FFPE dataset: expression data for *CA9* alone had an accuracy, specificity and sensitivity of 94%, respectively, and achieved similar performance in the RNAlater dataset. Transforming growth factor-ß1 (*TGFB1)*-regulated genes, epithelial to mesenchymal transition (EMT) and *NOTCH* signaling cascade may support novel

decision to publish, or preparation of the manuscript. The library preparation, sequencing and partly the bioinformatics analysis were provided by the Genomics Core Facility (GCF), which is funded by the Faculty of Medicine at NTNU and Central Norway Regional Health Authority. The specific roles of this author is articulated in the 'author contributions' section." Note, O.E. was an essential part in performing experiments and manuscript writing. Andreas Scherer (A. S.) is the sole owner/employee of Spheromics (http://spheromics.com/) and provided help in RNA sequencing data analysis and preparation of respective manuscript parts. Illumina Inc. was not involved in the present study at all; accordingly, no funding was obtained. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The affiliation with Spheromics does not alter the authors' adherence to PLOS ONE policies on sharing data and materials. Illumina Inc. was not involved in the study.

therapeutic strategies. In conclusion, in this proof of concept study, RNAseq data obtained from FFPE kidney biopsies are comparable to data obtained from fresh stored material, thereby expanding the utility of archival tissue specimens.

Introduction

Clear cell renal cell carcinoma (ccRCC) makes up the majority of primary renal neoplasms with increasing incidence and considerable morbidity and mortality. Metastasis reflects a major cause of patient death [1, 2]. Renal cell cancer ranks among the ten most frequent cancers in women and men accounting for up to 2–3% of all adult cancers or malignancies [2–6].

The ccRCC is only curable by early surgical tumor removal. Thus, efforts to unravel molecular mechanisms of this disease for the search of prognostic markers and novel drug targets are important, e.g. by applying gene expression detection technologies to develop molecular signatures of disease progression.

In this study, we applied RNA sequencing (RNAseq), a method for measuring mRNA abundance based on next generation sequencing (NGS) technology. NGS can identify transcripts even at a low expression level and provides an increased dynamic range for gene expression measurements compared to microarrays [7, 8].

Current technologies for whole genome gene expression analyses are largely dependent on "high quality" RNA with low level of degradation. We wanted to test whether lower quality, partially degraded RNA obtained from archival formalin-fixed and paraffin-embedded (FFPE) renal tissues could serve as appropriate source of information.

The quality of RNA extracted from FFPE samples can vary widely among different specimens, or within different samples from the same specimen. RNA undergoes substantial chemical modification during formalin fixation, nucleic acids are cross-linked to proteins and RNA transcripts are degraded to smaller fragments [9]. Differences in formalin fixation methods and age of archival tissue samples add further variation to RNA quality. The Illumina TruSeq RNA Access Kit® holds promise to overcome these challenges for RNA sequencing applications by isolating mRNA through a sequence-specific capture protocol resulting in reduced ribosomal RNA and enriched exonic RNA sequences. The TruSeq RNA Access library preparation kit was designed to ensure high quality RNA sequencing data from degraded FFPE samples and to allow comparison across samples that vary in quality.

Transcriptome sequencing of RNA from concurrently harvested FFPE and fresh stored kidney biopsies with subsequent analysis of transcripts and pathways underlying ccRCC in our patient group served as indication of the comparability of the two sources of RNA. The comparison to published data helped to estimate the biological and clinical plausibility of our results.

Results

Study design

This study includes 16 adult patients from Haukeland University Hospital with ccRCC undergoing partial (n = 10) or full (n = 6) nephrectomy between November 2013 and August 2014 (Table 1). Each patient donated four core biopsies, including two with ccRCC and two from adjacent non-affected tissue ("normal"). One pair of ccRCC and normal tissue per patient was then stored in FFPE, the other pair in RNAlater®. This paired design allows comparison of mRNA abundance level differences between ccRCC and normal in FFPE and in RNAlater®, and to evaluate the impact of storage condition on expression profiles using RNAseq.

Table 1. Characteristic patient features at the time of surgery. eGFR was calculated with the MDRD formula. The staging was performed based on the EAU Guidelines on renal cell carcinoma: 2014 update [43].

Patient number	Age, yr	Gender	BMI	Nephrectomy type	eGFR (ml/min/1.73m^2)	TNM-stage	Size (mm)	Fuhrmann grade	Stage
9	70	Male	24	Partial	>60	pT1AcN0cM0	18	2	I
10	69	Male	34	Partial	>60	pT3AcN0cM0	15	2	III
11	37	Male	27	Partial	>60	pT1AcN0cM0	19	2	I
13	63	Male	24	Full	40	pT3AcN0cM0	69	4	III
15	68	Male	28	Partial	>60	pT1AcN0cM0	21	2	I
16	53	Male	33	Full	56	pT3bN0M1	100	2	IV
18	78	Male	27	Full	47	T3AcN0cM0	60	2	III
19	71	Female	22	Full	>60	pT2aN0cM0	90	1	II
21	53	Female	25	Full	55	pT1BcN0cM0	65	2	I
22	49	Male	25	Partial	>60	pT1BcN0cM0	50	2	I
24	69	Male	27	Partial	>60	pT1AcN0cM0	25	2	I
27	46	Male	31	Full	>60	pT2BcN0cM0	117	3	II
29	54	Female	29	Partial	>60	pT1AcN0cM0	15	2	I
31	67	Male	25	Partial	>60	pT1AcN0cM0	18	1	I
32	36	Male	23	Partial	>60	pT1AcN0cM0	18	3	I
33	48	Male	28	Partial	>60	pT1AcN0cM0	38	1	I

doi:10.1371/journal.pone.0149743.t001

Quality of Extracted RNA

To assess the quality of the 64 samples of extracted RNA we determined the Agilent RNA integrity number (RIN). Currently, the RIN is the most commonly used measure to determine RNA quality for gene expression analysis [10]. However, RIN values from FFPE samples are not a sensitive measure of RNA quality nor are they a reliable predictor of successful library preparation. Accordingly, previous investigators have used mean RNA fragment size as a determinant of RNA quality for the RNA sequencing library preparation (Illumina TruSeq RNA Access Kit®) when working with RNA obtained from FFPE tissues [11–13].

We have therefore also used the DV_{200} metric, the percentage of RNA fragments >200 nucleotides to evaluate the RNA quality according to the recommendation of the manufacturer and as described [11–13]. Using DV_{200} to accurately assess FFPE RNA quality, and by adjusting RNA input amounts, high-quality libraries can be prepared from poor-quality FFPE samples. In this respect, a sufficient DV_{200} value of as low as 30% was reported [13].

The mean Agilent RNA integrity number (RIN) and mean DV_{200} values (95% CI) were 5.7 (5.10–6.30) and 61% (58–64) for RNAlater® samples and 2.53 (2.33–2.73) and 75% (72–79) for FFPE samples, respectively.

Gene Expression (mRNA Abundance)

The number of detected genes, which passed an expression filter of more than 15 cpm in at least 8 samples per dataset, for FFPE was n = 9164 and for RNAlater® n = 9205. Notably, about 94% of the genes in each dataset (n = 8893) were common to both FFPE and RNAlater® datasets; correlation of the logarithmic fold change was $R^2 = 0.93$, and correlation of the average expression $R^2 = 0.97$, as shown in S1 Fig.

To find sources of similarity in the dataset consisting of all 64 samples and the expression values of expression-filtered 8893 genes, we applied multidimensional scaling (MDS). Samples segregate into two large groups along the leading log-fold change in the dimension 1 of the MDS plot. The leading log-fold change is the average (root-mean-square) of the largest

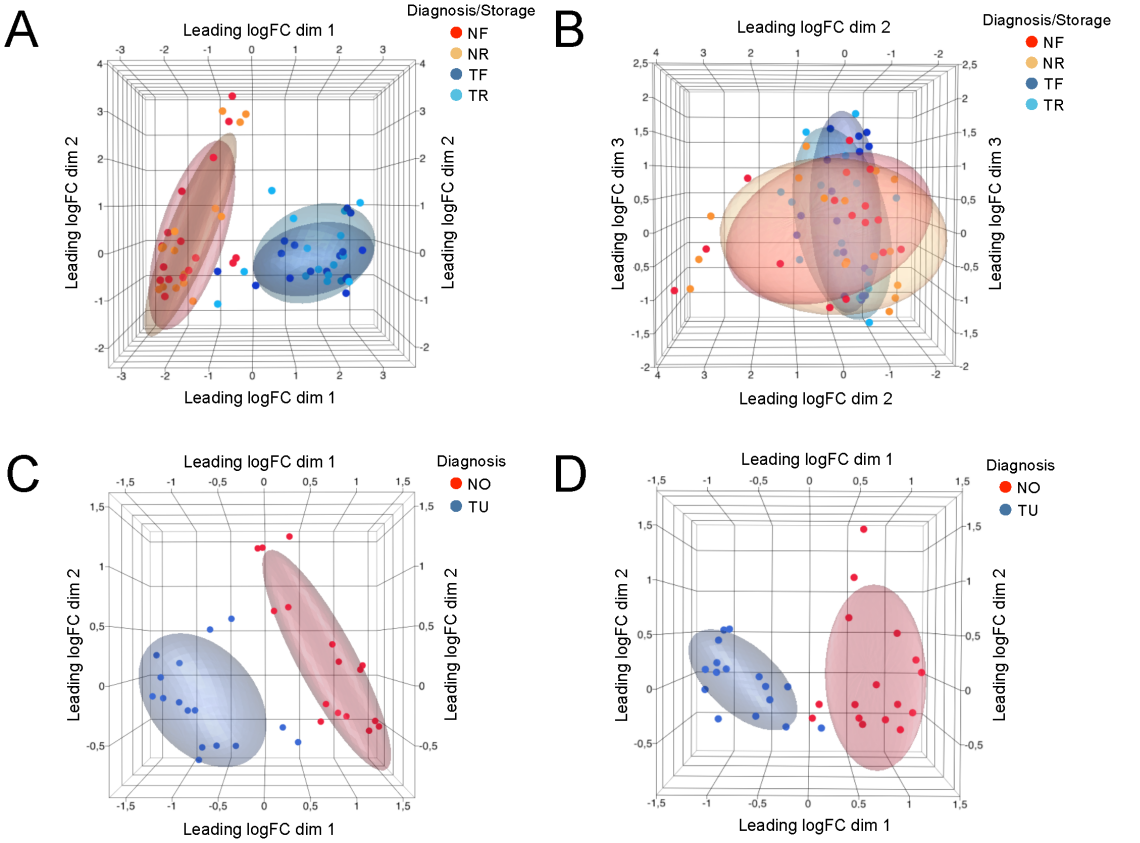

Fig 1. Multidimensional scaling (MDS) analysis of gene expression data. MDS analysis based on all commonly detected genes shows that samples segregate by diagnosis (A) and not by storage condition (B). Distances correspond to leading log-fold-changes between each pair of samples. MDS based on differentially expressed genes demonstrates less within-group variance compared to MDS with all detected genes in the RNAlater® (C) and FFPE (D) datasets. *NF*: *Normal*, *FFPE; NR*: *Normal*, *RNAlater®; TF*: *Tumor*, *FFPE; TR*: *Tumor*, *RNAlater®*. *NO = Normal; TU = Tumor*.

doi:10.1371/journal.pone.0149743.g001

absolute log-fold change between each pair of samples. As deducible from sample annotation in Fig 1A, the major known factor explaining the similarity of biopsy samples was attributed to "Diagnosis" (i.e. tumor and normal). Storage condition (FFPE or RNAlater®) did not appear to cause sample segregation (Fig 1B).

In a next step, we identified for each dataset the genes with differential expression changes between ccRCC and normal, and compared the two sets. The FFPE dataset demonstrated 1367 differentially regulated genes and the RNAlater® dataset 1418 genes (Benjamini-Hochberg adjusted p value ≤ 0.05, and abs FC ≥ 2); comparison of the non-tumorous, normal FFPE tissues versus the corresponding normal tissues from the RNAlater® group revealed a very high concordance with only 37 differentially expressed genes (data not shown).

In the MDS analysis, plotting values for differentially expressed genes indicates less within-group variance compared to the analysis of all detected genes, and the shrinkage of log-fold

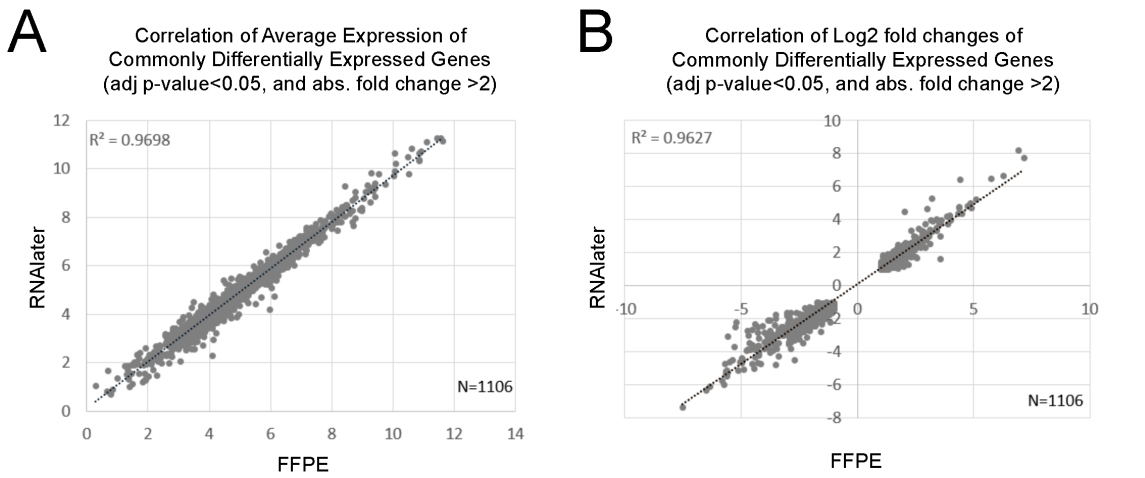

Fig 2. Correlation of gene expression data. The correlation of commonly differentially expressed genes is given with respect to (A) average expression and (B) log2 fold changes.

doi:10.1371/journal.pone.0149743.g002

changes indicates that some non-differentially expressed genes can have quite large fold changes (Fig 1C and 1D).

Each of these two datasets shared 1106 (about 80%) of differentially expressed genes with each other. The correlation of the average expression of these 1106 genes was $R^2 = 0.97$ (Fig 2A). The log2 fold changes of these differentially expressed genes correlated by $R^2 = 0.96$ (Fig 2B). All those genes in both datasets had the same direction of change. Table 2 shows the 20 most significantly affected genes with largest absolute fold changes in the FFPE dataset and the corresponding values of the RNAlater® dataset; 17 of these 20 genes were differentially expressed in both datasets, 3 did not pass the expression filter in the RNAlater® dataset. Amongst the 17 genes, 14 were among the top 20 ranking differentially expressed genes in the RNAlater® dataset. Vice versa, all top 20 differentially expressed genes of the RNAlater® dataset were differentially expressed in the FFPE dataset, 14 of which ranking among the top 20 in both datasets (not shown).

Immunohistochemistry

Immunohistochemistry of the three most regulated genes according to Table 2 confirmed strong overrepresentation of neuronal pentraxin-2 (NPTX2) and carbonic anhydrase 9 (CA9) as well as the underrepresentation of uromodulin (UMOD) in ccRCC [14–16]. The results are depicted in Fig 3, which also presents respective mRNA abundance plots.

Pathway Analyses

To test whether disease-relevant pathways have been captured in our experiment, we performed Ingenuity Pathway Analyses (IPA) of differentially expressed genes. 91 canonical pathways in the FFPE dataset and 109 pathways in the RNAlater® dataset were affected (adjusted p-value$\leq$0.05) with an overlap of 75%. The most affected pathways to a good extent reflect humoral and adaptive immune responses (Table 3). Sorting the pathways by smallest adjusted

Table 2. Gene expression analyses. The 20 most up- or down-regulated genes in the FFPE data set with corresponding RNAlater® values (upper panel), and the 20 most up- or down regulated genes in the RNAlater® dataset with corresponding FFPE values (lower panel), filtered by adjusted p-value≤0.05. Rank indicates the rank of the gene within the list of differentially genes sorted by largest to smallest absolute fold change. 14 genes are shared between the two lists. *TU*: *tumour*, *NO*: *normal*, *FC*: *fold change*, *ND*: *not detected*, *did not pass the expression filter*.

FFPE TU vs NO

		FFPE		RNAlater®		rank	
Ensembl Gene ID	**HGNC symbol**	**FC (TU vs NO)**	**adj. p-val.**	**FC (TU vs NO)**	**adj. p-val.**	**FFPE**	**RNAlater®**
ENSG00000169344	UMOD	-183,2	2,40E-07	-158,7	8,06E-08	1	3
ENSG00000106236	NPTX2	140,9	6,67E-07	220,1	2,29E-08	2	2
ENSG00000107159	CA9	121,2	5,50E-06	304,4	3,65E-09	3	1
ENSG00000074803	SLC12A1	-91,9	1,59E-07	-78,5	1,15E-07	4	7
ENSG00000169550	MUC15	-82,1	3,20E-07	-66,6	1,27E-06	5	8
ENSG00000142319	SLC6A3	76,6	2,17E-06	101,7	6,53E-07	6	4
ENSG00000169347	GP2	-57,2	1,13E-06	-52,7	4,52E-07	7	10
ENSG00000107165	TYRP1	-56,1	5,91E-06	ND	ND	8	ND
ENSG00000088836	SLC4A11	-54,2	1,14E-07	-62,7	2,16E-05	9	9
ENSG00000130822	PNCK	53,3	1,42E-06	92,0	1,89E-07	10	5
ENSG00000198691	ABCA4	-52,4	3,12E-07	ND	ND	11	30
ENSG00000165973	NELL1	-51,4	2,72E-07	-35,8	8,78E-07	12	16
ENSG00000186510	CLCNKA	-50,3	1,61E-08	-39,7	9,73E-08	13	13
ENSG00000215644	GCGR	-49,7	1,52E-07	-33,9	2,68E-06	14	18
ENSG00000164893	SLC7A13	-49,3	3,87E-04	-43,6	9,14E-06	15	11
ENSG00000138798	EGF	-47,9	1,43E-07	-37,3	2,26E-07	16	15
ENSG00000150201	FXYD4	-47,8	1,89E-05	-8,1	1,51E-02	17	134
ENSG00000184956	MUC6	-47,1	1,14E-05	ND	ND	18	ND
ENSG00000100362	PVALB	-45,7	5,83E-07	ND	ND	19	ND
ENSG00000130829	DUSP9	-45,0	7,90E-07	-24,4	1,56E-06	20	36

RNAlater® TU vs NO

		RNAlater®		FFPE		rank	
Ensembl Gene ID	**HGNC symbol**	**FC (TU vs NO)**	**adj. p-val.**	**FC (TU vs NO)**	**adj. p-val.**	**RNAlater®**	**FFPE**
ENSG00000107159	CA9	304,4	3,65E-09	121,2	5,50E-06	1	3
ENSG00000106236	NPTX2	220,1	2,29E-08	140,9	6,67E-07	2	2
ENSG00000169344	UMOD	-158,7	8,06E-08	-183,2	2,40E-07	3	1
ENSG00000142319	SLC6A3	101,7	6,53E-07	76,6	2,17E-06	4	6
ENSG00000130822	PNCK	92,0	1,89E-07	53,3	1,42E-06	5	10
ENSG00000185633	NDUFA4L2	87,6	6,30E-10	20,9	5,88E-06	6	50
ENSG00000074803	SLC12A1	-78,5	1,15E-07	-91,9	1,59E-07	7	4
ENSG00000169550	MUC15	-66,6	1,27E-06	-82,1	3,20E-07	8	5
ENSG00000088836	SLC4A11	-62,7	2,16E-05	-54,2	1,14E-07	9	9
ENSG00000169347	GP2	-52,7	4,52E-07	-57,2	1,13E-06	10	7
ENSG00000164893	SLC7A13	-43,6	9,14E-06	-49,3	3,87E-04	11	15
ENSG00000130208	APOC1	40,0	7,15E-09	9,1	6,01E-05	12	136
ENSG00000186510	CLCNKA	-39,7	9,73E-08	-50,3	1,61E-08	13	13
ENSG00000123610	TNFAIP6	37,8	2,98E-08	33,6	1,68E-07	14	26
ENSG00000138798	EGF	-37,3	2,26E-07	-47,9	1,43E-07	15	16
ENSG00000165973	NELL1	-35,8	8,78E-07	-51,4	2,72E-07	16	12
ENSG00000113889	KNG1	-34,9	7,04E-07	-35,6	5,31E-07	17	25
ENSG00000215644	GCGR	-33,9	2,68E-06	-49,7	1,52E-07	18	14
ENSG00000008196	TFAP2B	-32,5	4,04E-06	-29,9	7,56E-06	19	31
ENSG00000184661	CDCA2	32,4	1,77E-07	28,9	1,13E-06	20	33

doi:10.1371/journal.pone.0149743.t002

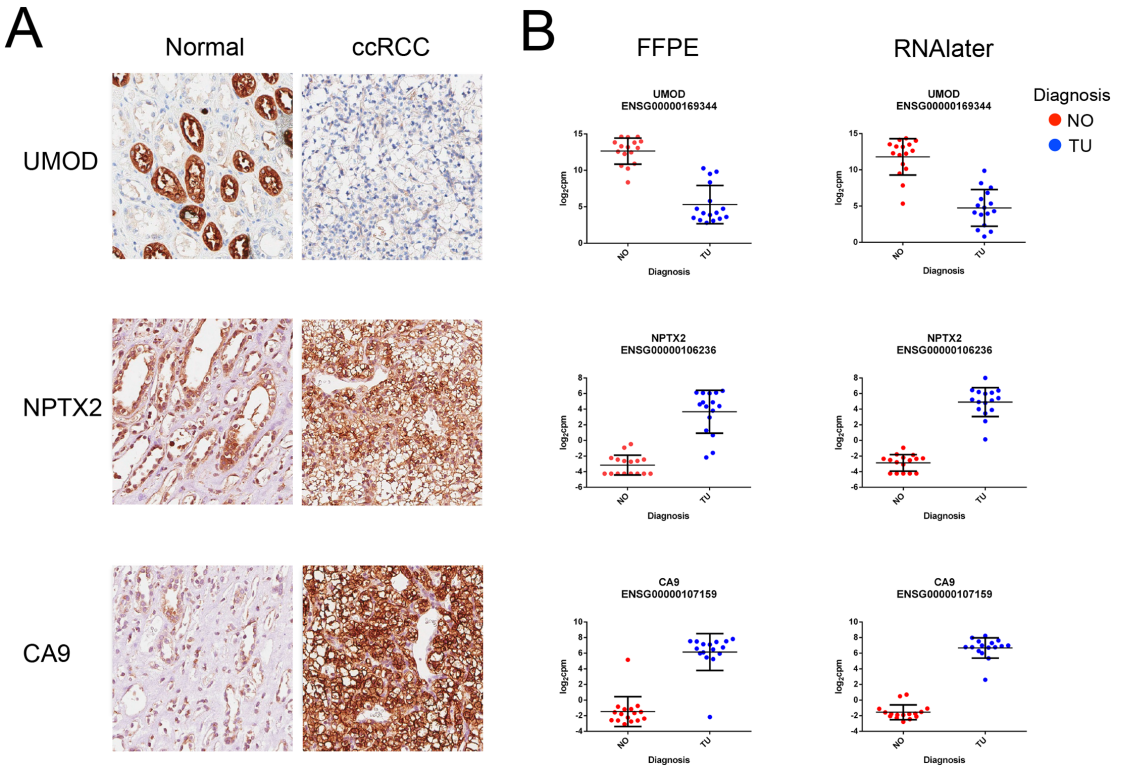

Fig 3. Immunohistochemistry and mRNA plots. (A) Immunohistochemistry of UMOD, NTPX2 and CA9. *Magnification x20*, *scale bar 50 μm*. (B) Respective mRNA abundance plots in the FFPE and in the RNAlater® datasets.

doi:10.1371/journal.pone.0149743.g003

p-values, 12 of the top 20 in the FFPE dataset rank among the top 20 pathways in the RNAlater® dataset.

Comparison with Published Data

We compared our ccRCC gene expression changes with findings described in a recently published meta-analysis of ccRCC datasets [17]. All 10 most up-regulated genes and 7 of the 10 most down-regulated genes from Zaravinos et al. [17] were found in the present study and are differentially expressed in FFPE and RNAlater® datasets (Table 4). The remaining genes did not pass our expression filter. The direction of fold changes was identical for all listed genes.

We further compared the findings from the FFPE and the RNAlater® datasets in relation to the known involvement of vascular endothelial growth factor (VEGF) in ccRCC [18, 19]. As demonstrated in Fig 4, many genes of the VEGF and NOTCH signaling cascades were retrieved in the FFPE and the RNAlater® datasets with very similar fold changes and agreement in direction of changes. We can also confirm a link to epithelial to mesenchymal transition (EMT) by the overrepresentation of mesenchymal markers, e.g. vimentin (*VIM*), endothelin 1 (*EDN1*), fibronectin 1 (*FN1*), or transforming growth factor-β *(TGFB1)*, and

Table 3. Pathway analysis. The 20 most affected canonical pathways in each NGS dataset with the corresponding values and ranks. Rank indicates the place of the pathway within the list of pathways sorted by largest to smallest –log(adjusted p-value). 12 of 20 pathways are shared between both datasets. *TU*: *tumour*, *NO*: *normal*, *FC*: *fold change*, *ND*: *not detected*, *did not pass the expression filter*.

FFPE	**-log(adj. p-value)**		**rank**	
	FFPE	**RNAlater®**	**FFPE**	**RNAlater®**
Antigen Presentation Pathway	13,90	9,13	1	3
Hepatic Fibrosis / Hepatic Stellate Cell Activation	13,90	14,60	2	2
LXR/RXR Activation	7,53	6,67	3	4
Leukocyte Extravasation Signaling	7,13	4,55	4	9
Coagulation System	6,78	6,59	5	5
Communication between Innate and Adaptive Immune Cells	6,60	3,58	6	17
Caveolar-mediated Endocytosis Signaling	6,54	3,69	7	12
Atherosclerosis Signaling	6,50	6,04	8	6
Dendritic Cell Maturation	6,50	4,18	9	10
Crosstalk between Dendritic Cells and Natural Killer Cells	6,31	3,62	10	14
Graft-versus-Host Disease Signaling	5,80	3,02	11	35
Complement System	5,78	4,55	12	8
Autoimmune Thyroid Disease Signaling	5,78	3,49	13	23
Virus Entry via Endocytic Pathways	5,78	2,92	14	38
OX40 Signaling Pathway	5,78	3,34	15	28
Intrinsic Prothrombin Activation Pathway	5,44	4,15	16	11
Allograft Rejection Signaling	5,44	3,49	17	25
Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes	4,85	3,52	18	22
Granulocyte Adhesion and Diapedesis	4,36	2,74	19	47
iCOS-iCOSL Signaling in T Helper Cells	4,35	2,86	20	41
RNAlater®	**-log(adj. p-value)**		**rank**	
	RNAlater®	**FFPE**	**RNAlater®**	**FFPE**
EIF2 Signaling	14,60	ND	1	ND
Hepatic Fibrosis / Hepatic Stellate Cell Activation	14,60	13,90	2	2
Antigen Presentation Pathway	9,13	13,90	3	1
LXR/RXR Activation	6,67	7,53	4	3
Coagulation System	6,59	6,78	5	5
Atherosclerosis Signaling	6,04	6,50	6	8
LPS/IL-1 Mediated Inhibition of RXR Function	5,23	3,87	7	27
Complement System	4,55	5,78	8	12
Leukocyte Extravasation Signaling	4,55	7,13	9	4
Dendritic Cell Maturation	4,18	6,50	10	9
Intrinsic Prothrombin Activation Pathway	4,15	5,44	11	16
Caveolar-mediated Endocytosis Signaling	3,69	6,54	12	7
Ethanol Degradation II	3,62	1,49	13	77
Crosstalk between Dendritic Cells and Natural Killer Cells	3,62	6,31	14	10
Histamine Degradation	3,58	0,41	15	267
B Cell Development	3,58	3,15	16	32
Communication between Innate and Adaptive I Immune Cells	3,58	6,60	17	6
eNOS Signaling	3,58	2,78	18	41
Valine Degradation I	3,57	1,48	19	78
mTOR Signaling	3,57	ND	20	ND

doi:10.1371/journal.pone.0149743.t003

Table 4. Comparison of our gene expression data with data from literature [17]. Twenty genes with smallest p-values and largest absolute fold changes in a meta-analysis of five microarray studies are compared to the corresponding genes and their fold changes and p-values of the NGS datasets. The median fold changes and standard deviations for the meta-analysis are presented. All shown genes were differentially expressed in only 2 or 3 microarray datasets. Large standard deviations indicate a large spread of values in the individual microarray studies. 17 of the 20 genes were found differentially expressed in both NGS datasets, 13 of these with fold changes within the fold change range of the microarray meta-analysis. *ND: not detected*, *did not pass initial expression filter*.

Zaravinos et al. [17]			Eikrem et al. (present study)				
Ten most significantly up-regulated genes			**FFPE**		**RNAlater®**		
HGNC symbol	**Median fold change ± SD (TU vs NO)**	**p-value**	**Fold change (TU vs NO)**	**p-value**	**Fold change (TU vs NO)**	**p-value**	**Fold change within range of [17]**
NDUFA4L2	53,94±58,53	<0.01	20,9	4,09E-07	87,6	6.85E-14	yes
PLIN2	27,86±27,89	<0.01	4,6	2,82E-05	4,7	1,03E-04	yes
NNMT	20,86±9,84	<0.01	9,0	2,25E-07	15,8	5,47E-10	yes
ENO2	19,97±9,82	<0.01	6,3	7,10E-08	7,3	1,39E-10	no
AHNAK2	16,62±2,23	<0.01	12,2	8,66E-09	16,0	1,96E-08	yes
NETO2	15,8±13,8	<0.01	10,6	5,10E-10	11,7	5,06E-13	yes
CA9	14,48±4,40	<0.01	121,2	3,72E-07	304,4	3,17E-12	no
VWF	13,06±2,61	<0.01	4,9	3,84E-08	13,7	1,06E-09	yes
COL23A1	12,75±5,10	<0.01	22,1	6,99E-09	20,9	5,05E-09	no
EHD2	12,70±13,94	<0.01	3,9	2,26E-10	4,0	2,96E-08	yes
Ten most significantly down-regulated genes			**FFPE**		**RNAlater®**		
HGNC symbol	**Median fold change ± SD (TU vs NO)**	**p-value**	**Fold change (TU vs NO)**	**p-value**	**Fold change (TU vs NO)**	**p-value**	**Fold change within range of [17]**
ATP6V0A4	-19,70±32,54	<0.01	-10,4	5,39E-08	-7,4	2,28E-05	yes
CA10	-21,45±8,80	<0.01	ND		ND		
SLC12A3	-23,67±31,69	<0.01	-10,5	6,59E-05	-18,9	1,39E-06	yes
CLDN8	-27,11±95,38	<0.01	ND		ND		
SERPINA5	-35,45±32,90	<0.01	-13,7	3,34E-05	-16,4	9,39E-07	yes
KNG1	-38,45±51,67	<0.01	-35,6	1,15E-08	-34,9	9,64E-09	yes
KCNJ1	-50,79±59,09	<0.01	-2,4	1,48E-09	-2,1	1,48E-04	yes
RALYL	-53,58±11,02	<0.01	ND		ND		
CALB1	-103,68±156,0	<0.01	-12,00	1,45E-03	-8,4	7,88E-05	yes
NPHS2	-159,10±155,4	<0.01	-3,8	4,63E-03	-4,4	1,76E-03	no

doi:10.1371/journal.pone.0149743.t004

underrepresentation of epithelial markers such as epithelial cell adhesion molecule (*EPCAM*) or E-cadherin (*CDH1*). The transcription factor grainyhead-like 2 (*GRHL2*), which inhibits EMT, is about 10 fold underrepresented [20].

IPA revealed *TGFB1* as one the most important regulator of gene expression in our ccRCC datasets, as shown in Fig 5. Of the 1367 differentially expressed genes in the FFPE dataset, the expression levels of 237 genes (17%) are influenced by *TGFB1* in the FFPE dataset (Fig 5A), and 253 of the 1418 (18%) differentially affected genes in the RNAlater dataset (Fig 5B). *TGFB1* itself was overrepresented 2.3 fold and 2.8 fold in the FFPE and the RNAlater dataset, respectively (Fig 4).

Classifier Analysis

We further wanted to test whether the RNAseq data from the FFPE dataset could be used to develop a molecular classifier for ccRCC. Hence, in a proof of concept approach, we first selected 100 genes with the largest absolute fold change and smallest adjusted p-value among the group of differentially expressed genes in the FFPE dataset. To avoid overfitting, we initially tested the performance of classifier models with 15 or fewer genes, where we preferred those

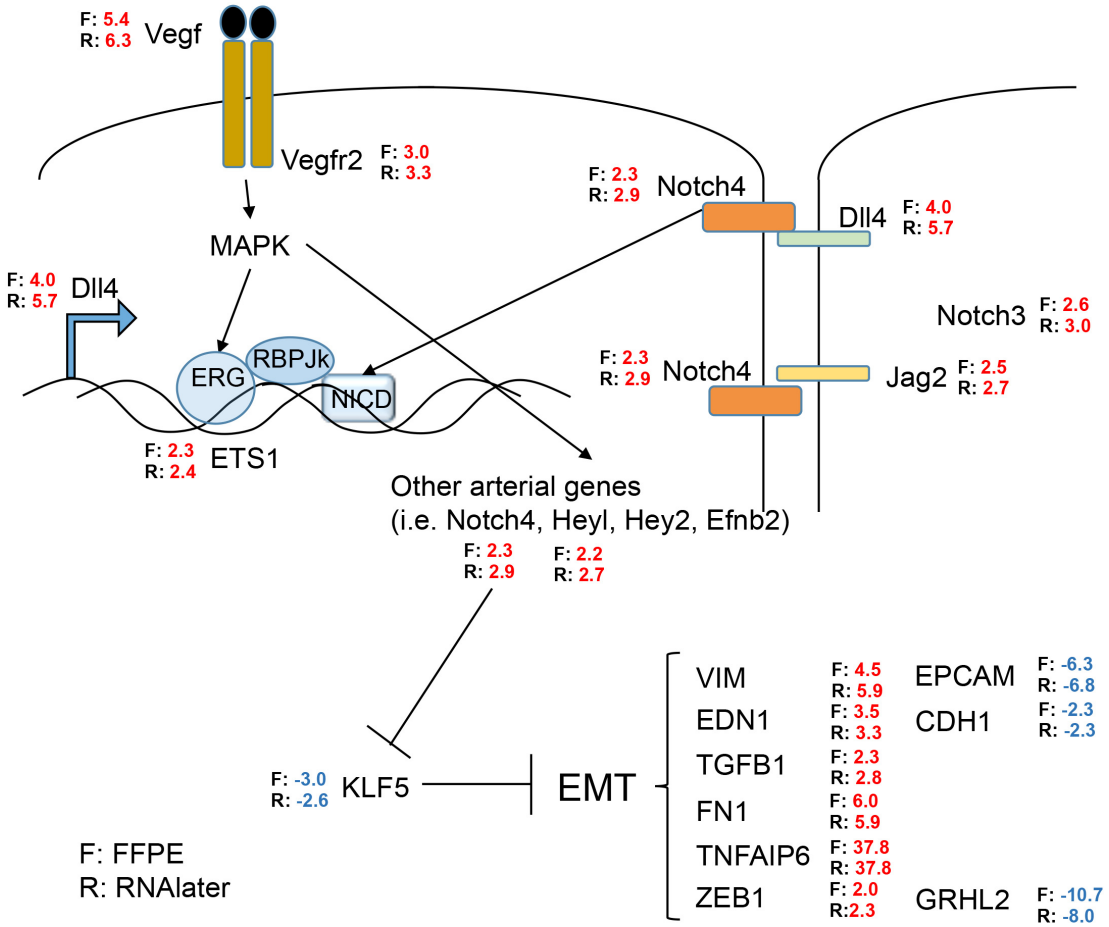

Fig 4. Pathway signature of VEGF and NOTCH mediated EMT in ccRCC. Comparison of gene expression data from the FFPE and from the RNAlater® dataset with published results [20] and between themselves. *F = FFPE samples*, *R = RNAlater® samples*, *Numbers = fold change of up-regulation (red) or down-regulation (blue)*.

doi:10.1371/journal.pone.0149743.g004

with few genes, as they would allow simpler testing in a clinical setting. *CA9* alone correctly classified 30 of 32 samples in the FFPE according to our annotation with an accuracy of 93.8% and area under the ROC curve (ROC AUC) of 0.96. Results of *CA9* from our patients are shown in Fig 6A–6C. One misclassified sample was a normal sample classified as tumor. However, importantly, this specimen contained some admixture of tumor tissue detected at a second look. The other misclassified sample from a different patient was a tumor sample with some adjacent tissue that was judged to be normal.

In the RNAlater® dataset, the single gene classifier model assigned one sample with the histological classification "normal" to the group of tumor samples, yielding an accuracy ACC = 96.8%, AUC = 1.0, and a specificity of 93.8% and a sensitivity of 100%.

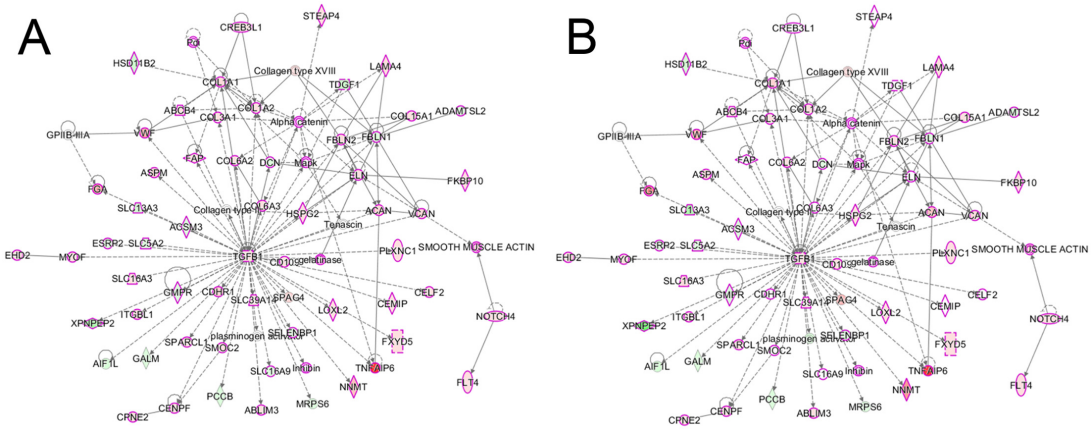

Fig 5. Gene network. The most differentially affected network with the central role of *TGFB1* in (A) FFPE samples and B) RNAlater data sets. *Proteins with cancer involvement are marked with purple outline*. *Red fill indicates overrepresentation of the gene in ccRCC*, *green indicates under-representation*. *Color intensity reflects range of fold change*.

doi:10.1371/journal.pone.0149743.g005

We then tested the single gene classifier model in an external dataset on a different technology platform. The publically available Gene Expression Omnibus (www.ncbi.nlm.gov/geo/) dataset GSE53757 contains Affymetrix HG-U133 microarray gene expression data from 72 human renal biopsies with four stages of ccRCC, and 72 matched normal samples [14]. The CA9-model correctly classified 139 of 144 samples independent of cancer stage (ACC = 96.5%, ROC AUC = 0.98). Results of this CA9 validation are shown in Fig 6D and 6E.

Serum Analyses of CA9 Levels

Optimally, biomarkers such as the gene panel classifiers are further developed into clinically applicable tests. In our simulation study, we wanted to examine, whether CA9-assisted detection of ccRCC could be translated into a less-invasive clinical application going beyond the information obtainable from tissue samples. To that end, we measured CA9 protein in the serum of our patients with early T1a tumor stage and compared the results of these subjects with patient groups suffering from a more advanced disease, because a strong association between serum levels of CA9 with tumor stage has recently been reported [15].

Accordingly, ELISA analyses of serum samples from patients from our institution showed the following values: Increased CA9 levels (95% CI) of 237 (31–443) pg/ml in metastatic patients (n = 9), and of 112 (74–151) pg/ml in non-metastatic patients with high tumor load (tumors larger than 9 cm; n = 15), as compared to a concentration of 54 (26–83) pg/ml in subjects with T1a stage tumors (n = 14); p = 0.0069.

The between group analyses showed significant differences between patients with T1a tumor stage and either with high tumor load (p = 0.0031) or with metastases (p = 0.0158). The comparison between the latter two groups showed no significant difference.

Additional potential novel classifiers have been found, but await further examination and validation. For example, expression values of the highly up-regulated *TNFAIP6* (tumor necrosis factor, alpha-induced protein 6; Fig 4) showed similar performance as *CA9* in the FFPE, RNAlater®, and the microarray dataset (ACC = 96.9%, 96.7%, 94.4%, respectively). We are presently collecting more material and data to expand and confirm these findings.

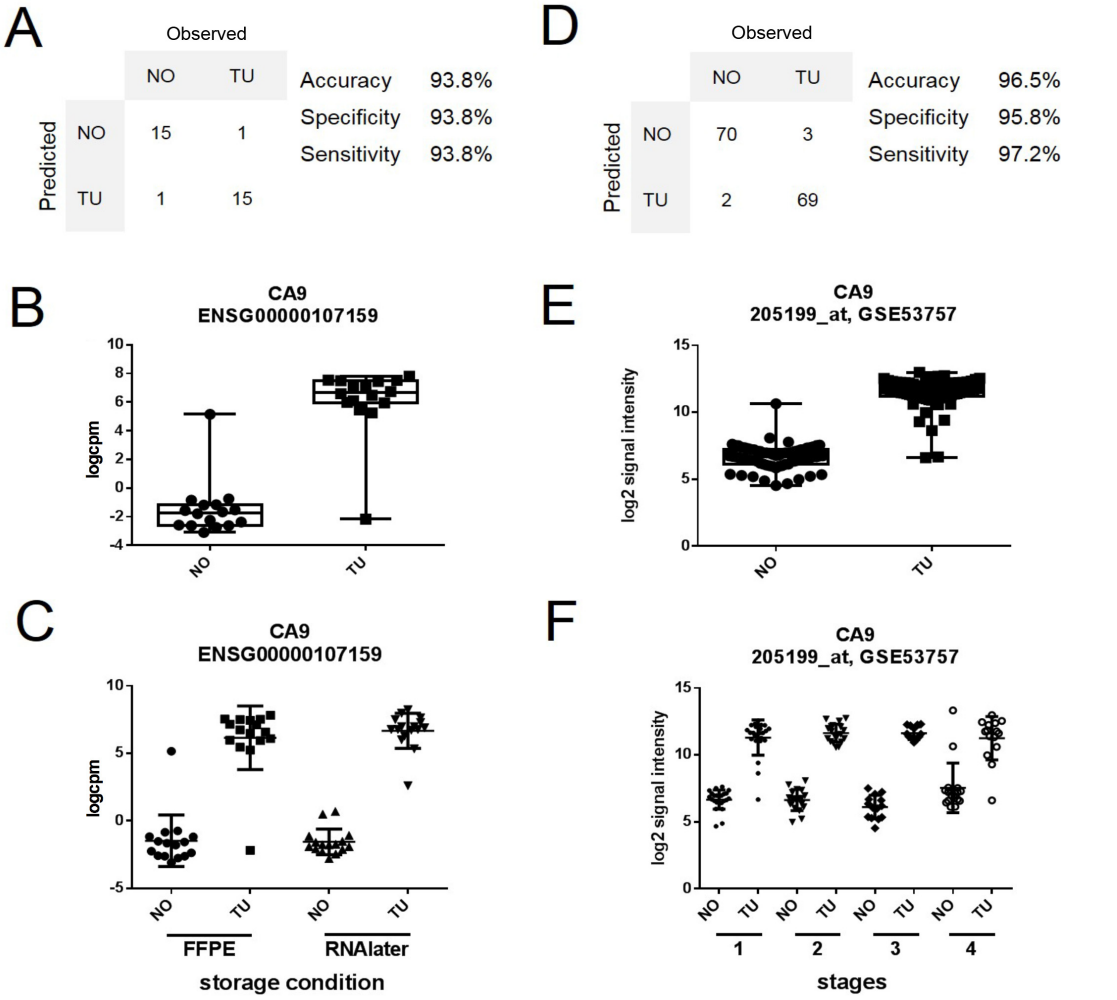

Fig 6. Development of a candidate marker for ccRCC. (A) Expression values of *CA9* correctly classified 30 of 32 samples in our FFPE dataset. (B) Whisker plot of expression value distribution in our FFPE dataset for *CA9*. (C) Scatterplot for the expression values of *CA9* in our FFPE and in our RNAlater dataset. (D) *CA9* expression values correctly classify 139 out of 144 samples in a microarray dataset of ccRCC (GSE53757). (E) Distribution of *CA9* expression values for normal (NO) and ccRCC tumor samples (TU) in the GSE53757 dataset. (F) Stratification of the expression values of overexpressed *CA9* into all four stages of ccRCC [14].

doi:10.1371/journal.pone.0149743.g006

Discussion

Our proof of concept study compares transcriptome sequencing of RNA extracted from human renal biopsies of ccRCC and matched adjacent non-tumorous tissue; samples were preserved in two different storage conditions (FFPE and RNAlater®). High similarity of the two datasets indicates that archival FFPE-samples can be utilized in respective studies.

We chose RNAlater® storage as the comparator. RNAlater® is considered to be an excellent RNA stabiliser [21] and many studies show that RNA yields and gene RNA abundance with RNAlater® are comparable to those obtained using frozen tissues [22]. Furthermore, the utilization of RNAlater® is more practical allowing also decentralized tissue harvesting without special equipment [22, 23].

To the best of our knowledge, there has been no in depth report yet comparing matched RNAlater® and FFPE storage conditions for parallel RNA sequencing and we are among the first to demonstrate the usability of the new Access kit (Illumina) also allowing low FFPE RNA amounts to generate RNA sequencing libraries. A related study has also demonstrated good concordance of RNA sequencing between the two storage conditions but has used different technology for only two renal cancers [24]. Obviously, the TruSeq Access kit is focused on studying mature mRNA levels in biological samples. A recent study has shown that other approaches, such as DSN (Duplex-specific nuclease)-seq and Ribo-zero-Seq can be used to investigate intergenic and intronic RNA species, reportedly giving information on slightly more mRNA species than polyA-enrichment methods, but at the expense of requiring more sequencing effort [25]. Where it is sufficient to study the human transcriptome coding regions, the TruSeq Access kit provides a cost-effective, highly reliable method, as our study shows.

Recent publications have studied the effect of storage time (up to 10 years) in FFPE on RNA quality and quantity, and the usability in mRNA expression experiments, both microarrays and RNAseq [26–28]. In concordance with our own unpublished data where we measured RNA quality and quantity from up to 30 year-old FFPE samples indicating their suitability for RNA sequencing, the publications agree that, RNA is still usable for RNAseq transcriptome studies although the RNA quality suffers with increasing time of FFPE-preservation.

Our approach is further supported by a recent publication showing that a newly developed exon capture RNAseq library preparation protocol for highly degraded RNA provided accurate estimates of RNA abundance, uniform transcript coverage and broad dynamic range investigating FFPE and flash frozen cancer tissues [29].

However, for the genome-wide detection of novel transcripts, whole exome enrichment of RNA might be a necessary additional step [30].

We detected a high degree of similarity between the gene expression results for the two datasets: 94% of the transcripts passing the initial expression filter were shared between the FFPE and RNAlater® sample groups, 80% of differentially expressed genes were in common, and 75% of the differentially affected pathways were found in both datasets. The differences in gene expression can probably be mostly explained by the cell-composition variation of the respective biopsies. This well described intra-tumor heterogeneity precluded the detection of an even higher number of common, differentially regulated genes and pathways [31]. Also, the capture process during library preparation could be different depending on the RNA quality. However, the very high concordance between FFPE non-tumor, normal tissue *vs*. normal tissue stored in RNAlater® further emphasizes the high similarity of the two data sets.

Despite some limitations, we have shown a striking similarity between the FFPE and the RNAlater® datasets, maintaining biologically relevant information at large. Immunohistochemistry confirmed the three most regulated genes of both data sets. CA9 is essentially not expressed in the normal nephron but specifically in ccRCC [5]. Thus, CA9 is an extensively investigated biomarker of ccRCC and also a predictor of outcome following anti-VEGF therapy [19, 32]. In a microarray study with nine patients, UMOD was the gene with the strongest under-representation in RCC [16]. The over-representation of NPTX2 is in accordance with the literature [14].

We also show good concordance with microarray gene expression profiling studies of ccRCC (Table 4). Directions of gene expression changes between ccRCC and normal samples

were identical for a set of differentially expressed genes in the microarray studies (14) and in the NGS studies. 17 of the 20 genes with largest absolute fold changes in the microarray meta-analysis were also differentially expressed in the NGS datasets (Table 4), and most fold changes were within the same range across the studies.

However, limitations and uncertainties in this comparison come from the large discrepancy in the fold changes detected in the microarray studies, and from the fact that all genes in the Table 4 were differentially expressed in only 2 or 3 of five microarray studies used in the meta-analysis. Different amplitudes in fold changes between the microarray dataset and the NGS dataset have been reported before [33]. The authors believe, one reason is that microarray probes might hit some, but not all, isoforms of a gene, and as a result the reported fold change of the probe set does not necessarily represent the expression change of the entire gene [33]. Furthermore, NGS is more sensitive in measurement of abundance differences of lowly or highly expressed genes. Microarrays reach a saturation level in the case of highly expressed genes, but NGS technology with its wider dynamic range of detection is more likely to detect fold changes. This may explain some of the fold change differences observed in the comparison of microarray and NGS data. Nevertheless, our dataset confirmed the trend of expression changes observed in microarray studies.

Our data also support and in part confirm novel therapeutic avenues, such as targeted at activated VEGF /NOTCH /DLL4 signaling cascades [18, 34–37]. The up-regulated NOTCH ligand Delta 4 (DLL4) is stimulated by VEGF and plays a role in tumor progression also predicting bad outcome [36, 38, 39]. EMT is augmented in our cancer data and is known to be a relevant feature in ccRCC [40]. Up-regulated *TGFB1* was the most significantly affected gene regulator in our study. Accordingly, *TGFB1* inhibition was shown to attenuate the invasive capacity of ccRCC cells [34]. However, potential cancer therapy targeted at *TGFB1* remains to be developed.

Classifier models consisting of features such as gene expression data in combination with a decision algorithm are powerful tools to support diagnostic and prognostic evaluation of patient data. Gene expression data for *CA9*—supplemented by CA9 serum protein data—showed an excellent performance both in our datasets and in an independent ccRCC microarray dataset. Thus, our data expand previous reports, which promote *CA9* as a diagnostic tool in ccRCC [5, 19, 41, 42].

Taken together, we show that in our hands RNAseq FFPE data are comparable to matched RNAlater® data. We used the proof of concept data to explore and to confirm published biological findings, and findings which may be worth following up in larger cohorts, leading to possible novel therapeutic strategies, e.g. based on *TGFB1*-regulated genes, the *NOTCH* signaling cascade, and EMT. Also of note, FFPE tissues have the distinctive advantage that material designated for RNA sequencing can be concurrently investigated by light microscopy.

Conclusions: Our study opens the door to transcriptome analyses of the archival, FFPE stored tissues from patients with ccRCC and supports *CA9* as a potential marker for ccRCC.

Materials and Methods

Patients

Adult patients (n = 16) from Haukeland University Hospital with ccRCC undergoing partial (n = 10) or full (n = 6) nephrectomy and with the possibility to undergo biopsies for this project were included consecutively from November 2013 until August 2014. Patients had a mean age of 58.2±6.8 years (3 females and 13 males). Patients had pT tumor stages T1a (n = 10), T2a or b (n = 2) and T3a or b (n = 4) [43]; additional patient characteristics can be found in Table 1. The regional ethics committee of Western Norway has approved our studies (REC West no. 78/05). All participants provided written consent as requested by our ethics committee.

Kidney Biopsies

Core biopsies have been obtained by O.E., L.L. and T.S. with a 16g needle from 16 patients undergoing (partial) nephrectomy in the operating room itself exactly at the time of surgery. Four paired biopsies from each patient with histologically-confirmed clear cell renal cell carcinoma (ccRCC) and adjacent non-tumorous ("normal") tissue were either stored as FFPE tissue or in an RNA-stabilizing agent (RNAlater®, Qiagen, Germany). Total RNA was extracted with miRNeasy FFPE kit or miRNeasy micro kit (Qiagen), respectively.

RNA Library Preparation and Sequencing

RNA sequencing libraries were prepared using TruSeq RNA Access library kit (Illumina, Inc., San Diego, CA, USA) according to the manufacturer`s protocol.

Initially total RNA concentration was measured using Qubit® RNA HS Assay Kit on a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). Integrity was assessed using Agilent RNA 6000 Nano Kit on a 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA) and the percentages of fragments larger than 200 nucleotides were calculated.

Thereafter, RNA samples (100 ng total RNA) were fragmented at 94°C for 8 minutes on a thermal cycler. First strand cDNA syntheses were performed at 25°C for 10 minutes, 42°C for 15 minutes and 70°C for 15 minutes, using random hexameres and SuperScript II Reverse Transcriptase (Thermo Fisher Scientific Inc., Waltham, MA, USA). In a second strand cDNA synthesis the RNA templates were removed and a second replacement strand was generated by incorporation dUTP (in place of dTTP, to keep strand information) to generate ds cDNA. AMPure XP beads (Beckman Coulter, Inc., Indianapolis, IN, USA) were used to clean up the blunt-ended cDNA from the second strand reaction mix. The 3`ends of the cDNA were then adenylated to facilitate adaptor ligation in the next step. After ligation of indexing adaptors, AMPure XP beads were used to clean up the libraries. In a first PCR amplification step, PCR (15 cycles of 98°C for 10 seconds, 60°C for 30 seconds and 72°C for 30 seconds) were used to selectively enrich those DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library. After validation of the libraries, using Agilent DNA 1000 kit on a 2100 Bioanalyzer instrument, the first hybridization step were performed using exome capture probes. Before hybridization a 4-plex pool of libraries were made, by combining 200 ng of each DNA library. The hybridization was performed by 18 cycles of 1 minute incubation, starting at 94°C, and then decreasing 2°C per cycle. Then streptavidin coated magnetic beads were used to capture probes hybridized to the target regions. The enriched libraries were then eluted from the beads and prepared for a second round of hybridization. This second hybridization (18 cycles of 1 minute incubation, starting at 94°C, and then decreasing 2°C per cycle) were required to ensure high specificity of the capture regions. A second capture with streptavidin coated beads were performed, followed by two heated wash procedures to remove non-specific binding form the beads. The enriched libraries where then eluted from the beads and cleaned up by AMPure XP beads prior to a second PCR amplification. The amplification step were performed by 10 cycles (98°C for 10 seconds, 60°C for 30 seconds and 72°C for 30 seconds) followed by a second PCR clean up using AMPure XP beads. Finally, the libraries were quantitated by qPCR using KAPA Library Quantification Kit—Illumina/ABI Prism® (Kapa Biosystems, Inc., Wilmington, MA, USA) and validated using Agilent High Sensitivity DNA Kit on a Bioanalyzer. The size range of the DNA fragments were measured to be in the range of 200–650 bp and peaked around 270 bp.

Libraries were normalized to 22 pM and subjected to cluster and single read sequencing was performed for 50 cycles on a HiSeq2500 instrument (Illumina, Inc. San Diego, CA, USA),

according to the manufacturer's instructions. Base calling were done on the HiSeq instrument by RTA 1.17.21.3. FASTQ files were generated using CASAVA 1.8.2 (Illumina, Inc. San Diego, CA, USA). Data are available in the repository Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76207.

Statistics and NGS Data Processing

We have a sample size of 64 samples, which is equivalent to 32 paired samples *(tumor samples vs. normal samples)*. Within both the FFPE and in the RNAlater dataset, we have 16 sample pairs *(tumors vs. normals)*. This sample size is sufficient to achieve a power of 0.85, where we apply a standard deviation of 0.7 of the expressed genes, an effect size of 2, and an alpha of 0.05 (R package RNASeqPower in https://www.bioconductor.org).

Assembly of reads and alignment of the contigs to the Human genome assembly GRCh38 was guided by Tophat and Bowtie. An empirical expression filter was applied, which left genes with more than 15 counts per million (cpm) in more than 8 samples per dataset. Comparative analysis was done using voom/Limma R-package. Differential gene expression was defined as Benjamini-Hochberg adjusted p-value ≤ 0.05, and an absolute fold change of ≥ 2. Pathway analysis was performed with Ingenuity Pathway Analysis (Qiagen, USA; version 24718999). The Ingenuity Knowledge Base information was used as reference set. Canonical pathways were sorted by smallest Benjamini-Hochberg-adjusted p-value.

Classifier analysis was performed with the KNNX Validation package in GenePattern (http://www.broadinstitute.org/cancer/software/genepattern). Leave-one-out method was used as internal cross validation method. Euclidean distance was used as distance measure, where three neighbors were considered. Data visualization was performed with JMP Pro 11 (www.sas.com), and Graphpad (www.graphpad.com).

Histology and Immunohistochemistry

Immunohistochemistry was performed on 4 μm thick FFPE sections from the tumor and adjacent non-tumorous tissue. The following primary antibodies were used: Carbonic anhydrase IX (CA9, polyclonal, rabbit, NB100-417, Novus Biologicals), neuronal pentraxin 2 (NPTX2, polyclonal, rabbit, NBP1-50275, Novus Biologicals) and uromodulin (UMOD, polyclonal, rabbit, sc-20631, Santa Cruz Biotechnology). For positive controls, tissues with known positive reactivity were used, for negative controls the primary antibody was omitted. Slides were scanned with ScanScope® XT (Aperio) at ×40 and viewed in ImageScope 12.

ELISA for CA9 Serum Levels

CA9 serum concentrations of 38 patients was measured using the Quantikine Human Carbonic Anhydrase IX Immunoassay (R&D Systems, Minneapolis, USA, catalogue number DCA900) according to instructions of the manufacturer, but with an overnight incubation at 4°C after having added the serum. Results were assessed with the Kruskal-Wallis and Dunn's test [44].

Supporting Information

S1 Fig. Correlation of the average expression of the commonly expressed genes in both FFPE and RNAlater datasets. Genes with an average expression of counts per million (cpm) >8 in at least 15 samples per dataset were considered.
(TIF)

Acknowledgments

We thank Dagny Ann Sandnes for help with immunohistochemistry and the other local urologists for participation in biopsy harvesting.

The library preparation, sequencing and partly the bioinformatics analysis were provided by the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine at NTNU and Central Norway Regional Health Authority.

Author Contributions

Conceived and designed the experiments: HPM CB OE. Performed the experiments: OE LL KH TS SL. Analyzed the data: AF AS VB HPM. Contributed reagents/materials/analysis tools: OE LL TS. Wrote the paper: OE HPM AS LL. Renal biopsy Processing: SL.

References

1. Eisengart LJ, MacVicar GR, Yang XJ. Predictors of response to targeted therapy in renal cell carcinoma. Archives of pathology & laboratory medicine. 2012; 136(5):490–5. doi: 10.5858/arpa.2010-0308-RA PMID: 22229848.
2. Ljungberg B, Campbell SC, Choi HY, Jacqmin D, Lee JE, Weikert S, et al. The epidemiology of renal cell carcinoma. European urology. 2011; 60(4):615–21. doi: 10.1016/j.eururo.2011.06.049 PMID: 21741761.
3. Tun HW, Marlow LA, von Roemeling CA, Cooper SJ, Kreinest P, Wu K, et al. Pathway signature and cellular differentiation in clear cell renal cell carcinoma. PloS one. 2010; 5(5):e10696. doi: 10.1371/journal.pone.0010696 PMID: 20502531; PubMed Central PMCID: PMC2872663.
4. Maher ER. Genomics and epigenomics of renal cell carcinoma. Seminars in cancer biology. 2013; 23 (1):10–7. doi: 10.1016/j.semcancer.2012.06.003 PMID: 22750267.
5. Oosterwijk E. Carbonic anhydrase expression in kidney and renal cancer: implications for diagnosis and treatment. Sub-cellular biochemistry. 2014; 75:181–98. doi: 10.1007/978-94-007-7359-2_10 PMID: 24146380.
6. Rydzanicz M, Wrzesinski T, Bluyssen HA, Wesoly J. Genomics and epigenomics of clear cell renal cell carcinoma: recent developments and potential applications. Cancer letters. 2013; 341(2):111–26. doi: 10.1016/j.canlet.2013.08.006 PMID: 23933176.
7. Zwiener I, Frisch B, Binder H. Transforming RNA-Seq data to improve the performance of prognostic gene signatures. PloS one. 2014; 9(1):e85150. doi: 10.1371/journal.pone.0085150 PMID: 24416353; PubMed Central PMCID: PMC3885686.
8. Wu C, Wyatt AW, Lapuk AV, McPherson A, McConeghy BJ, Bell RH, et al. Integrated genome and transcriptome sequencing identifies a novel form of hybrid and aggressive prostate cancer. The Journal of pathology. 2012; 227(1):53–61. doi: 10.1002/path.3987 PMID: 22294438; PubMed Central PMCID: PMC3768138.
9. Xiao YL, Kash JC, Beres SB, Sheng ZM, Musser JM, Taubenberger JK. High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. The Journal of pathology. 2013; 229(4):535–45. doi: 10.1002/path.4145 PMID: 23180419; PubMed Central PMCID: PMC3731037.
10. Roberts L, Bowers J, Sensinger K, Lisowski A, Getts R, Anderson MG. Identification of methods for use of formalin-fixed, paraffin-embedded tissue samples in RNA expression profiling. Genomics. 2009; 94 (5):341–8. doi: 10.1016/j.ygeno.2009.07.007 PMID: 19660539.
11. Walther C, Hofvander J, Nilsson J, Magnusson L, Domanski HA, Gisselsson D, et al. Gene fusion detection in formalin-fixed paraffin-embedded benign fibrous histiocytomas using fluorescence in situ hybridization and RNA sequencing. Laboratory investigation; a journal of technical methods and pathology. 2015; 95(9):1071–6. doi: 10.1038/labinvest.2015.83 PMID: 26121314.
12. Puls F, Hofvander J, Magnusson L, Nilsson J, Haywood E, Sumathi VP, et al. FN1-EGF gene fusions are recurrent in calcifying aponeurotic fibroma. The Journal of pathology. 2015. doi: 10.1002/path.4683 PMID: 26691015.
13. Huang W, Goldfischer M, Babyeva S, Mao Y, Volyanskyy K, Dimitrova N, et al. Identification of a novel PARP14-TFE3 gene fusion from 10-year-old FFPE tissue by RNA-seq. Genes, chromosomes & cancer. 2015. doi: 10.1002/gcc.22261 PMID: 26032162.

14. von Roemeling CA, Radisky DC, Marlow LA, Cooper SJ, Grebe SK, Anastasiadis PZ, et al. Neuronal Pentraxin 2 Supports Clear Cell Renal Cell Carcinoma by Activating the AMPA-Selective Glutamate Receptor-4. Cancer research. 2014; 74(17):4796–810. doi: 10.1158/0008-5472.CAN-14-0210 PMID: 24962026; PubMed Central PMCID: PMC4154999.

15. Takacova M, Bartosova M, Skvarkova L, Zatovicova M, Vidlickova I, Csaderova L, et al. Carbonic anhydrase IX is a clinically significant tissue and serum biomarker associated with renal cell carcinoma. Oncology letters. 2013; 5(1):191–7. doi: 10.3892/ol.2012.1001 PMID: 23255918; PubMed Central PMCID: PMC3525455.

16. Feng JY, Diao XW, Fan MQ, Wang PX, Xiao Y, Zhong X, et al. Screening of feature genes of the renal cell carcinoma with DNA microarray. European review for medical and pharmacological sciences. 2013; 17(22):2994–3001. PMID: 24302177.

17. Zaravinos A, Pieri M, Mourmouras N, Anastasiadou N, Zouvani I, Delakas D, et al. Altered metabolic pathways in clear cell renal cell carcinoma: A meta-analysis and validation study focused on the deregulated genes and their associated networks. Oncoscience. 2014; 1(2):117–31. PMID: 25594006; PubMed Central PMCID: PMC4278286.

18. Iacovelli R, Sternberg CN, Porta C, Verzoni E, de Braud F, Escudier B, et al. Inhibition of the VEGF/VEGFR pathway improves survival in advanced kidney cancer: a systematic review and meta-analysis. Current drug targets. 2015; 16(2):164–70. PMID: 25410406.

19. Stewart GD, O'Mahony FC, Laird A, Rashid S, Martin SA, Eory L, et al. Carbonic anhydrase 9 expression increases with vascular endothelial growth factor-targeted therapy and is predictive of outcome in metastatic clear cell renal cancer. European urology. 2014; 66(5):956–63. doi: 10.1016/j.eururo.2014.04.007 PMID: 24821582.

20. Cieply B, Farris J, Denvir J, Ford HL, Frisch SM. Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. Cancer research. 2013; 73(20):6299–309. doi: 10.1158/0008-5472.CAN-12-4082 PMID: 23943797; PubMed Central PMCID: PMC3806457.

21. Weber DG, Casjens S, Rozynek P, Lehnert M, Zilch-Schoneweis S, Bryk O, et al. Assessment of mRNA and microRNA Stabilization in Peripheral Human Blood for Multicenter Studies and Biobanks. Biomarker insights. 2010; 5:95–102. PMID: 20981139; PubMed Central PMCID: PMC2956623.

22. Medeiros F, Rigl CT, Anderson GG, Becker SH, Halling KC. Tissue handling for genome-wide expression analysis: a review of the issues, evidence, and opportunities. Archives of pathology & laboratory medicine. 2007; 131(12):1805–16. PMID: 18081440.

23. Mutter GL, Zahrieh D, Liu C, Neuberg D, Finkelstein D, Baker HE, et al. Comparison of frozen and RNA-Later solid tissue storage methods for use in RNA expression microarrays. BMC genomics. 2004; 5:88. doi: 10.1186/1471-2164-5-88 PMID: 15537428; PubMed Central PMCID: PMC534099.

24. Li P, Conley A, Zhang H, Kim HL. Whole-Transcriptome profiling of formalin-fixed, paraffin-embedded renal cell carcinoma by RNA-seq. BMC genomics. 2014; 15:1087. doi: 10.1186/1471-2164-15-1087 PMID: 25495041; PubMed Central PMCID: PMC4298956.

25. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC genomics. 2014; 15:419. doi: 10.1186/1471-2164-15-419 PMID: 24888378; PubMed Central PMCID: PMC4070569.

26. Webster AF, Zumbo P, Fostel J, Gandara J, Hester SD, Recio L, et al. Mining the Archives: A Cross-Platform Analysis of Gene Expression Profiles in Archival Formalin-Fixed Paraffin-Embedded Tissues. Toxicological sciences: an official journal of the Society of Toxicology. 2015. doi: 10.1093/toxsci/kfv195 PMID: 26361796.

27. Hedegaard J, Thorsen K, Lund MK, Hein AM, Hamilton-Dutoit SJ, Vang S, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. PloS one. 2014; 9(5):e98187. doi: 10.1371/journal.pone.0098187 PMID: 24878701; PubMed Central PMCID: PMC4039489.

28. Ribeiro-Silva A, Zhang H, Jeffrey SS. RNA extraction from ten year old formalin-fixed paraffin-embedded breast cancer samples: a comparison of column purification and magnetic bead-based technologies. BMC molecular biology. 2007; 8:118. doi: 10.1186/1471-2199-8-118 PMID: 18154675; PubMed Central PMCID: PMC2233637.

29. Cieslik M, Chugh R, Wu YM, Wu M, Brennan C, Lonigro R, et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. Genome research. 2015; 25(9):1372–81. doi: 10.1101/gr.189621.115 PMID: 26253700; PubMed Central PMCID: PMC4561495.

30. Halvardson J, Zaghlool A, Feuk L. Exome RNA sequencing reveals rare and novel alternative transcripts. Nucleic acids research. 2013; 41(1):e6. doi: 10.1093/nar/gks816 PMID: 22941640; PubMed Central PMCID: PMC3592422.

31. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. The New England journal of medicine. 2012; 366(10):883–92. doi: 10.1056/NEJMoa1113205 PMID: 22397650.

32. Stillebroer AB, Mulders PF, Boerman OC, Oyen WJ, Oosterwijk E. Carbonic anhydrase IX in renal cell carcinoma: implications for prognosis, diagnosis, and therapy. European urology. 2010; 58(1):75–83. doi: 10.1016/j.eururo.2010.03.015 PMID: 20359812.

33. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PloS one. 2014; 9(1):e78644. doi: 10.1371/journal.pone.0078644 PMID: 24454679; PubMed Central PMCID: PMC3894192.

34. Bostrom AK, Lindgren D, Johansson ME, Axelson H. Effects of TGF-beta signaling in clear cell renal cell carcinoma cells. Biochemical and biophysical research communications. 2013; 435(1):126–33. doi: 10.1016/j.bbrc.2013.04.054 PMID: 23618868.

35. Funakoshi T, Lee CH, Hsieh JJ. A systematic review of predictive and prognostic biomarkers for VEGF-targeted therapy in renal cell carcinoma. Cancer treatment reviews. 2014; 40(4):533–47. doi: 10.1016/j.ctrv.2013.11.008 PMID: 24398141.

36. Huang QB, Ma X, Li HZ, Ai Q, Liu SW, Zhang Y, et al. Endothelial Delta-like 4 (DLL4) promotes renal cell carcinoma hematogenous metastasis. Oncotarget. 2014; 5(10):3066–75. PMID: 24931473; PubMed Central PMCID: PMC4102792.

37. Erdmann R, Ozden C, Weidmann J, Schultze A. Targeting the Gremlin-VEGFR2 axis—a promising strategy for multiple diseases? The Journal of pathology. 2015; 236(4):403–6. doi: 10.1002/path.4544 PMID: 25875212.

38. Wang W, Yu Y, Wang Y, Li X, Bao J, Wu G, et al. Delta-like ligand 4: A predictor of poor prognosis in clear cell renal cell carcinoma. Oncology letters. 2014; 8(6):2627–33. doi: 10.3892/ol.2014.2554 PMID: 25364440; PubMed Central PMCID: PMC4214437.

39. Noguera-Troise I, Daly C, Papadopoulos NJ, Coetzee S, Boland P, Gale NW, et al. Blockade of Dll4 inhibits tumour growth by promoting non-productive angiogenesis. Nature. 2006; 444(7122):1032–7. doi: 10.1038/nature05355 PMID: 17183313.

40. Zhang X, Ren J, Yan L, Tang Y, Zhang W, Li D, et al. Cytoplasmic expression of pontin in renal cell carcinoma correlates with tumor invasion, metastasis and patients' survival. PloS one. 2015; 10(3): e0118659. doi: 10.1371/journal.pone.0118659 PMID: 25751257; PubMed Central PMCID: PMC4353622.

41. Tostain J, Li G, Gentil-Perret A, Gigante M. Carbonic anhydrase 9 in clear cell renal cell carcinoma: a marker for diagnosis, prognosis and treatment. European journal of cancer. 2010; 46(18):3141–8. doi: 10.1016/j.ejca.2010.07.020 PMID: 20709527.

42. Gimenez-Bachs JM, Salinas-Sanchez AS, Serrano-Oviedo L, Nam-Cha SH, Rubio-Del Campo A, Sanchez-Prieto R. Carbonic anhydrase IX as a specific biomarker for clear cell renal cell carcinoma: comparative study of Western blot and immunohistochemistry and implications for diagnosis. Scandinavian journal of urology and nephrology. 2012; 46(5):358–64. doi: 10.3109/00365599.2012.685493 PMID: 22571179.

43. Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Hora M, et al. EAU guidelines on renal cell carcinoma: 2014 update. European urology. 2015; 67(5):913–24. doi: 10.1016/j.eururo.2015.01.005 PMID: 25616710.

44. Idzenga T. Variability and repeatability of perineal sound recording in a population of healthy male volunteers. Neurourology and urodynamics. 2008; 27(8):802–6. doi: 10.1002/nau.20594 PMID: 18551575.

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 978-82-308-3000-0