University of Glasgow

Fisher, Simon David (2022) *Investigating the association of the UMOD gene with primary hypertension.* PhD thesis.

http://theses.gla.ac.uk/83078/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

Investigating the Association of the UMOD Gene with Primary Hypertension

Simon David Fisher MRes BSc(Hons)

This thesis is being submitted for the

degree of Doctor

of Philosophy

(Ph.D.) in the College of Medical, Veterinary & Life Sciences at the University of

Glasgow

January 2022

© S D Fisher

Abstract

Hypertension, the chronic elevation of blood pressure, remains a sub-optimally managed condition for a large proportion of people and combined with its high incidence, is a large contributor to morbidity and mortality worldwide. Hypertension precision medicine aims to tailor treatment protocols to individuals in order to reduce side-effects and cost whilst improving efficacy of treatment. A hypertension GWAS by Padmanabhan et al, identified a locus on the cis-promoter of the human UMOD gene, indexed at rs13333226, as associating with reduced urinary UMOD levels, reduced systolic pressure and reduced cardiovascular disease events per copy of the 'protective' haplotype. UMOD, a gene expressed almost exclusively in the thick ascending limb of mammalian kidneys was subsequently identified as correlating positively with bioavailability or activity of SLC12A1 (NKCC2), a renal sodium potassium chloride co-transporter, thus associating UMOD with blood pressure through natriuresis and volume retention. In this thesis, we aimed to refine the understanding of the variants which differentially drive expression of UMOD and to elaborate on the mechanisms with associate UMOD levels with blood pressure regulation, such that our findings may inform the rationale for downstream precision medicine experimentation.

Whilst rs13333226 was the discovery variant in the initial GWAS, we hypothesised the nearby variant rs4997081 functionally drives differential expression between alleles, based on previous data. Firstly, we sought to determine the binding affinity of 50-mer oligonucleotides, centered on our target variants, for renal cell nuclear lysate, in order to infer transcriptional activity. By performing electrophoretic mobility shift assays on material derived from both HEK293 and HK2 cells, on rs13333226 and rs4997081, we showed that rs13333226 does not display differential binding affinity for nuclear lysate between genotypes, whilst the 'risk' (G) genotype at rs4997081 displayed significantly greater binding affinity versus the protective (C) genotype. Using mass spectrometry on purified binding complexes at rs4997081, we identified a number of potential proteins, including PARP1 as complexing at this site. The addition of TNF-α to these cells caused an abrogation in the differences in binding

affinity between the risk and protective alleles at rs4997081. Together these data suggest that PARP1 differentially complexes at rs4997081 in a TNF-α dependent manner to drive differential expression of UMOD.

Following this, we aimed to characterise the association between UMOD mRNA and the renal transcriptome. We performed a low sample-size RNA-sequencing experiment (n=3/group) on human individuals stratified by expression of UMOD. We detected strong positive coexpression between UMOD and NKCC2, which we validated by qRT-PCR in 84 human samples, showing a highly significant correlation. Alongside this coexpression, we also detected significant increases in WNK1/4, KCNJ1/5 and decreases in SGK1 mRNA levels, with each of these genes implicated in renal blood pressure regulation independently in the literature, these findings may indicate additional complementary mechanisms by which UMOD associates with hypertension. By gene set enrichment and pathway analysis we showed enrichment toward NFKß and TNF-α signaling as associating with UMOD levels, corroborating our prior findings by EMSA and mass spectrometry. Furthermore, analysis of long non-coding RNAs both using bulk RNA sequencing and independent single cell data mining indicated lnc01762 as associating with UMOD expression and thus potentially associating with blood pressure regulation. Together, these data further reinforce the canonical understanding between UMOD and blood pressure regulation through NKCC2, but also indicate a number of novel mechanisms which may be potential therapeutic targets.

Seeking to identify relationships between the UMOD promoter haplotype, UMOD expression and patient phenotype, we performed analysis of a cohort of n=84 human renal samples characterized by Sanger sequencing at both rs13333226 and rs4997081 and for expression of UMOD by qRT-PCR. Our findings suggest that protective haplotype females have significantly reduced systolic blood pressure than either risk haplotype females or risk or protective haplotype males, indicating a potential UMOD sex effect. We stress that the single-measurement blood pressure recordings collected during this sampling do not qualify as reliable, however we believe this finding requires additional research. We furthermore sought to determine relationships between antihypertensive treatments and UMOD mRNA levels and show

that treatment with both bendroflumethiazide and amlodipine in this cohort associates with significant reductions in UMOD expression, independently of haplotype. When we removed individuals using these medications from our analysis, the risk genotype at rs4997081 was associated with significantly increased expression of UMOD versus the protective allele.

Aiming to expand on previous work conducted in 2014 using male, 12-week Sv129-*Umod*$^{-/-}$ mice, where it was shown that knock-out mice have lower baseline blood pressure versus wild type counterparts and blood pressure tolerance to 2% sodium chloride loading, we performed a mirrored study on Sv129-*Umod*$^{-/-}$, Sv129-*Umod*$^{-/+}$ and wild type Sv129 mice (n=6-8 per group). By weekly blood pressure plethysmography across a 6-week 2% sodium chloride loading period, we show no significant differences in baseline systolic or diastolic blood pressure and no differences in blood pressure tolerance in response to sodium chloride, between homozygous knock-out, heterozygous or wild type mice. By follow-up taqman qRT-PCR on mouse renal RNA, we show that UMOD mRNA expression in the homozygous knock-out animal is not fully abolished, and furthermore that *Nkcc2* expression was significantly reduced in sodium chloride treated mice in a *Umod* gene-dose-dependent manner suggesting compensatory mechanisms were occurring to diminish differences. Based on these initial findings we emphasise the need for further validation studies with increased power.

Finally, we sought to provide novel resources toward the study of UMOD, by generating a stably transfected UMOD human cell line, as currently, no commercially available lines express UMOD. Using full length human UMOD cDNA integrated into a pTARGET plasmid, we generated a HEK293-UMOD cell line with abundant expression of both UMOD mRNA and UMOD protein, significantly elevated above the negligible level of endogenous UMOD expression. Whilst not possible within the time constraints of this thesis, we suggest a number of downstream experiments based on the protein:protein interactions of UMOD within the cell, experiments which could be performed using this novel and valuable experimental resource.

Together, these data provide novel, multi-omics insights as to the relationship between UMOD variants and UMOD mRNA levels, with further basic-science evidence toward possible blood pressure mechanisms. Using a combination of base-directed experiments, transcriptomics and proteomics, for the first time, we provide empirical evidence linking rs4997081 to transcriptional apparatus driving expression of the UMOD gene and identified PARP1 as possibly driving this, in a TNF-α dependent manner. We also show that the expression of UMOD is strongly positively associated with the expression of NKCC2, suggesting this coexpression mechanism underpins the relationship between UMOD and blood pressure which is potentially enhanced by a number of novel targets surrounding WNK and SGK based signaling. We further enrich understanding of the relationship between UMOD and blood pressure by suggesting that rs4997081, not rs13333226, should be targeted for precision medicine and that the mechanism driving the pathway between TNF-α, UMOD and NKCC2 may be a candidate for precision pharmacogenomics. Together, these data provide insights toward the mechanism of blood pressure regulation at the UMOD locus and set a basis for research with increased sample size to further validate these novel findings.

Contents

Abstract II
Contents VI
List of Figures XI
List of Tables XIII
Acknowledgement XV
Authors Declaration XVI
List of Publications, Conferences and Awards XVII
Oral Presentation XVII
Poster Presentation XVII
Journal Articles XVII
Grants and Awards XVIII
Definitions/ Abbreviations XIX
1 Introduction 1
1.1 Cardiovascular Disease 1
1.2 Primary hypertension and the Global Burden 1
1.2.1 Primary Hypertension Burden in the United Kingdom 2
1.2.2 Risk Factors of Primary Hypertension and Cardiovascular Implications . 3
1.2.3 Regulation of Blood Pressure and Development of Primary Hypertension 5
1.2.4 Diuretics 8
1.3 Precision Medicine 10
1.3.1 Precision Medicine 10
1.3.2 Economics of Precision Medicine 11
1.3.3 Applications of Precision Medicine 12
1.4 Genomics as a Research Paradigm 14
1.4.1 Genome Wide Association Studies 14
1.4.2 The Genotype Tissue Expression Project 17
1.4.3 Linkage Disequilibrium in GWAS 18
1.4.4 Hypertension GWAS 19
1.4.5 Eukaryotic Gene Expression 22
1.5 The UMOD Gene and Hypertension 23
1.5.1 Padmanabhan et al Hypertension GWAS 23
1.5.2 The Thick Ascending Limb 24

1.5.3 The UMOD Gene 26
1.5.4 Association Studies, GTEx and the UMOD Promoter 28
1.5.5 The UMOD Protein 32
1.5.6 Physiological Functions of the UMOD Protein 33
1.5.7 UMOD and Blood Pressure Regulation 35
1.5.8 The Sodium Potassium Chloride Cotransporter 38
1.6 The UMOD Precision Medicine Hypothesis 40
1.6.1 Thesis Hypothesis 41
1.6.2 Thesis Aims 41
2 General Methods 42
2.1 Extraction of Genetic Material and Polymerase Chain Reaction 42
2.1.1 Human Renal Biopsy Processing 42
2.1.2 DNA Extraction 43
2.1.3 RNA Extraction 44
2.1.4 End-point Polymerase Chain Reaction and Electrophoresis 45
2.1.5 cDNA Preparation 52
2.1.6 Taqman Quantitative Real-time PCR 53
2.1.7 Sample Preparation of Next Generation Sequencing Samples 55
2.2 Next Generation Sequencing 55
2.2.1 Targetted DNA Sequencing of the Human UMOD Gene 55
2.2.2 DNA Sequence Data Processing Pipeline 55
2.2.3 Total RNA Sequencing between Low and High human UMOD Expressers 56
2.2.4 RNA Sequence Data Processing Pipeline 57
2.3 Computational Biology 57
2.3.1 Software 57
2.3.2 PadPlot 58
2.4 Ingenuity Pathway Analysis 63
2.5 Single Cell Data Mining 63
2.5.1 Automated EVOS Cell Detection 63
2.6 Plasmid Generation 64
2.6.1 Luria Bertani Medium, Glycerol Stocks and Agar Plates 64
2.6.2 Insert Preparation and Ligation into Plasmid 65
2.6.3 Transformation of Competent Cells 65

2.6.4 Plasmid Purification 66
2.7 Cell Culture 67
2.7.1 Media Preparation and General Methods 67
2.7.2 Transfection 68
2.7.3 MTT Assay 68
2.7.4 RNA and Protein Harvesting from Cells 69
2.8 Electrophoretic Mobility Shift Assay and Mass Spectrometry 69
2.8.1 Extraction of Nuclear Protein 69
2.8.2 Generation of the EMSA Product 69
2.8.3 Visualisation and Quantification 71
2.9 Western Blotting and Co-immunoprecipitation 72
2.9.1 Protein Extraction and Quantification 72
2.9.2 Western Blotting 72
2.10 *In-vivo* Experimentation 74
2.10.1 General Methods 74
2.10.2 Pilot Study Design 74
2.10.3 Tail Cuff Plethysmography 75
2.10.4 Metabolic Caging 76
2.10.5 Sacrificing 76
2.11 Statistical Analyses 77
3 Chapter 3 The UMOD Promoter and rs4997081 79
3.1 Introduction 79
3.2 Aims 81
3.3 Methods 83
3.3.1 Magnetic Purification of EMSA Product 83
3.3.2 Mass Spectrometry 83
3.3.3 TNF-α Treatment of HK2 Cells 84
3.3.4 Analysis of Electrophoretic Mobility Shift Assay Images 86
3.4 Results 87
3.4.1 Analysis of Human Renal Samples Stratified by Genotype 87
3.4.2 Next Generation Sequencing of the UMOD Promoter 95
3.4.3 Electrophoretic Mobility Shift Assay 98
3.4.4 Magnetic Purification and Mass Spectrometry 100
3.4.5 Analysis of TNF-α Treatment in the Context of Genotype 103

3.5 Discussion 106
4 UMOD and the Transcriptome 113
4.1 Introduction 113
4.2 Aims 116
4.3 Specific Methods 117
4.3.1 RNA Sequencing 117
4.3.2 Deseq2 Based Identification of Differentially Expressed Genes 117
4.3.3 Alternative Statistical Methodology 118
4.3.4 Utilisation of Gene-sets and Pathway Analysis 118
4.3.5 Single Cell Data Mining 118
4.3.6 Custom Co-expression Analysis 119
4.4 Results 121
4.4.1 Selection of the RNA-sequencing Subset 121
4.4.2 Quality Control 124
4.4.3 Analysis of RNA-Sequencing Data 127
4.4.4 Pathway Analysis 132
4.4.5 Co-expression Analysis 137
4.5 Discussion 141
4.6 Chapter Specific Self-authored Code 148
4.6.1 PadPlot- Bulk Data 148
4.6.2 ScGress- Single Cell 159
5 In-vitro study of UMOD 163
5.1 Introduction 163
5.2 Aims 166
5.3 Specific Methods 167
5.3.1 Design of the UMOD Plasmid 167
5.3.2 Selection of Stable Transfectants Using Geneticin 167
5.3.3 Culturing of Cells Prior to Experiments 168
5.4 Results 170
5.4.1 Generation and Purification of the UMOD pTARGET Plasmid 170
5.4.2 Maintenance Geneticin Concentrations and Impact of Transfection 172
5.4.3 qRT-PCR Analysis of Transfected HEK293 Cells 176
5.4.4 Protein Analysis of Transfected HEK239 Cells 178
5.5 Discussion 181

5.6 Chapter Specific Code 186
5.6.1 EVOS-Microscopy Cell Classification 186
6 Exploring the UMOD Knock-out Mouse Model 188
6.1 Introduction 188
6.2 Aims 191
6.3 Specific Methods 192
6.3.1 Taqman qRT-PCR on Mouse Gene Targets 192
6.3.2 Determining UMOD Expression in Ang-II Treated C57BL/6 Mice 193
6.3.3 Imputation of Diastolic Pressure 194
6.3.4 Retrospective Power Calculations 194
6.4 Results 195
6.4.1 Analysis of Blood Pressure Differences Between Umod Allele and Sodium Chloride Treatment 195
6.4.2 Organ Weights at Cull 204
6.4.3 Taqman qRT-PCR Gene Expression 206
6.5 Discussion 210
7 General Discussion 216
7.1 Conclusion 216
8 Reference List 226

List of Figures

Figure 1.2-1. Simplified diagram of blood pressure regulatory mechanisms in humans. 8
Figure 1.2-2 Single nephron visualization of renal acting antihypertensives. 10
Figure 1.3-1 Applications of precision medicine in cardiovascular disease. 14
Figure 1.4-1 The iceberg of missing heritability in GWAS. 16
Figure 1.4-2 GWAS Catalogue described human variants for blood pressure (top) and primary hypertension (bottom) by year. 21
Figure 1.5-1 UMOD human variants and their associated traits (GRCh38), visualized by mined data from GWAS Catalogue (2021). 30
Figure 1.5-2 GTEx Variants which drive differential expression of UMOD (GRCh37). 31
Figure 1.5-3 Two evidence-based UMOD:blood pressure hypotheses. 38
Figure 2.1-0 Selection process of human samples 43
Figure 2.1-1 Primer annealing points (red) on the UMOD gene 49
Figure 2.2-1 Genomic fastqc processing pipeline 56
Figure 2.2-2 Transcriptomic fastqc processing pipeline 57
Figure 2.3-1 Padplot web application architecture 59
Figure 2.3-2 Padplot web-hosting architecture 62
Figure 2.10-1 Experimental design of the UMOD mouse study 75
Figure 3.3-1 Methods of electrophoretic mobility shift assay analysis. 86
Figure 3.4-1 Randomisation of Sex (Male/female) and medication use 89
Figure 3.4-2 Parametric randomization. 90
Figure 3.4-3. Stratification of SBP by genotype 91
Figure 3.4-4 Stratification of SBP by sex by genotype. 92
Figure 3.4-5 UMOD expression by genotype at rs13333226 and rs4997081 93
Figure 3.4-6 Statification of UMOD expression based on medication 94
Figure 3.4-7 Next generation sequencing of the UMOD promoter 96
Figure 3.4-8 Linux based assignment of reference SNP ID by dbSNP performed with the GATK-HaplotypeCaller (2020). 97
Figure 3.4-9 EMSA assay at rs4997081 and rs13333226 99

Figure 3.4-10 Magnetic purification and mass spectroscopy. 101
Figure 3.4-11 Analysis of mass spectroscopy data. 102
Figure 3.4-12 EMSA assay in the context of rs4997081 examining the effect of TNF-α on binding affinity between genotype 105
Figure 3.5-1 Proposed mechanism of differential gene expression between rs4997081 alleles 109
Figure 4.4-1 Correlation analysis of UMOD and blood pressure 122
Figure 4.4-2 Selection of the high and low UMOD expressor subset 123
Figure 4.4-3 Quality control 125
Figure 4.4-4 Assessment of biopsy location by transcriptomics 126
Figure 4.4-5 Volcano plot of high versus low UMOD. 128
Figure 4.4-6 Clusterplot of high versus low UMOD. 129
Figure 4.4-7 LncRNA analysis between high and low UMOD expressors 131
Figure 4.4-8 Gene set enrichment analysis between high and low UMOD 133
Figure 4.4-9 Specific enriched gene-sets obtained from Gene Ontology 2021. 134
Figure 4.4-10 Ingenuity pathway analysis of data-mined connections between differentially expressed genes 135
Figure 4.4-11 Ingenuity pathway analysis, cellular functions 136
Figure 4.4-12 A) Example plot from one human scRNA-seq dataset analysed. 138
Figure 4.4-13 Linear regression of UMOD expression (dCT) against NKCC2 expression (dCT) in RNA extracted from human renal samples (n=84). 140
Figure 5.1-1 Theorised feedback loop between UMOD, NKCC2 and sodium chloride, ... 164
Figure 5.3-1. pTARGET Mammalian Expression Vector design, ... 167
Figure 5.4-1 Agarose gel characterization of UMOD-plasmid design 171
Figure 5.4-2 Characterisation of antibiotic toxicity in HEK293 cells 174
Figure 5.4-3 Impact of the transfection process on cell viability 175
Figure 5.4-4 Assessment of transfection and ions on UMOD expression 177
Figure 5.4-5 Transfected UMOD HEK293 cells western blot experiments 179
Figure 5.4-6 Effect of sodium chloride treatment on UMOD protein 180
Figure 6.1-1 Graham et al *Umod* sv129 blood pressure analysis 190
Figure 6.4-1 Genotyping of *Umod* $^{+/+, +/-, -/-}$ sv192 mice. 197

Figure 6.4-2 Measurements of systolic blood pressure (mmHg), stratified on genotype 198
Figure 6.4-3 Assessment of summarised systolic blood pressure 199
Figure 6.4-4 Measurements of diastolic blood pressure (mmHg), stratified on genotype 200
Figure 6.4-5 Assessment of summarised diastolic blood pressure 201
Figure 6.4-6 Retrospective power calculations 202
Figure 6.4-7 Assessment of urine output and fluid intake by met-cage 203
Figure 6.4-8 Cull organ weights 205
Figure 6.4-9 qRT-PCR analysis of genes relative to Actb 207
Figure 6.4-10 Correlation matrix of dCT between individual genes assessed by Taqman qRT-PCR. 208
Figure 6.4-11 Assessment of Taqman qRT-PCR *Umod* expression vs Actb in 14-week male c57bl6 mice treated with vehicle or AngII 209
Figure 7.1-1 Proposed pathway between rs4997081 and differential regulation of natriuresis in human TAL cells. 221

List of Tables

Table 1.2-1 Primary hypertension risk factors, delineated by modifiability and directionality of relationship. 4
Table 2.1-1 Genotyping Standard Operating Procedure 46
Table 2.1-2 Targetted Resequencing Standard Operating Procedure 49
Table 2.1-3 Full length UMOD cDNA Synthesis Standard Operating Procedure 50
Table 2.1-4 Mouse Genotyping Standard Operating Procedure 51
Table 2.1-5 Pre-Taqman cDNA Synthesis Standard Operating Procedure 53
Table 2.1-6 Taqman qRT-PCR Standard Operating Procedure 54
Table 2.6-1 Media and stock recipies 64
Table 2.6-2 Ligation reagents 65
Table 2.8-1 Electrophertic Mobility Shift Standard Operating Procedure 70
Table 4.3-1 Single cell dataset mining accession ids 119

Table 4.4-1 Top 30 differentially expressed genes between high and low UMOD. 130
Table 4.4-3 Five human renal single cell datasets were analysed for linear regression coexpression with UMOD.. 139

Acknowledgement

Firstly, I would like to thank the British Heart Foundation for their funding of my PhD, including for their understanding of the impact of the Covid19 pandemic.

I would like to thank my primary supervisor, Martin McBride. A figure I was initially drawn to for his scientific background, but a person who I came to realise combines research excellence and integrity with warmth, patience, and kindness. While these values cost nothing, they meant a great deal to me. I would also like to thank my secondary supervisor Delyth Graham for her limitless patience and understanding, even during the tougher moments of my project. While Martin and Del were my bosses for the last four years, I would also gladly now call them my friends. I would also like to thank Lesley Graham for her mentorship and continual support during the project, I know that she volunteered her help and expected nothing in return, for that I am grateful.

My friends and officemates. Antoniya and Eleni were there with me for the great moments, the funny moments and the hard moments and will always be important to me no matter where they are in the world. Arun, Jim, Daniel, Erin, Gabbie and Talha, I am thankful to you all too. To Alice, my oldest academic friend, for your bottomless patience with all of my daft questions and your endless moral support, thank you. To those I have not mentioned by name, every person who gave a smile or chatted rubbish with me in the office or lab when I may have been having a bad day, I am grateful.

Finally, to my family, to Ewan, Louise and Ollie. My mum and dad, Celia and Jim, for always believing in me and most importantly for instilling me with the confidence to believe in myself. Lastly, to Julie, my partner, the person who means everything to me and who deserves the greatest thanks. You make it all worth it. I know that I could not have done this without you.

Authors Declaration

I declare that this thesis has been written by myself and is a record of research performed myself. For certain experiments, other students were involved in practical aspects of the work and were present during these processes. Haya Khan BSci(Hons) was present during the preparation of UMOD amplicons for targeted resequencing. Liam McQueen (MSci) was present during the preparation of the pTarget-UMOD plasmid. Manshi Zhou (MSci) and Iain Black (MSci) were present during the Taqman qRT-PCR experiments on mouse tissue.

Lesley Graham performed all cell culture relating specifically to the generation of HK2 nuclear material for electrophoretic mobility shift assay (EMSA), but did not perform the EMSA and did not perform any other cell culture.

Glasgow Polyomics facility performed all library preparations and all next-generation sequencing reactions within this thesis but did not process samples prior to library preparations. Glasgow polyomics performed the initial analysis of the RNA-sequencing experiment, but not the downstream analysis or visualisation, and performed none of the targetted resequencing analysis.

Contribution from others has been clearly referenced and reproduced with permission. This work has not been submitted previously for a higher degree and was supervised by Dr Martin W. McBride and Dr Delyth Graham.

List of Publications, Conferences and Awards

Oral Presentation

- Functional Characterisation Of Uromodulin Promoter Risk Variants Associated With Blood Pressure. Simon Fisher, Lesley A Graham, Delyth Graham, Martin McBride, , September 2021, Hypertension, 78(Suppl_1),DOI: 10.1161/hyp.78.suppl_1.08

Poster Presentation

- Human Renal RNA-Seq Identifies Relationship Between UMOD Expression and Medullary Ion Transporters. Simon Fisher, Rafael Crespo Molina, Delyth Graham, Martin Mcbride, April 2021, Journal of Hypertension 39(Supplement 1):e263, DOI: 10.1097/01.hjh.0000747464.02907.d0
- Targeted resequencing using the MinION long read sequencing platform, McBride, M.W., Ellis, H., Fisher, S., Asirvatham, A. and Jones, M. (2019. The FASEB Journal, 33: 637.3-637.3. doi:10.1096/fasebj.2019.33.1_supplement.637.3

Journal Articles

- Distinct uterine artery gene expression profiles during early gestation in the stroke-prone spontaneously hypertensive rat. Kayley Scott, Hannah L Morgan, Christian Delles, Simon Fisher, Delyth Graham, and Martin W McBride Published March 2021, https://doi.org/10.1152/physiolgenomics.00159.2020
- Direct Delivery of MicroRNA96 to the Lungs Reduces Progression of Sugen/Hypoxia-Induced Pulmonary Hypertension in the Rat, Docherty CK, Denver N, Fisher S, et al. Mol Ther Nucleic Acids. 2020;22:396-405. Published 2020 Sep 6. doi:10.1016/j.omtn.2020.09.002

Grants and Awards

- Lead Applicant: Wellcome Trust Institutional Strategic Support Fund (ISSF [204820/Z/16/Z].
- Short-listed: Scottish Cardiovascular Forum Poster Presentation (2021).
- Turbo-talk winner (2nd place): British Heart Foundation Conference Oral Presentation (2021);
- Abstract placed 2nd: British and Irish Hypertension Society Conference (2020/ covid cancellation);

Definitions/ Abbreviations

Abbreviation	Full
ACE	Angiotensin coverting enzyme
ACN	Acetonitrile
ACTB	Beta-Actin
ADAMTS	A Disintegrin and Metalloproteinase with Thrombospondin motifs
ADP	Adenosine diphosphate
ADTKD	Autosomal dominant tubulointerstitial kidney disease
ALGGEN	Virtual laboratory for the identification of putative transcription factor binding sites
ALMS1	Centrosome and basal body associated protein
AMP	Adenosine monophosphate
ANKRD	Ankyrin repeat domain 11
ANOVA	Analysis of variance
API	Application to program interface
APOE	Apolipoprotein E
AQP	Aquaporin
ARHGAP	Rho GTPase activating protein 1
ARRIVE	Animal Research: Reporting of In Vivo Experiments
ATP	Adenosine Triphosphate
AVPR1a	Vasopressin receptor 1 alpa
AWS	Amazon web services
BAM	Binary Alignment Matrix
BCA	Bicinchoninic acid
BHF	British Heart Foundation
BMI	Body Mass Index
BMP	Bone Morphogenetic Protein
BPTF	Bromodomain PHD Finger Transcription Factor
BSA	Bovine serum albumin
CACNA1A	Calcium Voltage-Gated Channel Subunit Alpha1 A
CAD	Coronary Artery Disease
CASR	Calcium sensing receptor
CEU	Utah residents with ancestry from Northern and Western Europe
CHI	Community Health Index number
CID	Histone H3-like centromeric protein
CKD	Chronic Kidney Disease
CLAHE	Contrast limited adaptive histogram equalisation
CLCN1	Chloride voltage-gated channel 1
CLCNKB	Chloride Voltage-Gated Channel Kb
CLDN	Claudin
CMV	Cytomegalovirus
COOH	Methanoic Acid

CPNE	Copine 1
CPU	Central Processing Unit
CSVD	Cerebral Small Vessel Disease
CTDSPL	CTD Small Phosphatase Like
CVD	Cardiovascular Disease
CYP	Cytochrome P450
DALY	Disability adjusted life years
DAPK	Death Associated Protein Kinase 1
DBP	Diastolic blood pressure
DEG	Differentially expressed gene
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAH	Dynein axonemal heavy chain
DTT	DL-Dithiothreitol; Clelands reagent
EDTA	Ethylenediaminetetraacetic acid
EFHD1	EF-Hand Domain Family Member D1
EGF	Epidermal growth factor
EMSA	Electrophoretic mobility shift assay
ENCODEDB	Encyclopedia of DNA Elements database
ERK	Extracellular signal-regulated kinase
FAM	Probe identifier
FDR	False discovery rate
FUSBP	FUSE binding protein 1
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GATC	Eurofins sequencing provider
GFP	Green fluorescent protein
GFR	Glomerular filtration rate
GMP	Guanosine monophosphate
GOI	Gene ontology identifier
GPI	Glycosylphosphatidylinositol
GRM	Genetic relationship matrix
GRP	Gastrin-releasing peptide
GSE	Gene set enrichment
GSEA	Gene set enrichment analysis
GWAS	Genome wide association study
GWS	Genome wide significance
HEK	Human embryonic kidney
HIC1	HIC ZBTB Transcriptional Repressor 1
HIF1A	Hypoxia Inducible Factor 1 Subunit Alpha
HMOD	Hypertension-mediated organ damage
HNF	Hepatocyte nuclear factor-1 beta
HPLC	High-Performance Liquid Chromatography
HPSF	Hypoxanthine-guanine phosphoribosyltransferase 1
HRP	Horseradish Peroxidase
HSD	Honestly significant difference
IDE	Interactive development environment

IHC	Immunohistochemistry
IKMC	International Knockout Mouse Consortium
IPA	Ingenuity pathway analysis
IQR	Interquartile range
ISSF	Institutional Strategic Support Fund
JAK	Janus kinase 2
KCNIP	voltage-gated potassium (Kv) channel-interacting proteins
KCNJ	ATP-sensitive potassium (K-ATP) channel
KCTD	potassium channel tetramerizationdomain
KEGG	Kyoto Encyclopedia of Genes and Genomes
KINEMATICA	Homogenizer
KLHL	Kelch-like gene
LDL	Low density lipoprotein
LIC	Low income country
LINC	Long noncoding ribonucleic acid
LINE	Long interspersed nuclear element
LMIC	Lower and middle income countries
LNC	Long noncoding
LQT	Long QT
MAF	Minor allele frequency
MAGI	Membrane Associated Guanylate Kinase, WW And PDZ Domain Containing 1
MAML1	Mastermind Like Transcriptional Coactivator 1
MKTAL	Mouse kidney thick ascending limb
MONICA	monitoring cardiovascular disease
MST	Macrophage Stimulating 1
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NCBI	National Center for Biotechnology Information
NCC	Sodium-chloride symporter
NEB	New england biolabs
NEFL	Neurofilament Light Chain
NEPER	Cell Lysis Buffer
NFKB	Nuclear factor kappa beta
NGS	Next generation sequencing
NHE3	Na+/H+ exchanger isoform
NHS	National health centre
NKCC2	Sodium potassium chloride cotransporter
NMI	Nearest mutual information
NORDIL	the Nordic Diltiazem (NORDIL) study
NOVEX	Protein separation reagent
OLS	Ordinary least squares
OSR1	Odd-skipped related 1
PADI	Peptidyl arginine deiminase, type I
PAGE	Polyacrylamide gel electrophoresis
PAMELA	Pressioni Arteriose Monitorate E Loro Associazioni
PARP1	Poly [ADP-ribose] polymerase 1

PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDE	Phosphodiesterase
PDILT	Protein Disulfide Isomerase Like, Testis Expressed
PFA	polymer of formaldehyde
PHACTR1	phosphatase actin regulator-1
PMN	Peripheral mononuclear
POU3F3	POU Class 3 Homeobox 3
PRDM	PR domain zinc finger protein 1
PRKY	Protein Kinase Y-Linked (Pseudogene)
PSIP1	PC4 And SFRS1 Interacting Protein 1
PVDF	Polyvinylidene fluoride
PXDNL	Peroxidasin Like
QTL	Quantitative trait loci
RAAS	Renin angiotensin aldosterone system
RALYL	RALY RNA Binding Protein Like
RAM	Random access memory
RCF	relative centrifugal force
RDD	RNase-Free DNase Set buffer
RIN	RNA integrity number
RIPA	Lysis and Extraction Buffer
RNA	Ribonucleic acid
RNAL	Ribonucleic acid-like
ROMK	product of the KCNJ1 gene
RPKM	Reads per kilo base per million mapped reads
RPM	Reads per million
SBP	Systolic blood pressure
SCNN1G	Sodium Channel Epithelial 1 Subunit Gamma
SEABLOCK	Blocking buffer
SGIP1	SH3GL Interacting Endocytic Adaptor 1
SGK1	Serum/Glucocorticoid Regulated Kinase 1
SHRSP	Spontaneously hypertensive stroke prone rat
SKIPOGH	Swiss Kidney Project on Genes in Hypertension
SLC12A1	Sodium potassium chloride cotransporter
SNARE	SNAP REceptor
SNP	Single nucleotide polymorphism
SPAK	STE20/SPS1-related proline-alanine-rich protein kinas
STAT	Signal transducer and activator of transcription protein
TAL	Thick ascending limb
TBE	Tris/Borate/EDTA
TBS	Tris-Buffered Saline
TBST	Tris-Buffered Saline tween
TGFB	Transforming growth factor-beta
THSD	Thrombospondin type-1
TIGD	Pogo superfamily of DNA-mediated transposons

TLR	Toll like receptor
TMEM	Transmembrane protein
TNFA	Tumor necrosis factor alpha
TPM	Transcripts per million
TRANSFAC	Transcription factors database
TRPM	Transient receptor potential ion channels
TRPV	Transient receptor potential cation channel subfamily V member 1
TSS	Transcriptional start site
UCSC	University of California, Santa Cruz
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique molecular identifier
UMOD	Uromodulin
UTR	Untranslated ragion
VCF	Variant caller file
WDR	WD Repeat Domain
WHO	World health organisation
WKY	Wistar kyoto
WNK	WNK Lysine Deficient Protein Kinase
ZNF	Zinc-finger proteins

1 Introduction

1.1 Cardiovascular Disease

Cardiovascular disease (CVD) describes a collection of disorders involving pathophysiological events within the heart and vasculature. Broadly categorized, CVD can be subdivided into four main categories: coronary heart disease, myocardial disease, cerebrovascular disease, and peripheral artery disease. In order, these conditions describe disease in the vasculature of the heart, the muscle of the heart, the brain and lastly the peripheral vasculature of the body, with each condition carrying risk factors to both patient morbidity and mortality which scale with disease severity. The majority of CVD events which lead to large impacts on disability-adjusted-life-years (DALYs) and even mortality involve acute injury, usually attributed to ischemic damage in the heart and brain. However, these events are often preceded by chronic pathophysiological adaptations. Though these mechanisms are known to develop in many tissue types, involving a variety of processes, hypertension, the chronic elevation of blood pressure caries the highest burden.

According to the World Health Organisation (WHO) in 2021, CVD is the primary cause of death globally, with disproportionately high impacts in low and middle-income countries. In 2019, 17.9 million deaths were attributed to CVD, with 15.2 million attributed to heart attack and stroke. Between 1990 and 2019, the absolute number of concurrent CVD cases nearly doubled, from 271million to 523million individuals, whilst DALYs also nearly doubled from 17.7million to 34.4million years (Roth *et al.*, 2020). In the US, up to 30% of the household median income is spent on healthcare, primarily in the management and treatment of CVD, with the condition being termed “the world’s most urgent healthcare problem”, both in terms of health and economics (Emanuel *et al*, 2017; Slavin *et al.*, 2021).

1.2 Primary hypertension and the Global Burden

Primary hypertension is defined as sustained elevation of both systolic (SBP) and diastolic blood pressure (DBP) above a healthy physiological baseline, distinct from secondary hypertension which is driven by underlying separate disease. Currently,

Stage I hypertension is accepted as readings of >130mmHg and >80mmHg and Stage II hypertension consisted of readings of ≥140mmHg SBP and ≥90mmHg DBP (versus 120mmHg and 80mmHg respectively in healthy individuals) (Unger *et al.*, 2020)(Whelton *et al.*, 2018). Increases in SBP and DBP values past the healthy range are known to correlate positively in a log-linear fashion with cardiovascular disease (CVD) incidence risk, with incremental increases of 20mmHg SBP or 10mmHg DBP associated with a doubling in risk of CVD death at ages 40-69 years (Lewington *et al.*, 2002). Globally, hypertension is the leading cause of death and is the number one modifiable CVD mortality risk factor and number two global all-cause mortality risk factor as of 2017 (secondary only to smoking) (Danaei *et al.*, 2009)(Forouzanfar *et al.*, 2017).

As of 2020, hypertension, whilst treatable if addressed effectively, is an expanding pandemic with prevalence in adults of lower and middle income countries (LICs, MICs) at 31.5% and high income countries (HICs) at 28.5% with these values increasing linearly per each decade from 1980 to present (Mills *et al*, 2020). Interestingly, hypertension is commonly a cryptic condition as the primary symptoms do not noticeably manifest, with a recent study suggesting that globally 46.5% of suffers are unaware of their condition. This further diminishes effective control which is currently rated as 19% controlled in HICs and 15.6% and 10.8% In MICs and LICs respectively (Chow and Gupta, 2019). Due to the extensive and ubiquitous nature of the hypertension pandemic, alongside its capacity to cause morbidity and mortality, hypertension has a large socioeconomic burden worldwide. Out-of-pocket expenditure per capita (PPP) ranges from $100 to $1000- constituting an estimated global burden of hundreds of billions of dollars a year (Wierzejska *et al.*, 2020). Though these studies acknowledge the caveats of obtaining such population level metrics from LMICs, hypertension is clearly one of the most prominent global health concerns of the 21st century to date.

1.2.1 Primary Hypertension Burden in the United Kingdom

Hypertension in the United Kingdom (UK) had around 30% prevalence as of 2013, with the British Heart Foundation estimating that as many as 5 million adults may be

undiagnosed individuals (Joffres *et al.*, 2013). The prevalence of hypertension in the UK, though extensive, does appear to be decreasing, with treatment efficacy increasing between 2003 and 2015 and concurrently average blood pressure values decreasing across this timescale (Zhou *et al.*, 2019). High blood pressure accounts for 12% of all NHS GP visits and is estimated to cost at least £2 billion per year. The prevalence of hypertension in the UK correlates with socioeconomic status, with the most deprived communities suffering both the highest proportion of patients and also the lowest access to treatment (Petersen and Benzeval, 2016).

Though estimating the impact of hypertension on morbidity and mortality in the UK is challenging, with limited evidence currently, a study in 2018 identified an odds-ratio of 2.8 for CVD related mortality in those suffering with hypertension versus healthy controls (Hammami *et al*, 2018), with a separate study indicating that up to 5000 stroke-deaths in the UK could be prevented by increasing population level control of blood pressure (*Health Survey for England 2017 Cardiovascular diseases*, 2018).

1.2.2 Risk Factors of Primary Hypertension and Cardiovascular Implications

Primary hypertension is driven by several risk factors, both modifiable (therefore potentially targetable) and immutable. Modifiable risk factors include body-mass-index (BMI), smoking, low-density lipoprotein serum values, alcohol intake, smoking status and diet (particularly sodium chloride intake). Contrastingly, a number of risk factors are fixed and include sex, age and genetics; this combination of factors, both environmental and genetic has led hypertension to be regarded as a *complex disease*. Diabetes also contributes to hypertension risk and may be categorized as both modifiable and immutable depending on the underlying disease mechanism (Unger *et al.*, 2020)(De Boer *et al.*, 2017). Together these covariates produce an increasing model of risk which positively correlates with these variates up to the age of 60, after which age correlates negatively with both SBP and DBP (Whelton *et al.*, 2018)(Pinto, 2007). The multifactorial aetiology of genetic architecture in hypertension is a particularly prominent aspect which will be discussed to greater detail in 1.4.4.

Risk Factors of Primary Hypertension		
Factor	**Type**	**Relationship**
Body Mass Index	Modifiable	Positive correlation
Is smoker	Modifiable	Increased
Dietary sodium chloride	Modifiable	Positive correlation
Alcohol intake	Modifiable	Positive correlation
Serum LDL	Modifiable	Positive correlation
Type II diabetes	Somewhat modifiable	Increased
Age	Immutable	Positive correlation (to Age 60)
Is male sex	Immutable	Increased
Genetics	Immutable	Cumulative

Table 1.2-1 Primary hypertension risk factors, delineated by modifiability and directionality of relationship.

Much of the morbidity and mortality caused by hypertension is driven by the process of hypertension mediated organ damage (HMOD) with organ specific pathophysiological processes developing over time. Firstly, the brain can be affected transiently due to potential ischemic events which are themselves increased in likelihood due to the long term increases in risk of thromboembolic type events in the circulatory system (Holst *et al*, 2010; Huang *et al*, 2016). HMOD in the brain is also known to manifest over longer, chronic time periods, with microbleeds and cerebral small vessel disease (CSVD) known to be increased in hypertensive patients over lifetimes. Furthermore, the increased risk of CSVD has been shown to increase linearly with the duration of hypertension in patients (Petrea *et al.*, 2020). Together these findings indicate prominent risk of both transient and chronic neurological damage due to hypertension.

Hypertension is also known to negatively affect both the heart and circulatory system. Firstly, the effects of heart HMOD manifest by driving both development of

hypertensive cardiomyopathy and heart-failure-with-preserved-ejection-fraction (HPEF), both structural disorders affecting the efficiency of the left ventricle (LV). Initially, LV remodeling is adaptive in response to increased peripheral pressure, however over time and without adequate control, this adaptation can become pathological ("transition to failure") leading to significant impact on mortality and disability adjusted life years (Drazner, 2011; Oh and Cho, 2020; Tackling and Borhade, 2021). With respect to the circulatory system, hypertension is a known risk factor in the development of atherosclerosis, with a direct positive correlation between SBP/ DBP and atheroma risk thought to be mediated by increased turbulent blood flow and inflammation (Rafieian-Kopaei *et al.*, 2014)(Prado *et al.*, 2008)(Kobayashi and Uesugi, 1995). Linking these two highly related but distinct conditions, coronary artery disease (CAD) is characterized by atherosclerosis of cardiac arteries progressing to myocardial infarction and is a known HMOD associated condition, with significant risk of mortality and morbidity with even short term reduction in SBP over months reducing CAD events by 16% (Cubrilo-Turek, 2003).

The kidneys are also a HMOD target, with hypertensive renal damage being a downstream risk even in individuals with moderately raised SBP and DBP due to the effect of raised pressure in the mechanically sensitive glomerular structure (Bidani and Griffin, 2004). Hypertensive damage to the kidneys and subsequent chronic kidney disease (CKD) is known to itself be a risk factor for CVD, creating a feedback loop of further pathophysiological processes in response to chronic increases in blood pressure (Griffn, 2017).

1.2.3 Regulation of Blood Pressure and Development of Primary Hypertension

A large component of blood pressure regulation and hypertension development can be attributed to the renin-angiotension-aldosterone system (RAAS), which is the primary regulator of Na+ retention and pressure natriuresis. The RAAS controls the pressure:volume homeostatic system in the kidneys where it mediates feedback between volume and natriuresis, with RAAS mechanisms known to be prominent in the development of hypertension (Yim and Yoo, 2008; Te Riet *et al.*, 2015; Oparil *et al.*, 2018). These physiological effects are mediated by decreases in perfusion

pressure and Na+ delivery stimulating renin, which in turn increases the pro-contractile and pro Na+ reabsorption effect of angiotensin II. In a healthy individual these mechanisms synergize to maintain blood pressure within a healthy physiological range but are known to become imbalanced in pro-hypertensive states. This leads to an abundance of perfusion pressure, combined with high Na+ load to the tubule, high Na+ reabsorption and in turn increased blood osmolarity and subsequent hypertension (Fountain and Lappin, 2018). Underpinning the strong link between Angiotensin II and HMOD are both the angiotensin-converting-enzyme (ACE) and the angiotensin-converting-enzyme II (ACE2). ACE catalyzes the formation of angiotensin II, driving disease, whilst ACE2 converts angiotensin II to the vasodilatory compounds angiotensin (1-7) with evidence indicating the pro-hypotensive effects of angiotensin (1-7) are as important as the pro-hypertensive effect of angiotensin-II in cardiovascular disease (Ferrarlo, 2011).

The endothelium of the vascular system is also a prominent contributor to blood pressure homeostasis through local and systemic effects on vessel contractile tone- primarily through the production of nitric oxide (NO). As a mechanical response to flow, endothelial cells are stimulated to produce NO which in turn causes vascular smooth muscle relaxation through initiation of the guanylate cyclase pathway (Spieker *et al*, 2006; Oparil *et al.*, 2018). Systemic reductions in NO production in both hypertensive and diabetic patients have been evidenced in human studies and contribute to disease development (Panza *et al.*, 1993; Ayub *et al.*, 2011). Alongside NO production, endothelial cells secrete several autocrine and paracrine vasoactive compounds including locally derived angiotensin-II and a second significant vasoconstrictor known as endothelin-1 (ET-1). ET-1 acts through activation of ET-A receptors in vascular smooth muscle cells, leading to a pro-contractile state. Significantly, targeting the ET-A is known to restore a normotensive state in both animal models and phase-2a human studies (Weber *et al.*, 2009; Lazich and Bakris, 2011).

Increasingly, the immune system is known to contribute toward the development and maintenance of hypertensive states in humans through effects on both the systemic vasculature and the kidneys (Oparil *et al.*, 2018). Through both innate and adaptive

mechanisms, cytokines, reactive-oxygen-species and macrophage derived NO act to increase vascular tone and promote arterial lumen thickening, most commonly through JAK-STAT, NF-κB, and Smad signalling pathways (Sprague and Khalil, 2009). Cytokines released by both vascular endothelium and circulating adaptive immune cells also promote responses to ET-1 and angiotensin-II, which in turn increase pro-inflammatory cytokine production (in particular IL-6), thus creating a positive feedback loop which further exacerbates the pro-contractile state observed in hypertension (Ruiz-Ortega *et al.*, 2002; Shinagawa *et al.*, 2017; Tanase *et al.*, 2019). Tumour necrosis factor alpha (TNF-α) is a prominent constituent of this family, known to exert a strong positive natriuretic response, possibly through the RAAS system (Sriramula *et al.*, 2008; Majid *et al*, 2015; Mehaffey and Majid, 2017). Contrastingly, an additional mechanism has been suggested, where silencing of renal TNF-α in mice was shown to reduce tolerance to salt sensitive hypertension in these animals, suggesting a more direct role through effects on sodium channels and transporters directly (Hao *et al*, 2018)(Figure 1.2-1).

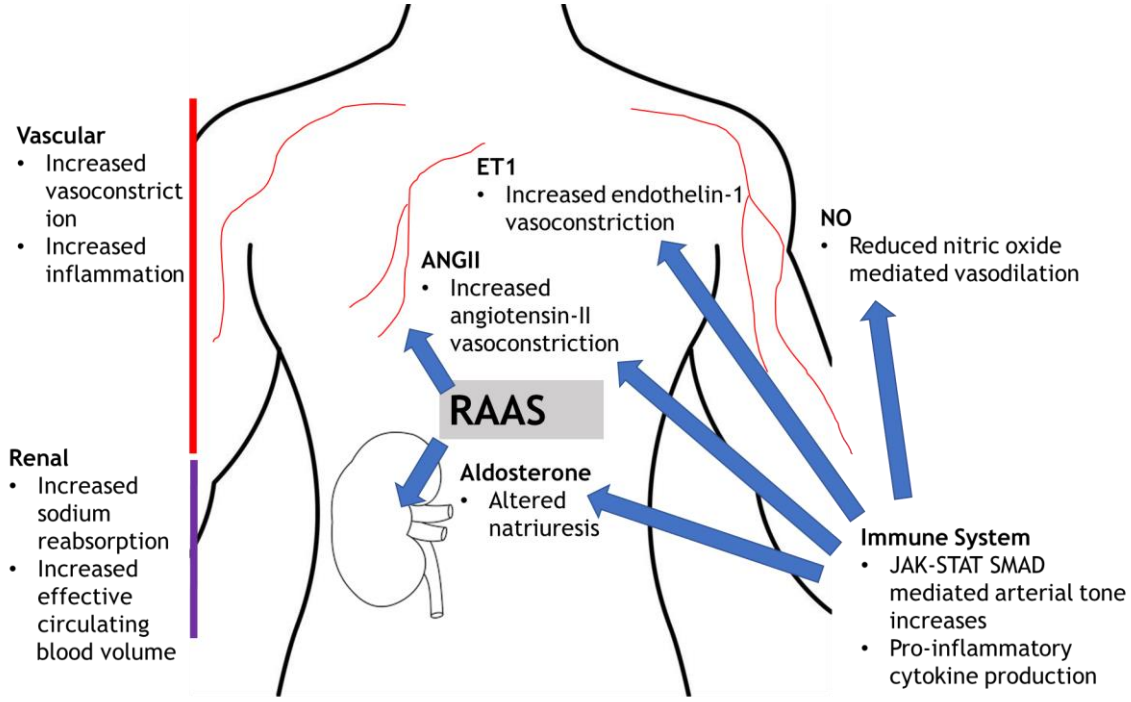


Figure 1.2-1. Simplified diagram of the primary blood pressure regulatory mechanisms in humans and the complementary effects of the immune system. Vascular mechanisms are highlighted by red (left), renal mechanisms are highlighted by purple (left). Contributions to mechanisms in both the vasculature and kidney are indicated by blue arrows, as are the effects of RAAS pathway metabolites on both the vasculature and kidneys.

1.2.4 Diuretics

Anti-hypertensive medications which target the renal system invariably attempt to attenuate natriuresis, though their mechanisms of action differ into three sub-categories. Firstly, thiazide diuretics aim to reduce the reabsorption of NaCl in the distal-convoluted-tubule (dCT), specifically achieving diuretic inhibition by inhibiting the Na^+ /Cl^- co-transporter (NCC). With the dCT constituting 7% of Na+ reabsorption, this results in reduced osmotic pressure back across the lumen and into the blood (Adrogué and Madias, 2007; Duarte and Cooper-Dehoff, 2010). Subsequent volume reductions result in stimulation of both the sympathetic nervous system and the hypotensive aspects of the RAAS, with studies emphasizing the contribution of Ang-(1-7) and Renin in both humans and animal models (Jessup *et al.*, 2008; Kramers *et*

al., 2020). Interestingly, though the reductive effects of thiazide diuretics are lower than their loop counterparts, resistance to rises in blood pressure as a response to titrating down the dosage appear to suggest there is an additional undiscovered mechanism of action which confers more chronic protection with seminal study suggesting this 'non-diuretic' effect was due to mechanisms in the systemic epithelia with total peripheral resistance falling (Duarte and Cooper-Dehoff, 2010)(Aleksandrow *et al*, 1959; FREIS *et al.*, 1960).

The second of three main renal antihypertensives are the loop diuretics, which target the sodium-potassium-chloride cotransporter (SLC12A1 (NKCC2)) of the tubular epithelial cells located in the thick-ascending-limb (TAL)- responsible for 25% of total sodium reabsorption (Jung *et al*, 2011). Specifically, loop diuretics are hypothesized to bind within the translocation cavity of this cotransporter (the same site as where the ions are shuttled), thus inhibiting function (Somasekharan *et al*, 2012). Much like thiazide diuretics, loop diuretics act on the lumen-side of the system, though they are distinguished both on the proteins they target and their location within the kidney (Ellison, 2019a). Loop diuretics, due to their mechanism of action targeting a transporter which associates with movement of not just sodium, but also chloride and potassium ions, are associated with cation disturbances in general, including hypokalemia, hypomagnesemia and hypocalcemia (Oh and Han, 2015).

The last family of renally acting antihypertensives are the potassium sparing diuretics, including mineralocorticoid receptor blockers (MRBS), which competitively inhibit the mineralocorticoid receptor (the target of aldosterone). Potassium sparing diuretics promote epithelial sodium channel degradation or inhibit reabsorption, which ultimately promotes natriuresis, primarily in the distal convoluted/connecting tubules (Sica, 2015). This family of drugs is particularly effective in resistant forms of primary hypertension, with spironolactone (the most prescribed MRB) being significantly more effective at treating resistant hypertension versus other antihypertensives. Interestingly, the magnitude of lowering of BP negatively correlated with plasma renin concentration in this analysis (Narayan and Webb, 2016; Yugar-Toledo *et al.*, 2017).

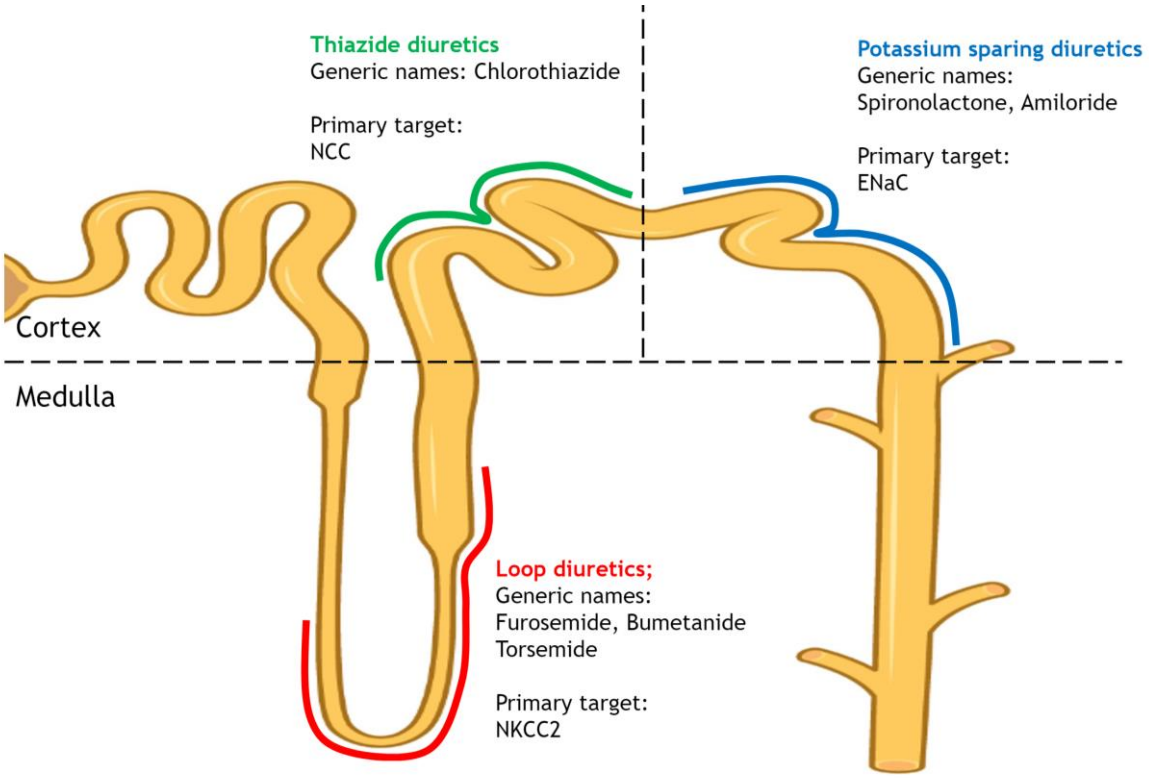


Figure 1.2-2 Single nephron visualization of diuretics.

Loop diuretics target mechanisms within the loop and ascending limb, particularly NKCC2 mediated natriuresis. Thiazide diuretics target NCC in the distal convoluted tubule, whilst potassium sparing diuretics target ENaC in the proximal convoluted tubule and distal collecting duct (shown in blue).

1.3 Precision Medicine

1.3.1 Precision Medicine

Precision medicine is potentially one of the most impactful healthcare paradigm shifts of the 21st century, where the fundamental aim is to use targeted treatment approaches to reflect the underlying differences in patients. In effect, using patient information to inform prescription increases efficacy and decreases side-effects and costs on a per-patient basis, delivering benefits to both the individual and the health

service. With an estimated 10% of all CVD healthcare being spent on drugs in the US, the impacts of reducing this are likely to be significant (Bode and Dong, 2018). Broadly speaking, precision medicine considers all of the underlying differences between patients: lifestyle, environment and deep-phenotyping (examining patients for biomarkers at the genomic, proteomic and metabolomic level). Specifically, deep-phenotyping (particularly via genomics) is the most promising of these fields because typically, clinician led changes to prescriptions are the easiest of the three to implement on a broad scale and, unlike lifestyle and environment, do not rely on patient-led action. Consideration of the genome as a method of informing treatment led to the emergence of pharmacogenomics as a research discipline, with 'pharmacogenomics' constituting a portion of 'precision medicine', though they are often used interchangeably. Essentially, pharmacogenomics aims to associate genetic variants with drug effects, both pharmacokinetically and pharmacodynamically. In future, large pharmacogenetic reference libraries may exist which could be used to inform per-patient drug choices and dosages, though such a widespread implementation would require both substantial infrastructure investment and ethics discussion.

1.3.2 Economics of Precision Medicine

The success of precision medicine development and implementation depends heavily on cost-effectiveness. Over the last decade, the primary economic barrier has been the cost of next-generation-sequencing- an essential aspect of pharmacogenomics. However, with the reduction in cost of whole-genome-sequencing from $22,000,000 in 2006 to around $1000 in 2020 these costs are beginning to be outweighed by benefits (Schwarze *et al.*, 2018, 2019). As early as 2014, studies began reporting on the cost-effectiveness of targeting variants in the treatment of both cardiomyopathy and colorectal cancer though they emphasized that at this point it was not cost effective to perform population level screening (CS *et al.*, 2015). In more recent years, a number of analyses (n=53) concluded that precision medicine was increasingly positively cost effective, although these analyses were significantly heterogenous in nature (Kasztura *et al.*, 2019). Possibly the most important finding from the large number of analytical studies attempting to characterize the economics

of precision medicine is that, whilst the impact is consistently positive, the methodologies of measuring these impacts require significant refinement in order to provide directly comparative results (Faulkner *et al.*, 2020).

1.3.3 Applications of Precision Medicine

Pharmacogenomic style precision medicine applications can be broadly characterized into two fields: precision oncology and non-cancer pharmacogenomics. In terms of current healthcare impacts, precision oncology has the largest effect on treatment. Precision oncology is the process by which tumours are sampled, and whole genome sequencing performed to identify tumour specific drug vulnerabilities, based on specific annotated variants within the sample. Precision oncology is a rapidly emerging treatment paradigm in HIC healthcare, though accessibility is more limited in poorer countries (Faulkner *et al.*, 2020). Using pre-determined genomic signals, the process optimizes the application of both small molecule inhibitors (such as those targeting kinases) and immune therapeutics by delivering them in a more targeted manner; with both of these drug classes having proven successes to date (Krzyszczyk *et al.*, 2018)(Bode and Dong, 2018). A growing reference library of specific tumour mutations now exists from which healthcare professionals are able to generate predictively effective drug combinations (Bode and Dong, 2018). However, notable challenges to generalized execution of precision oncology exist and even in world-leading countries such as the UK, disparity in social-status, age and race currently diminish the potential effectiveness of the discipline (Spratt, 2018).

Cardiovascular disease pharmacogenomics constitutes the majority of non-oncological precision medicine, with several streams of treatment currently in development, existing or in use. Variants on the Cytochrome P450 (CYP)2C19 (CYP2C19) gene are now known to contribute to adverse drug reactions to the anti-platelet drug clopidogrel by reducing enzymatic degradation and contribute to the risk of acute coronary syndrome in these patients (SA *et al.*, 2013). As a result, guidelines in both the UK and US were updated to reflect this risk, though CYP2C19 genotyping is not practiced as standard despite this genotyping having been shown to influence clopidogrel prescribing habits in small clinical trials (Tuteja *et al.*, 2020).

Similarly, the effect of cytochrome P450-2C9 (CYP2C9) and vitamin K-epoxide reductase complex (VKORC1) alleles on warfarin pharmacodynamics and kinetics have been evidenced, with variants on the former influencing degradation and sensitivity respectively (IY Gong, 2011)(JA *et al.*, 2011). However, in much the same way as clopidogrel, point-of-care genotyping, though clearly evidenced as effective, is not standard protocol in the UK currently (Jorgensen *et al.*, 2019).

Cardiovascular precision medicine also extends to lipid modulating drugs, with myopathy noted as a moderately common side effect. A locus of variants on the organic anion polypeptide transporter 1B1 (SLCO1B1) was noted to associate with an increased risk of myopathy in patients with lower doses of statins recommended for the risk allele, though again, even with strong evidence point-of-care genotyping is not standard (LB *et al.*, 2014; Tuteja and Rader, 2018). Long QT syndrome type-3 (LQT3) has been observed as driven by gain-of-function mutations in SCN5A- the alpha subunit of the sodium channel NaV1.5. In these patients it was noted that they would be likely to respond positively to Mexiletine therapy, specifically targeted to block this sodium current in cardiomyocytes (due to their underlying genotype). This targeted therapy was shown to reduce arrhythmic events from 22% to 3% in patients, requiring only a simple genotyping assay for pre-selection (A *et al.*, 2016). Lastly, loss of function variants on CYP2D6 are known to reduce the therapeutic index of ß-blockers- a mechanism particularly important to heart failure patients with direct association with patient survival (M *et al.*, 2011). A more recent study evidenced a greater than 6-fold increase in trough concentration, though as with other CVD precision medicine targets, broad implementation is not the case (Anstensrud *et al.*, 2020) (Figure 1.3-1).

Though a number of pharmacogenomic precision applications for cardiovascular diseases have been identified and examined over the last decade, ranging from risk-of-death interventions to adverse side effects, their utilization in the UK and worldwide is still very limited. This reflects inadequacies in genotyping facilities and infrastructure rather than the lack of evidence toward their benefits.

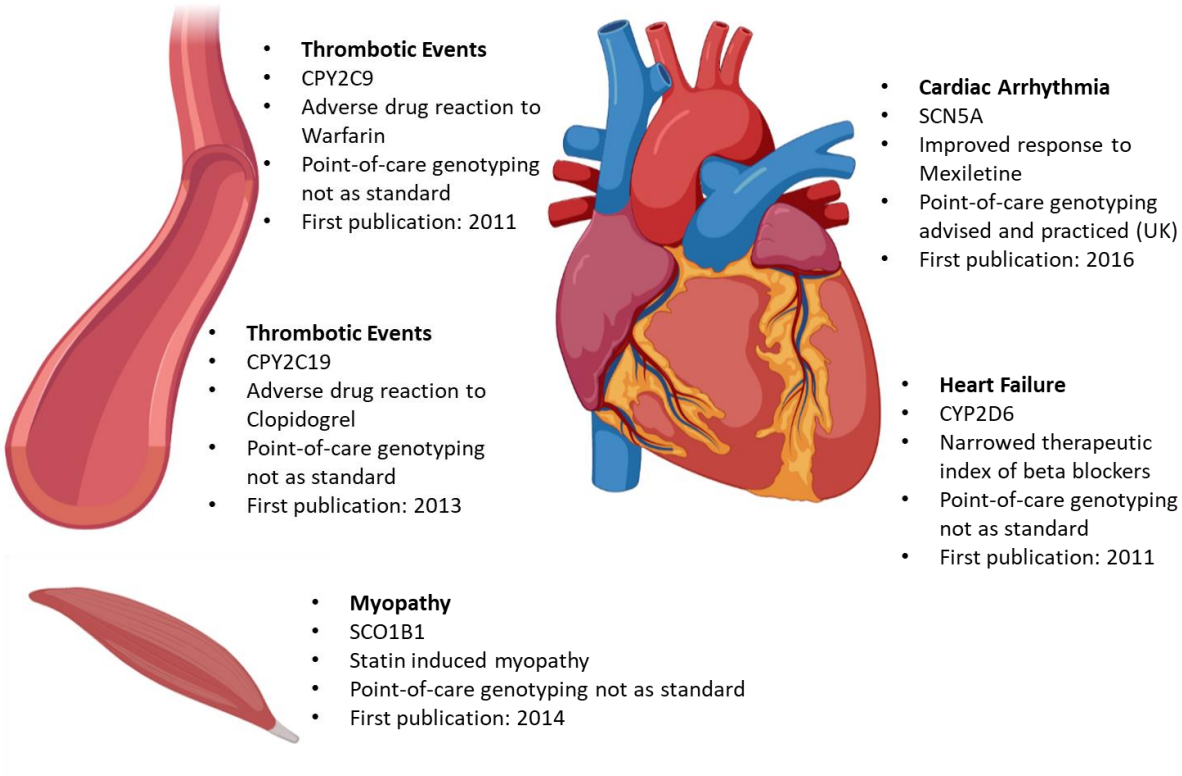


Figure 1.3-1 Applications of precision medicine in cardiovascular disease. Described elements per application are: precision medicine gene, genotype dependent reaction, health service action and initial publication date. Only cardiac arrythmia risk response to mexelitine derived from SCN5A alleles are commonly practiced precision medicine applications in public healthcare in the United Kingdom.

1.4 Genomics as a Research Paradigm

1.4.1 Genome Wide Association Studies

Genome wide association studies (GWAS) are high throughput population level studies resulting from the development of next-generation-sequencing (NGS) and its subsequent reduction in cost. GWAS largely facilitated the transition of genetics from size-limited Mendelian study to population level analysis. Either binomial in design (with or without a disease) or continuous (quantitative trait, such as blood pressure),

GWAS attempt to correlate a phenotype or disease with genetics. Specifically speaking, they attempt to associate variants on the genome with the target phenotype, such that the loci on which these variants exist can be used to identify novel risk markers, novel druggable targets and potentially novel precision medicine targets (Mattson and Liang, 2017; Ho *et al.*, 2019) . Early in their development, GWAS were predicted to radically improve medicine (JN, 2009; RJ *et al.*, 2010). However, in more recent years a necessary acknowledgement of the limitations has emerged, with the suggestion of the impact of GWAS on life expectancy and morbidity being 'minimal to none' (Joyner and Paneth, 2019). This lack of impact can be attributed to both inadequacy of interpretation and the comparatively weaker 'follow-up'- the translation of GWAS findings to the wet laboratory. With respect to the first caveat, this can largely be attributed to missing heritability, the aspect of genetics not properly controlled for in GWAS, particularly the impact of minor-allele-frequency (MAF) <10% variants on disease, alongside the majority (90%) of GWAS variants existing outside of protein coding regions (Manolio *et al.*, 2009)(Lonsdale *et al.*, 2013). In the context of blood pressure regulation, it was found using a meta-analysis study of 1.3 million individuals that the average effect size of rare variants was around 8 times greater than common variants at loci, though typically rare variants are more likely to be overlooked or not detected during analysis (Surendran *et al.*, 2020). However, in spite of the seemingly poor translation, GWAS continue to become increasingly large in number and scope (55,000 unique loci for nearly 5000 diseases and traits currently exist), suggesting that applications may simply be lagging behind the initial GWAS execution (Loos, 2020) (Figure 1.4-1).

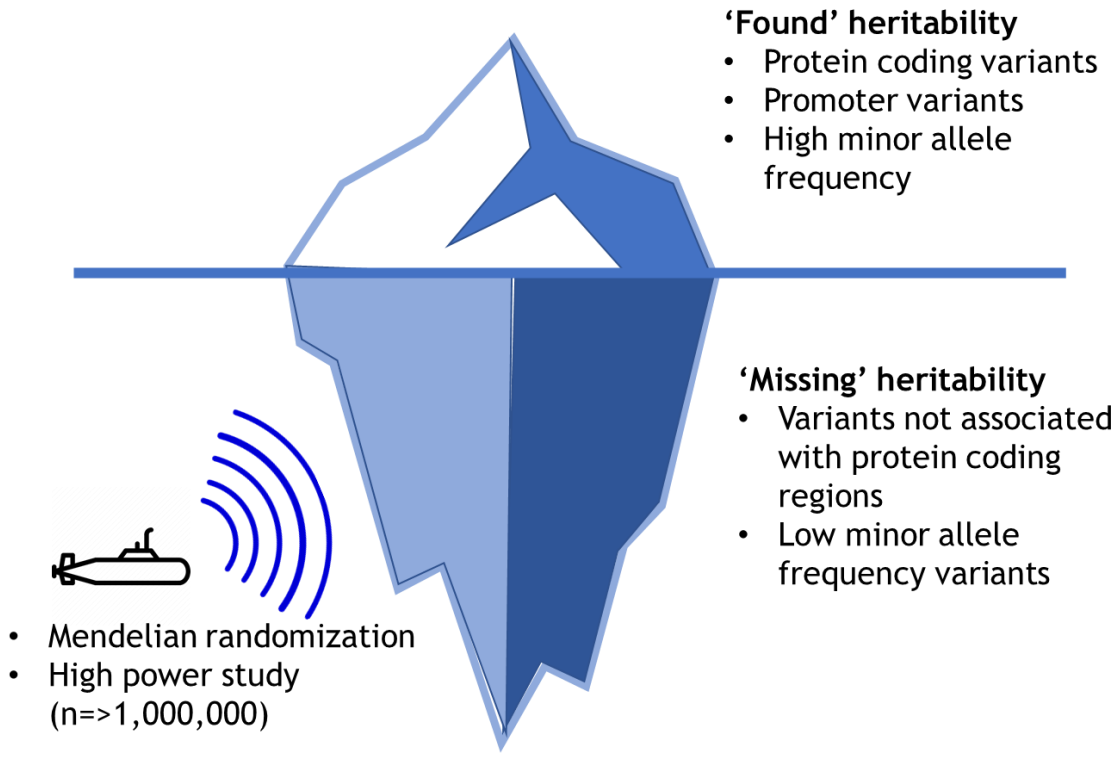


Figure 1.4-1 The iceberg of missing heritability in GWAS.
'Missing' heritability, whilst a major caveat, is considered to be increasingly addressable using large sample sizes to detect minor allele frequency variants, combined with the statistical methods of mendelian randomization to attribute causality.

Typically, GWAS are designed using a randomly sampled cohort, both for control and disease states- or one randomly sampled group for a continuous phenotype. These participants are then genotyped by SNP-array, an NGS method which directly characterizes around 350,000 variants, though this varies with the panel used. Up to an additional 3,000,000 variants are then imputed using known linkage-disequilibrium blocks (section 1.4.3) on the genome to provide around 3,500,000 variants (Loos, 2020)(Li *et al.*, 2015). Bayesian variable (correcting for familial data trends) multiple regression is then typically performed to provide associations between variants and traits, though statistical models are continually evolving in line with improved computational power and machine learning (Banerjee *et al.*, 2018). Genome wide

significance (GWS) p-values are thresholded far below the typical 95% test statistic due Bonferroni based correction of multiple testing, with the GWS p-value currently accepted at 5e10-8 (Roeder and Wasserman, 2009). In certain cases for quantitative traits, GWAS can be further enhanced by subsampling the extremes of the phenotype distribution, a method with strong potential to reduce the cost/power ratio by magnitudes of order (Li *et al.*, 2019).

1.4.2 The Genotype Tissue Expression Project

Whilst GWAS have provided an abundance of loci for both continuous traits and diseases, their translation to function remains challenging. Context within the cell, in particular the direct effect of GWAS variants on gene expression rapidly became the downstream high-throughput experiment after GWAS. Referred to as the Genotype Tissue Expression Project (GTEx), GTEx aim to determine tissue specific variation in gene expression in response to genotype (Lonsdale *et al.*, 2013). GTEx, unlike GWAS, require mRNA-based quantification of gene expression and are therefore generally centered on transcriptomic analysis (RNA-sequencing) and not genomic analysis- with alleles as the experimental variable as opposed to a specific phenotype. GTEx provide two-fold enhancement to GWAS. Firstly, they can provide numerical data on gene expression, which can allow effect sizes of GWAS variants for the same phenotype to be compared. Secondly, they can provide tissue specific information on GWAS loci- providing greater understanding of where these genes act, improving the translatability of identified targets (Lonsdale *et al.*, 2013; Basu *et al.*, 2021). GTEx catalogues provide a valuable resource for the field and continue to be improved and utilized in multiple different ways, including identifying inter-chromosomal effects and RNA splicing, two biological phenomenon not highlighted by GWAS alone (Aguet *et al.*, 2017; Ward and Gilad, 2017; Barbeira *et al.*, 2021). GTEx analyses have also been used in the context of blood pressure, where they were used to identify two subtypes of hypertension patients, based on their GTEx characteristics, a *Diffused group*, where dysregulation existed across a broad range of tissues and a *Localized group*, where dysregulation was limited to more specific tissues. Interestingly, comorbidities between these two subgroups were also

substantially different, indicating potentially separate streams of disease development and progression (Basu *et al.*, 2017).

1.4.3 Linkage Disequilibrium in GWAS

Linkage describes the non-random association of variants at or between specific loci on the genome, counterbalanced by the actions of genetic drift and recombination (Qu *et al.*, 2020). Linkage disequilibrium (LD) as a concept describes the proportion of correlation between alleles, with '1' representing perfect squared correlation and '0' representing no correlation (Qanbari, 2020). These 'haplotype blocks', regions of perfect or near-perfect linkage disequilibrium, are themselves powerful markers of both population growth and evolution (Wall and Pritchard, 2003). However, they also provide critical underpin to GWAS. Genome wide association studies typically publish one variant at a loci (attributed to the nearest plausible gene or other genomic marker), normally referred to as the 'index' or 'discovery' variant, effectively the genetic marker at this locus with the most significant p-value. However, in both concept and practice, GWAS do not discover a variant but rather they discover a locus (Joiret *et al.*, 2019). Whilst the discovery variant is used to tag the locus, that does not mean the casual variant is the discovery variant, this can only be determined with a degree of accuracy through 'fine mapping'. Fine mapping is a critical analysis step in GWAS which utilizes underlying information on haplotype blocks to determine the size of this locus, therefore providing an array of potential causal variants. Fine mapping, although conceptually simple, is also strongly influenced by localized, non-linear LD patterns, meaning that those variants most proximal to the discovery SNP are not always the most likely causal candidates (Schaid *et al* , 2018). This was evidenced by the APOE-4 locus detected in Alzheimers targeted genotyping having a non-linear LD at the target site, leading to potential causal variants being distributed across a relatively wide region around the discovery SNP (Martin *et al.*, 2000). The necessity for the inclusion of fine-mapping in GWAS analysis may help to further explain the poor translation of GWAS to clinical effect (Joyner and Paneth, 2019).

1.4.4 Hypertension GWAS

Both GWAS for blood pressure as a quantitative trait and for hypertension as a disease were phenotypes which were studied early in the progression of GWAS due to their population penetrance and genetic contribution (between 30-50% of blood pressure variance is thought to be attributed to genetics) (Ehret and Caulfield, 2013; RS *et al*, 2017; Wang and Wang, 2018). Data mining of all currently GWAS-catalogue archived studies, with respect to hypertension, not blood pressure, found to date, a total of 46 studies have been catalogued by the central system. The first GWAS was published in 2007 with a non-significant linear trend in publication frequency between 2007 and 2021 of an average of four publications per year. On average, these publications returned 17 GWS variants per study with an absolute range of 1:161 and a 25:75% interquartile range (IQR) of 2:16. This indicates a strong skewness in the right-tail of the distribution of 3.31AU, driven by a strong increase in the number of variants discovered in more recent years due to improvements in experimental design and sample size. With respect to sample size, a high degree of variability exists. The 46 catalogued hypertension GWAS enrolled between 112: 382147 participants, with a 25:75% IQR of 821:17022. This again indicates a strong degree of positive skewness in the dataset with a skewness value of 3.07AU- a feature of more recent studies enrolling higher number of participants.

When data were mined from GWAS-catalogue specifically and filtered to remove detection of identical variants between studies total of 696 unique variants have been annotated by GWAS as associating with hypertension. Of these, meta-analysis of the most significant variants shows that their top WGS finding is most likely to exist in intronic regions (45.9%), with the second most populated cohort existing in intergenic regions (18.7%). Following this, an even smaller proportion of variants were detected in regulatory (cis-promoter) regions (4.5%) with the remaining variants existing in both 5’ and 3’ UTRs and splice regions alongside a small proportion causing missense mutations. The number of associated loci and the heterogeny of their location in the human genome further emphasizes the complex polygenic architecture of hypertension with the disease contribution likely to be driven by small additions from each loci. It is theorized that a large majority of hypertension genetic contribution is

missing as the quantitative additions of discovered loci do not effectively constitute the known phenotypic variance of blood pressure. As of 2019, only 3.5% of the genetic contribution to blood pressure was thought to have been discovered (Wang and Wang, 2019). Much like the majority of GWAS, hypertension GWAS are thought to be lacking the crucial “rare variant” contribution, due the GWAS study design often not capturing these variants within typical analysis (A *et al.*, 2018). As described previously, low frequency variants are now considered to have large effect sizes in blood pressure regulation (Surendran *et al.*, 2020). However, because of their inherent low minor-allele-frequency (MAF), rare variants, though they may explain much of the population level variance, do not present the greatest potential translatability. This falls instead to hypertension GWS loci which have been identified as existing in the most physiologically plausible regions where mechanism to pathophysiology can be effectively resolved.

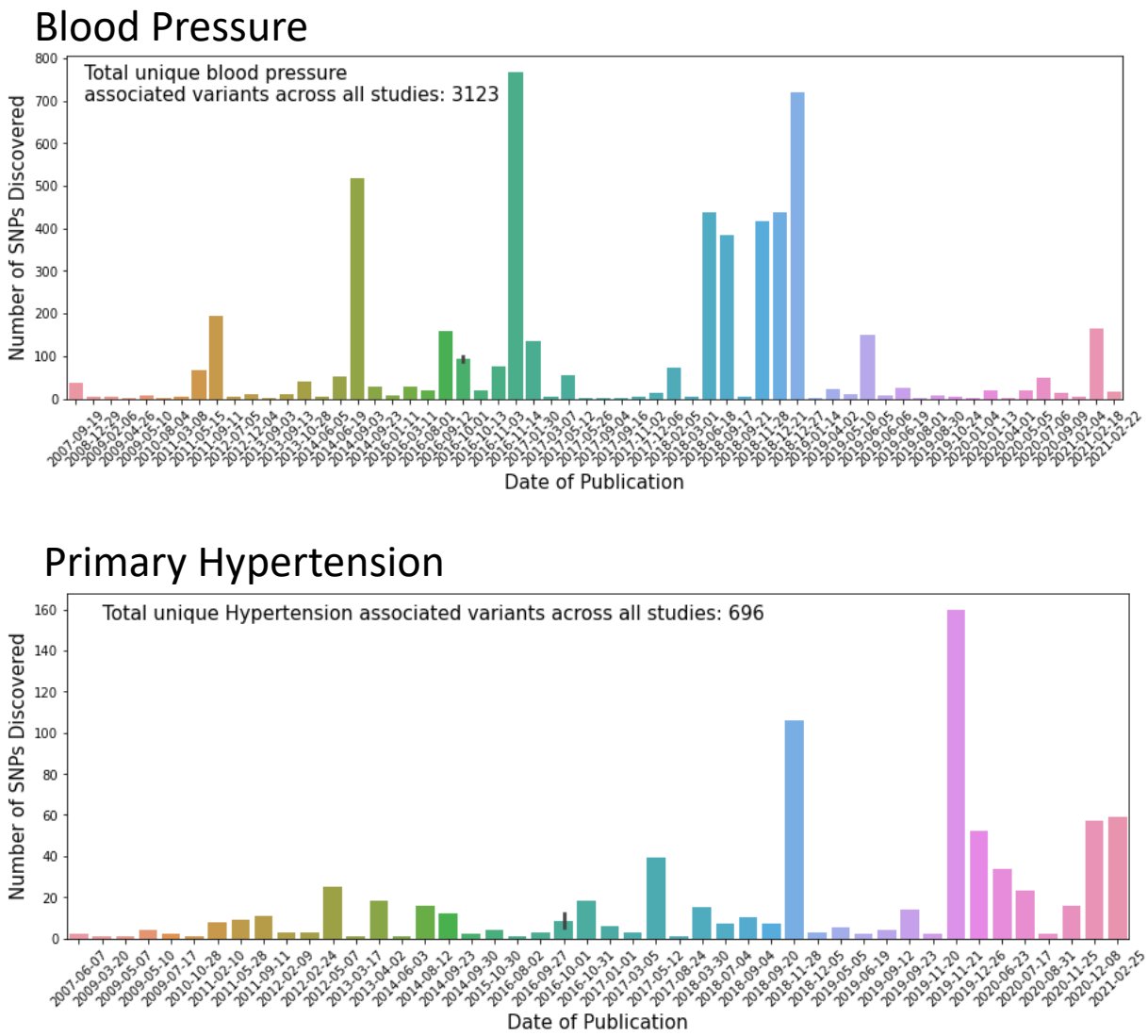


Figure 1.4-2 GWAS Catalogue described human variants for blood pressure (top) and primary hypertension (bottom) by year.

Data was filtered to prevent multiple reporting of the same variant by different publications. Exact date of publication is shown (x-axis). Original data were text mined in their original format as catalogued by GWAS Catalogue (*NHGRI-EBI Catalog of human genome-wide association studies, 2021*)

1.4.5 Eukaryotic Gene Expression

Broadly speaking, gene expression in eukaryotic cells is regulated and maintained by transcription factors (TFs). Transcription factors are proteins which interact by non-covalent mechanisms with genomic DNA, facilitating the formation of single-stranded messenger RNA (mRNA). Transcription factors typically associate with the cis-promoter of their target gene, generally 500-1000bp upstream of the TSS, where they usually 'dock' in a complex nature consisting of several proteins maintained in an active complex by non-covalent mechanisms. Mammalian genomes typically encode for around 1500 transcription factors, with each one recognizing a 'perfect' oligonucleotide sequence (response element (RE)) but also possessing binding affinity for non-perfect RE's (Tan *et al*, 2008; Todeschini, Georges and Veitia, 2014; Zhou *et al.*, 2017). When they perform an activatory function, transcription factors may be referred to as 'activators', conversely if they function to diminish gene expression, they may be referred to as 'repressors. The robustness of the binding match between the TF and the RE element, in terms of the topology of the interaction between the transcriptional complex and the associated DNA, is known to play a key role in the resulting gene expression (MacNeil and Walhout, 2011).

Whilst transcription can be initiated by RE interactions, eukaryotic promoters can also interact with secondary nuclear elements which repress transcription. Typically referred to as 'silencers', these elements correlated negatively with gene expression when activated. Silencer molecules are generally small-molecule oligonucleotides (micro RNAs (miRNAs) and antigene RNAs (agRNAs)), though recruitment of proteins to RE loci can also occur (Kolovos *et al.*, 2012; Leopold, Stirpe and Schalch, 2019). In turn, recruitment of silencing complexes to RE can result in chromatin remodelling effects, leading to more sustained effects on gene expression than more temporary small-molecule RNA interactions (E *et al.*, 2006). DNA methylation, a covalent modification to genomic DNA, can also be induced by gene silencing RE interactions, with DNA methylation study becoming an increasingly well recognised element to epigenetics (E *et al.*, 2006).

Both positive and negative regulation of gene expression can be affected by the magnitude of the non-covalent interaction between transcriptional activators, repressors and their target genomic loci. Thus, differential magnitude of electrostatic interactions at RE which contain variants explains a degree of differential gene expression in the population. With the advent of high-throughput genomics and transcriptomics, particularly over the last few years, the heterogeneous nature of transcription factor binding affinity between alleles has been widely suggested as being highly functionally relevant (Cusanovich *et al.*, 2014; Mitchelmore *et al.*, 2020; Abramov *et al.*, 2021). As recently as 2021, it was suggested by novel analysis of CHIp-Seq data, that >270,000 variants on the human genome accounted for the differential binding affinity of several hundred different transcription factors across several hundred different cell types (Abramov *et al.*, 2021). However, specific functional linking between GWAS and gene expression in the literature remains rare, possibly commenting on the disconnect between bioinformatics and functional physiological data.

1.5 The UMOD Gene and Hypertension

1.5.1 Padmanabhan et al Hypertension GWAS

One of the earliest and most impactful hypertension GWAS was conducted at the University of Glasgow and discovered a locus on the cis-promoter of the human UMOD gene using an extreme case control design. The GWAS highlighted the variant rs13333226 as associating with hypertension, in particular the minor G allele associated with a decreased risk of 13% (OR [95%CI]: 0.87 [0.84–0.91]) for each copy, 0.49mmHg lower SBP per allele and also associated with a 7.7% reduction for adjusted cardiovascular disease risk per allele (Padmanabhan *et al.*, 2010). The study initially enrolled 1621 cases and 1699 normotensive controls but this was expanded to 19845 cases versus 16541 controls upon follow-up. Both aspects of the study sourced their participants from Scandinavian populations. The discovery arm enrolled participants from the Nordic Diltiazem study (NORDIL) and the Malmo Diet and Cancer study- with both of these studies reflecting the inherent proportional representation of racial groups within Scandinavia, therefore featuring primarily White-European participants (BERGLUND *et al.*, 1993; Hedner, 2009). The validation study enrolled from the

MONItoring trends and determinants of CArdiovascular diseases (MONICA)/Pressioni Arteriose Monitorate E Loro Associazioni (PAMELA) and also from the aforementioned Malmo diet and cancer study (Keil, 2005; R *et al.*, 2005). The Padmanabhan study employed an extreme case control design, where they sampled the upper 2% and lower 9.2% of the distribution of blood pressure in the population, increasing their statistical power significantly and resulting in the identification of the GWS variant rs13333226 on the UMOD gene. Additional study, published in 2018 also further reflected and corroborated the initial finding, associating rs13333226 with blood pressure and hypertension (TJ *et al.*, 2017).

Rs1333226 is a A/G single-nucleotide-polymorphism which exists 1617 base pairs upstream of the canonical UMOD transcriptional start site in humans. In the Asian haplogroup the prevalence of the reference A-genotype is 93.4%, though this varies with haplogroup, decreasing to 62.5% in the African American subgroup. In the primary enrollment population, White-European, the prevalence of the reference A-genotype is 82.0%. Each copy of the minor G-allele in the study cohort was subsequently associated in further experiments with reduced urinary UMOD excretion (AA:53, AG:40, GG:17mg/24hr)(n=110). Significantly, it was also found that the minor G-allele was associated with decreased fractional excretion of sodium (AA:0.92, AG:0.95: GG:0.73 %) (n=256). These data together suggest that the allele at rs13333226 directly associates with the activity of the UMOD promoter. In turn the initial findings appeared to indicate that differences in UMOD expression between genotype drive differences in sodium excretion which, when further extrapolated, would suggest UMOD levels associate with blood pressure and hypertension through sodium homeostasis. This hypothesis, given the highly specific expression profile of the UMOD gene almost exclusively to the thick ascending limb cells of the mammalian kidney, strongly implicated UMOD with blood pressure regulation by this mechanism.

1.5.2 The Thick Ascending Limb

The thick ascending limb of the mammalian kidney performs a central role in renal physiology, particularly regarding ion homeostasis and urinary protein composition. Composed of water impermeable eponymous thick-ascending-limb (TAL) cells, these

cells form a polarized apical-basolateral complex between the apical pole facing lumen of the medulla containing pre-urine and the basolateral pole facing the blood (Mount, 2014). Arguably the most active mechanism within these cells is the apical transport of Na^+, K^+, and Cl^- ions, which plays a crucial role in sodium diuresis and is primarily mediated by NKCC2, the sodium potassium chloride co-transporter responsible for counter-gradient reabsorption of these ions from the lumen. The role of NKCC2 in TAL cells is responsible for 25-30% of all sodium reabsorption with homozygous protein coding mutations in NKCC2 known to drive Bartter syndrome type-1, characterized by pathologically low blood pressure and salt-wasting (GR, PS and PA, 2011). Constituting a smaller fraction of Na+ reabsorption in the lumen is the sodium/proton exchanger isoform 3 (NHE3), with tubule specific NHE3 *crelox* knockout mice displaying increased Na^+ wasting versus control and a 20% reduction in SBP (Fenton *et al.*, 2017). Additionally, adaptive mechanisms are known to exist in response to abrogation of the action of NHE3 with knockout mice shown to upregulate expression the γ-subunit of the epithelial sodium channel (ENaC), a passive sodium reabsorption channel in the proximal convoluted tubule (Brooks *et al.*, 2001).

Though Na+ homeostasis (particularly though NKCC2) appears to be the central regulatory role within TAL cells, the function of NKCC2 has been shown to be highly dependent on the activity of K^+ channels, with abrogation of this ion current shown to result in a significant decrease in the reabsorption of Na^+ and Cl^- ions. Removal of K^+ ions from the lumen displayed a decrease in sodium current of 215 to 133microA whilst additional blockade of K^+ conductance displayed the same effect (R and E, 1981). The movement of K^+ ions, unlike Na^+ or Cl^- ions is self-sustained due to the continual cycling of K^+ into the TAL and back out into the lumen (Mount, 2014). Potassium is cycled into the lumen by the renal outer medullary potassium channel (ROMK) and Maxi-K, with approximately 80% of the current being facilitated by the former (R and E, 1981). ROMK has such a central role to the activity of NKCC2 that ROMK loss-of-function mutations are also known to cause Bartters syndrome type-1 (DB *et al.*, 1996).

Additionally, TAL cells are responsible for approximately 55% of total magnesium reabsorption (GA, 1989). Magnesium reabsorption is not controlled or dominated by a specific channel or transporter but rather is believed to be mediated exclusively by 'paracellular shunt' mechanisms wherein passive reabsorption occurs across a highly positive electrochemical gradient at tight junctions between the cells through Claudin-16/19 heterodimers (Curry and Yu, 2018). Much like Mg^{2+} ions, Ca^{2+} ions are also reabsorbed via the paracellular pathway, with roughly 20% being by this manner. Significantly, claudin knockdown mice exhibit hypermagnesuria and hypercalciuria, with hypomagnesemia (N *et al.*, 2008). The calcium sensing receptor (CASR) is thought to play a role in calcium dependent paracellular transport regulation via its control over Claudin expression (H *et al.*, 2013). Whilst the majority of renal calcium reabsorption is paracellular, there is a lesser proportion attributed to the epithelial Ca^{2+} channels TRPV5 and TRPV6 which can be rapidly recycled between the sub-surface and apical cell membrane, allowing a degree of homeostatic regulation- an adaptation to the need for extracelluar Ca^{2+} to be tightly controlled in the body (De Groot *et al*, 2008).

1.5.3 The UMOD Gene

Sourced from Ensembl(2021) and UCSC(2021) against GRCh38, the UMOD gene encodes for the uromodulin protein in humans and is located on chromosome-16 at 20,344,373-20,367,623 in p12.3, consisting of 23,250 bases from the transcriptional start site (TSS) to the tail of the 3'UTR. The gene is encoded on the reverse strand and is preceded 5' by PDILT and followed 3' by GP2. The canonical reference promoter extends roughly -3000bases from 20,352,710. Within this canonical cis-promoter region there are 48 variants with a MAF >3% and 6 repeating elements (Repeat Masker 2021). The promoter is encompassed by a region of LD, which in the European population begins to extend from 20,353,000 to 20,358,500 (NIH LDAssoc 2021). Tight tissue specific control of promoter activity is apparent in mammals, an observation which has been examined in mice, where it was shown by transgenic design with human and bovine UMOD that the 3kb region upstream of TSS in mice appears to contain the necessary *cis* elements which govern this specificity (Zhu *et al.*, 2002). An additional transgenic study further evidenced this using a 3.9kb mouse

promoter sequence in conjunction with bovine UMOD (Kim *et al.*, 2003). UMOD is expressed almost exclusively in the thick-ascending-limb of mammals (90%), with 10% of total expression in the distal-convoluted tubule. The cis-promoter is highly conserved across humans, rats, mice and cattle (Schaeffer *et al* 2021).

Up to 12 transcripts have been identified and validated through a combination of Sanger sequencing and NGS. These may be broadly categorized into three distinct groups. Firstly, the 'full length' UMOD transcripts, of which there are three, consist of 11 exons. Two of these transcripts have their TSS within 30bases of each other, with the reference transcript TSS at 20,352,710. The third full length transcript has a TSS several thousand bases upstream at 20,356,301, potentially placing this transcript under the control of non-canonical regulation. Of these eleven exons, ten are <400 bases, whilst exon-3 is 801 bases long, constituting the 'long' exon of the transcript. The second category encompasses 'short transcripts', of which there are eight. These transcripts all contain the first two full length exons, but their defining feature is the foreshortening of exon three to varying degrees and the absence of further exons. The last category contains the one 'non-coding' annotated UMOD transcript which encompasses untranslated versions of exons 7-11. As a non-coding transcript, the biological function of this transcript remains unclear.

Analysis of the nascent pre-mRNA by miRabel- a prediction tool which uses a composite of novel methods and historic binding prediction algorithms- predicts the interaction of 176 human miRNAs with the canonical UMOD transcript, three of which are predicted to interact with the 5'UTR; 4419b, 4708-5p and 18a-3p. The majority of the rest of the signals interact with the 3' UTR, with the most enriched (miRabel score < 0.1) signals being for 16-5p, 548c-3p, 214-3p, 193a,3p, 125b-2-3p, 330-3p, 103a-3p, 107 and 195-5p in descending order of significance (Quillet *et al.*, 2020). Additionally, a strong enrichment for the transcription factor HNF1ß is apparent within the cis-promoter of human UMOD, with two putative interaction sites present. Furthermore, in kidney specific HNF1ß$^{-/-}$ mice, UMOD levels are decreased by >90% (G. L *et al.*, 2004; Schaeffer *et al.,* 2021).

1.5.4 Association Studies, GTEx and the UMOD Promoter

Variants on the UMOD gene have been associated with a number of phenotypes in humans, across variants extending from several thousand base pairs upstream of the canonical TSS to variants within a few hundred base pairs of this site. Particularly well categorized is rs12917707, the T-allele of which has been associated with decreased estimated glomerular filtration rate (eGFR), decreased renal function, increased risk of chronic kidney disease and significantly, increased urinary UMOD levels, versus the C-allele (A *et al.*, 2009, 2010; C *et al.*, 2016, 2016). Around 1000bp upstream (toward the TSS) of rs12917707, rs13329952 has been categorized in a similar manner to rs12917707, with the T-allele associated with increased eGFR versus C (C *et al.*, 2016). An additional finding at rs13329952 was the association with patient non-compliance for calcium channel blocker use, suggesting a potential precision medicine role for the UMOD promoter here (Y *et al.*, 2019). Around 600bp upstream of this, is rs13333226 which, alongside its previously discussed association with blood pressure, has been associated with type-2 diabetes risk (NR *et al.*, 2018). Around 350bp upstream of rs13333226 is rs4997081, the C-allele of which has been associated with increase urate levels in humans (A *et al.*, 2019). Additionally, a further 300bp upstream of rs4997081, rs4293393 was associated with a risk of chronic kidney disease and serum creatinine- indicating an effect on renal function (DF *et al.*, 2010). In total, seven variants on the *cis* human UMOD promoter have been associated with disease and physiological functions, with the majority of traits (14/38) attributed to rs12917707. Interestingly, further UMOD variants upstream of the canonical TSS have been associated with type-1 diabetes, though no functional pathways have been suggested by their parent publications (Figure 1.5-1).

Genotype tissue expression analysis of the human UMOD gene indicate >40 potential variants which associate with functional expression, including rs13333226 and rs12917707. However, the majority (>70%) of these highlighted variants exist in intronic regions. Research has shown the potential auto-regulatory effects of intronic variants on gene expression. Interestingly, one paper has shown that even with the deletion of the cis-promoter, introns can still stimulate ‘strong’ expression of their parent gene (Rigau *et al.*, 2019; Rose, 2019). Whilst available GTEx catalogues on the

human renal transcriptome do corroborate the GWAS findings at rs13333226 and rs12917707, the majority of the identified variants provide cryptic indications as to the underlying architecture of the variance in UMOD expression in humans (Gadegbeku *et al.*, 2013) (Figure 1.5-2).

The issue of translatability between high through-put omics experiments and functional data can again be exemplified by data mining the Encyclopedia of DNA Elements Database, which curates all submitted human CHIp-Seq datasets. For the region of the human UMOD promoter, encompassing 20,356,000:20,330,000 (containing rs1297707 and rs13333226), 1606 transcription factors and repressors have been typed *in silico* as associating along this locus, annotated in over 50 different human cell types. Of these, 710 were directly annotated as transcription factors, whilst the remaining 895 proteins were designated as possibly participating in histone modifications at the locus (ENCODEDB 2021). Thus, the signal to noise ratio of such datasets makes functional predictions as to transcription factor binding at variants extremely difficult, lending weight to the necessity that such questions must be addressed by laboratory based experiments.

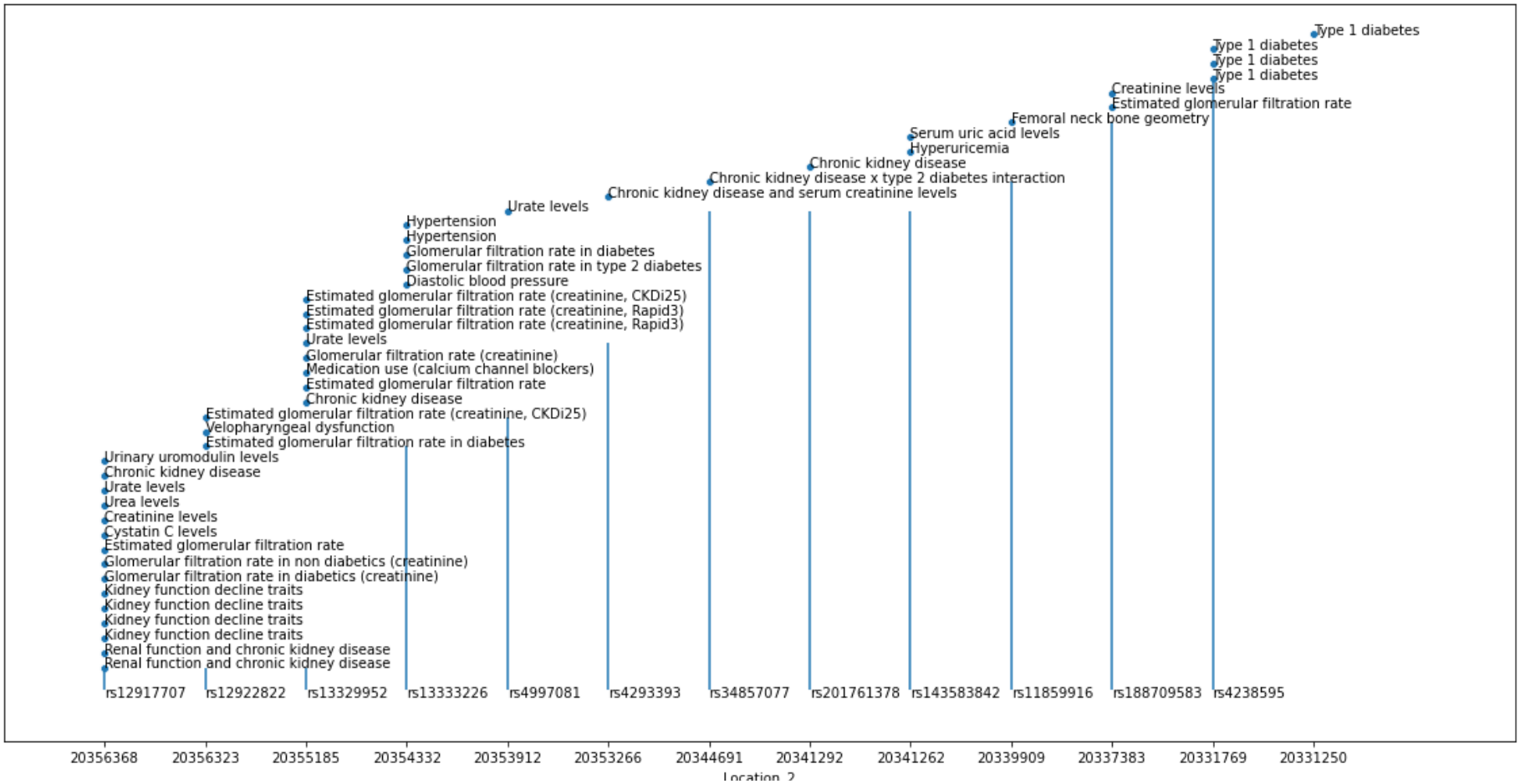


Figure 1.5-1 UMOD human variants and their associated traits (GRCh38), visualized by mined data from GWAS Catalogue (2021).

Coordinates are shown left (5') to right (3'), the UMOD TSS is upstream (right) of rs34857077. X-axis data are shown not to scale on coordinates but rather are displayed as equidistant for visualization purposes.

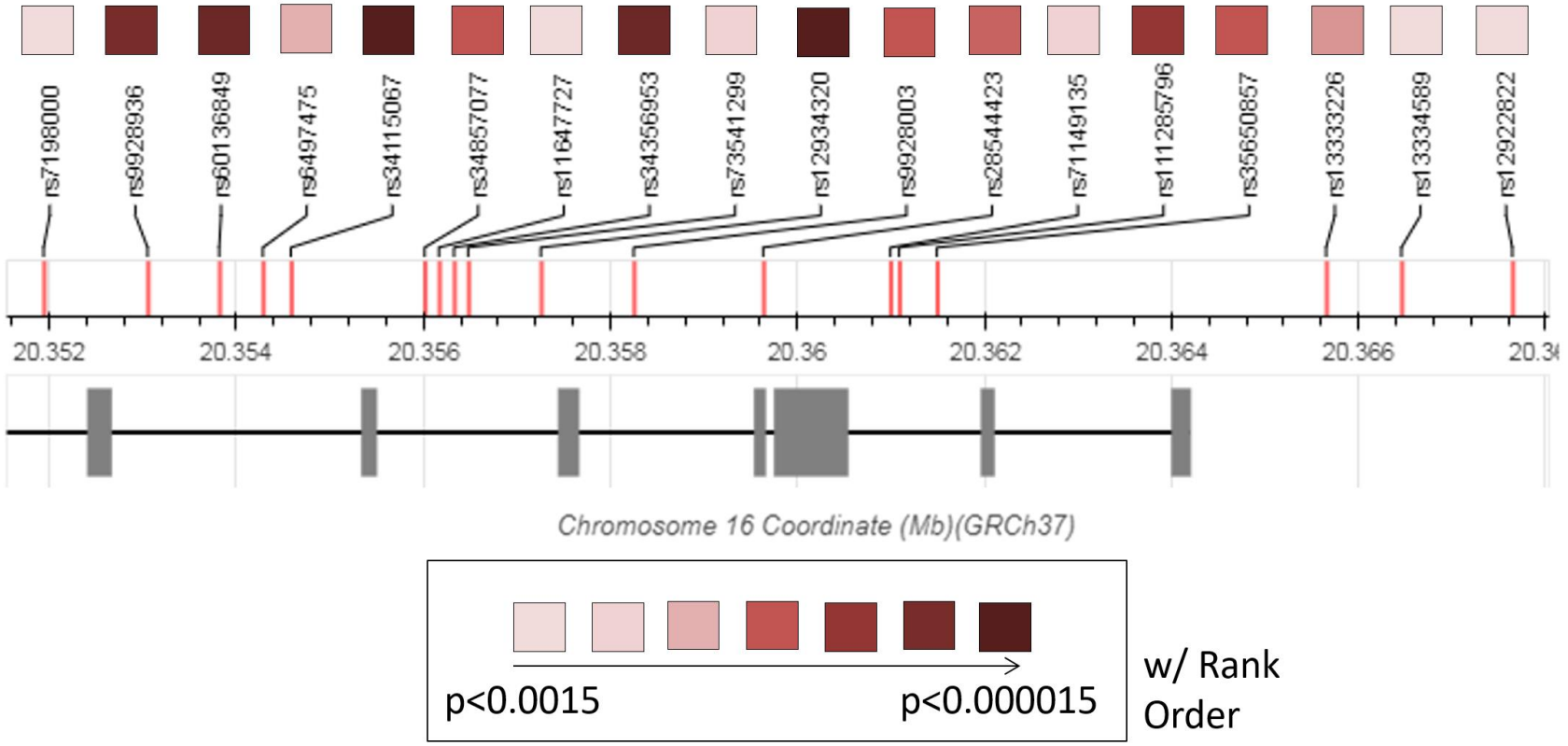


Figure 1.5-2 GTEx Variants which drive differential expression of UMOD (GRCh37).

GTEx visualisations are 5' (left) to 3' (right), the UMOD promoter is on the right. FDR (Bonferroni) p-values are colour coded. Of the top 18 variants on the UMOD gene which are associated with differential expression, 15 are located intronically, whilst 3 are in the UMOD *cis* promoter.

1.5.5 The UMOD Protein

The canonical UMOD protein is composed of 640 amino acids (AA). From the N-terminus there are several annotated structural motifs, beginning with four epidermal-growth-factor like domains (EGF) ending at AA148. Two of these EGF domains are known to bind calcium, leading some to regard the uromodulin protein as a medullary calcium sensor, providing important protection against calcium crystallization in filtrate (Johnson *et al.*, 2017a; R *et al.*, 2020). Leading from these EGF domains, toward the C-terminus, is a cysteine-rich domain which terminates into a zona pellucida and finally a C-terminal phosphatidylinositol (GPI) anchoring site at AA614 (Wolf, Zhang and Nie, 2019). Extensive glycosylation at up to eight potential N-linked sites is thought to account for up to 30% of the molecular weight of the UMOD protein, which is between 80-105kDaltons (kDa) in humans (F, N and D, 2003).

The UMOD protein is shuttled to the cell surface in a manner prioritising the apical surface (versus the basolateral) due to inherent epithelial polarity, underpinning the high degree of bias toward urinary uromodulin versus the much lower levels detected in serum (Scherberich *et al.*, 2018). Highlighted by immunogold labelling this is facilitated by cytoplasmic vesicles via the Golgi apparatus (Scherberich *et al.*, 2018; Boder *et al.*, 2021). From endoplasmic reticulum to cell surface, the UMOD protein is first folded, glycosylated and GPI anchored before it undergoes N-glycan maturation prior to entering a secretory stage at the cell surface (Wolf, Zhang and Nie, 2019). Correct protein folding is dependent on 48 well conserved cysteine residues within the protein (Schaeffer *et al*, 2021).

Unique to the apical surface, UMOD is cleaved by the activity of serine protease hepsin at AA589 on the C-terminus aspect of the zona pellucida with this activity shown to affect salt sensitive hypertension induction in mice, where hepsin deficient mice showed muted blood pressure responses to mice (Olinger *et al.*, 2019a). Upon cleavage, the apical UMOD secreted protein self-polymerises, whilst basolateral UMOD does not (C, A and WR, 1985; Olinger *et al.*, 2019a; Wolf *et al*, 2019; Boder, *et al.*, 2021). Although not specifically known, the homo-polymerisation site of UMOD is thought to be within an interdomain linker connecting the ZP-N and ZP-C domains

(M *et al.*, 2016). An internal motif within the zona pellucida, referred to as the internal hydrophobic patch (IHP), prevents internal polymerization, whilst the external hydrophobic patch (EHP) is not retained within the mature protein after cleavage, resulting in the observed extracellular polymerization (C *et al.*, 2009).

1.5.6 Physiological Functions of the UMOD Protein

1.5.6.1 Urinary Tract Infections

Uromodulin protein has been noted for several decades as providing protection against pathological bacterial infection in the nephron. Scanning electron microscope of the mature urine purified UMOD homopolymer was shown to display UMOD as a 'mesh' structure with a pore size of 100-1000nm which theoretically could entangle invasive bacteria prior to their proximity to TAL cells (R and PR, 1988). Additionally, it has been shown that the high-mannose glycan chain of glycosylated UMOD interacts with several *E.coli* adhesion molecules, preventing infiltration- with additional study extending this potential past *E.coli* to several strains of common pathological bacteria and yeast (P. J *et al.*, 2001; D *et al.*, 2004; Coady *et al.*, 2018). Subsequent population level study correlating UMOD levels with urinary tract infection incidence showed that the sub-population of those with the lowest urinary UMOD protein displayed an increased risk of bacteremia (PS *et al.*, 2017).

1.5.6.2 Kidney Stones and Calcium & Magnesium Homeostasis

Protection from kidney stone development has also been a noted function of uromodulin with both male and female *Umod*$^{-/-}$ mice displaying spontaneous intrarenal calcium crystalisation within 9 months of birth (M. L *et al.*, 2004). However, exact underlying mechanisms through which this formation and pathophysiological cell damage is mediated remain unclear (B, 1992). Several case controlled human studies have however shown that urinary uromodulin levels correlated negatively with kidney stone formation risk (A *et al.*, 2000; Lau *et al.*, 2008). Furthermore, interaction between UMOD and the calcium sensing receptor (CaSR) has been observed in both mice and humans, with cinacalet, a CaSR antagonist

shown to decrease urinary UMOD levels- an effect attributed to drug induced alterations in cAMP levels (Tokonami *et al.*, 2018a).

Whilst also acted upon by intracellular Ca2+ levels, UMOD is also known to affect levels of both Ca^{2+} and Mg^{2+} in TAL cells. With UMOD protein levels in vitro shown to diminish the caveolin mediated endocytotic mechanism of TRPV5 and TRPV6, increasing the cell surface bioavailability of both channels and subsequently calcium reabsportion. Decreased luminal TRPV5 was also observed in *Umod*$^{-/-}$ mice in the dCT (L. J *et al.*, 2001; T *et al.*, 2003; M *et al.*, 2018). As with the case for Ca^{2+} reabsorption, similar mechanisms can be observed with Mg^{2+}, where UMOD inhibits endocytosis of the magnesium channel TRPM5 and is hypothesized to play a role in hypomagnesemia (Baaij *et al.*, 2013; MTF, J and M, 2019).

1.5.6.3 Immune Regulation

Extracellular homo-polymeric UMOD protein is known to play both an immunogenic and immunosuppressive role, depending on surrounding physiology. Isolated human polymorphonuclear leukocytes (PMNs) were shown to become activated *in-vitro* in response to incubation with human UMOD protein where it was phagocytosed, with additional study indicating canonical neutrophil activation in response to polymerized UMOD (JK *et al.*, 1990; Thomas *et al.*, 1993). UMOD protein has also been shown to enhanced neutrophil infiltration and migration in renal epithelia (M *et al.*, 2010) and also to stimulate TNF-α secretion in renal F4/80 positive macrophages (Immler *et al.*, 2020a). Furthermore, in the context of ischemia-reperfusion injury, increased necrosis and inflammation was observed in *Umod*$^{-/-}$ mice due to enhancement of the expression of the innate immune mediator toll like receptor-4 (TLR4) in these animals versus controls (El-Achkar *et al.*, 2008).

Basolaterally secreted UMOD protein may act as a systemic immune modulator, potentially through basolateral mechanisms. *Umod*$^{-/-}$ mice were shown to have reduced IL17 expression- a key granulopoietic cytokine, causing neutrophilia. This study showed an increase in neutrophils in both the kidney and also systemically, a finding reversed by the addition of exogenous IL17 treatment (Micanovic *et al.*, 2015).

Additionally, an upregulation of a number of systemic cytokines in response to UMOD levels in the context of hypertension has been suggested, though this was only shown *in silico* and not biologically (Jian *et al.*, 2015).

1.5.6.4 Tubulointerstitial Kidney Disease

Homozygous protein coding mutations on the human UMOD gene have been shown to drive chronic kidney disease, particularly referred to as UMOD-autosomal-dominant-tubulointerstitial kidney disease (UMOD-ADTKD). UMOD-ADTKD a progressive disease which first manifests in teenage years and progresses to end stage renal failure, normally within three to six decades (Bleyer *et al.*, 2021). UMOD-ADTKD is typically characterized physiologically by hyperuricaemia, though with an absence of the haematuria or proteinuria normally associated with other renal disorders (Gast *et al.*, 2018). Furthermore, it was shown that UMOD-ATKD may or may not be associated with the presence of intra medullary cysts in patients (TC *et al.*, 2002).

It is understood that the protein coding mutations driving UMOD-ADTKD cause misfolding which allows the mutated uromodulin protein to accumulate with the endoplasmic reticulum (ER), causing increased intracellular accumulation, ER-stress and reduced uromodulin excretion (L *et al.*, 2003). It was also found that UMOD-ADTKD patients had distinct glycosylation patterns within the uromodulin protein that they express (P *et al.*, 2006). In an ADTKD-UMOD mouse model, it was shown that TAL cells of these animals had upregulation of Grp78 , a marker or ER-stress and additional upregulation of pathways indicating intracellular unfolded protein responses (SE *et al.*, 2017).

1.5.7 UMOD and Blood Pressure Regulation

Whilst the seminal GWAS associating rs13333226 and the UMOD gene with hypertension provided a potential link to blood pressure through sodium natriuresis, they acknowledge there were no causative experiments conducted at that point to provide data on mechanisms (Padmanabhan *et al.*, 2010). In the years since however, two fundamental papers have been published providing causative mechanisms, with both papers differing in their approach and interpretation.

Firstly, Trudu et al published in 2013, suggesting that UMOD affects blood pressure regulation through affecting the *activity* of NKCC2. In particular, they hypothesized that salt-sensitive hypertension is underpinned by aberrant over-activation of NKCC2. In turn, and consistent with this, they generated a human overexpressing UMOD transgenic mouse ($Tg^{Umodwt/w}$), which displayed increased phosphorylation of NKCC2 at two known activation sites (Thr96 and Thr101). These transgenic over-expressers were shown to have increased baseline blood pressures, and also appeared to respond to the removal of NaCl from diet more than their wild type counterparts. Of additional importance in the Trudu paper was the finding that in stable transfectants of NKCC2 in renal cells, activation of the cotransporter was observed in wildtype uromodulin cotransfectants but not in solubilized uromodulin (truncated at the anchor site) cotransfectants, implying that the action of uromodulin in relation to NKCC2 activation is dictated by cellular membrane anchored uromodulin (Trudu *et al.*, 2013).

Trudu et al built further on the mechanism behind their proposed effect, by showing that the modulation of NKCC2 activation by uromodulin in their $Tg^{Umodwt/w}$ mouse model was mediated in-part by increased activation of STE20/SPS1-related proline/alanine-rich kinase (SPAK) and oxidative stress response 1 kinase (OSR1)- themselves known regulators of NKCC2 phosphorylation. Thus, Trudu et al provided a clear causative mechanistic relationship between UMOD and blood pressure regulation, through uromodulin associating with increased phosphorylation/activity of NKCC2 via SPAK and OSR1. Additionally, Trudu et al showed that in the context of elevated *UMOD* expression in mice, loop-diuretic treatment was significantly more effective at reducing blood pressure in $Tg^{Umodwt/w}$ mice versus wildtype counterparts.

By contrast, in 2014, Graham et al employed a knock-out driven approach, using homozygous knock out mice on a Sv129 background ($UMOD^{-/-}$). Like Trudu et al, Graham et al concluded that UMOD associated with blood pressure through its effects on sodium excretion. However, they suggested a different mechanism. They hypothesised that TNF-α is central to regulation, and this was achieved through UMOD mediating the effect of TNF-α on NKCC2 function. To this end, they showed reduced levels of NKCC2 expression in UMOD knockout animals, and that TNF-α excretion in

these animals was also significantly increased in the presence of dietary sodium chloride. With UMOD potentially sequestering TNF-α in the lumen, TNF-α was then available to act in an autocrine manner to inhibit Na^+ uptake by NKCC2, in turn affecting blood pressure. Observation of this mechanism was also reflected in the phenotype of the animals, with *Umod*$^{-/-}$ displaying tolerance to dietary NaCl as a method of increasing this blood pressure. Interestingly, these animals displayed increased urinary excretion of NaCl, both at baseline and with dietary NaCl, suggesting constitutive effects of natriuresis, not limited only to period of elevated NaCl intake (Graham *et al*, 2014).

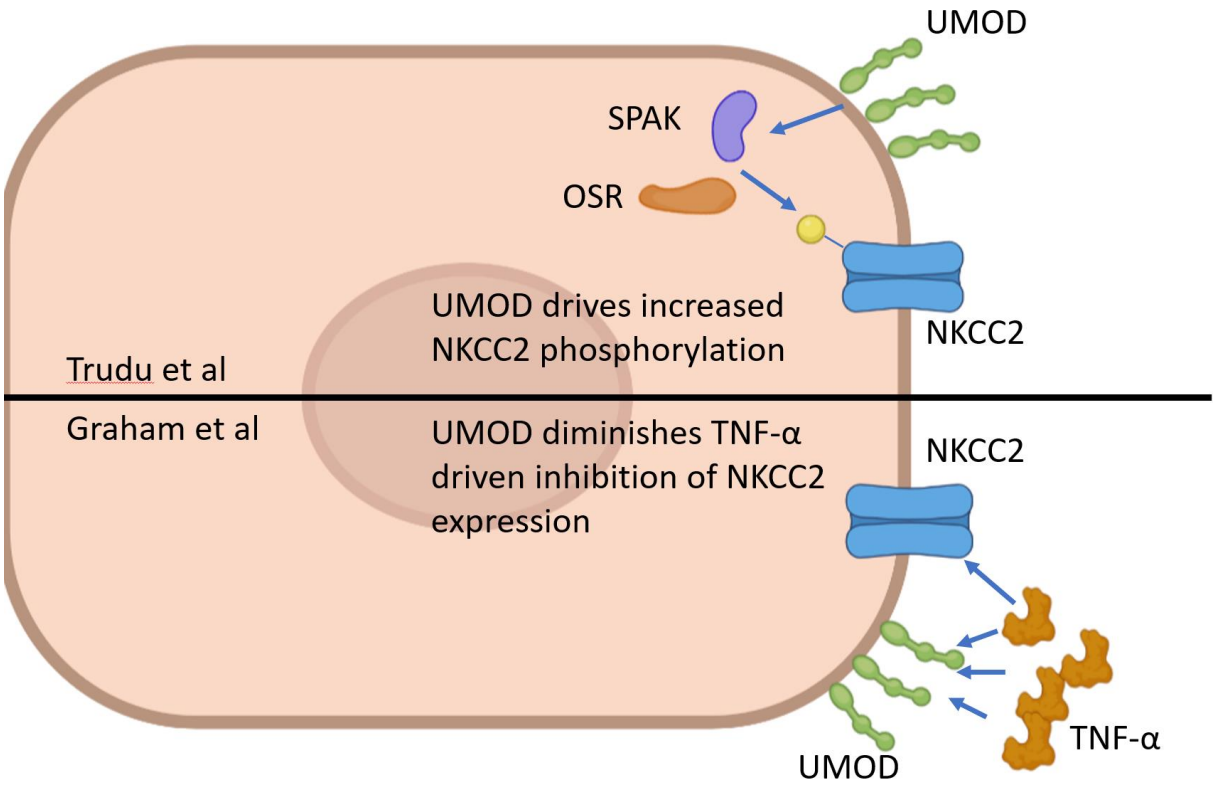


Figure 1.5-3 Two evidence-based UMOD:blood pressure hypotheses.
Trudu *et al* hypothesise that basolateral UMOD drives increased phosphorylation of NKCC2 by SPAK/OSR, leading to increased activity. Graham *et al* hypothesise that basolateral UMOD acts as a TNF-α sink, leading to reduced activatory effects of TNF-α on expression of NKCC2.

1.5.8 The Sodium Potassium Chloride Cotransporter

Both interpretations of the link between UMOD and blood pressure indicating mechanistic action through NKCC2, suggesting this transporter is highly likely to be the causative pathway. Responsible for the filtration and reabsorption of around 200-250g of NaCl per day in humans, NKCC2 is encoded by one gene, but has highly differentiated splicing, which is itself highly location specific along the nephron (Hebert and Gamba, 1994; SC and G, 1994; Castrop and Schießl, 2014). NKCC2

consists of 1099 aminos acids in humans, with a size of 131.4kDa (Markadieu and Delpire, 2014) with glycosylation increasing the molecular weight by 30-40kDa in humans (A, MN and RJ, 2009). As opposed to a single large 'pore like' channel, NKCC2 has a complex structure which contains ion specific interaction sites at a central protein motif, with sites for one Na+ and K+ and two Cl- ions per single transport (Markadieu and Delpire, 2014).

The membrane localization of NKCC2 in its activated form is primarily mediated in an inhibitory manner via cyclic-guanosine-monophosphate (cGMP) and in a stimulatory manner by cyclic-adenosine-monophosphate, with the activatory cAMP pathway known to be modulated by vasopressin (Hebert *et al.*, 1987; A *et al.*, 1991). Specifically speaking, this subsequent finding identified that the presence of vasopressin in TAL cells underpins the requirement of NKCC2 to function as a potassium transporter (such that in the absence of vasopressin, potassium was not transported by the cotransporter). Additionally, in a separate mechanism, Ca2+ and Mg2+ stimulation the CaSR causes modulation of cAMP in TAL cells potentially allowing the CaSR to affect the activity of NKCC2 in an ion-dependent manner (G and PA, 2009). Furthermore, TNF-α has been shown to affect the bioavailability and function of NKCC2, reflecting the *in vivo* findings of Graham et al in *Umod*$^{-/-}$ mice (S *et al.*, 2011). Specifically, this study examined the TAL cells of *TNF*$^{-/-}$ mice, where it was observed that in the knockout strain, both protein and mRNA expression of NKCC2A were significantly upregulated. Like Graham et al, this study suggests that TNA-α acts as an endogenous inhibitor of NKCC2. Lastly, NKCC2 activity has been shown to respond to NaCl levels, which can be attenuated by TNF-α, with low salt diet leading to increased phosphorylation of NKCC2 and upregulation of NKCC2B. This upregulation was subsequently diminished by treatment with endogenous TNF-α (S *et al.*, 2020).

Loop diuretics (Figure 1.2-2) are highly potent actors at NKCC2, in particular, loop diuretics bind to the chloride binding site on the cotransporter, consequently inhibiting sodium reabsorption (Ellison, 2019b). However, despite the high efficacy of loop diuretics through their effects on NKCC2, these drugs are not normally recommended for long-term interventions due to their notable side effects-

particularly their ability to induce rapid uresis leading to discomfort in patients. Building on this, studies have suggested the need for further refinement of the NKCC2 induced diuretic effects of loop diuretics in order for their side effects to be better managed (Caceres and Ortiz, 2019).

1.6 The UMOD Precision Medicine Hypothesis

Whilst there is a lack of agreement in the field on the potential mechanisms underpinning the relationship between UMOD and the activity of NKCC2, the overlying hypothesis is clear; UMOD associates with blood pressure primarily through exerting effects on NKCC. Thus, UMOD may itself be a hypertension precision medicine target.

With volume overload via natriuresis considered a major driver of hypertension, loop diuretics should be optimized as much as possible as a therapeutic, and administered as first line treatment where appropriate. In the context of UMOD, the hypertensive population can be stratified into two groups, differentiated by their potential response to loop diuretic medication. Firstly, the UMOD 'risk' genotype (those with increased UMOD levels) can be considered 'optimal responders' to loop diuretics due to the previously discussed relationship between UMOD, NKCC2 and the hypertensive phenotype. Contrastingly, those individuals with the UMOD 'protective' genotype can be regarded as likely to have diminished NKCC2 driven sodium reabsorption, resulting in weakened responses to loop-diuretics versus the risk genotype.

Such a hypothesis has led to the development and execution of an on-going multicentre trial (n=240), using torasemide as the target medication and rs13333226 as their identifying variant. The BHF-UMOD study intends to use the change in 24-hour ambulatory SBP area under the curve between baseline and end of treatment as their primary study metric, though they also hope to capture mechanism validation data via urinary UMOD excretion and serum ion concentrations (McCallum *et al.*, 2021).

1.6.1 Thesis Hypothesis

We hypothesise that the UMOD gene is a precision medicine target. However, unlike McCallum et al, we believe that the underlying genetic contribution on the cis-UMOD promoter may not exclusively be affected (or affected at all) by rs13333226. Additionally, we believe that understanding the transcriptional mechanisms underlying genotype dependent UMOD expression may further inform and refine the UMOD precision medicine hypothesis. Lastly, we hypothesise that whilst UMOD affects NKCC2 protein quantity or activity, the drivers of this relationship have not yet been fully elucidated and may be multifaceted. Greater understanding of the mechanisms associating blood pressure with the UMOD locus are likely to benefit any downstream clinical investigation. Therefore, we propose a basic science approach to provide new insights to this phenomenon.

1.6.2 Thesis Aims

The stated aims of this thesis are as follows;

- To provide additional insights into which variants on the UMOD promoter drive expression, and to study the mechanisms underlying their activity
- To further examine the relationship between the expression of UMOD and NKCC2
- To establish a system of stable expression of the UMOD gene to facilitate trafficking studies
- To perform *in vivo* experimentation examining the relationship between sodium chloride and UMOD expression
- To use the impact of the Covid-19 pandemic to develop a series of computational biology tools to assist in experimentation and hypothesis development

2 General Methods

2.1 Extraction of Genetic Material and Polymerase Chain Reaction

2.1.1 Human Renal Biopsy Processing

Diseased human kidneys were extracted from renal cancer patients by invasive surgery with the patient under general anesthetic during the procedure. During *ex-vivo*, a 1mm wide bore was used to sample tissue from the opposite (non-diseased) pole of the cancerous kidney, to an average depth of 10mm. Tissue was immediately stored in RNALater (Sigma-Aldrich, Dorset, UK) and shipped on the same day to the Institute of Cardiovascular and Medicinal Sciences, University of Glasgow. Upon receipt, samples were immediately removed from solution and stored at -80˚C until use. Accompanying each sample was phenotypic information consisting of systolic blood pressure, pulse rate, age, sex and body-mass-index. Systolic blood pressure and pulse rate were not considered reliable variables due to their single-sampling methodology in the period when the patient was due to undergo invasive surgery.

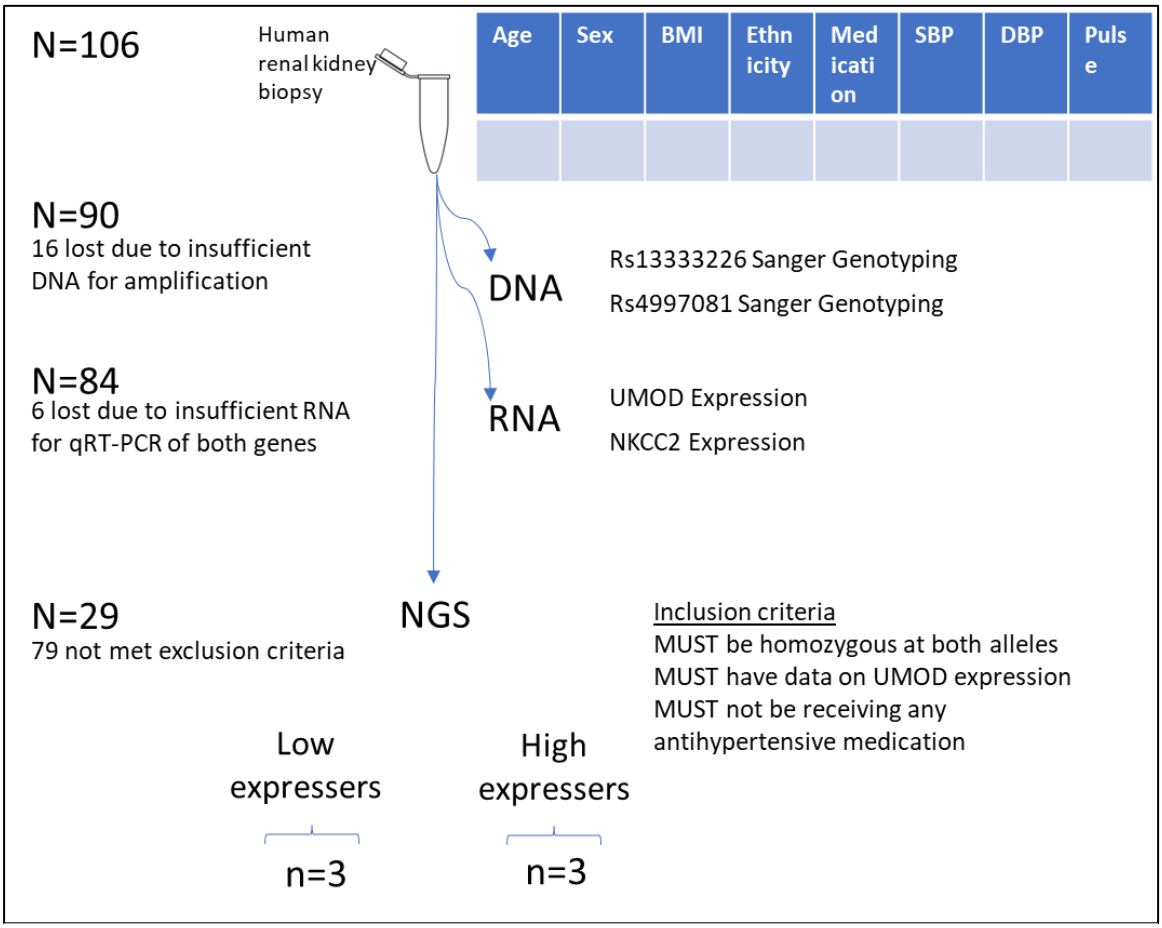


Figure 2.1-0 Selection process of samples for experimental methodologies with attrition shown. Due to low yields of nucleotide material, some samples were lost at both DNA and RNA extraction steps. For NGS experiments, an inclusion criterion was developed, from this subset of n=29 appropriate samples, n=3 per group were selected based on their expression of UMOD, as determined by qRT-PCR.

2.1.2 DNA Extraction

All samples were processed using DNeasy Blood and Tissue Kit (Qiagen, Manchester, UK) on a RNAzap (Sigma-Aldrich, Dorset, UK) and 70% ethanol (dH_2O) cleaned lab-bench, using filter pipette tips only to minimize contamination. Where appropriate, frozen samples were fragmented using a liquid nitrogen cooled mortar and pestle, to obtain chipped products which amassed to less than 30mg, to prevent overloading of the column. Samples were first lysed in 220µL 1:10 Proteinase-K:Buffer AL for 1-hour

at 56˚C and vortexed every 10min. Following this, 200µL 100% ethanol was added and the mixture thoroughly mixed. The resulting mixture was then transferred to a DNeasy Mini-spin column and centrifuged at 8000 x g (RCF) for 1min and flow-through discarded. Buffer AW1 was then applied to the column and centrifuged at 8000 x g (RCF) for 1min with flow-through again discarded. Finally, buffer AW2 was applied to the column and centrifuged at 8000 x g (RCF) for 3min. The column insert was then removed and placed in a fresh 1.5ml eppendorf. To the membrane of the column directly, 30µL nuclease-free-water (Thermofisher, Paisley, UK) was applied. The column was rested at room temperature for 1min and then centrifuged at 8000 x g (RCF) for 1min. The flow through of this centrifugation was then re-applied to the membrane of the column and then re-centrifuged under the same conditions, increasing yield. For yield quantification, 1µL of sample was pipetted onto the receptacle of a NanoDrop ND-1000 Spectrophotometer (ThermoFisher, Paisley, UK) with a 260/280 ratio of >1.8 accepted as 'pure' in accordance with machine use guidelines. Product DNA was then stored at -20˚C until use.

2.1.3 RNA Extraction

All samples were processed using DNeasy Blood and Tissue Kit (Qiagen, Manchester, UK) in an RNAzap (Sigma-Aldrich, Dorset, UK) and 70% ethanol (dH_2O) cleaned fume hood, using filter pipette tips only to minimize contamination. Where appropriate, frozen samples were fragmented using a liquid nitrogen cooled mortar and pestle, to obtain chipped products which amassed to less than 30mg, to prevent overloading of the column. No more than $1e10^6$ were used, suspended in PBS, where applicable. To tissues, 700µL QIAzol Lysis Reagent (Qiagen, Manchester, UK) was added alongside one sterile stainless steel ball, and these samples disrupted and homogenized using a Polytron PT2100 Benchtop Homogeniser (KINEMATICA, Luzern, Switzerland) set to a 30Hz cycling period for 2min. Where cells were used, media was removed from adherent cells, which were then washed with 3x PBS washes. Following this, cells were then directly lysed on-plate using 700µL QIAzol. Homogenised QIAzol suspensions were then incubated at room temperature (19-22˚C) before 140µL chloroform was added to each sample and tubes thoroughly mixed by hand-shaking. Product was then incubated for 3min at room temperature before being centrifuged

at 12000 x g (RCF) at 4˚C for 15min. The upper aqueous phase of three specific layers was then extracted by pipette and transferred to a fresh 1.5 eppendorf where it was mixed by pipetting with 525µL 100% ethanol. A maximum of 700µL of this product was then transferred to RNeasy Mini Spin columns and centrifuged at 8000 x g (RCF) for 15s (at room temperature), with flow through discarded, any left over product greater than 700µL was then subjected to the same method.

Following this, and on-column DNAse digest was performed by first adding 350µL Buffer RWT then centrifuging at 8000 x g (RCF) for 1min. To the column, a total of 80µL DNase-I: Buffer RDD was added (10µL:70µL) and incubated at room temperature for 15min. After 15min, 350µL Buffer RWT was added to the column and then centrifuged at 8000 x g (RCF) for 1min To each dried column, 500µL buffer RPE was added and this centrifuged at 8000 x g (RCF) for 15s with flow-through discarded. An additional 500µL buffer RPE was added and this again centrifuged at 8000 x g (RCF) for 2min with flow-through discarded. To the membrane of the column directly, 30µL nuclease-free-water (Thermofisher, Paisley, UK) was applied. The column was rested at room temperature for 1min and then centrifuged at 8000 x g (RCF) for 1min. The flow through of this centrifugation was then re-applied to the membrane of the column and then re-centrifuged under the same conditions, in theory increasing yield. For yield quantification, 1µL of sample was pipetted onto the receptacle of a NanoDrop ND-1000 Spectrophotometer (ThermoFisher, Paisley, UK) with a 260/280 ratio of >2 accepted as ‘pure’ in accordance with machine use guidelines. Product RNA was then stored at -80˚C until use.

2.1.4 End-point Polymerase Chain Reaction and Electrophoresis

All end-point PCR reactions were performed on genomic-DNA products obtained from extraction using Qiagen DNeasy Blood and Tissue Kit as detailed in Section 2.1.2. Input gDNA yield to the reaction was 10ng, suspended in an appropriate volume of nuclease-free-water. HotStarTaq DNA Polymerase (Qiagen, Manchester, UK) was selected as the polymerase enzyme for these reactions. A series of different reactions were performed, depending on the required product. Amplification of the human UMOD promoter for Sanger genotyping (2.1.4.1), amplification of the human UMOD

gene for next-generation sequencing (2.1.4.2), amplification of full length human UMOD cDNA (2.1.4.3) and finally amplification of the mouse UMOD promoter for UMOD alleles in the $UMOD^{-/-}$:$UMOD^{wt}$ colony (2.1.4.4). Human primers were designed against GRCh38 and mouse primers against Mm9 genome builds and ordered from Eurofins Genomics (Ebersberg, Germany) where they were assembled and purified by HPSF before being shipped lyophilized and resuspended in-lab to concentrations of 100mM in nuclease-free-water. All end-point PCR reactions were amplified on a MJ Research PTC 225 Tetrad thermal cycler (Marshall Scientific, Hampton NH, USA) using 96-well PCR plates (Starlab, Milton Keynes, UK) .

An agarose gel consisting of 0.8-1.2% Agarose (Invitrogen, Carlsbad, USA) in 120ml Tris-borate EDTA (TBE) was prepared using a microwave and upon cooling, 1.5µL 10mg/ml ethidium bromide solution ((Sigma-Aldrich, Dorset, UK) was carefully added in a fume hood before gentle mixing. Molten agarose was poured into a gel template with a 15-well comb before being solidified over 1.5hours. Upon solidification the gel was transferred to an electrophoresis system (BioRad, Hampstead, UK). 2µL 5X loading dye (Thermofisher, Paisley, UK) was added to 8µL PCR product and the resulting 10µL added to wells on the gel. In lane 1 8µL of DNA ladder of appropriate size (either 100bp or 1kb) was added (NEB, Ipswich MA, USA). Electrophoresis was performed on the samples by running the gel in 1X TBE solution at 100V for 1-hour using a BioRad Powerpac 300 direct-current system. Following this, gels were visualized using a BioRad Molecular Imager XRS+ transillumination system (BioRad, Hampstead, UK). Gel products were then discarded in specific ethidium bromide contaminated wastebins.

2.1.4.1 Amplification for Sanger Genotyping of rs13333226 and rs4997081

Reagents	
Mastermix Reagents	Volume (µL)
Template gDNA	2 (5ng/µL stock)
dNTP (10mM)	0.6
10X PCR Buffer	3
HotStarTaq DNA Polymerase	0.3

Nuclease Free Water	22.1
Forward Primer (10nM)	1
Reverse Primer (10nM)	1
Cycling Conditions	
Time	Temperature (˚C)
15min	95
1min	95
1min	58
1min	72
Cycle back to step-2 34 times	
10min	72
Primers (GRCh38)	
Sequence	Tm (˚C)
5'-GAAACTGGTGAGTAGTGTTGGT-3 (Forward)	59.6
5'-CATCAACAATGGCACATGTATA-3 (Reverse)	60.7

Table 2.1-1 Genotyping Standard Operating Procedure

2.1.4.2 Amplification for Targetted Next Generation Sequencing of the Human UMOD Gene

Reagents	
Mastermix Reagents	Volume (µL)
Template gDNA	2 (10ng/µL stock)
dNTP (10mM)	0.6
10X PCR Buffer	3
HotStarTaq DNA Polymerase	0.3
Nuclease Free Water	21.7
Forward Primer (10nM)	1.2
Reverse Primer (10nM)	1.2
Cycling Conditions	

Time	Temperature (˚C)	
15min	95	
1min	95	
1min	At least 5˚C lower than primer Tm	
1min	72	
Cycle back to step-2 34 times		
10min	72	
Primers (GRCh38)		
Sequence	Fragment ID	Tm (˚C)
5'-TCATCCTTACACCTACTTCC-3' (Forward)	1	55.6
5'-ACGAACCTTAATCTCTCAGC-3' (Reverse)		57.6
5'-AAAACACAAATTAGCCGGGC-3' (Forward)	2	64.7
5'-AAAAGTCAGTGTGAGAGTGG-3' (Reverse)		56.8
5'-GACAAGTTAATGGGTGCAGC-3' (Forward)	3	62.1
5'-GATGCGACCTAAAACACTGC-3' (Reverse)		62.3
5'-CTCTTTGATCCTCTCTGTCC-3' (Forward)	4	57.9
5'-TGTGAACAGAGATGGATGGG-3' (Reverse)		63.7
5'-GGGATGTTGGTGAGGTAAGG-3' (Forward)	5	63.1
5'-TCCAAAACAAGGAAAGAAGC-3' (Reverse)		60.8
5'-CCAGGAATTGGAGGCTATGG-3' (Forward)	6	65.4

5'-TGAGCTGATTGGTGGACTGG-3 (Reverse)		66.4
5'-GAGGAGTGCAGTATAGACGAGG-3' (Forward)	7	61.8
5'-TGTCATTGGCCCCACATTCC-3' (Reverse)		69.9

Table 2.1-2 Targetted Resequencing Standard Operating Procedure

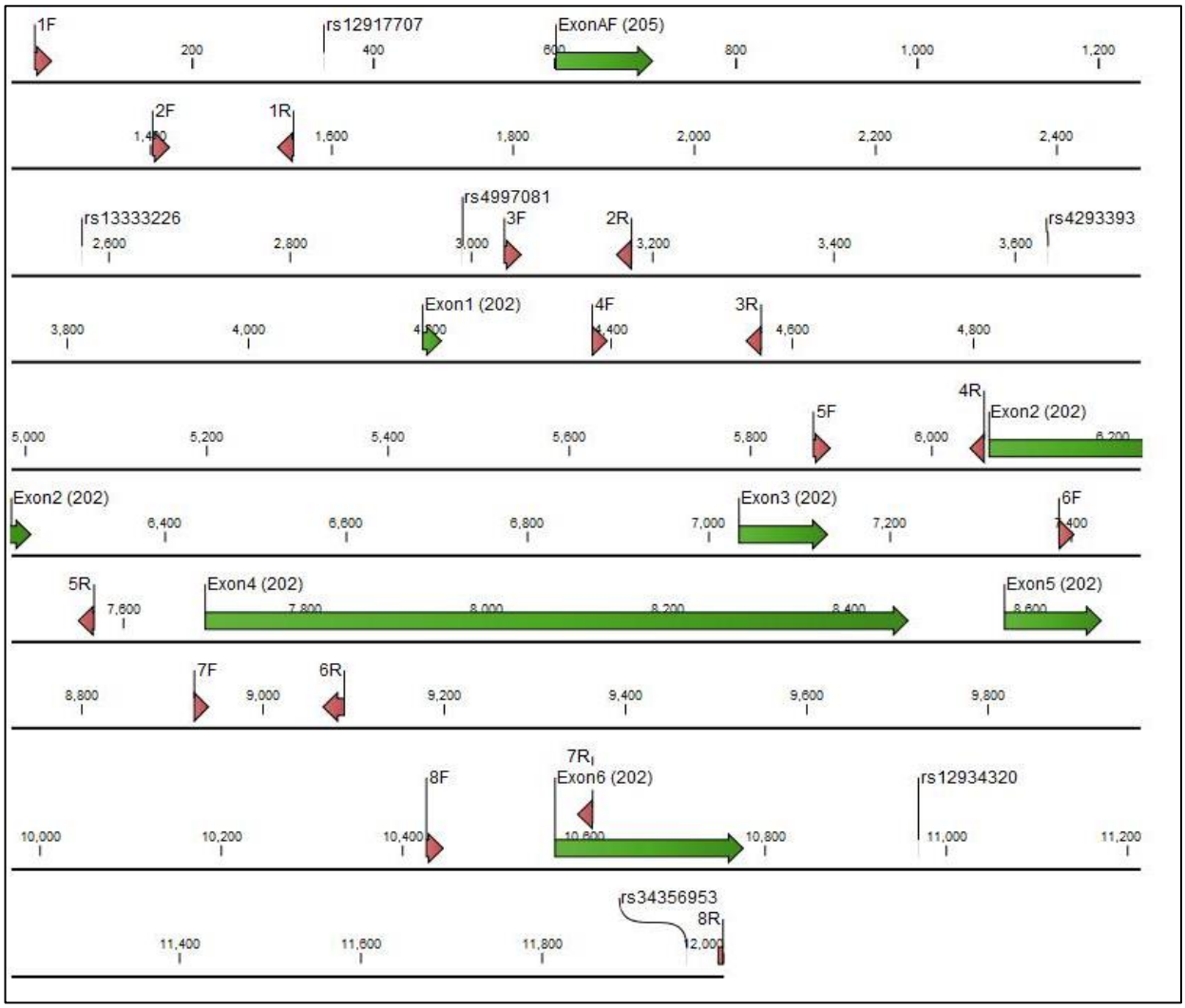


Figure 2.1-1 Primer annealing points (red) on the UMOD gene

2.1.4.3 Amplification of full-length human UMOD cDNA

Reagents		
Mastermix Reagents	Volume (µL)	
Template cDNA	2µL (not quantified)	
dNTP (10mM)	0.6	
10X PCR Buffer	3	
HotStarTaq DNA Polymerase	0.3	
Nuclease Free Water	22.1	
Forward Primer (10nM)	1	
Reverse Primer (10nM)	1	
Cycling Conditions		
Time	Temperature (˚C)	
15min	95	
1min	95	
1min	58	
1min	72	
Cycle back to step-2 34 times		
10min	72	
Primers (GRCh38)		
Sequence	Technique	Tm (˚C)
5'-AGACTAACTCTACCTTTCTG -3 (Fwd [5'UTR])	cDNA amplification	49.1
5'- GTACACCGTCACAAGTCCCAT-3 (Rev [exon-11])		52.4
5'- TAATACGACTCACTATAG- 3' (Fwd [T7])	Insert orientation verification	40.9
5'-GTACACCGTCACAAGTCCCAT-3' (Rev [exon-11])		63.7

Table 2.1-3 Full length UMOD cDNA Synthesis Standard Operating Procedure

2.1.4.4 Amplification for Characterisation UMOD alleles of SV129 Mice

Reagents		
Mastermix Reagents	Volume (µL)	
Template gDNA	2 (10ng/µL stock)	
dNTP (10mM)	0.6	
10X PCR Buffer	3	
HotStarTaq DNA Polymerase	0.3	
Nuclease Free Water	22.1	
Forward Primer (10nM)	1	
Reverse Primer (10nM)	1	
Cycling Conditions		
Time	Temperature (˚C)	
15min	95	
1min	95	
1min	At least 5˚C lower than primer Tm	
1min	72	
Cycle back to step-2 34 times		
10min	72	
Primers (Mm9)		
Sequence	UMOD Fragment	Tm (˚C)
5'- AGGGCTTTACAGGGGATGGTTG -3' (Forward)	Wild-type	68.9
5'- GATTGCACTCAGGGGGCTCTGT -3' (Reverse)		59.1
5'- AGGGCTTTACAGGGGATGGTTG -3' (Forward)	Knock-out	68.9
5'- GGCTATTCGGCTATGACTGGG -3' (Reverse)		66.6

Table 2.1-4 Mouse Genotyping Standard Operating Procedure

2.1.4.5 PCR Product Purification and Sanger Sequencing

Where appropriate, in the context of samples being analysed either by Sanger or next-generation-sequencing, products were purified by QIAquick PCR Purification Kit (Qiagen, Manchester, UK). To each 30µL PCR product, 150µL Buffer PB was added before the solution was transferred to a QIAquick column and centrifuged for 1min at 8000 x g (RCF) and flow through discarded. To the dried column, 750µL Buffer PE was added and again centrifuged for 1min at 8000 x g (RCF). Following this, 30µL nuclease-free-water was applied directly to the membrane and the column left for 1min, after which columns were centrifuged at 8000 x g for 1min to collect the eluate. The flow through of this centrifugation was then re-applied to the membrane of the column and then re-centrifuged under the same conditions, in theory increasing yield. For yield quantification, 1µL of sample was pipetted onto the receptacle of a NanoDrop ND-1000 Spectrophotometer (ThermoFisher, Paisley, UK) with a 260/280 ratio of >1.8 accepted as 'pure' in accordance with machine use guidelines. Purified product was stored at -20˚C until use. The total volume of purified samples were sent to GATC (discontinued, now Eurofins (Ebersberg, Germany)) for Sanger sequencing. Returned AB1 files which contain the electropherogram signal were analysed for genotype at rs13333226 and rs4997081 by eye using CLC Genomics Software (Qiagen, Manchester, UK). Homozgous signals were distinguished the electropherogram displaying a single peak for a specific base, heretozygous signals were characterized by the electropherogram displaying a double peak of half-height.

2.1.5 cDNA Preparation

All cDNA reactions were prepared in the absence of gene-specific-primers on a RNAzap (Sigma-Aldrich, Dorset, UK) and 70% ethanol (dH_2O) cleaned bench, using filter pipette tips only to minimize contamination. Superscript-IV (Thermofisher, Paisley, UK) reverse transcription reactions were performed using 200ng RNA extracted by methods detailed in 2.1.3. The resulting 20µL cDNA product was diluted in 80µl nuclease-free-water to give a final dilution of 1:5 and stored at -20˚C until further use.

Reaction Vessel 1	
Reagents	Volume (µL)
Template RNA	2 (100ng/µL stock)
dNTP (10mM)	1
Random Hexamers (50µM)	1
Nuclease free water	9
Cycling Conditions (Vessel 1)	
Time	Temperature (˚C)
5min	65
1min	4
Reaction Vessel 2	
Reagents	Volume (µL)
5x Superscript IV Buffer	4
100mM DTT	1
RNAseOUT (Thermofisher)	1
Superscript IV Reverse Transcriptase	1
Vessel 1 Product	13
Cycling Conditions (Vessel 2)	
Time	Temperature (˚C)
10min	23
10min	55
10min	80

Table 2.1-5 Pre-Taqman cDNA Synthesis Standard Operating Procedure

2.1.6 Taqman Quantitative Real-time PCR

Diluted cDNA from 2.1.5 was combined with the desired Taqman qRT-PCR probe (Thermofisher, Paisley, UK) and Taqman Mastermix (Thermofisher, Paisley, UK) was added to wells on a 384-well optical plate (Thermofisher, Paisley, UK). All reactions were performed in triplicate and plates featured both an RT (reverse-transcription) negative control and no-RT negative control. Reactions were duplexed, with the gene-of-interest assayed using a FAM labelled probe and a housekeeper gene assayed

with a VIC labelled probe in the same well. Plates were amplified using a QuantStudio 12K Flex Real-Time PCR System (Thermofisher, Paisley, UK). Raw data was displayed as Ct values, in triplicate. Any biological n with a technical replicate Ct standard deviation value > 0.5 was excluded from analysis, in the exception that it was not immediately apparent there was a user error in one of the 3 triplicate wells. If there was a clear user-error in a well, this well was excluded from the analysis Triplicates were reduced to means and mean Ct values used to calculate relative quantification via the double-delta-Ct method. The ddCT method first involved subtracting the Ct of the gene of housekeeper from the gene of interest (dCT), before normalizing each sample to the average dCT of the control group (ddCT). Following this, the second exponent of the negative ddCT was calculated to determine in the fold change in gene expression.

ΔCt = Ct (gene of interest) – Ct (housekeeping gene)

ΔΔCt = ΔCt (Sample) – ΔCt (Control average)

Fold gene expression = 2^-(ΔΔCt)

Reagents	
Mastermix Reagents	Volume (µL)
Universal Mastermix (no UNG)	2.5
Gene of Interest (FAM Fluorophore)	0.25
Housekeeper (VIC Fluorophore)	0.25
cDNA Template	2
Cycling Conditions	
Method	Selection
Cycling Style	Standard
Well volume	5 µL

Table 2.1-6 Taqman qRT-PCR Standard Operating Procedure

2.1.7 Sample Preparation of Next Generation Sequencing Samples

Relative quantification for UMOD mRNA of all 86 human renal samples was performed. Of these data, 3 individuals with UMOD mRNA expression in the lower 20th percentile of the gaussian distribution were selected as the 'low UMOD expression' group and conversely individuals with UMOD expression in the upper 20th percentile selected as the 'high UMOD expression' group. Due to insufficient evidence on the correlation between anti-hypertensive medication and UMOD expression, any individual samples which were recorded as taking medication at the time of surgery were excluded. Both next-generation-sequencing of DNA and RNA was performed on the same samples. Prior to RNA-sequencing, samples were analysed using Agilent 2100 Bioanalyzer with a Eukaryote Total RNA Nano assay (Glasgow Polyomics). RIN values greater than 6 were considered acceptable.

2.2 Next Generation Sequencing

2.2.1 Targetted DNA Sequencing of the Human UMOD Gene

Seven consecutive 1.5kb fragments of the human UMOD gene, starting 2kb upstream of the UMOD-201 human transcriptional-start-site were amplified by 2.1.4.2 and purified by 2.1.4.5. These purified fragments were pooled to produce a final equimolar solution of 1000ng total DNA in 100µL, suspended in nuclease-free-water. Samples were then shipped on dry ice to Glasgow Polyomics (Garscube, Glasgow, UK). An operator at Glasgow Polyomics subsequently fragmented this solution into 150bp reads before ligating adapters and beads provided with the TruSeq DNA Nano kit (Illumina, SanDiego CA, USA). The prepared library was then sequenced on the MiSeq2 platform in paired end style (Illumina, SanDiego CA, USA). Data was returned in the form of raw reads which were trimmed and analysed by the author.

2.2.2 DNA Sequence Data Processing Pipeline

Raw reads were obtained and loaded into a Linux operating system (Ubuntu 16.04.1 LTS). Cutadapt (Version 3.4), running in Python 3.7 was used to trim adapter sequences from reads (5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCA-3' and 5'-AGATCGGAAGAGCGTCGTGTGTAGGGGAAAGAGTGT-3'). Reads were then aligned to

GRCh38 (Release 78) using Burrow-Wheeler-Aligner (Version 0.7.17) to produce a binary-alignment-file (BAM). Picard tools (version 2.20.6) was then used to mark PCR duplicates in the BAM file, following which duplicates were removed. Genome-analysis-toolkit (GATK) (Version 4.0.7.0) haplotype-caller was then used to detect SNPs from these BAM files, with base calls excluded if the Phred score of the base was <20 (5% chance of a mis-call). Haplotype caller provided an endpoint tabular file containing all detected variants, their coordinates, db-SNP annotation and allele count.

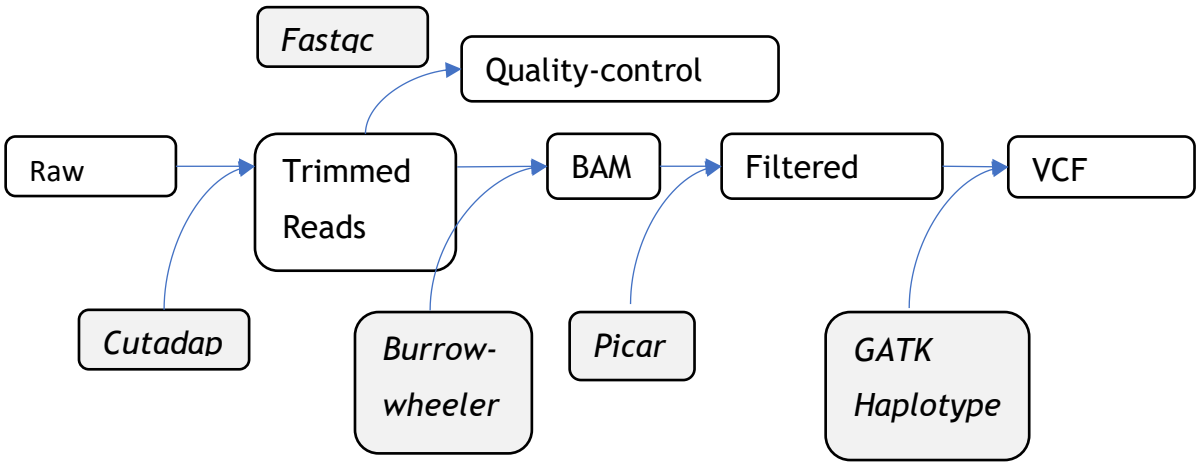


Figure 2.2-1 Genomic fastqc processing pipeline

2.2.3 Total RNA Sequencing between Low and High human UMOD Expressers

Following verification of integrity of RNA, 30µL of human renal RNA, concentration 30µg/µl, was shipped on dry ice to Glasgow Polyomics (Garscube, Glasgow, UK). Upon receipt, an operator fragmented RNA to 75bp reads and libraries were prepared using Illumina TruSeq Stranded mRNA Sample Preparation Kit adapters (Illumina, SanDiego CA, USA). The prepared library was then sequenced on the NextSeq500 platform in paired end style using the in-house sequencer at Glasgow Polyomics (Illumina, SanDiego CA, USA).

2.2.4 RNA Sequence Data Processing Pipeline

Raw reads had their adapters (5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC'-3', 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3') trimmed and bases with Phred score < 15 were removed using FastP (Version 0.20.0). Reads were then assessed for potential contaminants by Fastq-Screen (Version 0.13.0). Following this, reads were aligned to GRCh38 (Release 78) using the splice-sensitive aligner Hisat2 (Version 2.1.0). Transcript abundances were then determined by Stringtie (Version 1.3.6) using the GRCh38 Gene-feature-file (Release 79). Stringtie tabular counts were then ported to R (version 3.6.0) using TxImport where samples were assessed for differential expression between groups using Deseq2. Further discussion of the statistical methods of Deseq2 can be found in 4.3.2.

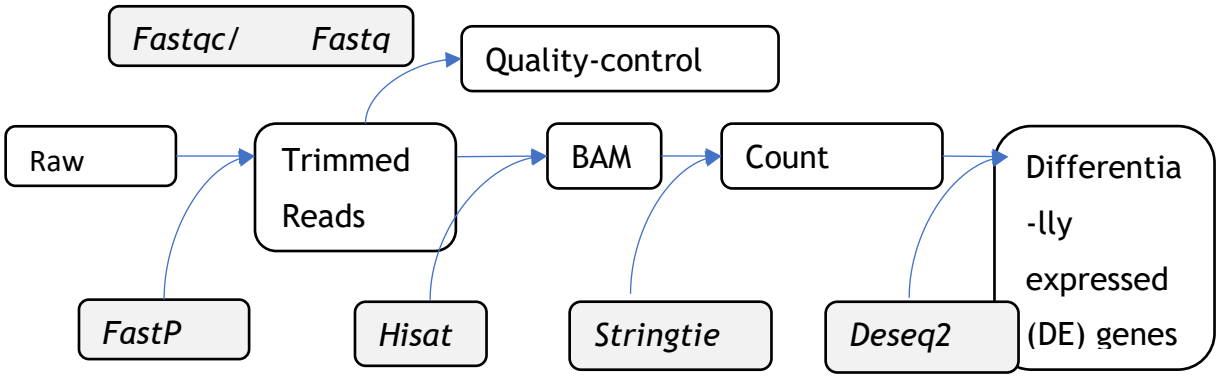


Figure 2.2-2 Transcriptomic fastqc processing pipeline

2.3 Computational Biology

2.3.1 Software

During the Covid19 disruption and based on several potential data analysis applications using data already generated (particularly RNA-seq), Python 3.8 was learned and executed across several open-source programs by the author. Python was chosen over R due to its inherent flexibility and more widely applicable use-cases in biology. The author's Integrated Development Environment (IDE) was created around

Jupyter Client (Version 6.1.5), though as applications became larger in scope, Microsoft Visual Studio Code 2019 was transitioned into. In total over 5000 lines of code were written to generate these novel applications. All underlying code for applications can be found in appendix or otherwise is fully available at https://github.com/SimonF92. Modules were installed into the IDE using Pip (Version 21.1.2) or Anaconda, depending on requirements.

2.3.2 PadPlot

Padplot describes a web-application authored in full by the writer during the initial covid19 pandemic. Padplot is a modular program for visualization and analysis of bulk RNA-sequencing data. No other software developers contributed to the writing and hosting of Padplot. Padplot was hosted on a linux based virtual machine provided by Amazon Web Services.

2.3.2.1 Data Processing

Deseq2 output files were loaded into a Python environment using Pandas (Version 1.0.5) following which log2FoldChanges, adjusted-pvalues and gene-names were created as individual list variables. Adjusted-pvalues were transformed to both negative log-pvalues and negative log-10-pvalues by Numpy (Version 1.19.5) before being reintegrated with the original dataframe variable. Padplot was then constructed around this central ‘prepdata’ function which could be used via flow control to send necessary values to each of the individual plotting functions.

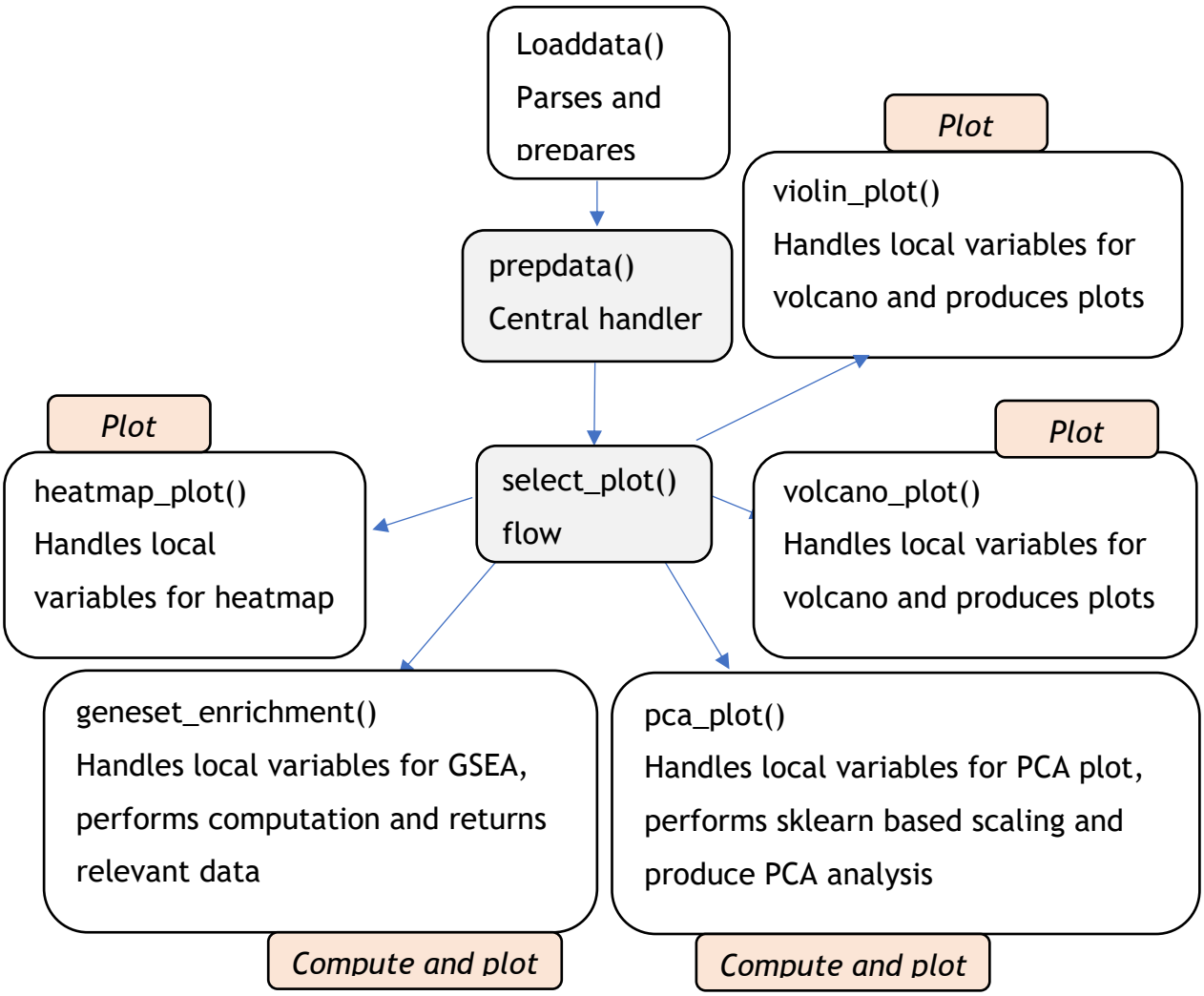


Figure 2.3-1 Padplot web application architecture

2.3.2.2 Principal Component Analysis

Principal component analysis was performed using sklearn.preprocessing StandardScaler and sklearn.decomposition PCA (Version 0.23.1). A dataframe was created from raw counts at each gene for each sample, with an option provided via flow control to convert all values to their log(count+1) counterpart- to provide a log-transformation alternative. A 'features' list was subsequently created producing a range of values between 0 and the total count of genes in the sample. Each gene per sample was transformed via pca_fit into PCA scaled counterparts. An 'explained

variance' variable was then created by taking the explained variance value at each of the components and converting this to a percentage against the sum of all variances, with PC1 representing the component with the greatest variance in the data. From the list of components, a new dataframe was created containing both PC1 and PC2 for each sample, with these values providing the x and y coordinates of a scatterplot visualization and the sample name providing a text variable for label at each of these points. This scatterplot was created using matplotlib (Version 3.3.1), seaborn (Version 0.10.1) and text labels were dodged from each other using adjustText (Version 0.7.3).

2.3.2.3 Graphical Visualisations

Both log10 and log2 options were provided to the user in terms of p-value transformations, with the discretionary choice relating to the skewness in their data. A figure canvas was created with Matplolib.pyplot and the volcano-plot cast to it using Seaborn.scatter, with each point being represented by local variables of log2foldchange on the x-axis and negative log-pvalue on the y-axis. Individual points on the scatter were coloured by Boolean depending on their significance. Genes with the highest negative log-pvalue were also tagged with this respective gene names using plt.text, creating a variable which was then passed through AdjustText in order to prevent overlap of gene names within the final visualization. In the case of human data, an additional option was provided to merge gene names within the dataset with Gene Ontology definitions (using Pandas), generating volcano plots in which the user could define and highlight genes relating to a particular process within their data. Options for both vertical and horizontal guidebars to indicate significance were also added be assigning plt.hline and plt.vline variables to these.

For heatmap visualisations, users were first asked to provide data grouping following which the basic dataframe obtained from prepdata() was melted around grouping using Pandas. Data was then plotted using the clustering algorithm of Seaborn due to its inherent ability to produce hierarchy before being cast to a Matplotlib canvas. All data was plotted on z-score of individual samples relating to the mean of counts at that gene, as opposed to raw counts.

Lastly, for violinplot visualisations, as for heatmap, the user was given the option to log(counts +1) transform their data in order to reduce skewness, creating a new dataframe variable. Again these counts were melted around grouping, following which a Matplotlib canvas was prepared to which a Seaborn violinplot object was cast. Additional options were given to the user to add a swarmplot jitter based on their requirements alongside sample grouping.

2.3.2.4 Gene Set Enrichment Analysis

Gene set enrichment within Padplot was performed using the Gsea.py library (0.10.2). A subsetted dataframe of gene counts at each sample for each gene was created from 'prepdata()' before this new 'gene_exp' variable was subsetted to the top-7000 genes in terms of expression. Using caching, gene sets were pulled via the Enrichr API and geneset enrichment performed using a permutation analysis across all genes- with no thresholding. One hundred permutations were performed and enriched pathways determined by their signal to noise ratio. The gene-set variable created was parsed into a dataframe from which both total gene-sets and individual gene-sets could be accessed. In the case of individual gene-sets, visualization was based on creating the ratio of experimental to control gene expression and then representing this ratio on a sliding scale with an underlying colourmap from blue to red. The user was given control over all of the statistical methods and gene-sets against which to compare.

2.3.2.5 Interactivity

Streamlit (Version 0.69.2) was used to provide interactivity to the user with input and output (IO) being pushed through the Streamlit python to css and html conversion process. All variables, including those relating to plotting were modifiable through Streamlit to give the end user control over plots. The webpage was continuously cached on the host machine to improve speeds for the user.

2.3.2.6 Web hosting

A Linux (Ubuntu 16.04.1 LTS) machine with 1 CPU core and 1Gb RAM was reserved on Amazon Web Services (AWS, Seattle WA, USA). Via controlled access through PuTTy, to this, a Python 3.8 environment was constructed with appropriate libraries installed via Pip. Git fetch was used to pull code from https://github.com/SimonF92 on the machine. Streamlit was used via port 8501 to run Padplot via the dynamic IP of the AWS console. Users requesting access to Padplot were given remote access via this dynamic IP and port, allowing them to run Padplot from their own computers in-browser without having to install any python-based applications natively. No security steps were taken to secure Padplot, as a result, the machine was turned off after use and no user data was stored on the machine at any time. A fully working web address can be provided to readers on request although the website is not running at all times to prevent fees from Amazon. This web-application may become non-functional in 2022 due to inherent costs.

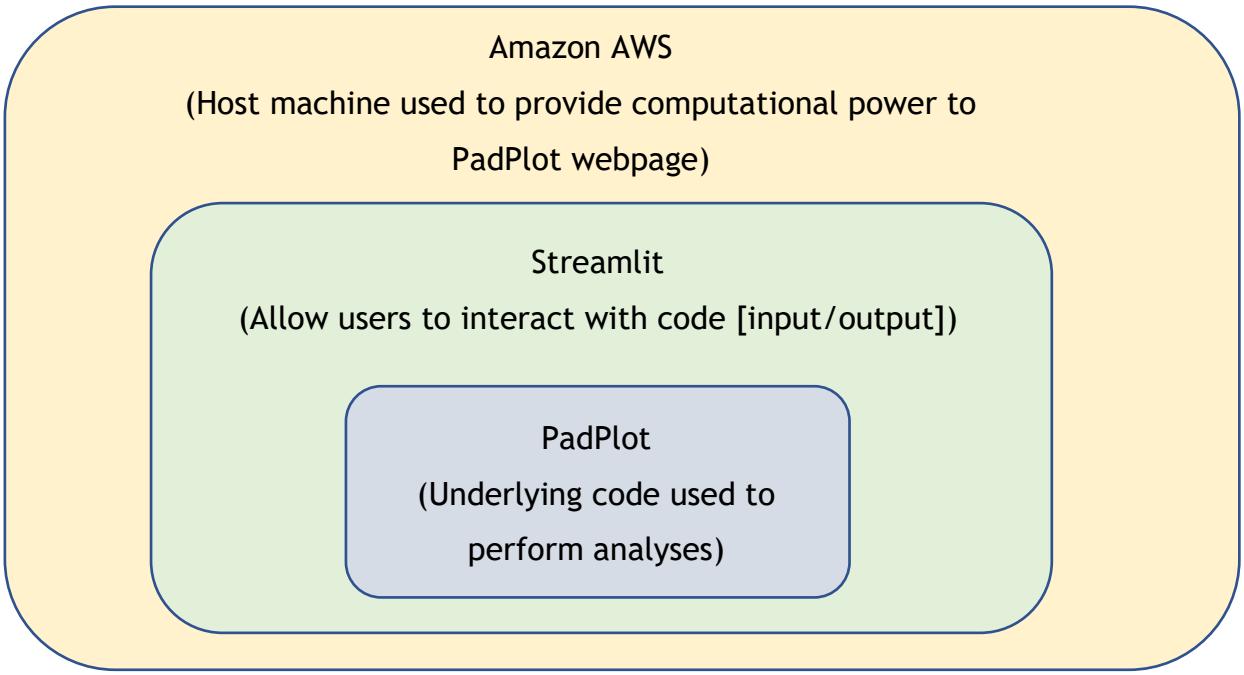


Figure 2.3-2 Padplot web-hosting architecture

2.4 Ingenuity Pathway Analysis

Ingenuity Pathway Analysis (Qiagen, Manchester, UK) Winter-2020 version was used. Datasets used in Ingenuity Pathway Analysis were loaded into the application in comma-separated-values 'csv' format, filtered on log2FoldChange of +1/-1 and adjusted p-value of (<0.05). Gene-sets within IPA consisted of independently curated, private gene sets, distinct from the gene-ontology based publicy curated gene-sets used in gene set enrichment analysis within the study.

2.5 Single Cell Data Mining

Human renal single cell data (10X Genomics) was mined from GEO accession GSM4145205. Features, matrix and barcode files were loaded into Python using Scanpy (Version 1.6.1). Cells expressing <200 genes, cells expressing >10% mitochondrial genes or genes which were expressed in <10 cells were removed from analysis. Counts were normalized by log transformation before nearest-neighbour clustering was performed (10 neighbours, 40 components) and these data then dimensionally reduced into 2 dimensions by UMAP for initial visualization. Leiden algorithm was then applied to this to generate clustering. UMOD expressing cells were then sub-filtered by selecting only cells with log-expression of the UMOD gene > 0.2, this enriched the dataset for theoretically primarly TAL cells, increasing the signal to noise ratio of the data. This subset of data was subsequently further reduced by reanalysing the components by nearest-neighbours (10 neighbours, 50 components) in order to stratify UMOD expressing cells from one another.

2.5.1 Automated EVOS Cell Detection

EVOS .tif images were imported using openCV (Version 4.4.0.40) and converted to 16bit greyscale before Clahe based normalisation was applied. Subsequent normalized images were read into the environment using Skimage (Version 0.17.2) and filtered using multiotsu thresholding (Skimage) with 5 brightness classes applied. Dead cells were characterized as those occupying a brightness threshold by Otsu (Skimage) algorithm of >2. If cells were highly confluent, a variable was created using morphology.white_tophat (Skimage) in order to increase accuracy of background detection. Distance transformation was prepared using ndi.transform_edit (Skimage)

on these thresholded images. Following this, local max features were detected using feature.peak_local (Skimage) with a min_distance of 1. Markers were then create from this matrix and this variable used to create segmented cells using segmentation.watershed. Total quantity of segmented features was used to give a count of total dead cells, which could automatically recognize and dissect clumped cell clusters. The ratio of total cell coverage was normalized to dead cells in order to give a marker of plate viability.

2.6 Plasmid Generation

2.6.1 Luria Bertani Medium, Glycerol Stocks and Agar Plates

All transformed and stock bacteria were cultured using Luria Bertani medium. All products were sterlised using the autoclave method before use.

Luria Bertani Medium	
Component	Mass or Volume for 1-litre
Tryptone	10g
dH_2O	950ml (+1000- volume NaOH)
Yeast Extract	5g
Sodium Chloride	5g
Sodium Hydroxide (to pH 8)	Variable
Ampicillin	100mg (100µg/ml)
Glycerol Stock	
Component	Mass or Volume for 10ml
Glycerol	5ml
dH_2O	5ml
LB-Agar XGal Plate	
Component	Mass or Volume for one plate
Lucia Bertani Medium from above	30ml
Agar Powder	0.3g
X-Gal (5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside)	0.6g

Table 2.6-1 Media and stock recipies

2.6.2 Insert Preparation and Ligation into Plasmid

Full length human UMOD RNA (500ng) was reverse transcribed to cDNA by the steps described in 2.1.5 and amplified by the steps described in 2.1.4.3. Cloning of this PCR product was performed using the pTargeT Mammalian Expression Vector System (Promega, Wisconsin, USA). PCR product was used at either a 1:1 or 3:1 ratio with pTtargeT Vector. Reaction vessels were incubated overnight at 4˚C to facilitate ligation of product into the vector.

Ligation of human UMOD into pTargeT System	
Reagent	Volume (µL)
T4 DNA Ligase	1
pTargeT Vector	1
PCR Product	1 or 3
Control Insert	2 (Absent from standard reaction)
T4 DNA Ligase	1
Nuclease-free-water	To 10µL

Table 2.6-2 Ligation reagents

2.6.3 Transformation of Competent Cells

All transformations were performed using aseptic technique. LB-Agar XGal Plates were equilibrated to room temperature prior to plating and JM109 High Efficiency Competent Cells, which produce at minimum 1×10^8cfu/µg DNA, were removed from storage at -80˚C and thawed on ice immediately prior to reaction. 50µL JM109 cells were combined with 2µL of ligation product and vessels were gently flicked to disperse the mixture before being incubated on ice for 20minutes. The cells were then heat-shocked for 50 seconds at 42 ˚C before being immediately returned to ice for 2minutes. 950µL room temperature LB media was then added to each reaction vessel. This mixture was then incubated for 1.5hours at 37˚C with shaking at 150RPM.

100µL of each transformation reaction was then plated onto duplicate LB-Agar XGal plates which were then incubated overnight at 37˚C. The following day, colonies were screened by eye for blue/white colour selection. White colonies indicate recombinant events and so were selected for further analysis by scraping these colonies with a sterile filtered pipette tip and splitting the scraped product between starter culture and PCR amplification. Starter cultures were prepared by incubating this contaminated pipette tip in 10ml LB-media overnight with shaking. The following day 1ml of overnight starter culture was added to ml glycerol stock and stored at -80˚C for future use. Insert orientation was verified by amplifying the plasmid product again the T7 promoter 5'-CATTATGCTGAGTGATATC-3'.

2.6.4 Plasmid Purification

1ml of amplified product from a single colony, clonally expanded through start culture was added to 999ml LB-media and incubated overnight at 200 RPM at 37˚C with gentle shaking (at 40rpm), prior to purification using the Qiagen Plasmid Midi Kit (Qiagen, Manchester, UK). The following morning product was purified using centrifugation at 6000rcf (g) for 15min at 4˚C and the supernatant discarded. Pelleted transformed bacterial cells were then resuspended in 0.3ml of Buffer P1. To this 0.3ml buffer P2 was added and the mixture then vigorously inverted multiple times to ensure homogenisiation. To the resulting mixture, 0.3ml chilled buffer P3 was added and this immediately mixed 5 times before being incubated on ice. The resulting solution was centrifuged at maximum speed (13000 x g (RCF))) for 10min. The supernatant (containing plasmid) of this centrifugation was then retained.

A QIAGEN-tip 20 was equilibrated using 1ml Buffer QBT, the solution was allowed to perfuse into the tip and emptied via gravity flow. Following equilibration, the supernatant from above was added to the column and allowed to empty via gravity flow. Following this, 2x2ml Buffer QC was allowed to flow through the column. Finally, the DNA retained in the column was eluted using Buffer QF. DNA was then precipitated from the eluent by adding 0.56ml per 0.8ml elution volume of room temperature isopropanol. The resulting solution was then immediately centrifuged by ultracentrifugation at 15000 x g (RCF) for 30min. The supernatant was carefully

removed and the resulting pellet was then washed with 1ml of 70% ethanol before being centrifuged again at 15000 x g (RCF) for 10min. The supernatant was again carefully removed and the resulting pellet air dried for 5-10min before the final DNA pellet was resuspended in 10ml nuclease free water. This resuspended solution of DNA was then characterized for concentration by Nanodrop Spectrophotometer using the same standards for quality as described prior.

2.7 Cell Culture

2.7.1 Media Preparation and General Methods

Cells were initially obtained by donation from PhD researcher Ahmad Mohammed A Alhajoj. All experiments were conducted under aseptic technique. HEK293 cells were cultured in Minimal Essential Medium (MEM) (Thermofisher, Paisley, UK), supplemented with 10% fetal-bovine-serum, 1% L-glutamate, 1% Sodium pyruvate and 1% Penicillin-Streptomycin (all Thermofisher, Paisley, UK). Cells were cultured in a Class-II Biosafety laminar flow hood (Thermofisher, Paisley, UK) and incubated at a constant 37˚C with 5% CO_2 . All cultures were performed using Corning Flasks of varying size (T25 (5ml), T75 (13ml) and T150 (20ml)) (Corning, Arizona AZ, USA). Where appropriate, cells were analysed for viability using light microscopy and counted using a hemocytometer (Sigma, Dorset, UK). Cell counting was completed by first incubating cells with 0.4 trypan blue solution (Sigma, Dorset, UK), each quadrant on the finder was used to count cells, and average count was taken as the mean of the four quadrants.

Cells were raised from -80˚C by rapid transference from cryostorage vials to flasks containing pre-warmed media. Media was replaced on growing cells every 48-72 hours, all solutions which were to come into contact with cells were first heated to 37˚C using a waterbath. Between uses, all solutions were stored at 4˚C. Upon confluency or prior to experimental necessity, cells were split and passaged by first removing media and then performing 3x gentle washes with sterile PBS with the last PBS wash aspirated leaving only cells. Upon completion of washes, cells were

incubated with 5ml 10% trypsin solution in PBS (Thermofisher, Paisley, UK) for 5minutes in an incubator to detach them from plate surface. This 5ml solution was mixed with 5ml of supplemented MEM to neutralize the trypsin. The resulting 10ml solution was centrifuged at 1000 x g (rcf) for 5minutes to create a pellet. Having discarded the supernatant, this pellet was resuspended in 1ml of supplemented media. If these cells were to be frozen, 100µL of sterile Dimethyl sulfoxide (DMSO) was added and mixed, following which the resulting solution was added to a cryostorage vial before being rapidly cooled to -80˚C using a MrFrosty Freezing Container (Thermofisher, Paisley, UK) overnight. After overnight cooling, cryostorage vials were transferred to liquid nitrogen for long term storage.

2.7.2 Transfection

Transfection was performed usingt PolyFect Transfection Reagent (Qiagen, Manchester, UK). Cells were seeded at $6e10^5$ density into 6-well tissue culture plates with 3ml supplemented MEM before being grown to 70% confluency. 2µg of purified DNA product from 2.4.4 was dissolved in TE buffer with basic MEM to a volume of 100µL. To this, 20µL PolyFect Transfection Reagent was added and solution mixed by pipetting. These samples were incubated for 10min at room temperature (21-23˚C). Media was aspirated from cells before 600µl supplemented MEM combined with the 120µL polyfect product using mixing was transferred to wells. The resulting plate was incubated overnight, following which this media:polyfect mixture was aspirated and replace with supplemented MEM.

2.7.3 MTT Assay

Confluent cells post-transfection were assessed for viability using MTT Assay Kit (Sigma, Dorset, UK). Cells seeded in 96-well TC plate wells were treated with 10µL MTT labelling reagent and incubated for 4hours at 37˚C (5% CO_2). Following this, 100µL Solubilisation solution was added to each well and the plate incubated overnight at 37˚C to permit the MTT reaction (5% CO_2). Spectrophotometrical absorbance was then assessed at 550nm wavelength using a PerkinElmer VICTOR plate reader (PerkinElmer, USA).

2.7.4 RNA and Protein Harvesting from Cells

Where appropriate, following media aspiration with subsequent 3x PBS washes before aspiration of the final PBS volume, cells were incubated in either Lysis Buffer (Qiagen, Manches, UK) or RIPA buffer (Thermofisher, Paisley, UK) for RNA and protein extraction respectively for 5 minutes in an incubator at 37˚C. Following this, the cells were scratched from the dish using sterile cell scraper and the resulting solutions harvest and stored at -80˚C (RNA) or -20˚C (protein) for future use.

2.8 Electrophoretic Mobility Shift Assay and Mass Spectrometry

2.8.1 Extraction of Nuclear Protein

Wild type HEK293 cells were trypsinised and pelleted as described in 2.5.1 with one confluent T75 producing a packed cell volume of 20µL with the supernatante discarded. From this, nuclear protein was extracted using the NEPER kit (Thermofisher, Paisley, UK). For packed cell volumes of 20µL, 200µL ice cold CER-I solution was added to this pellet and the resulting mixture was vortexed for 15 secondard prior to a 10 minutes incubation on ice. Following this, 11µL ice cold CER-II was added and the mixture vortexed for 5 seconds before a 1 minute incubation on ice. The mixture was then vortexed for 5 seconds again before being centrifuged at 13,000 x g (rcf) for 5 minutes. Supernatant from this spin was discarded before the resulting pellet was resuspended in 100µL ice cold NER solution with repeated vortexing over a 45 minute period between periods of incubation on ice. After this process, the tube was centrifuged at 13,000 x g (rcf) for 10 minutes. The supernatant, containing nuclear proteins was then stored at -80˚C until required.

2.8.2 Generation of the EMSA Product

Electrophoretic mobility shift assay was performed using LightShift Chemiluminescent kit (Thermofisher, Paisley, UK). The current GRCh38 human

reference genome was used to determine the 50-mer flanking sequences (+25bp, -24bp) of both rs4497081

(CCCAGGGGCCTATTGTGGGATG[C/G]GGGGAGCGGGGAGGGATAGCATTAGGA) and rs13333226
(CTGTTTGGGAAGAGGAGTCAATAT[T/C]CCTACAGCTGTGCTACCTCTTTGAC). Of these four sequences, both standard and 5-prime biotinylated were obtained from Eurofins (Eurofins Genomics, Ebersberg, Germany) along with their reverse and complement sequences giving a total of eight oligonucleotides. Both the biotinylated and un-biotinylated oligos were incubated with their respective reverse complement sequences for 30min at room temperature (21-23˚C) to product double stranded product. Following this, EMSA products were produced by combining reagents in a 1.5ml Eppendorf tube with gentle mixing and a 20min incubation at room temperature.

Generation of EMSA Product			
Reagent	Negative Control-A (No protein) (µL)	Experimental Assay (µL)	Negative Control-B (out-competition) (µL)
dH_2O	10	7	3
10X Binding Buffer	2	2	2
1µg/µL Poly (dIdC)	1	1	1
50% Glycerol	1	1	1
1% NP-40	1	1	1
1M KCl	1	1	1
100mM $MgCl_2$	1	1	1
200mM EDTA	1	1	1
Unlabelled Target Oligo (1pmol)	0	0	4
Protein Extract	0	3	3
Biotinylated Target Oligo (10fmol)	2	2	2

Table 2.8-1 Electrophertic Mobility Shift Standard Operating Procedure

2.8.3 Visualisation and Quantification

The products of EMSA reaction were visualized by combining 20µL of product with 5µL bromophenol blue dye and loading these products into pre-flushed wells of a NOVEX DNA retardation gel (Thermofisher, Paisley, UK) and running the system at 150V in 0.5% TBE (dH_2O) using a BioRad Powerpac 300 direct-current system for approximately 1-hour, or until the dye migrated to the end of the gel. This gel was then transferred onto a Biodyne™ B Nylon Membrane (Thermofisher, Paisley, UK) in a wet process using pre-cooled 0.5% TBE solution at 380mA for 30 minutes. The membrane was subsequently cross linked using a UV crosslinker equipped with 254nm bulbs, with the dye-side facing down for 10 minutes.

This crosslinked membrane was then blocked using 1X Blocking Buffer for 15 minutes with gentle shaking (40RPM) before being incubated with 20ml Blocking Buffer containing 66.7µL Stabilised Streptavidin-HRP Conjugate for 15 minutes with shaking. The membrane was then washed in 1X wash buffer four times, with 5 minutes shaking periods in each wash. Following this, the membrane was then incubated with 30ml Substrate Equilibration Buffer for 5 minutes with shaking. Lastly, the membrane was incubated face down in a mixed combination of 6ml Luminol Enhancer solution and 6ml Stable Peroxide Solution for 5 minutes. Following this, and within 20 minutes of this stage, the membrane was imaged using a Licor cDigit scanner (Licor Biosciences, Nebraska, USA).

Data were exported in zip format and unzipped into Image Studio Light (Licor Biosciences, Nebraska, USA). Experimental lanes of the membrane were analysed by densitometry, drawing equal sized squares by hand around both the upper and lower bands. The upper band represented oligonucleotides bound to nuclear protein whilst the lower band represented free oligonucleotides. Therefore, the ratio of upper to lower band was used to quantify binding affinity. Log ratios of these binding affinities

were then used to visualise relative binding affinity and perform statistical calculations.

2.9 Western Blotting and Co-immunoprecipitation

2.9.1 Protein Extraction and Quantification

Proteins were extracted from HEK-293 by performing an on-plate lysis of cells. 50µL RIPA buffer (Thermofisher, Paisley, UK) was added to 1x 10^6 washed cells and a 30min incubation performed at room temperature. Cells were then scraped using a blunted P1000 pipette tip and lysates harvested and store at -80˚C until further use.

Protein concentration was then determined using Pierce Gold Rapid BCA kit (Thermofisher, Paisley, UK). Bovine Serum Albumin (BSA) standards were first prepared by diluting 2000µg/ml BSA in nuclease-free-water serially, to concentrations of [2000,1500,1000,750,500,250,125,25,0] µg/ml. All BCA assay reactions were performed in wells of a 96-well TC plate. To each well, 196µL Reagent A was added, to which 4µL Reagent B was added, and then wells mixed by pipetting. To these wells, 20µL of RIPA lysates were added with mixing and these reactions incubated at room temperature for 5min. All reactions were performed in duplex, including the standard controls, with additional blank wells included. Following a 5min incubation, this plate was then analysed at 480nm wavelength using a PerkinElmer VICTOR plate reader (PerkinElmer, USA). Raw absorbance data from this analysis was obtained. From all wells, the blank control value was subtracted to normalize. Mean values were obtained from duplicates. The BSA ladder was used to generate a linear regression equation in the form $y = mx + c$, from which unknown sample concentrations were calculated. Samples were then diluted to standard concentrations in dH_2O.

2.9.2 Western Blotting

Samples diluted to stocks in 2.7.1 were mixed 1:4 with NuPAGE LDS-sample buffer and heated to 95˚C for 5min. All western blotting was performed using the

Invitrogen™ Novex™ mini system (Thermofisher, Paisley, UK). A total of 30ug protein lysate in LDS-sample buffer was loaded into the wells of a Bolt™ 4 to 12%, Bis-Tris, 1.0 mm, Mini Protein Gel (Thermofisher, Paisley, UK), with 3µL PageRuler Prestained Protein Ladder(Thermofisher, Paisley, UK) loaded in the rightmost well. NuPAGE running buffer (1X) (Thermofisher, Paisley, UK), was used as the running solvent with the system placed in ice while it ran. Proteins were electrophoresed using a BioRad Powerpac 300 direct-current system, initially at 100V for 10min before increasing the voltage to 120V until the dye had migrated to the bottom of the gel.

Prior to transfer, transfer buffer was prepared by combining 50ml 20X NuPAGE transfer buffer with 850ml dH2O and 100ml methanol. This solution was cooled to 4˚C following which either Powerblotter Membrane (Thermofisher, Paisley, UK) or PVDF membrane (Thermofisher, Paisley, UK) and filter sponges were activated in 1X transfer buffer for 10min. Following activation, the gel was paired with the membrane with continuous wetting and the system transferred using a Novex cage in a cold room at 4˚C at 120V for one hour.

Ensuring a continually wet membrane, this membrane was removed and blocked in 50:50 SEABLOCK (Thermofisher, Paisley, UK) in Tris-buffered-saline (TBS) (6.05 g Tris and 8.76 g NaCl in 800 mL of H2O) with rotary shaking at 40 RPM for 1 hour. Following this, the membrane was incubated at 4˚C overnight at 40 RPM in 1:800 rabbit anti-UMOD antibody (Abcam, Cambridge MA, USA) diluted in 1X TBS-Tween (TBST) (TBS with 1% Tween solution) 50:50 with SEABLOCK. The following day, the membrane was washed in TBST for 5min at 40 RPM, with this process repeated 3 times, each time discarding the wash. The membrane was then incubated at room temperature in a tinfoil wrapped vessel for 1 hour in 1:10,000 donkey Anti-Rabbit IgG H&L (Alexa Fluor® 488) (Abcam, Cambridge MA, USA) diluted in 1X TBST. Following this, the membrane was washed in a tinfoil wrapped vessel in TBST for 5min at 40 RPM, with this process repeated 3 times, each time discarding the wash. The membrane was then imaged face down on a Licor cLX system at high-quality in dual channel mode, detecting both 488nm and 647nm channels. After this, the membrane was incubated at 4˚C overnight at 40 RPM in 1:4000 mouse anti-Actin antibody (Abcam, Cambridge MA, USA) diluted in 1X TBS-Tween (TBST) (TBS with 1% Tween solution) 50:50 with

SEABLOCK. The following day, the membrane was washed in TBST for 5min at 40 RPM, with this process repeated 3 times, each time discarding the wash. The membrane was then incubated at room temperature in a tinfoil wrapped vessel for 1 hour in 1:10,000 donkey Anti-mouse IgG H&L (Alexa Fluor® 647) (Abcam, Cambridge MA, USA) diluted in 1X TBST. Following this, the membrane was washed in a tinfoil wrapped vessel in TBST for 5min at 40 RPM, with this process repeated 3 times, each time discarding the wash. The membrane was then imaged face down on a Licor cLX system at high-quality in dual channel mode, detecting both 488nm and 647nm channels.

Together these steps produced an image of firstly raw UMOD densitometry and secondly UMOD and Actin densitometry, with Actin used as a housekeeper. Data were exported in zip format and unzipped into Image Studio Light (Licor Biosciences, Nebraska, USA). Experimental lanes of the membrane were analysed by densitometry, drawing equal sized squares around signals. UMOD densitometry was normalized to loading by dividing it by the Actin densitometry values.

2.10 *In-vivo* Experimentation

2.10.1 General Methods

All animal procedures were approved by the Home Office according to the Animals (Scientific Procedures) Act (1986) (Project Licences 60/9021 and PP0895181) and followed ARRIVE guidelines and subject to local ethics approval. Sv129-WT and Sv129-*Umod*-/- , *Umod* +/- and *Umod* +/+ mice were housed with an ambient temperature of 21°C±2 on a 12-hour light/dark cycle and given ad libitum access to standard chow and tap water. One male *Umod* +/- and one female *Umod* +/- were cage mated and litters reared with the mother until weaning. At 6 weeks, pups were genotyped using tissue derived from ear notching using an ear punch.

2.10.2 Pilot Study Design

At 12 weeks male *Umod*-/- , *Umod* +/- and *Umod* +/+ mice were randomly assigned to one of two groups where they were separated into paired housing and given ad libitum access to standard chow and either tap water or 2% sodium chloride in tap water. Animals were monitored to ensure their tolerance to sodium both by weighing

them and observing their behaviour. Animals had their baseline blood pressures assessed by tail cuff plethysmography for training in the week before beginning the procedure and started on either tap water or sodium chloride solution for the next 6 weeks. Blood pressure was taken every Tuesday morning at 9am by tail cuff plethysmography. This produced a total of 7 measurements- one training measurement and 6 actual measurements. At week 3 and week 6 of being on this procedure, the animals were housed in metabolic cages for 24 hours. After the final blood pressure measurement, the animals were sacrificed and tissues harvested.

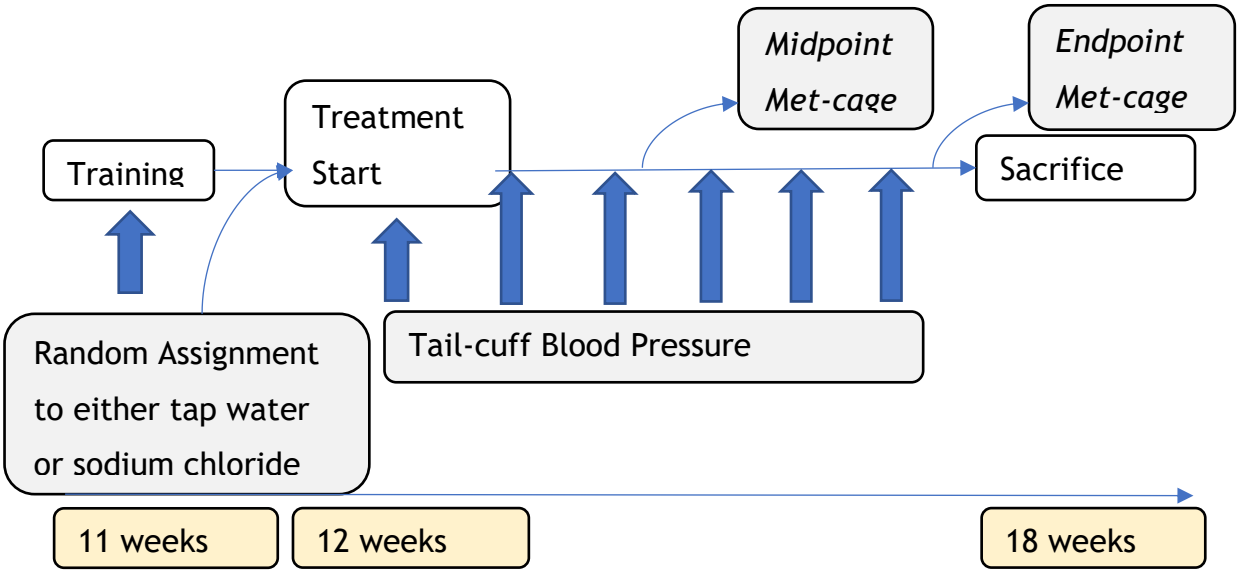


Figure 2.10-1 Experimental design of the UMOD mouse study

2.10.3 Tail Cuff Plethysmography

A BP-2000 Series II machine (Therassay/ Visitec Systems/ Bioseb) in mouse configuration was used to measure all tail-cuff blood pressure readings, preheated to 37˚C to vasodilate the tail prior to study. Animals were taken in their home cages to a dedicated, quiet room. A cardboard tube was used to pick up the mice non-aversively before they were placed in a free moving, closed vessel on the BP-2000 system and their tails had a cuff placed over the base. These mice were then secured in place by gently taping down the end of their tails. A total of 5 induction and 15

actual measurements were taking with 30 second intervals and a maximum pressure of 220mmHg. Data were exported in their raw format.

Data was firstly assessed for 0 values which were discarded. From non-zero values, the 5th and 95th percentiles were calculated and values lying out with this range discarded as outlier. Using percentiles and not Gaussian distribution boundaries accounted for the high incidence of positive outliers. Mean values for each set of 15 measurements were then calculated and these mean values used as values for any subsequent analysis.

2.10.4 Metabolic Caging

One mouse was assigned to a single metabolic cage with ad-libitum access to powdered chow and 100ml of their assigned drinking water. Prior to the first 24-hour metabolic cage period, a 4-hour acclimatization period was given to the animals to allow them to acclimatize to the cage. Over the 24-hour period, all urine was collected and at the end of the period, drinking water was assessed for volume. Per animal, volume of liquid consumed, and volume of urine excreted were obtained as metrics, whilst the urine itself was stored at -80˚C until further use. All transfers of animals between home cage and metabolic cage was performed using a tube under non-aversive handling guidelines.

2.10.5 Sacrificing

After 6 weeks of treatment, at 18 weeks of age, animals were anaesthetized with 5% isoflurane in oxygen mix in an induction chamber until non-responsive to pressure applied forcefully to the tip of the tail with a pair of forceps. These animals were then weighed before being placed on an operating table in a supine position and attached to a face mask to maintain general anesthesia. Blunt forceps were then used to raise the skin above the diaphragm before the skin and fur covering the area was cut off, leaving muscle tissue overlaying the chest exposed. A blunt incision was made into the peritoneum before the tissue was cut away on either side of the Xiphoid process of the sternum, exposing the diaphragm. A second blunt incision was made into the diaphragm before this was cut away. Following this, the rib cage 5mm either

side of the sternum was cut, exposing the organs of the chest. A 23G needle fitted to 2ml syringe (Sigma-Aldrich, Manchester, UK) was inserted into the apex of the heart, targeting the left ventricle (LV). The animal was then ex-sanguinated by removing blood from the LV as the chamber filled. Following exsanguination, watchmakers forceps were used to lift and hold the aorta from the posterior wall of the thoracic cavity before this was cut away from the base of the spine upward toward the heart until it was cut away completely at the aortic arch. One half of this aortic section was placed in 10% paraformaldehyde solution with the other half snap frozen using liquid nitrogen. The heart was then removed from the animal and connective tissue stripped from it before it was weighed. The heart was then cut laterally (horizontally if standing up) into equally sized pieces half which was placed in 10% paraformaldehyde solution with the other half snap frozen using liquid nitrogen. With the aorta and heart removed, the abdomen wall was then cut to expose the organs of the abdomen. Both the right and left kidney were removed, weighed and then cut in half laterally. Three of the four sections were then snap frozen in liquid nitrogen with the last section placed in 10% paraformaldehyde solution (PFA). Lastly the liver was removed and weighed, with a 100mg section dissected away and snap frozen in liquid nitrogen.

All snap frozen components were stored at -80˚C until further use. All PFA stored components were kept in 10% PFA at room temperature for 24 hours before being transferred to 70% ethanol and stored at 4˚C in the fridge.

2.11 Statistical Analyses

Data was analysed in Python 3.8 using the Stats and Scikit-learn libraries. Continuous data was tested for normal distribution prior to analysis. Appropriate tests were then selected based on the required analysis. For comparisons between two groups, students t-test was applied, with Welch's correction applied to non-normally distributed data. For comparisons between multiple groups, one-way ANOVA was used with Tukey's HSD post-hoc. For linear regression analysis, two-sided p-values were calculated against a hypothesis test whose null hypothesis was that the slope was zero. For comparisons involving multiple testing, Benjamini-Hochberg correction

was applied. 95% confidence interval values were calculated and represented alongside mean values where appropriate. For all data, $p<0.05$ was considered the lowest threshold of statistical significance. For data involving more than one significance test, multiple testing correction was applied.

3 Chapter 3 The UMOD Promoter and rs4997081

3.1 Introduction

With the Padmanabhan et al study discovering the association of the UMOD gene with blood pressure and Graham et al & Trudu et al confirming this association *in-vitro* and *in-vivo* as acting via NKCC2 centred natriuresis, focus moved toward proving genetic functional causation (Padmanabhan *et al.*, 2013; Graham *et al*, 2014). Specifically, 'functional causation' attributed to base and not haplotype would be defined by empirical data showing that the specific base change at a specific variant on the locus resulted in a difference in UMOD transcription between alleles which associated with differences in blood pressure. Interestingly, causality at the discovery variant, rs13333226, has not been shown to date; rather it was shown that the loci containing rs13333226 is causative. Due to linkage disequilibrium (LD) resulting in strong or perfect correlation between rs13333226 and variants in proximity, effectively any variant in LD may be causative. Previous data assessing a range of variants in LD with rs13333226 firstly showed that rs13333226 assessed by luciferase assay, does not produce a difference in luminosity between alleles. Significantly however, this data indicated that a variant in near perfect LD with this variant, rs4997081, may be functional due to it displaying increased promoter activity in the risk allele (Graham, unpublished). Specifically, Graham et al, showed that, in this series of experiments on constructs containing the human UMOD promoter, with site directed mutagenesis to specify alleles at rs133333226 and rs4997081, a significant ($p<0.0001$) increase in luminosity was observed at the risk allele versus the protective allele. They performed these experiments across several relevant cell lines, HeLa, NRK, HEK293 and HK2, observing the same finding across each cell.

It is possible that the variant array used in the initial GWAS in 2010 did not include rs4997081 as a SNP, and furthermore that the UMOD promoter locus was not annotated to the extent which allowed rs4997081 to be imputed, therefore it was not included in the initial analysis. Based on the combination of our previous luciferase data and our finding that rs4997081 was masked in the initial analysis, we believe at present, that preliminary indications suggest that rs4997081 is most likely

a causative variant. Critically, rs13333226 and rs4997081 are not in perfect LD, rather they are in 91.9% linkage (National Cancer Institute, Division of Cancer Epidemiology and Genetics). Currently, UMOD precision medicine studies genotype rs13333226 for analysis, therefore, an error rate of up to 8.1% may exist between characterised and true causal variants; possibly leading to decreases in accuracy of analysis.

Showing functionality by luciferase assay between alleles at target variants provides important preliminary data. However, luciferase assay design uses reporter constructs in a system isolated from underlying human biology. As a result of this, the specifics of gene transcription cannot be interrogated to provide enhanced understanding of causality at the biochemical level. Electrophoretic mobility shift assay conversely assesses the binding affinity of eukaryotic transcriptional apparatus to target oligonucleotides, and can be modified to stratify binding by genotype. The significant benefit of this approach is that we believe a magnetic based isolation system can be developed in order to capture and characterise the transcription factors binding at the target locus. The benefits of this design therefore are two-fold, in that we believe we can use this to determine causation between rs13333226 and rs4997081, and also show the mechanism of this function. This in turn is likely to further enhance potential precision medicine applications downstream.

As we rationalise it is possible that further causative variants may exist cryptically in non-annotated regions at the locus, the work in this chapter initially intends to examine the human promoter for novel variants by targeted resequencing stratified against UMOD mRNA levels in our human renal samples. Following this, we intend to design, optimise and execute a protocol for determining the differential binding of transcription factors to our variant-containing oligonucleotides and building on this, to isolate and characterise these proteins by Mass Spectroscopy. For these experiments, we elected to use the HK2 cell line, which is an immortalised proximal epithelial cell line (Ryan *et al.*, 1994). HK2 cells, while not exact physiological matches for TAL cells, are the most appropriate cell model of choice as they remain highly representative of renal tubular cells and are a well-researched model for drug induced nephrotoxicity due to their broad expression of renal ion transporters (Jenkinson *et al.*, 2012). Alongside this they are known to exhibit stable properties

across passage, making them good choices for long term studies on renal cells, where material will be harvested longitudinally (Handl *et al.*, 2020).

Finally, UMOD has been shown repeatedly to associate with TNF-α across various mechanisms, including where blockade was shown to inhibit functional decline in UMOD mutant (C147W/+) mice and also where it has been shown to stimulate secretion of TNF-α in monocytes (Su *et al.*, 1997; Johnson *et al.*, 2017b). Furthermore, in wild-type mice, intraperitoneal injection of polyomerised UMOD was shown to produce this effect *in-vivo,* suggesting a relationship between UMOD and TNF-α (Immler *et al.*, 2020b). Graham et al in their publication relating Umod to blood pressure in mice also show an effect of Umod on TNF-α, where they suggest basolateral UMOD acts as a TNF-α sink. In early 2021, we became aware of highly significant, unpublished data implicating TNF-α treatment in the context of UMOD expression. This novel data showed TNF-α treatment associated with increased levels of UMOD expression in human cells, which could be reversed using Etanercept (specifically at rs4997081). The researchers involved with these experiments however, based their assays on Taqman qRT-PCR, thus our newly optimised EMSA protocol had the potential to corroborate and enhance their findings. Based on the pre-existing and annotated relationship between UMOD and TNF-α alongside this new and highly encouraging data, we pivoted to specifically focus on TNF-α, with our fundamental aim being to uncover targetable pathways in the expression of the UMOD gene in the context of blood pressure regulation.

3.2 Aims

- To examine our bank of human DNA/RNA samples genotypes for rs13333226 and rs4997081 for expression of the UMOD Gene

- To assess by NGS the promoter of low and high UMOD expressers for novel variants which differ between groups

- To investigate by EMSA, both rs13333226 and rs4997081 risk and protective alleles, and any further discovered variants for differential binding affinity to nuclear protein lysates

- To design and execute a protocol for enrichment of bound transcription factors

- To use these data to characterise the role of TNF-α in UMOD expression and build on the significant new data we were made aware of

3.3 Methods

3.3.1 Magnetic Purification of EMSA Product

Magnetic purification of products prepared from EMSA were purified using a protocol based on the m-280 Streptavidin Dynabead system (Thermofisher, Paisley, UK). Pooled EMSA reactions totalling 100µL were used in conjunction with 20µL M-280 Streptavidin Dynabeads. Prior to purification, these beads were resuspended in the original vial by vortexing, following which 20µL beads were transferred to a new 1.5ml Eppendorf which was placed on a Dynamag-Spin magnet system for 2min. With the tube still attached to the magnet, the supernatant was discarded. The tube was then removed from the magnet and the string of beads which had migrated toward the magnet were resuspended in 1X B&W buffer (Thermofisher, Paisley, UK). Resuspended beads were then placed back on the magnet for 2min and the process repeated 3x times in order to wash the beads, each time using 1X B&W buffer as the resuspension media. The last wash was resuspended in 40µL 1X B&W buffer. 100µL EMSA product was then added to this 40µL of washed beads and this volume incubated for 15min at room temperature whilst under gentle vertical rotation to ensure thorough mixing (10 RPM). Following this, the product of this reaction was placed on the magnetic rack for 3min to allow migration of beads onto the magnet. The supernatant was removed and the product resuspended in 100µL 1X B&W buffer. This wash step was repeated 3 times using 2min magnetic incubations before the beads from the last wash were resuspended in 100µL nuclease free water. The biotin-streptavidin bond was then broken using a 5min incubation at 65°C and these products submitted in full for mass spectrometry.

3.3.2 Mass Spectrometry

A volume of 30µL purified sample from 2.6.4 was mixed with 200µL UA solution (8M Urea (Sigma, Dorset, UK) in 0.1M Tris/HCl, pH 8.5) inside an Ultracel YM-10 (Millipore) vessel and the unit centrifuged at 14,000 x g (RCF) for 40min. Following this, an additional 200µL UA solution was added to the column and the column centrifuged again at 14,000 x g (RCF) for 40min, after which flow through was discarded. Subsequently, 100µL IAA (0.05M iodoacetamide in UA) was added to the column which

was then mixed at 600 RPM in a Thermo-mixer (Thermofisher, Paisley, UK) for 1min before a 5min incubation. The unit was then centrifuged for 30min at 14,000 x g (RCF). 100µL UB (8M Urea (Sigma, Dorset, UK) in 0.1M Tris/HCl, pH 8.0) was then added to the filter and this centrifuged for 40min at 14,000 x g (RCF), with this step repeated twice. 40µL of UB with Lys C (enzyme to protein ratio 1:50) was added to the column and mixed for 1min at 600 RPM using a thermo-mixer. The column was then incubated in a wet chamber overnight at room temperature. Filter units from these columns were then transferred to new column with 120µL ABC (0.05 M NH4HCO3) in water with Trypsin (enzyme to protein ration 1:100) added and mixed at 600 rpm in thermo-mixer for 1 min. These units were then incubated at 4 hours at room temperature. Filter units were then centrifuged at 14000 x g (RCF) for 40min. 50 ul of 10% ACN was then added and the units centrifuged at 14,000 x g for 20 min. Lastly, the filter unit was acidified with CF3COOH and dried in a 96-well plate with vacuum centrifuged. These dried peptides were solubilized in 20 µL 5 % acetonitrile with 0.5 % formic acid using the nanoflow uHPLC system (Thermofisher, Paisley, UK). Peptide ions were detected by electrospray ionization mass spectrometry (MS/MS) with an Orbitrap Elite MS (Thermofisher, Paisley, UK). 5µL of sample was desalted and concentrated for 10min on a trap column with 1% acetonitrile and 0.1% formic acid. Peptides were separated on a Pepmap C18 phase column (50x75cm, 3µm particle size, 100A pore size). The Orbitrap Elite acquired full-scan MS in the range 300 to 2000 m/z for a high-resolution precursor scan at 60,000 RP (at 400 m/z), while simultaneously acquiring up to the top 15 precursors which were isolated at 0.7 m/z width and subjected to CID fragmentation (35 % NCE) in the linear ion trap using rapid scan mode. Singly charged ions were excluded from selection, while selected precursors are added to a dynamic exclusion list for 30s. Proteins were assigned identifications using the Mascot Search Engine (v 2.6.2, Matrix Science) and further identified using the Swissprot database. A mass tolerance of 10ppm was used with 0.3 Da MS/MS matching.

3.3.3 TNF-α Treatment of HK2 Cells

HK2 cells were cultured as per 2.5, in complete MEM, in T75 flasks. When 70% confluent, media was aspirated and replaced with media containing 20ng/ml human

recombinant TNF-α (Thermofisher, Manchester, UK) or 20ng/ml TNF-α supplemented with 10mg/ml Etanercept (Thermofisher, Manchester, UK) for prescribed timepoints. Upon completion of timepoints, the media was immediately aspirated and cells were washed with sterile PBS before trypsin-based harvesting as per 2.5.1. Pelleted cells were then harvested for nuclear protein as per 2.6.1. The culturing of HK2 cells specific to this experiment was performed in full by Dr Lesley Graham.

3.3.4 Analysis of Electrophoretic Mobility Shift Assay Images

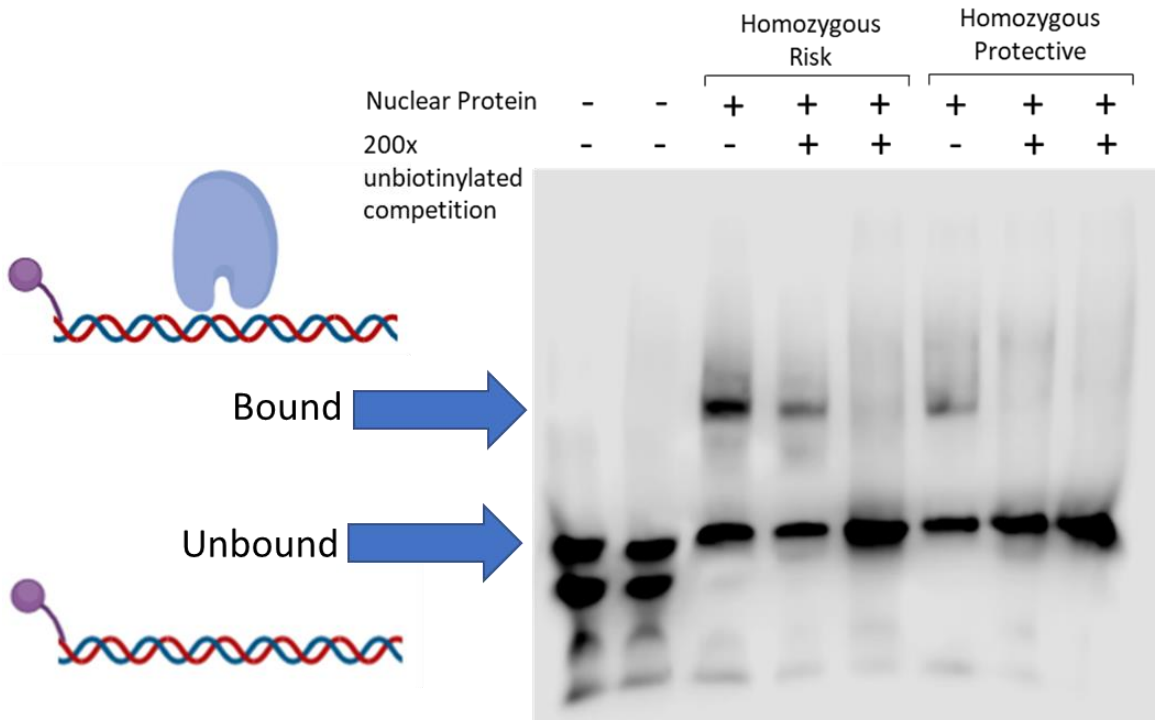


Figure 3.3-1 Methods of electrophoretic mobility shift assay analysis.

Electrophoretic Mobility Shift Assays quantify binding affinity through the generation of a fragment of oligonucleotide which is non-covalently associated with a transcription factor. These migrate more slowly during electrophoresis, producing a signal of both bound (upper) and unbound (lower) biotinylated-oligonulceotide. Through stoichiometry using the ratio of bound:unbound densitometric signal, binding affinity can be inferred. As this metric is produced as a ratio, the log of this ratio is used for downstream analysis.

3.4 Results

3.4.1 Analysis of Human Renal Samples Stratified by Genotype

The human samples we assessed were well randomised for sex, medication use, BMI and age, with no differences between groups observed (Figure 3.4-1)(Figure 3.4-2). Analysis of the systolic blood pressures of participants stratified by genotype at rs13333226 and rs4997082 were significant, with risk genotype at both alleles having significantly greater SBP than protective genotype. The confidence intervals of the differences to protective genotypes were (0.747 15.554, 95%CI) mmHg in the case of rs13333226 and (1.098 15.514, 95%CI) mmHg in the case of rs4997081 (Figure 3.4-3). These differences, when stratified by sex, were only maintained in female participants (Figure 3.4-4). We stress that only a single blood pressure measurement was taken, thus we will not interpret these findings further. Our reasons for this exclusion are discussed to greater detail in 3.5.

A total of 76 participants were genotyped for rs13333226 and rs4997081. By Sanger sequencing, at rs13333226, 52 were homozygous risk (AA), 21 were heterozygous (AG) and 3 were homozygous protective (GG). At rs4997081, by Sanger sequencing, 50 were homozygous risk (GG), 21 were heterozygous (GC) and 3 were homozygous protective (CC). A total of 2 (2.6% of) participants did not have matched genotypes between rs13333226 and rs4997081. Taqman qRT-PCR for UMOD expression on 84 samples indicated the rs13333226 genotype did not associate with differences in UMOD expression between alleles, though homozygous risk individuals did trend toward increased expression (Figure 3.4-5)(A/B).In the context of rs4997081 there were no differences in UMOD expression between alleles, though homozygous risk individuals did again trend toward increased expression (Figure 3.4-5)(C/D).

Analysis of the data suggested that the use of antihypertensive medication may associate with differences in UMOD expression. When groups were matched for the haplotype at rs13333226/rs4997081, individuals prescribed amlodipine had significantly lower ($p<0.01$) expression of UMOD versus unprescribed counterparts (Figure 3.4-6)(A) with the same observation made the context of Bendroflumethiazide use ($p<0.001$) (Figure 3.4-6)(B). When individuals prescribed

these medications were removed, homozygous risk individuals at rs4997081 had significantly greater ($p<0.05$) expression of UMOD versus their wild type counterparts of fold change between (1.189, 9.38, 95%CI), but this difference was not reflected at rs133333226, although showed a strong trend (Figure 3.4-6)(D/E).

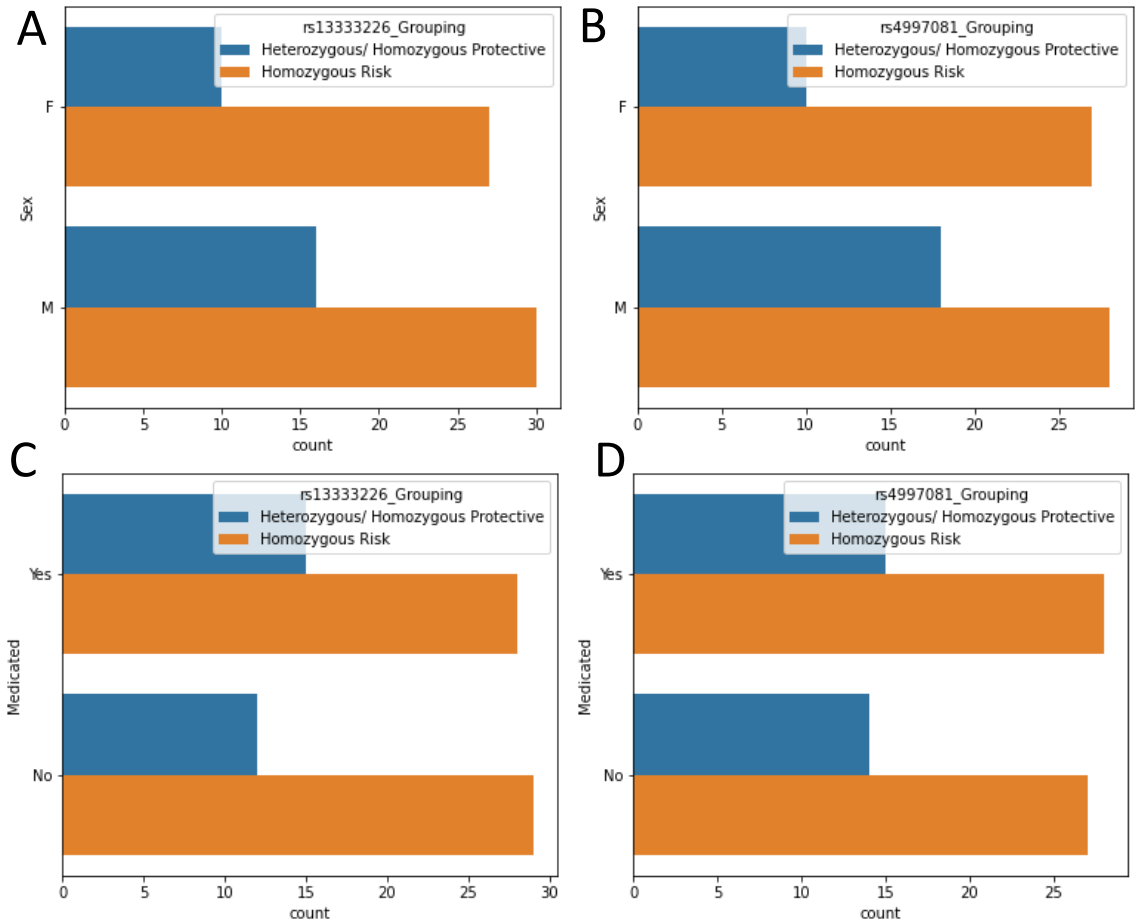


Figure 3.4-1 Randomisation of Sex (Male/female) and medication use (Yes/no), between rs13333226 (1A, C) and rs4997081 (1B, D) genotypes. At rs13333226, Homozygous 'protective' (GG (forward strand)) and heterozygous (GA (forward strand)) alleles were combined. Homozyous 'risk' consists of the remaining AA genotypes. At rs4997081, homozygous 'protective' (CC (forward strand)) and heterozygous (CG (forward strand)) alleles were combined. Homozyous 'risk' consists of the remaining GG genotypes. Non-significant, Pearsons chi-squared.

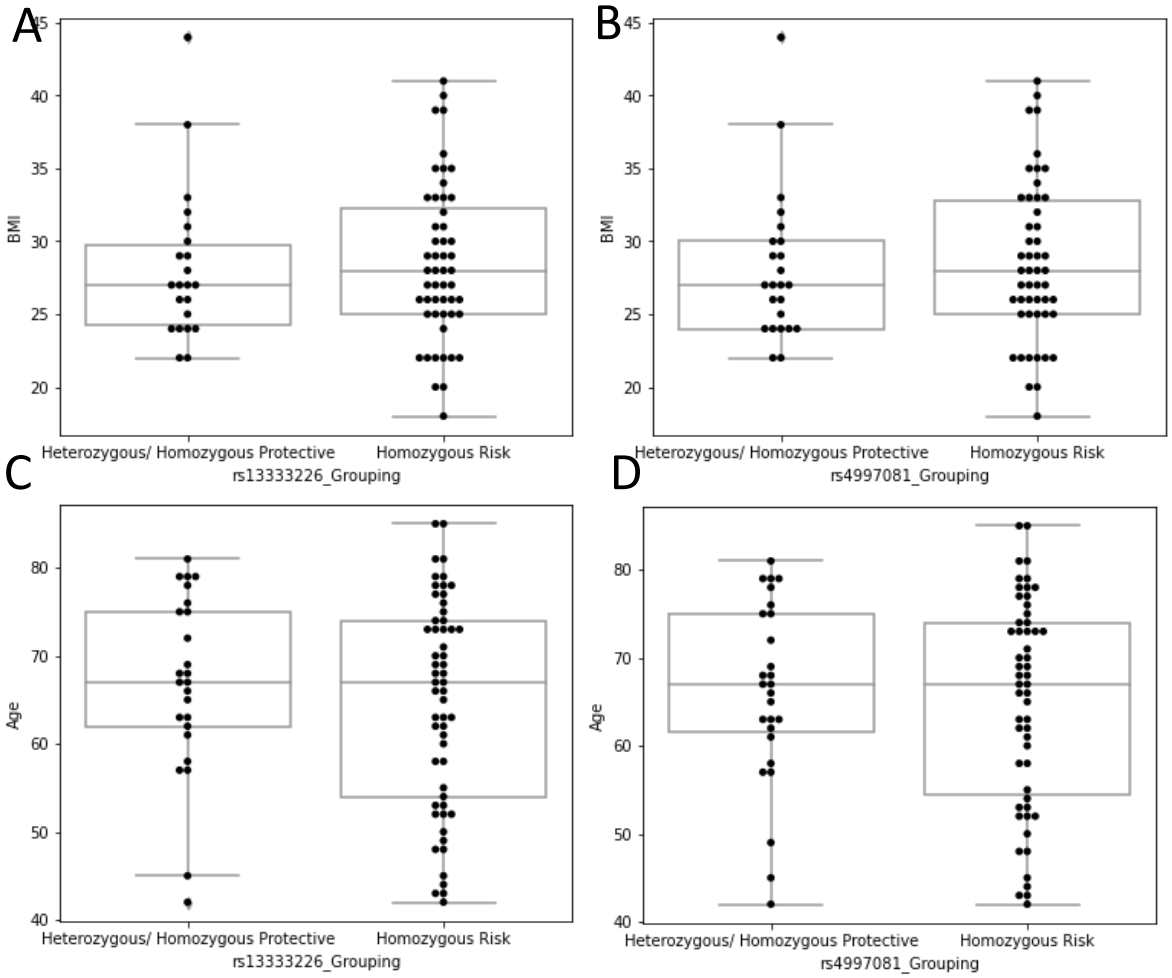


Figure 3.4-2 Parametric randomization of continuous variables.

Randomisation of BMI (AU) and age (Years), between rs13333226 (1A, C) and rs4997081 (1B, D) genotypes. Grouping by genotype as described in Figure 1. Non-significant, students unpaired t-test.

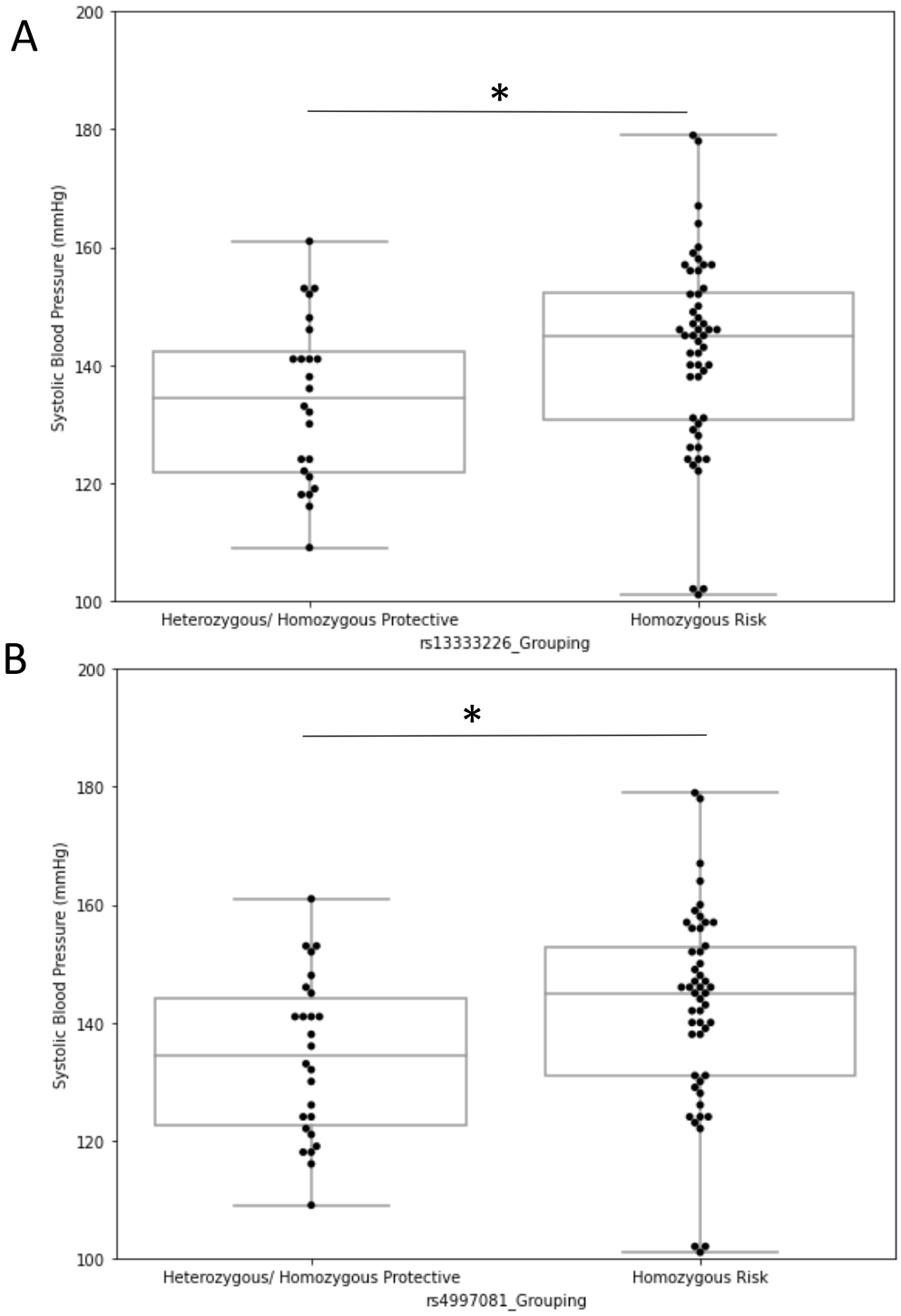


Figure 3.4-3. Stratification of SBP by genotype

A) Distribution of systolic blood pressures (mmHg) by rs13333226 genotype. B) Distribution of systolic blood pressures (mmHg) by rs4997081 genotype. Grouping by genotype as described in Figure 1. Students unpaired ttest, * $p<0.05$

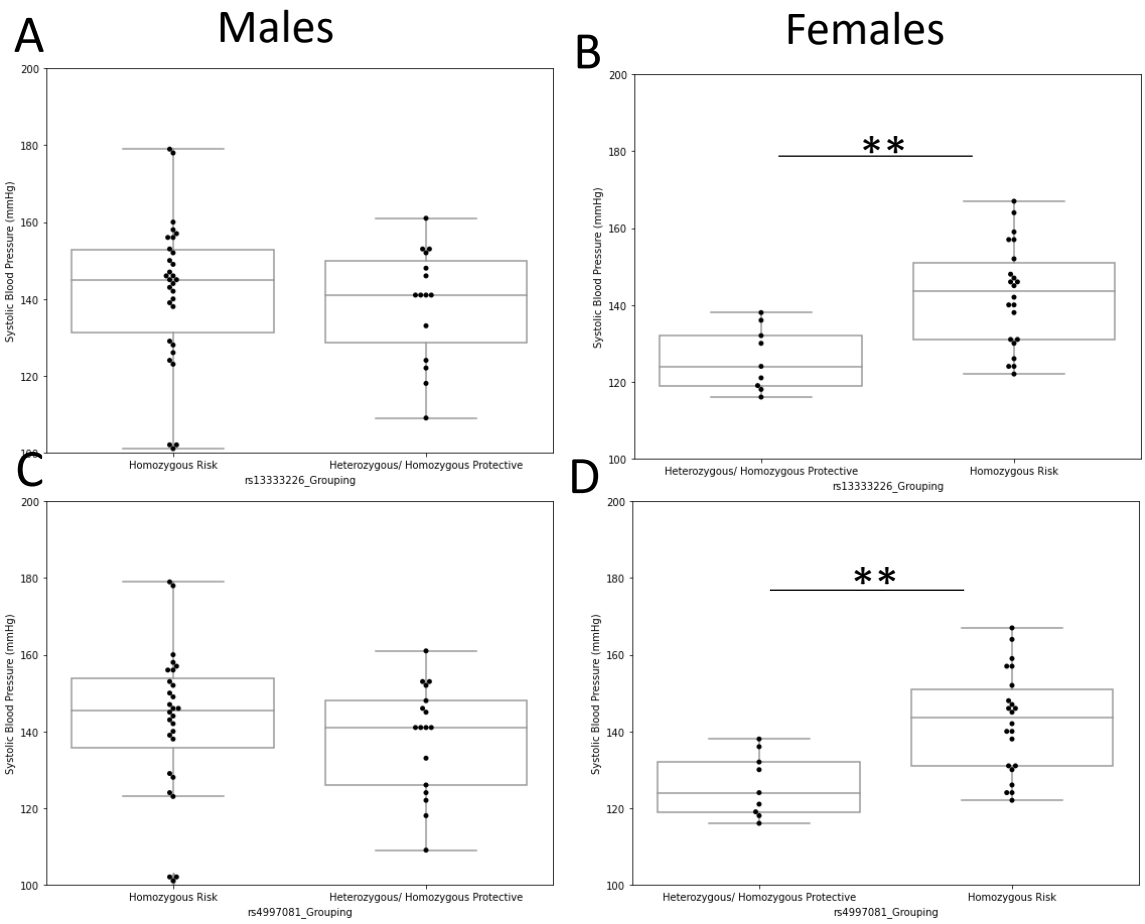


Figure 3.4-4 Stratification of SBP by sex by genotype.

A) & B) Distribution of systolic blood pressures (mmHg) by rs13333226 genotype by sex. C) & D) Distribution of systolic blood pressures (mmHg) by rs4997081 genotype by sex. Grouping by genotype as described in Figure 1. Students unpaired t-test, ** $p<0.01$

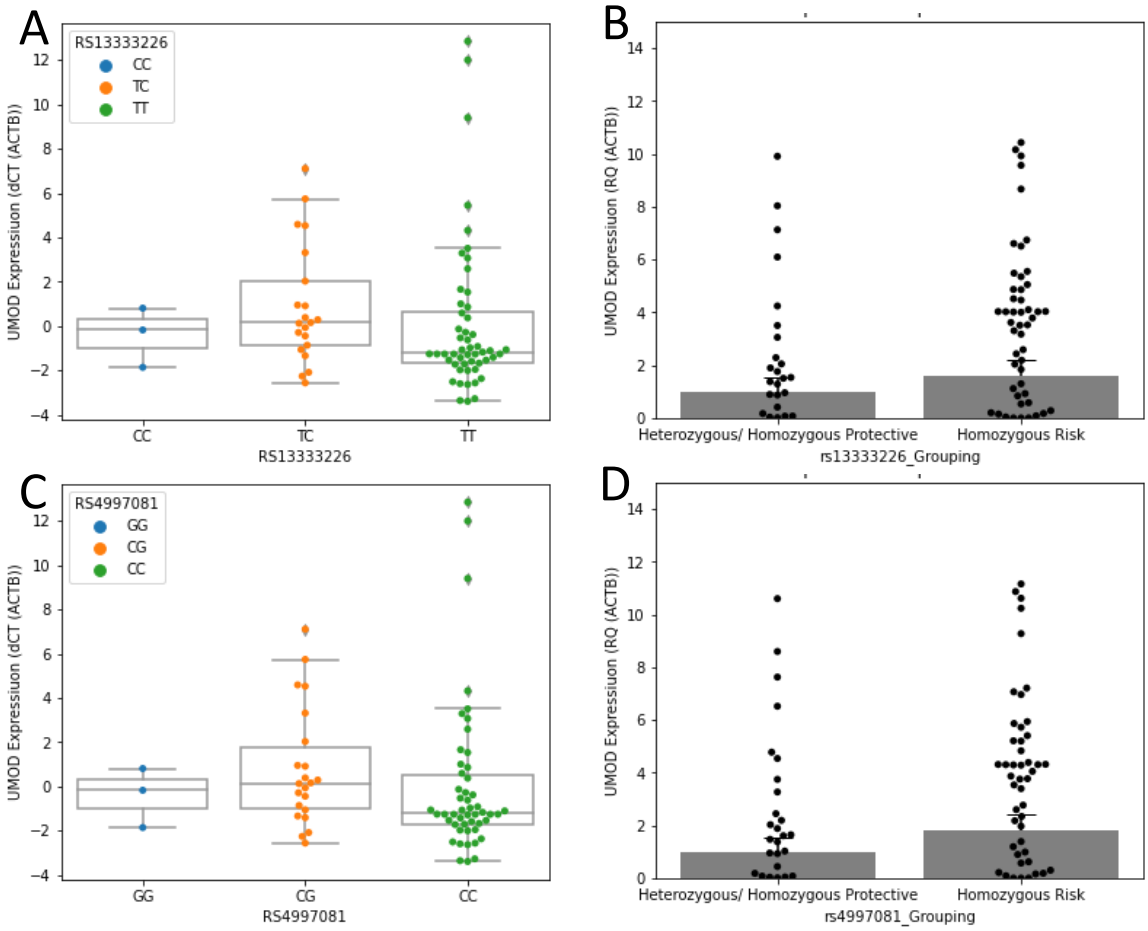


Figure 3.4-5 UMOD expression by genotype at rs13333226 and rs4997081

A) Umod expression (dCT vs ACTB) by rs13333226 genotype (CC/ CT/ TT), genotype shown as actual base change (reverse strand). B) Relative quantification of umod expression (RQ vs ACTB) by grouped rs13333226 genotype, grouping as described in Figure 1. C) Umod expression (dCT vs ACTB) by rs4997081 genotype (GG/ GC/ CC), genotype shown as actual base change (reverse strand). D) Relative quantification of umod expression (RQ vs ACTB) by grouped rs13333226 genotype, grouping as described in Figure 1. Non-significant, students unpaired t-test.

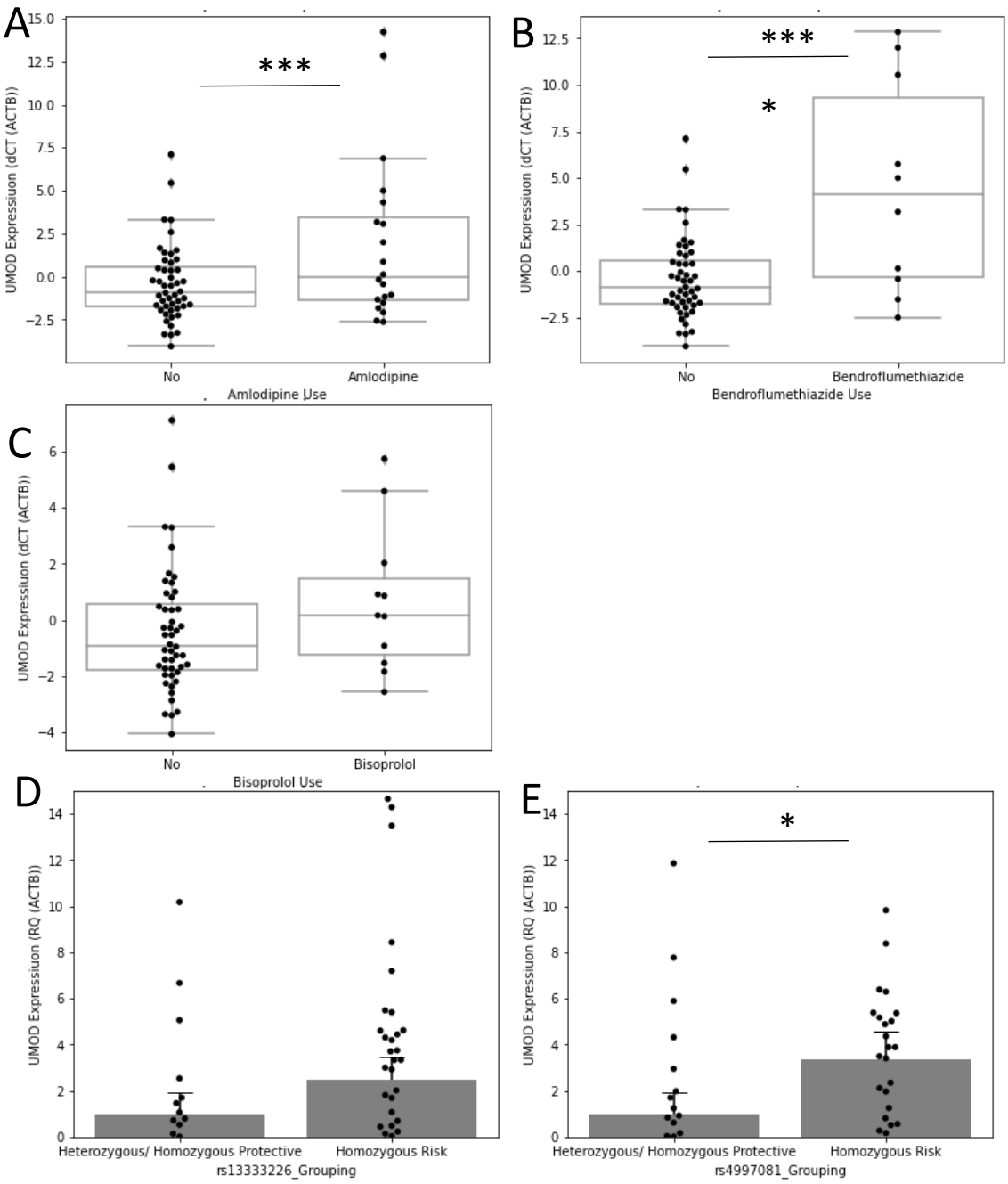


Figure 3.4-6 Stratification of UMOD expression based on medication

A) Expression of UMOD (dCT vs ACTB) between individuals not-using, or using amlodipine. B) Expression of UMOD (dCT vs ACTB) between individuals not-using, or using bendroflumethiazide. C) Expression of UMOD (dCT vs ACTB) between individuals not-using, or using bisoprolol. D)/E) expression of UMOD (RQ vs ACTB) between rs13333226/rs4997081 genotype when individuals using antihypertensive medication were removed from analysis. * $p<0.05$, ***$p<0.001$, **** $p<1e-4$

3.4.2 Next Generation Sequencing of the UMOD Promoter

The UMOD gene was successfully resequenced from -4000bp upstream of the TSS to intron-5 using PCR amplified amplicons of 1800bp (as anticipated from the primer design) from a subset of high and low UMOD expressers with the amplicons amplifying cleanly, only producing one band, indicating good primer specificity. (Figure 3.4-7)(A/B).

A number of variants were detected in reference to GChR38 in all samples, though none were significantly enriched in either high or low UMOD expressers (Figure 3.4-7)(C). On average, 126 variants were annotated differently the reference genome (95%CI 14.52, 17.48) and there was no difference between variant detection number between groups. Of detected variants, the majority were assigned dbSNP identities, though four detected variants could be considered specific to the local west-Scotland haplogroup and as such were not catalogued within dbSNP (Figure 3.4-8). None of the detected variants, by text mining were associated with blood pressure.

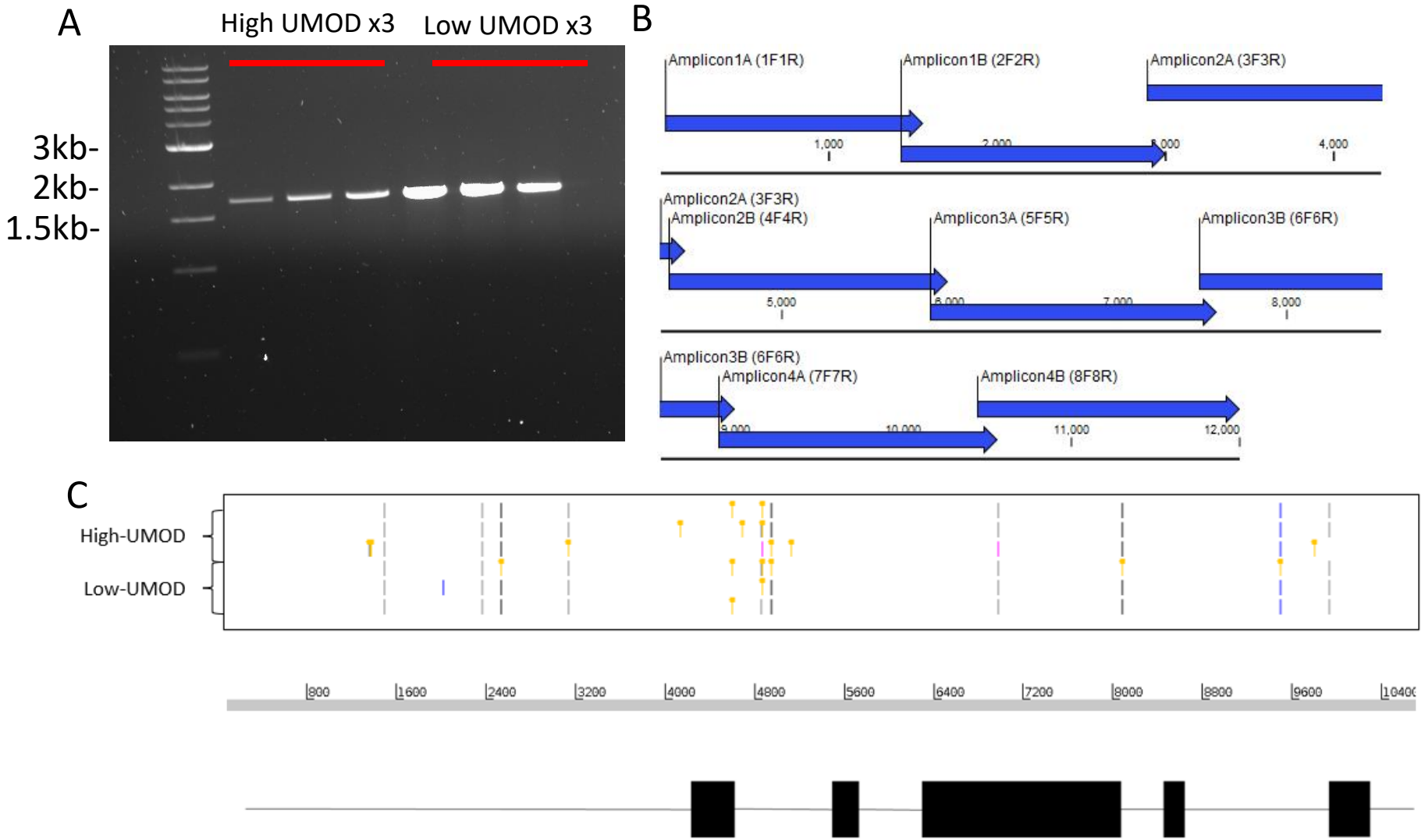


Figure 3.4-7 Next generation sequencing of the UMOD promoter.

A)UV luminescence of the 1F1R amplicon across samples, prior to NGS. B) Design of the overlapping amplicons for targeted resequencing. C) Detected variants across the samples between high and low UMOD expressers, displayed relative to the UMOD gene (reverse strand (promoter to right)).

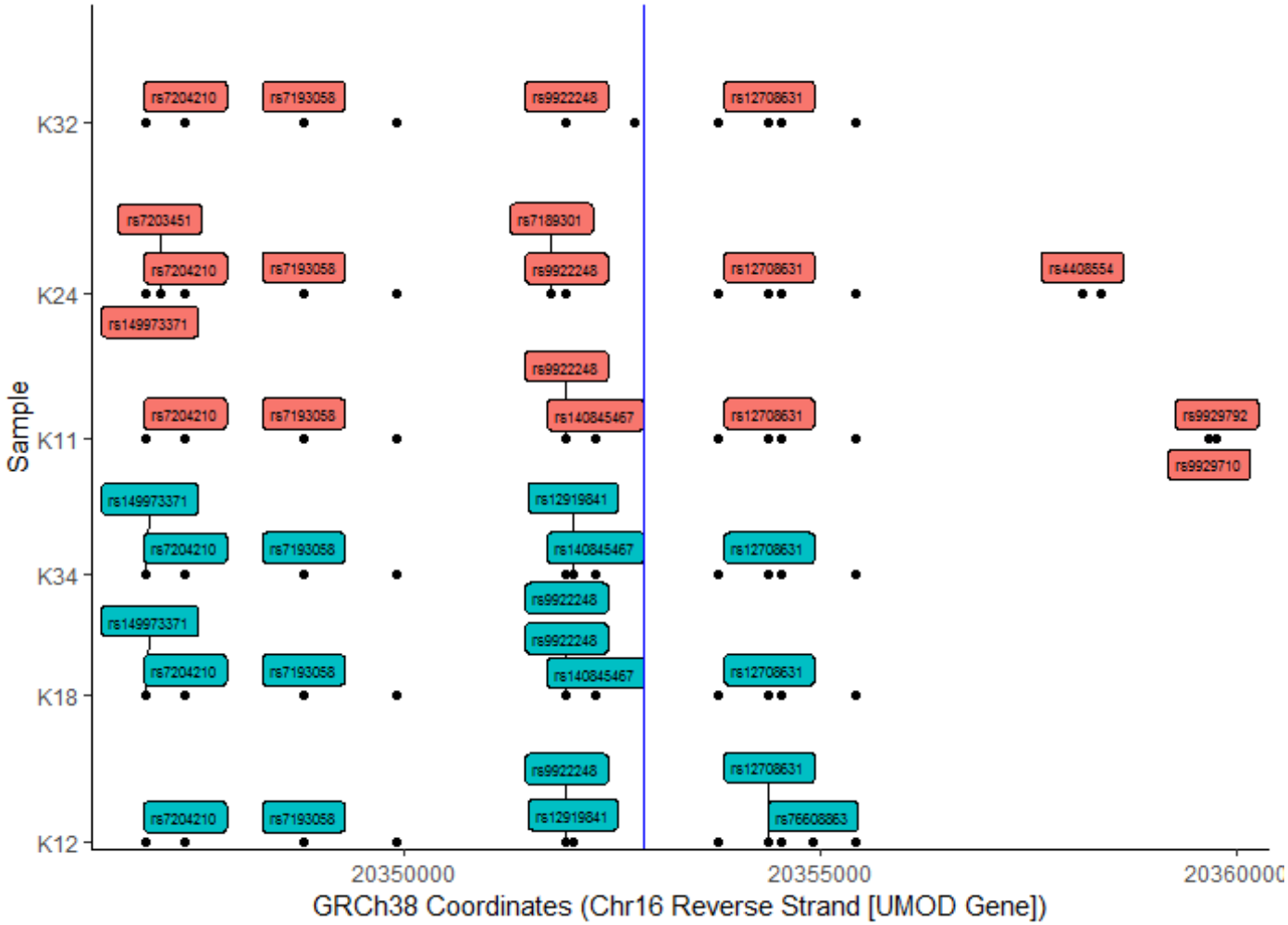


Figure 3.4-8 Linux based assignment of reference SNP ID by dbSNP performed with the GATK-HaplotypeCaller (2020). Shown 5' (left) to 3' (right). Blue line indicates the canonical TSS. Individuals with high UMOD expression designated by red colour, individuals with low UMOD expression designated by teal colour. Variants (black dots) not annotated within the dbSNP database are shown without labels.

3.4.3 Electrophoretic Mobility Shift Assay

Successful optimisation of the EMSA method was performed over several months in HeLa cells (data not shown), following which assays were performed using HEK293 total nuclear lysates. In the context of rs4997081, the C (risk) allele incorporated into a 50-mer oligonucleotide of the flanking sequence of this variant displayed significantly greater binding affinity (log-ratio of upper to lower band) versus lysate than the G (protective) genotype of between (0.133, 1.398, 95%CI) arbitrary units (Figure 3.4-9)(A/B)(lanes 3-5 and 7-9 are experimental lanes, with further lanes control). Furthermore, addition of the out-competition non-biotinylated oligonucleotide reversed the band shift binding observed (final lane), indicating this reaction was specific to the design of the experiment. There was no difference in the binding affinity of either genotype at rs13333226 (-1.22, 1.39, 95%CI) (Figure 3.4-9)(C/D). Critically, there was consistently an absence of any binding at rs13333226, whilst both genotypes at rs4997081 displayed binding, even when both variants were paired in the same experiment and electrophoresed on the same gel, indicating the observations are not determined by experimental batch (Figure 3.4-9)(E).

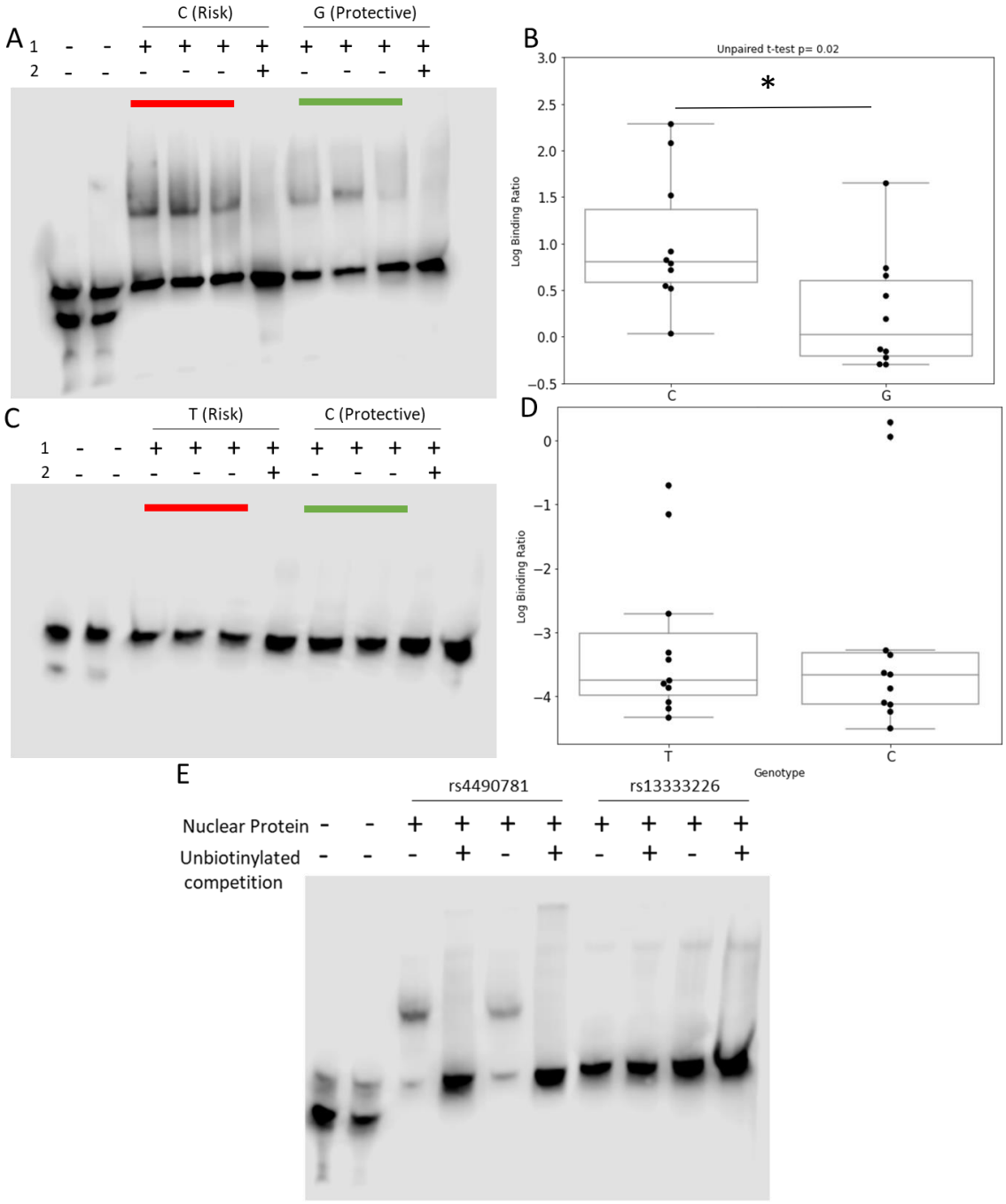


Figure 3.4-9 EMSA assay at rs4997081 and rs13333226

A) Example image of binding affinity for HEK293 nuclear lysate at the rs4997081 genotype. Lane 1&2 are controls for biotin signal, lanes 3-5 are experimental lanes (red) for the C genotype. Lanes 6 and 10 are negative controls for out-competition and lanes 7-9 are experimental lanes for G-genotype (green). B) Boxplot of total experiments at rs4997081. C) Example image of binding affinity for HEK293 nuclear lysate at the rs13333226 genotype. D) Boxplot of total experiments at rs13333226. E) Contemporary electrophoresis of both rs4997081 and rs13333226 variants.

3.4.4 Magnetic Purification and Mass Spectrometry

Based on conversations with Thermofisher, the supplier of magnetic dynabeads, we anticipated that successful purification of the complex containing biotin:streptavidin conjugated complex would produce a band at 29kDa, and that any of these complexes associated with transcription factors would produce a fragment greater than this (Figure 3.4-10)(A). During our purification process, we successfully eluted protein contaminants via washes, and evidence purified complexes which we then submitted for mass spectrometry. Monitoring of our wash fraction also indicated a successful wash-wise purification process as total protein fraction decreased with wash step (Figure 3.4-10)(B). With Coomassie blue staining, we detected a band at 29kDa in our purified product, however, we did not observe any other bands within the gel; likely attributed to the sensitivity of Coomassie staining.

Submission of these products for Tandem Mass Spectrometry in an n=1 per-group discovery phase returned characterization of 46 proteins in our negative control sample and 117 proteins in the rs4997081 oligonucleotide. A total of 29 of the protein matches unique to rs4997081 were enriched with >3 peptides and a MascotDb score of greater than 100, we selected these metrics as marked of reliability and plotted these on x and y axes in order to stratify proteins for visualisation (Figure 3.4-10)(C)(Figure 3.4-11)(A). We then performed text mining for for Uniprot descriptors, hypothesizing that these proteins may be involved in gene expression in the context of UMOD, from this we found that 15 of these descriptors for our detected proteins included the term ‘expression’ (Figure 3.4-11)(B).

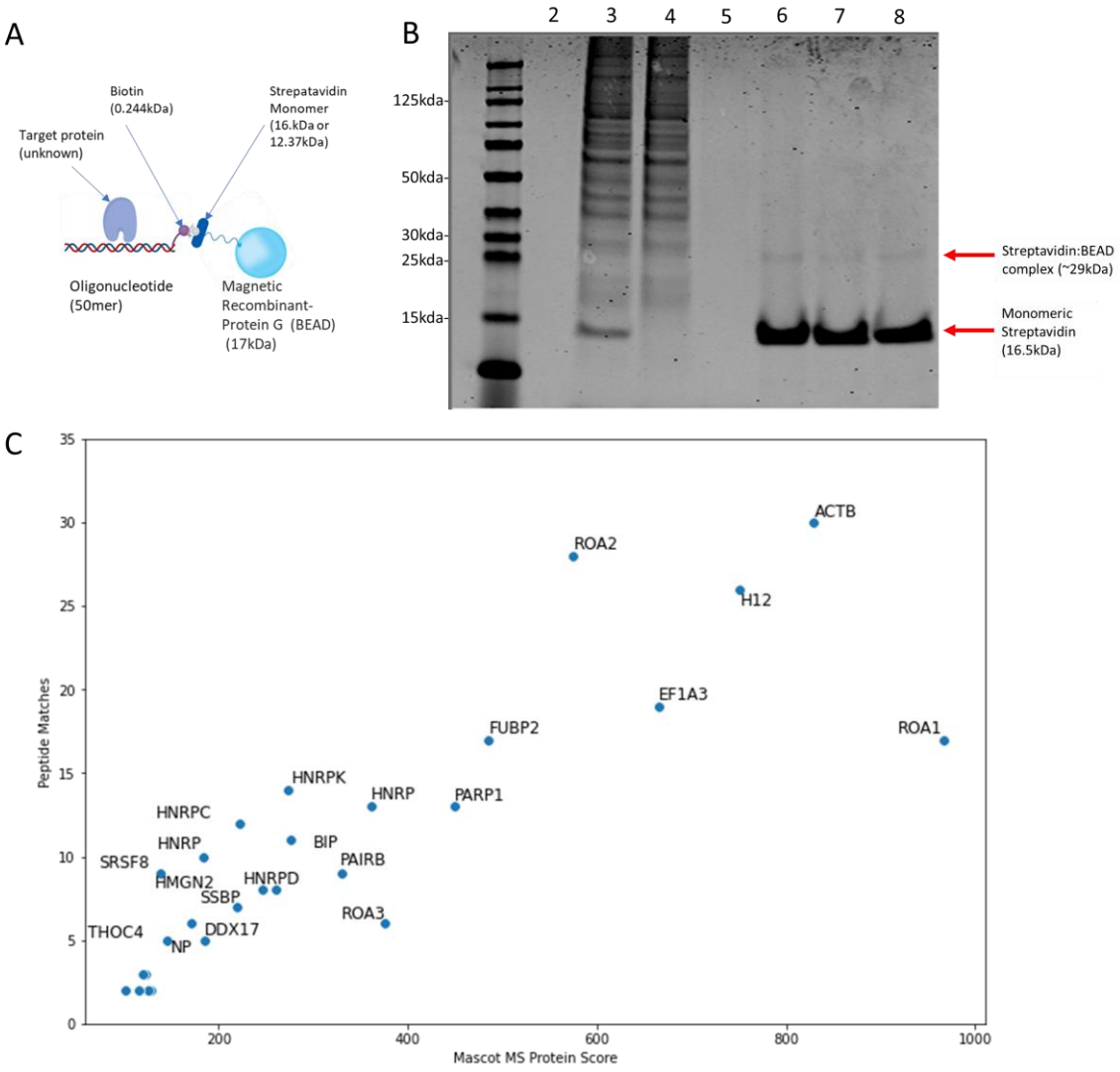


Figure 3.4-10 Magnetic purification and mass spectroscopy.

A) Schematic of the proposed complex generated by the magnetic purification process. Target proteins associate with the rs4997081 genotype. The 5' biotin of the oligonucleotide potently non-covalently associates with streptavidin of the m280 beads which are themselves part of the larger BEAD recombinant protein. B) Purification process by lane, L2- raw wash buffer, L3- initial wash supernatant using non-magnetised EMSA product, L4- wash supernatant after initial magnetic purification, L5- wash supernatant after 3 magnetic purification steps. L6-8 purified EMSA product submitted for mass spectrometry. C) Tandem Mass Spectrometry characterized proteins (derived from HEK293 cells) from the purified product, unique to rs4997081, versus control. These products were filtered on Peptide matches >3 and MS Score >50.

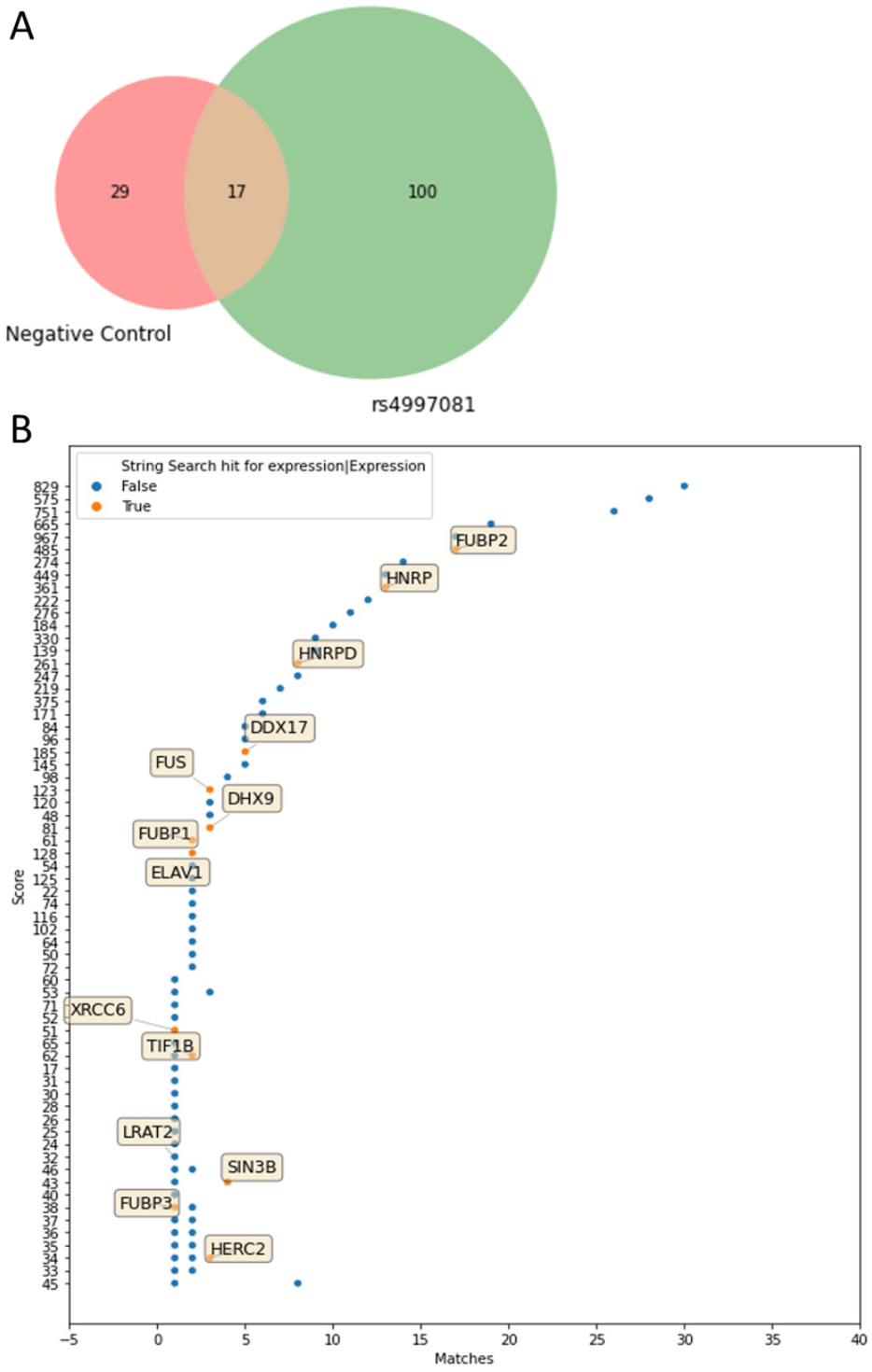
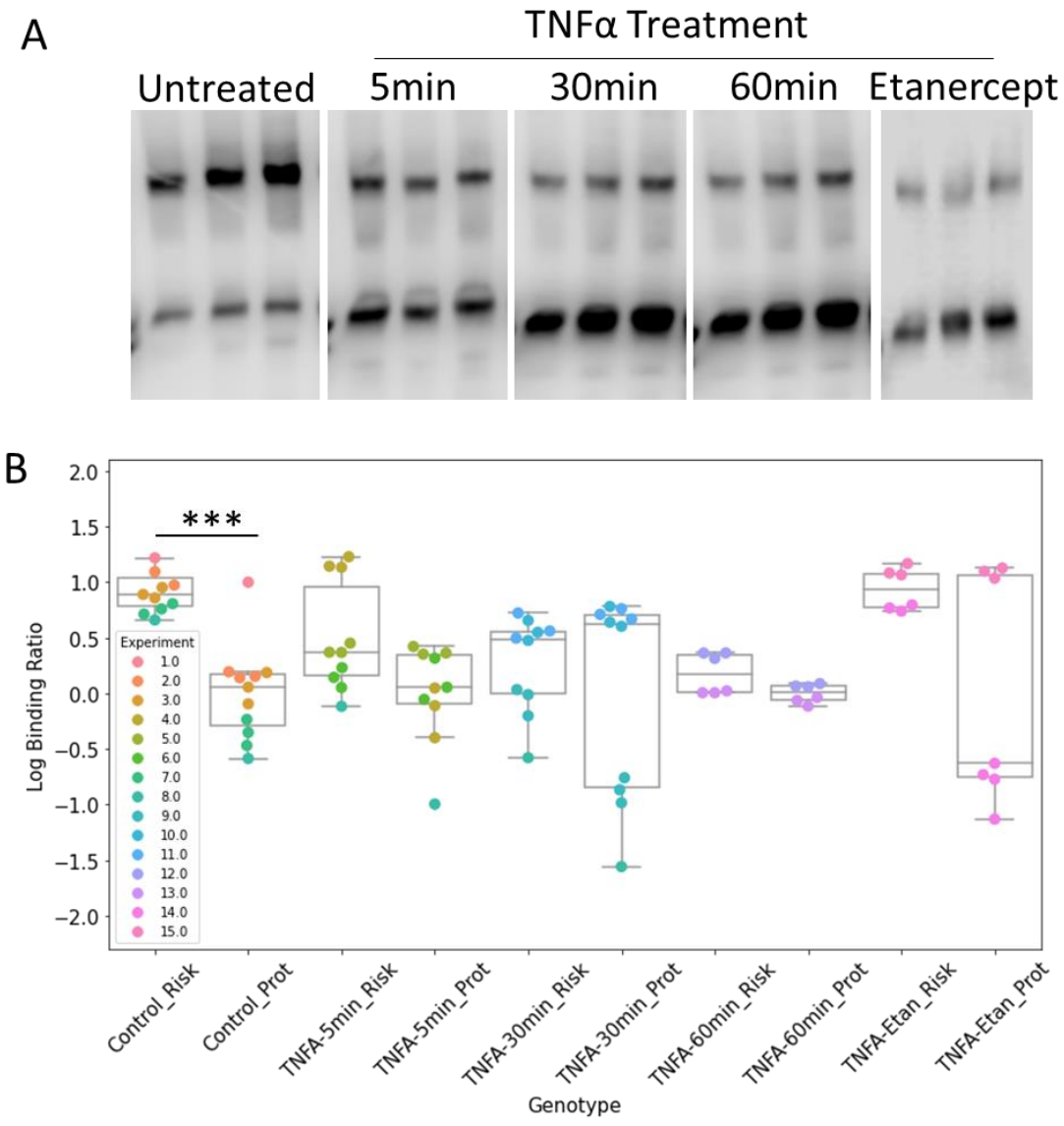


Figure 3.4-11 Analysis of mass spectroscopy data.

A) Venn diagram of detected peptides between negative control and rs4997081. B) String search on Uniprot descriptions of identified proteins for the strings ‘expression’ and ‘Expression’. Labels indicate string hit and the identified protein.

3.4.5 Analysis of TNF-α Treatment in the Context of Genotype

When HK2 cells were treated with TNF-α prior to extraction of their nuclear material, a time dependent negative effect was observed in terms of the binding affinity at the risk (C) genotype at rs4997081, but not the protective (G) genotype (Figure 3.4-12)(A). Comparing within treatment groups, by Bonferroni adjusted t-test showed a highly significant ($p<0.001$) decrease in binding affinity in the protective genotype versus the risk genotype in control HK2 cells, but no differences between genotypes when TNF- α treatment was added to the cells, and no differences with the addition of Etanercept and TNF- α (Figure 3.4-12)(B). Over the progression of this experiment, we began to consider that different treatments were exerting effects on the transcriptional complement, therefore for analysis purposes we chose to normalise within groups. When normalised within treatment groups, data indicate that the control risk genotype maintained significantly greater binding affinity versus the protective genotype of all treatments. Furthermore, there were no differences between genotype comparisons out-with the control experiment. A time-dependent, non-significant, trend could be observed, with the risk genotype becoming more similar to the protective genotype between 5 and 60 minutes, which appeared to reverse given the addition of etanercept, though this was also non-significant. However, a high degree of variability was observed in the TNF-α +Etanercept group, meaning we believe our interpretation of these data should remain cautious (Figure 3.4-12)(C).


A
TNFα Treatment
Untreated
5min
30min
60min
Etanercept
B
***
Log Binding Ratio
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
Experiment
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
Control_Risk
Control_Prot
TNFA-5min_Risk
TNFA-5min_Prot
TNFA-30min_Risk
TNFA-30min_Prot
TNFA-60min_Risk
TNFA-60min_Prot
TNFA-Etan_Risk
TNFA-Etan_Prot
Genotype

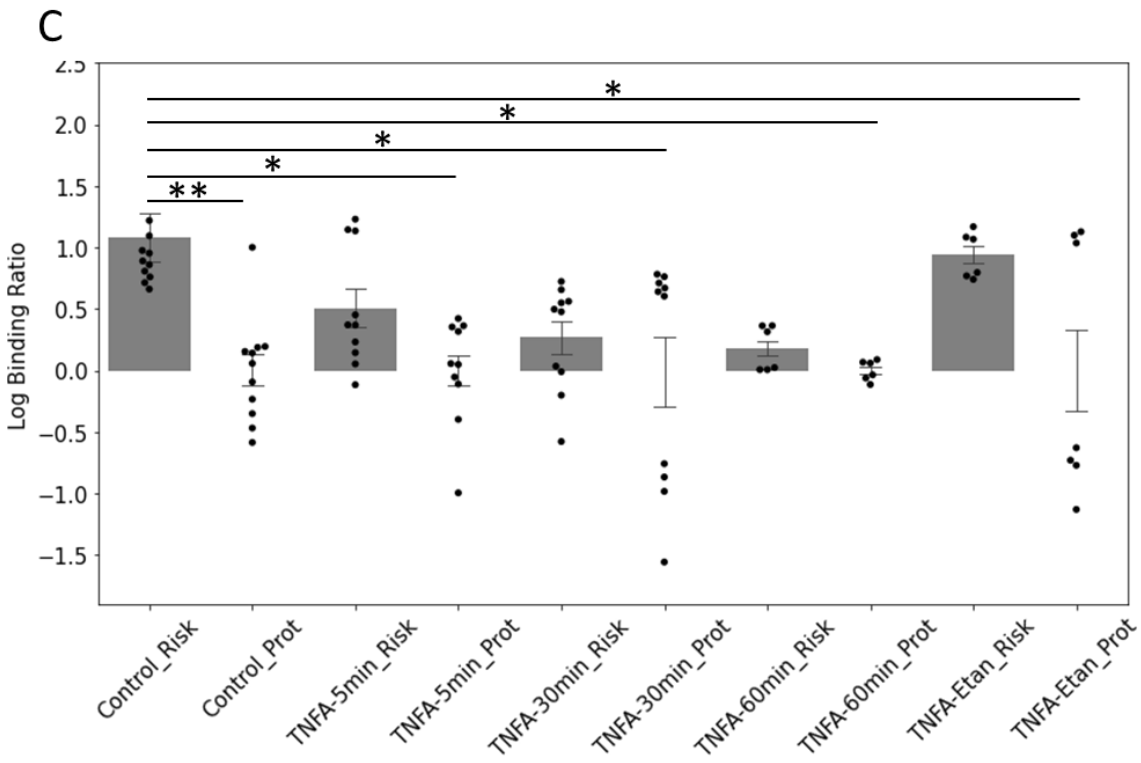


Figure 3.4-12 EMSA assay in the context of rs4997081 examining the effect of TNF-α on binding affinity between genotype

A) Example images of the risk (C 'Risk') genotype at rs4997081 across a range of 20ng/ml TNF-α treatments, administered to HK2 cells via media supplementation. Control= basic complete MEM media. TNF-α- Etanercept indicates 20ng/ml TNF-α supplemented with 10mg/ml Etanercept for 6-hours. *All cell culture and treatment experiments involving TNF-α were performed by Dr Lesley Graham. All EMSA reactions and analyses were performed by the author*. B) Log binding ratios across treatment and between genotype (C (risk, 'Risk') and G (protective, 'Prot')), experiment number is displayed by colour, *** $p<0.001$. C) Log binding ratios across treatment and between genotype (C (risk) and G (protective)), normalized within treatment groups. Anova, Tukey's post-hoc, * $p<0.05$, **, $p<0.01$.

3.5 Discussion

Whilst understanding of the relationship between UMOD and blood pressure regulation has been extensively developed following the seminal Padmanabhan et al GWAS publication, no group has ever shown using base-directed experiments, functional causality in terms of the relationship between UMOD variants and transcription factor binding. We show, for the first time, that rs4997081 displays significantly greater binding affinity for nuclear protein at the risk genotype versus the protective genotype, in multiple renal human cell lines. Conversely, we evidence for the first time, that the genotype at rs13333226 does not associate with differential binding affinity, and furthermore that, in the context of HEK293 nuclear lysate does not appear to interact with human transcriptional apparatus at rs13333226. Our data signal that a shift in the understanding of the UMOD promoter is necessary, and that rs4997081, not rs13333226, indicates the more likely causative variant. The strong linkage disequilibrium, combined with the likely masking of rs4997081 during initial analysis, explain why the haplotype was initially indexed at rs13333226. However, with a disparity of up to 8% in white Europeans between the two variants, combined with the high incidence of hypertension in the population, we strongly believe that the UMOD:blood pressure hypothesis can be enhanced be considering the potential functionality of rs4997081 at the locus.

Whilst we did not detect binding of transcriptional apparatus at rs13333226, we detected differential binding at rs4997081. Therefore, we further developed the discovery at rs4997081 by characterising proteins associated with this variant through mass spectrometry where we identified several potential transcriptional modulators. Our experimental design consisted of a novel operating procedure which delineated a complete process from variant to mass spectrometry characterisation, which does not exist elsewhere in the public domain. Notable findings from these data include the transcription factor FUSE binding protein 1/3 (FUSBP1/3) and Zinc Finger Protein 296 (ZNF296), which are known to be a regulators/ enhancers of gene expression by complexing with RNA Polymerase II (Gaudet *et al.*, 2011; Debaize and Troadec, 2018). Additionally, PC4 And SFRS1 Interacting Protein 1 (PSIP1) was detected, a protein which has been observed by GWAS as associating with systolic blood pressure (Chen

et al., 2014). However, of the findings from this experiment, we were most interested in the characterisation of Poly(ADP-ribose) polymerase-1 (PARP1) as enriched at the rs4997081. PARP1 is typically associated with DNA-repair, particularly in the context of cancer (Weaver and Yang, 2013). However, PARP1 is increasingly recognised as driving gene expression, with notable interactions with both HIF1-α and HNF1ß, where PARP-1 depleted melanoma cells showed reduced gene expression through HIF1-α (Rosenbluh *et al.*, 2016; Martí *et al.*, 2021)(Ray Chaudhuri and Nussenzweig, 2017). The fact that PARP-1 is noted as interacting with these transcription factors becomes ultimately important when considering that HNF1ß and HIF1-α have been found through meta-analysis as both existing within the top 20 most significant genes in terms of UMOD:creatinine ratio in humans (Olden *et al.*, 2014a). We did not detect either HNF1ß or HIF1-α within our mass spectrometry analysis. We argue therefore, that it is possible that differential PARP1:complex interactions at rs4997081 cause differential binding of either/or HNF1ß and HIF1-α at *cis* sites on the UMOD promoter which are themselves proximal, but not proximal enough that they were detected within our 50-mer oligonucleotide used in this study. Thus, PARP1 may exert effects on UMOD expression through interactions with other response elements on the cis-promoter. Fundamentally, we highlight the novelty of the PARP1 finding at the UMOD locus, and suggest this should be considered as potentially functional in terms of expression, particularly due it its differential binding at rs4997081.

PARP1 has also been noted as interacting with TNF-α signalling, in particular through nuclear factor-κB (NF-κB) (Alves-Lopes and Touyz, 2018; Liang *et al.*, 2018a). Nuclear factor kappa-light-chain-enhancer of activated B cells (NF-Kß) a transcriptional activator ubiquitously expressed in human cells, is critical to development, with knock-out mice being embryo-lethal (Hayden and Ghosh, 2014). In the context of Ang-II treated mice, PARP-1 inhibition by siRNA reduced the inflammatory response through inhibition of the phosphorylation of extracellular signal-regulated kinase (ERK) and NF-κB. Of additional interest is that NF-κB inhibition significantly decreases expression of HNF1ß in mice (Wu *et al.*, 2017). Crucially, it is a factor known to be activated through signal transduction by TNF-α (Dhingra *et al.*, 2009). Blockade of NF-Kß chronically in rats has been shown to decrease blood pressure, with this paper

attributing these effects to angiotensin-II receptor changes (Luo *et al.*, 2015). Additionally, sodium chloride treatment in rats has been shown to activate NF-Kß and subsequently increase blood pressure in these animals differentially between superoxide scavenging protocols (Dornas *et al.*, 2017).

We treated HK2 cells with TNF-α, which caused abrogation of the differences between the risk and protective alleles for their affinity for nuclear lysate. Though not significant, this reduction toward protective genotype in binding affinity at the risk genotype appeared to increase with treatment duration. This suggests that TNF-α treatment reduces the binding of transcriptional modulators (we propose PARP1) *in the risk genotype* at rs4997081, in turn reducing expression of UMOD. This proposed mechanism aligns closely with Heitmeier et al, where they showed that recombinant TNF-α significantly, and in a dose-dependent manner, reduced UMOD expression in mouse kidney thick ascending limb (MKTAL) cells (Heitmeier *et al.*, 2014). Heitmeier et al consider HNF1ß to be the transcription factor responsible for UMOD expression, so they additionally assessed HNF1ß levels, and show no change in response to TNF-α treatment. Furthermore extracellular UMOD has been shown to stimulate secretion of TNF-α in macrophages, suggesting a possible feedback loop between TNFα and UMOD, which we propose may be regulated by PARP1 (Immler *et al.*, 2020b). Importantly, HEK293 cells transfected to over-express wild-type UMOD display strong reductions in NFκB activation; suggesting negative feedback exists within this pathway to regulate UMOD expression (Dinour *et al.*, 2014). Critical to the understanding of our proposed mechanism is that risk genotype individuals have higher baseline binding affinity of PARP1 to rs4997081, that this differential affinity exerts effects on UMOD expression through NF-κB mediated activation of HNF1ß and that this increased binding affinity at the risk genotype negatively correlates with the presence of TNF-α in these cells. Curiously, one of the most documented side effects of PARP1 inhibition as a therapy for cancer is urinary tract infection, a condition known to be attributed to low urinary UMOD (Jiang *et al.*, 2019). We acknowledge that this proposed pathway involves extrapolation between indirect experiments, based on publication. Here we provide for the first time, a data driven explanation as to the differential expression of UMOD (Figure 3.5-1).

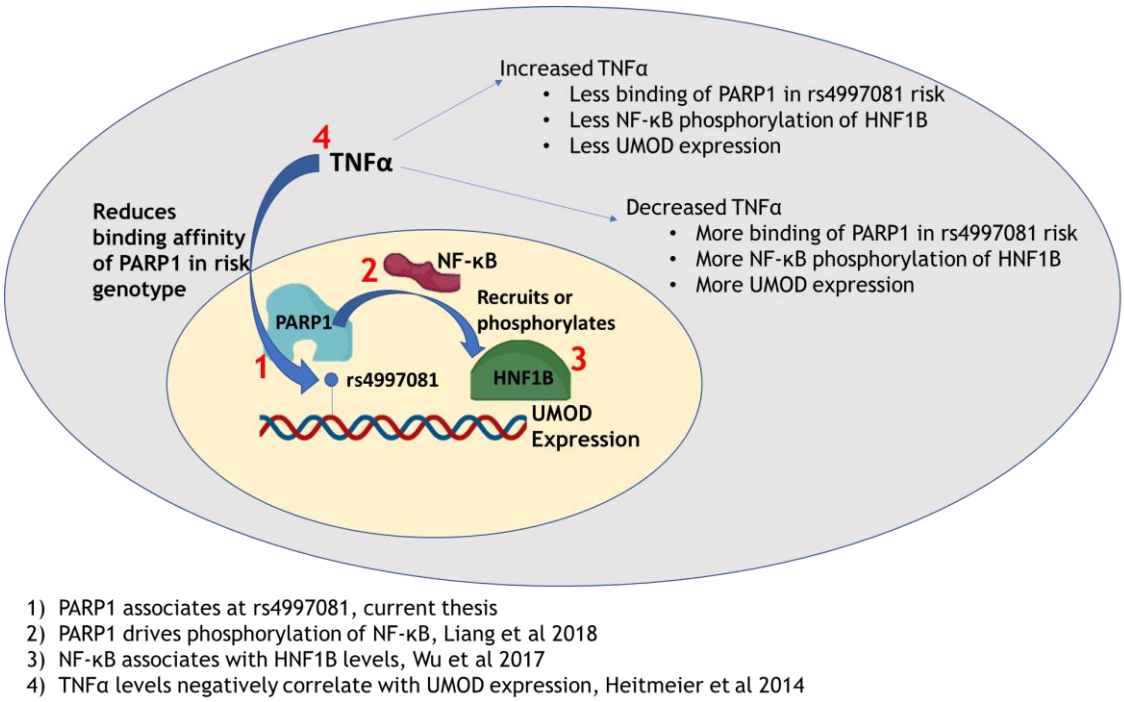


Figure 3.5-1 Proposed mechanism of differential gene expression between rs4997081 alleles

, based on thesis data and previous publication (indicated by red numbers).

- TNFα regulates differential association of PARP1 to rs4997081
- Associated PARP1 phosphorylates NFκB
- Phosphorylated NFκB activates HNF1β proximal to the PARP1 complex
- Activated HNF1β transcribes UMOD

Within our analysis of our bank of 84 human kidney samples, we detected a significant decrease in expression of UMOD in individuals prescribed amlodipine or bendroflumethiazide. This finding suggests an association of decreasing UMOD in response to antihypertensive use. Amlodipine is an angioselective L-type calcium channel blocker, with known off-target implications on the kidney. In particular, a study found that amlodipine (but not other classes on antihypertensives) was a significant risk factor in chronic kidney disease mortality (a disease which strongly associates with UMOD) (Haider *et al.*, 2015). Furthermore, in WKY rats, it was shown that nifedipine significantly decreased secretion of UMOD (Boder *et al.*, 2021). Conversely, Bendroflumethiazide inhibits NaCl reabsorption in the distal convoluted

tubules via NCC. Our finding that Bendroflumethiazide prescription appears to associate with decreased levels of UMOD in humans inversely reflects, but corroborates with, a finding in *Umod*$^{-/-}$ mice, which showed that these mice had increased levels of NCC mRNA (Bachmann *et al.*, 2005). We believe that these data may indicate that UMOD acts as a regulator within the renal system, responding to the disruptions to homeostasis induced by medication use. Not only should this be considered when analysing human population level data in the context of UMOD expression, but it could also be used in future experiments to further refine the precision medicine hypothesis surrounding UMOD. We should however, stress that the group size of blendroflumethiazide users was relatively small (n<15) which may have exposed this group to outliers; thus this initial finding should be regarded as requiring further validation.

The finding that both risk alleles at rs13333226 and rs4997081 were associated with significantly increased blood pressure versus control forms the basis of what we consider a limitation rather than a finding. The American Heart Association as of 2019 states that as a minimum, recordings by auscultation (cuffing) should be performed at least two times on two individual sittings, and the average of these >4 measurements should be recorded for research purposes (Muntner *et al.*, 2019). The information provided to us by the NHS used one pre-surgical measurement, and as a result would be considered unreliable. Were we to assume that these covariates were randomised across genotypes, we could interpret blood pressure readings, we are however, unwilling to do this. As a result, we cannot reliably link our findings from our human cohort back to blood pressure, even though blood pressure is the fundamental base-metric of the hypothesis. In terms of additional limitations, we detected the homozygous protective genotypes for rs13333226 and rs4997081 at an incidence of 2.8%, meaning we had very few (3) of these homozygous protective genotypes to include in our analysis. If the effect is allele-dose-dependent, then we believe we would have had significantly greater power over the study to detect differences had we been able to incorporate greater numbers of homozygous protective individuals. As we discussed in 3.4.5, the EMSA protocol we optimised remains a highly sensitive and challenging assay in terms of binding conditions, and

we are aware of potential drift, possibly attributed to treatment. Ideally, we would have performed all treatment points on the same day, and built our n-number in this manner, to eliminate treatment/batch effects. However, due to the nature of cell culture, and time restraints of the PhD, we could only perform experiments as the material was generated. A method of overcoming our concerns regarding the effect of treatment on cells would be to pair the generation of material with a cell viability assay such as an MTT assay, to ensure that cells remained metabolically similar between treatments, eliminating the potential need to within-group normalisation.

In terms of future work, this chapter presents novel pilot data implicating rs4997081 as functionally relevant in terms of UMOD expression. However, the mass spectrometry data we present was a preliminary discovery qualitative experiment, in that we only examined n=1 (and negative control), due to time constraints. We believe that antibody directed super-shift of anti-PARP1 on the purified EMSA products would provide quantitative data which we could cost-effectively expand to multiple samples. Additionally, we could use this technique to further enhance our understanding, by assessing PARP1 densitometry between rs4997081 genotypes. Furthermore, there are two more variants on the UMOD promoter which have been investigated in terms of UMOD expression by other groups, Trudu et al previously implicated rs4293393 and Olden et al describe rs12917707 as the most likely functional variant; though neither of these studies employ the base-directed experiments we used, rather they focus on haplotypes (Trudu *et al.*, 2013a; Olden *et al.*, 2014b). In terms of completeness, we would recommend that EMSA experiments also be performed against the genotypes of rs4293393 and rs12917707. Therefore, while these pilot findings are biologically interesting, they require considerable validation and follow-up leading on from this thesis to further determine causality. Furthermore, the low sample size, due to the inherent costs of these experiments, must be addressed, ideally providing a statistically addressable sample size for quantitative inference to be made, as the current n=1 should be regarded as qualitative only.

The work in this chapter highlights rs4997081 as functionally causative in driving differential associating of transcriptional apparatus between alleles. Conversely, we

provide evidence that rs13333226 is not functional by the same assay, though we did not examine other potentially relevant variants during this process. Using mass spectrometry, we identified a PARP1 containing complex as binding at the rs4997081 allele. Our experiments using TNF-α suggest that TNF-α treatment abrogates the differences in risk and protective genotypes. We believe these findings are linked to expression of UMOD via NF-κB induced HNF1ß, in a PARP1 and subsequently genotype dependent manner. To our knowledge, these experiments form the basis of the first functionally described relationship between UMOD and blood pressure.

4 UMOD and the Transcriptome

4.1 Introduction

The combination of research by both Trudu et al and Graham et al toward understanding the mechanisms underpinning the association of UMOD with blood pressure have provided substantial insights, but carry limitations. Trudu et al took an *activation-centric* approach to their experimentation, particularly focusing on the hypothesized role of UMOD in increasing activity and membrane localization of NKCC2 in TAL cells by phosphorylation (Trudu *et al.*, 2013). In contrast, Graham et al showed that the *Umod*$^{-/-}$ mice exhibited decreased expression of NKCC2 mRNA, which they hypothesise was driven by changes in TNF-α bioavailability in the lumen (LA *et al.*, 2014). Both approaches provide validated functional roles, however, neither was able to interpret the multifactorial role of UMOD in TAL cells, as these experiments were directed at both the mRNA and protein level by hypothesis-led rationale. Thus, an experimental protocol which is not directed by prior hypothesis may lead to a more refined understanding of the global roles of UMOD in blood pressure regulation. Ideally, such a study would make use of hypothesis-free, high-throughput omics style in order to provide a systematic interpretation as opposed to a specific one.

Building from Microarray analysis, which uses a panel of pre-selected genes and proceeds by gene-specific amplification on a larger scale than qRT-PCR, bulk RNA sequencing (RNA-Seq) was the first developed method by which researchers could theoretically examine the entire transcriptome of a sample, though these libraries are typically lacking in long non-coding components (Engelen *et al.*, 2006). Typically, bulk RNA-Seq performs non-specific RNA selection and non-specific reverse transcription on a homogenized sample. Fragmentation is then performed, typically generating fragments between 75-150bp, following which size selection occurs. After this, non-specific PCR is performed with a discretionary level of cycles depending on input RNA (in order to load the sequencer with adequate material), before the sample is sequenced, typically by Illumina next-generation-sequencing. Transcriptomic sequencing is, by contrast to previous methods of RNA quantification (qRT-PCR and Microarray), the first method which does not rely on a prior hypothesis. Instead, it is

expected to return a broader characterization of the constituent molecules within the analysed sample (Wang, Gerstein and Snyder, 2009). The catalogued transcriptomic data can then be interpreted for potential pathways which can be developed by gene-set-enrichment analysis (GSEA) and pathway analysis to provide analyses supporting linkage of the underlying hypothesis to diseases and functions.

With the rapid development and utilization of bulk RNA-Seq came the subsequent bioinformatics-based development of tools used to provide functional interpretation of these returned data. Central to this rational are 'gene-sets'. Gene-sets are effectively libraries of groups of genes which contain the summarized knowledge of genes associated with cellular function. For example, a gene-set may contain three genes (X (decreased expression),Y (increased expression) and Z (increased expression)) and be correlated with diabetes. At its smallest dimension, if a researcher produces an RNA sequencing dataset in which gene X is decreased and Z is increased, then by gene set enrichment analysis (GSEA), they may label their dataset as associating with diabetes, as the majority of the genes within the gene-set were differentially expressed in their experiment (Chen *et al.*, 2013). These libraries exist in a maintained and curated set on MSigDB (A *et al.*, 2011). Building on GSEA, Ingenuity Pathway Analysis (IPA)(Qiagen, Manchester, UK) uses custom algorithms to perform GSEA, however they are less scrutable due to their proprietary nature. The author is aware that IPA use a Fishers Exact Test to detect differences by gene-set overrepresentation analysis but cannot view the source code. Regardless of openness in their algorithms, they themselves are powerful tools which when used alongside typical GSEA can provide additional insights to biological systems and pathways as they also incorporate GWAS data and microRNA based mirTar databases into their analysis (Hsu *et al.*, 2011).

Single-cell sequencing (scRNA-Seq), developed in the decade leading from bulk RNA sequencing, further refined the technique, by examining not only the whole transcriptome on a sample-wise level, but rather on a single cell level (Hwang, Lee and Bang, 2018). Though there are multiple service providers for scRNA-Seq, each with somewhat distinct protocols, broadly summarized, scRNA-Seq methods isolate cells into individual 'reaction vessels', generally droplet based, before performing

in-droplet reverse transcription, which involves the tagging of transcripts with both a cell-based unique-molecular-indicator (UMI) and a transcript based UMI, after which these samples are sequenced by NGS. Due to the UMI tagging nature of the library preparation, genetic material can then be derived back to individual cells, maintaining the cellular precision advantages of scRNA-Seq versus bulk RNA-Seq Unlike its predecessor, bulk RNA sequencing, scRNA-Seq can elucidate complex relationships between cell types within the dataset and return additional information on cell trajectory. In particular, scRNA-Seq can be utilized to examine heterogenous gene expression within a sample containing the 'same' cell types (L and H, 2010; AK *et al.*, 2014). Increasingly, whereas previously this variation would have been considered 'noise', it is now considered to be biologically relevant and informative (S, 2009).

Of relevance to this study is the potential for UMOD to form coexpression networks- where the presence of UMOD mRNA correlates either negatively or positively with the presence of other genes. Whilst no formal bioinformatics data-analytics methodology exists for generating these correlations, potential does exist for the development of one. Given access to high-throughput data in the context of UMOD would allow us to investigate co-expression. Co-expressed gene subnetworks are known to play a role in blood pressure regulation with Huan et al identifying four gene co-expression networks which were associated with blood pressure regulation in a sample of 3679 humans (Huan *et al.*, 2015). Building on this, Cao et al used the same algorithm but this time in the context of salt sensitive hypertension, again in humans, and identified gene coexpression networks which associated with salt sensitive systolic blood pressure (Cao *et al.*, 2019). While these findings do not directly implicate UMOD within these networks, they do infer that this would be an import analyses to perform.

UMOD, in the context of blood pressure, presents as a suitable candidate for RNA-sequencing analysis because previous investigation was only able to focus on single-gene or single-protein analysis (derived from an NKCC2 focused hypothesis) and therefore was only able to provide interpretation based on this hypothesis. We believe that UMOD may contribute to blood pressure regulation through multiple

mechanisms in these cells, which provide a cumulative contribution. In particular, further data on potential signalling molecules between UMOD levels and the function of NKCC2 would be highly informative. The experimental work of this chapter investigates RNA sequencing (both bulk and independent single-cell) in the context of UMOD expression, with the end goal of extrapolating this to blood pressure regulation. Using human renal RNA stratified on UMOD expression, we aim to investigate these differences. In doing so, we aimed to refine prior knowledge generated by both Trudu et al and Graham et al and in turn provide additional insights associating the UMOD gene with blood pressure in a hypothesis-generating manner.

4.2 Aims

- To examine the relationship between UMOD levels and the renal transcriptome in humans using RNA sequencing

- To use pathway analysis to investigate the relationship between UMOD and renal physiology

- To use a combination of wet-laboratory and single cell mining to validate these observations

4.3 Specific Methods

4.3.1 RNA Sequencing

Samples were selected and processed as per section 2.1.7, 2.2.3 and 2.2.4.

4.3.2 Deseq2 Based Identification of Differentially Expressed Genes

Deseq2 was used on count files generated during 4.3.1 to identify differentially expressed genes. Deseq2 models these counts as following a negative binomial distribution- a distribution which models the probability of extreme log2 fold changes occurring by chance across replicate number. Deseq2 normalizes read counts using the median-of-ratios method- a statistical normalization method distinct from Transcripts-per-million (TPM) or Reads-per-kilobase-million (RPKM). Specifically, whilst TPM/RPKM type normalization methodologies which take derivation from the total genetic material *within* a sample, median-of-ratios uses a *between* sample approach. Mathematically stated, Deseq2 normalisation calculates the within-gene geometric mean μ and calculates a size factor estimate s as the median of the ratios to this μ.

Following normalization of read counts, Deseq2 performs hypothesis testing to determine differentially expressed genes (DEGs) (those genes where the null hypothesis of differences between groups can be rejected). Specifically, Deseq2 takes the shrunken estimates of Log2FC values derived from normalized counts and divides these by the standard error for this gene, producing a z-score which is compared against the null distribution (statistically speaking, DEGs would differ significantly from this distribution (FDR<0.05)). This Wald test can model both coefficient distributions for individual genes and coefficients for all contrasts. The p-values returned for per-gene comparisons by Deseq2 are then corrected for multiple testing effects using the Benjamini Hochberg false-discovery-rate (FDR) adjustment (Anders and Huber, 2010; Benjamini, 2010; Love, Huber and Anders, 2014)

4.3.3 Alternative Statistical Methodology

Where applicable, the Wald test was noted to decrease in resolving power as the number of genes within the model reduce, due to its reliance on the compounding effect of multiple genes on the accuracy of the model. As a result, a simple statistical algorithm was constructed in python3.8. This novel method consisted of firstly a within-group normalization, using the log(counts-per-million+1) (log(cpm+1) method. The log(cpm) method was performed by initially calculating the total counts for a sample/1,000,000, after which the counts for a specific gene were divided by this constant. The log of these values were then calculated, with any 0-values being converted to 1-values to prevent infinitely negative floating point errors. Following this, between-sample hypothesis testing was performed using a t-test using arrays of normalized count data, under the assumption that the variance was not equal between groups (Welch's t-test).

4.3.4 Utilisation of Gene-sets and Pathway Analysis

Gene set enrichment analysis was always performed using the latest editions of all gene-sets, as curated by MSigDb. Data were loaded into GSEA using normalized counts as quantitative metrics. Deseq2 output files were loaded into IPA, where FDR<0.05 and log2FC > 1 were considered statistically significant DEGs. For pathway analysis within IPA, connections between genes were constructed only where individual genes could be linked on a pathway by one or less intermediary genes.

4.3.5 Single Cell Data Mining

For data assessed for cell specific co-expression using the custom written Python3.8 application described in 1.1.1 (sic) and 4.3.6, data was mined from public database repositories. In particular, all datasets were obtained from the SRA (Read Archive) using the following accession numbers. In order to provide biologically relevant inference, all mined datasets had to be human in origin, and had to be derived from renal tissue from a healthy individual. Additionally, only verified upon initial analysis, each dataset had to contain the presence of UMOD+ cells, indicating the inclusion of TAL cells in the sample originally submitted for single cell analysis. Due to the limited

availability of appropriate datasets, the haplogroup of the individual was not considered physiologically meaningful.

Single Cell Dataset Mining	
Database	**Accession**
NCBI:GEO	GSE118545
NCBI:GEO	GSM4572192
NCBI:GEO	GSM4572193
NCBI:GEO	GSM4572194
NCBI:GEO	GSM4572195
NCBI:GEO	GSM4572196

Table 4.3-1 Single cell dataset mining accession ids

4.3.6 Custom Co-expression Analysis

While Scanpy libraries for providing single cell analysis provide meaningful insights into differences in cell populations between healthy and disease states, they cannot to date address the question of co-expression. ScGress, a custom authored python3.8 application attempts to perform this function. ScGress initially performs standard UMAP and Louvain algorithm clustering via Scanpy in python. However, following this the program allows select of a ‘gene of interest’ (GOI). ScGress then thresholds the entire dataset, only selecting cells which have a positive expression of the GOI >2 in log-normalised counts. ScGress additionally allows for the selection of cluster specific co-expression, by providing the option to filter for only a single Louvain labeled cluster. However, in the context of UMOD, because there is marginal expression of the gene in the dCT alongside the TAL, cluster specific co-expression was not utilized. ScGress then creates a list of all genes contained within the dataset, and following this performs statistical testing in a step-wise manner, between all of the genes (target genes) (TGs) in the dataset and the GOI. This is calculated by taking the expression levels of the TG and GOI on a per-cell basis and performing linear regression on these expression levels.

Statistically speaking, ScGress gathers metrics on the linear regression p-value (OLS method), the R of the regression, the beta-coefficient (slope) and finally the proportion of GOI expressing cells which also express the TG. This final metric is gathered to prevent TGs being labelled as physiologically interesting when in reality they are only expressed in a very small proportion of GOI cells, causing a type-II error in the statistical interpretation. The proportion of cells which must be TG+ was thresholded at 70% in the context of UMOD expression, though the application allows for this to be changed.

Following competition of the program, the user is provided with a dataframe containing all of the TGs tested (<10,000 genes long), their relationship to GOI expression (positive or negative correlation), the fit of that correlation (via the R metric) and lastly the significance of this correlation. Though not standard, it is advised that Bonferroni multiple-testing-corrections are applied to these p-values, however, as a discovery method it is noted that this may lead to type-I errors in interpretation due to the highly heterogeneous nature of single cell gene expression, even within clusters.

4.4 Results

4.4.1 Selection of the RNA-sequencing Subset

All 92 human samples exhibited expression of UMOD mRNA under a normal distribution with a somewhat skewed tail of lower expressors (these data are dCT, therefore higher values indicate lower expression) (4.4.1)(A). During exploratory analyses, the expression of UMOD by blood pressure was observed. Expression of UMOD mRNA did not significantly correlate with either diastolic (Figure 4.4-1)(B) or systolic (Figure 4.4-1)(C) blood pressure, though in both cases a non-significant positive trend between UMOD expression (normalized to beta-actin) and blood pressure was observed. Within the total cohort, the use of antihypertensive medication, a potential confounding variable, was widespread and heterogenous with 55 of 106 patients being medicated against hypertension (Figure 4.4-2)(A). However, selection of individuals not prescribed antihypertensive medication at the time of data entry were still successfully stratified on UMOD expression (Figure 4.4-2)(B/C). These n=3 sub-groups were not significantly different in terms of either their SBP or DBP. Furthermore, within the limits of n=3 analysis BMI, age and sex were not different between group (data not shown). The flow chart of the selection process for individual samples for next generation sequencing can be observed in Figure 2.1-0.

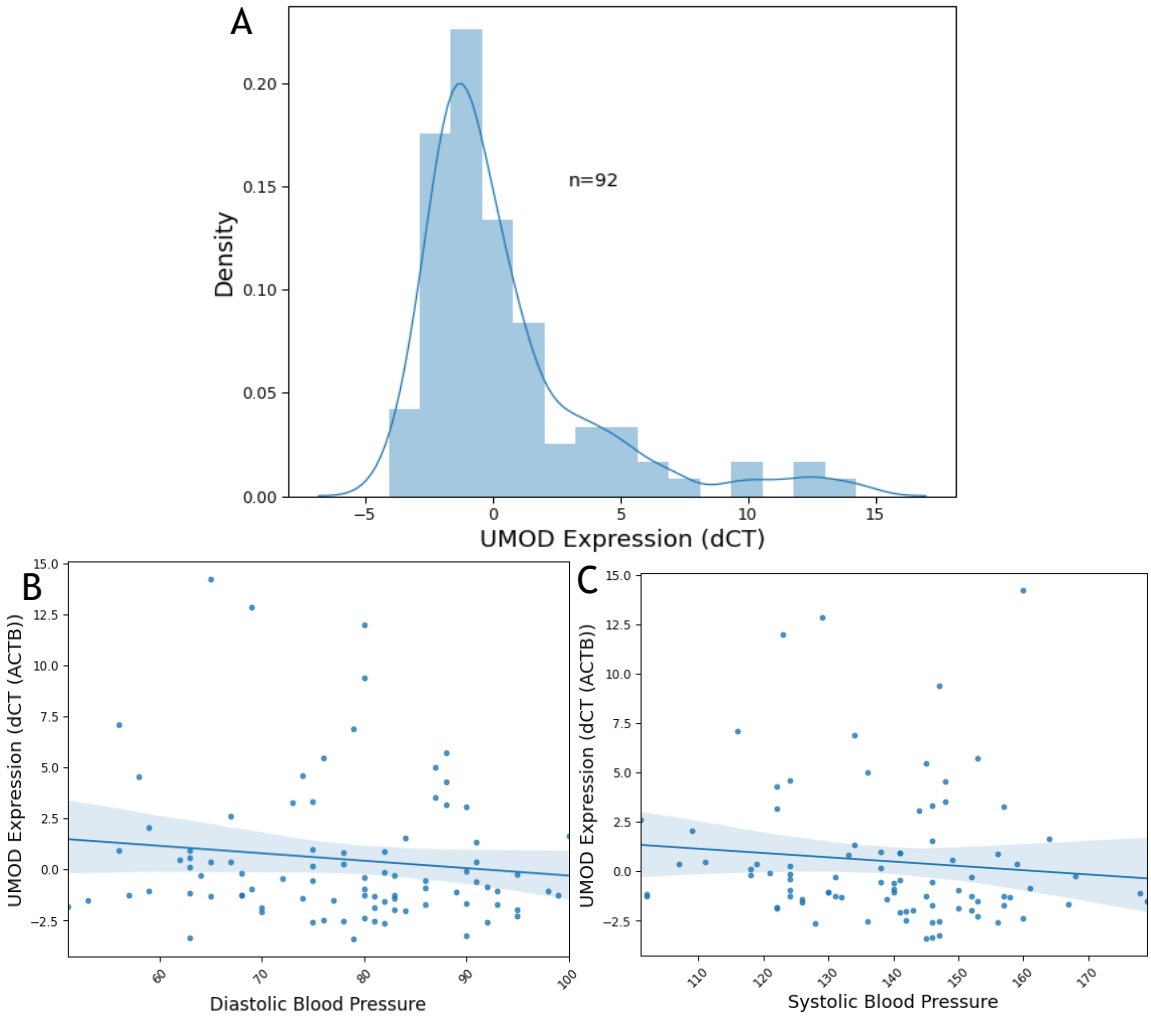


Figure 4.4-1 Correlation analysis of UMOD and blood pressure

A) Distribution of UMOD expression (dCT vs ACTB) in RNA obtained from human renal biopsy (n=92) as shown by density histogram analysis with fitted polynomial. D'Agostino and Pearson's test for normality indicates normal distribution of UMOD expression across samples p=1.65e-11. B) UMOD expression regressed against diastolic blood pressure (n=92), no significant trend by OLS (p>0.05). C) UMOD expression regressed against systolic blood pressure, (n=92), no significant trend by OLS (p>0.05).

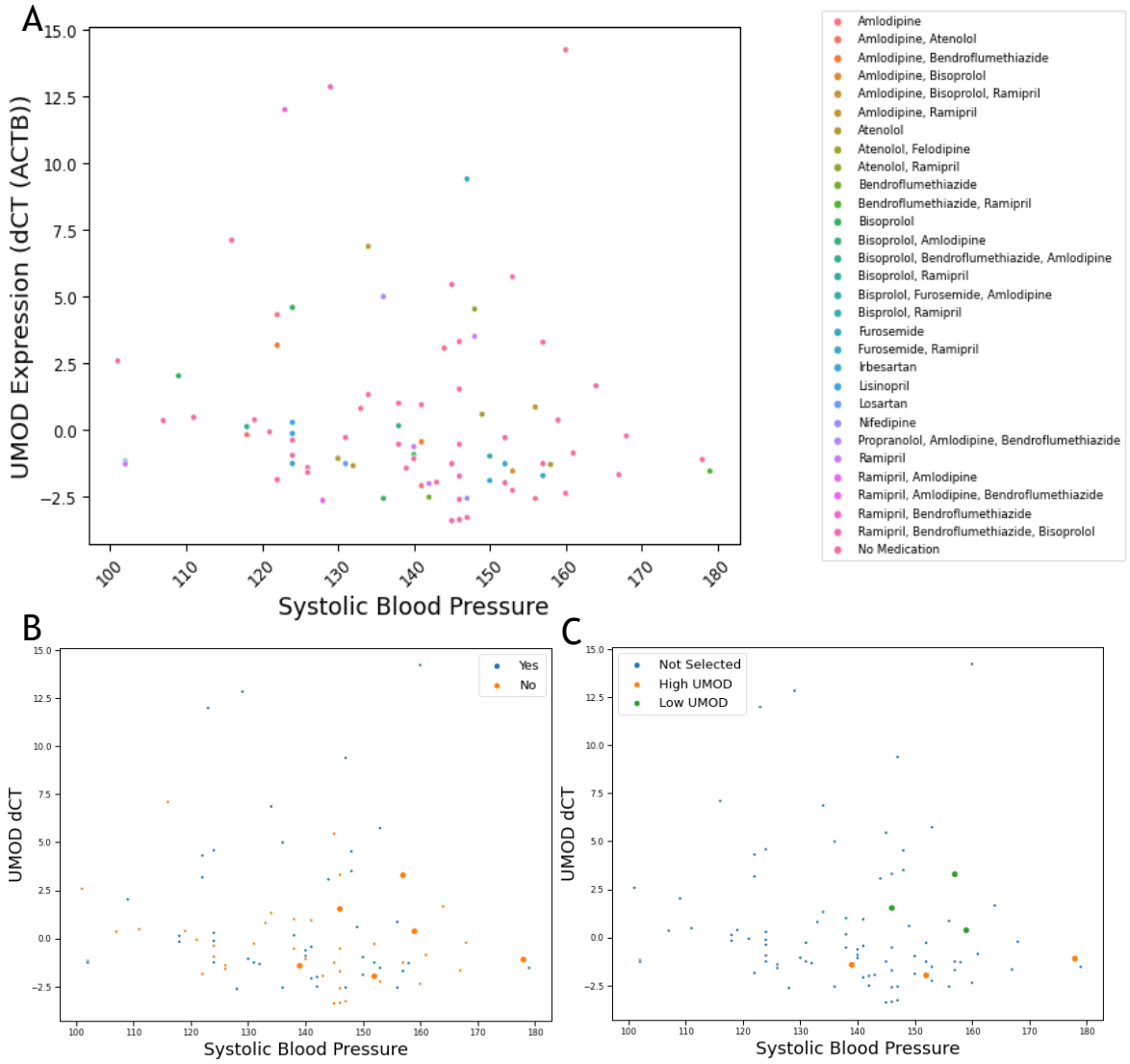


Figure 4.4-2 Selection of the high and low UMOD expressor subset

A) Distribution of antihypertensive medication usage across the dataset, using UMOD expression as the y-axis variable and systolic blood pressure as the x-axis variable for visualization purposes. Those not prescribed antihypertensive are also shown as 'no medication'. B) Binary classification of antihypertensive medication usage, all samples submitted for RNA-sequencing were not prescribed any hypertensive medication. C) Selection of the 'low' (green) and 'high' (orange) UMOD expressers (NB data shown as dCT meaning expression levels are the inverse of shown values). inverse of shown values).

4.4.2 Quality Control

Samples submitted for RNA-sequencing analysis displayed marginal degradation when assessed by Bioagilent (represented by the ratio of 18S and 28S subunits to total). Quantitative assessment of the RIN values indicated a RIN of 7.42 (6.91, 7.91 95%CI), indicating RNA submitted for assessment by RNA-sequencing was of acceptable quality (Figure 4.4-3)(A/B). Distribution of unfiltered read counts across individuals samples was uniform and normally distributed when log transformed, though a proportion of genes were filtered due to 0-counts in post-processing (an accepted process in RNA-sequencing analysis), which explains the marginal bimodality observed at 0-counts in the raw data (Figure 4.4-3)(C). As component influence is expected to logarithmically decay from PC1 though the additional components, principal component analysis displaying that PC1 and PC2 explained 32% and 22.6% respectively was reasonable (Figure 4.4-3)(D). Principal component analysis produced acceptable separation between groups, though grouping indicated a possibility that >2-means clustering would be superior to the prescribed 2-means in terms of Normalized Mutual Information score (NMI). Specifically, 2 of the 6 samples did not cluster well within their perceived grouping by PCA indicating that they are somewhat biologically distinct from the other samples within their grouping.

Due to the TAL specific expression of UMOD, it was necessary to confirm that 'high' UMOD expressors were not simply samples which sampled more thick ascending limb structures than 'low' UMOD expressors. Confirmation of the absence of sampling region variance was confirmed by data-mining TAL specific markers in mice and converting these to human orthologues before assessing our RNA-sequencing data for differences in these markers between groups (Figure 4.4-4). We hypothesized that if sampling was from similar region, TAL markers would not be differentially expressed. Of 96 mTAL markers assessed, only four were differentially expressed, indicating differential biopsy locations between groups was unlikely to be a confounder.

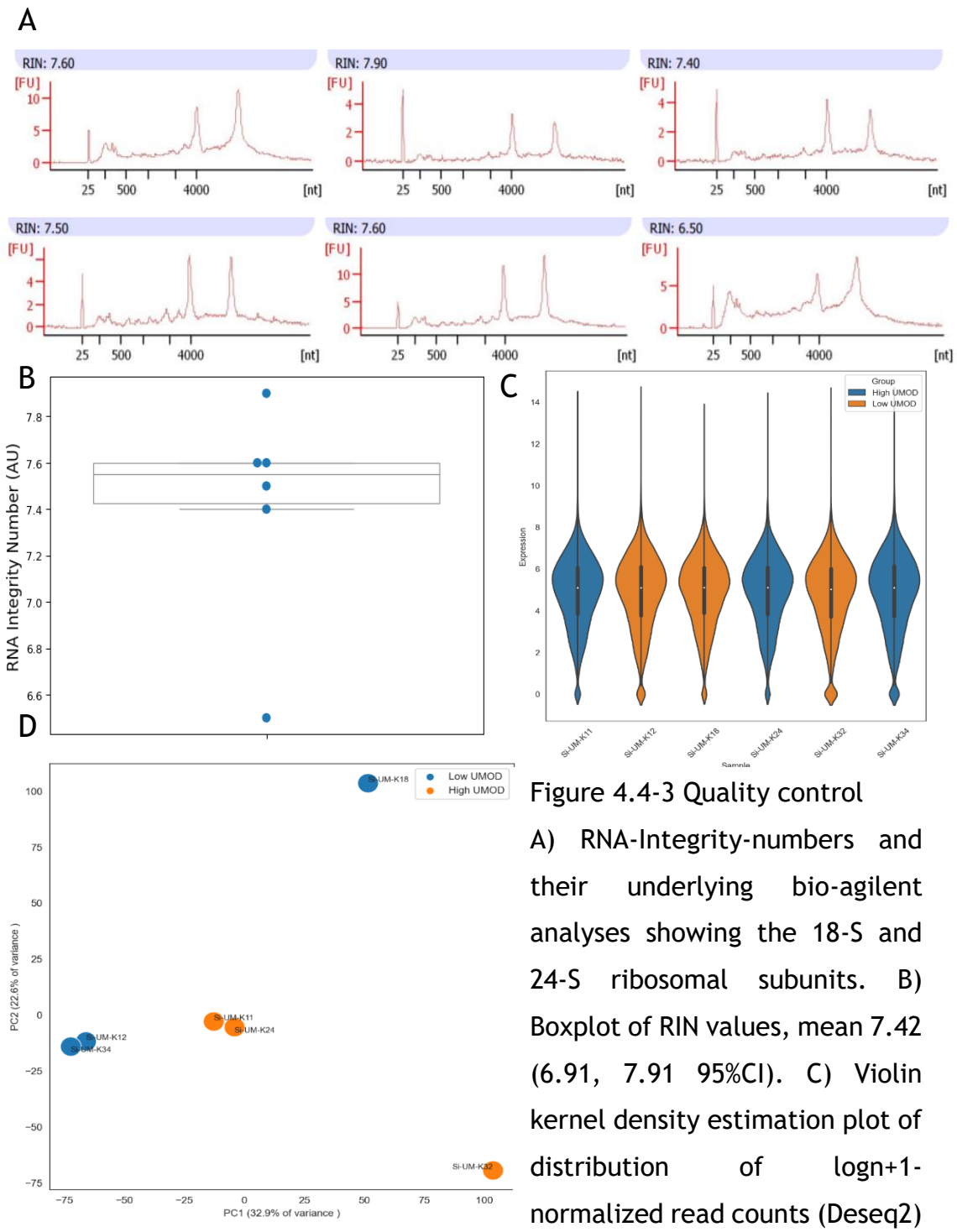


Figure 4.4-3 Quality control A) RNA-Integrity-numbers and their underlying bio-agilent analyses showing the 18-S and 24-S ribosomal subunits. B) Boxplot of RIN values, mean 7.42 (6.91, 7.91 95%CI). C) Violin kernel density estimation plot of distribution of logn+1-normalized read counts (Deseq2) D) Principal component analysis (log) of distribution of normalized (sklearn) read counts per sample. PC1 visualized on x-axis (32.0% of variance), PC2 visualized on y-axis (22.6% of variance).

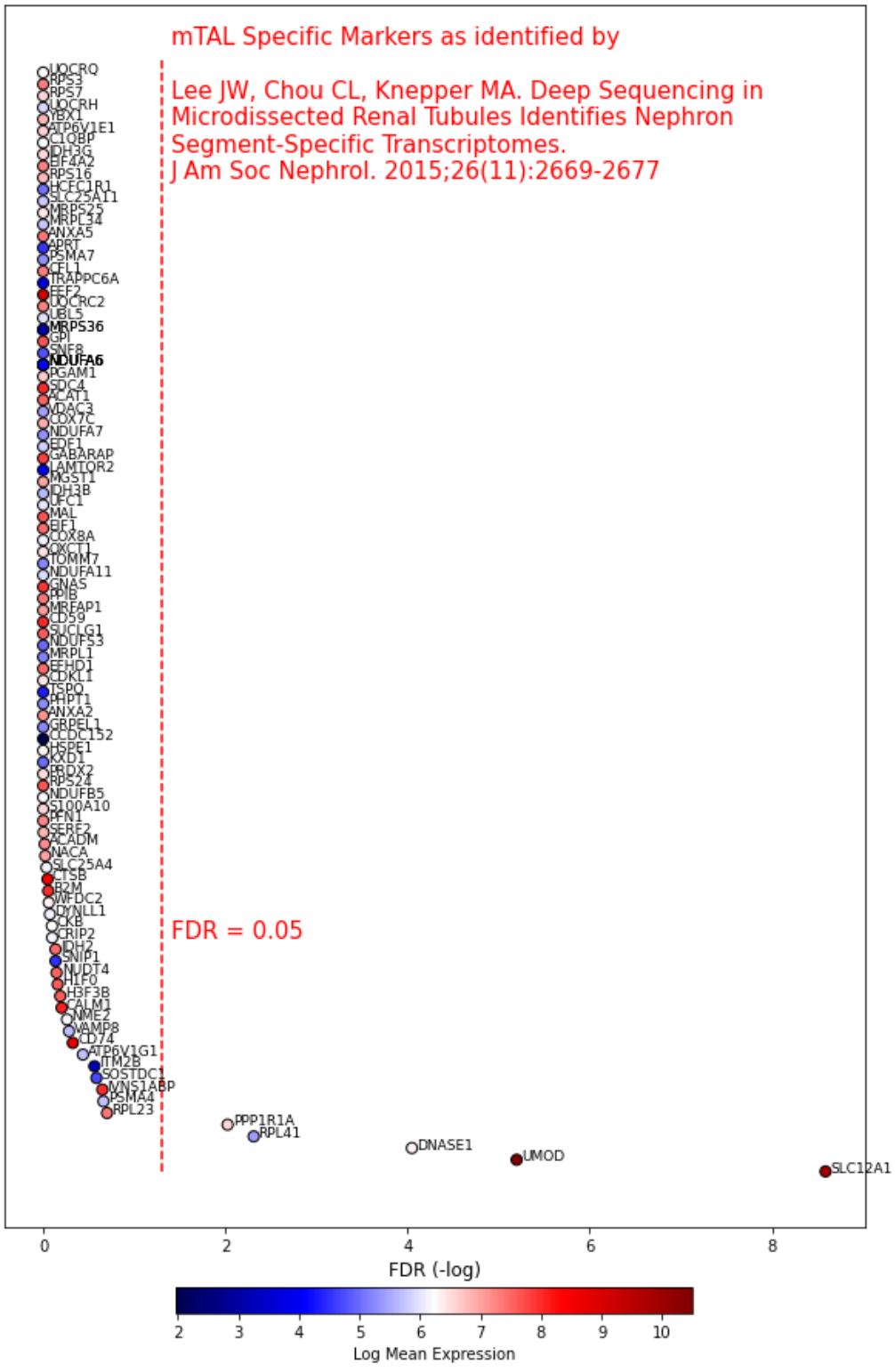


Figure 4.4-4 Assessment of biopsy location by transcriptomics

Adjusted p-values of all mTAL specific markers (Lee et al 2015). 5/96 mTAL markers were differentially expressed between group. Points coloured by logMean expression (n=3 group,2 groups).

4.4.3 Analysis of RNA-Sequencing Data

A total of 34760 genes were detected in at least one sample, of the total of 61393 currently annotated genes. The expression of genes followed the expected exponentially decaying signal, such that the majority of genes had zero of low expression counts which some genes had very high counts. A total of 163 genes were differentially expressed (DE) (FDR < 0.05) between high and low UMOD groups. The distributions of differentially expressed genes by chromosome was uniform across all gene biotypes. Of these differentially expressed genes, 39 were decreased and 124 were increased in the high-UMOD group. In total, 19 differentially expressed genes (DEGs) were lncRNAs, 139 were protein coding and the final five were a mixture of unprocessed pseudogenes and small-nuclear (SN)-RNAs (Figure 4.4-5)(Table 4.4-1)(Figure 4.4-6). Of the 19 lncRNAs differentially expressed, two of these, TMEM72-AS1 and LNC01762 had GWAS hits for diastolic blood pressure as described by GWAS catalogue. Analysis of lncRNAs located on chromosome-16, the same chromosome as UMOD (and therefore with cis-acting potential), indicated that whilst several were differentially expressed, none of these were within 50Mb of the boundaries of the UMOD gene (Figure 4.4-7).

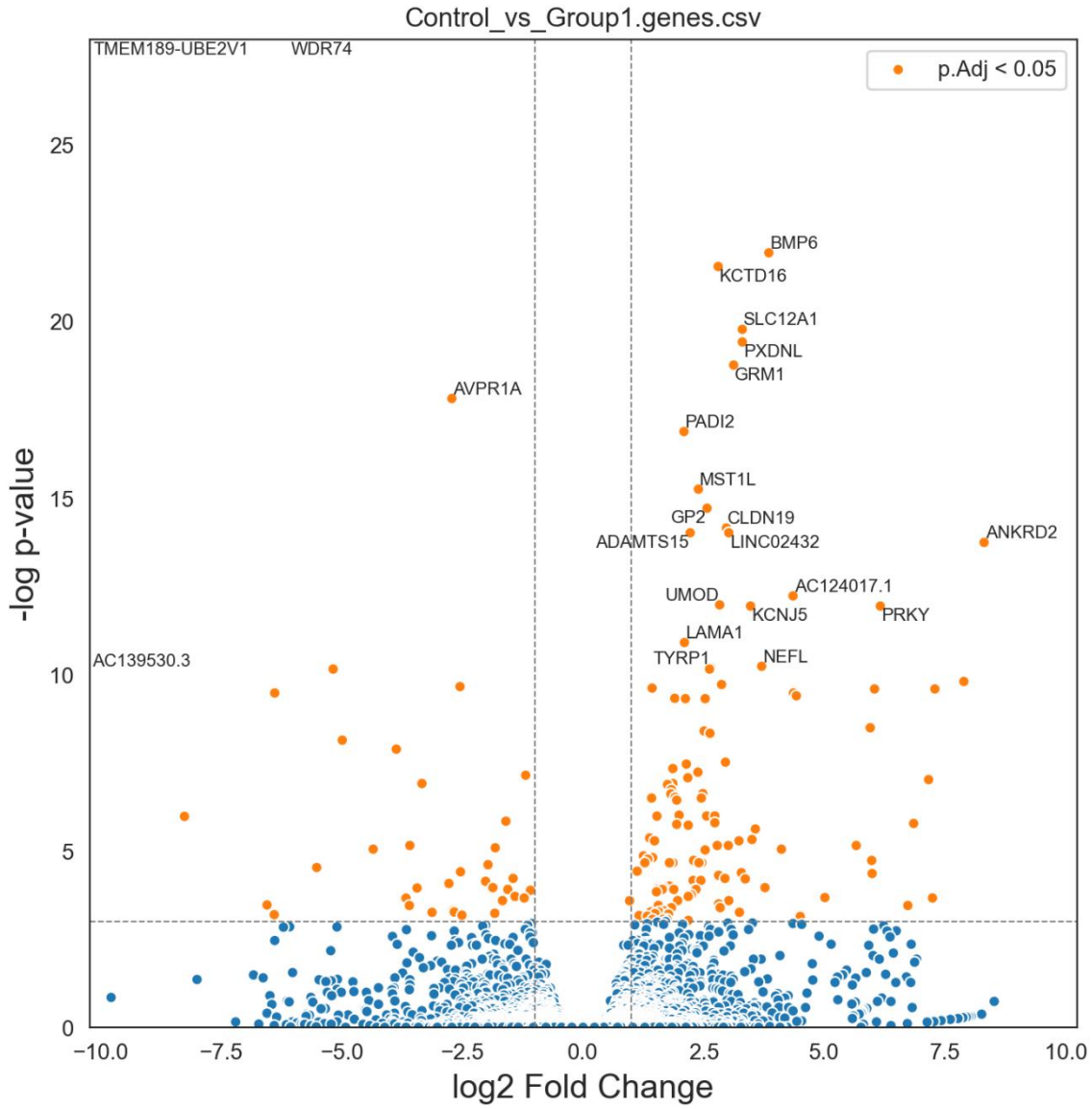


Figure 4.4-5 Volcano plot of high versus low UMOD.

genes, stratified on the y-axis by False-discovery-rate (BH) (-log transformation) and on the x-axis by Fold Change in gene expression between groups (log transformation). Guidebars (grey dashes) indicate significance thresholds of FDR 5% and log2 Fold Change values of -1, +1. The ensemble gene-names of the 25 most significantly differentially expressed genes are shown with labels. (n=3 group,2 groups).

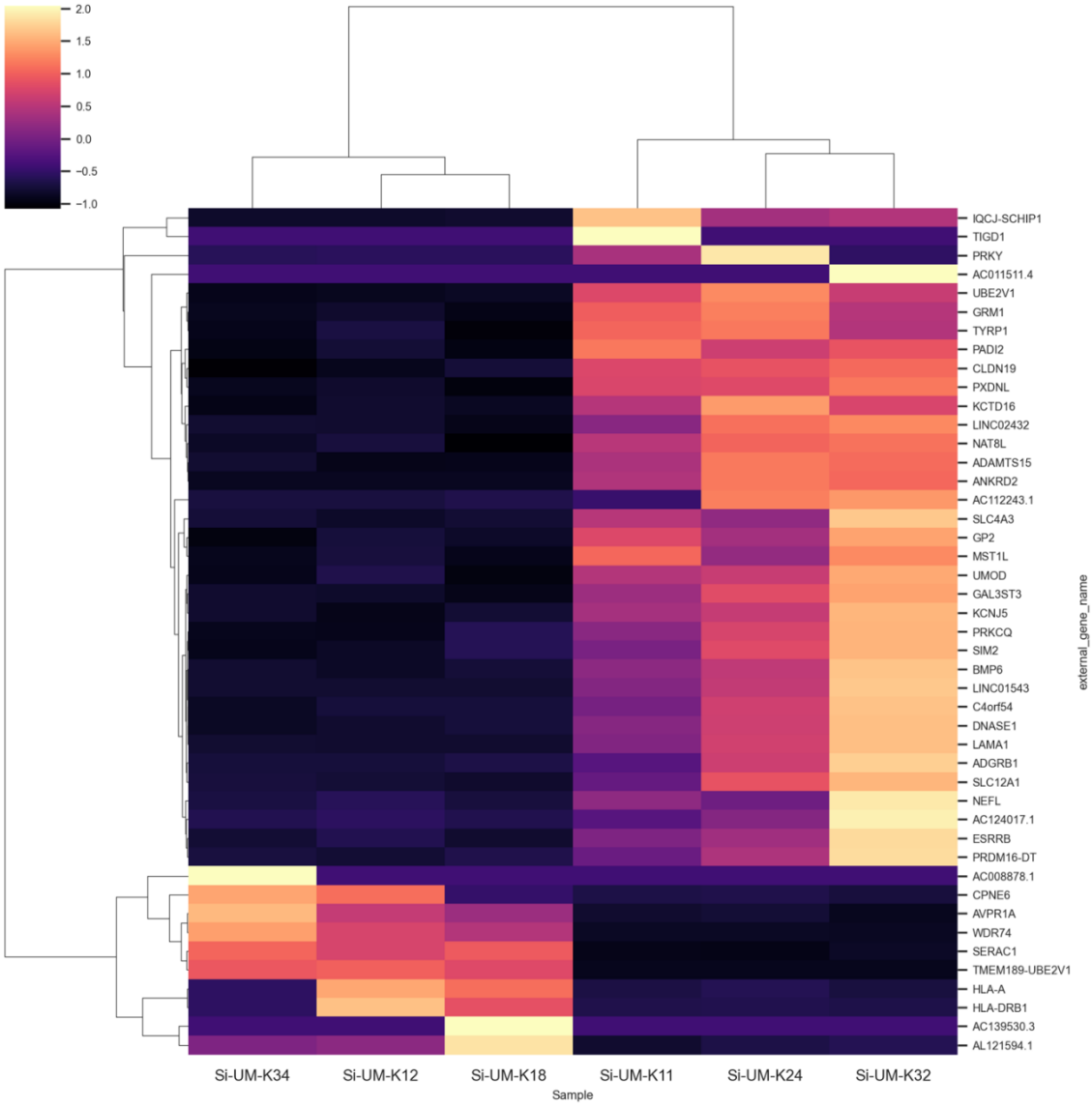


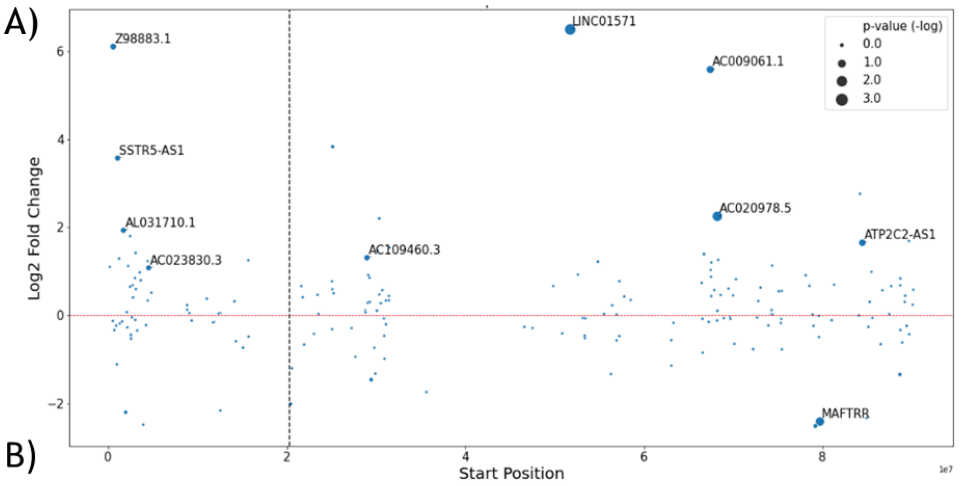
Figure 4.4-6 Clusterplot of high versus low UMOD.

Subsetted the most differentially expressed genes within the dataset (thresholded at p=1e-5). Calculated dendrograms are shown by black connecting lines and gene labels are shown on the right axis. Gene expression values were converted to within-gene z-scores to provide colour intensity

High vs Low UMOD RNA-Sequencing Analysis		
Gene Name	**Log2- Fold Change**	**Adjusted p-value**
WDR74	-6.06	5.52E-56
TMEM189-	-12.10	4.82E-19
BMP6	3.85	2.95E-10
KCTD16	2.80	4.35E-10
SLC12A1	3.30	2.58E-09
PXDNL	3.30	3.69E-09
GRM1	3.12	7.1E-09
AVPR1A	-2.72	1.83E-08
PADI2	2.09	4.67E-08
MST1L	2.39	2.39E-07
GP2	2.57	4.1E-07
CLDN19	2.97	7.19E-07
ADAMTS15	2.22	8.22E-07
LINC02432	3.02	8.22E-07
ANKRD2	8.31	1.08E-06
AC124017.1	4.35	4.91E-06
UMOD	2.83	6.34E-06
PRKY	6.16	6.56E-06
KCNJ5	3.47	6.56E-06
LAMA1	2.10	1.84E-05
NEFL	3.70	3.61E-05
TYRP1	2.65	3.78E-05
AC139530.3	-21.40	3.91E-05
PRDM16-DT	2.62	3.91E-05
AC008878.1	-21.40	3.91E-05
CPNE6	-5.18	3.91E-05
TIGD1	21.20	0.000051
LINC01543	7.89	5.55E-05
AC011511.4	21.00	6.05E-05

Table 4.4-1 Top 30 differentially expressed genes between high and low UMOD. Positive log2 Fold Changes indicates increased expression in high UMOD group. Adjusted p-values indicated benjamini-hochberg corrected p-values (Deseq2)

A)



B)

High vs Low UMOD RNA-Sequencing Analysis

Gene Name	Adjusted p-value	Blood Pressure GWAS
LINC02432	0.00000082	No
AC124017.1	0.00000491	No
PRDM16DT	0.00003910	No
LINC01543	0.00005550	No
TMEM72AS1	0.00251000	GWAS_Hit
LINC02437	0.00304000	No
CLDN10AS1	0.00579000	No
AC079310.1	0.00884000	No
KCNIP4-IT1	0.01250000	No
LINC02381	0.01860000	No
LINC01187	0.01910000	No
AC084125.2	0.03030000	No
AC104809.2	0.03170000	No
CYP4A22AS1	0.03170000	No
LINC01762	0.03370000	GWAS_Hit
LINC00659	0.04330000	No
AP000688.2	0.04690000	No

Figure 4.4-7 LncRNA analysis between high and low UMOD expressors

A) Analysis of lncRNAs in cis-proximity to UMOD. X-axis, chromosome-16 coordinates (100 Mbp ticks). Y-axis, level of differential expression (log2 fold change). Red horizontal line indicates 0-fold-change difference. Black horizontal dotted line indicates position of UMOD on chromosome-16. B) Table of all differentially expressed lncRNAs across all chromosomes with annotation as to any which had a previous GWAS association for blood pressure.

4.4.4 Pathway Analysis

Gene set enrichment analyses of data against KEGG and Gene-Ontology reference sets indicated a strong enrichment of DEG pathways toward transcriptional pathways and functions, with 24 individual processes identified (Figure 4.4-8)(A). The primary enriched process within gene transcription was the upregulation of RNA-polymerase II activity and binding in response to high UMOD. However, globally, significant pathways implicating gene expression, particularly in the context of RNA-polymerase II activity were ambiguous in terms of directionality, with both positive (6) and negative (18) normalized enrichment scores observed. Together these findings suggest an association between UMOD levels and transcriptional activity, though the potential specific gene targets are not described. Aside from transcriptional regulation, the primary schema identified by GSEA focused on processes involving the regulation of ionic gradients in cells, with 24 pathways identified. The majority of these (23) were upregulated in correlation with high-UMOD levels and transmembrane movement of 'cations', 'metals', 'chloride', 'calcium', 'hydrogen' and 'anions', though these gene sets also describe the transport of protein and carboxylic acids. Conversely, one pathway was highlighted as down-regulated in response to high-UMOD levels; sodium ion transmembrane transporter activity (Figure 4.4-8)(B).

A total of seven individual gene sets were described in detail (Figure 4.4-9) In particular, there was an enrichment of signaling pathways in correlation with changes in UMOD levels. TNF-α, TGF-ß and NF-kß signaling pathways were all differentially regulated. Additionally, the WNT and cAMP pathways were also implicated. Finally, data also suggest the alternate regulation of the RAAS system through vasopressin. In terms of IPA analysis shown in Figure 4.4-10 indicated POU3F3 was highlighted through *in-silico* mining as a transcription factor which may underpin the expression of both NKCC2 (SLC12A1) within these cells; although this transcription factor itself is not a DEG. Multiple DEGs facilitating natriuresis within TAL cells were detected in IPA alongside SLC12A1 (NKCC2); the sodium epithelial subunit 1-gamma (SCNN1G) and the Potassium Inwardly Rectifying Channel Subfamily J Member 1 (KCNJ1)(Figure 4.4-11). Two DEGs, SGK1 and WNK4, were highlighted as potentially interacting as signaling molecules, participating in the activity of natriuretic DEGs within the data.

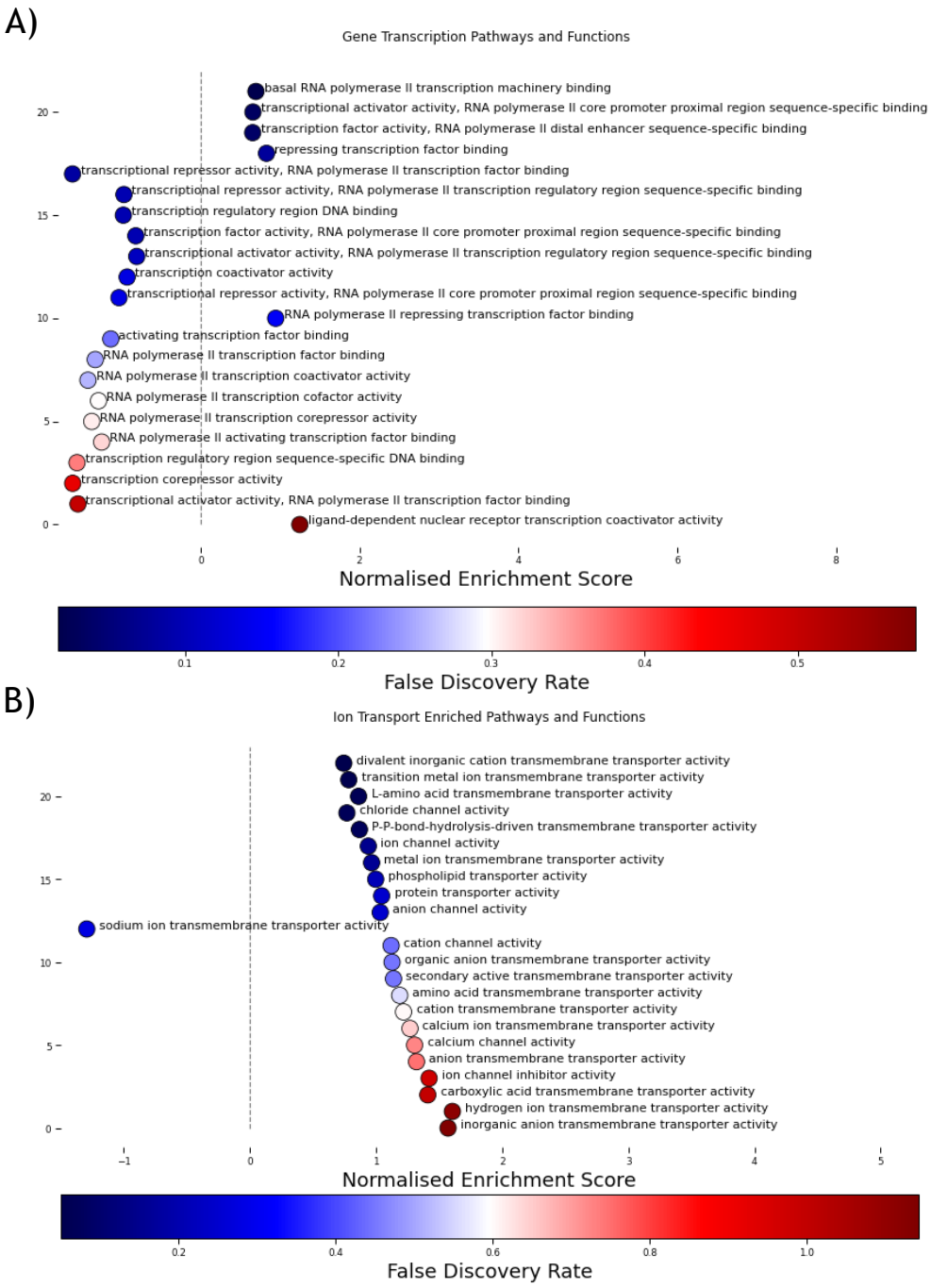


Figure 4.4-8 Gene set enrichment analysis between high and low UMOD

(A) Gene set enrichment arrays were filtered specifically through keywords 'transcription' by python regex, and visualised with enrichment score (x-axis) and FDR (colour intensity). (B) Gene set enrichment arrays were filtered specifically through keywords ' channel|transporter' by python regex, and visualised with enrichment score (x-axis) and FDR (colour intensity).

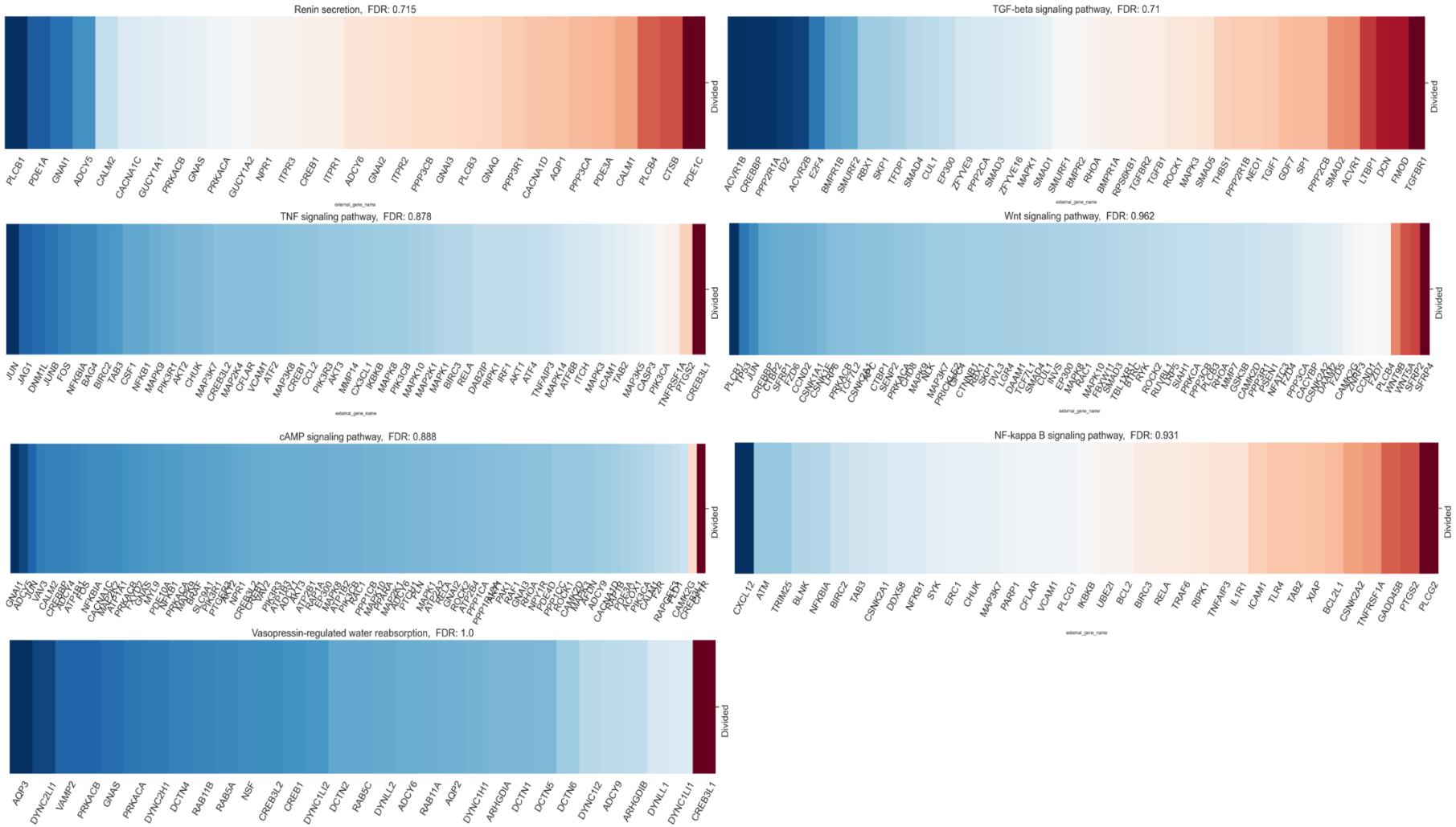


Figure 4.4-9 Specific enriched gene-sets obtained from Gene Ontology 2021.

Header describes the implicated process or function, genes are displayed below plots and the difference in counts of these genes between low and high UMOD are displayed by colour (blue= higher mean expression in low-UMOD, red= higher mean expression in high-UMOD). As this is a whole-sample GSEA, not all of the genes displayed within the pathway are statistically differentially expressed but they do contribute to the GSEA model.

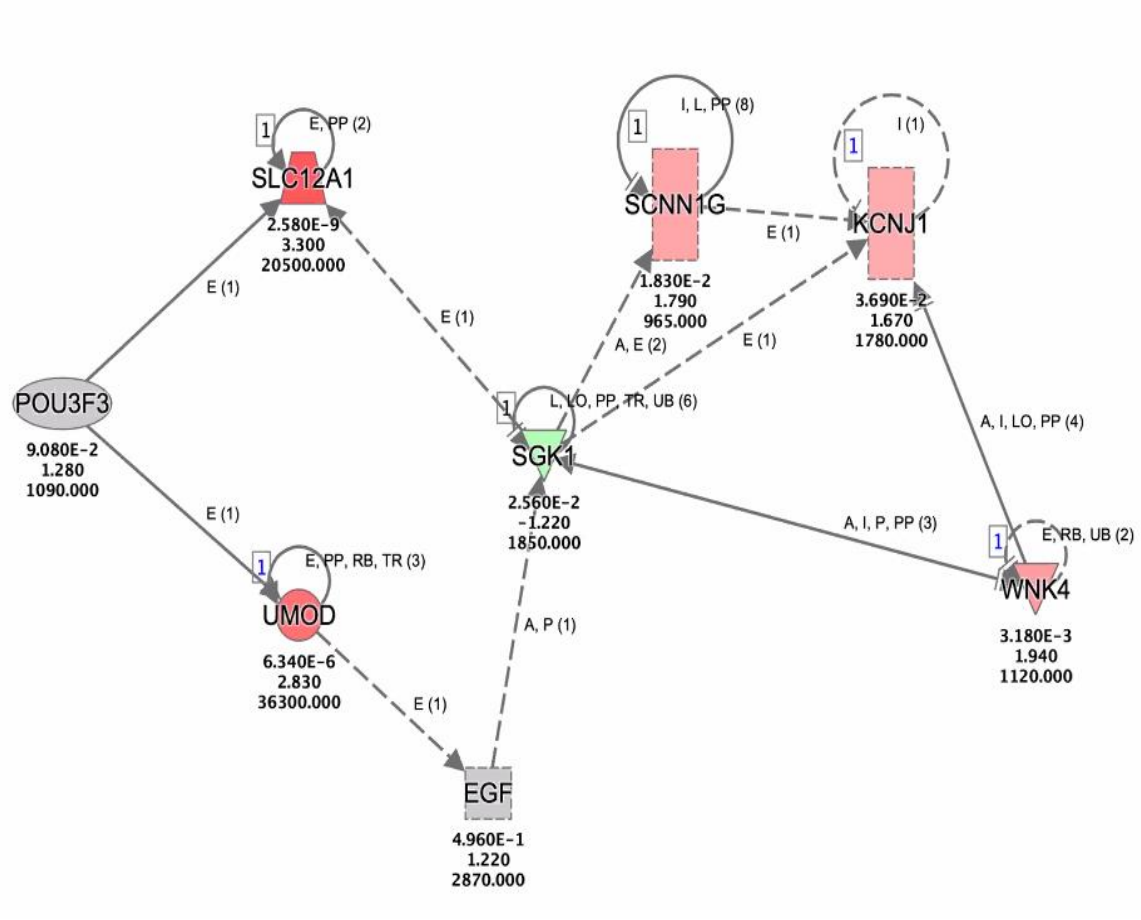


Figure 4.4-10 Ingenuity pathway analysis of data-mined connections between differentially expressed genes.

Genes which positively correlate with UMOD expression are coloured red, those with negative correlation are coloured green. The top-most numerical variable per gene indicates FDR adjusted p-value, the middle variable indicates log2-fold change, the bottom-most variable indicates mean expression in control (low-UMOD) group. Filled lines indicate published connections, dashed lines indicated *in silico* inferred connections.

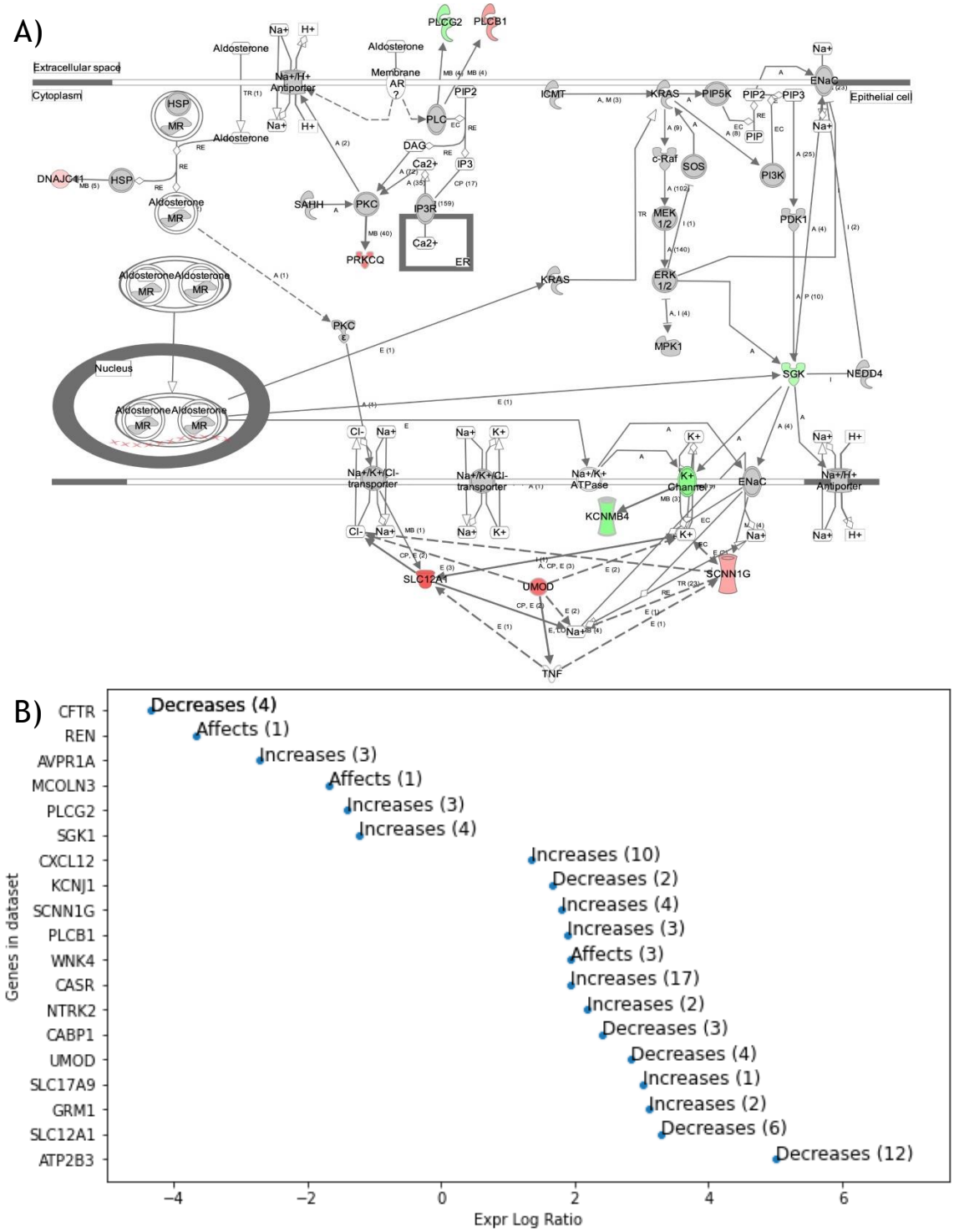


Figure 4.4-11 Ingenuity pathway analysis, cellular functions

A) IPA visualization of TAL cell ion handling functions with differentially expressed genes coloured by log2fold directionality. Genes which positively correlate with UMOD expression are coloured red, those with negative correlation are coloured green. B) IPA identified genes for enrichment phrase “movement of metal ion”, stratified by log2fold expression. Labelled by direction of enrichment.

4.4.5 Co-expression Analysis

Scanpy-based single-cell analysis (source code in 4.6.2) of five independent healthy human renal samples displayed clear $UMOD^+$ clustering within cell populations, presumably indicating these cells as either TAL or dCT components. Coexpression analysis was performed on these 5 samples by analysing the linear regression parameters of each gene against that of UMOD, on a cellular basis (Figure 4.4-12). As this is a fundamentally novel analysis method, the p-values of these Pearsons correlations were not adjusted for multiple testing due to the inherent variability and the risk of filtering interesting targets. To increase the signal to noise ratio of this analysis, only coexpressed signals which were detected in more than one sample were considered. The top 30 genes coexpressed at the single cell level with UMOD are shown in Table 4.4-3. Of these genes, WNK1 was the most significantly co-expressed, with an uncorrected p-value < 0.00001 and detectable expression in 71% of UMOD cells.

The coexpression between UMOD and NKCC2 (SLC12A1) which was observed in analysis of both bulk and single cell was verified using RNA from 84 human samples (described in 2.1-0) by Taqman qRT-PCR. A highly significant positive correlation in expression was detected by linear regression analysis (****, $p<0.0001$) (Figure 4.4-13).

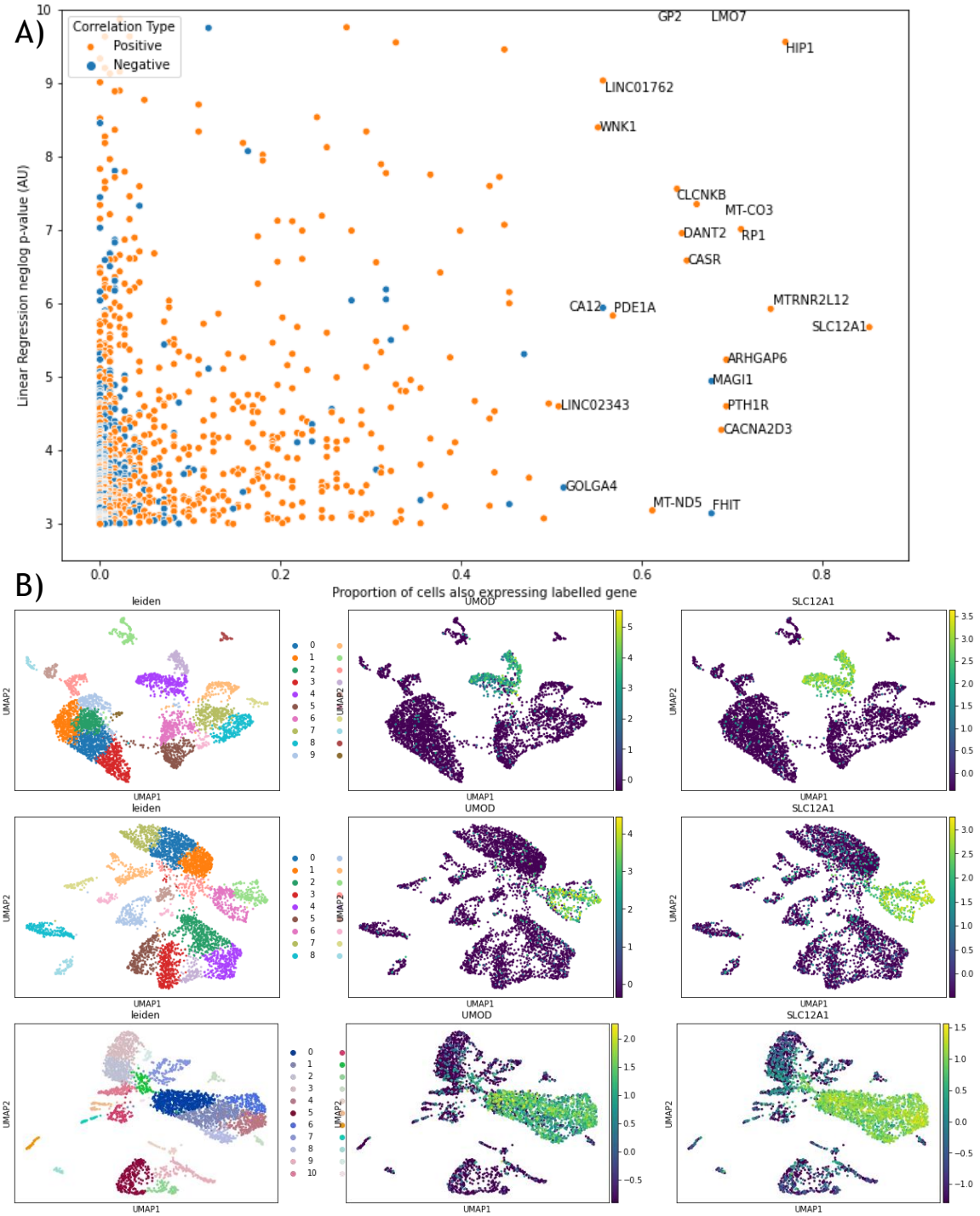


Figure 4.4-12 A) Example plot from one human scRNA-seq dataset analysed. Y-axis shows the p-value of the linear regression analysis between the target gene and UMOD (-log adjusted). A higher value is more significant. X-axis shows the proportion of UMOD+ cells which also express the target gene. A higher value here indicates a more robust correlation. B) Example visualisations of the significant co-expression between UMOD and NKCC2 at single-cell level.

Single Cell Coexpression Analysis		
Gene Name	**Number of Occurrences**	**Mean P-value**
WNK1	2	7.49E-07
CLCN5	2	5.46E-05
SLC12A1	2	0.000714668
DAPK1	2	0.000716177
AC019197.1	2	0.000733789
WNK4	2	0.00126575
PHACTR1	2	0.002679055
UGT8	2	0.003682381
THSD4	2	0.004895217
HIP1	4	0.0051493
CTDSPL	2	0.006152223
SGIP1	2	0.006185742
EFHD1	2	0.007186302
ACPP	2	0.007517833
PDE1A	2	0.007699149
GP2	3	0.008800284
RNPC3	2	0.009290538
AC092078.2	2	0.009360066
LINC01762	3	0.01153034
DNAH14	2	0.013810497
CACNA2D3	4	0.014578991
ATP1A1	2	0.016340168
LINC01606	2	0.017306249
STK39	2	0.017926789
SUSD4	2	0.019716807
BPTF	2	0.026456246
RALYL	2	0.027311588
CLCNKB	3	0.02959887
HS6ST2	2	0.029870208
LINC02343	3	0.03170051
ARHGAP6	2	0.032995793
KLHL13	2	0.033718924
MAML2	2	0.034779892
RP1	4	0.040472588
MAGI1	2	0.041484078
CASR	4	0.043297544

Table 4.4-3 Five human renal single cell datasets were analysed for linear regression coexpression with UMOD. Each coexpression matrix was filtered into genes which were detected in more than one dataset as coexpressors and p-value < 0.05. The number of significant occurrences of coexpression (/5) are shown in C2. Mean Pearson's test p-values are shown in C3 (non-adjusted p-values).

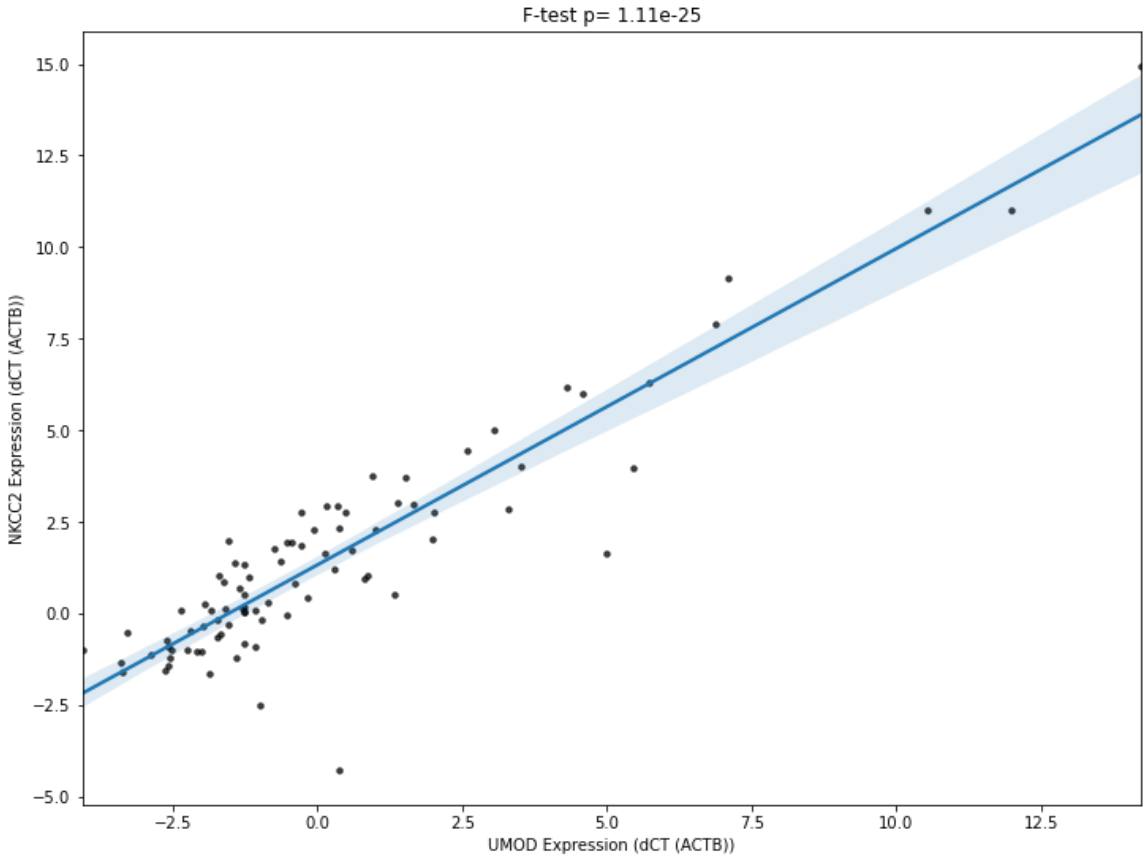


Figure 4.4-13 Linear regression of UMOD expression (dCT) against NKCC2 expression (dCT) in RNA extracted from human renal samples (n=84).
Blue filled area indicates 95% CI of the ß-coefficient. Pearson's correlation ****, p<0.0001.

4.5 Discussion

By stratifying human renal samples on expression of UMOD and then assessing the differences in the renal transcriptome between these groups, we show a strong, positive correlation between expression of UMOD and genes involved in natriuretic mechanisms. This provides, for the first time, transcriptomic evidence relating UMOD to the regulation of blood pressure. Our findings suggest that UMOD associates with blood pressure through several potential pathways, which may contribute varying effect-sizes to the relationship. Furthermore, we corroborate our bulk data with analysis of single cell data, from independent datasets. We believe that our most important finding is the detection of coexpression between UMOD and NKCC2 that we detected in both bulk and single cell transcriptomics. Building on this finding, we generated evidence of a significant relationship between UMOD expression and expression of NKCC2 in a qRT-PCR validation study (n=84). We emphasise that while these data do not show causation, they provide a number of pertinent indications as to the possible systematic effect of UMOD with respect to blood pressure regulation.

As discussed in 3.5, the majority of previous studies implicate the transcription factor HNF1β as the primary regulator of UMOD expression, with *Hnf1β* mice showing reduced *Umod* expression, and a binding site of Hnf1β on the *Umod* promoter at both −1.1 and −0.58 kb from the TSS shown in mice (Gresh *et al.*, 2004; Igarashi *et al.*, 2005; Han *et al.*, 2013; Verhave *et al.*, 2016). Interestingly, HNF1β has not been characterized in the literature as associating with expression of NKCC2; suggesting it does not factor in the coexpression we observed. However, it is unlikely that UMOD is under the explicit, singular control of this transcription factor. Our IPA analysis highlights POU class 3 homeobox 3 (POU3F3) as a potential coregulator of both UMOD and NKCC2. Crucially to this hypothesis, Pou3f3 (Brn1) knock-out mice have been shown to have reduced expression of *both Umod* and Nkcc2 (Nakai *et al.*, 2003; Kumar *et al.*, 2016; Rieger *et al.*, 2016). It should be noted that Reiger et al analysed blood pressures between *Pou3f3* missense mutation cohorts and did not observe a difference between these and wild type animals. However, Pou3f3 has been shown to exist in a quantitative trait locus for blood pressure on the rat genome, spanning 40Mb on Chromosome-9 ((Blood pressure QTL 392) Rattus norvegicus, 2021). In our

data, POU3F3 was not differentially expressed between the low and high UMOD groups. However,we do not believe this lack of differential expression to be an indication of a lack of relationship. Rather, we hypothesise the potential mechanisms between differential expression of both UMOD and NKCC2 between groups are likely under-pinned by more cryptic processes involving POU3F3. We would suggest the most likely mechanism here being that POU3F3 coding mutation(s) may induce differential binding to the RE on both the UMOD and NKCC2 promoter between the high and low UMOD groups. Less likely, though also possible, is that variants on both the UMOD and NKCC2 promoters drive this differential coexpression. In theory, if it can be shown that POU3F3 binds to a RE on the UMOD promoter that centers on rs4997081 we believe this may be a causative finding as to the UMOD:blood pressure hypothesis, however, this has not been proven to date and our preliminary analysis by ALGGEN-TRANSFAC prediction does not suggest a binding site at this locus. Fundamentally and empirically, our novel data here suggest that POU3F3 is likely to be the primary driver of coexpression between UMOD and NKCC2. To further build on these findings, we would suggest both a targeted CHiP-Seq of POU3F3 binding sites and additionally anti- Pou3f3 siRNA treatment in mice. Such experiments could show both potential promoter interactions and downstream physiological effects.

Following from our observations of UMOD and NKCC2 coexpression, we identified several genes and pathways associating differential UMOD expression with blood pressure regulation. By IPA we identified serum/glucocorticoid regulated kinase 1 (SGK1) levels as negatively correlated with expression of NKCC2 and SCNN1G. SGK1 is a kinase known to drive differential activation of renal ion transporters in humans, with variants (MAF 3-10%) identified as correlating with differences in blood pressure in humans (Lang *et al*, 2010)(Van Beusecum *et al.*, 2019). Interestingly, it has been noted that the effect of SGK1 on the activation of ion transporters in renal cells is not fundamental to their function, rather that SGK1 acts as a secondary regulator which may drive pathophysiology as opposed to basic function (Lang and Shumilina, 2013). By IPA we show that the relationship between UMOD and these ion transporters/channels, is secondary through EGF, which has itself been shown to play a role in salt sensitive hypertension (Ying and Sanders, 2005;

Graham *et al,*. 2014a). Within this pathway matrix, we also identified lysine deficient protein kinase 4 (WNK4) as differentially expressed. WNK4, a renally expressed kinase similar to SGK1, is also heavily implicated in the activation of ion channels and transporters (Lang and Shumilina, 2013; Terker *et al.*, 2018). Furthermore, human studies have shown a relationship between WNK4 and blood pressure (Murthy, Kurz and O'Shaughnessy, 2017). Additionally, analysis of *Wnk4*$^{-/-}$ mice found these animals were highly hyponatremic and additionally displayed both decreased *Nkcc2* phosphorylation and decreased absolute quantities of *Nkcc2* (Terker *et al.*, 2018). Finally, cross-phosphorylation between WNK4 and SGK1 has also been detected, suggesting feedback mechanisms may exist to further enhance pathophysiological states and provide homeostatic correction to aberrant pathways (Ring *et al.*, 2007).

By gene set enrichment analysis, we detected enrichment within our DEGs for pathways implicating TNF-α, TGF-ß and the RAAS system in association with differential UMOD levels between groups. Based on our Chapter 3 hypothesis, we predicted the enrichment of TNF-α signaling, as we theorised TNF-α was fundamental to the expression of UMOD. We believe our findings in this chapter reinforce those of Chapter 3, as we directly implicate both TNF-α and NF-Kß signaling, independently of our Chapter 3 experiments. It is possible that the differences in TNF-α and NF-Kß signaling we detected between the high and low UMOD expressors in this study may actually be underpinned by the mechanism we suggest in Chapter 3. Furthermore, the findings in this chapter suggest to us that the inhibitory activity of TNF-α in TAL cells correlates with levels of UMOD, suggesting a possible regulatory or negative feedback mechanism here with TNF-α signaling itself dependent on UMOD levels.

Both our observations of the TGF-ß and RAAS system in the context of UMOD expression can be considered 'non-canonical' as they are based on novel findings from this study. Interestingly, step-wise changes in TGF-ß expression in mice in five steps from 10% to 300% normal was shown to correlate positively and proportionately with both blood pressure and sodium transport in the kidneys of these animals (Kakoki *et al.*, 2013; Matsuki *et al.*, 2014). It is currently hypothesized that TGF-ß signaling alters blood pressure via aldosterone and thus is a cofactor in the RAAS system (Kakoki *et al.*, 2013). The finding that the RAAS system appears to associate with UMOD levels

was unexpected, with limited published evidence to support this finding. However, we would emphasise that potential synergy or inhibition between UMOD and RAAS should be considered as functionally relevant in TAL cells. Arginine Vasopressin Receptor 1A (AVPR1A), a critical receptor within the RAAS system, was the third most negative DEG within our dataset; decreasing in expression with increasing levels of UMOD. It is possible that the significant downregulation of AVPR1A provides evidence of homeostatic adaptations to the increase natriuretic load induced by elevated UMOD levels in these cells.

Analysis of coexpressed genes by single cell datamining using our custom self-authored python applications detected a number of targets which cross-validate the findings observed in our bulk data experiment. Specifically, we detected coexpression of both WNK1/4, alongside NKCC2. Interestingly, the calcium sensing receptor (CASR) was also detected by this analysis; as it was within the bulk dataset also. The excretion of UMOD is known to be dependent upon activation of intracellular CASR (Tokonami *et al.*, 2018b). Whilst we do not believe that the CASR is a prominent factor in the UMOD:blood pressure hypothesis, we do believe that it is notable that CASR was enriched multiple times within our data analytics for this chapter. Of single-cell coexpressed genes, the lncRNA LINC01762 appealed to us for further analysis because it was also a DEG within our bulk dataset, and furthermore due to its strong consistency of enrichment in single cell analysis, with 4/5 datasets detecting this gene as coexpressed. Significantly, rs12078697-C on LINC01762 has been shown to associate with diastolic blood pressure in humans (GWAS Catalogue, 2021). We performed a burrow-wheeler alignment for each exon (plus reverse complement) of LINC01762 to determine if there were any potential interactions with other DEGs within our dataset, and found none. However, the cryptic nature of lncRNA effects on downstream targets means we do not rule out causation based on a lack of direct interaction

In terms of limitations, our primary aim in Chapter 4 was to address the transcriptomic relationship between UMOD *genotype* and renal pathways. In these experiments we assessed the relationship between UMOD *levels* and renal pathways. Had it been possible, we would have targeted both homozygous alleles of rs4997081;

this would have provided a more clinical, precision-medicine centric interpretation of the data. However, our grouping and selection protocol (combined with the low prevalence of homozygous 'protective' alleles) resulted in us not having access to adequate tissues for this experimental design. Additionally, our RNA was consumed during the transcriptomic application due to its inherently very low yield (we received on average a yield of >1µg total RNA per sample). Such limitation in tissue availability produced caveats to qRT-PCR based follow-up/ validation study of gene expression on the same human samples. Furthermore, we would like to note that this tissue was biopsied from the healthy pole of cancerous kidneys. As a result, it should be considered that our experiment is not specifically representative of an a-morbid, healthy population. It is reasonable to suggest that stress-response and immune pathways within these samples is likely to be upregulated versus true healthy controls due to both the proximity of a tumour and the therapeutic interventions administered to these patients. Lastly, in terms of caveats, we acknowledge that we did not apply multiple-testing-correction to our single-cell linear regression application. We feel that doing so when the application exists in a nascent state would be deleterious to our initial interpretations. The data, in theory, require a significantly more robust normalization strategy than the simple log-expression normalization applied. Inherently reducing the range and variability in expression of each cell within the data with appropriate normalization would increase the power of each regression analysis. We do not interpret the single cell data as 'significant', but rather, 'indicatory' of coexpression events.

In future, we believe an RNA-sequencing experiment stratified on UMOD genotype and not expression levels may provide highly relevant data in terms of informing the hypothesis with respect to blood pressure regulation at the UMOD locus. We would also emphasise that such a study should attempt to use samples which have an abundance of material. We would stress that, if possible, this material should yield enough to stratify across both qRT-PCR and western blot experiments as mRNA differences are not always perfectly correlated with protein differences. Such a resource of human tissue is likely to be highly challenging to procure. Therefore, it may be advisable to translate this study to *in vivo* using mouse or rat tissue, though,

in doing so this would prevent stratification across genotype as neither rs13333226 nor rs4997081 exist on the mouse or rat UMOD promoters. Lastly, of our complement of DEGs from this study, an appreciable proportion (19/163) were lncRNAs, and of these 2/19 had previously attributed GWS to blood pressure. LncRNA annotation, particularly in terms of diseases and functions, is generally poor but we believe the contribution of lncRNA with our dataset is not negligible and it should be considered that at least one or several of the lncRNAs we detected may play a causative role in blood pressure regulation via UMOD. LncRNAs are widely theorized to play a significant (and targetable) but 'dark-matter' style role in blood pressure regulation (Murakami, 2015; Collins *et al.*, 2020; Jiang and Ning, 2020). Therefore, we recommend that the use of specific siRNAs against our set of lncRNAs followed by the monitoring of UMOD expression may be potentially relevant downstream experiments.

All transcriptomics visualisations in this thesis were made with the self-authored Padplot. The impact of Covid19 was substantial as it caused significant lab closures, with 4-months of absolute closure and 6 months of additional disruption. During this time, the author worked on building and deploying Padplot. Preexisting bioinformatics tools for analysing bulk RNA-Seq data exist, therefore Padplot was conducted with a specific set of goals which would allow it to occupy a niche and provide novelty. Firstly, to be free and open-source, so that it can be used by anyone. In this way, Padplot superseded both subscription based tools (which cost money) and Galaxy (which is free but cannot be modified). Secondly, Padplot was required to be lightweight, so that it can be deployed cheaply on Amazon storage with no costs incurred, to protect the first aim. Padplot, due to its Streamlit integration, incorporates <1500 lines of code and furthermore only requires <10 external libraries to run making it lightweight, significantly more so than other open-source tools which do not deploy directly to a browser. Finally, Padplot was required to be used by anyone, but in particular, wet-lab biologists who do not have experience with code. Therefore, Padplot was not written to run directly in python, nor did it come in a docker container, but rather it could be accessed using an IP address. These specifics (the easily accessible source, Padplot being free and with a shallow learning curve),

made Padplot a useful tool, as was consistently fed-back by users. However, in December of 2021, users began to incur the author with costs from Amazon, therefore Padplot was discontinued until funding can be found to restart the server- as charging customers a fee is out-with the remit of the work.

The transcriptomic work conducted within this chapter reinforces previous work both internally at the University of Glasgow and in the wider academic community, linking UMOD to blood pressure. However, we show for the first time that it is likely that the nature of this relationship is multifactorial. We evidence the canonical pathway suggesting UMOD links to blood pressure through NKCC2 is likely the primary driver, and we also provide increased resolution on the relationship between TNF-α and UMOD expression, with our evidence indicating this is driven by NF-Kß activation. Furthermore, we build upon these findings by suggesting that UMOD may associate with the RAAS system, and that this may be driven through TGF-ß. We also provide seminal data suggesting lncRNAs may associate with blood pressure through UMOD. Finally, we emphasise that the significant contributions of these data to the understanding of UMOD and blood pressure would benefit from further validation through *in-vivo* and *in-vitro* follow-up study.

4.6 Chapter Specific Self-authored Code

4.6.1 PadPlot- Bulk Data

```
import numpy as np
import pandas as pd
import streamlit as st
import matplotlib.pyplot as plt
import seaborn as sns; sns.set_style("white")
import adjustText
from adjustText import adjust_text
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import gseapy as gp
from gseapy.plot import gseaplot, heatmap
import matplotlib.gridspec as gridspec
st.beta_set_page_config(layout="centered")
initial_sidebar_state="expanded"
def _max_width_():
    max_width_str = f"max-width: 1300px;"
    st.markdown(
        f"""
    <style>
    .reportview-container .main .block-container{{
        {max_width_str}
    }}
    </style>
    """,
        unsafe_allow_html=True,
    )
def loaddata():
    global df
    global headers
    global log2s
    global paddys
    global genenames
    global temp
    global gene
    global controlg
    global expg
    global title
    global pim
    st.header("Next point PadPlot to the Relevant Columns")


    df=pd.read_csv(data)
    #st.dataframe(df)

    headers= list(df.columns.values)
    title=data.name

def loaddata():
    global df
    global headers
    global log2s
    global paddys
    global genenames
    global temp
```

```
    global gene
    global controlg
    global expg
    global group1_title
    global group2_title
    global schema
    global statmethod
    global gene_set
    global permtype
    #st.dataframe(df.head(20))
    st.header("Next select a Global Style for Plots")
    STYLES = {
        "White":"white",
        "Dark grid": "darkgrid",
        "Dark": "dark",
        "White Grid":"whitegrid",
        "Ticks":"ticks",


    }
    schemes= list(STYLES.keys())
    sch = st.selectbox("Define a plot schema", schemes)
    schema= STYLES.get(sch)
    st.text("")
    st.text("")
    st.text("")
    st.header("Finally point PadPlot to the Relevant Columns")

    data.seek(0)
    df=pd.read_csv(data)
    #dfprocesses=pd.read_csv('mart_export.txt')
    headers= list(df.columns.values)

stat_methods=['signal_to_noise','t_test','ratio_of_classes','diff_of_classe
s','log2_ratio_of_classes']

gene_sets=['GO_Molecular_Function_2018','GO_Biological_Process_2018','KEGG_
2019_Human','MGI_Mammalian_Phenotype_Level_4_2019','miRTarBase_2017','WikiP
athways_2019_Human']


    col1, col2, col3= st.beta_columns(3)
    genenames = col1.selectbox("Select Gene Names", headers)
    paddys = col2.selectbox("Select Adjusted p-values", headers)
    log2s = col3.selectbox("Select Log2 Fold Changes", headers)

    controlg = st.multiselect("Select Control Group", headers)
    expg = st.multiselect("Select Experimental Group", headers)
    col4, col5= st.beta_columns(2)
    group1_title = col4.text_input("Define Group 1 Title Here", 'ie
"Control"')
    group2_title = col5.text_input("Define Group 2 Title Here", 'ie
"Experimental"')
    st.text("")
    st.text("")
    st.header("Customise Gene Set Enrichment Parameters (Human Data Only)")
    col6, col7, col8=st.beta_columns(3)
    statmethod = col6.selectbox("Gene Set Enrichment Method", stat_methods)
```

```
    gene_set= col7.selectbox("Gene Set Enrichment Reference Database",
gene_sets)
    permtype= col8.selectbox("Gene Set Enrichment Permutation Type
(Phenotype only if samples > 15", ['gene_set','phenotype'])

def selectplot():
    st.text("")
    st.text("")
    st.text("")
    options = [
        "PadPlot Volcano",
        "PadPlot Heatmap",
        "PadPlot PCA",
        "PadPlot Violin",
        "PadPlot Gene Set Enchrichment (Human Only)",


    ]
    #padplots= list(OPTIONS.keys())
    st.header("Choose a plot:")
    plot = st.selectbox("Define a plot schema", options)

    if plot== "PadPlot Volcano":
        volcanoplot()

    elif plot== "PadPlot Heatmap":
        heatmapplot()
    elif plot== "PadPlot PCA":
        pcaplot()
    elif plot== "PadPlot Violin":
        violinplot()
    elif plot== "PadPlot Gene Set Enchrichment (Human Only)":
        geneset_enrichment()
    else:
        st.header("No Plot Currently Selected")
def pcaplot():
    st.sidebar.header("Change PCA-plot Parameters")


    pca_df=df[controlg+expg]
    genes=len(pca_df)
    pca_df=pca_df.transpose()
    logdata = st.sidebar.checkbox('Log PCA Input Values')
    if logdata:
        pca_df=np.log((pca_df+1))
        title='Logged Values'
    else:
        pca_df=pca_df
        title='Raw Values'
    pca_df['Sample']=pca_df.index
    #st.dataframe(pca_df)
    labelsize=st.sidebar.slider('Size of text labels', 2, 20, 8)

    g1= [group1_title]
    g1=g1* len(controlg)

    g2= [group2_title]
    g2=g2*len(expg)
    grouping=g1+g2
```

```
    pca_df['Group']= grouping


    #st.dataframe(pca_df)
    features = range(0,genes,1)
    x=pca_df.loc[:,features].values
    y=pca_df.loc[:,['Sample']].values
    #x=StandardScaler().fit_transform(x)
    pca=PCA()
    principalComponents=pca.fit_transform(x)
    ex_variance=np.var(principalComponents,axis=0)
    ex_variance_ratio=ex_variance/np.sum(ex_variance)
    pc1=round((ex_variance_ratio[0] * 100), 1)
    pc2=round((ex_variance_ratio[1] * 100), 1)
    pcx= [i[0] for i in principalComponents]
    pcy= [i[1] for i in principalComponents]
    #pc2
    principaldf=pd.DataFrame(data=zip(pcx,pcy),columns=['PC1 ' + '(' +
str(pc1) + '% of variance' + ' )','PC2 ' + '(' + str(pc2) + '% of variance'
+ ' )'])#,raise_missing=False)
    principaldf['Sample']=pca_df.index
    principaldf['Group']= grouping
    principaldf['x']=principaldf['PC1 ' + '(' + str(pc1) + '% of variance'
+ ' )']
    principaldf['y']=principaldf['PC2 ' + '(' + str(pc2) + '% of variance'
+ ' )']
    #st.dataframe(principaldf)

    with sns.axes_style(schema):
        fig = plt.figure(figsize=(8,8))
        ax = fig.add_subplot(1, 1, 1)
        sns.scatterplot(data= principaldf,x= 'PC1 ' + '(' + str(pc1) + '%
of variance' + ' )',
                        y= 'PC2 ' + '(' + str(pc2) + '% of variance' + '
)',
                        ax=ax,s=400, hue=principaldf.Group.tolist())
        plt.title(title,loc='left')
        texts= []
        for x,y,s in zip(principaldf.x,(principaldf.y),principaldf.Sample):
            texts.append(ax.text(x,y,s,size=labelsize))
        adjust_text(texts,force_points=0.1, force_text=0.2,
expand_points=(1,1))
    st.pyplot(fig)
def violinplot():
    st.sidebar.header("Change Violin-plot Parameters")

    df2=df
    samples=controlg+expg


    cols = [x for x in headers if x not in samples]
    gp1=group1_title
    gp2=group2_title
    g1= [gp1]
    g2= [gp2]
    grouping= (g1 * len(controlg)) + (g2 * len(expg))
    temp = (df2[headers].melt(id_vars = cols, var_name =
'Sample',value_name = 'Expression'))
    temp=temp.dropna()
```

```
    logviolin = st.sidebar.checkbox('Log Expression Input Values')

    if logviolin:
        temp['Expression']=np.log((temp['Expression']+1))
        title='       Logged Values'
    else:
        temp['Expression']=temp['Expression']
        title='       Raw Values'

    #st.markdown(controlg)
    low_samples = controlg
    temp['Group'] = temp['Sample'].apply(lambda x : gp1 if x in low_samples
else gp2)

    with sns.axes_style(schema):
        fig = plt.figure(figsize=(8,8))
        ax = fig.add_subplot(1, 1, 1)
        grouped = st.sidebar.checkbox('Sample Grouping')
        if grouped:
            sns.violinplot(data=
temp,x='Sample',y='Expression',hue='Group',dodge=False)
        else:
            sns.violinplot(data= temp,x='Sample',y='Expression')
        jitter = st.sidebar.checkbox('Add Jitter')

        if jitter:
            jittersize=st.sidebar.slider('Size of points', 0, 10, 1)
            jitteralpha=st.sidebar.slider('Opacity of points', float(0),
float(1), float(0.2))
            sns.stripplot(data=
temp,x='Sample',y='Expression',alpha=jitteralpha,s=jittersize,color='black'
,jitter=0.4)

        else:
            pass

        locs, labels = plt.xticks()
        plt.setp(labels, rotation=45)
        plt.title(title,loc='left')

    st.pyplot(fig)
def prepdata():
    global l2fc
    global logp
    global df
    global ps
    df= df.sort_values(by=[paddys])
    Fold_changes = log2s
    padjs = paddys
    gene_names = genenames
    fc_df = df.filter([Fold_changes])
    fc_df.columns= ['L2FC']
    p_df = df.filter([padjs])
    p_df.columns= ['p-value (Adjusted)']
    gene_df = df.filter([gene_names])
    gene_df.columns= ['Gene']
    df['L2FC'] = fc_df
    df['padj'] = p_df
    df['Gene'] = gene_df
```

```
    df['neglogp'] = - np.log(df['padj'])
    df['p(-log10)'] = - np.log10(df['padj'])
    ps=df["padj"].tolist()
    l2fc=df["L2FC"].tolist()
    logp=df["neglogp"].tolist()
    #st.dataframe(df)
def volcanoplot():
    global logp



    st.sidebar.header("Change Volcano-plot Parameters")
    option = st.sidebar.selectbox('Choose Volcano Type',('Standard', 'Gene
Ontology (Human Only)'))

    p=st.sidebar.slider('p-value threshold', float(min(ps)),
float(max(ps)), 0.05)
    top_genes=st.sidebar.slider('Label top genes', 0, 30, 5)
    textsize=st.sidebar.slider('Size of labels', 0, 20, 8)
    x=st.sidebar.slider('Adjust x-axis', float(0), float(max(l2fc)+2),
float(max(l2fc)+2))
    y=st.sidebar.slider('Adjust y-axis', float(0), max(logp)+10,
(max(logp)+10))
    size=st.sidebar.slider('Point Size', float(1), float(50), float(25))
    logtrans = st.sidebar.checkbox('Log base-10 Transformation')

    if logtrans:
        logp=df["p(-log10)"].tolist()
        yaxistitle="-log10 p-value"
        guidebar=-np.log10(p)
        df['x']= df.L2FC
        df['y']= df['p(-log10)']
        df['Gene2']= df.Gene

        dfmod20=df.head(top_genes)
    else:
        yaxistitle="-log p-value"
        guidebar=-np.log(p)
        df['x']= df.L2FC
        df['y']= df.neglogp
        df['Gene2']= df.Gene

        dfmod20=df.head(top_genes)

    if option=='Gene Ontology (Human Only)':
        st.error('Currently In Development')
        '''
        st.sidebar.header("Biological Process Labelling Input (Human
Only)")
        user_input = st.sidebar.text_input("Enter GO Processes here
separated by commas, ie 'Sodium, Receptor, Membrane, Pregnancy'")
        user_list=user_input.split(',')
        st.text(user_list)



        @st.cache(suppress_st_warning=True)
        def prepmartdata():
```

```
            st.write("Cache miss: expensive_computation ran")

            dfprocesses=pd.read_csv('mart_export.txt')
            df2=dfprocesses.merge(df, right_on='Row.names', left_on='Gene
stable ID', how='outer')
            df2=df2.dropna(subset=['Unnamed: 0', 'Row.names','padj','GO
term name'])
            return df2



        df2=prepmartdata()


        remove_list = user_list
        df2['flagCol'] = np.where(df2['GO term
name'].str.contains('|'.join(remove_list)),1,0)
        df2['flagCol'] = df2['flagCol'].replace(0, np.nan)
        dfflagged=df2.dropna(subset=['flagCol'])
        genes=dfflagged['external_gene_name'].values.tolist()
        genes = set(genes)
        genes = list(genes)
        df2['Flagged by keywords ' + str(remove_list)]=
df2['external_gene_name'].isin(genes)
        df3= df2[df2['Flagged by keywords ' +
str(remove_list)].astype(str).str.contains('True')]
        df3=df3.drop_duplicates(subset=['external_gene_name'])

        df3=df3.sort_values(by=['padj'])
        df3['x']= df3.log2FoldChange
        df3['y']= df3.neglogp
        df3['Gene2']= df3.external_gene_name

        dfgolabels=df3.iloc[0:top_genes]


        df['Flagged by keywords ' + str(remove_list)]=
df['external_gene_name'].isin(genes)
        coly=df['Flagged by keywords ' + str(remove_list)].values.tolist()
        with sns.axes_style(schema):
            fig = plt.figure(figsize=(8,8))
            ax = fig.add_subplot(1, 1, 1)
            sns.scatterplot(l2fc,logp,hue=coly,s=size,ax=ax)
            ax.set_ylabel("-log p-value",fontsize=15)
            ax.set_xlabel("log2 Fold Change",fontsize=15)
            plt.ylim(0, y)
            plt.xlim((-x), x)
            plt.title(title)
            legendlabel='Flagged by keywords ' + str(remove_list)
            ax.legend([legendlabel],loc="upper right")
            texts=[]
            props= dict(boxstyle='round', facecolor='wheat', alpha=0.9)
            for x,y,s in zip(dfgolabels.x,dfgolabels.y,dfgolabels.Gene2):

texts.append(ax.text(x,y,s,size=textsize,bbox=props))
            adjust_text(texts)
```

```
            st.pyplot(fig)
            '''


    else:


        df['coly']=df["padj"]<p
        coly=df['coly'].values.tolist()

        with sns.axes_style(schema):
            fig = plt.figure(figsize=(8,8))
            ax = fig.add_subplot(1, 1, 1)
            sns.scatterplot(l2fc,logp,hue=coly,s=size,ax=ax)
            ax.set_ylabel(yaxistitle,fontsize=15)
            ax.set_xlabel("log2 Fold Change",fontsize=15)
            guides = st.sidebar.checkbox('Toggle Guidebars')
            if guides:
                ax.hlines(guidebar,(-x),x,linestyles='--
',color='grey',linewidth=0.7)
                ax.vlines(-1,0,y,linestyles='--
',color='grey',linewidth=0.7)
                ax.vlines(1,0,y,linestyles='--',color='grey',linewidth=0.7)
            else:
                pass

            plt.ylim(0, y)
            plt.xlim((-x), x)
            plt.title(title)
            legendlabel="p.Adj < " + str(p)
            ax.legend([legendlabel],loc="upper right")
            texts=[]
            #st.dataframe(dfmod20)
            for x,y,s in zip(dfmod20.x,dfmod20.y,dfmod20.Gene2):
                        texts.append(ax.text(x,y,s,size=textsize))
            adjust_text(texts)




            st.pyplot(fig)
def heatmapplot():
    #p2=0.05
    #delta= st.sidebar.number_input("Adjust p-value:",
min_value=float(0.0001), max_value=float(1),value=float(0.05),step=0.005)
    #st.write(delta)
    #p2=st.slider('p-value threshold', min(ps), 0.05, 0.0005)
    st.sidebar.header("Change Heatmap Parameters")
    user_input = st.sidebar.text_input("Threshold adjusted p-value at
(enter a value)", 0.05)
    delta=float(user_input)
    fontsize=st.sidebar.slider('Set gene names font size', 0.01,
float(1.5), float(1))
    xfont=st.sidebar.slider('Set sample names font size', 2, 20, 10)
    sns.set(font_scale=fontsize)

    MODES = {"Standard": "magma",
        "Yellow, Orange and Red":"YlOrRd",
```

```
        "Orange and Red":"OrRd",
        "Blue and Red":"bwr",
        "Red and Purple":"RdPu",
        "Hot":"hot"


    }
    x= list(MODES.keys())
    c = st.sidebar.selectbox("Define a colour-scheme", x)
    colour= MODES.get(c)


    df2=df.where(df[paddys]<delta)
    samples=controlg+expg
    cols = [x for x in headers if x not in samples]
    gp1="placehld"
    gp2="placehld_2"
    g1= [gp1]
    g2= [gp2]
    grouping= (g1 * len(controlg)) + (g2 * len(expg))
    temp = (df2[headers].melt(id_vars = cols, var_name =
'Sample',value_name = 'Expression'))
    #st.markdown(controlg)
    low_samples = controlg
    temp['Group'] = temp['Sample'].apply(lambda x : gp1 if x in low_samples
else gp2)
    gene=genenames
    global temp_adj
    def plot():
        g= sns.clustermap(temp.groupby([gene,
'Sample'])['Expression'].mean().unstack(),cmap=colour,metric =
'correlation',z_score=0)
        g.ax_heatmap.set_xticklabels(g.ax_heatmap.get_xmajorticklabels(),
fontsize = xfont)
        return g.fig
    #st.table(temp_adj)
    figure= plot()
    st.pyplot(figure)
def geneset_enrichment():

    gseaselect=[genenames]+controlg+expg
    gene_exp=df[gseaselect]
    gene_exp=gene_exp.sort_values(by=gene_exp.columns[2])
    gene_exp=gene_exp.tail(7000)
    #st.dataframe(gene_exp.head(5))
    #
gene_sets='GO_Biological_Process_2018','GO_Molecular_Function_2018','KEGG_2
019_Human'
    # gene_sets='MGI_Mammalian_Phenotype_Level_4_2019'
    # gene_sets='GO_Molecular_Function_2018'
    gp1="placehld"
    gp2="placehld_2"
    g1= [gp1]
    g2= [gp2]
    grouping= (g1 * len(controlg)) + (g2 * len(expg))
    @st.cache(suppress_st_warning=True,allow_output_mutation=True)
    def calculate_genesets():
        gs_res = gp.gsea(data=gene_exp, # or
data='./P53_resampling_data.txt'
```

```
                        gene_sets=gene_set, # enrichr library names
                        cls= grouping, # cls=class_vector
                        # set permutation_type to phenotype if samples >=15
                        permutation_type=permtype,
                        permutation_num=100, # reduce number to speed up
test
                        outdir=None,  # do not write output to disk
                        no_plot=True, # Skip plotting
                        method=statmethod, # or t_test
                        processes=0, seed= 7,
                        format='png')
        return(gs_res)


    gs_res=calculate_genesets()
    st.success('Gene Set Enrichment Successful')
    terms = gs_res.res2d.index
    fdrs= gs_res.res2d['fdr']
    reindexed=gs_res.res2d.reset_index()
    st.dataframe(reindexed)
    processes = st.selectbox("Enriched Processes in Your Dataset (Choose
one to visualise)", terms)
    keepGO= str(processes[-len(processes):])
    #st.text(keepGO)

    #ids =
np.unique(reindexed.stack()[reindexed.astype('str').str.contains(keepGO)].i
ndex.get_level_values(0))
    ids= reindexed.index[reindexed['Term'] ==keepGO].tolist()
    #st.text(ids)

    def plot(n):
        #st.text(n)
        genes = gs_res.res2d.genes[n].split(";")
        #st.dataframe(genes)
        dfprocess= gs_res.heatmat.loc[genes]
        control_mean= dfprocess.iloc[:,:3].mean(axis=1)
        exp_mean= dfprocess.iloc[:,3:].mean(axis=1)
        dfprocess_mean=pd.DataFrame()
        dfprocess_mean['Control_mean']=control_mean
        dfprocess_mean['Experimental_mean']=exp_mean

dfprocess_mean['Divided']=(dfprocess_mean['Control_mean']/dfprocess_mean['E
xperimental_mean'])
        dfprocess_mean=np.log(dfprocess_mean[['Divided']])

dfprocess_mean=dfprocess_mean.sort_values(by='Divided',ascending='False')


        grid=sns.clustermap(data=dfprocess_mean.T ,z_score=0,
col_cluster=False,row_cluster=False, figsize=(18,4),cmap='RdBu_r',
dendrogram_ratio=0.0,cbar_pos=None)
        labels=grid.ax_heatmap.get_xticklabels()
        grid.ax_heatmap.set_xticklabels(labels=labels, rotation=60,
fontsize = 16)
        labels=grid.ax_heatmap.get_yticklabels()
        grid.ax_heatmap.set_yticklabels(labels=labels, fontsize = 16)
        grid.ax_heatmap.set_title(terms[n] + ',  FDR: ' +
str(round(fdrs[n],3)),size=17)
```

```
        return grid
    indexes=ids
    for value in indexes:
        figure=plot(value)
        st.pyplot(figure)



    def fullgenesets():
        fig, ax = plt.subplots(figsize=(3,len(fullset_head)/4))

scatterplot=ax.scatter(y=fullset_head.index,x=fullset_head['neglogp'],c=ful
lset_head['nes'],cmap='seismic',edgecolors='black',s=60)
        cbar= plt.colorbar(scatterplot, orientation="horizontal",
pad=2.8/len(fullset_head))
        cbar.set_label("Normalised\nEnrichment Score")
        plt.title('Enriched Pathways and Functions',size=12)
        plt.xlabel('FDR (-log)')

        st.pyplot(fig)

    gseaplot = st.sidebar.checkbox('Display GSEA Bubble Plot')
    if gseaplot:
        fullset=gs_res.res2d

        fullset['neglogp']= -np.log(fullset['fdr'])
        fullset=fullset.sort_values(by='neglogp',ascending=False)
        head=st.sidebar.slider('Number of Pathways', 0, 50, 20)
        fullset_head=fullset.head(head)
        fullset_head=fullset_head.sort_values(by='neglogp')
        fullgenesets()

_max_width_()
#load=st.checkbox("Load Data")
st.title("Welcome to PadPlot")
st.header("Start by Providing your Dataset")
data = st.file_uploader("Upload a Dataset", type=["csv", "txt","tsv"])
if not data:
    st.warning('Please input a file.')
    st.stop()
st.success('Dataset loaded.')
#st.success('Dataframe loaded')
title=data.name
st.text("")
st.text("")
st.text("")
loaddata()
prepdata()
selectplot()
#if load:
#    loaddata()
#    prepdata()
# if data not in globals():
#     st.error('No File Loaded')
st.text("")
st.text("")
st.text("")
```

```
st.text("")
st.text("")
st.text("")
st.text("")
c1, c2 = st.beta_columns((2))
c3, c4 = st.beta_columns((2))
```

4.6.2 ScGress- Single Cell

```
import scanpy as sc
import numpy as np
from adjustText import adjust_text
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
from scipy import stats
import shutil
import os

#############################################################################
#############################################################################
# to be changed by the user, everything else is automatic
dataset_name='GSM4572192'
filepath='GSM4572192_Control1_filtered_feature_bc_matrix.h5'
reference_gene='UMOD'
#############################################################################
#############################################################################
newdir='figures\\umap' + dataset_name +'\\'
plotdir=dataset_name
if not os.path.exists(newdir):
    os.makedirs(newdir)

if not os.path.exists(plotdir):
    os.makedirs(plotdir)
def read_and_prepare_data(filepath, dataset_name,reference_gene):
    dataset=sc.read_10x_h5(filepath)
    dataset.var_names_make_unique
    sc.pl.highest_expr_genes(dataset, n_top=20, )
    dataset.var['mt'] = dataset.var_names.str.startswith('MT-')
    sc.pp.calculate_qc_metrics(dataset, qc_vars=['mt'], percent_top=None,
log1p=True, inplace=True)
    sc.pl.violin(dataset, ['n_genes_by_counts', 'total_counts',
'pct_counts_mt'],
                 jitter=0.4, multi_panel=True)
    adata=dataset
    adata = adata[adata.obs.n_genes_by_counts < 2500, :]
    adata = adata[adata.obs.pct_counts_mt < 10, :]
    sc.pp.normalize_total(adata, target_sum=1e4)
    sc.pp.log1p(adata)
    sc.pp.scale(adata, max_value=10)
    sc.tl.pca(adata, svd_solver='arpack')
    sc.pl.pca(adata, color=reference_gene)
    sc.pp.neighbors(adata, n_neighbors=10, n_pcs=40)
    sc.tl.umap(adata)
    sc.tl.leiden(adata)
```

```
    sc.pl.umap(adata, color=['leiden',
reference_gene],use_raw=False,save=dataset_name + '\\Full_data.png')
    adata.layers['scaled'] = sc.pp.scale(adata, copy=True).X
    reference_adata = adata[adata[: , reference_gene].X > 1, :]
    sc.pp.neighbors(reference_adata, n_neighbors=10, n_pcs=50)
    sc.tl.umap(reference_adata)
    sc.tl.leiden(reference_adata)
    sc.pl.umap(reference_adata,
color=['leiden',reference_gene],save=dataset_name + '\\' + reference_gene +
'_cells.png')
    sc.tl.rank_genes_groups(reference_adata, 'leiden', method='t-test')
    sc.pl.rank_genes_groups(reference_adata, n_genes=25, sharey=False)

    create_peicewise_linear_regressions(reference_adata)

#specific_cluster=reference_adata[reference_adata.obs['leiden'].isin(['0'])
,:]

def create_peicewise_linear_regressions(reference_adata):
    genes=list(reference_adata.var.index)
    genes
    gene_list=[]
    p_vals=[]
    intercepts=[]
    expression_level=[]
    r_val=[]
    slopes=[]
    for gene in genes:
        reference_subset=reference_adata[: , reference_gene].X
        reference_subset=reference_subset.tolist()
        target_subset=reference_adata[: , gene].X
        target_subset=target_subset.tolist()
        reference_flat= [item for sublist in reference_subset for item in
sublist]
        target_flat= [item for sublist in target_subset for item in
sublist]
        if len(target_flat)==len(reference_flat)*2:
            target_flat=[1]*len(reference_flat)
        else:
            pass

exp_level=(reference_adata[reference_adata.obs['leiden'].isin(['0']),:][:,g
ene].X > 0).mean(0)
        df_pvals = pd.DataFrame({reference_gene:reference_flat,
                            "Gene":target_flat})
        df_pvals=df_pvals[df_pvals['Gene']>1]
        reference_flat=df_pvals[reference_gene].values.tolist()
        target_flat=df_pvals['Gene'].values.tolist()
        try:
            slope, intercept, r_value, p_value, std_err =
stats.linregress(reference_flat,target_flat)
            p_value
        except:
            p_value=1
            intercept=1
            slope=1
            r_value=1
        gene_list.append(gene)
        print(gene)
```

```
        p_vals.append(p_value)
        print(p_value)
        intercepts.append(intercept)
        expression_level.append(exp_level)
        r_val.append(r_value)
        slopes.append(slope)
        if p_value < 0.1 and p_value > 0 and exp_level > 0.5:
            sns.scatterplot(reference_flat,target_flat)
            plt.ylabel(gene)
            plt.xlabel(reference_gene)
            plt.title('Correlation of ' + gene + ' with ' + reference_gene
+ ' p='+ str(round(p_value,6)))
            savetitle=dataset_name +
'\\{}_with_{}.png'.format(reference_gene,gene)
            plt.savefig(savetitle)
            plt.clf()
            print(df_pvals)
            #time.sleep(3)
        else:
            pass

    create_allgenes_plot(gene_list,p_vals,intercepts,expression_level,
r_val, slopes,dataset_name,reference_adata)

def create_allgenes_plot(gene_list,p_vals,intercepts,expression_level,
r_val, slopes,dataset_name,reference_adata):
    from adjustText import adjust_text
    plt.figure(figsize=(12,8))
    df_regression = pd.DataFrame({"Gene":gene_list,
                          "pval":p_vals,
                                'intercept':intercepts,
                                 'expression':expression_level,
                                 'r_value':r_val,
                                 'slope':slopes})
    df_regression=df_regression.sort_values(by='pval')
    df_regression=df_regression[df_regression['pval']<0.05]
    df_regression=df_regression[df_regression['pval']>0]
    df_regression=df_regression[df_regression['intercept']>0.1]
    exp_vals=df_regression['expression'].values.tolist()
    exp_vals_flat= [item for sublist in exp_vals for item in sublist]
    df_regression['expression']=exp_vals_flat
    df_regression['neglogp']=-np.log(df_regression['pval'])
    df_regression_labs=df_regression[df_regression['expression']>0.5]

#df_regression_labs=df_regression_labs[df_regression_labs['expression']>0.5
]
    df_regression['Positive Correlation'] = df_regression['slope'] > 0
    df_regression['rSq']=df_regression['r_value']**2
    cmap=sns.diverging_palette(250, 30, l=65, center="dark", as_cmap=True)

sns.scatterplot(x=df_regression['expression'],y=df_regression['neglogp'],hu
e=df_regression['Positive Correlation'].values.tolist(),s=35)
    plt.legend(title='Correlation Type', loc='upper left',
labels=['Positive', 'Negative'])
    plt.ylabel('Linear Regression neglog p-value (AU)')
    plt.xlabel('Proportion of cells also expressing labelled gene')
    plt.ylim(2.5,10)
    #plt.hlines(0.05,0,max(df_regression['expression']),linestyles='--')
    #plt.vlines(0.5,0,max(df_regression['pval']))
```

```
    texts=[]
    for x,y,s in
zip(df_regression_labs['expression'],df_regression_labs['neglogp'],df_regre
ssion_labs['Gene']):
                texts.append(plt.text(x,y,s,size=10))
    adjust_text(texts,autoalign='x'
               )
    plt.savefig(dataset_name+'\Coexpressed_genes_regression.png')
    df_regression_labs.to_csv(plotdir+'\\' +
dataset_name+'Coexpressed_genes.csv')


    x=df_regression_labs.head(3)
    x=x['Gene'].values.tolist()
    sc.pl.umap(reference_adata,
color=[reference_gene,x[0],x[1],x[2]],save=dataset_name+'\\Coexpressed_gene
s.png')

    return df_regression_labs
read_and_prepare_data(filepath, dataset_name,reference_gene)

source_dir = newdir
target_dir = plotdir

file_names = os.listdir(source_dir)

for file_name in file_names:
    shutil.move(os.path.join(source_dir, file_name), target_dir)

shutil.rmtree(newdir)
```

5 In-vitro study of UMOD

5.1 Introduction

Whilst UMOD has been linked to the regulation of blood pressure in the context of sodium, TAL cells may also respond to the concentration of extracellular sodium chloride and react to changes via altered UMOD expression. In the clinical context, human studies have shown that the urinary excretion of UMOD correlates with sodium chloride intake. A dependent, time series study, using low or high doses of dietary sodium chloride (10mmol/day for the first week vs 240mmol/day for the second week) showed that urinary excretion of UMOD protein in humans was significantly below baseline for week one, and significantly above baseline for week two (11.7 microg/min) versus baseline (19.5 microg/min; $p < 0.05$) and high salt intake (23.1 microg/min; $p < 0.01$) (Torffvita *et al*, 2004). Significantly, this study also showed that the response blood pressure changes was also linearly related to the absolute excretion of UMOD, indicating a proportionate blood pressure response in relation to UMOD levels. Additionally, a study by the Swiss Kidney Project on Genes in Hypertension (SKIPOGH) performed multivariate linear regression on a cohort of 1020 human individuals and showed that UMOD levels correlate positively with sodium chloride and blood pressure in these individuals, though their data only show this distinct three-way relationship in individuals in the upper quartiles of UMOD expression which reflects further on the challenges of examining this relationship in the human population (Ponte *et al.*, 2021). Since the first UMOD promotor variant associated UMOD levels with hypertension, UMOD has been associated with effects on the movement of sodium ions, specifically through promoting sodium chloride reabsorption via NKCC2. However, a unidirectional UMOD:NKCC2 relationship in the content of sodium chloride is not fundamentally proven; particularly as these clinical studies are not able to provide cellular interpretations. Therefore, it may be relevant to also consider the existence of a cyclical feedback loop between UMOD, NKCC2 and sodium chloride within TAL cells.

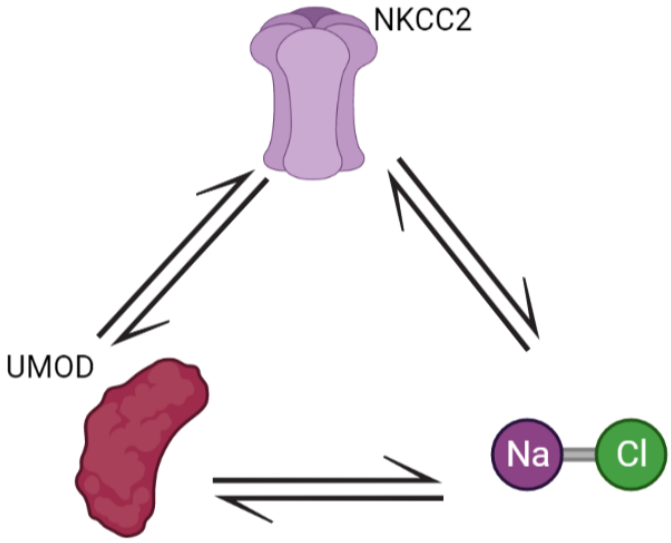


Figure 5.1-1 Theorised feedback loop between UMOD, NKCC2 and sodium chloride, in the context of protein - trafficking through Thick Ascending Limb cells.

In humans, the presence of UMOD in the urine has been shown to correlate with both the blood pressure phenotype and sodium chloride levels. However, for extracellular UMOD to exist, this first requires the protein to be trafficked-through and be cleaved-from TAL cells. Trafficking is the process by which nascent proteins are shuttled to other locations within the cell, including the cell membrane. Trafficking itself is a highly studied aspect of general pathophysiology in the context of wide and varied human disease models. The study of UMOD trafficking, as detailed in section 1.4.5 is also reasonably well examined, with a relatively robust understanding of the mechanisms which allow UMOD to reach the cell surface. In the case of UMOD, the most well-studied physiological functions (immune regulation, calcium sequestration) appear to be generated from extracellular, cleaved protein. However, it is possible that not only may UMOD protein manifest effects in the context of natriuresis at the cell surface, but also during the trafficking process. With the identical apical cycling and trafficking profile of both UMOD and NKCC2, it is reasonable to hypothesise a linked process. A recent paper examining the cell surface recycling of NKCC2, identified Alstrom-1 protein (ALMS1) as mediating trafficking and activity. ALMS1 knock out rats exhibited increased sodium natriuresis and were more sensitive to bumetanide- an NKCC2 inhibitor. Whilst this paper, noting the significance of UMOD in NKCC2 activity, attempted unsuccessfully to detect UMOD interactions, this does

not rule out potential interactions between UMOD and NKCC2 within the cell as they are trafficked (Allison, 2018).

Studying potential UMOD co-trafficking protein to protein interactions may be a valuable addition to understanding of the actions of UMOD in the context of blood pressure regulation. However, mechanistic interactions of the human UMOD protein are challenging to study due to the underlying properties of TAL cells. Primary human TAL cells can only be obtained by surgical biopsy, and due to limitations of operating theatres are challenging to obtain in viable states for cell culture; explaining why the majority of clinical studies only examine urinary UMOD levels. It is possible to use snap frozen primary tissue for protein work, however, yields from human surgery are often <5mg tissue, leading to poor signal in assay due to low concentration. Another approach involves the culturing of primary mouse TAL cells, however, the complement of transcription and trafficking cofactors between mouse and human is not entirely homologous and therefore mice signals may not translate well. This, when combined with the challenges in predicting eukaryotic gene expression mechanisms described in 1.3.5, may significantly confound analysis. Finally, no immortalised human cell lines have UMOD expression below semi-negligible cycle-threshold values of 34-35. Therefore, it is our opinion, that the most appropriate method for obtaining material for experimentation is to generate a UMOD stable transfected cell line.

Human Embryonic Kidney -293 (HEK293) cells are a hypotriploid immortalised renal derived human cell line (Graham *et al.*, 1977; Louis *et al*, 1997). Of the 64 immortalised human cell lines catalogued by the human protein atlas and clustered by the UMAP algorithm, HEK293 have the most relevant and varied renal gene complement (The Human Protein Atlas, 2021). For this reason, HEK293 cells are the most likely of available cell lines to express the required complement of transcription and trafficking factors required to be functionally relevant for mechanistic examination of UMOD, given that TAL cells are also derived from renal pathway pluripotents *in-utero*. Additionally, HEK293 cell lines have optimised transfection reagents and protocols, making them ideal choices for stable transfection.

Based on the observed relationship in humans between sodium chloride and urinary UMOD levels, we believe that UMOD levels and the magnitude of trafficking of UMOD may be affected by the presence of extra-cellular sodium chloride and that this is likely to provide mechanistic information as to the relationship between UMOD and blood pressure regulation. However, prior to this we require the development of a stable transfected HEK293 cell line. Therefore, we intend to produce a stable, reusable source of UMOD protein via the generation of a transfectant and to utilise this in downstream experiments.

5.2 Aims

- To establish a stable human UMOD transfectant HEK293 cell line as a resource

- To design and optimise an assay for measuring the effect of sodium chloride on UMOD protein expression via western blotting

- To examine trafficking of UMOD protein within the cell, in the context of sodium chloride

5.3 Specific Methods

5.3.1 Design of the UMOD Plasmid

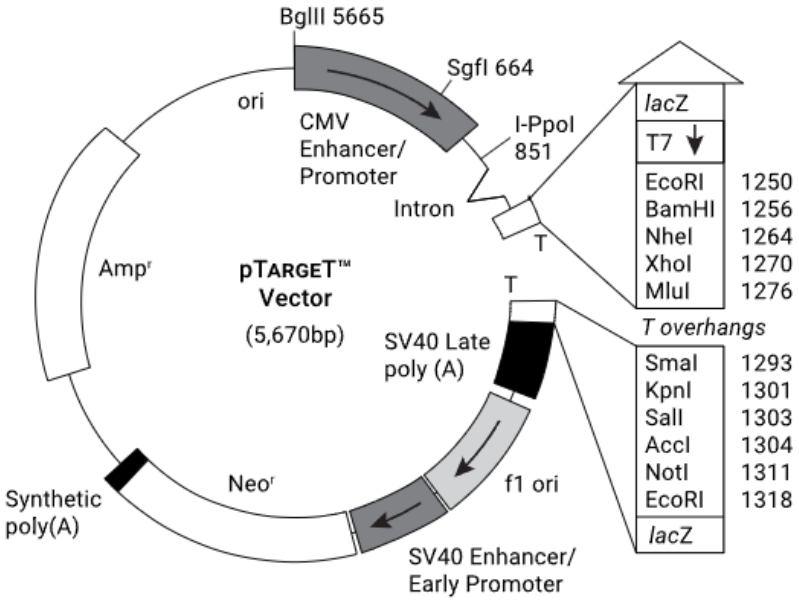


Figure 5.3-1. pTARGET Mammalian Expression Vector design, human UMOD cDNA ligated into the T-overhangs shown right.

The full length human UMOD transcript (UMOD-201 (Havana)) was produced as cDNA by 2.1.4.3 and ligated into the pTARGET Mammalian expression vector as per 2.4.2 at position 1276. This placed UMOD expression under the control of a cytomegalovirus (CMV) enhancer/promotor at position 669-1022 giving constitutive expression in transformed cells, independent of endogenous expression mechanisms. The pTARGET plasmid also encoded a Neomycin resistance gene (neomycin phosphotransferase, protein id:AAS48584.1) at position 2675-3469. Incorporation of neomycin resistance facilitated selection of $UMOD^{tg}$ cells from $UMOD^{wt}$ counterparts by applying antibiotic based selection pressure.

5.3.2 Selection of Stable Transfectants Using Geneticin

In order to prevent negative drift in retention of transfected genetic material over time, Geneticin™ Selective Antibiotic (G418 Sulfate) (Thermofisher, Manchester, UK) was used to maintain selective pressure on transfectants, as the plasmid contains a neomycin resistance insert. Prior to stable use of geneticin, there was a requirement

to establish the optimal dosing concentration of antibiotic in order to enrich for successful stable transfectants. This was determined by seeding individual T25 flasks with the same transfected population of cells and culturing for 24 hours in basic complete media. Following this 24-hour period, complete media with a range of Geneticin concentrations (0-1400ug/ml) was aspirated onto cells. Cells were imaged daily between 1 and 7 days by EVOS and dead cell proportions determined using a custom self-authored python program as per 2.3.5. The optimal concentration was determined as the highest concentration at which a population of cells survived over the course of the experiment.

5.3.3 Culturing of Cells Prior to Experiments

Twelve individual transfection assays were performed, producing a concurrent number of 'populations', which were individually stored in liquid nitrogen between uses. During these processes, cells were maintained in complete MEM as per 2.5.1.

For RNA based assays, immediately prior to experiments, stable transfectant cell populations were grown to confluency in T25 flasks before undergoing trypsinisation and reseeding into 6-well tissue-culture plates. Cells were cultured for 24-hours in basic media before being visually assessed for competency. Following this period, experimental assays were performed, with RNA extracted as per 2.4.5. Each individual T25 was treated as a biological n-number, while each well was treated as a technical replicate- for which averages were calculated. For protein-based assays, T25 flasks were grown to competency, with different transfection populations representing biological n-numbers. Cells were lysed in-flask using RIPA buffer (Thermofisher, Manchester, UK) and lysed homogenised solutions stored at -20˚C until use.

If appropriate, solutions of ionic compounds were added to media prior to aspiration onto cells for culturing, to reflect sodium chloride consumption in humans. Control media was labelled as basic complete media. Sodium-chloride-media was created by diluting a sterile stock solution of NaCl (1mg/ml) into 40mM concentration of basic complete media. Due to the potential ionic gradient pressure induced by the addition

of ions to media, an additional control media was created using 27.6mM $MgCl_2$ from a stock of 1mg/ml.

5.4 Results

5.4.1 Generation and Purification of the UMOD pTARGET Plasmid

Full length human UMOD cDNA was successfully reverse transcribed and subsequently amplified from human RNA, with the estimated size by PCR ladder of 2300bp reflecting the calculated size of 2353bp **(**Figure 5.4-1)(A). Blue/ white selection of colonies of transformed JM109 *e-coli* returned 42 white colonies and 29 blue colonies. White colonies indicate high probability of an interrupted lacZ gene, inhibiting metabolism, thus inferring successful integration of the UMOD insert **(**Figure 5.4-1)(B). Upon PCR amplification of expanded colonies, these colonies displayed successful amplification of the human UMOD gene, with successful amplification against the T7 forward primer at a fragment size of 2500bp; indicating successful orientation of the insert **(**Figure 5.4-1)(C). Plasmid purification from expanded bacterial cultures was successful, with high yield and retention between wash eluates, successfully producing purified plasmid for transfection (Figure 5.4-1)(D).

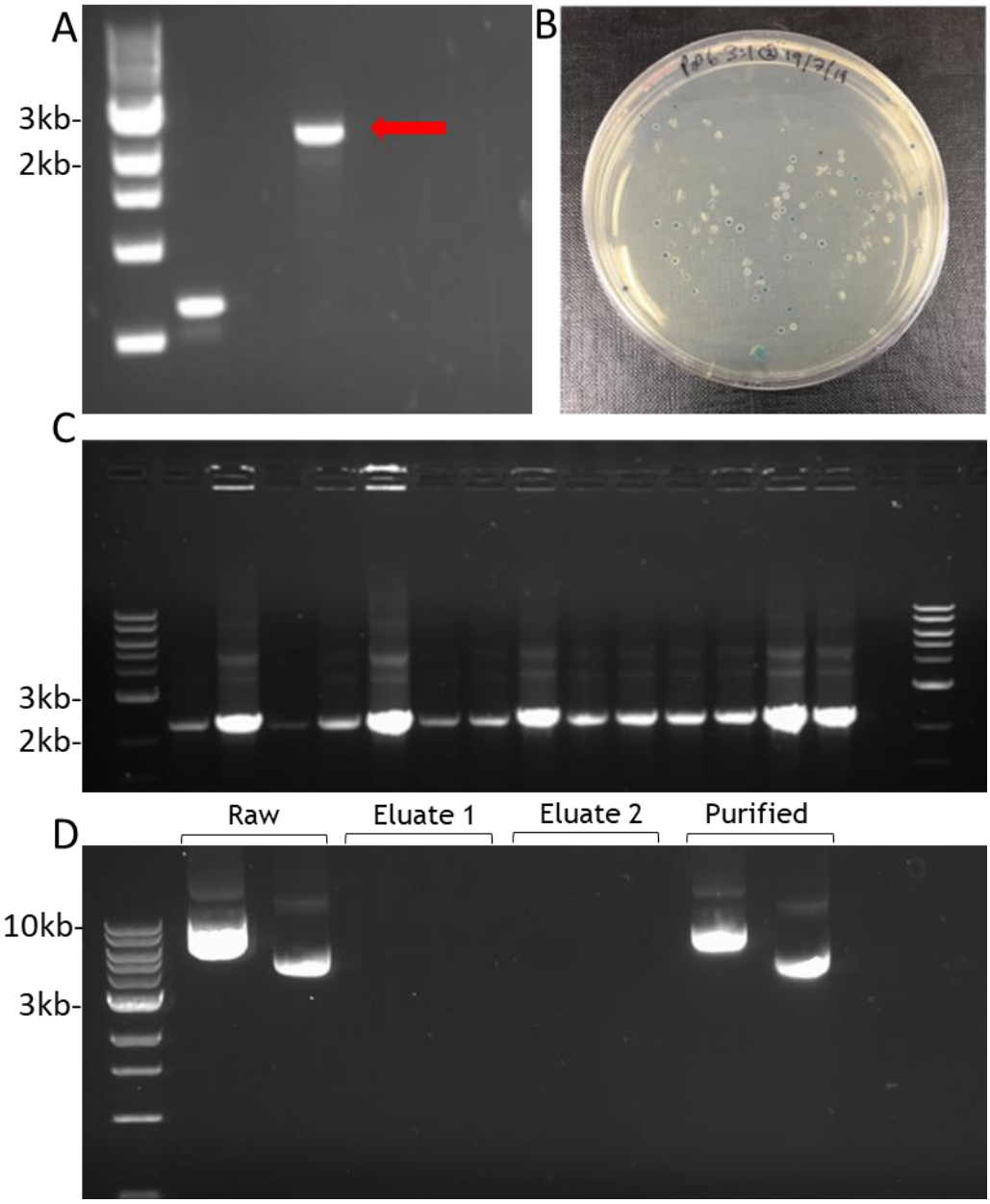


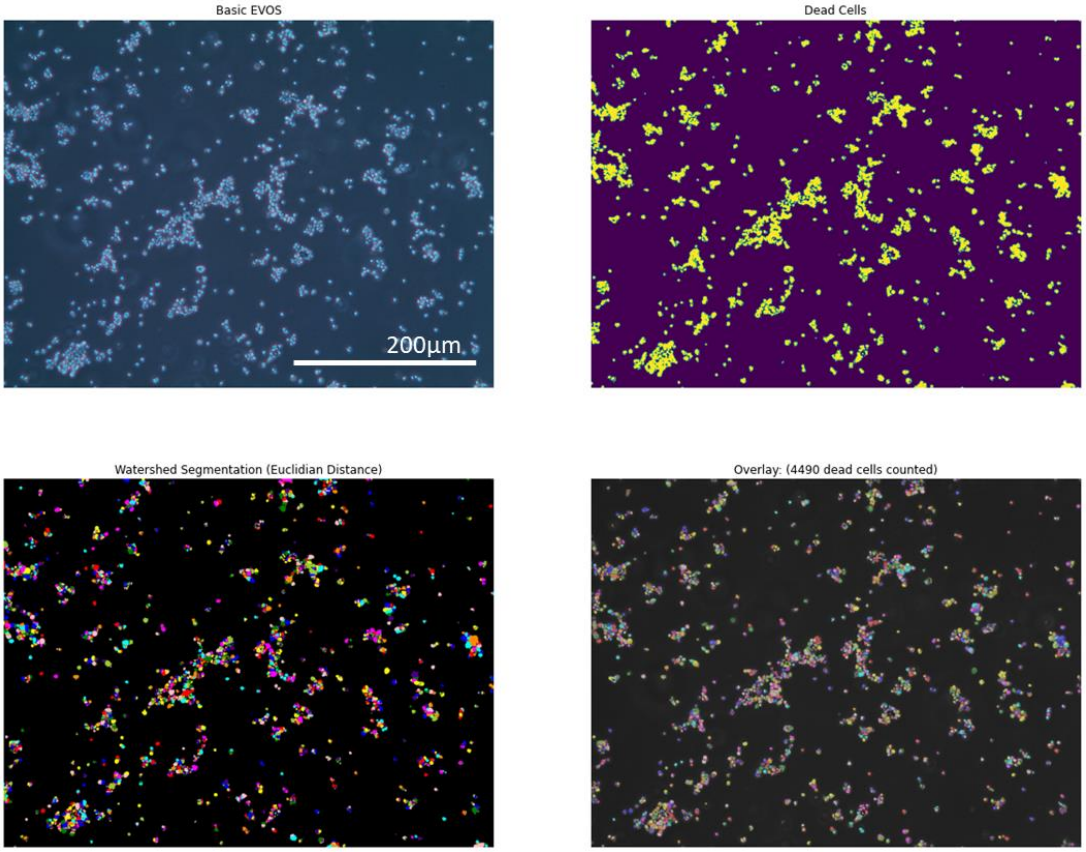
Figure 5.4-1 Agarose gel characterization of UMOD-plasmid design

A) End-point PCR of full length human UMOD cDNA (2353bp) (Lane 4) using the primer combination and reaction parameters described in 2.1.4.3, relativized against NEB-1kb ladder. B) Blue/white selection of successfully transfected JM109 competent Escherichia coli (e-coli) cells. C) End-point PCR of the UMOD transcript from selected colonies of JM109 transformed cells using the UMOD_rev to T7 pair, to verify orientation. D) direct UV luminescence of genetic material derived from plasmid purification; lane 2-3, raw material, lanes 4-7 wash eluates, lanes 8-9 purified plasmid.

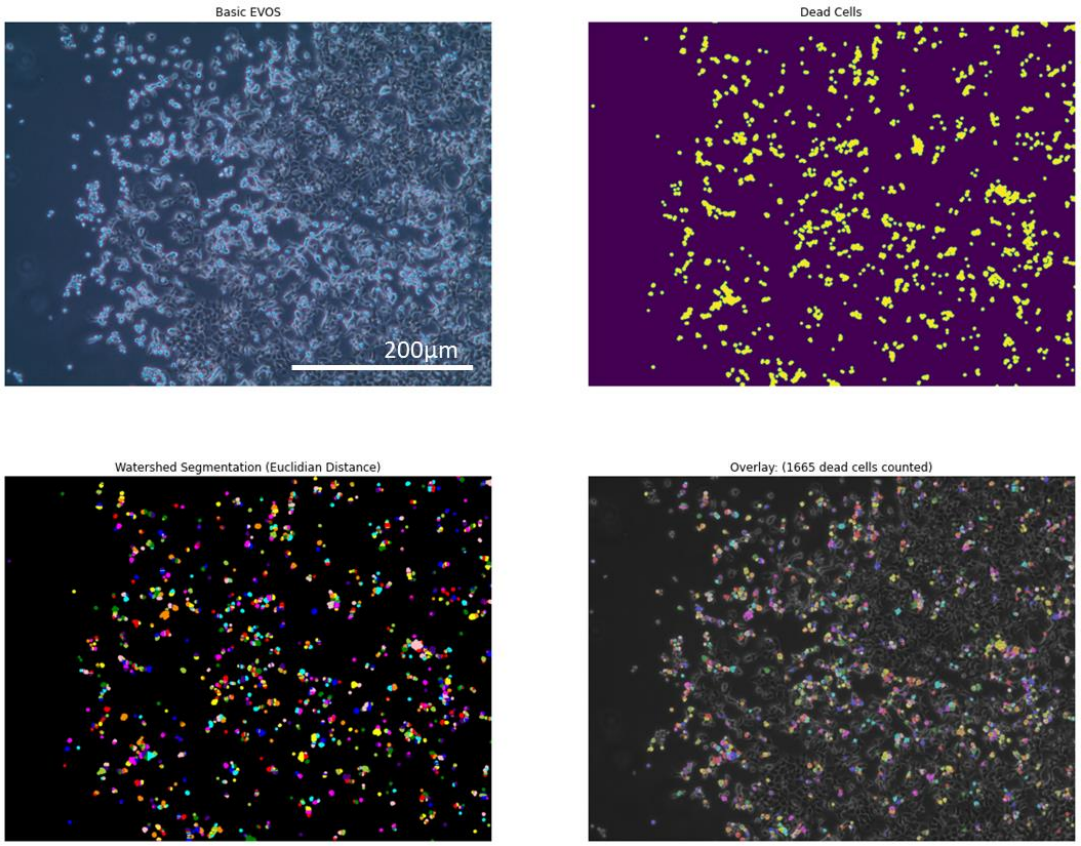
5.4.2 Maintenance Geneticin Concentrations and Impact of Transfection

Use of a custom created python application was successful in estimating proportions of dead HEK293 cells versus viable cells (Figure 5.4-2)(A-C). Cross-validation of python accuracy using random rolling 50μM windows was performed. Calculating python computed dead cells against actual manually (by eye) counted cells produced an error of 4.47% (2.58- 6.35% 95CI). With an error rate <5%, this python application was considered acceptable for use in quantifying toxicity. MTT assay 24-hours post-transfection with either pTARGET-UMOD plasmid or empty vector produced no difference in cytotoxicity in cells (Figure 5.4-3)(A) and there was no difference in ACBT expression between transfected and control cells (Figure 5.4-3)(B); both of these findings indicating that the presence and integration of the pTARGET-UMOD plasmid within HEK293 cells was not acutely deleterious to cell viability, though our data does not include an examination of the long term effects (weeks to months). Determining the antibiotic concentration against which pTARGET-UMOD plasmid transfected cells would survive, but their untransfected counterparts would not, suggested a concentration between 500-600μg/ml Geneticin. For this reason, 600μg/ml was selected, and cells closely observed with potential for decreasing to 500μg/ml considered upon detection of flask-wide toxicity (Figure 5.4-3)(C).

A


Basic EVOS
200µm
Dead Cells
Watershed Segmentation (Euclidian Distance)
Overlay: (4490 dead cells counted)

B


Basic EVOS
200µm
Dead Cells
Watershed Segmentation (Euclidian Distance)
Overlay: (1665 dead cells counted)

C

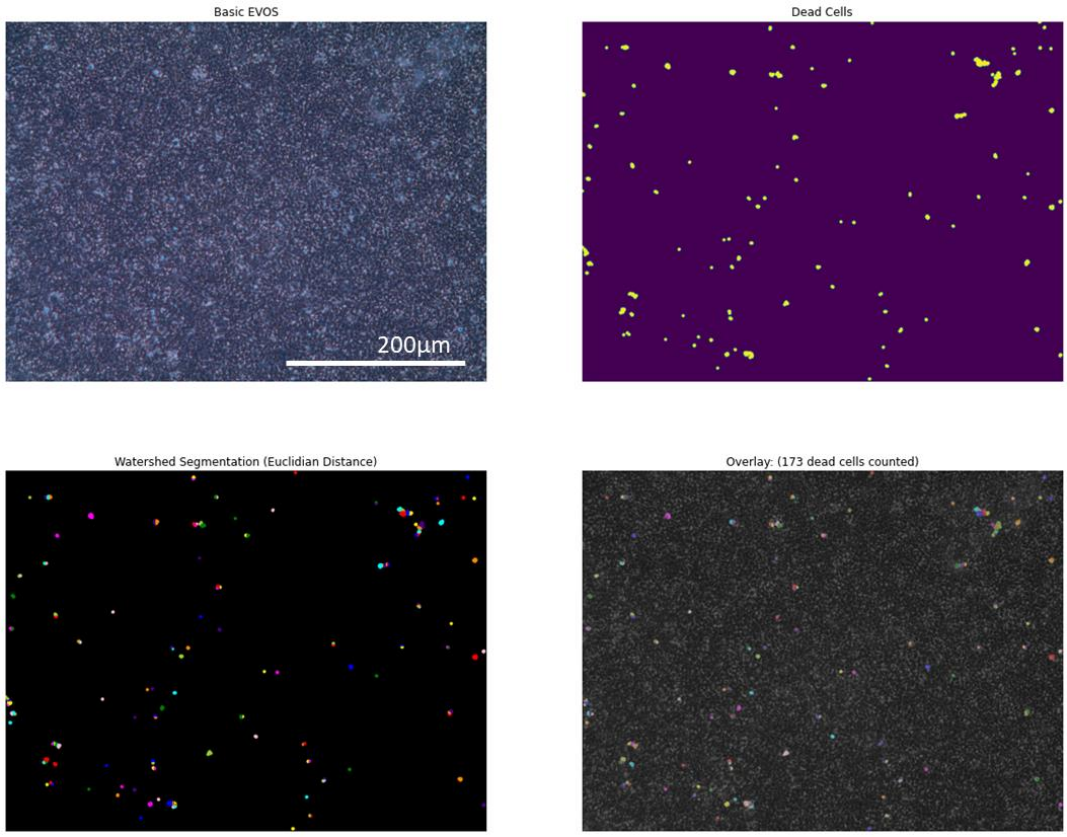


Figure 5.4-2 Characterisation of antibiotic toxicity in HEK293 cells

A) Python machine learning based quantification of dead HEK293 cells in response to 1000ug/ml Geneticin treatment at day-2 of experimentation. Frame 1- basic image obtained from EVOS microscope. Frame 2- intensity based thresholding of dead cells versus background. Frame 3- watershed segmentation via Euclidian distancing to individually segment clumped cells. Frame 4- super imposition of individual dead cells against basic EVOS image. B) 600ug/ml Geneticin treatment. C) 100ug/ml Geneticin treatment.

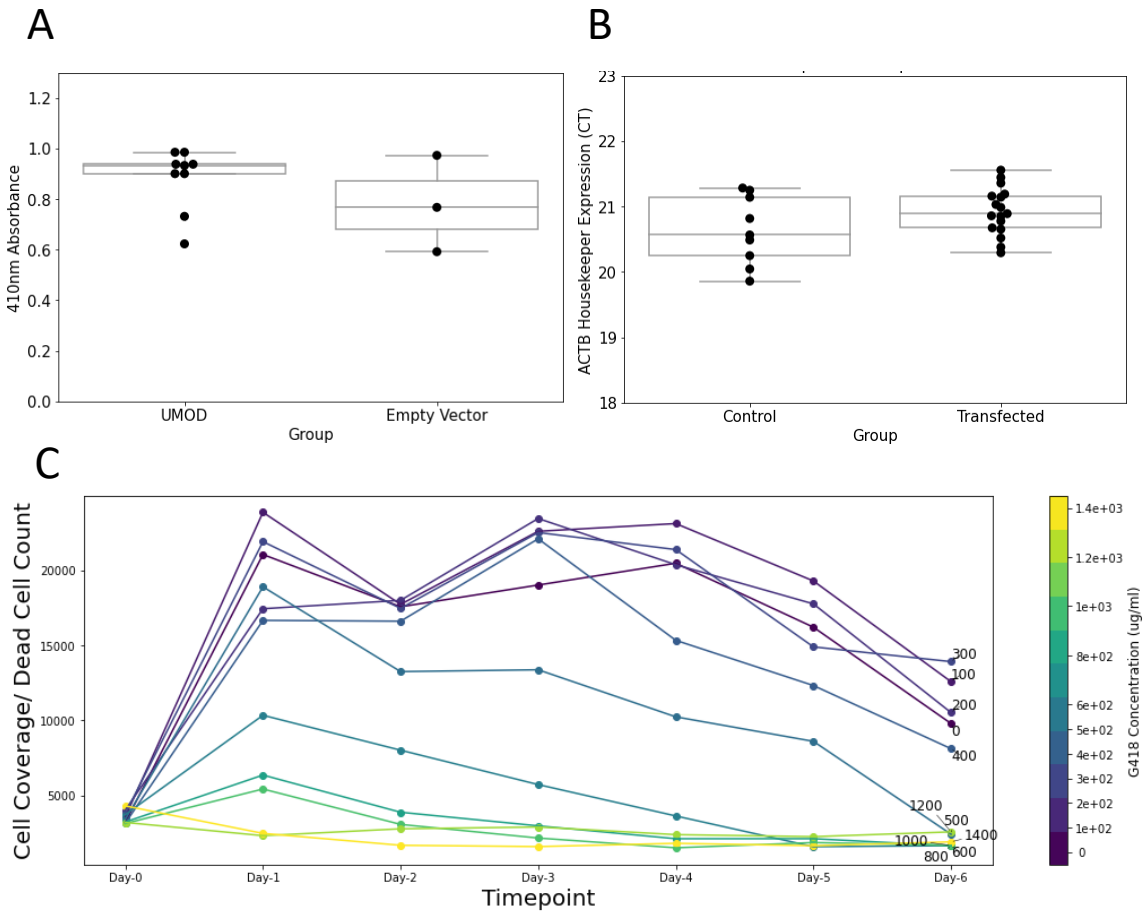


Figure 5.4-3 Impact of the transfection process on cell viability

A) MTT assay absorbance at 410nm between UMOD-plasmid cells and cells transfected with empty vector. B) Cycle threshold values of B-Actin by Taqman qRT-PCR between control and transfected cells. C) Cell coverage/ dead cell count across a range of antibiotic concentrations (100-1400ug/ml), imaged from day-0 to day-6. Cell coverage was quantified by taking the absolute pixel count above threshold, quantifying the absolute cell mass:*l_closed; = binary_closing(lowt,strel);l_closed_filled;nd.morphology.binary_fill_holes(l_closed);cell_mass= cv2.countNonZero(np.float32(l_closed_filled))*, thus normalizing each count to total cell coverage.

5.4.3 qRT-PCR Analysis of Transfected HEK293 Cells

Expression of human UMOD mRNA was significantly greater ($p<0.0001$) in pTARGET-UMOD transfected cells versus control, mean dCT 0.257 (-0.34, 0.89 95CI) versus 12.87 in untransfected cells (Figure 5.4-4)(A). There was no significant difference ($p>0.05$) in expression of SLC12A1 (NKCC2) between transfected and control cells mean dCT 17.322 (16.41, 18.22 95CI) versus 16.17 in untransfected cells, though these data did trend toward a decrease in expression in transfected cells (Figure 5.4-4)(B). Assessing the impact of differential ionic media supplementation on the expression of ACBT indicated there was no effect on expression of this housekeeper between standard, 40mM NaCl and 27mM MgCl2 supplementation (where MgCl2 was used as an ionic control) (ANOVA $p>0.05$) (Figure 5.4-4)(D). Examination of the relationship between transfected UMOD expression and endogenous NKCC2 expression by linear regression indicated no relationship in expression (Figure 5.4-4)(C). Time series analysis of the differences between expression of UMOD mRNA across a 14-day period, sampling every 3 days from day 2 (Figure 5.4-4)(E) suggested two possible effects; firstly, that pTARGET derived UMOD expression trends to decreasing over time and secondly, that there may be differences in response of UMOD expression between ionic solutions. This initial time series analysis was not statistically analysed due to nested triplicates within triplicates deriving from the same transfection event; therefore representing only one biological n.

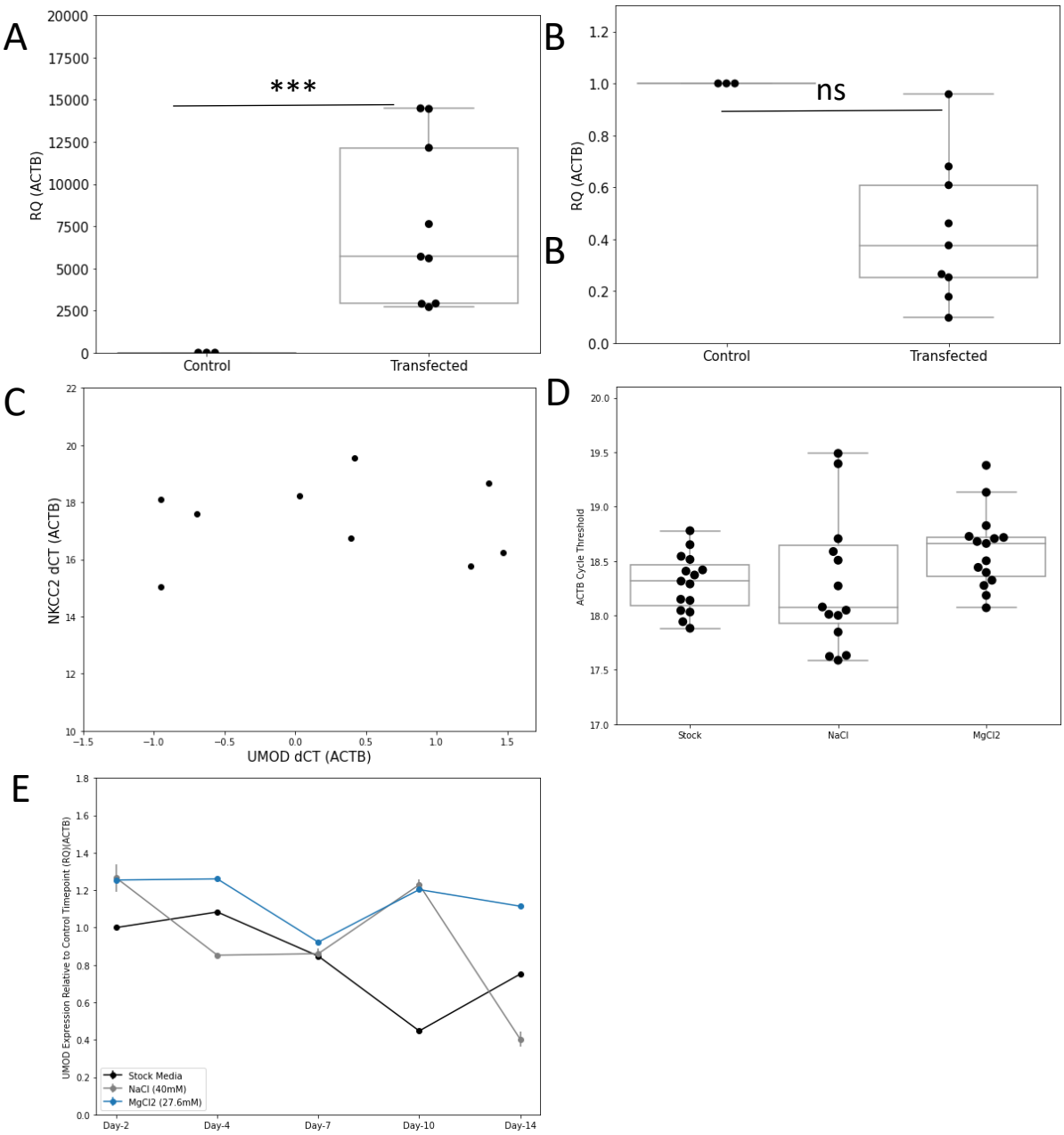


Figure 5.4-4 Assessment of transfection and ions on UMOD expression

A) Taqman Relative Quantification of UMOD versus B-Actin (ACBT) in individual transfection reactions with pTARGET-UMOD plasmid at 48-hours post transfection. B) Taqman Relative Quantification of SLC12A1 versus B-Actin (ACBT) in individual transfection reactions with pTARGET-UMOD plasmid at 48-hours post transfection. C) Regression of dCT for UMOD predictor versus NKCC2 response. D) ACTB cycle-threshold values between stock, 40mM NaCl and 27.6mM MgCl2 at 48-hours post culturing within supplemented media. E) Taqman Relative Quantification of UMOD versus B-Actin (ACBT) in transfected HEK293 cells, with technical triplicates over a two-week period, sampling every 3 days via well splitting.

5.4.4 Protein Analysis of Transfected HEK239 Cells

Protein extracted individual transfectant populations was assessed for UMOD expression with successful optimisation and detection at the expected molecular weight of 60kDa **(**Figure 5.4-5)(A/B). Whilst protein expression could be detected in transfected cells and was absent in wild type cells, transfected media did not produce a signal for UMOD protein where we anticipated that UMOD may be secreted from the cells and therefore detectable **(**Figure 5.4-5)(C). The expression of GAPDH protein appeared to be affected by the addition of 40mM NaCl to media, however, B-Actin was not affected in the same manner (Figure 5.4-5)(D)**.**

Quantitative assessment of the UMOD protein in response to sodium chloride was challenging to interpret. B-Actin was detected at multiple isoforms (Figure 5.4-6)(C). Multiple isoforms of Actin are annotated in mammals, and possible antibody cross-reactivity also observed (Lubit and Schwartz, no date; Müller *et al.*, 2013; Vedula *et al.*, 2017). However, in this case subjective interpretation of how to quantify and analyze isoforms become challenging, as differential conclusions can be made. Quantifying against the dominant isoform (Figure 5.4-6)(A) indicates a non-significant ($p>0.05$) result, whilst quantifying against all detected isoforms (Figure 5.4-6)(B) indicates that UMOD expression is significantly decreased ($p<0.05$) in response to NaCl in HEK293 transfected cells.

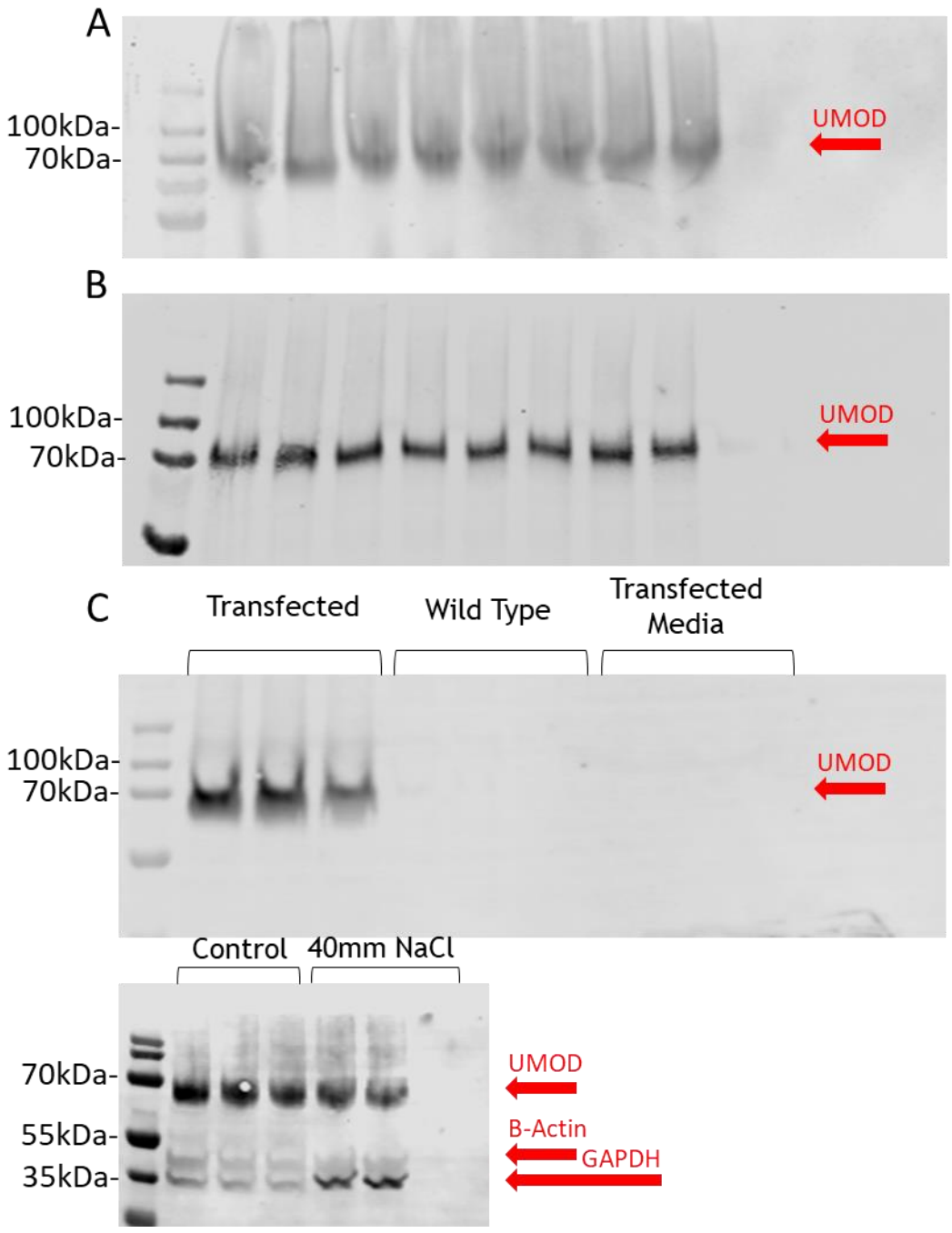


Figure 5.4-5 Transfected UMOD HEK293 cells western blot experiments

A) Western blot of protein derived from individual pTARGET-UMOD HEK293 transfectants, pre-optimisation, relativized against PageRuler Plus Ladder. B) Western blot of protein derived from individual pTARGET-UMOD HEK293 transfectants, post-optimisation. C) Individual transfected (lane 2-4) versus untransfected HEK293 cells (lane 5-7). Media aspirated directly from transfected cells (lane 8-10). D) Transfected control versus 40NaCl media, to optimize for housekeeper stability between B-Actin and GAPDH.

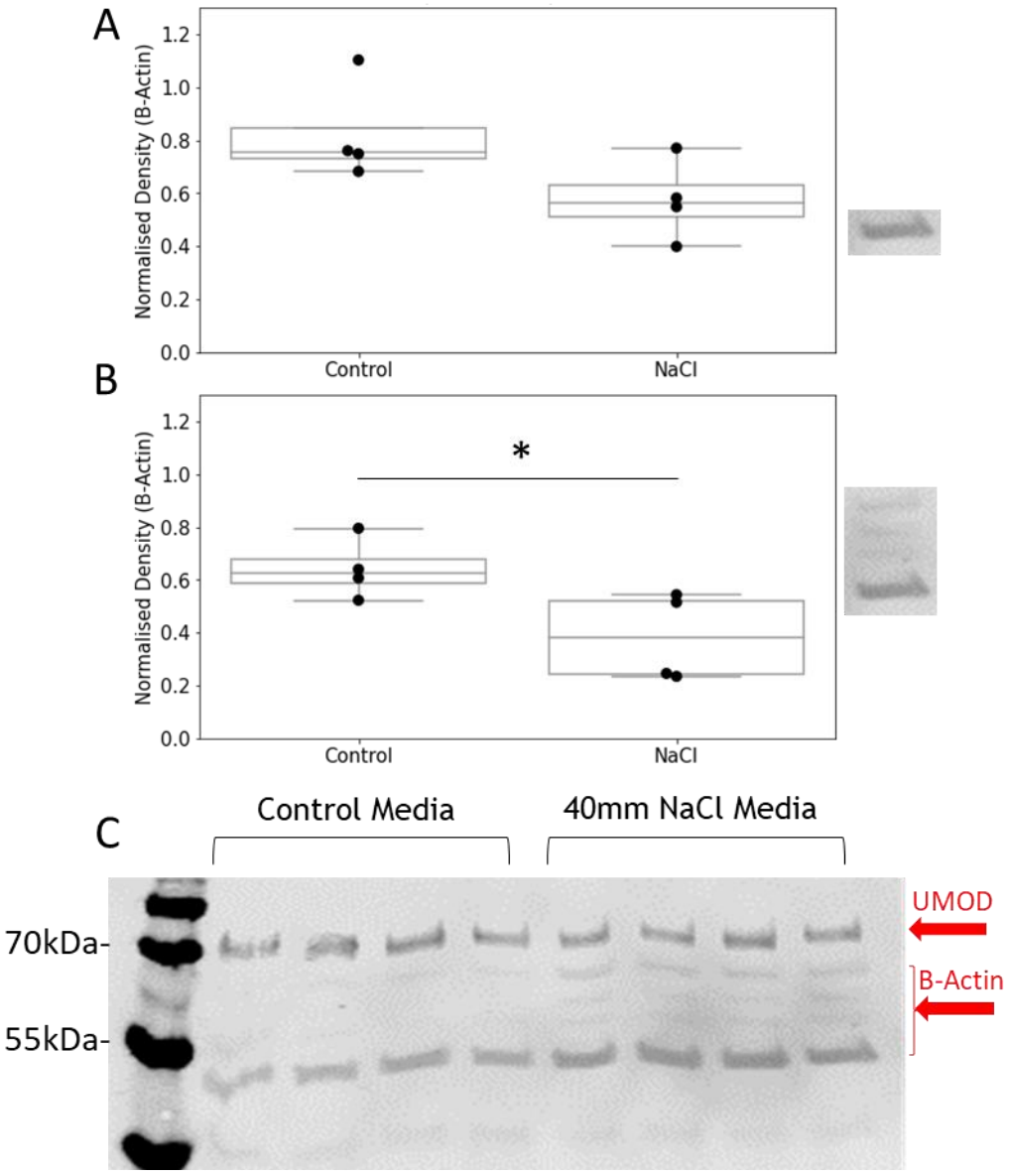


Figure 5.4-6 Effect of sodium chloride treatment on UMOD protein

A) Normalised density of UMOD protein (vs B-Actin) between control media and 40mM NaCl media. B-Actin produced a multi-isoform signal, 6A normalized UMOD expression to the dominant B-Actin isoform. Right side, dominant isoform. B) Normalized UMOD expression to the density of all four B-Actin isoforms ($p<0.05$). Right side, all four isoforms. C) Licor representative image of the observed relationship.

5.5 Discussion

Our attempt to generate a stable transfectant cell line which produces detectable quantities of both UMOD mRNA and UMOD protein was successful, providing us with a replenishing and abundant stock of material for downstream experiments. We evidenced by both MTT and housekeeper analysis that the transfection process and the presence of viral-promoter-expressed UMOD within HEK293 cells is not acutely (days to weeks) deleterious to these cells. Messenger RNA and protein UMOD expression were both equivalent to B-Actin levels and several thousand-fold greater at mRNA versus endogenous expression in untreated HEK292 cells. The variability in expression we observed in our individual transfectant populations is likely attributed to plasmid copy-number-variation and should be considered in downstream experiments (Kaufman *et al.*, 2008; Kim and Eberwine, 2010; Di Blasi *et al.*, 2021). It should be noted that such inherent variability in expression levels between 'biological n' (individual transfectants) could be exploited during experimental design, as this reflects the variability in the human population. We also show pilot data indicating that there is potential for expression levels to decrease over time, with our mRNA expression decreasing by roughly 15% on average over a two-week period of cell culture, in an n=3 experiment in different media.

Stable transfection involves random recombination events between the host genome and the introduced plasmid, thus downstream response adaptations may occur. Reduction in expression over time of target genes in stable transfectants is noted and may be attributed to resource competition (Szczesny *et al.*, 2018; Di Blasi *et al.*, 2021). Additionally, though not described in the literature, we would also consider reduced expression to be attributed to epigenetic responses in transfected cells. Although not specific to our experiment Rui et al suggest that episomal vectors undergo chromatinization in vivo, indicating the persistent effects of transgene expression are regulated in-part by histone modification (Riu *et al.*, 2007). Due to these possible effects, we would recommend that researchers making use of the pTARGET-UMOD HEK293 cells we have generated, only culture the cells when material is required.

We hypothesised that exogenous expression of UMOD may induce expression of endogenous NKCC2, as we had previously observed evidence of coexpression between these two genes in both bulk and single cell RNA sequencing and through qRT-PCR analysis. Our pilot data do not show a relationship in expression between transfected UMOD and endogenous NKCC2 in these cells. This is likely attributed to the fundamental differences between endogenous and induced gene expression. Several studies have reported interpretation caveats around induction of endogenous gene expression in response to transfection, with Jacobsen et al finding that less than 10% of transcriptomic changes in transient transfections are attributed to introduction of the target gene, rather are a cellular response to the insults associated with transfection (Lang *et al.*, 1995; Jacobsen *et al.,*, 2009). Whilst we performed our experimentation with stable transfectants, the UMOD expression in these cells was predominantly under the control of a CMV enhancer/promoter, producing non-eukaryotic expression, thus these HEK293-UMOD cells are likely to not be ideal models for coexpression analysis.

Our investigation into the effect of sodium chloride treatment in these cells returned challenging pilot data, due to interpretation of B-Actin expression. Should we choose to accept normalization against all B-Actin isoforms, sodium chloride significantly decreased expression of UMOD protein in these cells. This result is contrary to previously published data in humans which showed that urinary UMOD levels increased in response to elevated sodium chloride in humans (Torffvita *et al.*, 2004). We believe this is related to the contrasting specifics of our experimental design versus Torffvita et al. Torffvita et al measured urinary UMOD concentrations, therefore, they assessed extracellular UMOD protein levels; showing an increase in response to sodium chloride. In contrast, by purifying and harvesting cells, we assessed intracellular levels, showing a decrease in response to sodium chloride. We believe that sodium chloride stimulates the trafficking of UMOD through the cell, providing an increase in abundance of extracellular UMOD. The inverse of this effect being a decrease in intracellular UMOD, as it is more rapidly transported to the cell surface. In order to prove this within our experiments, we needed to detect extracellular UMOD, therefore optimization of detection of the media fraction in is

vital (Figure 5.4-5). We hypothesised that a sodium-chloride-dependent inverse relationship between intracellular and media fraction of the UMOD protein would be detected. Both serum-free and serum-containing media can be used for media protein analysis, however, we would recommend that future experiments attempting to analyse the media fraction of UMOD protein should incorporate a serum starvation step. We believe it is possible that the fetal-bovine-serum complement of standard media may be disrupting the western blot process (Shin *et al.*, 2019). In terms of experimental design, we elected to control specifically for the presence of both cations and anions, using MgCl2 as an osmolarity control, however, it may be recommended to use Mannitol to control for osmolarity as the introduction of magnesium may itself produce unconsidered effects (Traversari *et al.*, 2020)(Thermofisher, Technical Methods, 'Quantitative screening of the effects of hyper-osmotic stress on cancer cells cultured in 2-or 3-dimensional settings', 2019).

Our hypotheses relating to the trafficking of UMOD with the cell were not addressed during the series of experiments we conducted in this chapter, due to time constraints in conjunction with the effect of the Covid19 pandemic, which will be addressed in greater detail in Chapter 7. The effect of sodium chloride on trafficking has been explored before, in the context of blood pressure regulation. Collectrin, which interacts with the soluble N-ethylmaleiamide-sensitive factor attachment protein receptor (SNARE) was shown to promote membrane association of ENaC and aquaporin II (AQP2) in Wistar Kyoto (WKY) rats to a greater extent in those which were treated with 8% sodium chloride in chow versus 1%. Interestingly, collectrin is expressed under control of the same transcription factor known to regulate UMOD expression, HNF1B (Yasuhara *et al.*, 2008). Additionally, increased dietary salt is has been shown to increase phosphorylation of the oxidative stress response kinase-1/STE20/SPS1-related proline alanine–rich kinase (OSR1/SPAK), which in turn mediates vasopressin activity (Chiga *et al.*, 2008; Saritas *et al.*, 2013). This differential activity of vasopressin in response to sodium chloride has been shown to differentially phosphorylate and traffic NKCC2 (Mutig, 2017). We believe, based on evidence suggesting disturbances to cell-membrane bound UMOD being linked to NKCC2 activity, that UMOD and NKCC2 may be co-trafficked, and that sodium chloride

is functionally important here- possibly via our proposed pathway (Olinger *et al.*, 2019b). Based on these considerations, we believe trafficking assays are essential next steps in investigating UMOD.

We center our hypothesis on potential interactions between UMOD and NKCC2, as they traffic through TAL cells. However, using a hypothesis generating approach may be recommended to identify any potential binding partners. Co-immunoprecipitation approaches, using UMOD protein as the target should be the focal point of these experiments, but we would recommend purifying any pulldown and submitting these samples for mass spectroscopy. This would allow the characterization not only of individual binding partners, but also macromolecular complexes involving the interactions of multiple proteins. Such approaches have been discussed by other researchers and provide clear benefits over conventional antibody driven methods (Free *et al.*, 2009; Maccarrone *et al.*, 2017). We would like to highlight that Maccarone et al provide a comprehensive purification protocol for the steps prior to mass spectroscopy. It may even be recommended to pair these experiments with the RNA-sequencing data we obtained in Chapter 4, in a method similar to Zhang et al (Zhang *et al.*, 2020). Furthermore, as we have proven the very low level of endogenous NKCC2 in these cells, we would recommend the use of cotransfection with NKCC2 in order to produce adequate material for cotrafficking assays. Additionally, we emphasize that the tagging of our UMOD plasmid with green-fluorescence-protein (GFP) would provide additional benefits as this would facilitate visual tracking of UMOD in order to identify potential compartmentalisation (Kosugi-Tanaka *et al.*, 2006; Kuruppu *et al.*, 2013). However, we elected not to create chimeric constructs for these initial studies as we were wary of the effect of GFP on protein interactions between UMOD and targets, as chimera proteins featuring GFP have been noted to interact differently and be differentially modified to their canonical counterparts (Baens *et al.*, 2006; Skube *et al.*, 2010).

Within this chapter we developed a tool for classifying dead cells from living ones and background pixels. We did this as a repetitive task like cell counting lends itself highly applicably to code, and not to researcher based manual counting. The need for this kind of application is noted within literature, however, deployment lags

behind development, with Hu et al and Greca et al being the first groups to attempt to address this with advanced methods, though Greca's algorithm is more prognostic (Hu *et al.*, 2022)(Greca *et al.*, 2022). The cell classifier deployed here works to within 95% accuracy of ground truth annotation, however, it is unlikely to be transferable to other cell types due to its highly specific algorithm. We would suggest a deep learning approach be applied to this, reflected of the work previously described by Hu et al and Greca et al. However, our method may be a way of providing semi-supervised annotation for a deep learning method, making it potentially useful for other researchers.

In summary, we have successfully established a stable UMOD expressing human cell line via pTARGET-UMOD transfected HEK293 cells. Whilst our experiments within this chapter cover the successful generation of a cell-based resource, they do not comprehensively cover the role of UMOD protein within the cell. Based on preliminary data, we emphasise that these experiments should, as a priority, investigate UMOD trafficking in the context of sodium chloride. We recommend a coIP:mass-spectroscopy approach here, in order to elucidate as many cotrafficking targets as possible.

5.6 Chapter Specific Code

5.6.1 EVOS-Microscopy Cell Classification

```
import skimage.io
import skimage.feature
import skimage.morphology
import skimage.filters
from skimage.color import label2rgb
from skimage.measure import label, regionprops
from skimage.segmentation import clear_border
from scipy import ndimage as ndi
from matplotlib.collections import PatchCollection
import matplotlib.patches as mpatches
import matplotlib.pyplot as plt
import cv2
import imagecodecs
import numpy as np
from skimage.morphology import binary_closing, disk
from skimage.morphology import erosion, dilation, opening, closing,
white_tophat
from skimage.morphology import black_tophat, skeletonize, convex_hull_image
from skimage.morphology import disk
from skimage import (
    color, feature, filters, measure, morphology, segmentation, util
)
filepath='F:/EVOS/F.d8.500.0005.tif'
#filepath='F:/EVOS/F.d8.0000.0016.tif'
#filepath='F:/EVOS/F.d8.800.0008.tif'
image = cv2.imread(filepath)
gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
clahe = cv2.createCLAHE(clipLimit = 2.0, tileGridSize=(8, 8))
claheNorm = clahe.apply(gray)
cv2.imwrite(filepath+'claheNorm.tif', claheNorm)

fig=plt.figure(figsize=(50, 20))
img = skimage.io.imread(filepath+'claheNorm.tif')
img = img[:,:]

thresholds = filters.threshold_multiotsu(img, classes=5)
regions = np.digitize(img, bins=thresholds)
cells = img > thresholds[0]
deadcells = img > thresholds[2]
selem =  morphology.disk(7)
res = morphology.white_tophat(deadcells, selem)
highconfluency=deadcells ^ res
lowconfluency=deadcells
cells=highconfluency

distance = ndi.distance_transform_edt(cells)
local_maxi = feature.peak_local_max(distance, indices=False,
                                    min_distance=1)
markers = measure.label(local_maxi)
segmented_cells = segmentation.watershed(-distance, markers, mask=cells)


underlay = cv2.imread(filepath)
# remove artifacts connected to image border
```

```
cleared = clear_border(segmented_cells)
# label image regions
label_image = label(cleared)
image_label_overlay = label2rgb(label_image, image=underlay,
bg_label=0,alpha=0.4)

fig, ax = plt.subplots(2,2, figsize=(20, 16))
ax[0,0].imshow(image, cmap='gray')
ax[0,0].set_title('Basic EVOS')
ax[0,0].axis('off')
#ax[1].imshow(color.label2rgb(segmented_cells, bg_label=0))
ax[0,1].imshow(cells)
ax[0,1].set_title('Dead Cells')
ax[0,1].axis('off')
ax[1,0].imshow(color.label2rgb(segmented_cells, bg_label=0))
ax[1,0].set_title('Watershed Segmentation (Euclidian Distance)')
ax[1,0].axis('off')
ax[1,1].imshow(image_label_overlay)
ax[1,1].set_title('Overlay: (' + str(segmented_cells.max()) + " dead cells
counted)")
ax[1,1].axis('off')
plt.show()
fig.savefig('C:/Users/Administrator/Desktop/BlobinatorPro.png',
facecolor='w',bbox_inches='tight')
```

6 Exploring the UMOD Knock-out Mouse Model

6.1 Introduction

The *Umod* knock-out sv129 mouse is a highly important *in-vivo* asset in the study of the relationship between UMOD and blood pressure. Developed by Bates et al in 2004, these mice are based on an exon disruption event as opposed to promoter disruption, based on omega-type replacement targeting developed by Koller and Smithies in 1992 (Koller and Smithies, 1992; Bates *et al.*, 2004). Using a neomycin cassette as a selection medium, Bates et al produced an animal with a gene-dose dependent reduction (or in the case of knock-out animals, removal) of UMOD protein, specific to the first four exons plus the intervening introns of the mouse *Umod* gene. These *Umod* knock out animals were observed to grow and breed normally and their serum electrolytes, urinalysis, and kidney histology were normal. However, it was not until 2014 that these animals would be investigated in the context of blood pressure regulation.

Graham et al and Trudu et al explored UMOD expression in the context of hypertension by studying the contrasting effects of both transgenic *Umod* over-expression and the *Umod* knock-out model in mice, with both generating opposing findings which provided biologically relevant data in the context of the human UMOD hypothesis. However, both these studies utilized a 'homozygous' design, in effect, clinically relevant translatability in both studies may have been further emphasized by the inclusion of heterogenous genotype. Effectively, at a population level in humans, UMOD promoter genotypes are not binary, rather they are a Mendelian mixture of both homozygous and heterozygous individuals. Both rs13333226 and rs4997081 exist with predominantly homozygous genotypes, however, there exists a reasonably large population in both which are heterozygous (rs13333226 CEU: A/A: 67.3%, A/G: 30.1%, G/G: 2.7%)(dbSNP 2021). Reflecting this *in vivo,* using study of the heterozygous ($Umod^{+/-}$) genotype in mice may therefore provide novel insights and be more translatable to humans than study of the homozygous knock out ($Umod^{-/-}$) alone.

Physiological adaptions to reduced levels of gene expression, particularly germ-line changes, manifest differently to absent gene expression, thus leading to the potential identification of novel targets. Based on this understanding, heterozygous animal models have been explored in the context of other cardiovascular related diseases, producing novel findings not observed in homozygous animals. In the context of familial dilated cardiomyopathy, hetereozygous knock-outs of *Mylk3*, in mice displayed a 50% reduction in mRNA levels, yet interestingly the displayed a 75% reduction in protein expression, with a concurrent similar weighting in measured phenotype (a 23% fractional shortening reduction in heterozygous animals versus a 30% reduction in homozygous knockouts) (Tougas *et al.*, 2019). Additionally, cardiac myopathy in hetereozygous animals has been studied in zebrafish by examining the *tnnt2a* gene, where it was shown that hetereozygous mutations promote progressive cardiac remodelling in adult fish, whereas homozygous mutations are embryo-lethal and cannot be studied in mature organisms (Kamel *et al.*, 2021). Particularly relevant to this study is the finding in mice, where it was shown that global knock-out of *Hsd11b2* causes hypertension in mice, whilst a heterozygous knockout of the same gene in the same strain results in these animals exhibiting wild-type blood pressures but salt-sensitivity (Kotelevtsev *et al.*, 1999; Bailey *et al.*, 2011).

It is probable that heterozygous knock out ($Umod^{+/-}$) of UMOD expression affects both the blood pressure phenotype and underlying physiological drivers differently to a global knock out ($Umod^{-/-}$) in these animals, given that these differential relationships have been observed consistently in other models. Graham et al observed lower baseline blood pressures in Sv129 $Umod^{-/-}$ mice, and tolerance to dietary sodium loading as a method of increasing blood pressure versus wild type animals (Figure 6.1-1). Specifically, these animals were generated using a deletion of the mouse *Umod* by a specific deletion of a '2kb segment 5' of the capsite of the UMOD gene and the first four exons plus the intervening intron' (Bates *et al.*, 2004). Heterozygous animals from this colony were shown by Bates et al to have reduced expression of UMOD protein versus wild type counterparts and therefore the animals would plausibly exhibit the reduced blood pressure phenotype to a lesser (but still present) extent than their homozygous knockout counterparts.

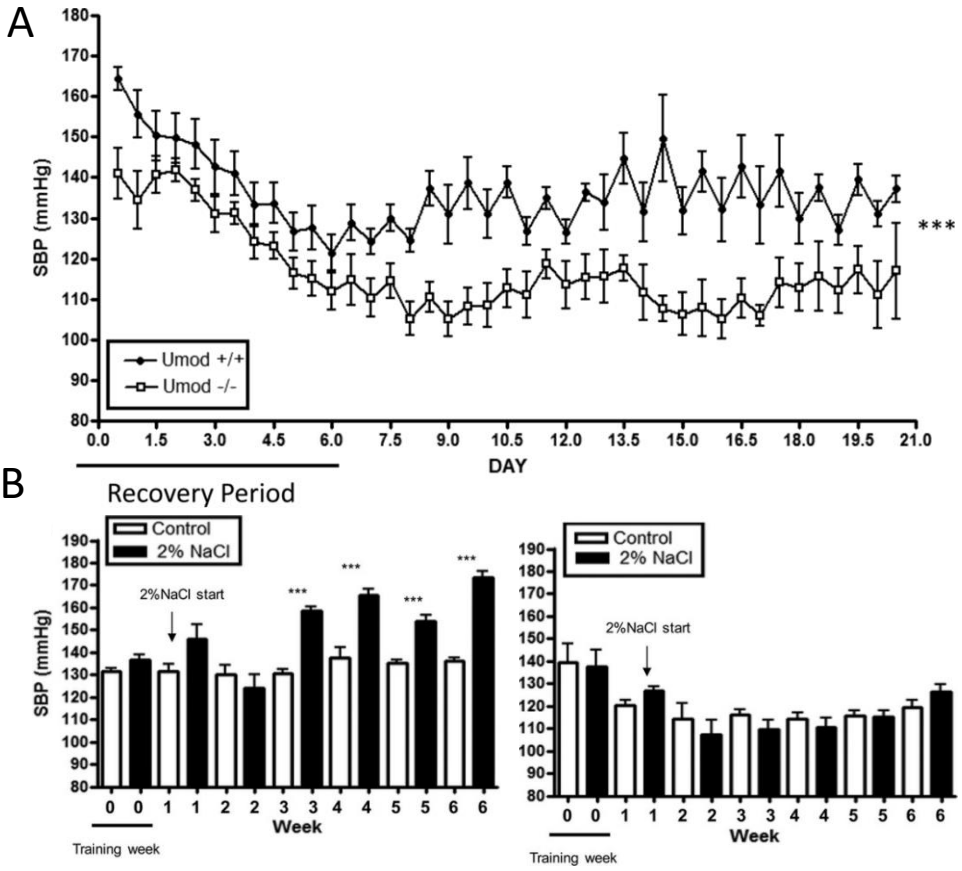


Figure 6.1-1 Graham et al Umod sv129 blood pressure analysis

Radiotelemetry observations of systolic blood pressures between Umod wild-type and knock-out male sv129 mice. Weekly tail cuff plethysmography systolic blood pressures between genotype and sodium chloride treatment. *Figures obtained directly from Graham LA, Padmanabhan S, Fraser NJ, Kumar S, Bates JM, Raffi HS, Welsh P, Beattie W, Hao S, Leh S, Hultstrom M, Ferreri NR, Dominiczak AF, Graham D, McBride MW. Validation of uromodulin as a candidate gene for human essential hypertension. Hypertension. 2014 Mar;63(3):551-8.*

As an initial priority, we would seek to validate the relationship between a germline homozygous knockout of UMOD (*Umod*$^{-/-}$) and reduction in blood pressure, using the tail-cuff plethysmography method. Following this, based on our hypothesis that the differences between TAL cell physiology between homozygous and heterozygous genotypes may drive the differential regulation of pathways, we intend to examine changes in gene expression between genotypes, in response to sodium induced

hypertension. In Chapter 4, we generated novel data associating UMOD with natriuresis and ion handling in the human kidney, which we would cross-validate using our animal model. Supplementing these data, Graham et al performed an RNA-Seq experiment on mTAL cells derived from their *Umod*$^{-/-}$ animals versus wild type. These data strongly corroborate our finding in humans, however, they also implicated two further notable targets; otopetrin 1 (*Otop1)* and Solute carrier family 7 (cationic amino acid transporter, y+ system), member 12 (*Slc7A12)* as potentially functional. Furthermore, whilst analysis of RNA may provide useful insights, protein-level interpretation may be more precise. In particular, it would be advisable to examine the presence of NKCC2 in both the homozygous and heterozygous genotypes by immunohistochemistry (IHC). As discussed in 5.1, current understanding is unclear on whether UMOD and NKCC2 colocalise during their trafficking through the cell, therefore we believe it would be pertinent to observe NKCC2 localisation by IHC to see if its distribution is dependent on UMOD in a dose-dependent manner.

6.2 Aims

- To repeat the observations made by Graham et al (2014) in the context of the *Umod*$^{-/-}$ sv129 mouse, in order to validate the model for present day study

- To enhance these observations by including the heterozygous genotype in analysis

- To follow up on any significant findings using qRT-PCR, western blot, immunohistochemistry and RNA-sequencing so as to infer mechanisms of blood pressure regulation by *Umod* in the mouse model

6.3 Specific Methods

6.3.1 Taqman qRT-PCR on Mouse Gene Targets

RNA was extracted from half mouse kidneys, sectioned laterally horizontally, as per 2.1.3. Superscript II reverse transcriptase (Thermofisher, Manchester, UK) was used to prepare cDNA, not Superscript IV, due to the relative abundance of high quality RNA versus applications using human RNA. Though the steps were identical to 2.1.5, the parameters of the reaction differed and were as follows (in particular, 600ng input RNA was used, with a 1:5 dilution of cDNA post-reaction);

Reaction Vessel 1	
Reagents	Volume (µL)
Template RNA	3 (200ng/µL stock)
dNTP (10mM)	1
Random Hexamers (50µM)	1
Nuclease free water	7
Cycling Conditions (Vessel 1)	
Time	Temperature (˚C)
5min	65
1min	4
Reaction Vessel 2	
Reagents	Volume (µL)
5x Superscript IV Buffer	4
100mM DTT	2
RNAseOUT (Thermofisher)	1
Superscript II Reverse Transcriptase	1
Vessel 1 Product	13
Cycling Conditions (Vessel 2)	
Time	Temperature (˚C)
10min	25
50min	42
15min	70

Taqman qRT-PCR reactions were performed against a number of genes-of-interest, all genes were normalised to the mouse ACTB housekeeper gene and analysed using the method described in 2.1.6;

Taqman qRT-PCR Probes	
Gene	**Thermofisher Assay ID**
Umod	Mm00447649_m1
Slc12a1	Mm01275821_m1
Sgk1	Mm00441380_m1
Pou3f3	Mm00843792_s1
Slc7a12	Mm00499866_m1
Otop1	Mm00554705_m1
Actb	Mm02619580_g1

Table 6.3-1 Taqman qRT-PCR Probe Ids

6.3.2 Determining UMOD Expression in Ang-II Treated C57BL/6 Mice

RNA was extracted from the kidneys of 14-week male C57BL/6 mice which were donated by Dr Ryszard Nosalski. These animals had undergone induced hypertension, confirmed by tail-cuff plethysmography using the same apparatus as this study, derived from the implantation of 14-day osmotic minipumps (Alzeta Corporation, San Jose, California) delivering a total of 490ng/kg/day mouse angiotensin-II (Sigma, Dorset, UK). Control animals consisted of sham treatment, with the osmotic pump delivering an infusion of vehicle for the same duration of time. Dr Nosalski culled these animals by terminal isoflurane anesthesia and cardiac puncture exsanguination before removing the kidneys and storing these at -80˚C until needed. The RNA derived from the kidneys of these animals was reverse transcribed as per 6.3.1 and subsequent Taqman qRT-PCR performed as per 2.1.6.

6.3.3 Imputation of Diastolic Pressure

Diastolic blood pressure measurements by tailcuff plethysmography were significantly more likely to return null values than systolic blood pressure, therefore imputation was used to infer missing measurements. In the case of missing values, these 'not-a-number' (NaN) readings were designated as 0. Any 0 values within the dataset were subsequently imputed based on the mean of the adjacent readings (+/- 1 week) for these animals. In total, 18/242 summarised measurements of DBP were imputed, representing 7.4% of the total number of readings.

6.3.4 Retrospective Power Calculations

Power calculations were performed to determine estimated sample sizes required per group in order to detect blood pressure differences, prior to the study, based on the standard deviations of blood pressure recordings by Graham et al 2014, in the initial study. New calculations were performed retroactively using the cumulative standard deviations of the set of blood pressure measurements at baseline in the present study. Prior to calculation, the Anderson-Darling test was used to confirm normal distribution of blood pressure recording within groups. A clinically meaningful difference of 5mmHg was assumed as a detectability threshold. A 1-β of 0.9 was used, with an α of 0.05. Where appropriate, a range of clinically meaningful differences were used from 2-20mmHg. All calculations were performed in Minitab (20.3.0.0).

6.4 Results

6.4.1 Analysis of Blood Pressure Differences Between Umod Allele and Sodium Chloride Treatment

Genotyping the *Umod* $^{+/+, +/-, -/-}$ sv192 mice via the novel developed assay, which was intended to captured and characterized heterozygotes alongside both the homozygous animal strains was successful, returning a clear and allele-specific banding (Figure 6.4-1).

Time series analysis of both systolic and diastolic blood pressures across the study between standard water and sodium chloride treatment detected no significant differences between alleles or treatment (Figure 6.4-2)(A)(Figure 6.4-4)(A). When these data were normalized to the starting blood pressure of the animal, using the δ-baseline method, no significant differences between allele or treatment were observed (Figure 6.4-2)(B)(Figure 6.4-4)(B). Analysis of both the SBP and DBP of 12-week male *Umod* $^{+/+, +/-, -/-}$ sv192 mice found no differences between *Umod* alleles at baseline (Figure 6.4-3)(A)(Figure 6.4-5)(A). A high degree of variability was detected in both measurements, with a mean standard deviation in SBP of 20.965 mmHg, roughly 20% of the blood pressure value. As a result, estimation of the population level statistics of the wild type sv129 mice we assessed suggests a baseline SBP of 131.79mmHg (119.96,143.62 95%CI), in heterozygous animals a baseline SBP of 131.49mmHg (118.32, 114.72 95%CI) and in knockout animals a baseline SBP of 138.66mmHg (115.99, 161.33 95%CI). The wide confidence margins of this estimation reflect the sample size calculations shown in (Figure 6.4-6) where, should the standard deviation remain stable with increasing sample size, a cohort of 371 animals would be required to detect a 5mmHg difference in SBP. Area under curve analysis, which captures the total effect of treatment across the study, of both the systolic and diastolic blood pressures of these animals (Figure 6.4-3)(B)(Figure 6.4-5)(B). indicated no difference between *Umod* alleles or 2% sodium chloride treatment across the study.

There was no difference in 24-hour urine output measured by metabolic cage between allele or treatment at 3 weeks (Figure 6.4-7)(A), however, at 6-weeks, 2% sodium chloride treated mice produced significantly more urine than their standard water counterparts ($p<0.05$) (Figure 6.4-7)(C). This difference, though not statistically significant, showed a similar trend in both the heterozygous and homozygous knock out animals at 6-weeks. There was no difference in fluid consumption between alleles or treatment at either 3 or 6 weeks, with a large degree of variability (Figure 6.4-7)(C/D).

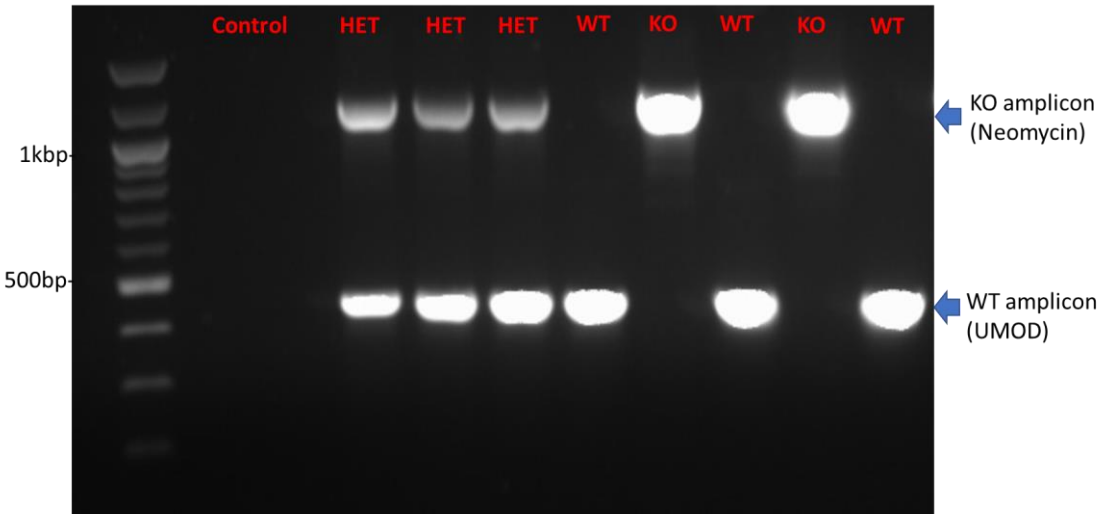


Figure 6.4-1 Genotyping of Umod $^{+/+, +/-, -/-}$ sv192 mice.

Amplification of the Neomycin reverse primer via the Neomycin selection cassette incorporated into the knock-out allele resulted in the amplification of a 1100bp fragment. Amplification of the wild-type primer pair against the canonical mouse Umod promoter to exon-2 resulted in the amplification of a 400bp fragment. Genotypes were distinguished by size. Fragments shown relation to New England Biolabs 100bp ladder.

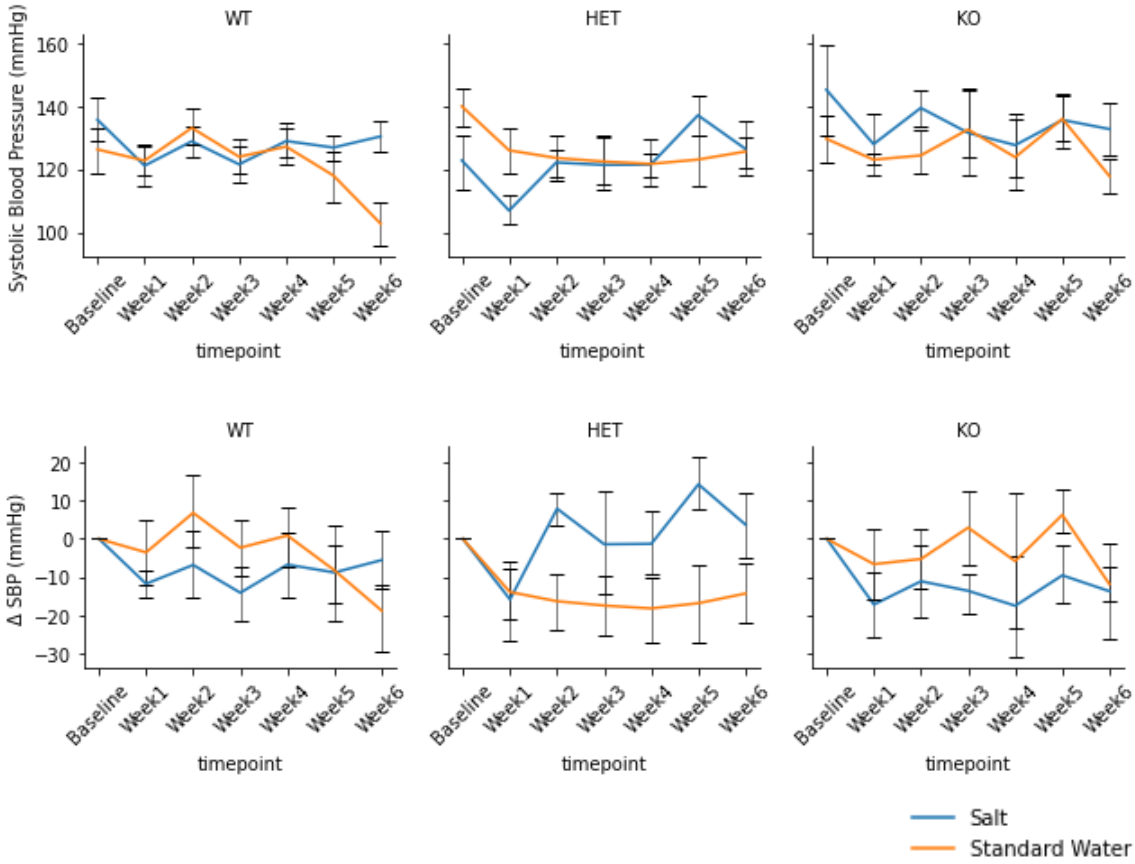


Figure 6.4-2 Measurements of systolic blood pressure (mmHg), stratified on genotype

(Graph columns) timepoint (X-axis) and treatment (colour [blue or orange]), data shown as mean +/- standard error (n measurements per group= 5-9). B) Delta (δ), difference in systolic blood pressure from baseline stratified on genotype (Graph columns) timepoint (X-axis) and treatment (colour [blue or orange]), data shown as mean +/- standard error (n measurements per group= 5-9)

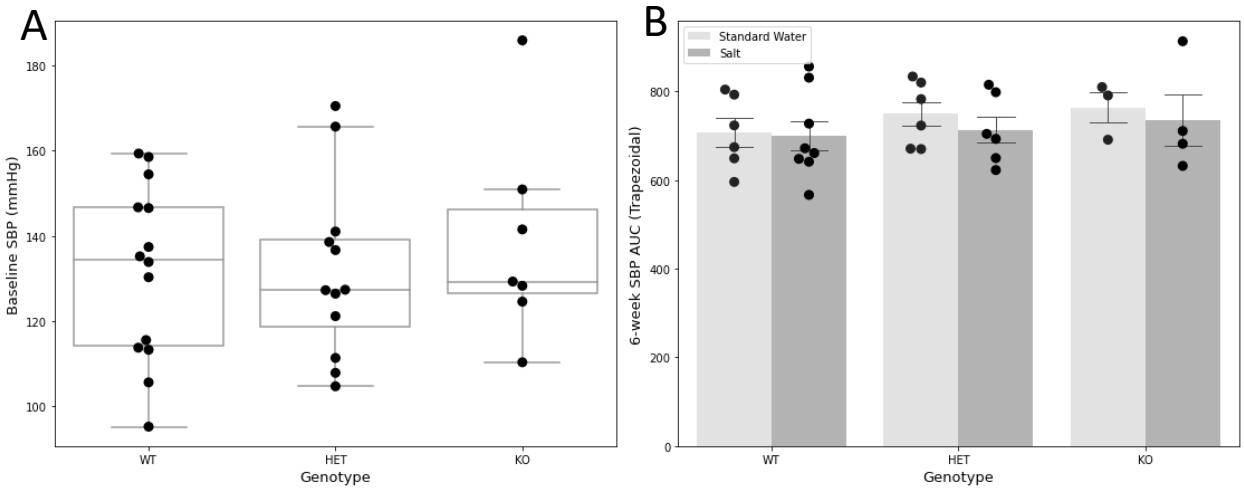


Figure 6.4-3 Assessment of summarised systolic blood pressure

A) Baseline systolic blood pressures (mmHg) of sv129 12-week male mice by Umod genotype; wild type, heterozygous and knock-out. B) Cumulative area-under-curve (Trapezoid) analysis of systolic blood pressure across the 6-week treatment period, stratified into standard water or salt (Sodium chloride 2%) treatments. Data shown as mean +/-standard error, n=3-9 per group.

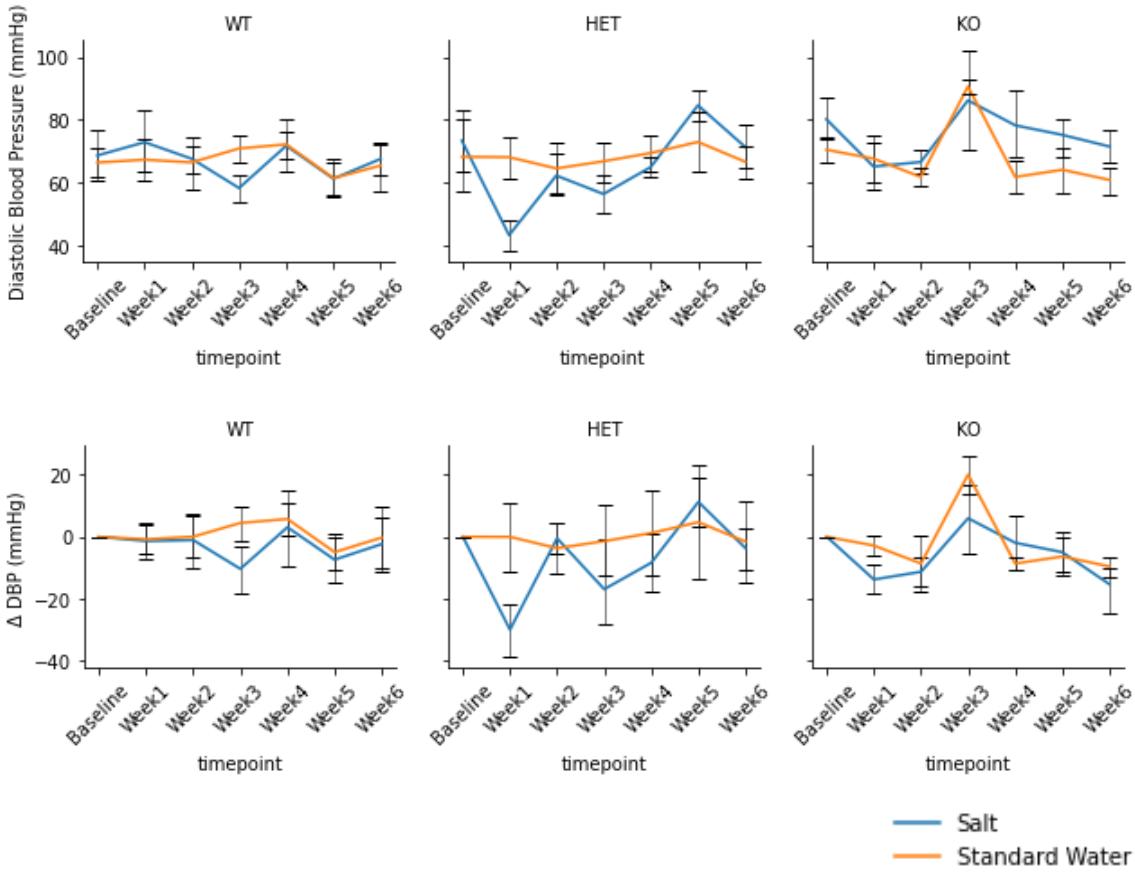


Figure 6.4-4 Measurements of diastolic blood pressure (mmHg), stratified on genotype

(Graph columns) timepoint (X-axis) and treatment (colour [blue or orange]), data shown as mean +/- standard error (n measurements per group= 5-9). B) Delta (δ), difference in systolic blood pressure from baseline stratified on genotype (Graph columns) timepoint (X-axis) and treatment (colour [blue or orange]), data shown as mean +/- standard error (n measurements per group= 5-9)

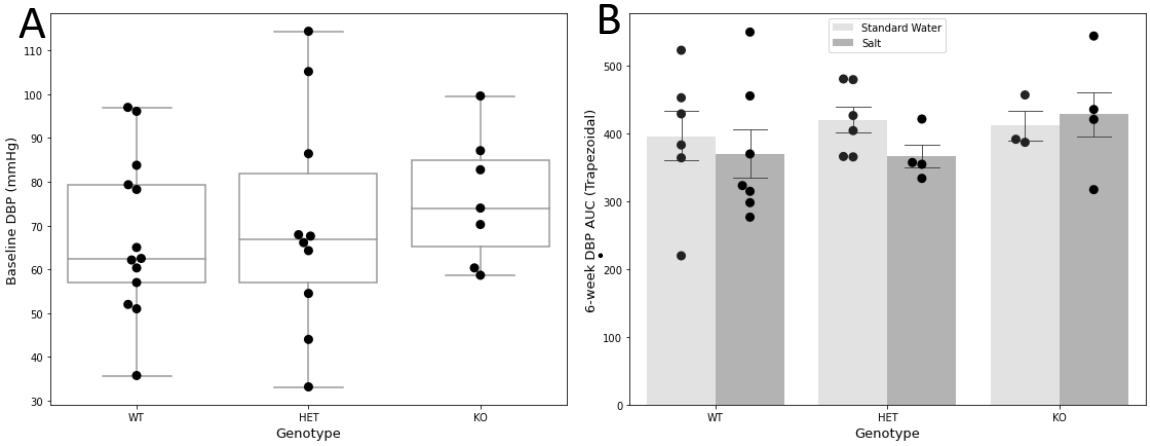
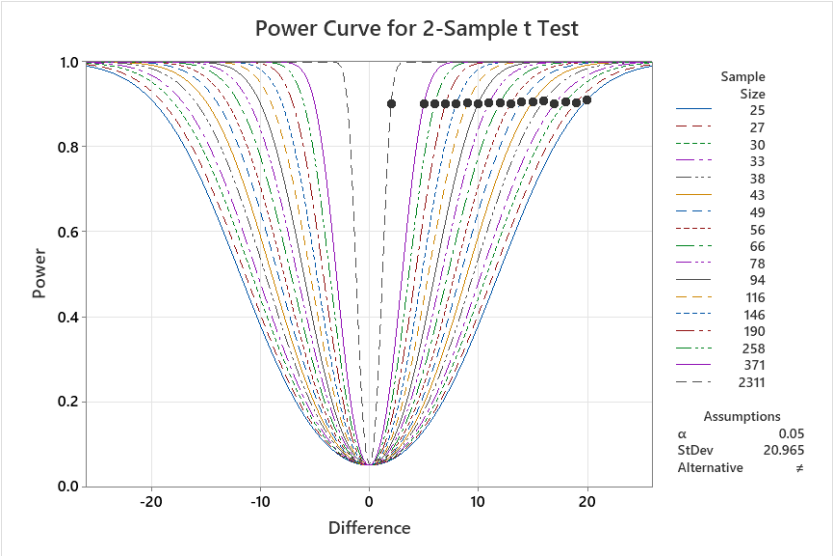


Figure 6.4-5 Assessment of summarised diastolic blood pressure

A) Baseline diastolic blood pressures (mmHg) of sv129 12-week male mice by Umod genotype; wild type, heterozygous and knock-out. B) Cumulative area-under-curve (Trapezoid) analysis of diastolic blood pressure across the 6-week treatment period, stratified into standard water or salt (Sodium chloride 2%) treatments. Data shown as mean +/-standard error. *7.4% of the values used to create the matrix for AUC analysis of DBP were imputed by adjacent median.*

A



B

Difference	Sample Size	Target Power	Actual Power
5	371	0.9	0.900438
6	258	0.9	0.900510
7	190	0.9	0.900798
8	146	0.9	0.901390
9	116	0.9	0.902462
10	94	0.9	0.902021
11	78	0.9	0.902544
12	66	0.9	0.903838
13	56	0.9	0.901894
14	49	0.9	0.905262
15	43	0.9	0.906484
16	38	0.9	0.907111
17	33	0.9	0.900431
18	30	0.9	0.904857
19	27	0.9	0.904528
20	25	0.9	0.910626

Figure 6.4-6 Retrospective power calculations

A) Absolute curves of two-sample power calculations for a range of theoretical systolic blood pressure differences between groups, based on the cumulative standard deviation of 20.965mmHg, α=0.05 and 1-ß of 0.9. B) Table of sample sizes ranging across theoretical blood pressure differences between groups, a difference of 5mmHg would require a sample size of 371 animals per group to detect, a difference of 20mmHg would require 25 animals per group.

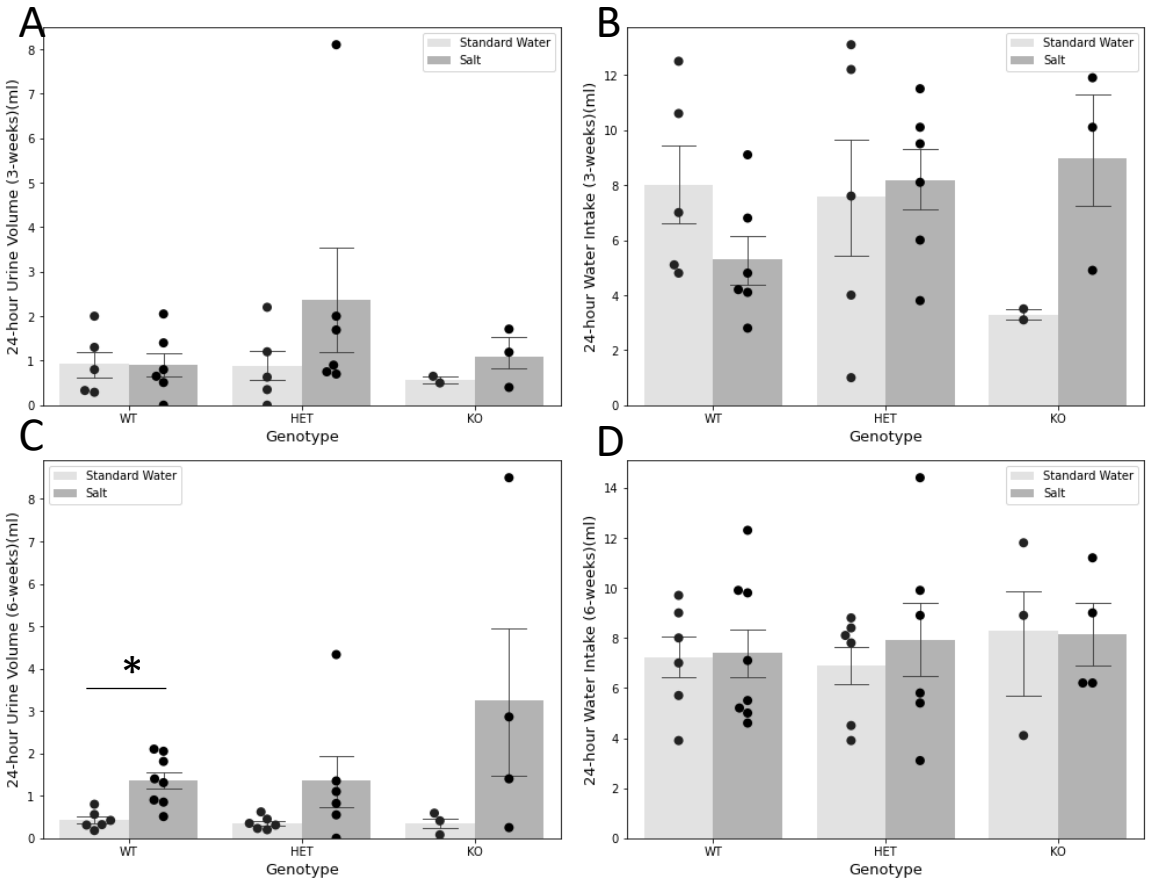


Figure 6.4-7 Assessment of urine output and fluid intake by met-cage

A) 24-hour metabolic caging urine output at 3-weeks of treatment (15- week old male Umod $^{+/+, +/-, -/-}$ sv192 mice). B) 24-hour metabolic caging water consumption at 3-weeks of treatment. C) 24-hour metabolic caging urine output at 6-weeks of treatment. D) 24-hour metabolic caging water consumption at 3-weeks of treatment. * = $p<0.05$ Bonferroni adjusted unpaired t-test. . n=3-8 per group.

6.4.2 Organ Weights at Cull

There was no difference between allele or treatment on the body weight at sacrifice of 12-week male *Umod* $^{+/+, +/-, -/-}$ sv192 mice nor was there a difference in the weight of the liver, left or right kidney or heart normalized to cull weight (Figure 6.4-8).

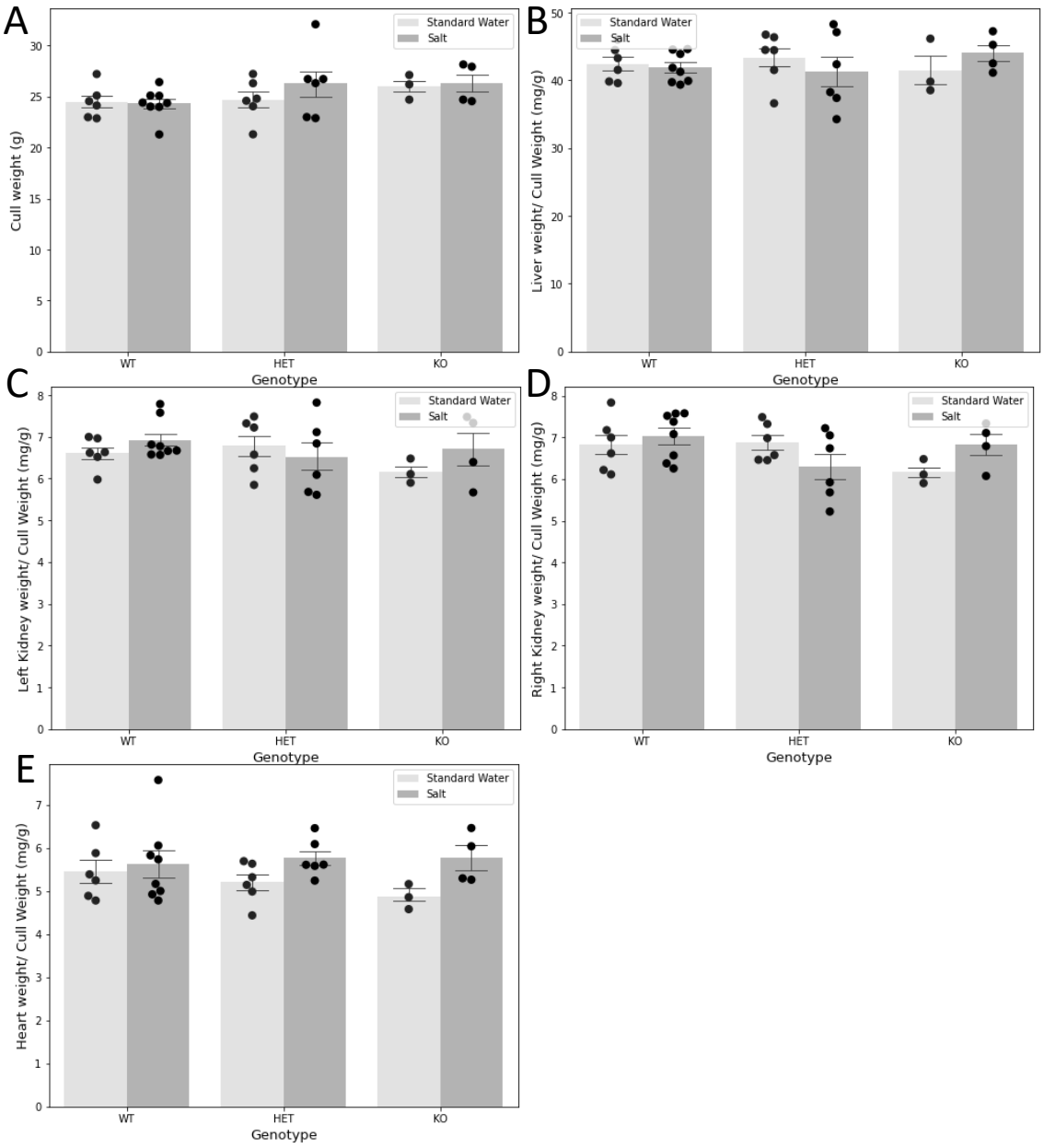


Figure 6.4-8 Cull organ weights

Measurements in grams (g) of body weights of 18 week old male Umod $^{+/+, +/-, -/-}$ sv192 mice. A) total weight, B) liver weight normalized to body weight (mg/g), C) left kidney weight normalized to cull weight (mg/g), D) right kidney weight normalized to body weight (mg/g), E) heart weight normalized to body weight. n=3-8 per group.

6.4.3 Taqman qRT-PCR Gene Expression

Expression of Umod was significantly decreased in the knock-out salt treated animal versus both treatment protocols in both wild type and heterozygous animals. However, neither treatment of heterozygous animals, nor standard water treated UMOD knockout animals displayed reduced *Umod* expression versus wild type counterparts ($p<0.01$ or $p<0.001$). Interestingly, expression of *Umod* was detected at ct22-25 knockout salt treated animals versus ct15-17 for wild type animals, indicating the non-negligible detection of *Umod* RNA in knock-out animals (Figure 6.4-9)(A). Expression of *Slc12a1* (Nkcc2) was significantly lower in salt treated heterozygous and knock-out animals versus their standard water counterparts (Figure 6.4-9)(A). *Sgk1* expression was significantly decreased in salt treated wildtype animals versus their standard water counterparts, with a similar trend in the knock-out animals, but not the heterozygous (Figure 6.4-9)(C). Expression of *Slc7a12* trended toward a decrease in salt treated knock-out animals (Figure 6.4-9)(D). There was no relationship between Otop1 expression and either allele or treatment (Figure 6.4-9)(E) and additionally no relationship between *Pou3f3* expression and allele or treatment was detected (Figure 6.4-9)(F).

When these data were cumulatively examined for correlations between the dCT values of each gene assessed by Taqman qRT-PCR, creating a linear correlation matrix using python, *Slc7a1* and *Slc12a1* (Nkcc2) displayed a significant positive correlation in expression. No other significant correlations were detected, although a trend for negative correlation between *Umod* and *Nkcc2* expression in these animals was detected (Figure 6.4-10).

Finally, the relationship between *Umod* and AngII treatment in independent 14-week male c57bl6 mice indicated no difference in expression between AngII treatment and vehicle control (Figure 6.4-11).

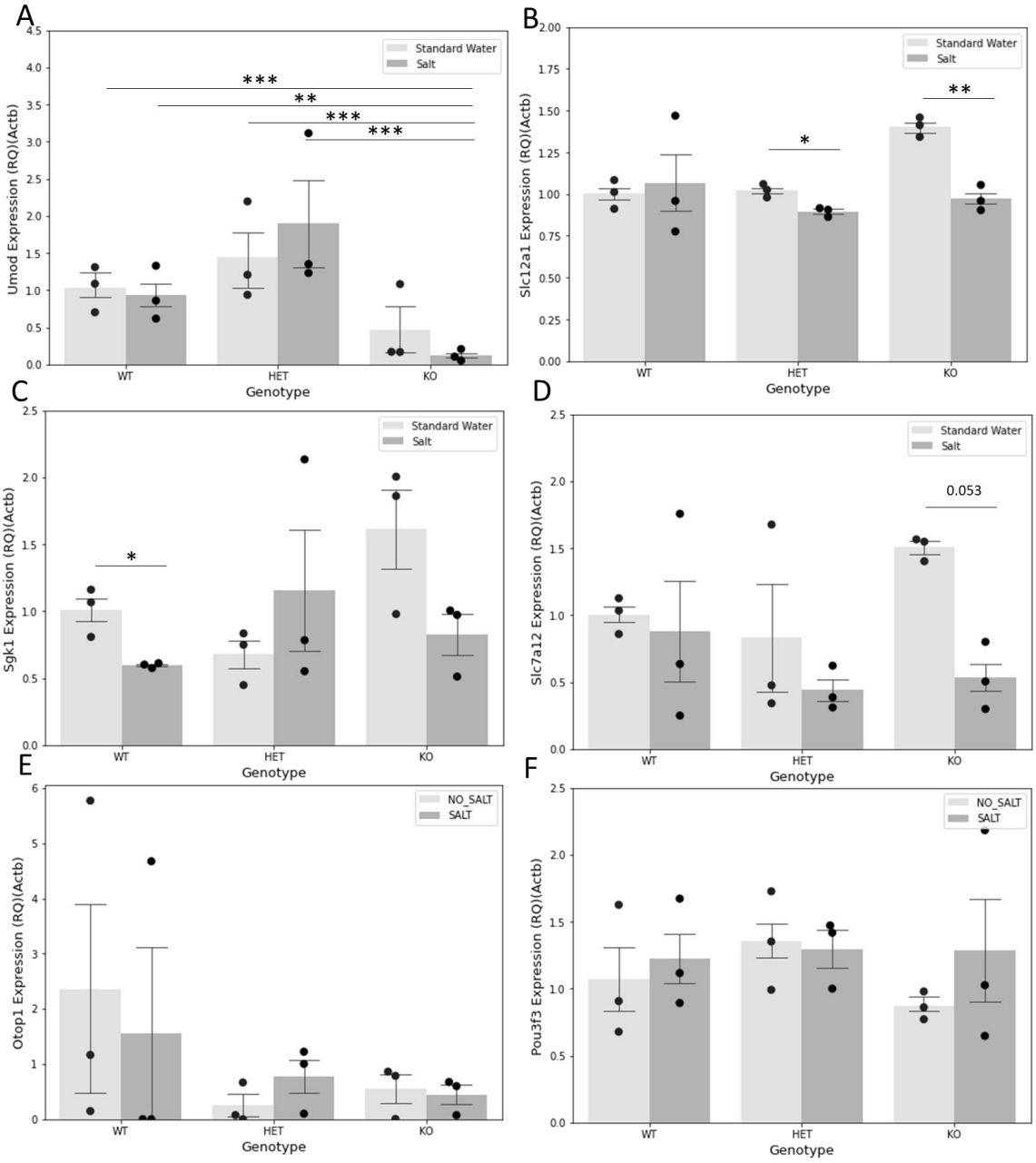


Figure 6.4-9 qRT-PCR analysis of genes relative to Actb

A) Taqman qRT-PCR expression of Umod relative to Actb in 18 week male Umod $^{+/+, +/-, -/-}$ sv192 mice. B) Taqman qRT-PCR expression of Slc12a1 (Nkcc2) relative to Actb. C) Taqman qRT-PCR expression of Sgk1 relative to Actb. D) Taqman qRT-PCR expression of Slc7a12 relative to Actb. E) Taqman qRT-PCR expression of Otop1 relative to Actb. F) Taqman qRT-PCR expression of Pou3f3 relative to Actb. * = $p<0.05$, ** = $p< 0.01$, ** = $p <0.001$, bonferroni adjusted unpaired t-test.

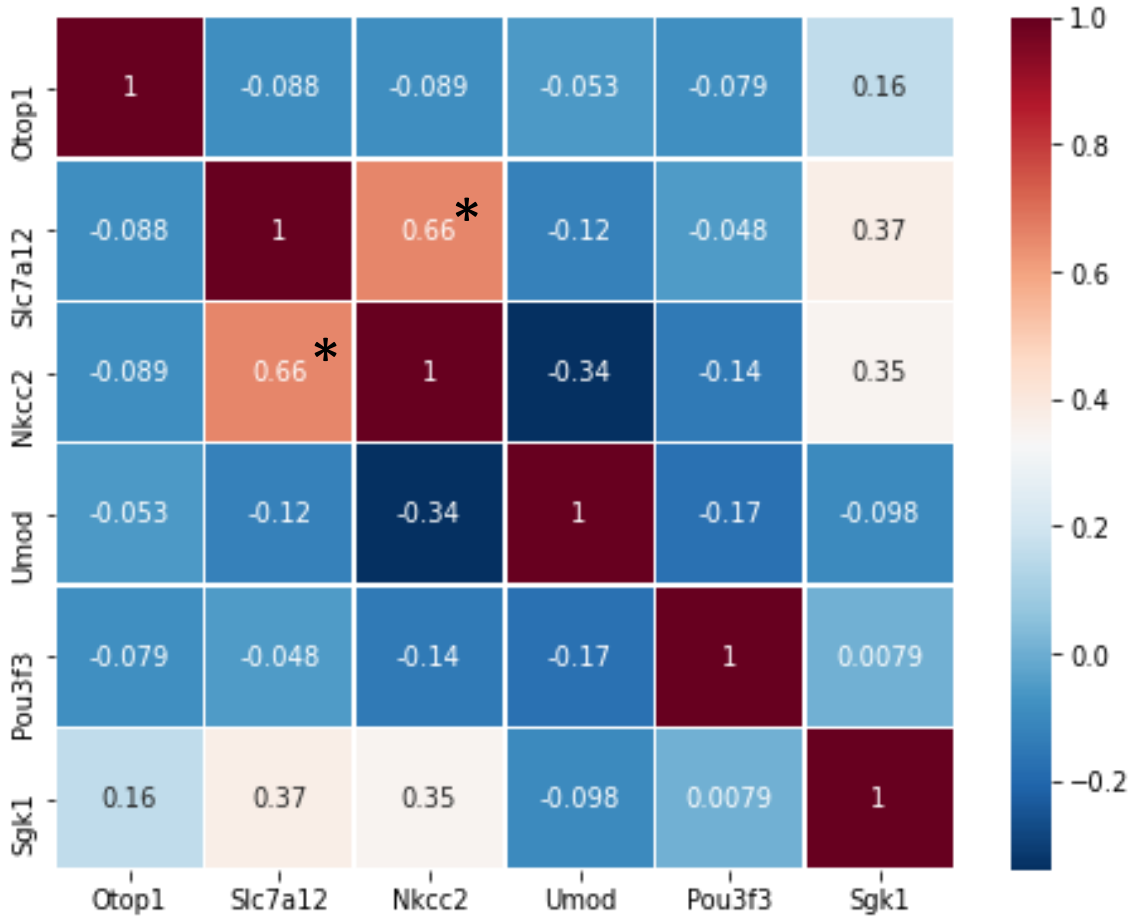


Figure 6.4-10 Correlation matrix of dCT between individual genes assessed by Taqman qRT-PCR.

Correlation performed on a per-sample basis, n=12 per sample. The positive correlation in dCT between Slc12a1 and Slc7a12 was significant Pearsons correlation, * = $p<0.05$

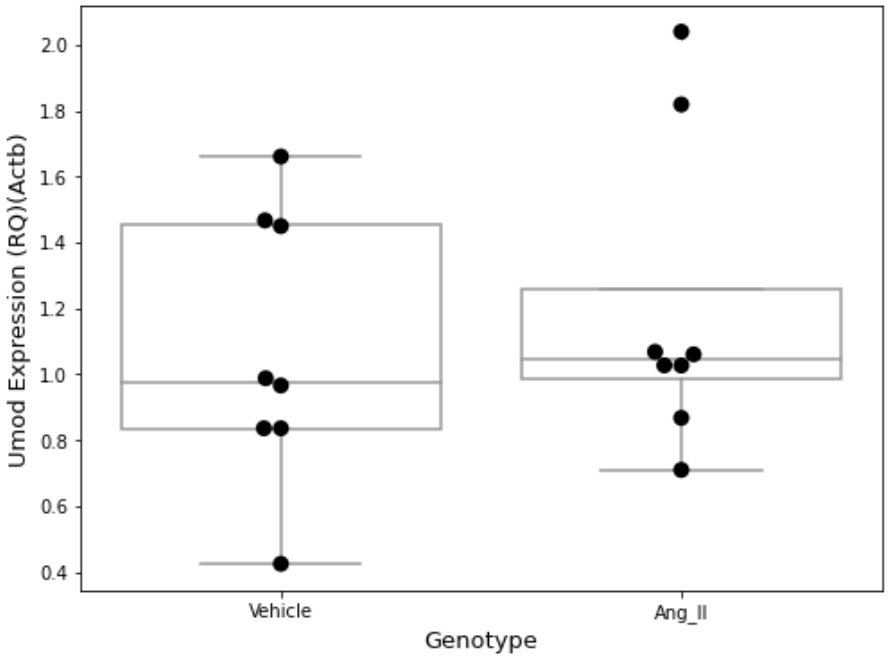


Figure 6.4-11 Assessment of Taqman qRT-PCR Umod expression vs Actb in 14-week male c57bl6 mice treated with vehicle or AngII
via ALZET Osmotic Pump (490ng/kg/day for 14 days). Renal RNA donated by Dr Ryszard Nosalski.

6.5 Discussion

The work conducted within this chapter form the basis of a critical shift between present-day analysis of *Umod* $^{+/+, +/-, -/-}$ sv129 mice and the previous study by Graham et al in 2014 (LA *et al.*, 2014). A disparity in results was observed in that the mice we assessed did not display allele dependent differences in blood pressure in response to sodium chloride treatment. The two foremost findings from the 2014 study were firstly that *Umod* knock-out mice display lower baseline blood pressures than their wild-type counter parts, and secondly, that the increases in blood pressures observed in wild-type animals in response to sodium chloride treatment are not observed in knock-out animals. Fundamentally, we did not observe a difference in baseline SBP between wild-type and knock-out animals with each analysis group in our study reflecting the wild-type genotype from Graham et al, clustering at 130mmHg **(**Figure 6.1-1)(A). With respect to their second finding regarding sodium chloride treatment, we did not observe a difference in response to sodium chloride treatment between genotypes across the 6-week period (Figure 6.1-1)(B). Graham et al specifically observed a significant increase in blood pressure in response to 2% sodium chloride in wild-type animals, which was not reflected in their knock-out animals. Crucially, we did not observe this wild-type specific increase in blood pressure in our study, thus, in effect our wild-type animals remained phenotypically similar to our knock-out animals. Whilst we did not observe the anticipated difference between wild-type and knock-out animals, we intended to build on the findings by Graham et al through including heterozygous animals in our study design. However, in terms of SBP at both baseline and 6-week sodium chloride treatments, we did not observe any significant difference to either wild-type or knock-out mice and nor did we observe a gene-dose-dependency effect.

The lack of replicability between Graham et al and the present study can be attributed to several factors, each of which may have varying causality. Firstly, our analysis of variability in tail-cuff plethysmography indicated a high degree of both within-sample and between-sample variance. As a result, the confidence intervals of our data between genotypes and also by treatments were wide across the study; this is reflected by the large sample sizes estimated retrospectively by power

calculations. The power calculations we performed prior to this study were correct, in that they reflected the standard deviation from the pilot data of the 2014 study,. However, it is possible that the technical aspects of sampling may have changed between 2014 and present. In 2004, a study reported that tail cuff method sampling can 'fail to accurately detect elevated blood pressures' and that these inadquencies cannot accurately be attributed to operator or technical issues (Whitesall *et al.*, 2004). Furthermore, these authors state they 'strongly advise investigators to obtain an independent and simultaneous validation of tail-cuff determinations of mouse blood pressure before making critical genotyping determinations'. Consequently, a number of subsequent papers note the inherent variability attributed to tail cuff plethysmography and how this can be an analytical confounder, with variability even attributed to the sex of the researcher performing the measurements (Wilde *et al.*, 2017; Drüeke and Devuyst, 2019). Our design elected to use the tail-cuff method based on a data-driven and educated decision; we had strong, peer-reviewed previous findings in the same strain of animals, in the same building, suggesting that this method produced adequately stable data for analysis. For undetermined reasons, our data were significantly more variable than the previous study. This may be attributed to animal anxiety, particularly via the handling procedure of the researcher, the presence of rats in proximity or even the the proximity of other individuals to the apparatus (Zhao *et al.*, 2011). Conversely, radiotelemetry based methods elimate the caveats around anxiety, due to it avoiding handling-induced stress. In terms of future work therefore, we would strongly recommend a transition to radiotelemetry based sampling due to its fundamental ability to eliminate a number of confounding issues in terms of variability (Feng *et al.*, 2008; Lemmer, 2016; Drüeke and Devuyst, 2019).

Within our analysis, we also observe a lack of blood pressure response in wild type animals to 2% sodium chloride, where it was previously observed by Graham et al after week-3 of treatment as significantly raising SBP. Whilst this may be attributed to the increased variability of measurements within our study, we do not rule out that these animals may not have responded to 2% sodium chloride due to physiological reasons. While we believe the increase in 6-week urine volume in response to 2%

sodium chloride in our animals is indicatory of physiological responses, it is possible that an increased level of salt-challenge may be required in order to induce detectable blood pressure changes. In theory, this would be acheivable by titrating the dose of sodium chloride past 2%, up to a level which is at the ceiling of physiological tolerance in these animals. Up to 6% sodium chloride in diet has been published as tolerable in male sv129 mice, however, this was only performed over a 1-week period (Michael *et al.*, 2008). Additionally, a further study found that 10% fructose/saline solution was also tolerated in male sv129 animal for four weeks (Ackermann *et al.*, 2011). Though these studies followed a shorter time course than our 6-week experiment, they did observe differences in blood pressures versus controls and they did not report a large number of animal deaths in their manuscripts (though this does not exclude the possibility of unreported high death rates). Although our decision to use 2% sodium chloride was based on the study by Graham et al, we would recommend in future that our sv129 strain be reassessed for its tolerance to sodium chloride across a range of concentrations higher than 2%.

The lack of gene dose dependent effects between *Umod* and blood pressure that we anticipated could also be attributed to genetic effects with this strain. Genetic drift describes the incorporation of spontaneous (de novo) and randomly sited mutations. Modifications which could in turn act to diminish or abrogate the defining characteristics of knock out strains versus their wild type counterparts (Cariappa *et al.*, 2009; Zeldovich, 2017). Genetic drift is typically characterised as significant when the expected characteristics no longer manifest, and typically develops over more than ten generations (Fontaine and Davis, 2016; Zeldovich, 2017). According to the International Knockout Mouse Consortium (IKMC), up to 58% of mouse strains do not fully represent the original knock-out phenotypically, highlighting the prevalance of this issue (Fontaine and Davis, 2016). Based on the present characteristics of our animals, if we disregard the discussed technical variability within the study, we believe it is likely that our *Umod* $^{+/+, +/-, -/-}$ Sv129 mouse model may have developed genetic drift over the seven years between the Graham et al study published in 2014 and present. Whilst the presence of genetic drift may be a confounding variable to the study presented in this chapter, it may also be beneficially exploited. If

previously, *Umod* knock-out mice had reduced blood pressures and now they do not, it means we presently have access to an environmentally controlled, case versus control difference in genome to explore between blood-pressure-different and blood-pressure-similar animals. We would recommend NGS be performed on samples between 2014 and now, with a sequencing depth of at least 30X, to identify these potential de-novo mutations. Such a study could be used to identify genetic markers of blood pressure regulation and could be extrapolated to the human population data present at UK Biobank.

Another confounding variable within this study, is our finding by Taqman qRT-PCR analysis that UMOD expression is non-negligible in the knock-out strain. We validated this across all knock-out samples indicating this finding cannot be attributed to technical errors within genotyping. This finding should therefore be regarding as a concern. Bates et al in 2004, when they developed this model, validated by northern blot a gene dose dependent decrease in *Umod* expression and no detectable expression in the knock-out (Bates *et al.*, 2004). They specifically state that the knockout is characterized by a deletion of a '2kb segment 5' of the capsite of the UMOD gene and the first four exons plus the intervening intron'. We mapped our TaqMan probe to the boundaries between exons-2 and 3 on the 2021 GRCm38 genome build. This indicates to us that the genome of our knockout animals must contain these exons in order for RNA to be detected by TaqMan. Additionally relevant is that there was a 10-fold decrease in the presence of this RNA versus either wild-type or heterozygous animals. This suggests that the *Umod* knock out animals may not be full knock-outs, and that marginal expression still remains. Due to the use of a cassette within exon-3, it is somewhat plausible that we may be measuring mRNA targeted for nonsense-mediated-decay (NMD); and this would go further to explain the significantly lower quantities of this RNA versus other alleles. However, Bates et al did not discuss this as part of their design, so this remains speculation without empirical data. We would comment that this study was conducted in 2004 and used the tools available at that time; present day sequencing of these animals would be necessary to prove this theory. Irrespective of mechanism, we would stress that it

would be highly pertinent to measure *Umod* protein levels in these animals as fundamentally physiological effects are exerted primarily by proteins and not mRNA.

Our detection of increased levels of *Nkcc2* in standard water treated animals versus wild type, in a *Umod* gene-dose-dependent manner, was highly interesting. This finding may underpin a large degree of the disparity between the Graham et al study and our study, given they did not detect this. Furthermore, it may explain the lack of observed baseline differences in systolic blood pressure between wild-type, heterozygous and knock-out animals. We discussed in Chapter 4 that the relationship between UMOD and blood pressure is likely to be multifactorial but predominantly underpinned by NKCC2. We speculate that the pathways and mechanisms affected by reducing *Umod* levels in these animals, which acted to reduce blood pressure in the 2014 study, may now be negatively balanced by the increases in Nkcc2 we observe. Of additional interest to use here was the significant correlation in expression between *Nkcc2* and *Slc7a12* in these animals. McBride et al showed in 2006 at the Annual Scientific Meeting of the British Hypertension Society via conference paper that *Slc7a12* was significantly enriched between spontaneously hypertensive stroke prone rats (SHRSPs) versus Wistar-Kyoto (WKY) control counterparts, suggesting a potentially functional role in blood pressure regulation. Furthermore, it has been shown that inhibition of *Slc7a12* expression leads to blood pressure decreases in Dahl Salt Sensitive rats (Crespo, 2015). Our data build on these studies by indicating that *Slc7a12* may exert these effects through *Nkcc2* in the kidneys.

Although the present study ultimately produced data showing that there is no relationship between *Umod* and blood pressure in our strain of Sv129 mice, we counter this by suggesting it is possible that the true relationship was not observed due to one or more of technical or biological effects discussed here. We would suggest that, due to the potential high variability of tail cuff plethysmography, future work incorporate the use of radiotelemetry either alone or alongside tail cuff plethysmography as a validatory metric. Furthermore, due to the apparent high inter-animal variability, it should be considered that an increased sample size be used in any further experiments here. Should the lack of differences in baseline blood pressure between alleles be attributed to genetic differences between the 2014

animals and the present animals, then we believe this is an asset which should be studied through NGS. Firstly, we suggest genomic analysis for de novo mutations. However, based on our unanticipated findings by TaqMan, which implicated *Nkcc2* as potentially countering *Umod*-based blood pressure changes, we would furthermore suggest a transcriptomic comparison between the 2014 and present day *Umod* knock-out animals. To this end, we applied for funding for this experiment via the Wellcome Trust Institutional Strategic Support Fund (ISSF). Funding was awarded with the work supported by Wellcome [204820/Z/16/Z] and though, outside the time restraints of this thesis, we expect to have greater clarity on these genetic changes shortly.

7 General Discussion

7.1 Conclusion

Understanding the mechanisms controlling blood pressure regulation in humans is a significant area of research due to the implications for patients. With the advent of precision medicine as a treatment paradigm, using basic science to link genetic variants with possible blood pressure regulatory mechanisms is likely to promote the development of personalized hypertension therapy. In 2010, Padmanabhan et al identified the UMOD gene, specifically the promoter, as a potential hypertension precision medicine target given their finding that the protective allele of rs133333226 was associated with decreased urinary UMOD levels, a 13% reduction in hypertension risk and a 7.7% reduction in the risk of incidence of cardiovascular disease (Padmanabhan, Melander, Johnson, Di Blasio, *et al.*, 2010). Following this discovery, both Trudu et al and Graham et al associated this increase in UMOD levels with effects on natriuresis via NKCC2 in mammals, though they proposed differing mechanisms driving this effect (Trudu *et al.*, 2013b; Graham *et al.,* 2014b). Where Trudu et al proposed that UMOD modulates the phosphorylation and thus activity of NKCC2, Graham et al suggested that UMOD modulates the TNFα regulated expression of NKCC2. The precision medicine hypothesis for UMOD states that, individuals who are risk-genotyped for rs13333226, will have greater levels of UMOD expression and therefore greater *levels or activity* of NKCC2. As a result, these individuals are more likely to respond more optimally to loop diuretics (which target NKCC2) than their protective or heterozygous counterparts.

Prior to precision medicine, there is a need for detailed understanding of underlying mechanisms, so that researchers can design clinical experiments. In this thesis, we identified potential knowledge-gaps in the UMOD precision medicine hypothesis, which we aimed to address using basic science, so they could be utilized downstream in precision medicine research. Firstly, functional causality had never been ascribed to rs13333226 in terms of this variant specifically driving expression of the UMOD gene; prior to this, only a 3-3.9kb region of the promoter in mammals had been characterized as being required for expression of UMOD (Zhu *et al.*, 2002; Kim *et al*

2003). The linkage disequilibrium block in humans which contains rs13333226, though varying between haplogroup, encompasses at least a 2kb region of this promoter (NCI, Division of Cancer Epidemiology and Genetics). We regard the finding by Padmanabhan as identifying a locus, or haplotype, as opposed to a specific variant. Based on prior unpublished data using promoter luciferase activity assays, we regarded the masked variant rs4997081 as being the driver of UMOD gene expression. However, luciferase experiments require transfection of target cells and therefore may not reflect the endogenous genomic activity of this promoter. Our first research question therefore focused on showing causality at rs4997081, using a method which retained conditions reflective of the environment in TAL cells. Electrophoretic mobility shift methodology bases all material on endogenously derived transcriptional apparatus. In utilizing this technique, we believed we could accurately identify transcription factors associating at the variant and identify this using mass spectrometry; thus allowing interpretation of transcriptional pathways.

Our second research question focused on the effects of UMOD in the cell. Firstly, the lack of consensus between the two published theories as to the relationship between UMOD and NKCC2 suggests that there may be additional experiments to undertake to further investigate this. Additionally, we were aware that no group had performed transcriptomic experiments in human samples; rather that prior to this, the focus had centered on NKCC2 using directed experiments. The ubiquitous nature of UMOD in the kidney is reflected by its involvement in varied processes, ranging across calcium handling, immune function and chronic kidney disease (Thomas *et al.*, 1993; L. J *et al.*, 2001; T *et al.*, 2003; M *et al.*, 2010, 2018; Baaij *et al.*, 2013; Micanovic *et al.*, 2015; MTF, J and M, 2019) . We believed in designing this aspect of the study that the relationship between UMOD and blood pressure may be multi-modal, with TAL cells associating with UMOD levels through a number of pathways, the primary of which was NKCC2. As a result, we elected to perform an RNA-sequencing style experiment, allowing us to generate data-led hypothesis relating to potential 'non-canonical' relationships between UMOD and blood pressure.

No blood pressure study to date has described a 'story', from GWAS variant to cellular processes underlying pathophysiological mechanisms. Such translations from GWAS

to function are reported to be lacking, and are regarded as a fundamental barrier to clinical translation (Cano-Gamez and Trynka, 2020). Through the two primary experimental directives in this thesis, we aimed to generate a 'story' of UMOD, combining our EMSA based analysis or promoter variants with our bulk-RNA sequencing transcriptomic analysis.

We showed, using material generated from two human renal cell lines (with a complement of *human* transcriptional apparatus) that rs4997081 has differential binding affinity between the risk and protective genotype; identifying the risk genotype as binding significantly greater amounts of nuclear lysate. Conversely, we showed that rs13333226 does not appear to bind transcription factors in a similar manner. Furthermore, we showed that with TNF-α treatment, the differences in binding between risk and protective genotype at rs4997081 normalise, with the increased binding in risk reducing toward the protective. In a n=1 pilot study using mass spectrometry on purified binding partners at rs4997081, we identified PARP1 as associating at this locus. We believe, consistent with this hypothesis and based on extrapolation through numerous studies, that TNF-α mediates the binding of PARP1 to rs4997081, which phosphorylates NF-κB, in turn activating HNF1ß to transcribe UMOD (Bachmann *et al.*, 2005; Weaver and Yang, 2013; Ray *et al*, 2017; Alves-Lopes and Touyz, 2018; Liang *et al.*, 2018b; Martí *et al.*, 2021). We examined the renal transcriptome between participants with high and low levels of UMOD expression and significantly show enrichment for TNF-α and NF-κB in the high-UMOD group, which provides independent corroborative data to this theorised pathway. In short, the presence of TNF-α appears to be protective, based on our theory and supporting evidence by Graham et al and Heitmeier et al, TNF-α negatively correlates with UMOD levels (LA *et al.*, 2014) (Heitmeier *et al.*, 2014).

Levels of UMOD mRNA in TAL cells correlated strongly with NKCC2 expression in our RNA-sequencing data, a finding which we expanded to 84 samples, further evidencing a very significant correlation in expression. We believe, based on these data, that UMOD and NKCC2 are coexpressed. We suggest that POU3F3 may mediate this coexpression, however this hypothesis requires further investigation. Uniquely to our investigation, we build on the findings relating UMOD to blood pressure through

NKCC2 by detecting the increased expression of WNK1/4, KCNJ1/5, and SCNN1G; signalling molecules and ion channels/transporters all known to cause increased activation of NKCC2 or facilitate sodium transport themselves (Ring *et al.*, 2007; Murthy, Kurz and O'Shaughnessy, 2017)(Welling and Ho, 2009). We also detected a decrease in expression of SGK1 and AVPR1, which may indicate homeostatic mechanisms acting to reduce phosphorylation of NKCC2 and inhibit the activation of the RAAS pathway in the as an adaptive mechanism to increased UMOD levels and the subsequent increases in natriuretic processes through NKCC2. Our 'story' of UMOD as we present, from variant to pathways can be viewed below (Figure 7.1-1). As studies move from mendelian disorders and their genes toward high-throughtput, complex genetic studies, the 'renal-polygenic' blood pressure rationale which we support with this thesis is increasingly considered to be important. Recently, in a multi-omics study of 430 human kidneys, 179 genes were enriched toward having causality with respect to blood pressure regulation in humans (Eales *et al.*, 2021). While Eales identified a strong enrichment toward UMOD pathways within their study, they also highlight a number of further gene pathways in the kidney which may have causality- this is important to consider, as whilst this thesis focuses on UMOD, it does not make the assumption that the UMOD locus is the only BP regulator within the kidney. Furthermore, with the much higher power of the Eales et al paper than our n=3 per group transcriptomics experiments, the data produced by Eales et al should be considered more robust.

In 1.4.7, we discussed the contrasting hypotheses of Graham et al and Trudu et al. Graham et al theorised that UMOD was linked to blood pressure by acting as a TNF-α sink, leading to reduced activator effects of TNF-α on expression of NKCC2. TNF-α was central to this hypothesis. Both our data from Chapter 3 and Chapter 4, the promoter and transcriptome studies respectively, independently yielded findings implicating TNF-α in the UMOD hypothesis. Furthermore, the findings in Chapter 4 were inherently unbiased and not directed toward TNF-α based experiments. In this thesis, we link TNF-α to UMOD expression through rs4997081. Reflective of Graham et al, we believe TNF-α to be a negative regulator of UMOD expression. As we believe UMOD and NKCC2 are coexpressed, based on strong evidence, we would also regard

TNF-α as being a negative regulator of NKCC2 expression. Thus, we believe our findings both build-on and reinforce the hypothesis presented by Graham et al. We stress however that additional experiments are required to prove the causality of TNF-α in this network, at this point we regard our novel data and pathways to be a pilot indication.

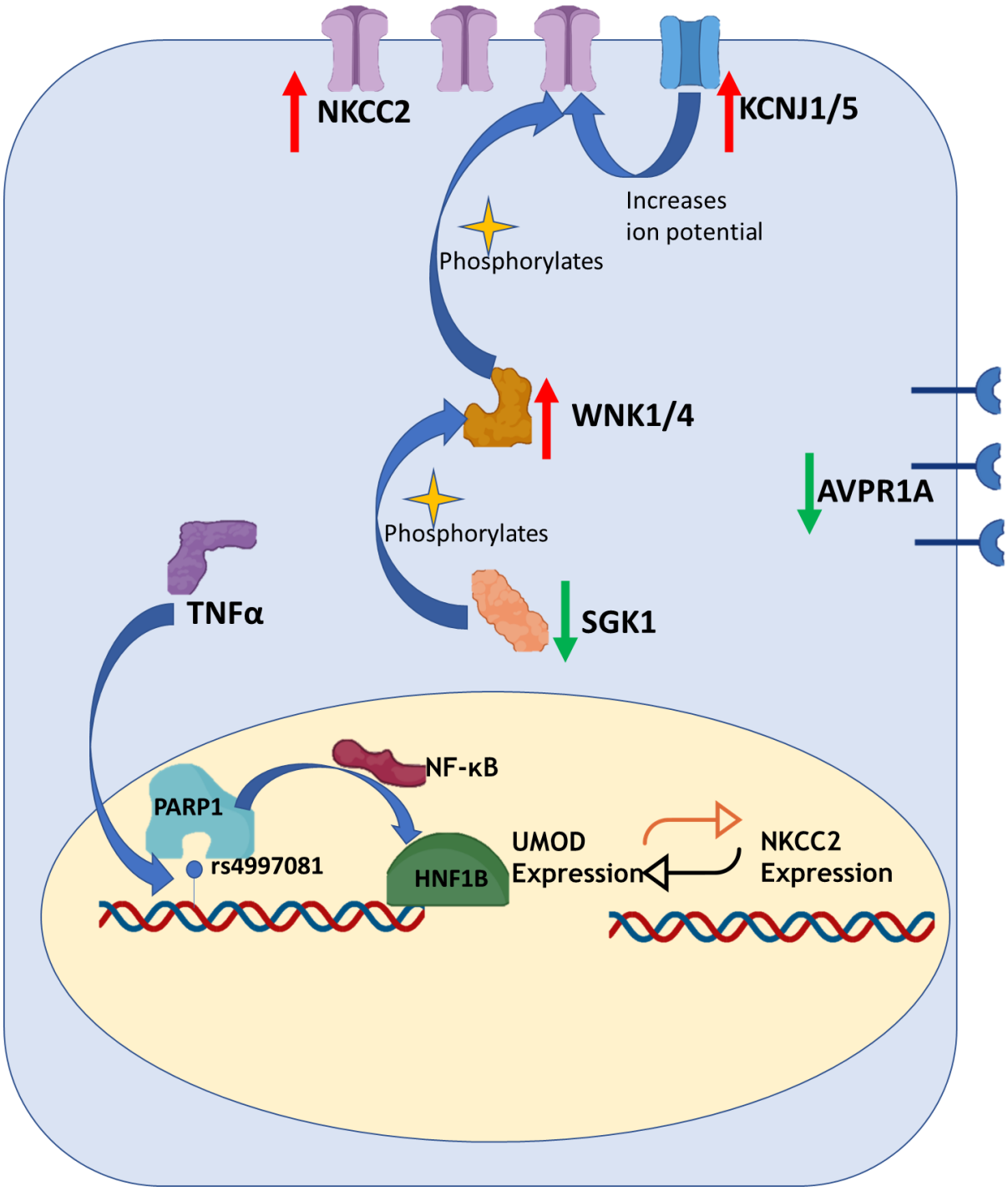


Figure 7.1-1 Proposed pathway between rs4997081 and differential regulation of natriuresis in human TAL cells.

Increased UMOD in response to differential PARP1 based activation of HNF1β drives increased UMOD expression. UMOD expression increases correlate positively with NKCC2 expression, alongside increases in expression of WNK1/4 and KCNJ1/5; which activate NKCC2. Decreases in SGK1 and AVPR1A expression may be compensatory mechanisms in response to heightened sodium reabsorption by NKCC2.

The research we present in this thesis, whilst novel and detailed, does have limitations. Firstly, our experiments on the sv129 *Umod*$^{-/-}$ mouse do not indicate a blood pressure phenotype in response to the disruption of expression of *Umod*. Additionally, we do not consider the blood pressure data of our human cohort to be clinically reliable and thus we cannot make conclusions based on blood pressure itself. As a result, no aspects of our study fundamentally link our discoveries back to blood pressure regulation *per-se*. We would however, counter this by suggesting that the extrapolation to blood pressure based on our highly significant findings of increased NKCC2 mRNA in correlation with increased UMOD mRNA is a strong and logical process. Additionally, we were restricted within this study by the lack of endogenous expression of UMOD within any commercially or academically available human cell lines. Human TAL cells are very difficult to obtain in viable states for cell culture. As a result, we chose to use renally derived HEK293 and HK2 cell lines to obtain nuclear material for our EMSA study. These cells are not strictly tubular epithelial cells, thus there is potential that the transcriptional apparatus within these cells does not reflect the complement with TAL cells. Lastly, our RNA sequencing experiment was stratified on expression of UMOD and not by rs4997081. As we discussed in 4.5, this was due to limitations within the cohort. As a result, we are not able to specifically relate the differences observed within our human RNA-sequencing experiment directly to rs4997081, rather these differences are indicative of changes to the transcriptome in correlation with differences in UMOD levels. Across the entire experiment we present highly novel data, yet we would stress that correlation is not causation; we have no evidence at present to indicate that UMOD directly causes these changes in the renal transcriptome, we suggest rather that UMOD correlates with these changes.

In terms of future work, we emphasise the necessity for rs4293393 and rs1297701 to also be investigated using the methodology we designed and optimized in Chapter 3. We recommend this in order to be inclusive of the total literature-highlighted variants on the UMOD promoter, though personally we feel rs4997081 is the foremost functional variant based on unpublished luciferase data by Graham et al. Furthermore, we believe that the sv129 *Umod*-/- needs extensive genomic

characterization to identify any potential drift, as this could be highly informative; particularly given that we detected significant increases in Nkcc2 expression in the *Umod*$^{-/-}$ strain. Additionally, we recommended that increased sample sizes be used during the mass spectroscopy analysis in order to provide greater reliability to these findings.

The prominence of the findings in both Chapter 3 and Chapter 4 lead us to believe that analysis of rs4997081 should be expanded to population level analyses. We discussed in Chapter 3, that rs4997081 was likely masked during the initial GWAS, thus we obtained access to UK Biobank data in collaboration with a colleague and assessed UK Biobank variant caller file (VCF) database for rs4997081. We discovered that rs4997081 is not genotyped directly meaning low-level multivariate regression between rs4997081 and the accurate UK biobank blood pressure recordings would not be possible. However, UK biobank also have 50,000 whole genome sequencing samples. From these, it would be possible to directly characterize rs4997081 and directly relate it to blood pressure. In theory, such an *in-silico* analysis could be performed on all variants on the UMOD promoter to construct a highly detailed understanding of the haplotype. This would, for the first time, link rs4997081 specifically to blood pressure and would be superior to the lower powered analysis we have performed from our cohort of 76 genotyped human sample. UK Biobank also have a repository of urine samples from their participants, from which UMOD and creatinine could be ratioed to infer UMOD levels in these patients. We also discovered a potential sex effect of the UMOD locus in relation to blood pressure, with females differentiating in blood pressure between alleles and not males. We strongly recommend this finding be expanded to population level in order to determine if this finding was artefactual due to our low sample size and potential unreliability with respect to this finding.

In Chapter 5 we laid out a framework for investigating the role of UMOD in trafficking, and the relationship between UMOD, trafficking and salt. Having developed a transfected stable UMOD cell line over several months, we had planned a number of trafficking experiments. However, we were made aware of highly exciting data regarding the UMOD promoter and resultingly pivoted our experiments toward this

research. We believe that the future work outlined in Chapter 5 should be regarded as a priority. Our experiments in this thesis did not address the actions of UMOD at the protein level, though fundamentally proteins exert physiological effects, not mRNA. Therefore, we would emphasize that cotrafficking assays in order to characterize potential interactions between UMOD and NKCC2 be performed in future.

The work in this thesis provides additional basic science insights as to mechanisms which may underpin the association of UMOD with blood pressure. We show empirically that rs4997081, not rs13333226, appears to bind transcription factors or complexes at the locus with differential affinity. Therefore, based on our findings, we would recommend that clinicians undertaking trials involving UMOD consider genotyping both rs13333226 and rs4997081 during these trials, as there is potential that the 8% disconnect in linkage between these two variants could present as a confounder. We suggest that rs4997081 exerts differential effects on expression of UMOD through the binding of PARP1 at this locus, in a potentially TNF-α regulated manner. We confirm that, at a basic-science level, the primary association between UMOD and blood pressure regulation appears to be through NKCC2, thus loop diuretics may present as the best precision medicine therapeutic agent during future work. However, this thesis emphasises the need to regard the prominence of TNF-α in this relationship, alongside the potential for SGK1, WNK1/4 and KCNJ1/5 to be therapeutically targetable. It should be noted that this thesis does not use reliable measurements of human blood pressure at any point, therefore we do not propose to frame these findings as ‘precision medicine’ themselves, rather that they provide a scientific basis for downstream experiments.

The work in this thesis was written during the Covid19 pandemic with a full laboratory closure lasting for 4 months and then further disruption due to the requisition of Biosafety cabinets by the Lighthouse Covid19 testing laboratory. The authors PhD spanned the disruption, which began during the experiments conducted within Chapter 5. These experiments were halted, and due to their cell-based nature, they could not be reinitialized until close to the end of the authors PhD. This did have a substantial impact on the flow and design of the experiments which could not be

avoided or worked around. During the pandemic the author aimed to maintain contribution by building and deploying open-source bioinformatics tools. These were successful in that the Padplot website was utilized by at least 6 individuals within the institute and, based on metrics from Amazon-Web-Services, many further remote researchers. The author is committed to the ongoing maintenance of these tools and will retain the right to keep these open source for academic researchers to use.

In conclusion, the work presented within this thesis provides additional insights into the mechanistic understanding of the relationship between the UMOD locus and blood pressure, surrounding rs4997081. These new insights are likely to contribute to the wider challenge of addressing hypertension management in humans, given they set a basis for new precision medicine experimentation. While our data are significant, there remains a need for further research with the further experiments we have outlined likely to feature prominently in future work.

8 Reference List

A, G. *et al.* (2000) 'Determinants of urinary excretion of Tamm-Horsfall protein in non-selected kidney stone formers and healthy subjects', *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, 15(10), pp. 1580–1587. doi: 10.1093/NDT/15.10.1580.

A, K. *et al.* (2009) 'Multiple loci associated with indices of renal function and chronic kidney disease', *Nature genetics*, 41(6), pp. 712–717. doi: 10.1038/NG.377.

A, K. *et al.* (2010) 'New loci associated with kidney function and chronic kidney disease', *Nature genetics*, 42(5), pp. 376–384. doi: 10.1038/NG.568.

A, L. *et al.* (2011) 'Molecular signatures database (MSigDB) 3.0', *Bioinformatics (Oxford, England)*, 27(12), pp. 1739–1740. doi: 10.1093/BIOINFORMATICS/BTR260.

A, M. *et al.* (2016) 'Gene-Specific Therapy With Mexiletine Reduces Arrhythmic Events in Patients With Long QT Syndrome Type 3', *Journal of the American College of Cardiology*, 67(9), pp. 1053–1058. doi: 10.1016/J.JACC.2015.12.033.

A, N., MN, P. and RJ, T. (2009) 'A conserved hydrophobic tetrad near the C terminus of the secretory Na+-K+-2Cl- cotransporter (NKCC1) is required for its correct intracellular processing', *The Journal of biological chemistry*, 284(11), pp. 6869–6876. doi: 10.1074/JBC.M804302200.

A, R. *et al.* (2018) 'Advances in the Genetics of Hypertension: The Effect of Rare Variants', *International journal of molecular sciences*, 19(3). doi: 10.3390/IJMS19030688.

A, S. *et al.* (1991) 'Vasopressin alters the mechanism of apical Cl- entry from Na+:Cl- to Na+:K+:2Cl- cotransport in mouse medullary thick ascending limb', *The Journal of membrane biology*, 120(1), pp. 83–94. doi: 10.1007/BF01868594.

A, T. *et al.* (2019) 'Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels', *Nature genetics*, 51(10), pp. 1459–1474. doi: 10.1038/S41588-019-0504-X.

Abramov, S. *et al.* (2021) 'Landscape of allele-specific transcription factor binding in the human genome', *Nature Communications 2021 12:1*, 12(1), pp. 1–15. doi: 10.1038/s41467-021-23007-0.

Ackermann, T. F. *et al.* (2011) 'EMD638683, a novel SGK inhibitor with antihypertensive potency', *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*, 28(1), pp. 137–146. doi: 10.1159/000331722.

Adrogué, H. J. and Madias, N. E. (2007) 'Sodium and Potassium in the Pathogenesis of Hypertension', *New England Journal of Medicine*, 356(19), pp. 1966–1978. doi: 10.1056/nejmra064486.

Aguet, F. *et al.* (2017) 'Genetic effects on gene expression across human tissues', *Nature 2017 550:7675*, 550(7675), pp. 204–213. doi: 10.1038/nature24277.

AK, S. *et al.* (2014) 'Single-cell RNA-seq reveals dynamic paracrine control of cellular variation', *Nature*, 510(7505), pp. 363–369. doi: 10.1038/NATURE13437.

Aleksandrow, D., Wysznacka, W. and Gajewski, J. (1959) 'Influence of Chlorothiazide upon Arterial Responsiveness to Nor-Epinephrine in Hypertensive Subjects', *New England Journal of Medicine*, 261(21), pp. 1052–1055. doi: 10.1056/nejm195911192612103.

Allison, S. J. (2018) 'ALMS1–NKCC2 interactions', *Nature Reviews Nephrology 2018 15:2*, 15(2), pp. 62–62. doi: 10.1038/s41581-018-0084-5.

Alves-Lopes, R. and Touyz, R. M. (2018) 'PARP-1 (Poly[ADP-Ribose] Polymerase-1) a novel target in aortic aneurysm', *Hypertension*, 72(5), pp. 1087–1089. doi:

10.1161/HYPERTENSIONAHA.118.11830.

Anders, S. and Huber, W. (2010) 'Differential expression analysis for sequence count data', *Genome Biology 2010 11:10*, 11(10), pp. 1–12. doi: 10.1186/GB-2010-11-10-R106.

Anstensrud, A. K. *et al.* (2020) 'Impact of genotype-predicted CYP2D6 metabolism on clinical effects and tolerability of metoprolol in patients after myocardial infarction – a prospective observational study', *European Journal of Clinical Pharmacology 2020 76:5*, 76(5), pp. 673–683. doi: 10.1007/S00228-020-02832-0.

Ayub, S. G. *et al.* (2011) 'Reduced nitrate level in individuals with hypertension and diabetes', *Journal of Cardiovascular Disease Research*, 2(3), pp. 172–176. doi: 10.4103/0975-3583.85264.

B, H. (1992) 'Tamm-Horsfall glycoprotein--inhibitor or promoter of calcium oxalate monohydrate crystallization processes?', *Urological research*, 20(1), pp. 83–86. doi: 10.1007/BF00294343.

Baaij, J. H. F. de *et al.* (2013) 'Elucidation of the distal convoluted tubule transcriptome identifies new candidate genes involved in renal Mg2+ handling', *https://doi.org/10.1152/ajprenal.00322.2013*, 305(11), pp. 1563–1573. doi: 10.1152/AJPRENAL.00322.2013.

Bachmann, S. *et al.* (2005) 'Renal effects of Tamm-Horsfall protein (uromodulin) deficiency in mice', *American journal of physiology. Renal physiology*, 288(3). doi: 10.1152/AJPRENAL.00143.2004.

Baens, M. *et al.* (2006) 'The Dark Side of EGFP: Defective Polyubiquitination', *PLOS ONE*, 1(1), p. e54. doi: 10.1371/JOURNAL.PONE.0000054.

Bailey, M. A. *et al.* (2011) 'Hsd11b2 Haploinsufficiency in Mice Causes Salt Sensitivity of Blood Pressure', *Hypertension*, 57(3), pp. 515–520. doi:

10.1161/HYPERTENSIONAHA.110.163782.

Banerjee, S. *et al.* (2018) 'Bayesian multiple logistic regression for case-control GWAS', *PLoS Genetics*, 14(12). doi: 10.1371/JOURNAL.PGEN.1007856.

Barbeira, A. N. *et al.* (2021) 'Exploiting the GTEx resources to decipher the mechanisms at GWAS loci', *Genome Biology 2021 22:1*, 22(1), pp. 1–24. doi: 10.1186/S13059-020-02252-4.

Basu, M. *et al.* (2017) 'Prediction and Subtyping of Hypertension from Pan-Tissue Transcriptomic and Genetic Analyses', *Genetics*, 207(3), p. 1121. doi: 10.1534/GENETICS.117.300280.

Basu, M. *et al.* (2021) 'Predicting tissue-specific gene expression from whole blood transcriptome', *Science Advances*, 7(14), p. eabd6991. doi: 10.1126/SCIADV.ABD6991.

Bates, J. M. *et al.* (2004) 'Tamm-Horsfall protein knockout mice are more prone to urinary tract infection Rapid Communication', *Kidney International*, 65(3), pp. 791–797. doi: 10.1111/J.1523-1755.2004.00452.X.

Benjamini, Y. (2010) 'Discovering the false discovery rate', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4), pp. 405–416. doi: 10.1111/J.1467-9868.2010.00746.X.

BERGLUND, G. *et al.* (1993) 'Design and feasibility', *Journal of Internal Medicine*, 233(1), pp. 45–51. doi: 10.1111/J.1365-2796.1993.TB00647.X.

Van Beusecum, J. P. *et al.* (2019) 'High Salt Activates CD11c + Antigen-Presenting Cells via SGK (Serum Glucocorticoid Kinase) 1 to Promote Renal Inflammation and Salt-Sensitive Hypertension', *Hypertension (Dallas, Tex. : 1979)*, 74(3), pp. 555–563. doi: 10.1161/HYPERTENSIONAHA.119.12761.

Bidani, A. K. and Griffin, K. A. (2004) 'Pathophysiology of hypertensive renal damage:

Implications for therapy', *Hypertension*. Lippincott Williams & Wilkins, pp. 595–601. doi: 10.1161/01.HYP.0000145180.38707.84.

Di Blasi, R. *et al*. (2021) 'A call for caution in analysing mammalian co-transfection experiments and implications of resource competition in data misinterpretation', *Nature Communications 2021 12:1*, 12(1), pp. 1–6. doi: 10.1038/s41467-021-22795-9.

Bleyer, A. J. *et al*. (2021) 'Autosomal Dominant Tubulointerstitial Kidney Disease – UMOD', *GeneReviews(®)*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK1356/ (Accessed: 23 August 2021).

Bode, A. M. and Dong, Z. (2018) 'Recent advances in precision oncology research', *npj Precision Oncology 2018 2:1*, 2(1), pp. 1–6. doi: 10.1038/s41698-018-0055-0.

Boder, P., Mary, S., Mark, P. B., *et al*. (2021) 'Mechanistic interactions of uromodulin with the thick ascending limb: perspectives in physiology and hypertension', *Journal of hypertension*, 39(8), pp. 1490–1504. doi: 10.1097/HJH.0000000000002861.

Boder, P., Mary, S., Graham, L., *et al*. (2021) 'THE ROLE OF CALCIUM IN UROMODULIN EXPRESSION AND SECRETION FROM RENAL MEDULLARY EPITHELIAL CELLS OF HYPERTENSIVE AND NORMOTENSIVE RATS', *Journal of Hypertension*, 39(Supplement 1), pp. e302–e303. doi: 10.1097/01.HJH.0000747960.25133.04.

De Boer, I. H. *et al*. (2017) 'Diabetes and hypertension: A position statement by the American diabetes association', *Diabetes Care*, 40(9), pp. 1273–1284. doi: 10.2337/dci17-0026.

Brooks, H. L. *et al*. (2001) 'Profiling of renal tubule Na+ transporter abundances in NHE3 and NCC null mice using targeted proteomics', *The Journal of Physiology*, 530(3), pp. 359–366. doi: 10.1111/J.1469-7793.2001.0359K.X.

C, P. *et al*. (2016) 'Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function', *Nature communications*, 7. doi:

10.1038/NCOMMS10023.

C, S. *et al*. (2009) ‘Analysis of uromodulin polymerization provides new insights into the mechanisms regulating ZP domain-mediated protein assembly’, *Molecular biology of the cell*, 20(2), pp. 589–599. doi: 10.1091/MBC.E08-08-0876.

C, T., A, D. and WR, C. (1985) ‘Human Tamm-Horsfall glycoprotein: urinary and plasma levels in normal subjects and patients with renal disease determined by a fully validated radioimmunoassay’, *Clinical science (London, England : 1979)*, 68(5), pp. 529–535. doi: 10.1042/CS0680529.

Caceres, P. S. and Ortiz, P. A. (2019) ‘Molecular regulation of NKCC2 in blood pressure control and hypertension’, *Current Opinion in Nephrology and Hypertension*, 28(5), pp. 474–480. doi: 10.1097/MNH.0000000000000531.

Cano-Gamez, E. and Trynka, G. (2020) ‘From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases’, *Frontiers in Genetics*, 11, p. 424. doi: 10.3389/FGENE.2020.00424/BIBTEX.

Cao, H. *et al*. (2019) ‘CeRNA network analysis and functional enrichment of salt sensitivity of blood pressure by weighted-gene co-expression analysis’, *PeerJ*, 2019(9). doi: 10.7717/PEERJ.7534/SUPP-3.

Cariappa, A. *et al*. (2009) ‘B cell antigen receptor signal strength and peripheral B cell development are regulated by a 9-O-acetyl sialic acid esterase’, *Journal of Experimental Medicine*, 206(1), pp. 125–138. doi: 10.1084/JEM.20081399.

Castrop, H. and Schießl, I. M. (2014) ‘Physiology and pathophysiology of the renal Na-K-2Cl cotransporter (NKCC2)’, *https://doi.org/10.1152/ajprenal.00432.2014*, 307(9), pp. F991–F1002. doi: 10.1152/AJPRENAL.00432.2014.

Chen, E. Y. *et al*. (2013) ‘Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool’, *BMC Bioinformatics 2013 14:1*, 14(1), pp. 1–14. doi:

10.1186/1471-2105-14-128.

Chen, H. *et al*. (2014) 'Rare genetic variant analysis on blood pressure in related samples', *BMC Proceedings*, 8(1), pp. 1–5. doi: 10.1186/1753-6561-8-S1-S35/TABLES/5.

Chiga, M. *et al*. (2008) 'Dietary salt regulates the phosphorylation of OSR1/SPAK kinases and the sodium chloride cotransporter through aldosterone', *Kidney International*, 74(11), pp. 1403–1409. doi: 10.1038/KI.2008.451/ATTACHMENT/353A5CC1-7E3D-4FDE-931A-5583C6D1B677/MMC1.PPT.

Chow, C. K. and Gupta, R. (2019) 'Blood pressure control: a challenge to global health systems', *The Lancet*. Lancet Publishing Group, pp. 613–615. doi: 10.1016/S0140-6736(19)31293-0.

Coady, A. *et al*. (2018) 'Tamm-Horsfall Protein Protects the Urinary Tract against Candida albicans', *Infection and Immunity*, 86(12). doi: 10.1128/IAI.00451-18.

Collins, L. *et al*. (2020) 'Regulation of Long Non-coding RNAs and MicroRNAs in Heart Disease: Insight Into Mechanisms and Therapeutic Approaches', *Frontiers in Physiology*, 11, p. 798. doi: 10.3389/FPHYS.2020.00798/BIBTEX.

Crespo, K. (2015) 'Modularisation épistatique des loci à trait quantitatif associés à la pression artérielle et identification de gènes candidats pour l'hypertension'. Available at: https://papyrus.bib.umontreal.ca/xmlui/handle/1866/13901 (Accessed: 16 December 2021).

CS, B. *et al*. (2015) 'The cost-effectiveness of returning incidental findings from next-generation genomic sequencing', *Genetics in medicine : official journal of the American College of Medical Genetics*, 17(7), pp. 587–595. doi: 10.1038/GIM.2014.156.

Cubrilo-Turek, M. (2003) ‘Hypertension and Coronary Heart Disease.’, *EJIFCC*, 14(2), pp. 67–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/30302078 (Accessed: 23 June 2021).

Curry, J. N. and Yu, A. S. L. (2018) ‘Magnesium Handling in the Kidney’, *Advances in Chronic Kidney Disease*, 25, pp. 236–243. doi: 10.1053/j.ackd.2018.01.003.

Cusanovich, D. A. *et al.* (2014) ‘The Functional Consequences of Variation in Transcription Factor Binding’, *PLOS Genetics*, 10(3), p. e1004226. doi: 10.1371/JOURNAL.PGEN.1004226.

D, C. *et al.* (2004) ‘Variation of high mannose chains of Tamm-Horsfall glycoprotein confers differential binding to type 1-fimbriated Escherichia coli’, *The Journal of biological chemistry*, 279(1), pp. 216–222. doi: 10.1074/JBC.M308821200.

Danaei, G. *et al.* (2009) ‘The preventable causes of death in the United States: Comparative risk assessment of dietary, lifestyle, and metabolic risk factors’, *PLoS Medicine*, 6(4). doi: 10.1371/journal.pmed.1000058.

DB, S. *et al.* (1996) ‘Genetic heterogeneity of Bartter’s syndrome revealed by mutations in the K+ channel, ROMK’, *Nature genetics*, 14(2), pp. 152–156. doi: 10.1038/NG1096-152.

Debaize, L. and Troadec, M. B. (2018) ‘The master regulator FUBP1: its emerging role in normal cell function and malignant development’, *Cellular and Molecular Life Sciences 2018 76:2*, 76(2), pp. 259–281. doi: 10.1007/S00018-018-2933-6.

DF, G. *et al.* (2010) ‘Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases’, *PLoS genetics*, 6(7), pp. 1–9. doi: 10.1371/JOURNAL.PGEN.1001039.

Dhingra, S. *et al.* (no date) ‘IL-10 attenuates TNF-a-induced NFkB pathway activation and cardiomyocyte apoptosis’. doi: 10.1093/cvr/cvp040.

Dinour, D. *et al.* (2014) 'Wild-type uromodulin prevents NFkB activation in kidney cells, while mutant uromodulin, causing FJHU nephropathy, does not', *Journal of nephrology*, 27(3), pp. 257–264. doi: 10.1007/S40620-014-0079-7.

Dornas, W. C. *et al.* (2017) 'Oxidative stress causes hypertension and activation of nuclear factor-κB after high-fructose and salt treatments', *Scientific Reports 2017 7:1*, 7(1), pp. 1–11. doi: 10.1038/srep46051.

Drazner, M. H. (2011) 'The progression of hypertensive heart disease', *Circulation*, 123(3), pp. 327–334. doi: 10.1161/CIRCULATIONAHA.108.845792.

Drüeke, T. B. and Devuyst, O. (2019) 'Blood pressure measurement in mice: tail-cuff or telemetry?', *Kidney International*, 96(1), p. 36. doi: 10.1016/J.KINT.2019.01.018.

Duarte, J. D. and Cooper-Dehoff, R. M. (2010) 'Mechanisms for blood pressure lowering and metabolic effects of thiazide and thiazide-like diuretics', *Expert Review of Cardiovascular Therapy*. NIH Public Access, pp. 793–802. doi: 10.1586/erc.10.27.

E, B. *et al.* (2006) 'Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin', *Molecular and cellular biology*, 26(7), pp. 2560–2569. doi: 10.1128/MCB.26.7.2560-2569.2006.

Eales, J. M. *et al.* (2021) 'Uncovering genetic mechanisms of hypertension through multi-omic analysis of the kidney', *Nature Genetics 2021 53:5*, 53(5), pp. 630–637. doi: 10.1038/s41588-021-00835-w.

Ehret, G. B. and Caulfield, M. J. (2013) 'Genes for blood pressure: an opportunity to understand hypertension', *European Heart Journal*, 34(13), p. 951. doi: 10.1093/EURHEARTJ/EHS455.

El-Achkar, T. M. *et al.* (2008) 'Tamm-Horsfall protein protects the kidney from ischemic injury by decreasing inflammation and altering TLR4 expression', *https://doi.org/10.1152/ajprenal.00083.2008*, 295(2), pp. 534–544. doi:

10.1152/AJPRENAL.00083.2008.

Ellison, D. H. (2019a) ‘Clinical pharmacology in diuretic use’, *Clinical Journal of the American Society of Nephrology*, 14(8), pp. 1248–1257. doi: 10.2215/CJN.09630818.

Ellison, D. H. (2019b) ‘Clinical Pharmacology in Diuretic Use’, *Clinical Journal of the American Society of Nephrology*, 14(8), pp. 1248–1257. doi: 10.2215/CJN.09630818.

Emanuel, E. J., Glickman, A. and Johnson, D. (2017) ‘Measuring the Burden of Health Care Costs on US Families: The Affordability Index’, *JAMA*, 318(19), pp. 1863–1864. doi: 10.1001/JAMA.2017.15686.

Engelen, K. *et al.* (2006) ‘A calibration method for estimating absolute expression levels from microarray data’, 22(10), pp. 1251–1258. doi: 10.1093/bioinformatics/btl068.

F, S.-C., N, M. and D, C. (2003) ‘Tamm-Horsfall glycoprotein: biology and clinical relevance’, *American journal of kidney diseases : the official journal of the National Kidney Foundation*, 42(4), pp. 658–676. doi: 10.1016/S0272-6386(03)00829-1.

Faulkner, E. *et al.* (2020) ‘Being Precise About Precision Medicine: What Should Value Frameworks Incorporate to Address Precision Medicine? A Report of the Personalized Precision Medicine Special Interest Group’, *Value in Health*, 23(5), pp. 529–539. doi: 10.1016/J.JVAL.2019.11.010.

Feng, M. *et al.* (2008) ‘Validation of volume-pressure recording tail-cuff blood pressure measurements’, *American Journal of Hypertension*, 21(12), pp. 1288–1291. doi: 10.1038/AJH.2008.301/2/AJH.1288.F1.JPEG.

Fenton, R. A. *et al.* (2017) ‘Renal tubular NHE3 is required in the maintenance of water and sodium chloride homeostasis’, *Kidney international*, 92(2), p. 397. doi: 10.1016/J.KINT.2017.02.001.

Ferrarlo, C. M. (2011) ‘ACE2: More of Ang-(1-7) or less Ang II?’, *Current Opinion in*

Nephrology and Hypertension, 20(1), pp. 1–6. doi: 10.1097/MNH.0b013e3283406f57.

Fontaine, D. A. and Davis, D. B. (2016) 'Attention to Background Strain Is Essential for Metabolic Research: C57BL/6 and the International Knockout Mouse Consortium', *Diabetes*, 65(1), p. 25. doi: 10.2337/DB15-0982.

Forouzanfar, M. H. *et al*. (2017) 'Global burden of hypertension and systolic blood pressure of at least 110 to 115mmHg, 1990-2015', *JAMA - Journal of the American Medical Association*, 317(2), pp. 165–182. doi: 10.1001/jama.2016.19043.

Fountain, J. H. and Lappin, S. L. (2018) *Physiology, Renin Angiotensin System*, *StatPearls*. StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29261862 (Accessed: 23 June 2021).

Free, R. B., Hazelwood, L. A. and Sibley, D. R. (2009) 'Identifying Novel Protein-Protein Interactions Using Co-Immunoprecipitation and Mass Spectroscopy', *Current protocols in neuroscience / editorial board, Jacqueline N. Crawley ... [et al.]*, 0 5(SUPPL. 46), p. Unit. doi: 10.1002/0471142301.NS0528S46.

FREIS, E. D. *et al*. (1960) 'Mechanism of the altered blood pressure responsiveness produced by chlorothiazide.', *The Journal of clinical investigation*, 39(8), pp. 1277–1281. doi: 10.1172/JCI104143.

G, G. and PA, F. (2009) 'Thick ascending limb: the Na(+):K (+):2Cl (-) co-transporter, NKCC2, and the calcium-sensing receptor, CaSR', *Pflugers Archiv : European journal of physiology*, 458(1), pp. 61–76. doi: 10.1007/S00424-008-0607-1.

GA, Q. (1989) 'Control of magnesium transport in the thick ascending limb', *The American journal of physiology*, 256(2 Pt 2). doi: 10.1152/AJPRENAL.1989.256.2.F197.

Gadegbeku, C. A. *et al*. (2013) 'Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multi-disciplinary

approach', *Kidney international*, 83(4), p. 749. doi: 10.1038/KI.2012.428.

Gast, C. *et al.* (2018) 'Autosomal dominant tubulointerstitial kidney disease-UMOD is the most frequent non polycystic genetic kidney disease', *BMC Nephrology*, 19(1). doi: 10.1186/S12882-018-1107-Y.

Gaudet, P. *et al.* (2011) 'Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium', *Briefings in bioinformatics*, 12(5), pp. 449–462. doi: 10.1093/BIB/BBR042.

GR, A., PS, C. and PA, O. (2011) 'Molecular regulation of NKCC2 in the thick ascending limb', *American journal of physiology. Renal physiology*, 301(6). doi: 10.1152/AJPRENAL.00396.2011.

Graham, F. L. *et al.* (1977) 'Characteristics of a human cell line transformed by DNA from human adenovirus type 5', *Journal of General Virology*, 36(1), pp. 59–72. doi: 10.1099/0022-1317-36-1-59/CITE/REFWORKS.

Graham, L. A., Padmanabhan, S., Fraser, N. J., Kumar, S., Bates, J. M., Raffi, H. S., Welsh, P., Beattie, W., Hao, S., Leh, S., Hultstrom, M., Ferreri, N. R., Dominiczak, A. F., Graham, D. and Mcbride, M. W. (2014a) 'Validation of uromodulin as a candidate gene for human essential hypertension', *Hypertension (Dallas, Tex. : 1979)*, 63(3), pp. 551–558. doi: 10.1161/HYPERTENSIONAHA.113.01423.

Graham, L. A., Padmanabhan, S., Fraser, N. J., Kumar, S., Bates, J. M., Raffi, H. S., Welsh, P., Beattie, W., Hao, S., Leh, S., Hultstrom, M., Ferreri, N. R., Dominiczak, A. F., Graham, D. and Mcbride, M. W. (2014b) 'Validation of uromodulin as a candidate gene for human essential hypertension', *Hypertension*, 63(3), pp. 551–558. doi: 10.1161/HYPERTENSIONAHA.113.01423.

Graham, L. A., Padmanabhan, S., Fraser, N. J., Kumar, S., Bates, J. M., Raffi, H. S., Welsh, P., Beattie, W., Hao, S., Leh, S., Hultstrom, M., Ferreri, N. R., Dominiczak, A. F., Graham, D. and McBride, M. W. (2014) 'Validation of Uromodulin as a Candidate

Gene for Human Essential Hypertension', *Hypertension*, 63(3), pp. 551–558. doi: 10.1161/HYPERTENSIONAHA.113.01423.

Gresh, L. *et al*. (2004) 'A transcriptional network in polycystic kidney disease', *The EMBO journal*, 23(7), pp. 1657–1668. doi: 10.1038/SJ.EMBOJ.7600160.

Griffn, K. A. (2017) 'Hypertensive kidney injury and the progression of chronic kidney disease', *Hypertension*, 70(4), pp. 687–694. doi: 10.1161/HYPERTENSIONAHA.117.08314.

De Groot, T., Bindels, R. J. M. and Hoenderop, J. G. J. (2008) 'TRPV5: an ingeniously controlled calcium channel', *Kidney International*, 74(10), pp. 1241–1246. doi: 10.1038/KI.2008.320.

H, D. *et al*. (2013) 'Activation of the Ca(2+)-sensing receptor increases renal claudin-14 expression and urinary Ca(2+) excretion', *American journal of physiology. Renal physiology*, 304(6). doi: 10.1152/AJPRENAL.00263.2012.

Haider, D. G. *et al*. (2015) 'Use of Calcium Channel Blockers is Associated with Mortality in Patients with Chronic Kidney Disease', *Kidney and Blood Pressure Research*, 40(6), pp. 630–637. doi: 10.1159/000368539.

Hammami, I., Lacey, B. and Lewington, S. (2018) '3028The burden of hypertension and associated risk for cardiovascular mortality in the UK biobank', *European Heart Journal*, 39(suppl_1). doi: 10.1093/eurheartj/ehy563.3028.

Han, J. *et al*. (2013) 'Common genetic variants of the human UMOD gene are functional on transcription and predict plasma uric acid in two distinct populations', *Kidney international*, 83(4), p. 733. doi: 10.1038/KI.2012.449.

Handl, J. *et al*. (no date) 'The Effect of Repeated Passaging on the Susceptibility of Human Proximal Tubular HK-2 Cells to Toxic Compounds'. doi: 10.33549/physiolres.934491.

Hao, S., Hao, M. and Ferreri, N. R. (2018) 'Renal-specific silencing of TNF (Tumor Necrosis Factor) unmasks salt-dependent increases in blood pressure via an NKCC2A (Na+-K+-2Cl- Cotransporter Isoform A)-dependent mechanism', *Hypertension*, 71(6), pp. 1117–1125. doi: 10.1161/HYPERTENSIONAHA.117.10764.

Hayden, M. S. and Ghosh, S. (2014) 'Regulation of NF-κB by TNF Family Cytokines', *Seminars in immunology*, 26(3), p. 253. doi: 10.1016/J.SMIM.2014.05.004.

Health Survey for England 2017 Cardiovascular diseases (2018). Available at: www.statisticsauthority.gov.uk/assessment/code-of-practice (Accessed: 23 June 2021).

Hebert, S. C. *et al.* (1987) 'The medullary thick limb: Function and modulation of the single-effect multiplier', *Kidney International*, 31, pp. 580–589. doi: 10.1038/ki.1987.38.

Hebert, S. C. and Gamba, G. (1994) 'Molecular cloning and characterization of the renal diuretic-sensitive electroneutral sodium-(potassium)-chloride cotransporters', *The Clinical Investigator*, 72(9), pp. 692–694. doi: 10.1007/BF00212991.

Hedner, T. (2009) 'The Nordic Diltiazem Study (NORDIL). A Prospective Intervention Trial of of Calcium Antagonist Therapy in Hypertension', *http://dx.doi.org/10.3109/08037059309077174*, 2(4), pp. 312–321. doi: 10.3109/08037059309077174.

Heitmeier, M. *et al.* (2014) 'The role of tumor necrosis factor alpha in regulating the expression of Tamm-Horsfall Protein (uromodulin) in thick ascending limbs during kidney injury', *American journal of nephrology*, 40(5), pp. 458–467. doi: 10.1159/000369836.

Ho, D. S. W. *et al.* (2019) 'Machine Learning SNP Based Prediction for Precision Medicine', *Frontiers in Genetics*, 0(MAR), p. 267. doi: 10.3389/FGENE.2019.00267.

Holst, A. G., Jensen, G. and Prescott, E. (2010) 'Risk factors for venous thromboembolism: Results from the copenhagen city heart study', *Circulation*, 121(17), pp. 1896–1903. doi: 10.1161/CIRCULATIONAHA.109.921460.

Hsu, J. B.-K. *et al.* (2011) 'miRTar: an integrated system for identifying miRNA-target interactions in human', *BMC Bioinformatics 2011 12:1*, 12(1), pp. 1–12. doi: 10.1186/1471-2105-12-300.

Hu, C. *et al.* (2022) 'Live-dead assay on unlabeled cells using phase imaging with computational specificity', *Nature Communications 2022 13:1*, 13(1), pp. 1–8. doi: 10.1038/s41467-022-28214-x.

Huan, T. *et al.* (2015) 'Integrative network analysis reveals molecular mechanisms of blood pressure regulation', *Molecular Systems Biology*, 11(4), p. 799. doi: 10.15252/MSB.20145399.

Huang, L., Li, J. and Jiang, Y. (2016) 'Association between hypertension and deep vein thrombosis after orthopedic surgery: A meta-analysis', *European Journal of Medical Research*, 21(1), p. 13. doi: 10.1186/s40001-016-0207-z.

Hwang, B., Lee, J. H. and Bang, D. (2018) 'Single-cell RNA sequencing technologies and bioinformatics pipelines', *Experimental & Molecular Medicine 2018 50:8*, 50(8), pp. 1–14. doi: 10.1038/s12276-018-0071-8.

Hypertension and coronary artery disease: cause and effect : Journal of Hypertension (no date). Available at: https://journals.lww.com/jhypertension/abstract/1995/08001/hypertension_and_coronary_artery_disease__cause.8.aspx (Accessed: 23 June 2021).

Igarashi, P. *et al.* (2005) 'Roles of HNF-1beta in kidney development and congenital cystic diseases', *Kidney international*, 68(5), pp. 1944–1947. doi: 10.1111/J.1523-1755.2005.00625.X.

Immler, R. *et al.* (2020a) 'Extratubular Polymerized Uromodulin Induces Leukocyte Recruitment and Inflammation In Vivo', *Frontiers in Immunology*, 0, p. 3305. doi: 10.3389/FIMMU.2020.588245.

Immler, R. *et al.* (2020b) 'Extratubular Polymerized Uromodulin Induces Leukocyte Recruitment and Inflammation In Vivo', *Frontiers in Immunology*, 11, p. 3305. doi: 10.3389/FIMMU.2020.588245/BIBTEX.

IY Gong, U. S. N. C. G. D. (2011) 'Clinical and genetic determinants of warfarin pharmacokinetics and pharmacodynamics during treatment initiation', *PLoS One*, 6(11), p. e27808. doi: 10.1371/journal.pone.0027808.

J, L. *et al.* (2001) 'Distribution of transcellular calcium and sodium transport pathways along mouse distal nephron', *American journal of physiology. Renal physiology*, 281(6). doi: 10.1152/AJPRENAL.0085.2001.

J, P. *et al.* (2001) 'Tamm-Horsfall protein binds to type 1 fimbriated Escherichia coli and prevents E. coli from binding to uroplakin Ia and Ib receptors', *The Journal of biological chemistry*, 276(13), pp. 9924–9930. doi: 10.1074/JBC.M008610200.

JA, J. *et al.* (2011) 'Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing', *Clinical pharmacology and therapeutics*, 90(4), pp. 625–629. doi: 10.1038/CLPT.2011.185.

Jacobsen, L. B., Calvin, S. A. and Lobenhofer, E. K. (2009) 'Transcriptional effects of transfection: The potential for misinterpretation of gene expression data generated from transiently transfected cells', *BioTechniques*, 47(1), pp. 617–624. doi: 10.2144/000113132/ASSET/IMAGES/LARGE/TABLE5.JPEG.

Jenkinson, S. E. *et al.* (2012) 'The limitations of renal epithelial cell line HK-2 as a model of drug transporter expression and function in the proximal tubule', *Pflügers Archiv - European Journal of Physiology 2012 464:6*, 464(6), pp. 601–611. doi: 10.1007/S00424-012-1163-2.

Jessup, J. A. *et al.* (2008) 'Differential effect of low-dose thiazides on the renin angiotensin system in genetically hypertensive and normotensive rats', *Journal of the American Society of Hypertension*, 2(2), pp. 106–115. doi: 10.1016/j.jash.2007.10.005.

Jian, L. *et al.* (2015) 'Functional analysis of UMOD gene and its effect on inflammatory cytokines in serum of essential hypertension patients', *International Journal of Clinical and Experimental Pathology*, 8(9), p. 11356. Available at: /pmc/articles/PMC4637676/ (Accessed: 16 August 2021).

Jiang, X. *et al.* (2019) 'Current status and future prospects of PARP inhibitor clinical trials in ovarian cancer', *Cancer Management and Research*, 11, p. 4371. doi: 10.2147/CMAR.S200524.

Jiang, X. and Ning, Q. (2020) 'Long noncoding RNAs as novel players in the pathogenesis of hypertension', *Hypertension Research 2020 43:7*, 43(7), pp. 597–608. doi: 10.1038/s41440-020-0408-2.

JK, H. *et al.* (1990) 'Activation of the inflammatory response of neutrophils by Tamm-Horsfall glycoprotein', *Kidney international*, 37(2), pp. 717–726. doi: 10.1038/KI.1990.38.

JN, H. (2009) 'Genomewide association studies--illuminating biologic pathways', *The New England journal of medicine*, 360(17), pp. 1699–1701. doi: 10.1056/NEJMP0808934.

Joffres, M. *et al.* (2013) 'Hypertension prevalence, awareness, treatment and control in national surveys from England, the USA and Canada, and correlation with stroke and ischaemic heart disease mortality: A cross-sectional study', *BMJ Open*, 3(8), p. e003423. doi: 10.1136/bmjopen-2013-003423.

Johnson, B. G. *et al.* (2017a) 'Uromodulin p.Cys147Trp mutation drives kidney disease by activating ER stress and apoptosis', *The Journal of Clinical Investigation*, 127(11),

pp. 3954–3969. doi: 10.1172/JCI93817.

Johnson, B. G. *et al.* (2017b) 'Uromodulin p.Cys147Trp mutation drives kidney disease by activating ER stress and apoptosis', *Journal of Clinical Investigation*, 127(11), pp. 3954–3969. doi: 10.1172/JCI93817.

Joiret, M. *et al.* (2019) 'Confounding of linkage disequilibrium patterns in large scale DNA based gene-gene interaction studies', *BioData Mining 2019 12:1*, 12(1), pp. 1–23. doi: 10.1186/S13040-019-0199-7.

Jorgensen, A. L. *et al.* (2019) 'Implementation of genotype-guided dosing of warfarin with point-of-care genetic testing in three UK clinics: a matched cohort study', *BMC Medicine 2019 17:1*, 17(1), pp. 1–11. doi: 10.1186/S12916-019-1308-7.

Joyner, M. J. and Paneth, N. (2019) 'Promises, promises, and precision medicine', *The Journal of Clinical Investigation*, 129(3), pp. 946–948. doi: 10.1172/JCI126119.

Jung, J., Basile, D. P. and Pratt, J. H. (2011) 'Sodium reabsorption in the thick ascending limb in relation to blood pressure: A clinical perspective', *Hypertension*. Lippincott Williams & WilkinsHagerstown, MD, pp. 873–879. doi: 10.1161/HYPERTENSIONAHA.108.120246.

Kakoki, M. *et al.* (2013) 'Primary aldosteronism and impaired natriuresis in mice underexpressing TGFβ1', *Proceedings of the National Academy of Sciences of the United States of America*, 110(14), pp. 5600–5605. doi: 10.1073/PNAS.1302641110/-/DCSUPPLEMENTAL/PNAS.201302641SI.PDF.

Kamel, S. M. *et al.* (2021) 'A Heterozygous Mutation in Cardiac Troponin T Promotes Ca2+ Dysregulation and Adult Cardiomyopathy in Zebrafish', *Journal of Cardiovascular Development and Disease 2021, Vol. 8, Page 46*, 8(4), p. 46. doi: 10.3390/JCDD8040046.

Kasztura, M. *et al.* (2019) 'Cost-effectiveness of precision medicine: a scoping

review', *International Journal of Public Health 2019 64:9*, 64(9), pp. 1261–1271. doi: 10.1007/S00038-019-01298-X.

Kaufman, W. L. *et al.* (2008) 'Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation', *Nucleic Acids Research*, 36(17), p. e111. doi: 10.1093/NAR/GKN508.

Keil, U. (2005) 'The worldwide WHO MONICA Project: Results and perspectives', *Gesundheitswesen*, 67(SUPPL. 1). doi: 10.1055/S-2005-858240.

Kim, H.-T., Song, I.-Y. and Piedrahita, J. (2003) 'Kidney-Specific Activity of the Bovine Uromodulin Promoter', *Transgenic Research 2003 12:2*, 12(2), pp. 191–201. doi: 10.1023/A:1022911124946.

Kim, T. K. and Eberwine, J. H. (2010) 'Mammalian cell transfection: the present and the future', *Analytical and Bioanalytical Chemistry*, 397(8), p. 3173. doi: 10.1007/S00216-010-3821-6.

Kobayashi, M. and Uesugi, S. (1995) 'The role of hypertension as a risk factor of atherosclerosis', *Rinsho byori. The Japanese journal of clinical pathology*, pp. 104–110.

Koller, B. H. and Smithies, O. (1992) 'Altering genes in animals by gene targeting', *Annual review of immunology*, 10, pp. 705–730. doi: 10.1146/ANNUREV.IY.10.040192.003421.

Kolovos, P. *et al.* (2012) 'Enhancers and silencers: an integrated and simple model for their function', *Epigenetics & Chromatin 2012 5:1*, 5(1), pp. 1–8. doi: 10.1186/1756-8935-5-1.

Kosugi-Tanaka, C. *et al.* (2006) 'Protein kinase A-regulated membrane trafficking of a green fluorescent protein-aquaporin 5 chimera in MDCK cells', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1763(4), pp. 337–344. doi:

10.1016/J.BBAMCR.2006.02.005.

Kotelevtsev, Y. *et al*. (1999) 'Hypertension in mice lacking 11ß-hydroxysteroid dehydrogenase type 2', *The Journal of Clinical Investigation*, 103(5), pp. 683–689. doi: 10.1172/JCI4445.

Kramers, B. J. *et al*. (2020) 'Thiazide diuretics and the rate of disease progression in autosomal dominant polycystic kidney disease: an observational study', *Nephrology Dialysis Transplantation*. doi: 10.1093/ndt/gfaa150.

Krzyszczyk, P. *et al*. (2018) 'The growing role of precision and personalized medicine for cancer treatment', *Technology*, 6(3–4), p. 79. doi: 10.1142/S2339547818300020.

Kumar, S. *et al*. (2016) 'Generation and Standardized, Systemic Phenotypic Analysis of Pou3f3L423P Mutant Mice', *PLOS ONE*, 11(3), p. e0150472. doi: 10.1371/JOURNAL.PONE.0150472.

Kuruppu, S., Tochon-Danguy, N. and Smith, A. I. (2013) 'Applicability of green fluorescence protein in the study of endothelin converting enzyme-1c trafficking', *Protein Science : A Publication of the Protein Society*, 22(3), p. 306. doi: 10.1002/PRO.2212.

L, G. *et al*. (2004) 'A transcriptional network in polycystic kidney disease', *The EMBO journal*, 23(7), pp. 1657–1668. doi: 10.1038/SJ.EMBOJ.7600160.

L, L. and H, C. (2010) 'Coexistence of quiescent and active adult stem cells in mammals', *Science (New York, N.Y.)*, 327(5965), pp. 542–545. doi: 10.1126/SCIENCE.1180794.

L, M. *et al*. (2004) 'Tamm-Horsfall protein is a critical renal defense factor protecting against calcium oxalate crystal formation', *Kidney international*, 66(3), pp. 1159–1166. doi: 10.1111/J.1523-1755.2004.00867.X.

L, R. *et al*. (2003) 'Allelism of MCKD, FJHN and GCKD caused by impairment of

uromodulin export dynamics', *Human molecular genetics*, 12(24), pp. 3369–3384. doi: 10.1093/HMG/DDG353.

LA, G. *et al*. (2014) 'Validation of uromodulin as a candidate gene for human essential hypertension', *Hypertension (Dallas, Tex. : 1979)*, 63(3), pp. 551–558. doi: 10.1161/HYPERTENSIONAHA.113.01423.

Lang, F. and Shumilina, E. (2013) 'Regulation of ion channels by the serum- and glucocorticoid-inducible kinase SGK1', *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 27(1), pp. 3–12. doi: 10.1096/FJ.12-218230.

Lang, F., Yang Huang, D. Y. and Vallon, V. (2010) 'SGK, renal function and Hypertension', *Journal of nephrology*, 23(0 16), p. S124. Available at: /pmc/articles/PMC4026186/ (Accessed: 6 December 2021).

Lang, J. A. *et al*. (1995) 'Endogenous Human Renin Expression and Promoter Activity in CALU-6, a Pulmonary Carcinoma Cell Line', *Hypertension*, 25(4 II), pp. 704–710. doi: 10.1161/01.HYP.25.4.704.

Lau, W. H. *et al*. (2008) 'Qualification and application of an ELISA for the determination of Tamm Horsfall Protein (THP) in human urine and its use for screening of Kidney Stone Disease', *International Journal of Biological Sciences*, 4(4), pp. 215–222. doi: 10.7150/IJBS.4.215.

Lazich, I. and Bakris, G. L. (2011) 'Endothelin antagonism in patients with resistant hypertension and hypertension nephropathy', in *Endothelin in Renal Physiology and Disease*. S. Karger AG, pp. 223–234. doi: 10.1159/000328988.

LB, R. *et al*. (2014) 'The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update', *Clinical pharmacology and therapeutics*, 96(4), pp. 423–428. doi: 10.1038/CLPT.2014.125.

Lemmer, B. (2016) 'Superiority of radiotelemetry over the tail-cuff method in evaluating control and drug-induced data in cardiovascular functions', *http://dx.doi.org/10.1080/09291016.2016.1209854*, 47(6), pp. 919–925. doi: 10.1080/09291016.2016.1209854.

Leopold, K., Stirpe, A. and Schalch, T. (2019) 'Transcriptional gene silencing requires dedicated interaction between HP1 protein Chp2 and chromatin remodeler Mit1', *Genes & Development*, 33(9–10), pp. 565–577. doi: 10.1101/GAD.320440.118.

Lewington, S. *et al.* (2002) 'Age-specific relevance of usual blood pressure to vascular mortality: A meta-analysis of individual data for one million adults in 61 prospective studies', *Lancet*, 360(9349), pp. 1903–1913. doi: 10.1016/S0140-6736(02)11911-8.

Li, Y. *et al.* (2019) 'Extreme sampling design in genetic association mapping of quantitative trait loci using balanced and unbalanced case-control samples', *Scientific Reports 2019 9:1*, 9(1), pp. 1–9. doi: 10.1038/s41598-019-51790-w.

Li, Y. R. *et al.* (2015) 'Concept and design of a genome-wide association genotyping array tailored for transplantation-specific studies', *Genome Medicine 2015 7:1*, 7(1), pp. 1–13. doi: 10.1186/S13073-015-0211-X.

Liang, E. S. *et al.* (2018a) 'PARP-1 (Poly[ADP-Ribose] Polymerase 1) Inhibition protects from ang II (angiotensin II)-induced abdominal aortic aneurysm in mice', *Hypertension*, 72(5), pp. 1189–1199. doi: 10.1161/HYPERTENSIONAHA.118.11184.

Liang, E. S. *et al.* (2018b) 'PARP-1 (Poly[ADP-Ribose] Polymerase 1) Inhibition Protects From Ang II (Angiotensin II)-Induced Abdominal Aortic Aneurysm in Mice', *Hypertension (Dallas, Tex. : 1979)*, 72(5), pp. 1189–1199. doi: 10.1161/HYPERTENSIONAHA.118.11184.

Lonsdale, J. *et al.* (2013) 'The Genotype-Tissue Expression (GTEx) project', *Nature Genetics 2013 45:6*, 45(6), pp. 580–585. doi: 10.1038/ng.2653.

Loos, R. J. F. (2020) '15 years of genome-wide association studies and no signs of slowing down', *Nature Communications 2020 11:1*, 11(1), pp. 1–3. doi: 10.1038/s41467-020-19653-5.

Louis, N., Evelegh, C. and Graham, F. L. (1997) 'Cloning and Sequencing of the Cellular–Viral Junctions from the Human Adenovirus Type 5 Transformed 293 Cell Line', *Virology*, 233(2), pp. 423–429. doi: 10.1006/VIRO.1997.8597.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology 2014 15:12*, 15(12), pp. 1–21. doi: 10.1186/S13059-014-0550-8.

Lubit, B. W. and Schwartz, J. H. (no date) 'An Antiactin Antibody that Distinguishes between Cytoplasmic and Skeletal Muscle Actins'.

Luo, H. *et al.* (2015) 'Chronic NF-κB blockade improves renal angiotensin II type 1 receptor functions and reduces blood pressure in Zucker diabetic rats', *Cardiovascular Diabetology*, 14(1), pp. 1–11. doi: 10.1186/S12933-015-0239-7/FIGURES/7.

M, Bokhove *et al.* (2016) 'A structured interdomain linker directs self-polymerization of human uromodulin', *Proceedings of the National Academy of Sciences of the United States of America*, 113(6), pp. 1552–1557. doi: 10.1073/PNAS.1519803113.

M, N. *et al.* (2018) 'Uromodulin regulates renal magnesium homeostasis through the ion channel transient receptor potential melastatin 6 (TRPM6)', *The Journal of biological chemistry*, 293(42), pp. 16488–16502. doi: 10.1074/JBC.RA118.003950.

M, P. *et al.* (2011) 'Association of beta-adrenergic receptor polymorphisms and mortality in carvedilol-treated chronic heart-failure patients', *British journal of clinical pharmacology*, 71(4), pp. 556–565. doi: 10.1111/J.1365-2125.2010.03868.X.

M, S. *et al.* (2010) 'Uromodulin facilitates neutrophil migration across renal epithelial

monolayers', *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*, 26(3), pp. 311–318. doi: 10.1159/000320554.

Maccarrone, G. *et al.* (2017) 'Characterization of a Protein Interactome by Co-Immunoprecipitation and Shotgun Mass Spectrometry', *Methods in molecular biology (Clifton, N.J.)*, 1546, pp. 223–234. doi: 10.1007/978-1-4939-6730-8_19.

MacNeil, L. T. and Walhout, A. J. M. (2011) 'Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression', *Genome Research*, 21(5), pp. 645–657. doi: 10.1101/GR.097378.109.

Majid, D., Prieto, M. and Navar, L. (2015) 'Salt-Sensitive Hypertension: Perspectives on Intrarenal Mechanisms', *Current Hypertension Reviews*, 11(1), pp. 38–48. doi: 10.2174/1573402111666150530203858.

Manolio, T. A. *et al.* (2009) 'Finding the missing heritability of complex diseases', *Nature*, 461(7265), p. 747. doi: 10.1038/NATURE08494.

Markadieu, N. and Delpire, E. (2014) 'Physiology and Pathophysiology of SLC12A1/2 transporters', *Pflugers Archiv : European journal of physiology*, 466(1), pp. 91–105. doi: 10.1007/S00424-013-1370-5.

Martí, J. M. *et al.* (2021) 'Selective modulation by PARP-1 of HIF-1α-recruitment to chromatin during hypoxia is required for tumor adaptation to hypoxic conditions', *Redox Biology*, 41, p. 101885. doi: 10.1016/J.REDOX.2021.101885.

Martin, E. R. *et al.* (2000) 'SNPing Away at Complex Diseases: Analysis of Single-Nucleotide Polymorphisms around APOE in Alzheimer Disease', *American Journal of Human Genetics*, 67(2), p. 383. doi: 10.1086/303003.

Matsuki, K. *et al.* (2014) 'The Role of Transforming Growth Factor ß1 in the Regulation of Blood Pressure', *Current hypertension reviews*, 10(4), p. 223. doi:

10.2174/1573402110041503191233313.

Mattson, D. L. and Liang, M. (2017) 'From GWAS to functional genomics-based precision medicine', *Nature Reviews Nephrology 2017 13:4*, 13(4), pp. 195–196. doi: 10.1038/nrneph.2017.21.

McCallum, L. *et al.* (2021) 'Rationale and Design of the Genotype-Blinded Trial of Torasemide for the Treatment of Hypertension (BHF UMOD)', *American Journal of Hypertension*, 34(1), pp. 92–99. doi: 10.1093/AJH/HPAA166.

Mehaffey, E. and Majid, D. S. A. (2017) 'Tumor necrosis factor-α, kidney function, and hypertension', *American Journal of Physiology - Renal Physiology*. American Physiological Society, pp. F1005–F1008. doi: 10.1152/ajprenal.00535.2016.

Micanovic, R. *et al.* (2015) 'Tamm-Horsfall Protein Regulates Granulopoiesis and Systemic Neutrophil Homeostasis', *Journal of the American Society of Nephrology : JASN*, 26(9), p. 2172. doi: 10.1681/ASN.2014070664.

Michael, S. K. *et al.* (2008) 'High blood pressure arising from a defect in vascular function', *Proceedings of the National Academy of Sciences*, 105(18), pp. 6702–6707. doi: 10.1073/PNAS.0802128105.

Mills, K. T., Stefanescu, A. and He, J. (2020) 'The global epidemiology of hypertension', *Nature Reviews Nephrology*. Nature Research, pp. 223–237. doi: 10.1038/s41581-019-0244-2.

Mitchelmore, J. *et al.* (2020) 'Functional effects of variation in transcription factor binding highlight long-range gene regulation by epromoters', *Nucleic Acids Research*, 48(6), pp. 2866–2879. doi: 10.1093/NAR/GKAA123.

Mount, D. B. (2014) 'Thick Ascending Limb of the Loop of Henle', *Clinical Journal of the American Society of Nephrology*, 9(11), pp. 1974–1986. doi: 10.2215/CJN.04480413.

MTF, W., J, Z. and M, N. (2019) ‘Uromodulin in mineral metabolism’, *Current opinion in nephrology and hypertension*, 28(5), pp. 481–489. doi: 10.1097/MNH.0000000000000522.

Müller, M. *et al.* (2013) ‘Distinct Functional Interactions between Actin Isoforms and Nonsarcomeric Myosins’, *PLOS ONE*, 8(7), p. e70636. doi: 10.1371/JOURNAL.PONE.0070636.

Muntner, P. *et al.* (2019) ‘Measurement of Blood Pressure in Humans: A Scientific Statement From the American Heart Association’, *Hypertension*, 73(5), pp. E35–E66. doi: 10.1161/HYP.0000000000000087.

Murakami, K. (2015) ‘Non-coding RNAs and Hypertension–Unveiling Unexpected Mechanisms of Hypertension by the Dark Matter of the Genome’, *Current Hypertension Reviews*, 11(2), p. 80. doi: 10.2174/1573402111666150401105317.

Murthy, M., Kurz, T. and O’Shaughnessy, K. M. (2017) ‘WNK signalling pathways in blood pressure regulation’, *Cellular and Molecular Life Sciences*, 74(7), p. 1261. doi: 10.1007/S00018-016-2402-Z.

Mutig, K. (2017) ‘Trafficking and regulation of the NKCC2 cotransporter in the thick ascending limb’, *Current opinion in nephrology and hypertension*, 26(5), pp. 392–397. doi: 10.1097/MNH.0000000000000351.

N, H. *et al.* (2008) ‘Salt and acid-base metabolism in claudin-16 knockdown mice: impact for the pathophysiology of FHHNC patients’, *American journal of physiology. Renal physiology*, 295(6). doi: 10.1152/AJPRENAL.90388.2008.

Nakai, S. *et al.* (2003) ‘Crucial roles of Brn1 in distal tubule formation and function in mouse kidney’, *Development (Cambridge, England)*, 130(19), pp. 4751–4759. doi: 10.1242/DEV.00666.

Narayan, H. and Webb, D. J. (2016) ‘New Evidence Supporting the Use of

Mineralocorticoid Receptor Blockers in Drug-Resistant Hypertension', *Current Hypertension Reports*. Current Medicine Group LLC 1. doi: 10.1007/s11906-016-0643-8.

NR, van Z. *et al*. (2018) 'A Genome-Wide Association Study of Diabetic Kidney Disease in Subjects With Type 2 Diabetes', *Diabetes*, 67(7), pp. 1414–1427. doi: 10.2337/DB17-0914.

Oh, G. C. and Cho, H.-J. (2020) 'Blood pressure and heart failure', *Clinical Hypertension*, 26(1), pp. 1–8. doi: 10.1186/s40885-019-0132-x.

Oh, S. W. and Han, S. Y. (2015) 'Loop diuretics in clinical practice', *Electrolyte and Blood Pressure*. Korean Society of Electrolyte and Blood Pressure Research, pp. 17–21. doi: 10.5049/EBP.2015.13.1.17.

Olden, M. *et al*. (2014a) 'Common variants in UMOD associate with urinary uromodulin levels: A meta-analysis', *Journal of the American Society of Nephrology*, 25(8), pp. 1869–1882. doi: 10.1681/ASN.2013070781.

Olden, M. *et al*. (2014b) 'Common variants in UMOD associate with urinary uromodulin levels: A meta-analysis', *Journal of the American Society of Nephrology*, 25(8), pp. 1869–1882. doi: 10.1681/ASN.2013070781/-/DCSUPPLEMENTAL.

Olinger, E. *et al*. (2019a) 'Hepsin-mediated Processing of Uromodulin is Crucial for Salt-sensitivity and Thick Ascending Limb Homeostasis', *Scientific Reports 2019 9:1*, 9(1), pp. 1–15. doi: 10.1038/s41598-019-48300-3.

Olinger, E. *et al*. (2019b) 'Hepsin-mediated Processing of Uromodulin is Crucial for Salt-sensitivity and Thick Ascending Limb Homeostasis', *Scientific Reports*, 9(1). doi: 10.1038/S41598-019-48300-3.

Oparil, S. *et al*. (2018) 'Hypertension', *Nature Reviews Disease Primers*. Nature Publishing Group, p. 18014. doi: 10.1038/nrdp.2018.14.

P, V. *et al.* (2006) 'Alterations of uromodulin biology: a common denominator of the genetically heterogeneous FJHN/MCKD syndrome', *Kidney international*, 70(6), pp. 1155–1169. doi: 10.1038/SJ.KI.5001728.

Padmanabhan, S., Melander, O., Johnson, T., Blasio, A. M. Di, *et al.* (2010) 'Genome-Wide Association Study of Blood Pressure Extremes Identifies Variant near UMOD Associated with Hypertension', *PLOS Genetics*, 6(10), p. e1001177. doi: 10.1371/JOURNAL.PGEN.1001177.

Padmanabhan, S., Melander, O., Johnson, T., Di Blasio, A. M., *et al.* (2010) 'Genome-Wide Association Study of Blood Pressure Extremes Identifies Variant near UMOD Associated with Hypertension', *PLOS Genetics*, 6(10), p. e1001177. doi: 10.1371/JOURNAL.PGEN.1001177.

Panza, J. A. *et al.* (1993) 'Effect of increased availability of endothelium-derived nitric oxide precursor on endothelium-dependent vascular relaxation in normal subjects and in patients with essential hypertension', *Circulation*, 87(5), pp. 1475–1481. doi: 10.1161/01.CIR.87.5.1475.

Petersen, J. and Benzeval, M. (2016) 'Untreated hypertension in the UK household population - Who are missed by the general health checks?', *Preventive Medicine Reports*, 4, pp. 81–86. doi: 10.1016/j.pmedr.2016.05.007.

Petrea, R. E. *et al.* (2020) 'Mid to Late Life Hypertension Trends and Cerebral Small Vessel Disease in the Framingham Heart Study', *Hypertension*, pp. 707–714. doi: 10.1161/HYPERTENSIONAHA.120.15073.

Pinto, E. (2007) 'Blood pressure and ageing', *Postgraduate Medical Journal*. BMJ Publishing Group, pp. 109–114. doi: 10.1136/pgmj.2006.048371.

Ponte, B. *et al.* (2021) 'Uromodulin, Salt, and 24-Hour Blood Pressure in the General Population', *Clinical Journal of the American Society of Nephrology*, 16(5), pp. 787–789. doi: 10.2215/CJN.11230720.

Prado, C. M. *et al.* (2008) 'Turbulent blood flow plays an essential localizing role in the development of atherosclerotic lesions in experimentally induced hypercholesterolaemia in rats', *International Journal of Experimental Pathology*, 89(1), pp. 72–80. doi: 10.1111/j.1365-2613.2007.00564.x.

PS, G. *et al.* (2017) 'Urinary Uromodulin and Risk of Urinary Tract Infections: The Cardiovascular Health Study', *American journal of kidney diseases : the official journal of the National Kidney Foundation*, 69(6), pp. 744–751. doi: 10.1053/J.AJKD.2016.08.022.

Qanbari, S. (2020) 'On the Extent of Linkage Disequilibrium in the Genome of Farm Animals', *Frontiers in Genetics*, 0, p. 1304. doi: 10.3389/FGENE.2019.01304.

Qu, J. *et al.* (2020) 'Exact Distribution of Linkage Disequilibrium in the Presence of Mutation, Selection, or Minor Allele Frequency Filtering', *Frontiers in Genetics*, 0, p. 362. doi: 10.3389/FGENE.2020.00362.

'Quantitative screening of the effects of hyper-osmotic stress on cancer cells cultured in 2-or 3-dimensional settings' (2019). doi: 10.1038/s41598-019-50198-w.

Quillet, A. *et al.* (2020) 'Improving Bioinformatics Prediction of microRNA Targets by Ranks Aggregation', *Frontiers in Genetics*, 0, p. 1330. doi: 10.3389/FGENE.2019.01330.

R, B. and PR, F. (1988) 'Scanning electron microscopy of human urine and purified Tamm-Horsfall's glycoprotein', *Scandinavian journal of urology and nephrology*, 22(4), pp. 313–315. doi: 10.3109/00365598809180806.

R, G. and E, S. (1981) 'Presence of luminal K+, a prerequisite for active NaCl transport in the cortical thick ascending limb of Henle's loop of rabbit kidney', *Pflugers Archiv : European journal of physiology*, 392(1), pp. 92–94. doi: 10.1007/BF00584588.

R, M. *et al.* (2020) 'Uromodulin (Tamm-Horsfall protein): guardian of urinary and

systemic homeostasis', *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, 35(1), pp. 33–43. doi: 10.1093/NDT/GFY394.

R, S. *et al.* (2005) 'Prognostic value of ambulatory and home blood pressures compared with office blood pressure in the general population: follow-up results from the Pressioni Arteriose Monitorate e Loro Associazioni (PAMELA) study', *Circulation*, 111(14), pp. 1777–1783. doi: 10.1161/01.CIR.0000160923.04524.5B.

Rafieian-Kopaei, M. *et al.* (2014) 'Atherosclerosis: Process, indicators, risk factors and new hopes', *International Journal of Preventive Medicine*. Isfahan University of Medical Sciences(IUMS), pp. 927–946. Available at: /pmc/articles/PMC4258672/ (Accessed: 23 June 2021).

Ray Chaudhuri, A. and Nussenzweig, A. (2017) 'The multifaceted roles of PARP1 in DNA repair and chromatin remodelling', *Nature Reviews Molecular Cell Biology 2017 18:10*, 18(10), pp. 610–621. doi: 10.1038/nrm.2017.53.

Rieger, A. *et al.* (2016) 'Missense Mutation of POU Domain Class 3 Transcription Factor 3 in Pou3f3L423P Mice Causes Reduced Nephron Number and Impaired Development of the Thick Ascending Limb of the Loop of Henle', *PLoS ONE*, 11(7), p. 158977. doi: 10.1371/JOURNAL.PONE.0158977.

Te Riet, L. *et al.* (2015) 'Hypertension: Renin-Angiotensin-Aldosterone System Alterations', *Circulation Research*. Lippincott Williams and Wilkins, pp. 960–975. doi: 10.1161/CIRCRESAHA.116.303587.

Rigau, M. *et al.* (2019) 'Intronic CNVs and gene expression variation in human populations', *PLOS Genetics*, 15(1), p. e1007902. doi: 10.1371/JOURNAL.PGEN.1007902.

Ring, A. M. *et al.* (2007) 'An SGK1 site in WNK4 regulates Na+ channel and K+ channel activity and has implications for aldosterone signaling and K+ homeostasis',

Proceedings of the National Academy of Sciences of the United States of America, 104(10), pp. 4025–4029. doi: 10.1073/PNAS.0611728104.

Riu, E. *et al*. (2007) 'Histone Modifications are Associated with the Persistence or Silencing of Vector-mediated Transgene Expression In Vivo', *Molecular Therapy*, 15(7), pp. 1348–1355. doi: 10.1038/SJ.MT.6300177.

RJ, K. *et al*. (2010) 'Successes of genome-wide association studies', *Cell*, 142(3), pp. 350–351. doi: 10.1016/J.CELL.2010.07.026.

Roeder, K. and Wasserman, L. (2009) 'Genome-Wide Significance Levels and Weighted Hypothesis Testing', *Statistical science : a review journal of the Institute of Mathematical Statistics*, 24(4), p. 398. doi: 10.1214/09-STS289.

Rose, A. B. (2019) 'Introns as Gene Regulators: A Brick on the Accelerator', *Frontiers in Genetics*, 0(FEB), p. 672. doi: 10.3389/FGENE.2018.00672.

Rosenbluh, J. *et al*. (2016) 'Genetic and Proteomic Interrogation of Lower Confidence Candidate Genes Reveals Signaling Networks in β-Catenin-Active Cancers', *Cell systems*, 3(3), pp. 302-316.e4. doi: 10.1016/J.CELS.2016.09.001.

Roth, G. A. *et al*. (2020) 'Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study', *Journal of the American College of Cardiology*, 76(25), pp. 2982–3021. doi: 10.1016/J.JACC.2020.11.010.

RS, P., S, M. and S, T. (2017) 'Understanding the role of genetics in hypertension', *European heart journal*, 38(29), pp. 2309–2312. doi: 10.1093/EURHEARTJ/EHX273.

Ruiz-Ortega, M. *et al*. (2002) 'Angiotensin II regulates the synthesis of proinflammatory cytokines and chemokines in the kidney', *Kidney International, Supplement*, 62(82), pp. S12–S22. doi: 10.1046/j.1523-1755.62.s82.4.x.

Ryan, M. J. *et al*. (1994) 'HK-2: an immortalized proximal tubule epithelial cell line from normal adult human kidney', *Kidney international*, 45(1), pp. 48–57. doi:

10.1038/KI.1994.6.

S, B. *et al*. (2011) 'Tumor necrosis factor-alpha is an endogenous inhibitor of Na+-K+-2Cl- cotransporter (NKCC2) isoform A in the thick ascending limb', *American journal of physiology. Renal physiology*, 301(1). doi: 10.1152/AJPRENAL.00650.2010.

S, H. (2009) 'Non-genetic heterogeneity of cells in development: more than just noise', *Development (Cambridge, England)*, 136(23), pp. 3853–3862. doi: 10.1242/DEV.035139.

S, H. *et al*. (2020) 'Regulation of NKCC2B by TNF-α in response to salt restriction', *American journal of physiology. Renal physiology*, 318(1), pp. F273–F282. doi: 10.1152/AJPRENAL.00388.2019.

SA, S. *et al*. (2013) 'Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update', *Clinical pharmacology and therapeutics*, 94(3), pp. 317–323. doi: 10.1038/CLPT.2013.105.

Saritas, T. *et al*. (2013) 'SPAK differentially mediates vasopressin effects on sodium cotransporters', *Journal of the American Society of Nephrology : JASN*, 24(3), pp. 407–418. doi: 10.1681/ASN.2012040404.

SC, H. and G, G. (1994) 'Molecular cloning and characterization of the renal diuretic-sensitive electroneutral sodium-(potassium)-chloride cotransporters', *The Clinical investigator*, 72(9), pp. 692–694. doi: 10.1007/BF00212991.

Schaeffer, C., Devuyst, O. and Rampoldi, L. (2021) 'Uromodulin: Roles in Health and Disease', *Annual Review of Physiology Annu. Rev. Physiol. 2021*, 83, pp. 477–501. doi: 10.1146/annurev-physiol-031620.

Schaid, D. J., Chen, W. and Larson, N. B. (2018) 'From genome-wide associations to candidate causal variants by statistical fine-mapping', *Nature reviews. Genetics*, 19(8), p. 491. doi: 10.1038/S41576-018-0016-Z.

Scherberich, J. E. *et al.* (2018) 'Serum uromodulin—a marker of kidney function and renal parenchymal integrity', *Nephrology Dialysis Transplantation*, 33(2), pp. 284–295. doi: 10.1093/NDT/GFW422.

Schwarze, K. *et al.* (2018) 'Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature', *Genetics in Medicine 2018 20:10*, 20(10), pp. 1122–1130. doi: 10.1038/gim.2017.247.

Schwarze, K. *et al.* (2019) 'The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom', *Genetics in Medicine 2019 22:1*, 22(1), pp. 85–94. doi: 10.1038/s41436-019-0618-7.

SE, P. *et al.* (2017) 'A mouse model for inherited renal fibrosis associated with endoplasmic reticulum stress', *Disease models & mechanisms*, 10(6), pp. 773–786. doi: 10.1242/DMM.029488.

Shin, J. *et al.* (2019) 'Comparative analysis of differentially secreted proteins in serum-free and serum-containing media by using BONCAT and pulsed SILAC', *Scientific Reports 2019 9:1*, 9(1), pp. 1–12. doi: 10.1038/s41598-019-39650-z.

Shinagawa, S. *et al.* (2017) 'T cells upon activation promote endothelin 1 production in monocytes via IFN-γ and TNF-α', *Scientific Reports*, 7(1), pp. 1–8. doi: 10.1038/s41598-017-14202-5.

Sica, D. A. (2015) 'Mineralocorticoid Receptor Antagonists for Treatment of Hypertension and Heart Failure', *Methodist DeBakey cardiovascular journal*. Methodist DeBakey Heart & Vascular Center, pp. 235–239. doi: 10.14797/mdcj-11-4-235.

Skube, S. B., Chaverri, J. M. and Goodson, H. V. (2010) 'Effect of GFP tags on the localization of EB1 and EB1 fragments in vivo', *Cytoskeleton (Hoboken, N.J.)*, 67(1), pp. 1–12. doi: 10.1002/CM.20409.

Slavin, S. D. *et al.* (2021) 'Financial burden, distress, and toxicity in cardiovascular disease', *American Heart Journal*, 238, pp. 75–84. doi: 10.1016/J.AHJ.2021.04.011.

Somasekharan, S., Tanis, J. and Forbush, B. (2012) 'Loop diuretic and ion-binding residues revealed by scanning mutagenesis of transmembrane helix 3 (TM3) of Na-K-Cl cotransporter (NKCC1)', *Journal of Biological Chemistry*, 287(21), pp. 17308–17317. doi: 10.1074/jbc.M112.356014.

Spieker, L. E., Flammer, A. J. and Lüscher, T. F. (2006) 'The vascular endothelium in hypertension', *Handbook of Experimental Pharmacology*, 176(PART2), pp. 249–283. doi: 10.1007/3-540-36028-x_8.

Sprague, A. H. and Khalil, R. A. (2009) 'Inflammatory cytokines in vascular dysfunction and vascular disease', *Biochemical Pharmacology*. Elsevier Inc., pp. 539–552. doi: 10.1016/j.bcp.2009.04.029.

Spratt, D. E. (2018) 'Are we inadvertently widening the disparity gap in pursuit of precision oncology?', *British Journal of Cancer 2018 119:7*, 119(7), pp. 783–784. doi: 10.1038/s41416-018-0223-6.

Sriramula, S. *et al.* (2008) 'Involvement of tumor necrosis factor-α in angiotensin II-mediated effects on salt appetite, hypertension, and cardiac hypertrophy', *Hypertension*, 51(5), pp. 1345–1351. doi: 10.1161/HYPERTENSIONAHA.107.102152.

Su, S. J. *et al.* (1997) 'Uromodulin and Tamm-Horsfall protein induce human monocytes to secrete TNF and express tissue factor.', *The Journal of Immunology*, 158(7).

Surendran, P. *et al.* (2020) 'Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals', *Nature Genetics 2020 52:12*, 52(12), pp. 1314–1332. doi: 10.1038/s41588-020-00713-x.

Szczesny, R. J. *et al.* (2018) 'Versatile approach for functional analysis of human

proteins and efficient stable cell line generation using FLP-mediated recombination system', *PLOS ONE*, 13(3), p. e0194887. doi: 10.1371/JOURNAL.PONE.0194887.

T, N. *et al*. (2003) 'Localization and regulation of the epithelial Ca2+ channel TRPV6 in the kidney', *Journal of the American Society of Nephrology : JASN*, 14(11), pp. 2731–2740. doi: 10.1097/01.ASN.0000094081.78893.E8.

Tackling, G. and Borhade, M. B. (2021) *Hypertensive Heart Disease*, *StatPearls*. StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/pubmed/30969622 (Accessed: 23 June 2021).

Tan, K., Tegner, J. and Ravasi, T. (2008) 'Integrated approaches to uncovering transcription regulatory networks in mammalian cells', *Genomics*, 91(3), pp. 219–231. doi: 10.1016/J.YGENO.2007.11.005.

Tanase, D. M. *et al*. (2019) 'Arterial Hypertension and Interleukins: Potential Therapeutic Target or Future Diagnostic Marker?', *International Journal of Hypertension*. Hindawi Limited. doi: 10.1155/2019/3159283.

TC, H. *et al*. (2002) 'Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy', *Journal of medical genetics*, 39(12), pp. 882–892. doi: 10.1136/JMG.39.12.882.

Terker, A. S. *et al*. (2018) 'With no lysine kinase 4 modulates sodium potassium 2 chloride cotransporter activity in vivo', *American journal of physiology. Renal physiology*, 315(4), pp. F781–F790. doi: 10.1152/AJPRENAL.00485.2017.

Thomas, D. B. L. *et al*. (1993) 'Tamm Horsfall protein binds to a single class of carbohydrate specific receptors on human neutrophils', *Kidney International*, 44(2), pp. 423–429. doi: 10.1038/KI.1993.260.

TJ, H. *et al*. (2017) 'Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation', *Nature genetics*,

49(1), pp. 54–64. doi: 10.1038/NG.3715.

Todeschini, A.-L., Georges, A. and Veitia, R. A. (2014) ‘Transcription factors: specific DNA binding and specific gene regulation’, *Trends in Genetics*, 30(6), pp. 211–219. doi: 10.1016/J.TIG.2014.04.002.

Tokonami, N. *et al.* (2018a) ‘The excretion of uromodulin is modulated by the calcium-sensing receptor’, *Kidney International*, 94(5), pp. 882–886. doi: 10.1016/J.KINT.2018.07.022.

Tokonami, N. *et al.* (2018b) ‘The excretion of uromodulin is modulated by the calcium-sensing receptor’, *Kidney International*, 94(5), pp. 882–886. doi: 10.1016/J.KINT.2018.07.022.

Torffvita, O., Melanderb, O. and Hulténb, U. L. (2004) ‘Urinary excretion rate of Tamm-Horsfall protein is related to salt intake in humans’, *Nephron. Physiology*, 97(1). doi: 10.1159/000077600.

Tougas, C. L. *et al.* (2019) ‘Heterozygous Mylk3 Knockout Mice Partially Recapitulate Human DCM With Heterozygous MYLK3 Mutations’, *Frontiers in Physiology*, 0(JUN), p. 696. doi: 10.3389/FPHYS.2019.00696.

Traversari, S. *et al.* (2020) ‘Osmotic adjustments support growth of poplar cultured cells under high concentrations of carbohydrates’, *Plant cell reports*, 39(7), pp. 971–982. doi: 10.1007/S00299-020-02542-Y.

Trudu, M., Janas, S., Lanzani, C., Debaix, H., Schaeffer, C., Ikehata, M., Citterio, L., Demaretz, S., Trevisani, F., Ristagno, G., Glaudemans, B., Laghmani, K., Dell’Antonio, G., *et al.* (2013a) ‘Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression’, *Nature medicine*, 19(12), pp. 1655–1660. doi: 10.1038/NM.3384.

Trudu, M., Janas, S., Lanzani, C., Debaix, H., Schaeffer, C., Ikehata, M., Citterio, L.,

Demaretz, S., Trevisani, F., Ristagno, G., Glaudemans, B., Laghmani, K., Dell'Antonio, G., *et al.* (2013b) 'Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression', *Nature medicine*, 19(12), pp. 1655–1660. doi: 10.1038/NM.3384.

Trudu, M., Janas, S., Lanzani, C., Debaix, H., Schaeffer, C., Ikehata, M., Citterio, L., Demaretz, S., Trevisani, F., Ristagno, G., Glaudemans, B., Laghmani, K., Dell'antonio, G., *et al.* (2013) 'Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression Europe PMC Funders Group', *Nat Med*, 19(12). doi: 10.1038/nm.3384.

Tuteja, S. *et al.* (2020) 'Prospective CYP2C19 Genotyping to Guide Antiplatelet Therapy Following Percutaneous Coronary Intervention', *Circulation: Genomic and Precision Medicine*, 13, p. 2640. doi: 10.1161/CIRCGEN.119.002640.

Tuteja, S. and Rader, D. J. (2018) 'SLCO1B1 and Statin Therapy', *Circulation: Genomic and Precision Medicine*, 11(9), p. e002320. doi: 10.1161/CIRCGEN.118.002320.

Unger, T. *et al.* (2020) '2020 International Society of Hypertension Global Hypertension Practice Guidelines', *Hypertension*, 75(6), pp. 1334–1357. doi: 10.1161/HYPERTENSIONAHA.120.15026.

Vedula, P. *et al.* (2017) 'Diverse functions of homologous actin isoforms are defined by their nucleotide, rather than their amino acid sequence', *eLife*, 6. doi: 10.7554/ELIFE.31661.

Verhave, J. C. *et al.* (2016) 'Hepatocyte Nuclear Factor 1ß–Associated Kidney Disease: More than Renal Cysts and Diabetes', *Journal of the American Society of Nephrology : JASN*, 27(2), p. 345. doi: 10.1681/ASN.2015050544.

Wall, J. D. and Pritchard, J. K. (2003) 'Haplotype blocks and linkage disequilibrium in the human genome', *Nature Reviews Genetics 2003 4:8*, 4(8), pp. 587–597. doi:

10.1038/nrg1123.

Wang, Y. and Wang, J.-G. (2018) ‘Genome-Wide Association Studies of Hypertension and Several Other Cardiovascular Diseases’, *Pulse*, 6(3–4), pp. 169–186. doi: 10.1159/000496150.

Wang, Y. and Wang, J.-G. (2019) ‘Genome-Wide Association Studies of Hypertension and Several Other Cardiovascular Diseases’, *Pulse*, 6(3–4), p. 169. doi: 10.1159/000496150.

Wang, Z., Gerstein, M. and Snyder, M. (2009) ‘RNA-Seq: a revolutionary tool for transcriptomics’, *Nature reviews. Genetics*, 10(1), p. 57. doi: 10.1038/NRG2484.

Ward, M. C. and Gilad, Y. (2017) ‘Cracking the regulatory code’, *Nature 2017 550:7675*, 550(7675), pp. 190–191. doi: 10.1038/550190a.

Weaver, A. N. and Yang, E. S. (2013) ‘Beyond DNA repair: Additional functions of PARP-1 in cancer’, *Frontiers in Oncology*, 3 NOV, p. 290. doi: 10.3389/FONC.2013.00290/BIBTEX.

Weber, M. A. *et al.* (2009) ‘A selective endothelin-receptor antagonist to reduce blood pressure in patients with treatment-resistant hypertension: a randomised, double-blind, placebo-controlled trial’, *The Lancet*, 374(9699), pp. 1423–1431. doi: 10.1016/S0140-6736(09)61500-2.

Welling, P. A. and Ho, K. (2009) ‘A comprehensive guide to the ROMK potassium channel: form and function in health and disease’, *American Journal of Physiology - Renal Physiology*, 297(4), p. F849. doi: 10.1152/AJPRENAL.00181.2009.

Whelton, P. K. *et al.* (2018) ‘2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: Executive summary: A report of the American college of cardiology/American Heart Association task force on clinical

practice guidelines', *Hypertension*. Lippincott Williams and Wilkins, pp. 1269–1324. doi: 10.1161/HYP.0000000000000066.

Whitesall, S. E. *et al*. (2004) 'Comparison of simultaneous measurement of mouse systolic arterial blood pressure by radiotelemetry and tail-cuff methods', *American Journal of Physiology - Heart and Circulatory Physiology*, 286(6 55-6), pp. 2408–2415. doi: 10.1152/AJPHEART.01089.2003/ASSET/IMAGES/LARGE/ZH40060430850006.JPEG.

Wierzejska, E. *et al*. (2020) 'A global perspective on the costs of hypertension: a systematic review', *Archives of Medical Science*. Termedia Publishing House Ltd., pp. 1078–1091. doi: 10.5114/AOMS.2020.92689.

Wilde, E. *et al*. (2017) 'Tail-cuff technique and its influence on central blood pressure in the mouse', *Journal of the American Heart Association*, 6(6). doi: 10.1161/JAHA.116.005204.

Wolf, M. T. F., Zhang, J. and Nie, M. (2019) 'Uromodulin in Mineral Metabolism', *Current opinion in nephrology and hypertension*, 28(5), p. 481. doi: 10.1097/MNH.0000000000000522.

Wu, H. *et al*. (2017) 'Polychlorinated biphenyls-153 induces metabolic dysfunction through activation of ROS/NF-κB signaling via downregulation of HNF1b', *Redox biology*, 12, pp. 300–310. doi: 10.1016/J.REDOX.2017.02.026.

Y, W. *et al*. (2019) 'Genome-wide association study of medication-use and associated disease in the UK Biobank', *Nature communications*, 10(1). doi: 10.1038/S41467-019-09572-5.

Yasuhara, A. *et al*. (2008) 'Collectrin is involved in the development of salt-sensitive hypertension by facilitating the membrane trafficking of apical membrane proteins via interaction with soluble N-ethylmaleiamide-sensitive factor attachment protein receptor complex', *Circulation*, 118(21), pp. 2146–2155. doi:

10.1161/CIRCULATIONAHA.108.787259.

Yim, H. E. and Yoo, K. H. (2008) 'Renin-angiotensin system - Considerations for hypertension and kidney', *Electrolyte and Blood Pressure*. Korean Society of Electrolyte and Blood Pressure Research, pp. 42–50. doi: 10.5049/EBP.2008.6.1.42.

Ying, W. Z. and Sanders, P. W. (2005) 'Enhanced expression of EGF receptor in a model of salt-sensitive hypertension', *American journal of physiology. Renal physiology*, 289(2). doi: 10.1152/AJPRENAL.00003.2005.

Yugar-Toledo, J. C. *et al.* (2017) 'Managing resistant hypertension: Focus on mineralocorticoid-receptor antagonists', *Vascular Health and Risk Management*. Dove Medical Press Ltd., pp. 403–411. doi: 10.2147/VHRM.S138599.

Zeldovich, L. (2017) 'Genetic drift: the ghost in the genome', *Lab Animal 2017 46:6*, 46(6), pp. 255–257. doi: 10.1038/laban.1275.

Zhang, M. *et al.* (2020) 'Immunoprecipitation and mass spectrometry define TET1 interactome during oligodendrocyte differentiation', *Cell and Bioscience*, 10(1), pp. 1–11. doi: 10.1186/S13578-020-00473-5/FIGURES/6.

Zhao, X. *et al.* (2011) 'Arterial Pressure Monitoring in Mice', *Current protocols in mouse biology*, 1, p. 105. doi: 10.1002/9780470942390.MO100149.

Zhou, B. *et al.* (2019) 'Long-term and recent trends in hypertension awareness, treatment, and control in 12 high-income countries: an analysis of 123 nationally representative surveys NCD Risk Factor Collaboration (NCD-RisC)*'. doi: 10.1016/S0140-6736(19)31145-6.

Zhou, Q. *et al.* (2017) 'A mouse tissue transcription factor atlas', *Nature Communications 2017 8:1*, 8(1), pp. 1–15. doi: 10.1038/ncomms15089.

Zhu, X. *et al.* (2002) 'Isolation of mouse THP gene promoter and demonstration of its kidney-specific activity in transgenic mice',

https://doi.org/10.1152/ajprenal.00297.2001, 282(4 51-4), pp. 608–617. doi: 10.1152/AJPRENAL.00297.2001.