3’ UTR Structural Elements in *CYP24A1* are Associated with Infantile Hypercalcaemia Type 1

Nicole Ball

Thesis submitted to the University of East Anglia in accordance with the requirements for the degree of Doctor of Philosophy (PhD)

Norwich Medical School

Faculty of Medicine and Health Sciences

University of East Anglia

University of East Anglia

2022

ABSTRACT

Ribonucleic acids (RNAs) fold into complex structures that are critical for their function and regulation including post-transcriptional modification, localisation, translation and degradation. RNA structure and potential RNA misfolding has scarcely been studied in a clinical setting. Hypomorphic mutations in the cytochrome P450 family 24 subfamily A member 1 (*CYP24A1)* protein coding region causing inappropriately elevated active vitamin D metabolites have been observed in some cases of idiopathic infantile hypercalcemia and adult-onset nephrolithiasis. It is unclear why a subset of cases present with superficial *CYP24A1* mediated hypercalcemia (CMH) but do not exhibit protein-coding mutations in *CYP24A1*. This thesis presents a combination of biochemical profiling, next generation sequencing, bioinformatics, proteomic and molecular cytogenetic approaches to examine *CYP24A1* in a patient cohort with apparent CMH. This work identified several novel single nucleotide variants (SNVs) located in the *CYP24A1* 3' untranslated region (UTR). These SNVs led to *CYP24A1* messenger RNA (mRNA) misfolding. The mRNA structural abnormalities observed were associated with an over accumulation of an apparently less active CYP24A1 protein. The generation of a CMH cell line, using CRISPR, mimicking patients with *CYP24A1* 3' UTR variants causing mRNA structural alterations provided a model system for *in vitro* investigations into so-called non-canonical CMH. Subsequent single molecule fluorescence in situ hybridisation (smFISH) methods provided *CYP24A1* cellular localisation in addition to mRNA abundance *in vitro* and *ex vivo*. The important advancements presented in this thesis are valuable to understanding mRNA structure-function relationships and novel *CYP24A1* mutations, which affect mRNA translation and protein expression. The findings of this research provide a framework that can be used to better understand the molecular basis of pathogenesis in patients lacking protein coding region abnormalities.

TABLE OF CONTENTS

CHAPTER 1: AN INTRODUCTION INTO BONE AND RNA BIOLOGY

CHAPTER 2: MATERIALS AND METHODS

FIGURES

All figures in this thesis were created with BioRender.com

INTRODUCTION:

CHAPTER 3: THE RELATIONSHIP BETWEEN ACTIVE AND CATABOLIC VITAMIN D METABOLITE RATIOS ASSOCIATED WITH PTH REGULATION AND CALCIUM HOMEOSTASIS

CHAPTER 4: WHOLE EXOME SEQUENCING IN PATIENTS WITH SUSPECTED CMH

CHAPTER 5: mRNA STRUCTURAL ELEMENTS IN THE 3' UTR DICTATE CYP24A1 INTRACELLULAR ACTIVITY

CHAPTER 6: PROBE-BASED mRNA STRUCTURE ANALYSIS CONFIRMS 3' UTR VARIANTS IN CYP24A1 INDUCE mRNA STRUCTURE ALTERATIONS

CHAPTER 7: INTRODUCTION OF SNVs INTO HUMAN EMBRYONIC KIDNEY CELLS USING CRISPR-CAS 9

CHAPTER 8: A METHOD FOR SINGLE MOLECULE mRNA CYP24A1 DETECTION IN IN VITRO HUMAN CELL LINE AND EX VIVO SAMPLES

CHAPTER 9: DISCUSSION

TABLES

INTRODUCTION

MATERIALS AND METHODS

CHAPTER 3: THE RELATIONSHIP BETWEEN ACTIVE AND CATABOLIC VITAMIN D METABOLITE RATIOS ASSOCIATED WITH PTH REGULATION AND CALCIUM HOMEOSTASIS

CHAPTER 4: WHOLE EXOME SEQUENCING IN PATIENTS WITH SUSPECTED CMH

CHAPTER 5: mRNA STRUCTURAL ELEMENTS IN THE 3' UTR DICTATE CYP24A1 INTRACELLULAR ACTIVITY

DECLARATION

The research within this thesis was designed and conducted by the author under the guidance of supervisors Dr Darrell Green and Prof William Fraser. The author was at no time registered for any other university award, degree, or qualification during the registration for the degree of Doctor of Philosophy. All experiments were conducted between the laboratories of Dr Darrell Green and Prof William Fraser at the Norwich Medical School between September 2018 and December 2021. The research was funded by the internal FMH PhD Programme.

Word Count: 63,775

ETHICS STATEMENT

The human sample collection, analysis and storage within this thesis was conducted in accordance the Human Tissue Act 2004. The study in Chapter 3 of this thesis received ethics approval from the UK Ministry of Defence Research Ethics Committee (MODREC 165/Gen/10 and 692/MoDREC/15 ClinicalTrials.gov Identifier) and was conducted in accordance with the Declaration of Helsinki (2013). The author submitted a detailed application to the University of East Anglia Faculty of Medicine and Health Sciences Research Ethics Committee for project ethical approval. This was approved by Prof Alastair Forbes on 14th November 2019 (REF: 2018/19-100).

ACKNOWLEGEMENTS

First, I wish to thank my PhD supervisor Dr Darrell Green. Without his support and guidance over the past 4 years I would not have made it to the end of this journey. I have grown and developed the skills to become an independent researcher under his supervision and am grateful for the invaluable advice given during my studies. Thank you to my secondary supervisor Prof William Fraser and everyone at the Bioanalytical Facility for their invaluable advice and input during my PhD.

To Dr Yiliang Ding, Dr Yeuying Zhang and Dr Susan Duncan, thank you for the time you gave towards our collaboration and the direction you provided. I also would like to thank Dr Rocky Payet for his bioinformatics expertise.

Thank you to Dr Matthew Jefferson and Dr Yingxue Wang from whom I have learnt a huge amount. If ever I needed advice or guidance in the lab I knew I could turn to you both, for which I am grateful. I would have been lost without you both in those early days. To the BMRC PhD and Masters students I have met throughout the past couple of years, thank you so much for all your support and friendship.

Finally, to my friends and family, thank you for putting up with me being constantly stressed, always busy with PhD commitments, and being on an emotional rollercoaster (especially in the last year). Your understanding has meant everything and I couldn't have reached the end of this degree without your encouragement. Mum, Dad and James, you've had the brunt of this and I will be forever grateful to you for keeping me going.

GRANT AWARDS AND PUBLICATIONS

GRANT AWARDS DURING THIS PROJECT

- **Paget's Association £6,000** "Application of RIP-Seq technology to determine the role of SQSTM1 as an RNA binding protein" Ball, N. & Green, D.
- **Young Investigator Travel Award** (ASBMR 2022)
- **International Fellows Forum Grant** (EFF 2022)
- **Best Oral Poster Award** (British Endocrinology Society 2022)

PUBLICATIONS CONTAINING WORK FROM THIS THESIS

- Tang, J. C. Y., Jackson, S., Walsh, N. P., Greeves, J. & Fraser, W. D. The dynamic relationships between the active and catabolic vitamin D metabolites, their ratios, and associations with PTH. *Sci. Rep.* **9**, 6974 (2019).
- Ball, N., Duncan, S., Piec, I., Tang, J. C. T., Schoenmakers, I., Lopez, B., Chipchase, A., Kumar, A., Perry, L., Maxwell, H., Brewer, D., Ding, Y., Fraser, W. D., Green, D. mRNA structural elements in the 3' UTR dictate CYP24A1 intracellular activity. Journal of Bone and Mineral research. TBC (2023).

POSTER/ORAL PRESENTATIONS CONTAINING WORK FROM THIS THESIS

- Ball, N., Duncan, S., Piec, I., Tang, J. C. T., Schoenmakers, I., Lopez, B., Chipchase, A., Kumar, A., Perry, L., Maxwell, H., Brewer, D., Ding, Y., Fraser, W. D., Green, D. mRNA structural elements in the 3' UTR dictate CYP24A1 intracellular activity. Vitamin D Workshop Annual Conference – Oral Presentation September 2022

- Ball, N., Duncan, S., Piec, I., Tang, J. C. T., Schoenmakers, I., Lopez, B., Chipchase, A., Kumar, A., Perry, L., Maxwell, H., Brewer, D., Ding, Y., Fraser, W. D., Green, D. mRNA structural elements in the 3’ UTR dictate CYP24A1 intracellular activity. ASBMR Annual Conference – Oral Presentation September 2022
- Ball, N., Duncan, S., Piec, I., Tang, J. C. T., Schoenmakers, I., Lopez, B., Chipchase, A., Kumar, A., Perry, L., Maxwell, H., Brewer, D., Ding, Y., Fraser, W. D., Green, D. mRNA structural elements in the 3’ UTR dictate CYP24A1 intracellular activity. ASBMR Annual Conference – Poster Presentation September 2022
- Ball, N., Duncan, S., Piec, I., Tang, J. C. T., Schoenmakers, I., Lopez, B., Chipchase, A., Kumar, A., Perry, L., Maxwell, H., Brewer, D., Ding, Y., Fraser, W. D., Green, D. mRNA structural elements in the 3’ UTR dictate CYP24A1 intracellular activity. EFF Annual Conference – Poster Presentation September 2022
- Ball, N., Duncan, S., Piec, I., Tang, J. C. T., Schoenmakers, I., Lopez, B., Chipchase, A., Kumar, A., Perry, L., Maxwell, H., Brewer, D., Ding, Y., Fraser, W. D., Green, D. mRNA structural elements in the 3’ UTR dictate CYP24A1 intracellular activity. BES Annual Conference – Poster Presentation November 2022
- Ball, N., Duncan, S., Piec, I., Tang, J. C. T., Schoenmakers, I., Lopez, B., Chipchase, A., Kumar, A., Perry, L., Maxwell, H., Brewer, D., Ding, Y., Fraser, W. D., Green, D. mRNA structural elements in the 3’ UTR dictate CYP24A1 intracellular activity. BES Annual Conference – Oral Presentation November 2022

ABBREVIATIONS

$1,24,25(OH)_3D$ – 1,24,25 trihydroxyvitamin D

$1,25(OH)_2D$ – 1,25-dihydroxyvitamin D

143B – Human osteosarcoma cell line

$24,25(OH)_2D$ – 24,25-Dihydroxycholecalciferol

25OHD – 25-hydroxyvitamin D

7DHC – Pre-vitamin 7-dehydrocholesterol

Aca – Albumin-adjusted calcium

ALP – Total Alkaline Phosphatase

AUC – Area under the curve

AVRC – Arrhythmogenic right ventricular cardiomyopathy/dysplasia

BLAST – Basic local alignment search tool

BMD – Bone mineral density

BMI – Body mass index

BMP – Bone morphogenic proteins

Cas9 – CRISPR associated protein 9

CASR – Calcium-sensing receptor

CDC – Centres for Disease Control and Prevention

CE – Capillary electrophoresis

CGRP – Calcitonin gene related peptide

CHD – Congenital heart disease

CI – Confidence interval

CKD – Chronic kidney disease

CMH – CYP24A1 mediated hypercalcemia

CNV – Copy number variants

CRISPR – Clustered regularly interspaced short palindromic repeats

CSF1 – Colony-stimulating factor 1

CV – Coefficient of variation

CYP – Cytochrome P450

CYP24A1 – Cytochrome P450 family 24 subfamily A member 1

CYP27B1 – Cytochrome P450 family 27 subfamily B member 1

CYP2R1 – Cytochrome P450 family 2 subfamily R member 1

DBP – Vitamin D binding protein

DCSTAMP – Dendroyte expressed seven transmembrane protein

DEQAS – Vitamin D external quality assessment

DMS – Dimethyl sulfate

DNA – Deoxyribonucleic acid

DSB – Double stranded break

ECLIA – Electrochemiluminescence immunoassay

ESI – Electrospray ionisation

EtBr – Ethidium bromide

FDX – Iron-sulphur protein ferredoxin

FDXR – Flavoprotein ferredoxin reductase

FGF – Fibroblast growth factor

FGF23 – Fibroblast growth factor 23

FGFR – Fibroblast growth factor receptors

FISH – Fluorescence in situ hybridisation

GWAS – Genome wide association studies

HDR – Homologous directed repair

HEK293T – Human embryonic kidney cell line

HRP – Horseradish peroxidase

IGF – Insulin like growth factor

IIH – Idiopathic Infantile Hypercalcaemia

HCINF1 – Infantile Hypercalcaemia Type 1

IOM – Institute of Medicine

IQC – Internal quality control

IRES – Internal ribosome entry sites

KL – Klotho

LC-MS/MS – Liquid chromatography tandem mass spectrometry

LC3 – Microtubule-associated protein light chain 3

LD50 – Mean lethal dose of a drug or other substance that, when administered to a group of experimental animals, will kill 50 per cent of the group in a specified time

LloQ – Lower limit of quantification

MAPK – Mitogen activated protein kinase

MESOR – Midline estimate statistic of rhythm

MFE – Minimum Free Energy

miRNA – microRNA

MMP – Matrix matallopeptidases

MODREC – Ministry of Defence Research Ethics Committee

MRE – miRNA recognition element

mRNA – messenger RNA

NaPi-2a – Sodium-dependent phosphate transport protein 2A

NGS – Next generation sequencing

NHEJ – Non-homologous end joining

NIST – National Institute of Science and Technology

NMIA – N-methyl isotonic anhydride

NMR- Nuclear magnetic resonance

NNUH – Norfolk and Norwich University Hospital

ROS – Royal Osteoporosis Society

NUP205 – Nucleoporin 205

OPTN – Optineurin

OSX – Osteoblast-specific transcription factor

PBMCs – Peripheral blood mononuclear cells

PBS – Phosphate buffered saline

PHEX – Phosphate regulating endopeptidase homolog

PML – Promyelocytic leukemia protein

POLR2A – RNA Polymerase II Subunit A

PTH – Parathyroid hormone

PTH1R – Parathyroid receptor 1

RBP – RNA-binding proteins

RDA – Recommended daily allowance

RFLP – Restriction fragment length polymorphism

RNA – Ribonucleic acid

RIN3 – Ras And Rab Interactor 3

RMDB – RNA Mapping Data Band

ROC – Receiver Operating Characteristic

RUNX – Runt transcription factor family

RXR – Retinoid X receptor

SCCP – Secretory calcium binding phosphoproteins

SHAPE – 2'-hydroxyl acylation primer extension reverse transcription

shRNA – Short hairpin RNA

SLC34A1 – Solute carrier family 34 member 1

SLC34A3 – Solute carrier family 34 member 3

smFISH – Small molecule fluorescence in situ hybridisation

SMH – SLC34A1 mediated hypercalcemia

SNP – Single nucleotide polymorphisms

SNV – Single nucleotide variant

SOX9 – SRY-box transcription factor 9

SPSS – Statistical Package for the Social Science

SQSTM1 – Sequestosome-1

TALEN – Transcription activator-like effector nucleases

TGFβ – Transforming growth factor β

TNFRSF11 – TNF Receptor Superfamily Member 11b

TNFSF11 – Tumour necrosis factor ligand superfamily member 11

UL – Upper intake level

uORF – Upstream open reading frame

UTR – Untranslated region

UVB – Ultraviolet B

VCP – Valosin Containing Protein

VDR – Vitamin D receptor

VICtORy – Vitamin D and CardiOvascular Risk

VMR – Vitamin D metabolite relative ratio

WES – Whole exome sequencing

WGS – Whole genome sequencing

Wnt – Wingless-type genes

ZFN – Zinc finger nuclease

CHAPTER 1: AN INTRODUCTION INTO BONE AND RNA BIOLOGY

1.1 SKELETAL EVOLUTION, STRUCTURE AND DEVELOPMENT

Violent tectonic plate shifts over a billion years ago forced rock minerals such as calcium carbonate into the oceans. Oceanic organisms adapted to the increased mineral content using calcium carbonate to develop hard, protective body parts, e.g., shells and/or spines[1]. Exoskeleton adaptations have been observed in 500 million year old mineralised fossils that provide evidence for increased multicellular organism diversity and increased animal evolution rate[2]. While the newly established exoskeleton had a protective element, it limited the movement and organism locomotion and prevented the development of surface sensory organs, e.g., skin. Exoskeleton limitations influenced the dislocation of mineralised skeleton from the outside of the organism to inside the body in the next major evolutionary enhancement of skeletal development[1]. The evolutionary rise of vertebrates comprising endoskeletons allowed for increased radial activity and increased locomotion modes, e.g., running, swimming and flying. Surface sensory organs such as skin then developed, which provided interaction with physical surroundings. Newly adapted vertebrates favoured their endoskeleton mineral composition of calcium phosphate as opposed to calcium carbonate mineral composition observed in exoskeletons prior to this evolutionary shift. Calcium phosphate in skeletal systems is beneficial due to its porosity that allows more elasticity and strength for locomotion, increased vascularity and surface area of bone promoting metabolite exchange, plus increased chemical resistance and bone dissolution due to low solubility[3].

In the vertebrate lineage, the earliest known species to develop mineralised cartilaginous endoskeletons were *Chondrichthyes* over 422 million years ago[4]. It was not until two million years later that *Osteichthyes* developed an endoskeleton containing the same four mineralised tissues as modern mammals. These tissues include bone, cartilage, enamel and dentin. Controversy surrounded the initial skeletal tissue classification because primitive fossilised vertebral skeletons were scarce and difficult to characterise[1]. In recent years evolutionary genetics has supported fossil classification by identifying that the four mineralised tissues involved in skeletogenesis express specific secretory calcium-binding phosphoproteins (*SCPP*) genes that can be used for classification. Distinct core gene networks have been Identified in the four mineralised tissues that play a critical role in vertebrate phylogeny. For example, the Runt transcription factor family (*RUNX*) have been shown to regulate skeletogenesis, cartilage development and bone cell differentiation [5]. Identification of these networks have aided fossil classification. While the molecular mechanisms of mammalian skeletogenesis are generally well understood, the origin of genetic networks regulating skeletal development requires further research. Future investigations into as yet unknown core gene networks involved in skeletogenesis would expand the current understanding of skeletal evolution.

Bone has evolved to become a highly specialised endocrine organ that also provides internal support in higher vertebrates. Bone has key biological functions including support, protection, energy metabolism, male fertility, blood cell production, participation in calcium-phosphate homeostasis and provides a major inorganic ion source[6]. The extracellular bone matrix is mineralised providing increased skeletal rigidity and strength whilst still maintaining elasticity. The extracellular matrix consists of both mineral and organic phases. The organic bone matrix is predominantly formed from type I collagen (90%) while the remaining 10% consists of proteoglycans plus numerous non-collagenous proteins produced by bone cells[6]. The five known major types of bone cells include chondrocytes, osteoblasts, osteocytes, osteoclasts and bone lining cells. Osteoblasts, osteoclasts and bone lining cells reside on the bone surface, deriving from mesenchymal progenitor cells. Osteocytes in the interior of the bone are derived from osteoblasts that become trapped in the matrix that they secrete. Chondrocytes are the only cell observed in the cartilage[6] (Figure 1.1). Chondrocytes secrete extracellular matrix aiding the proliferation and cartilage maintenance. Chondrocyte maturation is tightly controlled by signalling molecules such as fibroblast growth factors (FGFs), bone morphogenetic proteins (BMPs), wingless-type (Wnt) genes, runt-related transcription factor (RUNX), paracrine signalling and osteoblast-specific transcription factor (OSX)[7] (Figure 1.1) (Table 1.1).

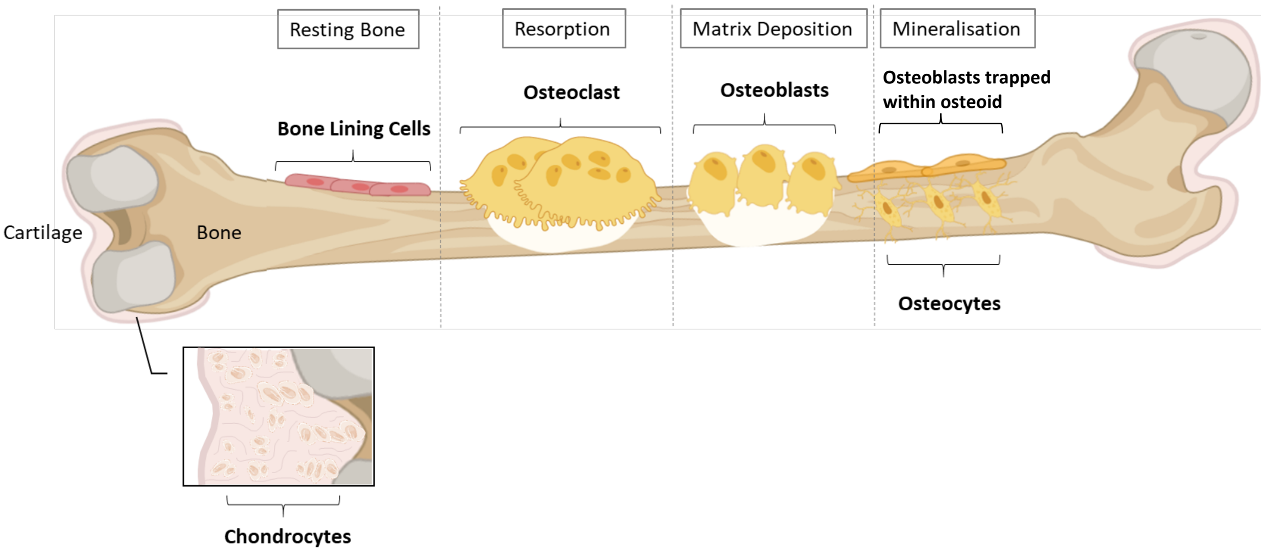


Figure 1.1: Bone cells and their function. Chondrocytes secrete extracellular matrix aiding the proliferation and cartilage maintenance[7]. Bone lining cells of osteoblast lineage regulate the transport of calcium into and out of the bone by stimulating osteoclast activity thereby regulating degradation and bone matrix resorption. Osteoclasts are multinucleated cells derived from haematopoietic stem cells. Osteoclasts degrade inorganic and bone protein molecules via the production of acidic hydrogen ions, cathepsin K and matrix metallopeptidases (MMPs). Calcium and phosphate are released from bone into the circulation via osteoclast bioactivity. Osteoblasts are mesenchymal-derived cells that produce an immature collagenous bone matrix (osteoid). Osteocytes are formed from osteoblasts trapped in the osteoid and so change their phenotype. Osteocytes provide strength to the bone matrix and along with osteoblasts produce important signalling molecules such as FGF23[6].

Table 1.1: Bone cells–- lineage, function and key signalling pathways[8]

Bone Cell	Lineage	Role	Signalling
Chondrocytes	Derived from pluripotent mesenchymal stem cells	Chondrocytes interact with surrounding matrix through cytoplasmic extensions[9] Chondroytes secrete type II collagen rich cartilage, which forms the endochondral skeleton Hypotrophic differentiation allows mineralisation of the collagen matrix forming the bone template Chondroyte activity continues in the proximal and distal ends of long bone[10,11] Chondroytes then apoptose or differentiate into osteoblasts[11]	Differentiation of chondrocytes is mediated by the IIH/PTHrP negative feedback loop, GFG signalling and key transcription factors SOX9 and Runx2[12,13] Chondrocytes also express RANKL, which controls resorption of mineralised cartilage[13].

Osteoblasts	Primarily, mesenchymal stem cells differentiate into osteoblasts, while bone lining cells and chondrocytes have also been shown to differentiate to osteoblasts[11] Osteoblasts then either i) differentiate to bone lining cells, ii) osteocytes, ii) or apoptose	Osteoblasts secrete type I collagen rich bone matrix Osteoblasts also regulate mineralisation of the collagen matrix[14]	Osteoblast differentiation is regulated by transcription factors e.g. SOX9 and Runx2[15,16]. The Wnt singalling pathway, FGF and BMP signalling regulate osteoblast genesis[17–19].
Osteoclasts	Derived from monocyte/macrophages[20]	Osteoclasts secrete hydrochloric acid to dissolve bone and cathepsin K to breakdown the matrix[21]	M-CSF and RANKL stimulate osteoclast differentiation[20] The decoy receptor, OPG, downregulates osteoclast genesis by competitively binding to RANKL to prevent RANK stimulation[22,23]

Osteocyte	Derived from osteoblasts trapped in bone matrix[24]	Respond to bone loading signals by stimulating bone modelling and remodelling through osteoclast and osteoblast activity[25,26]. Stimulate FGF23 secretion aiding mineral homeostasis by reducing serum phosphate, renal phosphate absorption and vitamin D activation[27–29]	Primary source of RANKL, which aids osteoclastgenesis[13]. Inhibit osteoblast formation by secreting Wnt singalling inhibitors e.g. SOST and DKK-1. In contrast, under mechanical loading, osteocytes secret SOST and DKK-1 to stimulate osteoblast differentiation and subsequent bone formation[30].

Two morphological forms of bone exist, termed cortical and cancellous bone. Cortical and cancellous bone comprise different structural and functional properties. Cortical bone is formed from compact collagen fibrils that provides protection, e.g., forming the outer shell surrounding cancellous bone in vertebrates. Cortical bone is dense with porosity between 5-10% and is found in the shaft of long bones. Cancellous bone is involved in metabolic functions and has a spongy and porous matrix (porosity 50-90%). Cancellous bone is found at the end of long bones and in flat bones, e.g., the vertebrae, skull and pelvis[31].

During the early stages of human embryonic development, the embryo's skeleton consists of a fibrous membrane and hyaline cartilage. Bone ossification is the bone formation process that begins in the 6-7th week of human embryonic development[32]. Two types of bone ossification processes produce mature bone, intramembranous and endochondral. In both processes a primary ossification centre is formed from which bone formation initiates. Intramembranous ossification involves the direct differentiation of mesenchymal cells to osteoblasts (Figure 1.2).

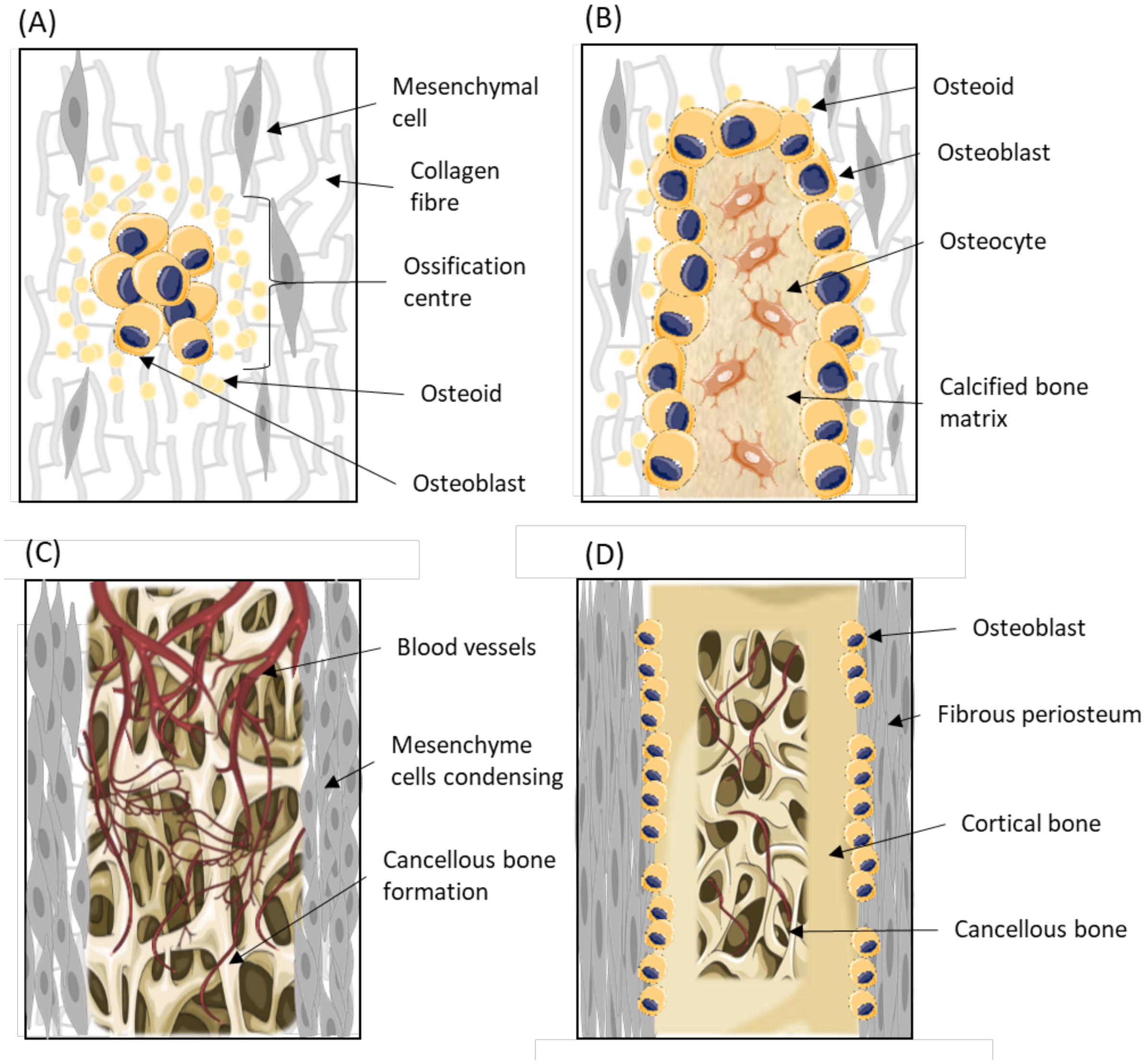


Figure 1.2: Skeletal development by intramembranous ossification. (**A**) Mesenchymal cells differentiate into osteoblasts to form the primary ossification centre from which new bone is formed. Osteoblast clusters secrete collagen-proteoglycan matrix (osteoid) that binds calcium salts. Calcium binding calcifies and hardens the osteoid matrix, entrapping osteoblasts. (**B**) Entrapped osteoblasts differentiate into osteocytes. Osteoblasts continue to secrete osteoid triggering the calcified bone matrix formation. (**C**) Blood vessels infiltrate the calcified matrix and aid cancellous bone formation. On the newly formed bone surface, mesenchyme cells form a dense layer of vascular connective tissue called the periosteum. (**D**). Osteoblasts on the inside of the periosteum continue to secrete osteoid in parallel to the existing matrix creating layers. The layers of bone produced from intramembranous ossification form the dense cortical bone.

Intramembranous ossification forms the flat bones in the skull, clavicle and cranial bones. The majority of long bones (e.g., femur, tibia and humerus) are formed by both intramembranous and endochondral ossification[32]. Endochondral ossification originates from primary ossification centres within the cartilage forming the diaphysis of bones during foetal development, developing the mature bone from an intermediate hyaline cartilage template. Hyaline cartilage forms the temporary embryonic skeleton during foetal development that guides endochondral ossification to form skeletal bone in its place (Figure 1.3). After birth, chondrocytes differentiate into their own secondary ossification centre forming the epiphyses of long bones, which is separated from the diaphysis by the epiphyseal plate. Longitudinal bone growth occurs at the epiphyseal plate where cartilage in this region undergoes mitosis. Secondary ossification results in the hyaline cartilage being completely replaced by bone (Figure 1.3). During puberty, oestrogen and testosterone release triggers epiphyseal closure as the epiphyseal cartilage is replaced with bone joining the epiphyseal plate and diaphysis known as the epiphyseal line.

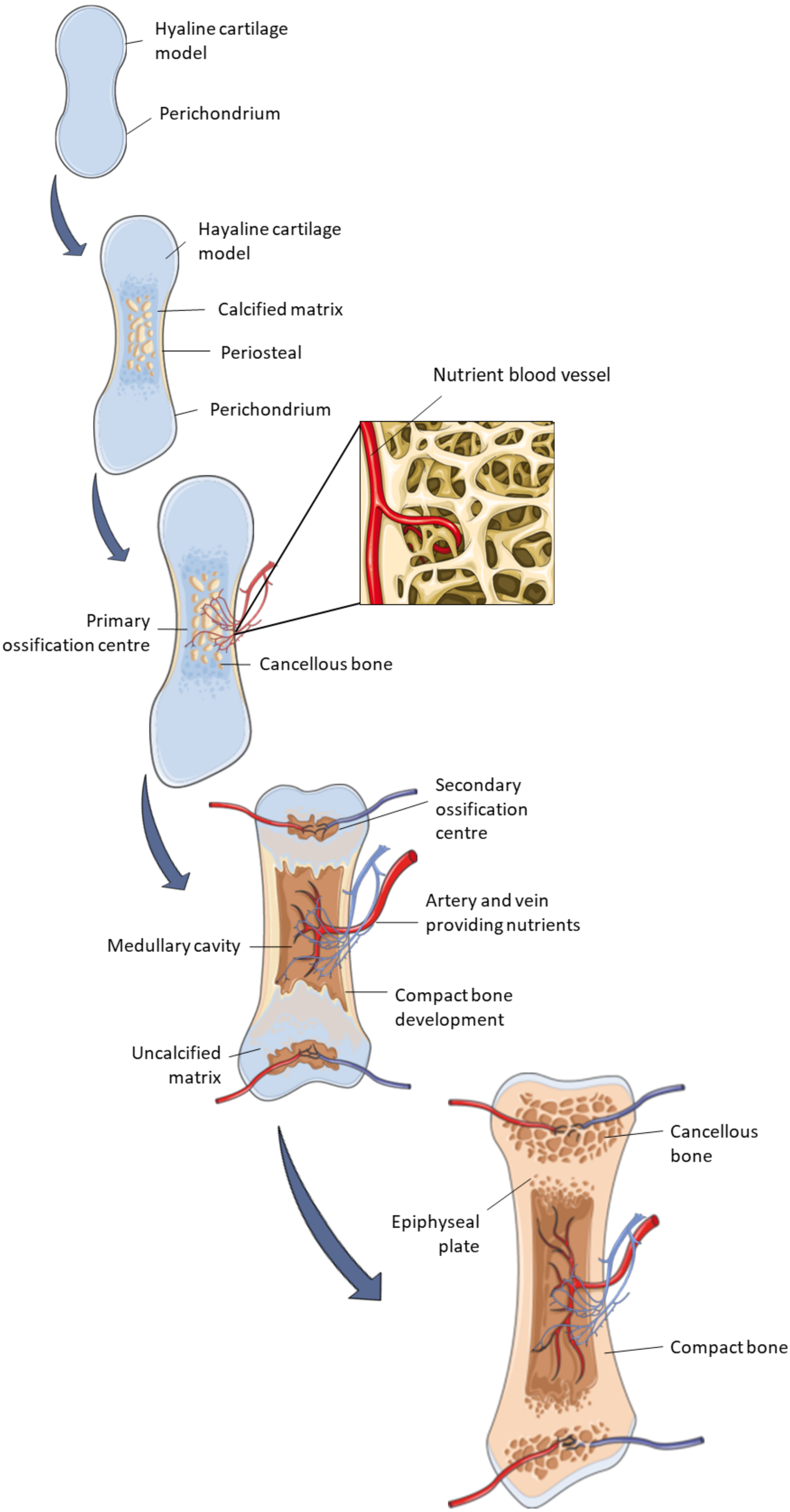

Hyaline cartilage model
Perichondrium
Hayaline cartilage model
Calcified matrix
Periosteal
Perichondrium
Nutrient blood vessel
Primary ossification centre
Cancellous bone
Secondary ossification centre
Artery and vein providing nutrients
Medullary cavity
Compact bone development
Uncalcified matrix
Cancellous bone
Epiphyseal plate
Compact bone

Figure 1.3: Skeletal development by endochondral ossification. Endochondral ossification develops mature bone from an intermediate hyaline cartilage template. Initially, mesenchymal cells differentiate into chondrocytes. Chondrocytes at the centre of the cartilage template proliferate and secrete extracellular matrix forming the cartilage model of the entire skeleton. Chondrocytes then undergo hypertrophy that increases the collagen and fibronectin content of the secreted matrix. Chondrocyte hypertrophy induces calcification. Decreased nutrient availability due to extracellular matrix calcification results in chondrocyte apoptosis, allowing blood vessels carrying osteoblasts to enter the cartilage model. Infiltrating blood vessels and osteoblasts aid the periosteum formation, a collar of compact bone that forms around the midsection (diaphysis) of long bones. The cartilage at the centre of the diaphysis degrades allowing osteoclasts to enter. Disintegrating cartilage is replaced with cancellous bone by infiltrating osteoblasts forming the primary ossification centre. Ossification continues from the primary centre to the ends of long bones. Once cancellous bone is formed in the diaphysis, osteoblasts degrade the new bone and form the medullary cavity. The medullary cavity is the central bone cavity where bone marrow and/or adipose tissue is stored[32]. After birth, secondary ossification occurs in the ends of long bones (epiphyses) where cancellous bone is retained rather than degraded by osteoclasts. In the epiphysis, chondrocytes differentiate in their own secondary ossification centre rather than utilising the primary ossification centre in the diaphysis. Secondary ossification results in the hyaline cartilage being completely replaced by bone[32].

Bone modelling, where bone resorption and formation occurs on different surfaces of existing bone e.g. as bone increases in length, occurs from birth to adulthood. Bone modelling is responsible for increased skeletal mass. The ossification process continues from the primary ossification centre to the ends of the newly formed bones. Bone ossification continues into early adulthood mid-twenties (~25 y)[32]. After bone mass has peaked in adulthood (~25 y), bone mass and structure are maintained by the bone remodelling process. Bone remodelling replaces bone with new tissue involving the coupling of bone formation and bone resorption and consists of five phases; differentiation of osteoclasts that digest mineral matrix, reversal signifying the end of resorption, osteoblast synthesis of new bone and quiescence where osteoblasts differentiate into bone lining cells on the newly formed surface of bone.

The bone ossification process is tightly regulated by numerous growth factors, transcription factors and proteases. BMP cytokines have a crucial role in inducing bone and cartilage formation plus osteoblast and chondrocyte differentiation. BMP regulates several transcription factors involved in bone ossification including RUNX2, OSX and human SRY-box transcription factor 9 (SOX9), which are essential for osteoblastogenesis and chondrogenesis. BMP signalling deregulation is associated with increased or decreased bone density leading to diseases such as heterotopic ossification and osteoporosis. Tight regulation of bone homeostasis is vital in preventing the musculoskeletal disease development.

2.1 BONE MINERAL HOMEOSTASIS

Bone is an endocrine organ that regulates systemic biological functions through peptide and steroid hormone secretion. Bone provides mechanical support to soft tissue plus leverage for muscle action, protection for the nervous system, maintaining a constant ionic environment in extracellular fluids and housing haematopoiesis in the marrow. To maintain endocrine function, bone must undergo constant remodelling by osteoclasts and osteoblasts. Osteoclast resorption and osteoblast formation imbalance alters the bone mineral density, which can be detrimental to skeletal strength. In adults, bone conformation is maintained through a balance in osteoclast and osteoblast activity. An imbalance of excessive resorption by osteoclasts or decreased osteoblast formation results in bone mass decreases. Bones can become weak and brittle over time leading to increased fracture risk. The imbalance of continual bone remodelling is linked to the pathophysiology of osteoporosis due to excess osteoclast activity or decreased osteoblast activity.

Regulatory factors and core gene networks determine the bone structure and homeostasis. Growth factors, e.g. insulin like growth factors (IGFs) and transforming growth factor β (TGFβ), stimulate osteoblast activity and bone formation. Humoral factors, e.g. parathyroid hormone (PTH) and prostaglandin E increase osteoclast activity and bone resorption. Skeletal hormone and cytokine production, such as fibroblast growth factor 23 (FGF23) and osteocalcin, regulate homeostatic functions including energy metabolism, male fertility, blood cell production and calcium-phosphate metabolism. Some studies have acknowledged bone as an endocrine organ[33–35]. Recognising bone as an endocrine organ has led to new areas of bone research drawing interest in recent years[36–41].

The skeleton acts as a mineral reservoir for calcium and phosphorus. Calcium and phosphate homeostasis are critical for human physiology and skeletal mineralisation. Calcium is the most abundant mineral in the human body with ~99% being stored in bone. Calcium ions are essential for bone mineralisation, heart rate regulation, blood coagulation, smooth and skeletal muscle contraction and nerve impulse conduction. The ionised form of calcium (1.1–1.3 mmol/l) is the active fraction and its measurement is not routine in many laboratories. Most hospital laboratories measure total serum calcium with plasma concentration ranging 2.2–2.6 mmol/L[42].

To maintain calcium homeostasis, adult calcium consumption of 1,000-1,300 mg is required per day from dietary sources, e.g. dairy products (Table 1.2). Calcium homeostasis is primarily dependent on interactions between PTH, 1,25-dihydroxyvitamin D (1,25$(OH)_2$D), which is the active form of vitamin D, and calcitonin under negative feedback control[42].

Table 1.2: Calcium Dietary Reference Intakes published by the Institute of Medicine 2011. Values correspond to recommended daily allowance in children and adults (*infant values correspond to Adequate Intake)[43]

Age and Sex	**Recommended Daily Allowance (mg)**
Infants	
0-6 months	200 *
6-12 months	260*
Children	
1-3 y	500
4-8 y	800
Females	
9-18	1300
19-50	1000
>51	1200
Male	
9-18	1300
19-70	1000
>70	1200

In hypocalcaemic conditions where the adjusted calcium concentration in the blood decreases outside of the lower limit (2.2 nmol/L), homeostasis is regulated in a series of anti hypocalcaemic events. The calcium-sensing receptor (CASR), a transmembrane G protein-coupled receptor, senses any fluctuation in ionised calcium and stimulates the synthesis and release of PTH into circulation when in a hypocalcaemic state[42]. PTH is secreted from chief cells in the parathyroid gland. PTH maintains ionised calcium and phosphate concentrations by triggering specific receptor-mediated responses. PTH increases osteoclast activity in the bone by binding to the parathyroid receptor 1 (PTH1R) on osteoblasts (Figure 1.4). PTH binding to osteoblasts triggers the release of receptor activator of nuclear factor kappa-B ligand (RANKL) and macrophage colony-stimulating factor 1 (CSF1), which stimulates osteoclastogenesis. Osteoclast activity induces bone resorption causing calcium and phosphate to be released into circulation from the bone tissue (Figure 1.4). RANKL is expressed by osteoblast and osteocytes. RANKL forms a heterodimer with the RANK receptor found on osteoclasts. RANKL binds to RANK to stimulate osteoclastogenesis, bone remodelling and calcium homeostasis (Figure 1.4). The RANK/RANKL activation stimulates the downstream intracellular signalling pathways MAPK, Nuclear factor kappa B subunit 1 (NFKB1), and PI_3K, subsequently promoting osteoclastogenesis and bone resorption. RANK/RANKL interactions are also regulated by cytokines (interleukin 1, 11 and TNF-a), plus $1,25(OH)_2D$, PTH and prostaglandin, which stimulates RANKL membrane expression causing increased bone resorption.

Additionally, RANK secreted from osteoclasts has been shown to promote bone formation by stimulating RANKL reverse signalling and RUNX2 activation[44]. OPG is secreted by osteoblasts and is a decoy receptor competing with the RANK receptor for RANKL binding. OPG/RANKL binding inhibits osteoclast formation and bone resorption by preventing RANK/RANKL binding (Figure 1.4). OPG/RANKL is an indicator of bone resorption and formation and therefore an indicator of bone health.

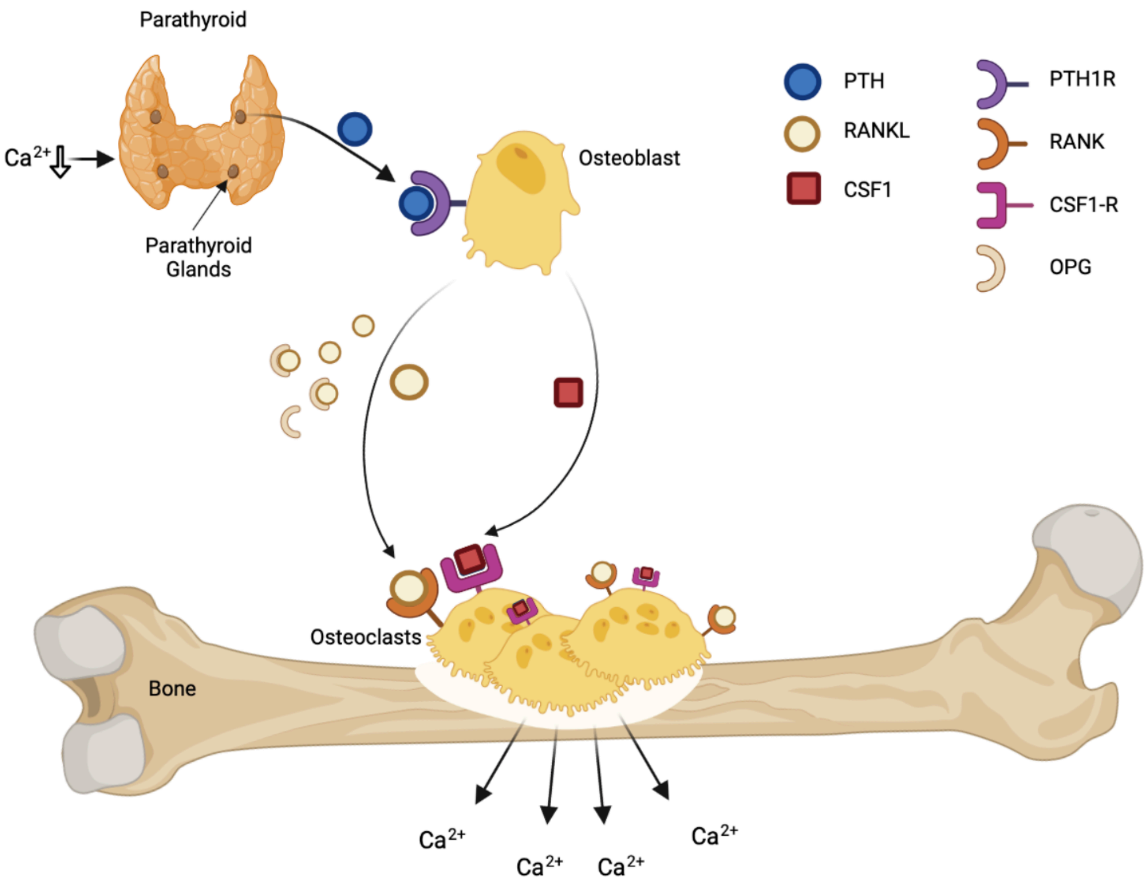


Figure 1.4: PTH stimulated osteoclastogenesis in hypocalcaemic conditions. Hypocalcaemia stimulates PTH secretion from parathyroid glands that binds to PTH1R receptors on osteoblasts. PTH/PTH1R stimulates osteoblasts to produce RANKL and CSF1. RANKL and CSF1 bind to RANK and CSF1-R osteoclasts receptors, respectively, which triggers osteoclast activity. Osteoclast activity increases bone resorption, which elevates serum calcium concentration to counteract hypercalcemia. OPG acts as a decoy receptor that competes with RANK for RANKL binding. OPG/RANKL binding inhibits osteoclastogenesis and bone resorption by preventing RANK/RANKL binding.

Studies have shown that mice with decreased RANKL expression have increased bone mass, osteopetrosis and a lack of osteoclasts[45–47]. Mice with reduced RANKL presented with reduced growth rate and reduced bone resorption. Mice with decreased OPG expression have been shown to present with early onset osteoporosis[48]. The RANKL, RANK and OPG interactions are therefore vial for bone homeostasis. RANK mutations in humans are associated with familial Paget's disease of bone and skeletal hyperphosphatasia[49,50]. OPG mutations are associated with hyperphosphatemia and osteoporotic fractures[51,52].

In hypocalcaemic conditions, PTH restores circulating ionised calcium concentrations by stimulating calcium reabsorption and inhibition of phosphate reabsorption in the kidneys. In the kidneys, PTH stimulates $1,25(OH)_2D$ production, which subsequently increases dietary calcium intestinal absorption (Figure 1.5). An endocrine feedback loop on the CASR decreases PTH secretion once ionised calcium concentrations normalise[53].

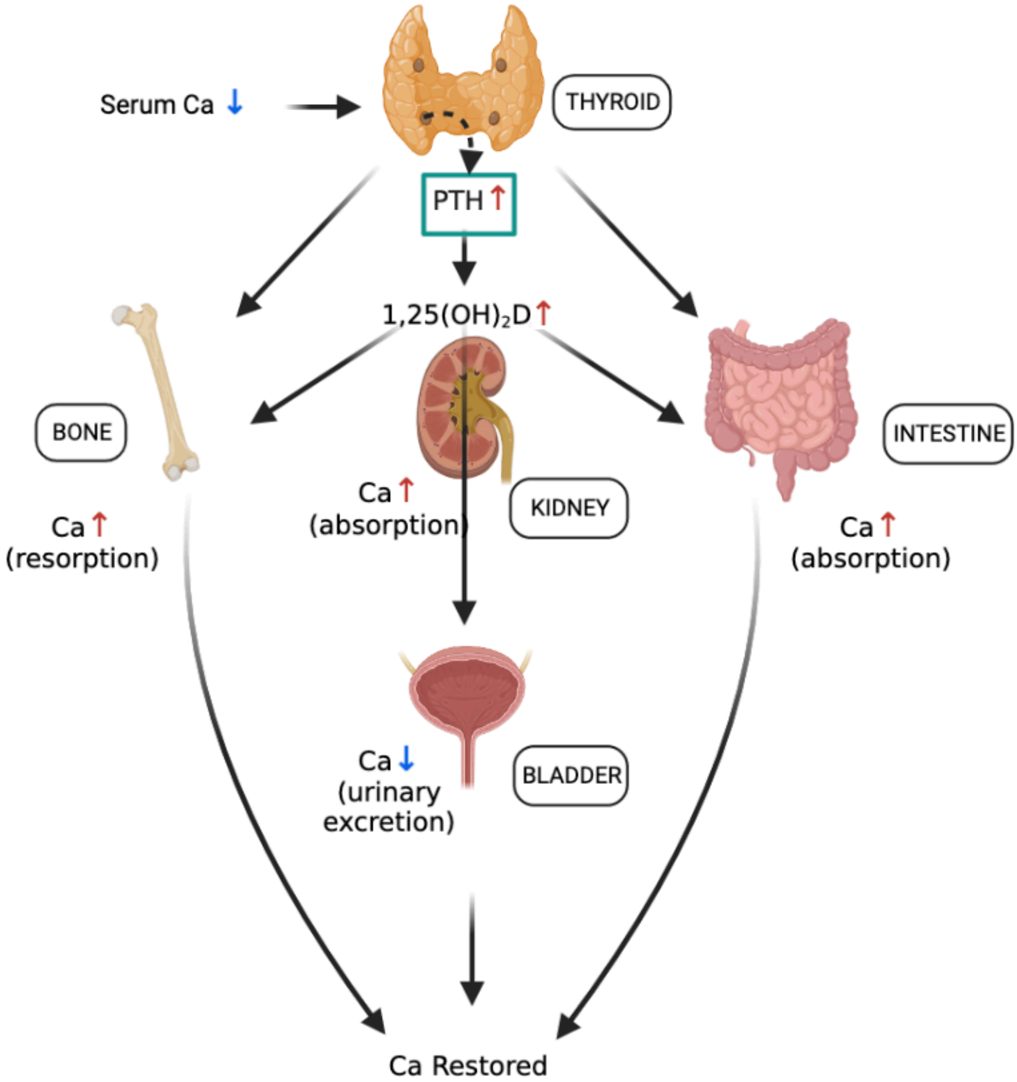


Figure 1.5: Calcium homeostasis in hypocalcaemic conditions. Decreased serum calcium triggers calcium sensing receptors in parathyroid cells to increase PTH production. PTH stimulates bone resorption elevating serum calcium and phosphate concentration. PTH triggers $1,25(OH)_2D$ hydroxylation in the kidneys, which stimulate calcium and phosphate absorption in the intestines. PTH stimulates calcium reabsorption in the kidneys and inhibiting phosphate reabsorption. The reduction in urinary calcium excretion aids restoration of serum calcium[54].

In hypercalcaemic conditions, calcitonin is synthesised and secreted from C-cells in the thyroid gland by parafollicular cells and acts to reduce circulating calcium concentrations. Calcitonin counteracts hypercalcaemic conditions by reducing osteoclast activity and bone resorption while suppressing renal tubular calcium reabsorption and increasing the calcium excretion into urine[54] (Figure 1.6). Through inhibiting bone resorption, calcitonin has a therapeutic role in the treatment of Paget's Disease of Bone[55].

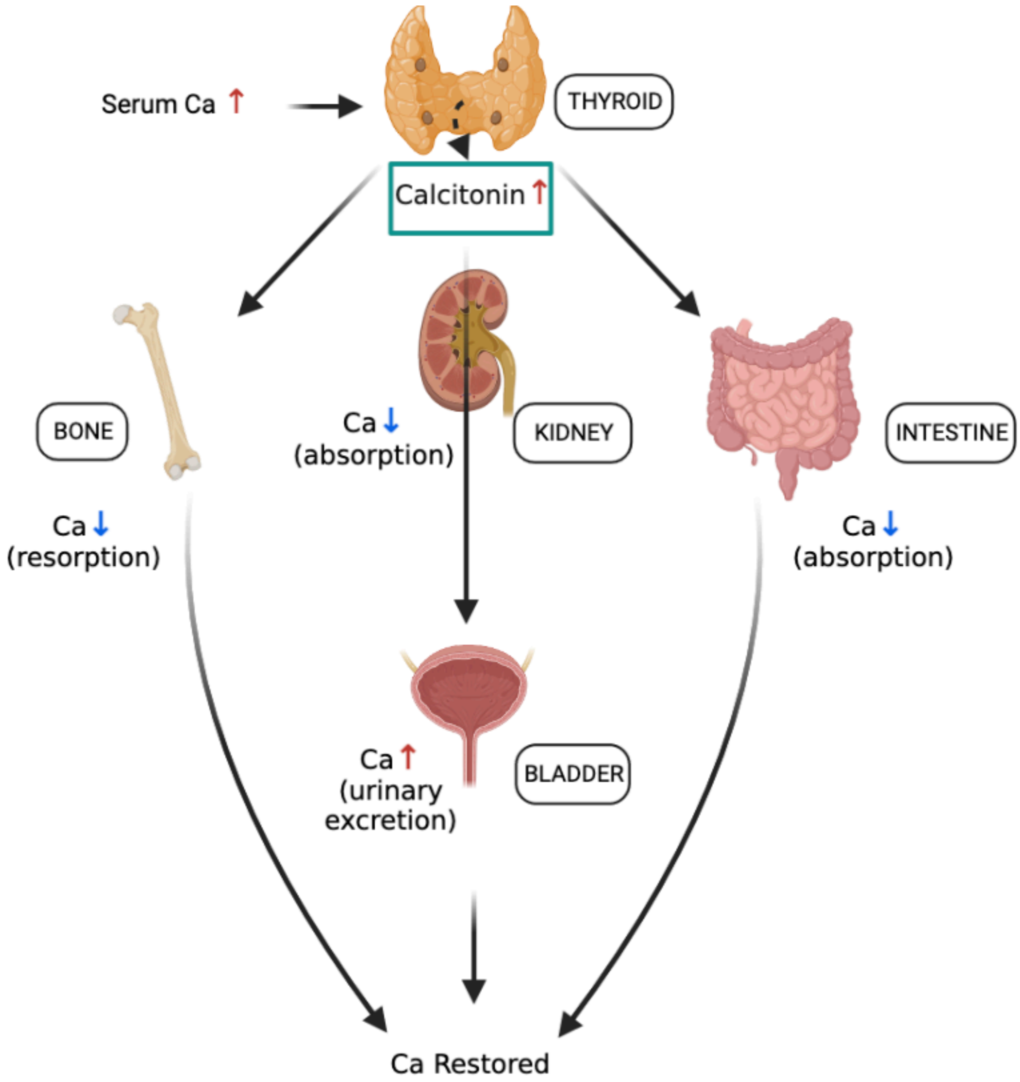


Figure 1.6: Calcium homeostasis in hypercalcaemic conditions. Increased serum calcium triggers C-cells in the thyroid gland to increase calcitonin production. Calcitonin inhibits bone resorption reducing serum calcium and phosphate concentration. Calcitonin also inhibits calcium reabsorption in the kidneys and intestines. The lack of calcium resorption and absorption results in an increase in urinary calcium excretion that aids restoration of serum calcium[54].

Phosphate plays an important role in bone formation, cell signalling, energy metabolism, nucleic acid synthesis and urinary buffering. Serum phosphate concentration range is maintained between 0.8-1.4 mmol/L in adults. The bone-kidney axis is the major regulator of phosphate homeostasis and is affected by FGF23 and the membrane-bound co-receptor protein Klotho (KL)[56] (Figure 1.7). In the kidneys, FGF23 binds to fibroblast growth factor receptors (FGFR). The FGFR family of receptors have a low affinity to FGF23 meaning the co-receptor KL is required to allow efficient ligand-receptor binding. Most human tissues express FGFRs while KL is only expressed in the kidney, parathyroid, pituitary and sinoatrial node. Hyperphosphatemia triggers FGF23 activity through the mitogen activated protein kinase (MAPK) pathway in a negative feedback loop to maintain phosphate homeostasis. FGF23 also increases phosphate excretion in urine. Once normal phosphate concentrations have been restored, calcitriol stimulates FGF23 in a negative feedback loop.

Bone has a key role in this phosphate homeostasis cycle. In hypophosphatemia conditions, PTH stimulates osteoclast activity increasing bone resorption causing additional phosphate reserves to be released from the bone to maintain homeostatic balance. This elevated resorption rate is directed by PTH and 1,25$(OH)_2$D activity[54].

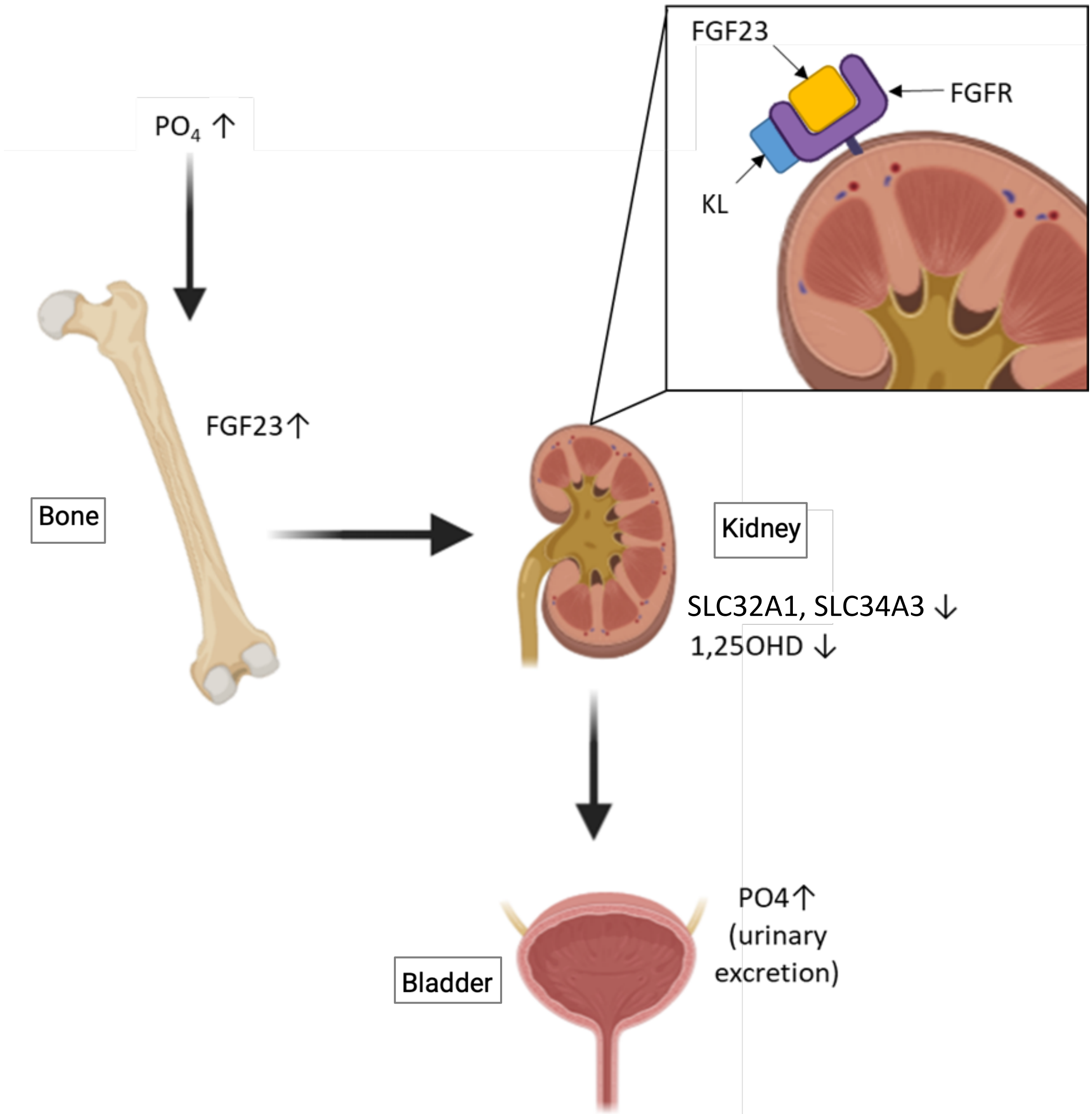


Figure 1.7: Phosphate homeostasis. Hyperphosphatemia stimulates increased FGF23 production from osteoblast lineage cells, particularly osteocytes. In the kidneys, FGF23 binds to FGFR, facilitated by co-receptor KL allowing efficient ligand-receptor binding. Increased FGF23/FGFR/KL binding inhibits renal phosphate reabsorption by inhibiting SLC34A1 and SLC34A3 cotransporters in the kidneys and reducing renal 1,25(OH)$_2$D. SLC34A1 and SLC34A3 co transporters in the renal proximal tubules are responsible for active update of phosphate into epithelial cells. Lack of phosphate reabsorption in the kidneys increases phosphate excretion in urine from the bladder, restoring serum phosphate concentrations.

The importance of calcium and phosphate homeostasis in skeletal development and bone metabolism is well addressed[57–59]. Bone remodelling and calcium homeostasis are highly interconnected as calcium is the main constituent that provides skeletal rigidity. To maintain calcium concentrations in the relatively narrow endogenous reference range, regulatory events in the intestine, kidney, bone and parathyroid gland must work in unison. Vitamin D and its metabolites play a crucial role in the maintanance of calcium and phosphate homeostasis.

1.3 VITAMIN D METABOLISM

The steroid hormone vitamin D plays a key role in classical calciotropic processes including bone and calcium homeostasis and is postulated to contribute to non-classical disorders[60]. Evidence has suggested that low serum 25OHD concentration is associated with several non-skeletal disorders[38,39–45]. Vitamin D deficiency has also been associated with cancer development, several metabolic syndromes, obesity, coronary heart disease, hypertension, diabetes type 1 and 2, tuberculosis severity, multiple sclerosis, dementia, pre-eclampsia and rheumatoid arthritis[69]. Sterol vitamin D is acquired in two forms as vitamin D_2 or vitamin D_3, collectively termed calciferol. Ergocalciferol (D_2) is present in naturally occurring irradiated steroid ergosterol in plants or from fortified food. Cholecalciferol (D_3) is primarily synthesised in humans by photochemical conversion of the intermediate compound pre-vitamin 7-dehydrocholesterol (7DHC) in the skin's epidermis after exposure to ultraviolet B (UVB) radiation (290-320 nm). UVB radiation of 7DHC induces an electrocyclic rearrangement in 7DHC producing pre-vitamin D. Thermal isomerisation of pre-vitamin D in the plasma membrane bilayer induces a hydrogen shift from C19 to C9 producing vitamin D_3. In humans, 80–90% of the body's vitamin D supply is produced by the skin's exposure to UVB radiation and 10–20% is acquired through diet.

Both vitamin D_2 and D_3 are prohormones and must be metabolised into their active forms regardless of source. Two consecutive hydroxylation reactions occur to activate the vitamin D prohormones, first in the liver, then in the kidneys (Figure 1.8). Sterol vitamin D binds to the vitamin D binding protein (DBP) in the blood and is transported to the liver. In the liver, the microsomal vitamin D hydroxylate cytochrome P450 family 2 subfamily R member 1 (CYP2R1) hydroxylates vitamin D_2 and D_3 prohormones into 25-hydroxyvitamin D (25OHD). The vitamin D metabolite 25OHD is the main circulating form of vitamin D. DBP binds to 25OHD, which is transported in the bloodstream to the kidneys for the second hydroxylation step when requiring activation. The majority of 25OHD in circulation is bound to DBP (85-90%) followed by albumin (10-15%) with a small percentage left in circulation as a free form of 25OHD[70]. The second hydroxylation step takes place in the proximal renal tubule where 25OHD is hydroxylated into the active systemic metabolite $1,25(OH)_2D$ by the enzyme cytochrome P450 family 27 subfamily B member 1 (CYP27B1). Once synthesised by CYP27B1 in the kidneys, $1,25(OH)_2D$ is transported in the bloodstream by DBP to target cells in the intestine, kidney and bone. Once $1,25(OH)_2D$ enters the target cell via diffusion, its biological action is mediated by the vitamin D receptor (VDR), a ligand-activated transcription factor. Interaction between $1,25(OH)_2D$ and the VDR prompts $1,25(OH)_2D$ translocation from the cytosol to the nucleus of target cells. A heterodimer is formed in the nucleus between the VDR and the retinoid X receptor (RXR). The VDR/RXR complex binds to specific sequences in the promotor region of target genes termed vitamin D response elements (VDRE). VDRE plus unique binding sites on the surface of VDR and RXR allows recruitment of further multi-protein co-regulatory complexes and basal transcription factors to increase or suppress the rate of gene transcription. The $1,25(OH)_2D$/VDR complex stimulates gene expression in (i) calcium homeostasis, (ii) 1,25$(OH)_2$D regulation and synthesis, iii) suppression of PTH synthesis, iv) immune response regulation and v) cell proliferation suppression[71].

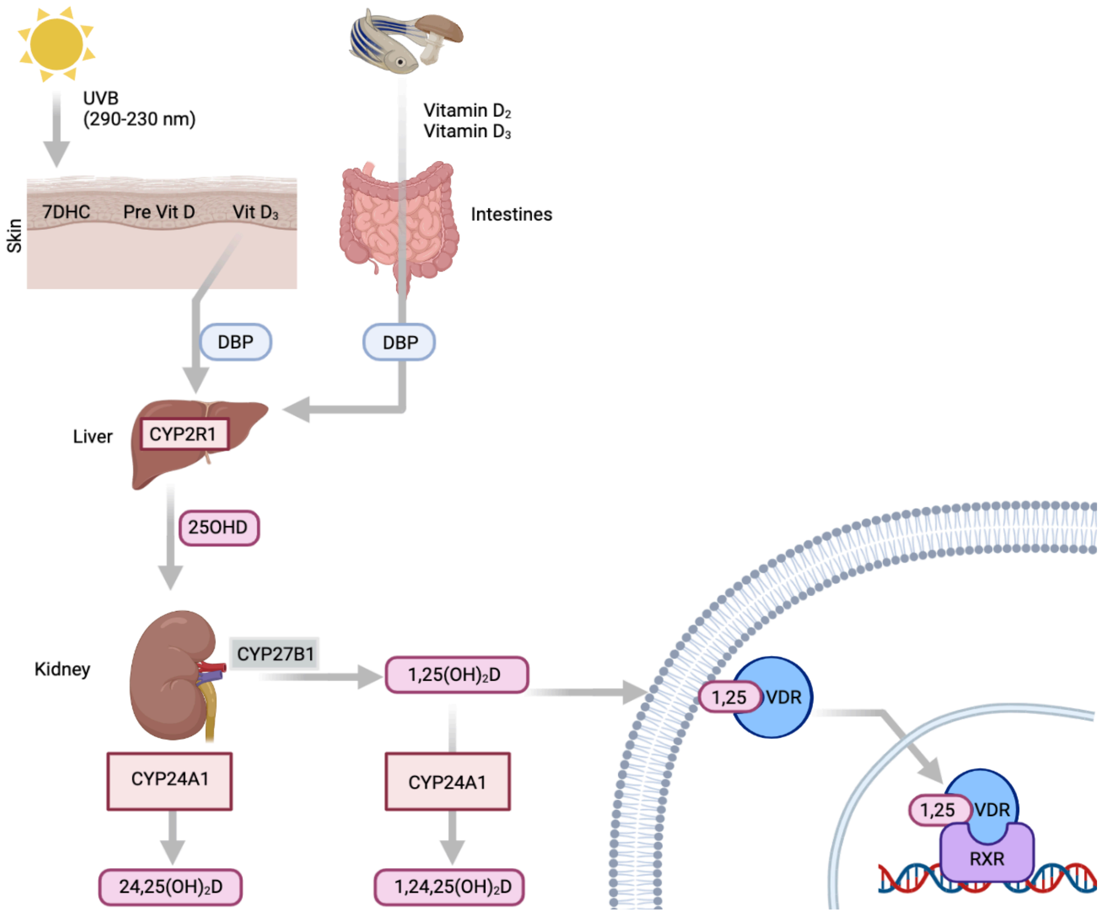


Figure 1.8: Vitamin D metabolism. UVB radiation (280-310nm) is absorbed by the double bond of 7DHC in the skin forming pre-vitamin D_3. Thermal isomerisation in the plasma membrane bilayer causes rearrangement of double bonds in pre-vitamin D_3 producing vitamin D_3. Vitamin D_2 and D_3 is present in food sources such as mushrooms and oily fish. The sterol vitamin D_3 diffuses out of the plasma membrane into the extracellular space. Vitamin D binding protein (DBP) then binds to sterol vitamin D_3 and D_2 and transports it through the circulatory system to the liver to begin hydroxylation into its active form. The main circulating vitamin D metabolite, 25OHD, is produced in the liver through CYP2R1 activity. DBP transports 25OHD to the liver for the second hydroxylation step into either 1,25$(OH)_2$D by CYP27B1 or 24,25-dihydroxycholecalciferol (24,25$(OH)_2$D) by CYP24A1. The metabolites 1,24,25$(OH)_2$D and 24,25$(OH)_2$D are further hydroxylated and excreted in bile. Metabolic activity of 24,25$(OH)_2$D is widely disputed[70,72,73] with recent evidence suggesting that 24,25$(OH)_2$D may have its own physiological actions Including the PTH suppression, fracture healing stimulation and growth plate cartilage regulation[74,75,76]. The biologically active 1,25$(OH)_2$D enters target cells and interacts with the VDR. The VDR/RXR heterodimer is formed in the nucleus allowing 1,25$(OH)_2$D to regulate gene expression[71].

$1\alpha,25(OH)_2D$ exerts its effect with the help of high-affinity VDR through a series of cell-signaling reactions, or as a ligand-activated transcription factor. In 2016 Maestro et al. published that the vitamin D receptor (VDR) has over 1000 target genes and is present in most human tissues[77]. Although the VDR is widely distributed, sites of calcium and phosphate homeostasis regulation have the greatest expression e.g. intestine, bones, kidneys and parathyroid glands. Additionally the VDR is found in fibroblasts, skin keratinocytes, epithelial cells, and numerous immune cells[78]. $1,25(OH)_2D$ initiates non-genomic or genomic effects by binding VDR. The target genes of vitamin D respond in a cell-specific manner (Table 1.3).

Table 1.3: Examples of direct targets of 1,25$(OH)_2$D, their cell type and roles[79].

Target Gene	Cell Type	Function
CYP24A1 (Cytochrome P450 24A1)	Kidney	Responsible for metabolism of 1,25$(OH)_2$D and 25OHD to prevent hypervitaminosis D [80]
PTHLH (Parathyroid hormone-like hormone)	Golgi complex Cytoplasm/ Nucleus	Regulates calcium transport, bone formation and resorption. Activates PLC signalling pathwaysl[81]
CAMP (Cathelicidin antimicrobial protein)	Keratinocytes White blood cells	Role in innate and adaptive immunity[82,83].
Trpv6 (Transient Receptor Potential Vanilloid 6)	Intestinal epithelia membrane	Mediates transcellular calcium transport maintaining calcium homeostasis[82]
CYP27B1 (Cytochrome P450 27B1)	Kidneys Endocrine glands Epithelial cells Lungs Breast Intestine Stomach Bone (osteoblasts and chondrocytes)	Expression of 1a-hydroxylase enzyme, which is responsible for the production of 1,25$(OH)_2$D from 25OHD[84]
Osteocalcin	Osteocytes Cartilages	Glucose metabolism through role in bone-pancreas endocrine loop[85]
FOXP3 (fork head box P3)	T Cells	Regulates activity of the immune system (T cells)[86]
CD14 (Cluster of differentiation)	Monocytes	Lipopolysaccharide (LPS)-binding protein[84]

The activity of $1,25(OH)_2D$ is integral in the endocrine system that maintains calcium and phosphate homeostasis (see chapter 1.2). In hypocalcaemic conditions (adjusted calcium concentration <2.2 nmol/L), $1,25(OH)_2D$ influences bone homeostasis in a series of anti hypocalcaemic events. Low serum calcium inhibits FGF23 and stimulates PTH synthesis. PTH stimulates $1,25(OH)_2D$ synthesis in the kidney by increasing CYP27B1 activity. $1,25(OH)_2D$ and VDR interaction, increases calcium resorption from bone increasing serum calcium concentrations. Once $1,25(OH)_2D$ concentration becomes elevated, $1,25(OH)_2D$ subsequently inhibits CYP27B1 activity creating a negative feedback loop to suppress PTH that maintains calcium and phosphorus homeostasis. Elevated $1,25(OH)_2D$ also stimulates CYP24A1 activity to prevent excessive $1,25(OH)_2D$ production. CYP24A1 activity is inhibited by hypocalcaemia and low PTH concentrations, which allows increased $1,25(OH)_2D$ production (Figure 1.9).

Regulation of $1,25(OH)_2D$ activity is tightly regulated by the mitochondrial enzyme CYP24A1. CYP24A1 activity is stimulated by elevated $1,25(OH)_2D$ serum concentration. Low serum calcium and PTH concentrations inhibit CYP24A1 activity, which aids $1,25(OH)_2D$ regulation. CYP24A1 hydroxylates $1,25(OH)_2D$ and 25OHD into $1,24,25(OH)_2D$ and $24,25(OH)_2D$, respectively to prevent overaccumulation of $1,25(OH)_2D$ and $1,24,25(OH)_2D$ and subsequent vitamin D toxicity. The metabolites $1,24,25(OH)_2D$ and $24,25(OH)_2D$ are further hydroxylated and excreted in bile (Figure 1.9).

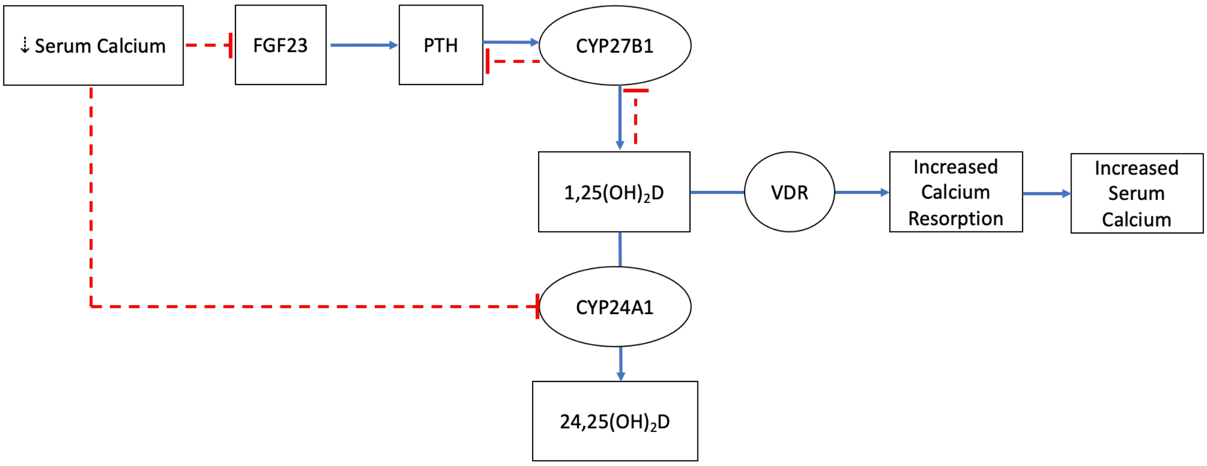


Figure 1.9: 1,25(OH)$_2$D negative feedback loop responsible for calcium homeostasis. Low serum calcium inhibits FGF23 and stimulates PTH synthesis. Elevated PTH increases 1,25(OH)$_2$D synthesis in the kidney through stimulating CYP27B1 activity. 1,25(OH)$_2$D and VDR interaction, increases calcium resorption from bone increasing serum calcium concentrations. Once elevated, 1,25(OH)$_2$D inhibits CYP24B1 activity to suppress PTH secretion. In hypocalcaemic conditions, CYP24A1 activity is inhibited, which allows increased 1,25(OH)$_2$D production. Once elevated, 1,25(OH)$_2$D stimulates CYP24A1 activity to prevent excessive 1,25(OH)$_2$D accumulation by hydroxylating 1,25(OH)$_2$D to 24,25(OH)$_2$D.

Metabolic activity of 24,25$(OH)_2$D is widely disputed[70,72,73] with recent evidence suggesting that 24,25$(OH)_2$D may have its own physiological actions including the PTH suppression, fracture healing stimulation and growth plate cartilage regulation[74,75,76]. Studies have indicated that 24,25$(OH)_2$D is elevated 10-12 days post fracture suggesting that the metabolite is required for fracture healing[75]. Additionally, CYP24A1-null mice who are unable to produce 24,25$(OH)_2$D have been show to have suboptimal fracture repair during the endochondral ossification phase as the absence of 24,25$(OH)_2$D impairs chondrocyte maturation[87]. Fracture repair was improved upon 24,25$(OH)_2$D treatment. The binding of 24,25$(OH)_2$D to the transmembrane protein FAM57B2 in the fracture callus induces lactosylceramide production, aiding callus formation and fracture healing. Questions remain unanswered as to whether 24,25$(OH)_2$D is optimally synthesised at the fracture site or surrounding circulation[74].

1.4 VITAMIN D REFERENCE RANGES AND TOXICITY

Vitamin D is a hormone critical to optimal musculoskeletal health. Evidence has suggested that low serum 25OHD concentration is associated with several non-skeletal disorders[61,62,62–68]. Vitamin D deficiency has also been associated with cancer development, several metabolic syndromes, obesity, coronary heart disease, hypertension, diabetes type 1 and 2, tuberculosis severity, multiple sclerosis, dementia, pre-eclampsia and rheumatoid arthritis[69]. Conditions linked to vitamin D deficiency are of worldwide major public health concern. Active vitamin D, 1,25$(OH)_2$D predominantly functions to regulate calcium homeostasis, e.g., promoting calcium absorption from the gut and enabling osteoid tissue mineralisation in the bone. In children, vitamin D deficiency can lead to rickets, resulting from defective mineralisation of the growth plate, whilst osteomalacia, resulting from defective mineralisation of the osteoid, has been observed in both adults and children with rickets. Many factors can influence vitamin D status including skin pigmentation, sun exposure and diet, the latter influencing ergocalciferol and cholecalciferol intake. Factors influencing vitamin D status particularly affects certain groups described as being at high risk of vitamin D deficiency, e.g., pregnant women, infants and children (>5 years), the elderly (>65 years) and individuals with high melanin pigmentation such as those with sub-Saharan African ancestry. Previous trials have reported inconsistencies in establishing adequate nutritional requirements to counteract extra-skeletal diseases associated with vitamin D deficiency[88–93]. While there is controversy as to the precise vitamin D requirements and the effect of 25OHD status on human health, vitamin D deficiency prevention is regarded as a public health priority[94–105].

Serum concentrations of 25OHD serve as a clinical indicator of vitamin D status. In 2011 it was estimated by the Institute of Medicine (IOM) that 25OHD serum concentration should exceed 50 nmol/L in healthy adults to maintain bone density, calcium absorption and reduce the risk of osteomalacia and rickets[43]. To achieve this concentration a required recommended daily allowance (RDA) of 400 IU for <12 months old, 600 IU for 1-70 years old and 800 IU for individuals aged >71 years is advised for both males and females. Pregnant and breastfeeding women are recommended 600 IU[43].

The IOM report defined vitamin D deficiency as a 25OHD serum concentration <30 nmol/L and vitamin D insufficiency as 30-50 nmol/L[43]. Reference ranges of 25OHD relating to vitamin D sufficiency stated in the IOM report have been supported by further organisations worldwide[106]. Vitamin D insufficiency results in a less severe phenotype than vitamin D deficiency that can cause secondary hyperparathyroidism, muscle weakness and high fracture rates in older patients[69]. In 2010, more than 50% of the UK adult population had vitamin D insufficiency and 16% had severe deficiency during winter and spring[107]. Evidence from the IOM report identified a positive correlation between vitamin D supplementation and clinical benefit, e.g., symptom relief, improved functionality or survival due to treatment intervention. Although vitamin D supplementation is effective at preventing vitamin D deficiency, it was concluded by the IOM that at extremes of both high and low 25OHD concentrations, patients were at risk of varying conditions associated with hypovitaminosis, e.g., rickets and hypervitaminosis, e.g., hypercalcemia[43]. Further research is required to determine a universal reference range for 25OHD in order to avoid over and under prescription of medicinal supplementation and the potential for vitamin D toxicity/deficiency.

There is a lack of consensus regarding vitamin D status between countries resulting in a range of terminology and reference values across current literature. An inconsistency in reference thresholds and use of terminology such as 'deficiency' and 'insufficiency' has caused difficulty in comparing reported prevalence of vitamin D deficiency. Previous inconsistencies with laboratory vitamin D metabolite measurements, including 25OHD, are improving with the implementation of the gold standard reference assay, liquid chromatography tandem mass spectrometry (LC-MS/MS). LC-MS/MS is now used by the UK National Laboratory, US Centres for Disease Control and Prevention (CDC) plus the US National Institute of Health, which established the Vitamin D Standardisation Program (VDSP)[108]. The standardisation of methods used for vitamin D status determination is an essential step forward in developing universal thresholds for vitamin D deficiency, insufficiency and toxicity.

Significant variation in dietary intake and vitamin D supplementation is observed between European countries[43]. Excessive vitamin D intake (> 4,000 IU per day) results in elevated circulating 25OHD triggering increased intestinal absorption and bone resorption of calcium leading to hypercalcemia. The development of hypercalcemia can present as vomiting, loss of appetite, dehydration, fatigue and renal stone development. The upper intake level (UL) for vitamin D supplementation set by the IOM was determined in different age groups[43] (Table 1.4).

Table 1.4: Vitamin D UL per age group as set by the IOM[43]

Age	**Upper Intake Level [µg/day] (IU)**
<6 months	25 (1,000)
7-12 months	37.5 (1,500)
1-3 years	62.5 (2,500)
4-8 years	75 (3,000)
>9 years	100 (4,000)

Vitamin D toxicity can occur in response to excessive vitamin D supplementation. Hypervitaminosis occurs when an increased amount of vitamin D metabolites, e.g., 1,25(OH)$_2$D and 25OHD, bind to the VDR in the nucleus of target cells causing exaggerated gene expression. Increased 1,25(OH)$_2$D concentrations saturate the DBP, which becomes compromised allowing alternate metabolites such as 25OHD to enter the cell nucleus and stimulate transcription[109]. Due to lipid solubility, supplement-derived vitamin D can be stored in the liver, muscles and fat tissues. Vitamin D stored in adipose tissue is available for release back into the plasma meaning that the use of daily, weekly and potentially monthly vitamin D supplementation may not be required to replenish vitamin D stores and could lead to vitamin D toxicity, Stored vitamin D can be progressively released into the bloodstream potentially inducing symptoms of hypercalcemia long after the removal of supplementation. Serum vitamin D assays do not account for stored vitamin D meaning that in some cases of hypercalcemia, symptoms may persist long after the removal of supplementation because of vitamin D's lipophilic properties resulting in absorption and storage in adipose tissue.

Previous clinical trial meta-analyses investigating vitamin D supplementation have shown that dosage and route of administration, e.g. orally or intramuscular, can vary greatly depending on the study. A recent meta-analysis in children aged 2-18 years presented results from 26 trials with interventions as little as 2.5 ug (100 IU) to 100 ug (4,000 IU) per day given as fortified food or supplements of bolus injection[110]. Although this meta-analysis concluded that vitamin D supplementation correlated with increased serum 25OHD in a dose-dependent manner, heterogeneity between the 26 trials studied was 99.9%. A random-effects model was used with I^2 statistics to assess heterogeneity, were <25% indicates homogeneity and >75% indicated high heterogeneity[110]. The analysis concluded that the greatest response in increased serum 25OHD concentrations per 100 IU vitamin D/d supplementation was seen in trials with a mean baseline of <30 nmol/L 25OHD and administered as fortified foods. Although there is large variation in administration and supplementation intake across current literature, the response to supplementation is greatly affected by baseline 25OHD concentration, delivery and dosage and less affected by latitude, age and sex of patients. The mean lethal dose (LD50) of vitamin D in humans has been estimated as 12 mg/kg (840,000 IU/kg)[109]. Previous reports summarising adult supplementation indicate that hypervitaminosis D develops once a daily intake of 40,000 IU is reached causing vitamin D toxicity by increasing serum 25OHD to above 200 nmol/L[111].

Careful dosing is required when prescribing vitamin D supplementation. Due to an increased vitamin D deficiency awareness, there has been an increase in both prescribed and non-prescription vitamin D supplementation in recent years. Supplementation effects need to be monitored to prevent exogenous hypervitaminosis D and hypercalcemia development. Although rare, vitamin D toxicity should be considered as a differential diagnosis in patients presenting with hypercalcemia.

1.5 CYTOCHROME P450 ENZYMES

The cytochrome P450 (CYP) superfamily of hemeproteins are involved in a highly diverse range of chemical reactions including catalysing organic substance oxidation, drug metabolism and hormone synthesis. CYP enzyme heme-containing monooxygenases are found across all kingdoms of life. In mammals, CYP enzymes are primarily found in hepatic microsomes where they are responsible for steroid, fatty acid and xenobiotic oxidation. CYP enzymes are also vital for hormone synthesis, cholesterol synthesis and have a key role in vitamin D metabolism. CYP enzymes were originally identified in 1954 by Klingenberg *et al*. while researching steroid hormone metabolism[112]. A decade later the function and significance of CYP enzymes was determined as catalysts in steroid hormone synthesis and drug metabolism[113]. The catalytic role of CYP in drug and steroid hydroxylation reactions was later corroborated by Cooper *et al*. in 1965[114]. Since initial identification of the CYP enzyme superfamily, several CYP proteins have been identified highlighting CYP enzyme superfamily's diverse range of roles in chemical activation, deactivation and tumorigenesis.

In humans, CYP proteins are functionally active in the mitochondrial inner membrane or the endoplasmic reticulum. In current human genome annotations there are 18 CYP genetic families based on similarities in amino acid sequence, coding for 57 biologically active CYP proteins[115]. CYP enzymes play important roles in protein ligand synthesis and/or degradation in pathways regulating differentiation, homeostasis, apoptosis, growth and neuroendocrine functions. CYP enzymes oxidise target substrates via their heme group using NADH or NADPH derived protons and electrons received from redox partners[116] (Figure 1.10).

In the mitochondrion, electrons from NADH or NADPH are donated to the membrane-bound flavoprotein ferredoxin reductase (FDXR). The electrons are then transported from the FDXR by the soluble iron-sulphur protein ferredoxin (FDX), which donates the electrons to CYP enzymes. The CYP enzyme heme group receives electrons from FDX, which are donated to molecular oxygen as the terminal electron acceptor to catalyse the reaction[116].

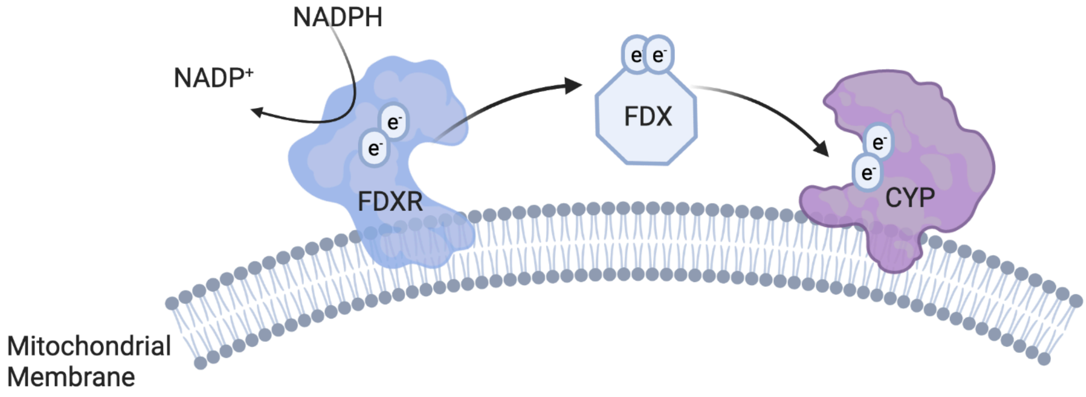


Figure 1.10: CYP electron transport chain in the mitochondria. Electrons are received from NADPH by the membrane bound protein FDXR. The electrons are donated to FDX that transports the electrons to the heme group in mitochondrial CYP proteins to facilitate target substrate hydroxylation[116].

During vitamin D metabolism three enzymes in the cytochrome P450 family, CYP27A1, CYP27B1 and CYP24A1, are primarily responsible for 1,25$(OH)_2$D regulation. Calcium homeostasis requires CYP protein regulation to prevent hypo- and hypercalcemia. Because CYP genes are critical to normal human health, mutations in these genes can affect enzyme function and cause CYP associated disease[116].

1.6 CYP24A1 ACTIVITY IN VITAMIN D METABOLISM

The mitochondrial inner membrane enzyme, CYP24A1, is a multifunctional enzyme that catalyses several pathways in vitamin D metabolism. By a classical endocrine negative feedback loop, CYP24A1 is induced in target tissues by 1,25$(OH)_2$D. In the kidneys, 25OHD is hydroxylated to become the active vitamin D metabolite 1,25$(OH)_2$D. 1,25$(OH)_2$D stimulation produces CYP24A1 upregulation, which offers an alternate 25OHD hydroxylation to prevent excess 1,25$(OH)_2$D production. This alternate 25OHD hydroxylation facilitated by CYP24A1 produces 24,25$(OH)_2$D. CYP24A1 also hydroxylates 1,25$(OH)_2$D into the short-lived metabolite 1,24,25$(OH)_2$D in response to elevated 1,25$(OH)_2$D concentration. Both 24,25OHD and 1,24,25$(OH)_2$D catabolic products are subjected to further hydroxylation and excretion, facilitated by CYP24A1. Hydroxylation of 1,25$(OH)_2$D by CYP24A1 aids prevention of vitamin D toxicity (Figure 1.11).

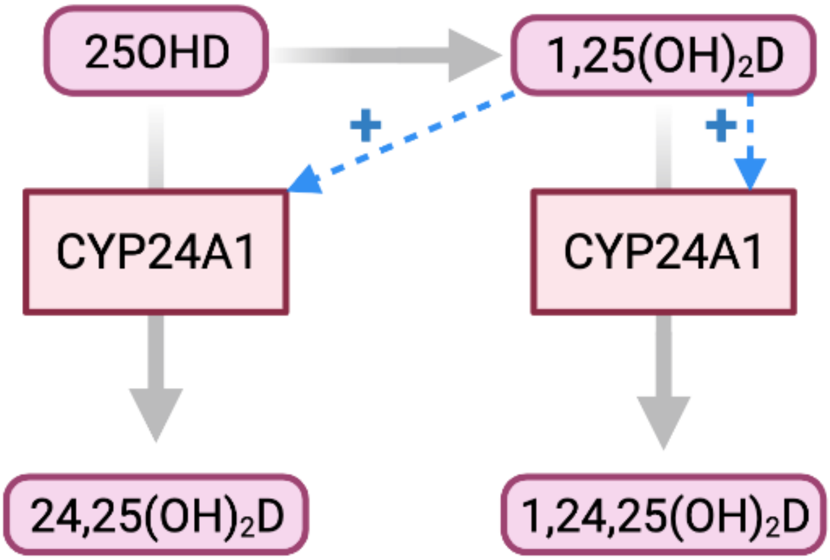


Figure 1.11: CYP24A1 regulation of 1,25(OH)$_2$D concentration. Increased 1,25(OH)$_2$D triggers CYP24A1 activity. Blue arrows indicating increased CYP24A1 stimulation from 1,25(OH)$_2$D. CYP24A1 hydroxylates 1,25(OH)$_2$D into 1,24,25(OH)$_2$D and 25OHD into 24,25(OH)$_2$D to prevent hypervitaminosis D.

CYP24A1 catalyses several hydroxylation reactions on the side chains of both 25OHD and 1,25$(OH)_2$D at carbon C-24 and C-23 (Figure 1.12). The CYP24A1 24-oxidation pathway from 1,25$(OH)_2$D is a five-step enzymatic pathway producing the biliary catabolite, calcitroic acid. The 23-hydroxylation pathway of 1,25$(OH)_2$D is also facilitated by CYP24A1 and results in 1,25$(OH)_2$D3-26,23-lactone. The preference of 23- or 24-hydroxylation is species dependent with human CYP24A1 predominantly executing 24-hydroxylation, rat displaying minimal 23-oxidation and guinea pig utilising 23-oxidation primarily[117–119]. The functional significance of 23- and 24-hydroxylation pathways and the degree to which each species performs them is unknown.

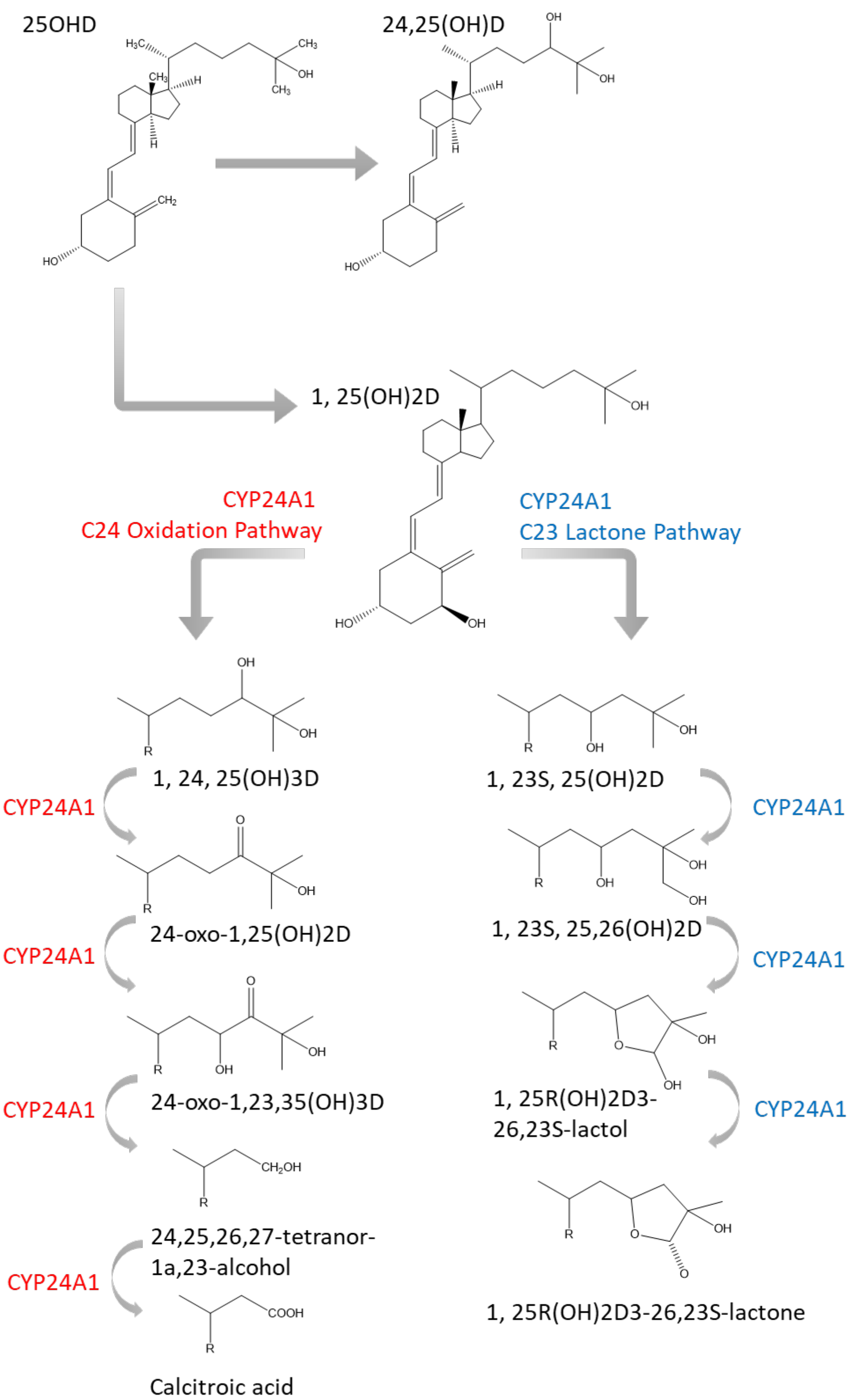


Figure 1.12: CYP24A1 hydroxylation pathways. CYP24A1 enzymatic hydroxylation of 1,25$(OH)_2$D via C-24 oxidation (in red) and C-23 hydroxylation (in blue). Water soluble end products from each pathway (calcitroic acid and 1,25$(OH)_2D_3$-26,23-lactone) are excreted in bile to prevent hypervitaminosis D.

CYP24A1 is expressed in all target cells containing the VDR, e.g., bone, kidney, small and large intestine. It was originally hypothesised that the CYP24A1 enzyme is triggered by 1,25$(OH)_2$D, which in turn restricts CYP24A1 activity in target cells as a negative feedback loop preventing excessive VDR pathway initiation. In 1999, work performed with the CYP24A1-null mouse supported this hypothesis confirming that the primary role of CYP24A1 in normal physiology is to catabolise both 25OHD and 1,25$(OH)_2$D to prevent toxic 1,25$(OH)_2$D concentrations[120]. *CYP24A1* null mice (*CYP24A1*$^{-/-}$) showed that in the absence of CYP24A1 protein, 1,25$(OH)_2$D clearance was inhibited and 1,25$(OH)_2$D half-life increased ten-fold demonstrating the role of CYP24A1 in catabolising 25OHD and/or degrading 1,25$(OH)_2$D in target cells to prevent its 1,25$(OH)_2$D biological activity[120]. Of the engineered *CYP24A1* deficient mice, 50% died before three weeks of age due to hypercalcemia resulting from lack of 1,25$(OH)_2$D clearance and decreased 1,24,25$(OH)_2$D and 24,25$(OH)_2$D production causing hypervitaminosis D. The 50% postnatal survival rate suggested that the surviving mice were able to utilise a potential as yet unknown alternate pathway for the 1,25$(OH)_2$D regulation[120]. Pharmacokinetic studies of labelled 1,25$(OH)_2$D in surviving *CYP24A1*$^{-/-}$ mice showed impaired 1,25$(OH)_2$D clearance plus an absence of 1,25$(OH)_2$D hydroxylated metabolites including calcitroic acid and 1,25$(OH)_2D_3$-26,23-lactone[120]. Lack of 1,25$(OH)_2$D metabolites observed in surviving CYP24A1$^{-/-}$ mice supported the hypothesis that CYP24A1 is responsible for both C-24 and C-23 side chain pathways. The *CYP24A1* -/- mouse model suggested an ability to adapt to CYP24A1 absence by utilising an as yet unidentified alternate catabolic route providing a reduced but sufficient active 1,25$(OH)_2$D synthesis to prevent potentially fatal hypercalcemia *in vivo*[120].

1.7 CYP24A1 MEDIATED HYPERCALCEMIA AND DEVELOPMENT OF IDIOPATHIC INFANTILE HYPERCALCEMIA (IIH)/ INFANTILE HYPERCALCEMIA TYPE 1 (HCINF1)

Hypercalcemia affects 1-2% of the general population with primary hyperparathyroidism being the main cause[121]. Primary hyperparathyroidism arises due to parathyroid adenomas (85%), hyperplasia (15%) or parathyroid cancer (<1%) resulting in overactive parathyroid glands and excess PTH secretion[121]. Increased serum PTH increases serum calcium concentration leading to hypercalcemia and other related conditions, e.g., kidney stone formation. In rare cases, vitamin D mediated hypercalcemia can occur due to abnormalities in the vitamin D metabolism pathway.

Hypercalcemia can be a consequence of abnormal calcium homeostasis (total serum calcium >2.6 mmol/L). Hypercalcemia is observed across the lifespan but development at infancy is rare. Many children with mild hypercalcemia remain asymptomatic while severe hypercalcemia presents in infants as vomiting, failure to thrive, colic and in rare cases neonatal death. Adult hypercalcemia presentations as depressed mood, musculoskeletal pain, abdominal pain linked to constipation or peptic ulceration, renal colic and nephrolithiasis. As hypercalcaemia progresses, adults can also experience nausea, vomiting, QT shortening that could cause ventricular fibrillation arrest, and in rare cases lead to a comatose state. Hypercalcemia can be PTH dependent or independent. PTH dependent hypercalcemia is the most common cause of infantile hypercalcemia where one or multiple parathyroid glands secrete excess PTH causing increased calcium within the blood leading to hypercalcemia. In cases of appropriate PTH secretion, where increased serum calcium is not due to excess PTH production, vitamin D metabolites should be investigated as a cause of abnormal calcium handling due to PTH-independent hypercalcemia. Hypercalcemia is a sign of vitamin D toxicity, which leads to a variety of conditions including granulomatous disease. The *CYP24A1* gene has been characterised as a genetic determinant underlying PTH independent hypercalcemia causing idiopathic infantile hypercalcemia (IIH).

IIH typically manifests in a child's first year and is characterised by hypercalcemia, vomiting, polyuria, dehydration, weight loss, failure to thrive, hypercalciuria and nephrocalcinosis. IIH was first reported in the 1950s when, in Great Britain, an epidemic of infantile hypercalciuria was linked to over fortification of vitamin D containing formula given to infants[122–124]. To rectify the epidemic the British Paediatric Association released guidance on reducing the prophylactic vitamin D dosage in fortified milk. Prior to this guidance it was common for infants to be prescribed vitamin D supplementation at 4,000 IU/day to prevent rickets development. The significant increase in IIH UK cases during the 1950s highlighted the correlation between high vitamin D supplementation in over fortified milk products and IIH disease penetrance. While initial IIH studies identified that high vitamin D supplementation unmasked many IIH cases among the vitamin D toxicity epidemic, an exact link between supplementation and the pathogenesis of IIH disease development remained unclear.

Although infant vitamin D supplementation dose has been reduced in the UK since the 1950s, many infants with suspected rickets are prescribed vitamin D supplementation without 25OHD profile assessment. Biochemical assessment of 25OHD during vitamin D supplementation can aid clinicians in preventing hypercalcemia in infants. The CYP24A1 enzyme is responsible for 25OHD and $1,25(OH)_2D$ hydroxylation to prevent vitamin D toxicity. In recent years IIH cases have been attributed to pathogenic variants in the *CYP24A1* gene sequence. While little is known of the exact prevalence of *CYP24A1* associated vitamin D toxicity, it has been reported that 35% of patients with hypercalcemia (>2.6 mmol/L serum calcium) and supressed PTH (<20 pg/mL) harbour *CYP24A1* variants[125]. An additional study estimated that the biallelic *CYP24A1* variant prevalence in the general population is 4-20%[126]. As *CYP24A1* loss-of-function mutations have been linked to IIH, the 'idiopathic' identifier of this condition is now deemed a misnomer. Vitamin D toxicity and/or sensitivity manifesting as hypercalcemia, hypercalciuria and/or nephrolithiasis caused by *CYP24A1* loss-of-function mutations resulting in elevated serum $1,25(OH)_2D$ is now refered to as infantile hypercalcaemia type 1 (HCINF1, OMIM #143880).

Hypomorphic mutations of CYP24A1, the enzyme responsible for hydroxylation of 25OHD and $1,25(OH)_2D$ (see section 1.6), can induce accumulation of circulating $1,25(OH)_2D$. Vitamin D toxicity and/or sensitivity is therefore a presenting feature of CYP24A1 mediated hypercalcemia (CMH) (Figure 1.13)[127]. Individuals with *CYP24A1* hypomorphic variants manifest increased sensitivity to vitamin D meaning vitamin D supplementation can trigger hypercalcemia. While CMH remains a rare condition, it highlights the significance of genotype on vitamin D metabolism and indicates a need for genetic evaluation when prescribed vitamin D supplementation leads to inappropriate calcium handling.

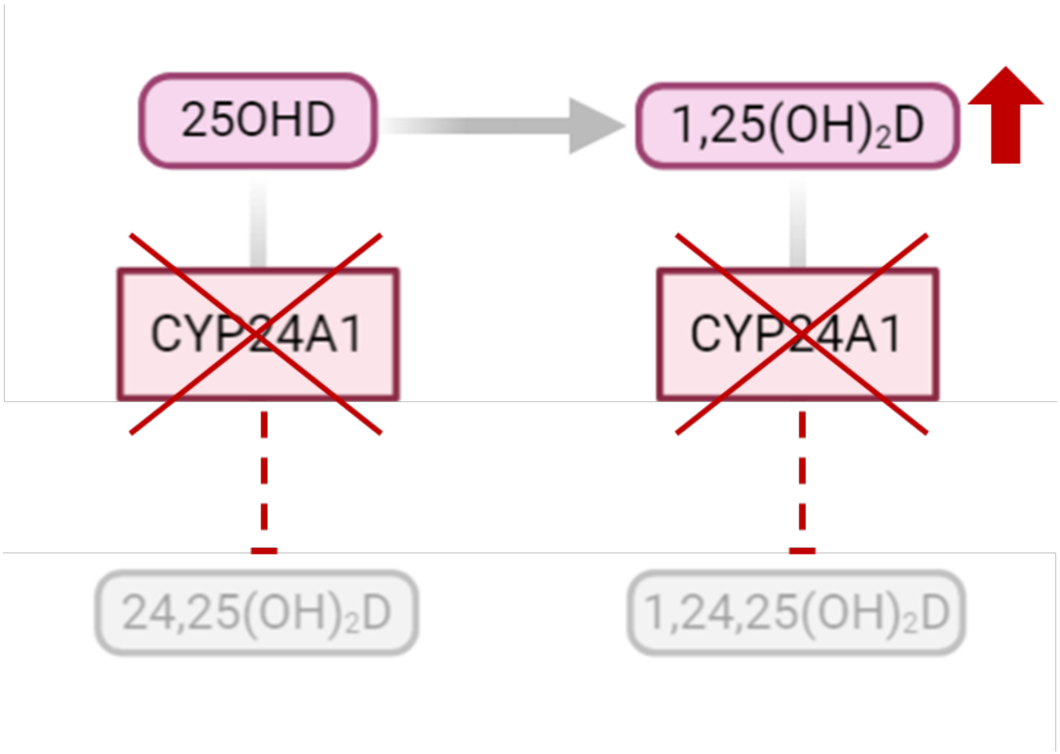


Figure 1.13: Loss-of-function mutations in *CYP24A1* causing hypercalcemia. Loss-of-function mutations in *CYP24A1* result in an inability to respond to elevated 1,25$(OH)_2$D concentrations. Reduced CYP24A1 activity causing 1,25$(OH)_2$D over accumulation can lead to vitamin D toxicity and hypercalcemia.

Several studies have reported adolescent and adult patients presenting with biallelic mutations in *CYP24A1* who remained asymptomatic in infancy. Adult presentation commonly manifests as renal stone formation, vomiting and can include flu-like symptoms[128]. CMH patients identified in adulthood occasionally have a history of recurrent kidney stone development throughout childhood. Often adult onset of CMH i78ocagered' by exogenous vitamin D supplementation. This 'triggering' of CMH in previously asymptomatic patients has been observed in females during and shortly after pregnancy[128]. This presentation is likely due to vitamin D supplementation and/or the natural physiological response to pregnancy. In order to meet the needs of the foetus, calcium metabolism changes to increase the $1,25(OH)_2D$ concentration for foetal skeletal mineralisation. This increase is due to increased CYP27B1 production from the placenta and kidneys. This alteration in metabolism should not normally cause maternal hypercalcemia although there are some rare cases of this occurring. When hypercalcemia does occur, it is important for clinicians to assess the $1,25(OH)_2D$ catabolism and whether a *CYP24A1* mutation could be the cause of excessive calcium production. Females reported to have biallelic mutations in *CYP24A1* were shown to have a high risk of developing severe clinical symptoms e.g., chronic kidney disease, nephrocalcinosis and nephrolithiasis either during pregnancy or shortly after labour[129].

Schlingmann *et al*. (2011) produced the primary evidence for hypomorphic mutations in the *CYP24A1* protein coding region associated in a HCINF1 cohort[127]. This study outlined the importance of CYP24A1 in the $1,25(OH)_2D$ hydroxylation in humans, which supported the previously reported mouse model findings[120,130,131]. Schlingmann performed mutational analysis on patients with suspected HCINF1 and identified homozygous and compound heterozygous mutations in *CYP24A1*. Subsequent studies identified patients with CMH plus associated *CYP24A1* inactivating mutations[132–135]. In 2018, Schlingmann *et al*. identified a 4-month-old infant with an unexplained hypercalcemia phenotype characteristic of HCINF1. Molecular genetic testing identified a *CYP24A1* homozygous mutation R396W. Subsequent studies have identified the same R396W mutation in several patients with HCINF1.

Studies have highlighted the phenotypic variability in patients with CMH by studying siblings with identical *CYP24A1* genetic abnormalities. One study followed an infant patient receiving 400 IU/d vitamin D supplementation to prevent the development of rickets[136]. After 10 months the patient was suspected of developing HCINF1 with renal ultrasound indicating severe nephrocalcinosis plus the development of hypercalcemia, hypercalciuria and suppressed PTH. The patient's sibling was also receiving vitamin D prophylaxis but was not diagnosed with rickets or HCINF1 at infancy. At age 13 y, the patient's sibling who was previously asymptomatic was found to have large concernments (a deposit of calcareous material) in both kidneys, with no prior nephrocalcinosis indication. The parents of these siblings were phenotypically unaffected as well as a third asymptomatic sibling[136]. Genetic testing revealed identical compound heterozygous mutations in *CYP24A1* in both children while the parents and third sibling were heterozygous carriers[136] (Figure 1.14). *CYP24A1* mutations causing E143del and R396W led to complete loss of enzyme function in both affected siblings indicating an autosomal recessive inheritance pattern[136]. This single nucleotide variant (SNV) loci observed in *CYP24A1* may be the cause of the extremely heterogeneous phenotype in HCINF1 patients due to the varying phenotype associated with identical mutations. Further studies showed that by omitting vitamin D supplementation in *CYP24A1* mutation positive patients, once HCINF1 is diagnosed in a sibling, there is an asymptomatic response regardless of whether they share identical biallelic mutations[132]. This finding indicates there is incomplete penetrance of inherited *CYP24A1* mutations with a clear link between HCINF1 penetrance and vitamin D dosage[136]. Infants who are diagnosed with rickets are often prescribed daily vitamin D supplementation. Supplementation often triggers early nephrocalcinosis from increased intestinal calcium uptake. The link between supplementation and HCINF1 has remained contradictory with some reports observing elevated $1,25(OH)_2D_3$[137], which is in contrast to other reports observing concentrations in the normal reference range[138].

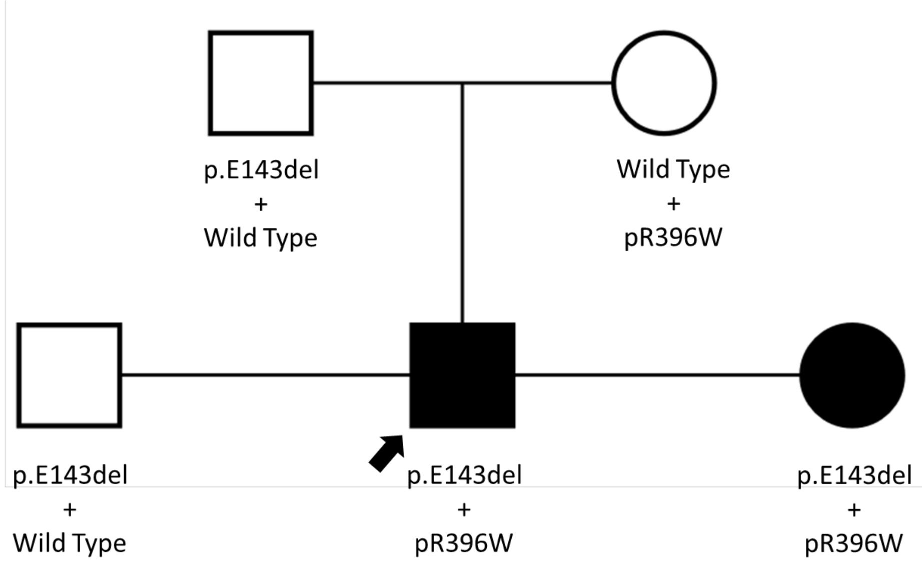


Figure 1.14: Family pedigree with *CYP24A1* mutations. Both affected children were compound heterozygous for p.E143del and p.R396W mutations in the *CYP24A1* coding region. Despite harbouring identical *CYP24A1* mutations, the proband displayed HCINF1 symptoms from 4 months while the affected sibling was asymptomatic until the age of 13 years suggesting incomplete disease penetrance and/or clinical symptoms only arising once cumulative dose of exogenous vitamin D supplementations is prescribed. Both parents and the third sibling carried one mutated allele and remained asymptomatic. The identified mutations lead to *CYP24A1* loss-of-function. These findings suggest a recessive inheritance pattern with incomplete penetrance of inherited *CYP24A1* defects[136].

It is clear from previous studies that CMH patients have a variable phenotype depending on the patient's genetics. Studies have shown that there is variation in the mutation phenotype between patients. Shah et al. concluded that normal calcium concentrations can be sustained even if the patient possesses a *CYP24A1* mutation, however once the calcium concentrations are modified to increase 1,25$(OH)_2$D concentration, e.g., during pregnancy, hypercalcemia can develop[139]. While *CYP24A1* loss-of-function mutations are typically deemed a PTH-independent cause of hypercalcemia, many studies have shown that patients with *CYP24A1* loss-of-function present with low-normal PTH concentrations[132–135,140–144]. Additionally, some studies demonstrate only partial penetrance in adults and an autosomal dominant pattern, others report autosomal recessive inheritance[127,145]. Monoallelic mutations have been shown to result in less severe phenotypes and are frequently asymptomatic. Monoallelic patients are more difficult to identify from biochemistry tests due to mild disease manifestations in comparison to biallelic patients who can be more conclusively diagnosed[145]. This supports the clinical need for biochemical testing to be performed alongside genetic analysis in these cases.

Some HCINF1 cases have been misdiagnosed due to an absence of protein coding *CYP24A1* mutations. In 2015 a study reported that *CYP24A1* mutations were absent in an HCINF1 patient cohort however a novel loss-of-function mutation was identified in *SLC34A1*[146]. *SLC34A1* encodes the sodium-dependent phosphate transport protein 2A (NaPi-2a) renal sodium co-transporter. Similarly to classical CMH patients, *SLC34A1* loss-of-function mutations result in elevated 1,25$(OH)_2$D causing hypercalcaemia (Figure 1.15). Due to NaPi-2a inactivation, serum PO_4 is reduced supressing FGF23 activity. Decreased FGF23 stimulates CYP27B1 activity facilitating 1,25$(OH)_2$D hydroxylation. In contrast to *CYP24A1* mutations, *SLC34A1* mediated hypercalcemia (SMH) is not resolved after removing vitamin D supplementation but is corrected by administering PO_4 replacement. Differential treatment for CMH and SMH highlights the different mutation mechanisms. It is therefore important to differentiate between patients carrying *CYP24A1* or *SLCS34A1* in both adult and infant patients, by PO_4 measurement, for appropriate treatment to be administered.

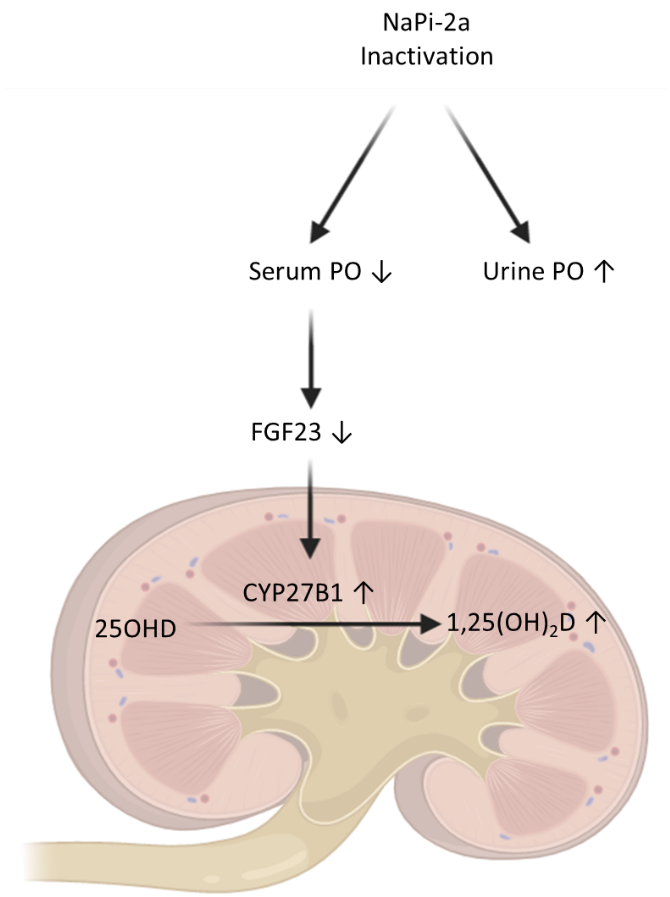


Figure 1.15: Loss-of-function mutations in *SLC34A1* cause NaPi-2a inactivation. Serum phosphate is decreased while phosphate expression in urine is increased. Decreased serum phosphate downregulates FGF23, which stimulates CYP27B1 activity in the kidneys driving 1,25(OH)$_2$D production leading to vitamin D toxicity and SMH.

Due to the increasing interest in vitamin D at a molecular level and its role in calciotropic processes such as HCINF1 as well as non-calciotropic disorders, e.g., cancer, standardisation of a clinical reference range for vitamin D metabolites and the vitamin D metabolite relative ratio (VMR) 24,25$(OH)_2$D:25OHD and 24,25$(OH)_2$D:1,24,25$(OH)_2$D relationship significance is required. Individuals with *CYP24A1* polymorphisms/variants manifest increased sensitivity to vitamin D meaning small supplementation doses can trigger hypercalcemia. While CMH remains a rare condition, it highlights the effect that genotype has on the vitamin D metabolism and indicates a need for genetic evaluation when prescribing vitamin D.

Biochemical analysis of vitamin D metabolites, PTH and corrected serum calcium is initially performed to diagnose and investigate hypercalcaemia. Genetic analysis should be considered to make a differential diagnosis of potential underlying *CYP24A1* variants in cases of inappropriately increased 1,25$(OH)_2$D and abnormal VMR. This is most important in cases where the patient has recurrent nephrolithiasis or nephrocalcinosis. CMH cases are rare and inheritance pattern and pathogenesis classification is inconclusive[147]. This further highlights how comparison of genetic analysis and biochemical findings can prevent the development of otherwise avoidable conditions such as nephrolithiasis and HCINF1 from unnecessary and unregulated supplementation in patients with underlying genetic variants e.g. *CYP24A1* loss-of-function.

Several studies have failed to observe *CYP24A1* mutations despite patients presenting with apparent CMH[125,148]. Recent case studies have failed to detect *CYP24A1* protein coding region SNVs all patients presenting with apparent CMH. Following proband identification with *CYP24A1* protein coding region SNVs, Dauber et al. recruited 27 patients with HCINF1 phenotypes. All 27 patients lacked protein coding region mutations despite displaying an analogous phenotype to the proband, indicating the HCINF1 heterogeneity plus the existence of unexplained disease mechanisms [148]. Patients who may have previously been overlooked due to absence of *CYP24A1* protein coding mutations despite CMH phenotypes often remain undiagnosed. Patients with CMH phenotypes plus no disease-causing variants upon sequencing require further investigation to identify any non-coding hypomorphic variants, plus further research is required to increase our understand of the disease mechanism and the presence of potential unexplained alternate pathways that exist in patients with non-classical CMH.

1.8 IDENTIFICATION AND TREATMENT OF *CYP24A1* MEDIATED HYPERCALCEMIA

Although rare, studies have shown that mutations in *CYP24A1* can be the underlying cause of some nephrocalcinosis cases. Reports from the dbSNP database have indicated that 4-20% of the general population (all ethnic groups and age ranges minus severe paediatric disease and first-degree relatives) harbour biallelic *CYP24A1* mutations that could lead to HCINF1 and/or lifelong nephrolithiasis[126]. The *CYP24A1* mutation frequency is expected to be much greater as the majority of nephrocalcinosis patients are asymptomatic meaning that the true frequency is difficult to estimate. A future discussion required in the field is the removal of the word “infantile” from HCINF1 given the increasing adult presentation frequency.

Intravenous bisphosphonates are typically prescribed when treating hypercalcaemic patients to attempt to counteract calcium stone formation. Bisphosphonates act by inhibiting bone resorption by osteoclasts, which suppresses calcium deposited into the bloodstream. In recent years the imidazole antifungal treatment ketoconazole has been identified as a potential alternative treatment for hypercalcaemic patients[138,147,149]. Previous studies reported that ketoconazole blocked the $1,25(OH)_2D$ synthesis in hypercalcaemic patients normalising their elevated serum calcium concentration[147]. The advantageous ketoconazole side effects can be harnessed for hypervitaminosis treatment. Further work demonstrated that by prescribing ketoconazole (200 mg) three times per day for sixty days normalised serum calcium and serum $1,25(OH)_2D$ in patients with hypervitaminosis[133]. Ketoconazole has also been shown to normalise suppressed serum PTH[133]. Ketoconazole treatment has been shown to be effective in treating patients with hypercalciuria and hyperphosphaturia although little evidence exists as to whether therapy will reverse reduced bone mineral density (BMD) observed in these patients[147]. A similar antifungal medication, Fluconazole, is suggested as an effective alternative treatment to ketoconazole for the vitamin D toxicity treatment[136]. Fluconazole is a cheaper alternative to ketoconazole with a higher potency. Low fluconazole dosage has been shown to decrease $1,25(OH)_2D_3$ and stabilise serum calcium levels. Further research is required into antifungal treatment for hypercalcaemia as little information is available on the effects of prolonged fluconazole treatment and its effect on steroidogenic pathways[136]. There is potential for fluconazole and ketoconazole to aid treatment of conditions associated with hypercalcemia, e.g., renal stone formation, by maintaining suitable vitamin D metabolite concentrations and serum calcium levels once these patients are identified[133].

1.9 VITAMIN D METABOLITE RELATIVE RATIO (VMR)

Investigations into hypercalcemia currently involve biochemical tests for 25OHD, $1,25(OH)_2D$, $24,25(OH)_2D$, PTH, serum calcium and phosphate. Recently, an LC-MS/MS method to measure 25OHD and its metabolite $24,25(OH)_2D$ simultaneously was developed [150]. The establishment of a quantitative assay for vitamin D metabolite assessment supported the determination of a diagnostic value for the ratio 25OHD:24,25$(OH)_2$D, termed the vitamin D metabolite relative ratio (VMR), in patients with *CYP24A1* inactivating mutations. VMR is achieved by dividing 25OHD serum concentration by 24,25$(OH)_2$D serum concentration to achieve a relative ratio. In patients with reduced CYP24A1 activity, 24,25$(OH)_2$D concentrations are decreased or undetectable while 25OHD concentrations remain normal or can be elevated. The decrease in 24,25$(OH)_2$D concentration increases the relative ratio between 24,25$(OH)_2$D and 25OHD, which increases the VMR allowing evaluation of CYP24A1 activity. Serum VMR measurement can therefore be used as a CYP24A1 status predictor. The VMR is more conclusive than serum 24,25$(OH)_2$D concentrations alone as it negates the possibility of vitamin D deficiency causing a reduced 24,25$(OH)_2$D concentration. The 24,25$(OH)_2$D concentration is reduced in CMH patients with *CYP24A1* mutations while serum 25OHD is increased. These opposite lineal concentrations suggests that the VMR (25OHD:24,25$(OH)_2$D) may be a more clinically relevant indicator for determining varying endogenous vitamin D dosage effects and detect when supplementation becomes detrimental. One case study presented a patient with hypercalcemia with a 25OHD:24,25$(OH)_2$D relative ratio of 132, which is significantly higher than the top end of the reference range (<25)[149]. Genetic testing confirmed that the 53-year-old Caucasian male harboured a compound heterozygous mutation in *CYP24A1*, which is likely the cause of the observed hypercalcemia plus unusual elevated VMR in this patient. The availability of initial diagnosis through VMR assessment has increased the current *CYP24A1* mutation understanding and HCINF1 frequency. Rapid serum screening has allowed for fast differential HCINF1 diagnosis that can be confirmed by next generation sequencing (NGS). Genetic testing for *CYP24A1* mutations is vital to identify similar patients who would be successfully treated before complications arise e.g., infantile hypercalcaemia, nephrocalcinosis, nephrolithiasis and chronic kidney disease.

Clinical biochemical/metabolite analysis is widely established. Functional genomics describes the entire cellular pathway from genotype to phenotype. By understanding each stage of the process from genome to epigenome, transcriptome, proteome and metabolome, a greater understanding of human disease will be obtained. Functional genomics revealed that the gene *CYP24A1* is associated with HCINF1 development and adult-onset hypercalcemia with high phenotypic variability between patients (e.g., mouse and human studies described above). One unanswered question remains. The 2012 study by Dauber et al indicates that some patients suspected of *CYP24A1* loss-of-function mutations do not display protein coding mutations[148]. Absence of protein coding *CYP24A1* mutations plus the *CYP24A1* null mouse study suggests alternative routes of vitamin D metabolism.

1.10 MOLECULAR MEDICINE AND GENE EXPRESSION

Molecular medicine is a broad research area that encompasses molecular and cell biology, genetics, proteomics and medical physics. Molecular medicine allows for identification of molecular structures, disease mechanisms, disease associated genetic abnormalities and assists in the development of molecular interventions for human disease. The study of gene expression in human disease aids disease classification and has advanced personalised medicine development. Genetic variability occurring between individuals causes patients to respond differently to treatment. By applying personalised medicine based on patient genetic makeup, patients can receive tailored therapy providing the optimal therapeutic response rather than universal treatments such as some cancer chemotherapy that may not benefit every individual.

Molecular genetics investigates gene expression among different organisms. This investigative approach deciphers the gene structure and function and how genetic mutations can alter specific phenotypes. Molecular genetics aids treatment development by bridging the gap between mutations in deoxyribonucleic acid (DNA) and molecular disease. Human genetics was greatly accelerated by the Human Genome Project completion in 2003[151]. This international collaboration provided researchers with the entire human genomic sequence (~3 billion base pairs) revealing the identity of 22,287 protein coding genes in the primary assembly[151].

The human genome completion initiated a need for high throughput and sophisticated sequencing techniques to be developed to facilitate the production of large data sets. Next generation sequencing (NGS) revolutionised the genomic and molecular biology field in 2009. NGS provides a high throughput technology that allows vast amounts of DNA sequencing data to be produced at a faster rate and more cost effective than the prior technique, Sanger or direct sequencing.

The human genome sequence revealed that fewer protein coding genes were present than originally anticipated (~100,000). While differential gene expression is regulated at a gene specific transcription level, post transcriptional events that chemically alter mRNA following transcription contribute to differing expression and function. Single protein coding genes can give rise to several different protein isoforms due to post-transcriptional modifications. Post-translational modifications can alter the protein after production by ribosomes. Post-translational modification occurs through proteolytic cleavage of one or multiple bonds in a target protein affecting its activity, glycosylation or phosphorylation. After undergoing transcription, nuclear processing steps produce mature mRNA. The process of preparing mRNA involves specific post transcriptional events including 5’ end capping, which produces a 7-methylguanosine cap, splicing of introns and 3’ cleavage or polyadenylation. Incorrect processing of mRNA leads to incorrect translation of the mRNA to the functional protein (Figure 1.16).

During 5’ capping the methylated 7-methylguanosine (m^7G) is linked by a 5’-5’ phosphodiester bond to the 5’ end of pre mRNA, which ‘caps’ the transcript[152]. The post transcriptional addition of the 5’ cap provides protection of the RNA transcript from ribonucleases degradation, facilitates transport from the nucleus to the cytoplasm for translation, RNA splicing and allows attachment of 40S subunit of cytoplasmic ribosomes during translation[152] (Figure 1.16). Polyadenylation, or 3’ cleavage of pre mRNA, results in the addition of adenine nucleotide bases to the 3’ end forming a poly (A) tail. The synthesis of the poly (A) tail binds to the specific poly (A) binding protein prevents mRNA degradation by ribonuclease, aid mRNA stabilisation, facilitates transport from the nucleus to the cytoplasm and enhance mRNA recognition by ribosome machinery[152] (Figure 1.16). The pre mRNA sequence contains both introns and exons. Post transcriptional removal of introns (splicing) is mediated by the RNA-protein complex called the spliceosome. The sites of splicing reactions are mediated by RNA-RNA base pairing between the pre mRNA molecule and the small nuclear RNA (snRNA) of the splicosome[152] (Figure 1.16). Alternative splicing is controlled by regulatory proteins and snRNAs and removes introns prior to mature messenger RNA (mRNA) generation. Different cell types express different regulatory proteins allowing various exon combinations to be produced in each cell type depending on the required protein function. Alternative splicing allows for several mRNA variations and subsequent protein isoforms to be produced from a single gene. For example, the calcitonin/calcitonin gene-related peptide-I (CALC-I) undergoes alternative splicing producing either CGRP, a neurotransmitter or calcitonin, a hormone involved in calcium homeostasis[153]. Post transcriptional modification of the pre mRNA can be altered by mutations within the RNA sequence. Mutations that alter post transcriptional processes are linked to a variety of human disease e.g. muscular dystrophy and myelodysplastic syndromes[154].

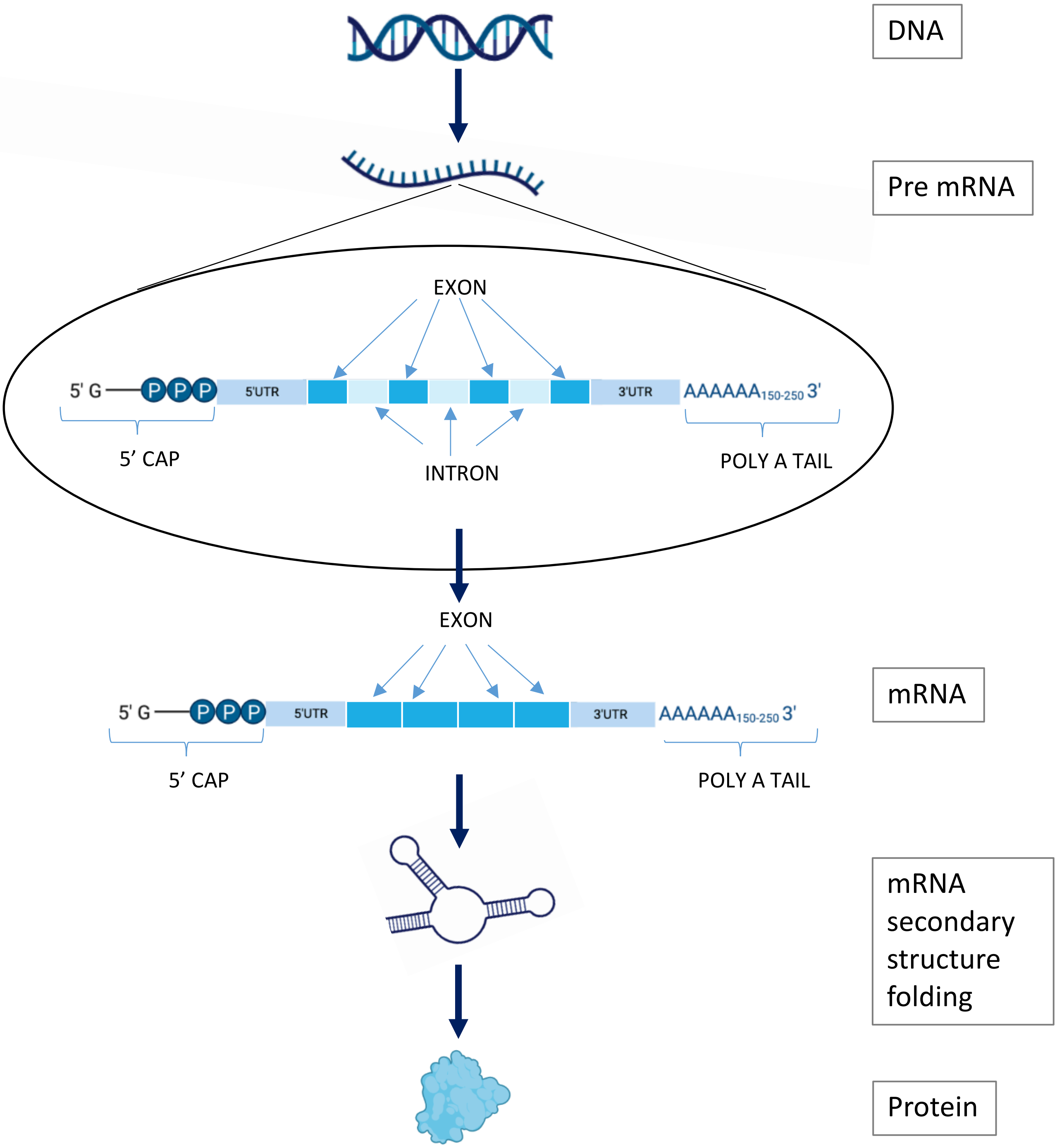


Figure 1.16: Summary of transcription and translation. DNA containing genetic material is first transcribed to pre mRNA. The resulting pre mRNA strand is a single stranded reverse-complement of the DNA. Post transcriptional modifications then occur to produce the mRNA molecule. The post transcriptional modifications include 5’ capping, 3’ poly A tail additions and splicing of the non-coding intron regions of the pre mRNA sequence. The resulting mRNA contains only coding exons and is folded into a complex secondary structure. The mRNA is then transported to the cytoplasm from the nucleus and undergoes translation via the ribosome to produce the desired protein.

Post translational modifications can occur at any stage of the protein's life cycle e.g. following translation where protein folding and stability are influenced, or once folding is completed to activate/inactivate the proteins biological activity. Post transcriptional modifications have different roles in altering protein function including activation through cleavage and tagging proteins for degradation. Additionally, modifications can aid interactions and localisation of the protein. Post translational modifications including phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation and proteolysis. Such processes increase the variability in protein function. Chemical groups are covalently bound to side chains of amino acids during post transcriptional chemical modification (Table 1.5) [155,152].

Lipidation occurs when fatty acyl or prenyl groups are added to proteins, typically membrane proteins. Lipidation is a post translational modification that aids membrane anchoring/binding affinity to the outer plasma membrane, which regulates subcellular localisation, interactions, folding and stability[155,152]. Proteolysis via the enzyme protease allows cleavage of peptide bonds of a protein sequence to produce smaller molecules. Therefore, a single mRNA transcript can produce multiple functional polypeptide chains due to post-translational cleavage. Proteolysis is critical for antigen processing apoptosis and cell signaling[155,152].

Table 1.5: Summary of Post Translational Modifications[155,152]

Modification	Amino Acid Target	Description
Phosphorylation	Try, Ser, Thr	Addition of PO_3 Aids many processes including cellular storage, regulating metabolism, transcription, cell cycle progression, transfer of free energy and differentiation
Methylation	Lys	Addition of methyl group Aids mRNA splicing, RNA damage signaling, mRNA translation and cell signaling.
Ubiquitination	Lys	Addition of ubiquitin Tags specific proteins for degradation by the proteosome, aids cellular localisation, regulates protein interactions
Acetylation	Lys	Addition of acetyl group

		Alters hydrophobicity, solubility and surface properties
Glycosylation	Asn (N- linked) Ser, Thr, Hyl (O- linked)	Addition of sugar moieties Aids localisation, cell-cell adhesion and determines protein structure

RNA modifications control processes ranging from alternative splicing to RNA stability, localisation, and translation. Regulation through complex networks of interactions lead to dynamic control of gene expression with deep implications for cellular physiology and pathology. Given the regulatory role in multiple cellular functions, alterations in genes encoding elements of post-transcriptional processing and their modifying enzymes have been observed in various human diseases[156–158], including cancer and neurological disorders. Post transcriptional and translational modifications have risen as major factors in regulating gene expression. The mechanism of action and effect on cell physiology and pathology are not fully understood and are the focus of epigenetic research. While post transcriptional or translational modifications have not previously been reported in CYP24A1 in association with abnormal calcium handling and/hypervitaminosis D, further investigations into the role of epitranscriptomics and disease progression is required to further understanding of disease pathogenesis and therapeutic strategies in association with human disease[159].

Investigations into monogenic disease, facilitated by GWAS, highlighted many cases where inherited traits were not explained by genetic variance. This confusion was previously thought to be due to an inability to efficiently detect rare single nucleotide variants (SNVs), structural defects or complex inheritance patterns. SNVs differ from SNPs in that these base alterations occur without population frequency limitations, i.e., SNVs occur in individuals rather than at >1% population level., which are described as SNPs[160]. Exome sequencing of the coding region allows for the measurement of variance in the DNA sequence. Investigations into genetic disease has previously focused largely on the protein coding region. Exome sequencing provided base-scale resolution of genetic variation that directly affects the subsequent protein. Exome sequencing limitations were revealed as GWAS identified that many informative variants that contributed to disease etiology resided in the non-coding region of the genome. Functional sites have been identified in non-coding regions of the genome that include cis-acting DNA transcriptional enhancers that assist gene transcription[161]. Transcriptional enhancers simultaneously regulate several genes and may not regulate proximal genes exclusively. It was hypothesised that missing heritability previously observed could be due to variants in DNA promoters, enhancers, structural elements and regulatory RNA coding regions[162]. Alterations in the 3’ and 5’ untranslated regions (UTRs) have been shown to impact transcription stability and translation by causing RNA structural alterations. It is estimated that ~35% of 5’ UTR and ~10% of 3’UTR harbor introns in the human genome[163]. Whole genome sequencing development confirmed that a significant amount of clinically relevant genetic information resides in the non-coding region, which was previously overlooked by diagnostic laboratories when sequencing exons exclusively[164]. Whole genome sequencing has allowed researchers to sequence both coding and non-coding regions of the genome producing large databases of genetic variation, e.g., 100,000 Genomes Project. Whole genome sequencing provided information on common and rare pathogenic variants revealing 98% of variants in the human genome reside in the non-coding region. These findings have increased research interest in the non-coding region to improve rare genetic disease diagnosis.

The study of single gene disorders was greatly advanced by the Human Genome Project genomic revolution. Single gene disorders arise from DNA variants in a single gene. Single gene disorders often have predictable inheritance patterns (recessive or dominant) and are relatively rare with only 1/100 human diseases known to be monogenic. Genetic testing is available for many familial single-gene disorders, e.g., some cancer types, musculoskeletal disorders, cystic fibrosis and sickle cell anemia[165]. Phenotypic heterogeneity has been observed where multiple different mutations in a single gene can cause varying phenotypic severity. Identical genetic variants have also been show to result in different phenotypes, which are thought to be influenced by the environment and/or additional genetic variations[136,166,167]. As previously described, the phenotypic variability in siblings with identical *CYP24A1* genetic abnormalities has been observed[136]. The phenotypic *CYP24A1* mutation spectrum is one example of how certain monogenic disorders can present at all age groups with varying severity. Non-coding variants have been linked to single cell disorders or associated simulating complex disease traits.

The combination of the human genome and new molecular tools has greatly expanded our understanding of disease-associated genes. Understanding single-gene disease can further advance our knowledge of all types of human disease with many discoveries also providing information on common human disease mechanisms.

1.11 RNA SECONDARY STRUCTURE

mRNA is responsible for directing protein synthesis by transporting genetic information from DNA to the ribosomes for translation. Features encoded in the mRNA primary sequence govern translational efficiency. The single stranded mRNA sequence determines its secondary and tertiary structure. Complementary sequences on the same RNA molecule fold back onto themselves to form partial self-base paired RNA. mRNA forms intramolecular hydrogen bonds between complementary bases due to the extra hydroxyl group, which DNA lacks. The 2'hydroxyl group on mRNA aids hydrogen bond formation that is important for stable structure formations. Interruptions in the complementary sequence remain single stranded and become looped into hairpin formations between stretches of nucleotide pairs. Hairpin regions can vary extensively. Hairpin loops are unpaired mRNA regions, which form as mRNA fold and forms base pairs with another section along the same mRNA strand causing a 'loop'. Short hairpin RNA (shRNA) sequences consist of stem-loop and microRNA-adapted shRNA. Short hairpins are located in parallel on long RNA molecules. Long hairpins (>100 bp) often contain bulges of mismatched regions. Hairpins and bulges in the RNA secondary structure often contain important recognition sites for many regulatory, structural and cleavage proteins, for example, RO60 and ANG[168]. Bulges can act directly as recognition sites for interacting molecules or indirectly by altering the RNA backbone to allow distant base pairs to become accessible for protein-RNA interactions. The intricate RNA secondary structure provides an added layer of regulation for gene expression depending on the cell type and required gene function[169]. The dynamic RNA structure affects post-transcriptional regulation including RNA maturation, translation and degradation. The 3D RNA secondary structure conformation has a central role in the biological function of RNA molecules, yet these structures have remained difficult to accurately predict.

SNVs in the mRNA sequence can alter the secondary structure conformation and subsequent mRNA function. Proper RNA folding can be affected by loss-of-function mutations that alter sites crucial for proper folding a RNA binding protein regognision. Mutations within the microtubule-associated protein tau (MAPT) destabilises the hairpin structure (exon 10), which alters protein interactions and splicing of the pre mRNA[170]. Gain-of-function RNA variants are associated with abnormal folding in locations that typically lack folding motifs. Additional folding can cause non-functional or toxic proteins to be produced during translation[170]. The translational regulation breakdown leading to dysfunctional protein production has also been linked to disease development. RNA misfolding is associated with many human diseases including Huntingdon’s disease, myotonic dystrophy and Fragile X syndrome[170,171]. mRNA translational ability is largely regulated by the 5’ and 3’ UTRs, which have key roles in mRNA stability, ribosome identification and interactions with translational machinery. Both the 5’ and 3’ UTRs are formed from complex secondary structure formations. The 5' UTR of mRNA supports ribosome recognition during translation due to a guanosine triphosphate nucleotide cap. The protein synthesis translation rate is affected by the 5' UTR length, secondary structure locations such as bulges, stems and loops, thermal stability, GC content, upstream open reading frames (uORFs) and internal ribosome entry sites (IRES). If specific protein binding sites in the 5' UTR that are required for translation are altered, translation can be hindered causing disease development and/or increased disease susceptibility[172]. The 3' UTR on the opposing end of mRNA includes a poly A tail to prevent enzymatic degradation[173]. Mutations affecting the 3' UTR are often overlooked in genetic screening but have been shown to play important roles in disease progression[161]. For example, a 3' UTR polymorphism in the human *CYP2A6* gene has been shown to affect mRNA stability and enzyme progression causing increased protein concentration and catalytic activity. The functional polymorphism found in the *CYP2A6* 3' UTR produced different secondary mRNA structures and distinct allele dependent difference in mRNA stability[174]. The identified *CYP2A6* polymorphism was shown to alter nicotine elimination and risk of smoking induced lung cancer[174]. The UTR protective and functional roles promote the stable mRNA secondary structure formation. Intramolecular mRNA secondary structure alterations plus UTR length have been linked with various diseases including breast cancer, congenital heart disease (CHD) and arrhythmogenic right ventricular cardiomyopathy/dysplasia (AVRC)[175,176].

UTRs facilitate mRNA interactions with molecules such as RNA-binding proteins (RBP) and ribosomes. RBPs and mRNA interactions control mRNA half-life, subcellular localisation, ribosome recruitment and translational efficiency. UTR variants causing disruption to proper protein-RNA interaction are associated with disease. Previous lack of UTR analysis has overlooked alterations in translational mechanisms by investigating coding regions exclusively. Further investigation into UTR mutations plus downstream effects of these alterations will provide more detailed information for diagnosis and prognosis of different diseases. Current research focusing on how UTR mutations alter mRNA secondary structure is being made possible by the development of new mRNA structure determination methodology. Development of techniques for rapid and accurate mRNA structure-function relationship characterisation is a major advancement in RNA biology.

Furthering our understanding of RNA structure is key to understanding RNA function in human disease. NGS development has aided the progression of RNA structural investigations from initial *in vitro* low throughout probing to high throughput *in vivo* and *ex vivo* profiling. Previous RNA structure determination methods included nuclear magnetic resonance spectroscopy (NMR), X-ray crystallography and cryo-electron microscopy, which require extensive resources but only produce limited data on high abundance, small RNAs (<200 nt and >1 μmol). Technologies using chemical structure probing methods have been developed to efficiently decipher *in vivo* structural information from the whole RNA transcript in a single reaction. Two classes of chemical probing reagents have been developed for structural determination. One class of chemical reagent modifies the Watson-Crick base pairing on the nucleobase to determine the single stranded mRNA regions. A common nucleobase reagent is dimethyl sulfate (DMS) that can efficiently penetrate the cell for *in vivo* probing in plant models. One DMS treatment limitation is that only bases A and C are reactive. The second class of chemical reagent modifies the ribose through SHAPE. Bioinformatics tools102ocalize folding algorithms and the probability that a nucleotide is single stranded or constrained to produce the mRNA secondary structure of mRNA. SHAPE reagents are able to react with all four nucleotide bases across the whole transcript in a single reaction (Figure 1.17).

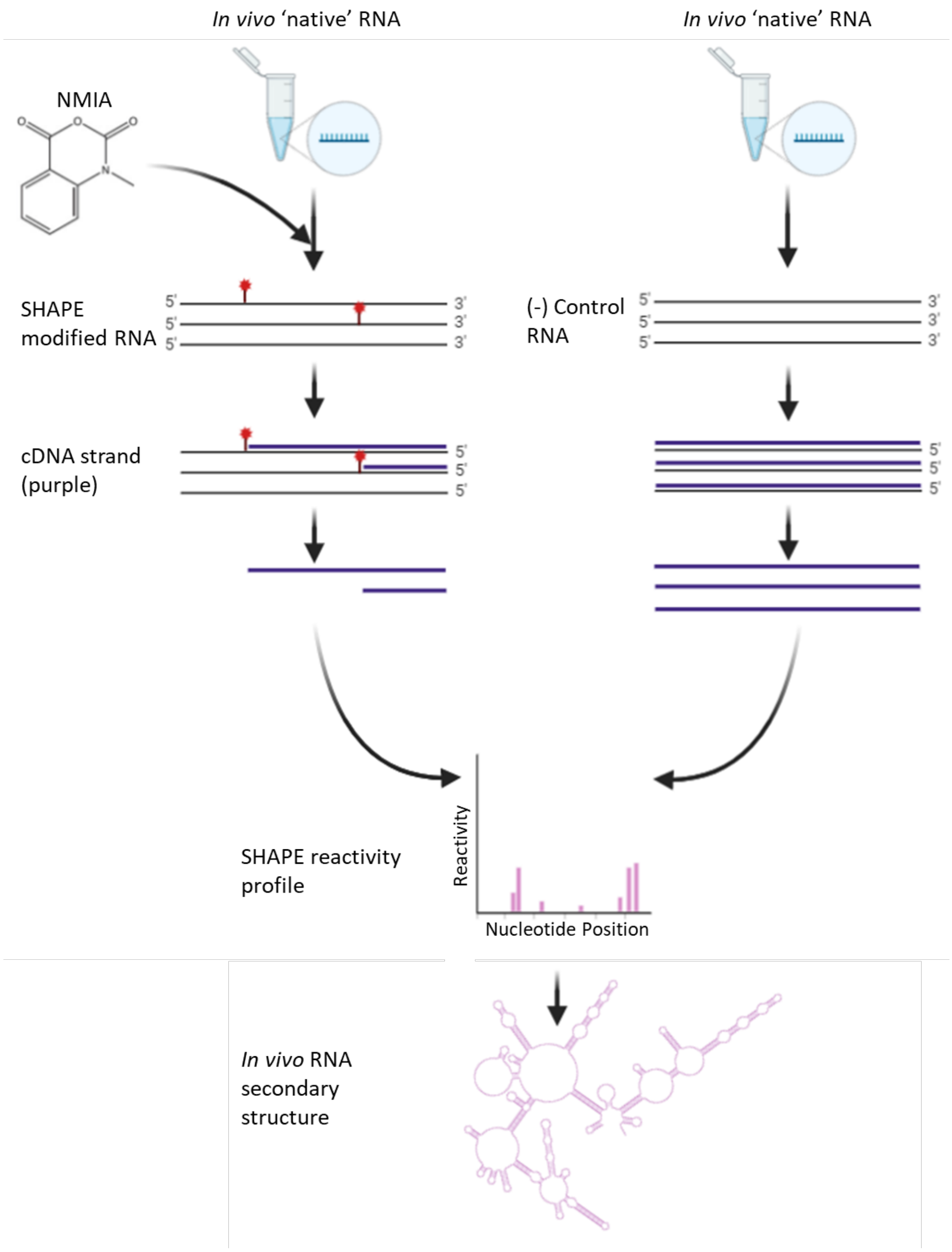


Figure 1.17: Workflow and SHAPE chemical modification. SHAPE reagents, e.g., N-methyl isotonic anhydride (NMIA) react with native RNA. Non constrained RNA regions (single stranded bases) react with NMIA in the 'treated' sample and are modified by the addition of an adduct at the 2'-OH position of RNA. Primer extension reverse transcription is used to identify termination positions where the reverse transcriptase reaches a nucleotide modified by NMIA. Modification positions correlate with mRNA structure formation. The cDNA pool produced maps to flexible nucleotide areas in the RNA and can be

used as pseudo-free energy change. Bioinformatics tools104ocalize folding algorithms and the probability that a nucleotide is single stranded or constrained to produce the mRNA secondary structure of mRNA. For SHAPE probing reactions using reverse transcription primers, a control reaction without NMIA reagent is performed in parallel with the probed modified reaction. The control data is used to normalise the SHAPE reaction data to give estimates of the SHAPE reactivity of each position without background interference.

The rise of genome wide technology has allowed accurate, quantitative and *in vivo* RNA structure mapping with single nucleotide resolution. In depth analysis of the impact of RNA structure on gene regulation is now possible with high throughput and minimal time constraints. Initial *in vivo* secondary structure resolution was performed in *Arabidopsis thaliana*[177]. By combining DMS and NGS technology, novel RNA structural features were observed that affected RNA function. This study identified a three nucleotide periodic pattern in the coding sequence that altered translational efficiency. It was also found that mRNA associated with abiotic stress response were mainly single stranded to maintain greater flexibility allowing varied response to environmental conditions. This *in vivo A. thaliana* study indicated that RNA secondary structure plays a key role in environmental adaptation supporting the findings that RNA structure affects all mRNA processing levels, e.g., translation, splicing and degradation[177]. The RNA structural diversity between difference species was first compared *in vivo* between plant species in 2018[178]. Rice and *A. thaliana* descended from a common ancestor 150 million years ago. The RNA structure conservation and diversity in *A. thaliana* and rice was used to assess the evolutionary adaptation in RNA secondary structure. Little correlation was observed in both sequence and *in vivo* structure[178]. These findings suggested that evolutionary selection modifies both RNA sequence and structure to regulate gene expression with RNA structure potentially providing an additional layer of regulation in plants meaning it is conceivable that the same may be true in animals. *In vivo* studies on rice also indicated that higher m^6A modification sites were less structured suggesting that m^6A association may facilitate gene regulation by altering the RNA structure to become more single stranded[178].

RNAs involvement in genetic regulation is well understood. RNA function is partly determined by its sequence plus the subsequent secondary structure formation. Development of techniques for rapid and accurate characterisation of structure-function relationships of mRNA is a key advancement in RNA biology. Understanding the genetic, transcriptomic and epigenetic basis of human diseases is a vital modern medicine goal since heritable traits were first observed. Each individual poses different disease susceptibility and progression. Their individual genetic makeup can also affect the treatment effectiveness depending on their specific response. Identification of genetic information, which affects disease phenotype, is vital for improved prevention, identification and disease treatment on a patient-by-patient basis. Progress in technological and conceptual investigation has expanded our understanding of complex human genome. Future research aims to expand the current knowledge of the noncoding genome and genomic variant effects through comprehensive genomic, transcriptomic and epigenomic profiling.

SNVs have been linked to many human diseases by affecting gene expression and protein binding plus potentially causing as yet unidentified effects. Most genes (93%) contain SNVs[179]. Non-coding region SNVs do not affect protein synthesis meaning the mechanism through which they cause disease is less understood. Understanding the non-coding SNV implications in human disease is a rising interest supported by newly developed *in vivo* investigative approaches, e.g., SHAPE. SNVs have previously been shown to affect protein and/or microRNA (miRNA) binding if the binding site or flanking regions are directly affected. Recent evidence has indicated that SNVs can also affect binding indirectly through RNA secondary structure alterations. A study published in 2020 identified how SNVs alter the secondary structure reduces the RNA affinity to its RNA binding protein/miRNA from outside the binding motif[180]. While these findings106ocalizad *in silico* RNA structure prediction tools, future studies investigating the SNVs effect on mRNA secondary structure *in vivo* would benefit from investigating a wide radius of potential target locations.

1.12 THESIS AIMS AND OBJECTIVES

In vivo and/or *ex vivo* investigations including the non-coding regions of *CYP24A1* will further our understanding on disease mechanism and phenotypic variability in conditions such as CMH. Structural changes in the mRNA transcript, caused by non-coding SNVs, in genes such as *CYP24A1* may be the fundamental mechanism behind CMH cases where pathogenic mutations in the protein coding region are not detected. Translation requires a specific and highly regulated sequence of events to enable successful protein production. Mutations causing changes to binding motifs in the *CYP24A1* transcript may result in nephrolithiasis and CMH[30]. Mutations affecting the 3' UTR are often overlooked in clinical genetic screening but these regions have been shown to play important roles in disease progression[21]. Further investigations into UTR mutations and downstream effects of these alterations could provide more detailed information for diagnosis and prognosis of CMH.

This thesis will investigate:

- *CYP24A1* hypomorphic variants in non-classical CMH cases where no protein coding mutations are observed.
- The effect of hypomorphic *CYP24A1* non-coding variants on mRNA secondary structure will be assessed *in silico, in vitro* and *ex vivo.*
- The effect of hypomorphic *CYP24A1* non-coding variants on CYP24A1 protein expression alterations to determine the effect on CYP24A1 functionality in non-classical CMH patients.
- The effect of hypomorphic *CYP24A1* non-coding variants on *CYP24A1* localisation *in vitro* and *ex vivo*.

By broadening our understanding of this rare patient cohort, this research aims to support identification and treatment of patients who may be misdiagnosed using classical CMH investigations.

CHAPTER 2: MATERIALS AND METHODS

Consumables, reagents and water used throughout this thesis were sterile and negative for dNase, rNase and pyrogen contaminants. All equipment and bench tops were sterilised with 2 % Chemgene (Thermo Fisher Scientific, Loughborough, UK, Loughborough, UK) before each experiment.

2.1 CELL CULTURE

Human osteosarcoma (143B) (Public Health England) and Human embryonic kidney (HEK293T) cells (gifted by Tom Wileman Laboratory, UEA) were cultured as a monolayer in DMEM GlutaMax medium (Thermo Fisher Scientific, Loughborough, UK) supplemented with 1% (volume/volume) penicillin-streptomycin (Thermo Fisher Scientific, Loughborough, UK) and 10% (volume/volume) fetal bovine serum (Sigma Aldrich, Gillingham, UK). Cells were cultured in 75 cm^3 flasks and maintained at 37 ^oC in 5 % carbon dioxide. Cells were split using 0.25% trypsin-EDTA (Thermo Fisher Scientific, Loughborough, UK) to dissociate from the flask and re-seeded with fresh medium once confluency reached 80-90% (every 2-3 days), to a new confluency of ~40%

2.2 CLINICAL SAMPLES

2.2.1 CHAPTER 3 CLINICAL SAMPLES

The LC-MS/MS method for measuring 25OHD and 1,25$(OH)_2$D to perform VMR relative ratio was developed using samples from an MOD study investigating vitamin D status and fracture risk during basic training. The study received ethics approval from the UK Ministry of Defence Research Ethics Committee (MODREC 165/Gen/10 and 692/MoDREC/15 ClinicalTrials.gov Identifier[181]) and was conducted in accordance with the Declaration of Helsinki (2013). The characteristics of the subjects included in chapter 3 are shown below (Table 2.1). In total, 2,252 new British Army recruits at the start of phase one training volunteered for the study. Written informed consent was obtained from all study participants, and each participant was required to complete a detailed health questionnaire including medical history and the use of dietary supplements. All recruits undertook physical and cognitive testing, and a detailed medical examination prior to joining the army. The British Army entry requirements restrict individuals with chronic medical conditions; therefore, the study population represents a medically screened, disease-free, and physically fit population. In the analysis, individuals who reported the use of calcium and vitamin D supplements (including multivitamins and cod liver oil) and participants who reported injury and illness prior to recruitment were excluded. Additionally, participants with conditions such as being underweight, eating disorders, or those with a history of bone fracture were excluded. A total of 940 participants were included in the final statistical analyses. The majority of participants were from the Caucasian population (92.9%), with a minority from a diverse ethnicity (Asian 1.6%, Black 1.7%, Chinese 0.1%, mixed 3%, others 0.7%).

Table 2.1: Baseline characteristics of the subjects included in chapter 3. *Data shown as the mean with ± SD in square brackets unless otherwise stated.

	Male	**Female**
No. of participants	652	288
Mean age, years [range]	21.7 [18–32]	22.1 [18–32]
Height, m [± SD]	1.77 [6.4]	1.66 [5.9]
body mass, kg [± SD]	75.9 [9.8]	64.7 [7.5]
Body mass index (BMI) [± SD]	24.1 [2.6]	23.4 [3.3]
Total body bone mineral density (BMD) (g/cm^2) [± SD]	1.24 [0.10]	1.16 [0.09]

Venous blood samples were obtained from the participants at the start of 14-week long basic military training. Sample collections were scheduled on a monthly basis to balance the seasonal variations because it is known that vitamin D status fluctuates during the year for individuals who live in the Northern hemisphere. Each recruitment intake comprised, on average (range), 86 (43–120) participants. Blood samples were collected into serum gel separator tube and EDTA plasma container (BD Vacutainer). Samples were centrifuged immediately after collection at 3,000 × g for 10 minutes. Plasma/serum layers were aliquoted into a separate polystyrene tube and stored at −20 °C until analysis. All samples were anonymised at the point of access.

2.2.2. CHAPTER 4-6 AND 8 CLINICAL SAMPLES

The University of East Anglia (UEA) Faculty of Medicine and Health Sciences Research Ethics Committee approved the collection and study of Human samples for non-clinical procedures investigating *CYP24A1* abnormalities (Reference: 2018/19–-100). Forty-seven patient serum samples were collected as part of routine requests for 25OHD LC-MS/MS analysis from the Department of Laboratory Medicine at the Norfolk and Norwich University Hospital between June 2016 and June 2017. Patients were referred from the metabolic or stone former clinics. Blood samples were collected into serum gel separator tubes (BD Vacutainer) and centrifuged immediately. The serum layer was the aliquoted and stored at -20 °C until analysis.

Whole blood from the Norfolk and Norwich University Hospital metabolic and stone former clinics for genetic analysis was obtained from patients identified with inappropriate 1,25(OH)$_2$D and/or VMR plus clinical presentation of nephrolithiasis and/or hypercalciuria serum (Patients 1-4 Chapter 5). Genomic DNA was obtained from Croydon Hospital from an infant presenting with nephrocalcinosis and Williams-

Beuren syndrome (Patient 5 Chapter 5) and from the Glasgow Children's Hospital Renal Unit from an infant presenting with nephrocalcinosis and polyurea (Patient 6 Chapter 5). All adults or infant parents/guardians provided written informed consent to donate samples to this study.

Negative control whole blood samples were collected at the Norfolk and Norwich University Hospital blood typing service (n=10). Exclusion criteria for control samples were those with a vitamin D, calcium or other metabolic disorder clinical history. Control samples were collected using the UK NHS Research Ethics Committee decision toolkit (http://www.hra-decisiontools.org.uk/ethics/).

2.3 LC-MS/MS STANDARD MATERIAL AND QUALITY CONTROL PREPARATION

Human lyophilised multilevel serum standards (Chromsystems, Gräfelfing, Germany) were used to calibrate $25OHD_3$ and $25OHD_2$. Standards for 25OHD were traceable to standard reference material SRM972a from the National Institute of Science and Technology (NIST). In house spiked standards for $24,25(OH)_2D_3$ and $24,25(OH)_2D_2$ were created traceable to NIST SRM972a.

Lyophilised calibrators were reconstituted in 1 mL LC-MS grade water (Thermo Fisher Scientific, Loughborough, UK), aliquoted and stored at -20 ^{o}C until use. Methanolic stock calibration standards were prepared by dissolving or spiking 1 mg into 1 mL of LC-MS grade methanol. The stock solutions were diluted 1:100 with methanol to form working standards. Vitamin D depleted serum was used as the blank serum calibration standard. All standards were aliquoted and stored at -20 ^{o}C until use.

Lyophilised Internal quality control (IQC) for $25OHD_3$ and $25OHD_2$ (Chromsystems, Gräfelfing, Germany), UTAK (Grifols, California, USA), calf serum pool containing endogenous $25OHD_3$ and $25OHD_2$ (Lorne Laboratories, Reading, UK) were used as quality control in each experiment. Vitamin D depleted serum (BBI Solutions, Kent, UK) was used as the base pool for spiking in-house vitamin D metabolites standards and controls. All QC material was prepared as instructed by the manufacturer. Chromsystems and UTAK controls were reconstituted in 1 mL and 5 mL of LC-MS grade water, respectively. Before introducing any new IQC material, working stocks were analysed over 20 separate runs alongside previous stocks to obtain the mean, standard deviation (SD) and coefficient of variation (CV) values.

2.4 LC-MS/MS MEASUREMENT OF SERUM 25OHD and $24,25(OH)_2D$

LC-MS/MS was performed using the Micromass Quattro Ultima Pt electrospray ionisation (ESI) tandem mass spectrometer (Waters Corp., Milford, MA, USA). MassLynx (v4.1) and QuanLynx (Waters Corp., Milford, MA, USA) was used for system control, data acquisition, baseline integration and peak quantification. The method quantified $25OHD_3$, $25OHD_2$, $24,25(OH)_2D_3$ and $24,25(OH)_2D_2$ simultaneously from a single injection. Linearity, inter/intra-assay coefficient of variation (CV) and lower limit of quantification (lLoQ) for each metabolite were assessed prior to patient sample analysis (Table 2.2).

Table 2.2: LC-MS/MS assay performance summary for the analysis of vitamin D metabolites $25OHD_3$, $25OHD_2$, $24,25(OH)_2D_3$ and $24,25(OH)_2D_2$

	$25OHD_3$	$25OHD_2$	$24,25(OH)_2D_3$	$24,25(OH)_2D_2$
Assay linearity (nmol/L)	0-250	0-250	0-25	0-25
Inter/intra-assay imprecision mean CV (%) across analytical range	≤11	≤8	≤15	≤11
ILoQ (nmol/L)	0.1	0.1	0.1	0.8

2.5 MEASUREMENT OF SERUM 1,25$(OH)_2$D

The DiaSorin LIAISON® XL 1,25$(OH)_2$D chemiluminescent immunoassay (DiaSorin Stillwater, USA) method was used to measure 1,25$(OH)_2$D in serum samples (chapter 3, 4, 5 and 7). The sandwich assay utilises a recombinant fusion protein for the capture of 1,25$(OH)_2$D molecule and a murine monoclonal antibody detection system. The assay measures total 1,25$(OH)_2$D between 12–480 pmol/L. The inter/intra-assay CV was ≤9.2% and mean assay recovery was 94 ± 2% across the analytical range. On the Vitamin D External Quality Assessment Scheme (DEQAS), the assay showed ≤8.5% bias against method-specific mean and ≤9.1% bias against all method mean and meets the performance characteristics for assay certification in DEQAS.

Certified pure standard (Isosciences 1,25(OH) 2D3: cat# S5154UNL, 1,25(OH)2D2: cat# S10002UNL) and certified pure internal standard (Supelco (Sigma) Cerilliant, 1,25(OH)D3-13C3 (25,26,27-13C3) cat# H-107) were used to calibrate the measurement of 1,25$(OH)_2$D by LC-MS/MS. Standards were spiked into vitamin D depleted serum (BBI Solutions, Cardiff, UK) to create a series of calibration standards with concentrations ranged between 30-900 pmol/L. Three pools of human sera containing 1,25$(OH)_2D_3$ and 1,25$(OH)_2D_2$ at 35, 75 and 300 pmol/L were analysed with each batch of samples as controls. ImmuTube® 1,25$(OH)_2$D LC-MS/MS kit (cat# LM1100, ImmunoDiagnostik, Bensheim, Germany) was used for sample extraction. Derivatisation reagent was synthesised by combining 40 mg of 4-[4-(Dimethylamino)phenyl]-1,2,4-triazolidine-3,5-dione (DAPTAD) (Santa Cruz Biotechnology, Dallas, TX, USA) with 60 mg of Iodobenzene diacetate (Sigma-Aldrich, Dorset, UK) in 40 mL of ethyl acetate, the mixture was placed on a magnetic mixer at room temperature for 3 hours.

LC-MS/MS was performed using the ACQUITY UPLC and Xevo TQ-S micro Mass Spectrometer (Waters Corp., Milford, MA, USA) (chapter 7). MassLynx (v4.2) and QuanLynx (Waters Corp., Milford, MA, USA) was used for system control, data acquisition, baseline integration and peak quantification. The method quantified 1,25$(OH)_2D_3$ a from a single injection. The assay measures total 1,25$(OH)_2D$ between 5-500 pg/mL. The inter/intra-assay CV was ≤9.2% and mean assay recovery was 94 ± 2% across the analytical range. On the Vitamin D External Quality Assessment Scheme (DEQAS), the assay showed ≤15% bias across the analytical range against method-specific mean and ≤9.1% bias against all method mean and meets the performance characteristics for assay certification in DEQAS. The intra and inter assay CV was <15% across the analytical range[182].

2.6 PTH, CALCIUM AND ALBUMIN ANALYSIS

Intact PTH and albumin-adjusted calcium (aCa) were analysed on the COBAS® (Roche Diagnostics, Burgess Hill, UK) platform. PTH in EDTA plasma was measured using electrochemiluminescence immunoassay (ECLIA), the inter-assay CV was ≤3.8% across an analytical range of 1.2–5000 pg/mL. Total calcium and albumin were measured based on spectrophotometric methods. The inter-assay CV for total calcium was ≤1.6% and albumin was ≤1.1% across the analytical range.

2.7 TOTAL DNA EXTRACTION

DNA was extracted from Human whole blood using the PureLink Genomic DNA kit (Invitrogen, Massachusetts, USA). In sterile 1.5 mL tubes, 200 µL of whole blood was combined with 20 µL of Proteinase K, 20 µL of rNase A and vortexed. Each sample was incubated for 2 minutes at room temperature followed by 55 °C for 10 minutes. Samples were lysed with 200 µL of lysis buffer and vortexed before incubating at 55 °C for a further 10 minutes. Samples were combined with 400 µL of ethanol and vortexed. Samples were transferred to a spin column and placed in a collection tube, both provided in the kit. The spin column was centrifuged at 2,000 x g for 1 minute. The flow through was discarded from each sample. Columns were washed with 500 µL of wash buffer one. The spin cycle was repeated and flow through was discarded. Columns were washed a second time with 500 µL of wash buffer two and centrifuged at 5,000 x g for 3 minutes. The columns were placed in sterile 1.5 mL tubes and 30 µL of nuclease free water was added. The column was incubated at room temperature for 1 minute before centrifuging at 5000 x g for 1 minute. DNA was quantified by Nanodrop and stored at -20 °C until further use.

2.8 TOTAL RNA EXTRACTION

RNA was extracted from Human whole blood using the miRNeasy mini kit (Qiagen). Equal volumes of whole blood and QIAzol lysis reagent were combined and briefly vortexed. Samples were incubated at room temperature for 5 minutes before adding 40 µL of chloroform and vortexing. Each sample was incubated at room temperature for 3 minutes before centrifuging at 12,000 x g for 15 minutes at 4 °C. The upper aqueous phase was transferred to a sterile 1.5 mL tube, combined with 525 µL ethanol and mix by pipetting. Samples were loaded onto rNeasy mini columns (supplied) and centrifuged at 8,000 x g for 15 seconds at room temperature. The flow through was discarded from each column. In a separate tube, 10 µL dNase I stock was combined with 70 µL of Buffer RDD before adding to the sample column. Columns were incubated at room temperature for 15 minutes. To each column, 700 µL of Buffer RWT was added before centrifuging at 8,000 x g for 15 seconds. The flow through was discarded from each collection tube. A second wash of 500 µL Buffer RPE was added to each column before centrifuging at 8,000 x g for 2 minutes. The flow through was again discarded before centrifuging at 14,000 x g for 1 minute to dry the column. The column was transferred to a sterile 1.5 mL tube and 30 µL of nuclease free water was added directly to the column and incubated at room temperature for 2 minutes. Columns were centrifuged at 8,000 x g for 1 minute to elute RNA. RNA was quantified by Nanodrop and stored at -80 ^{o}C until further use.

Total RNA was extracted from adherent cells in culture by removing all culture medium and washing cells with 7 mL phosphate buffered saline (PBS) (Thermo Fisher Scientific, Loughborough, UK). PBS was removed before adding 3 mL trypsin (Sigma Aldrich, Gillingham, UK). Flasks were incubated for 5 minutes at 37 ^{o}C to disassociate cells from the base. To halt trypsin-EDTA digestion, 7 mL of growth medium was added to each flask. Cell suspensions were transferred to 15 mL Falcon tubes and centrifuged at 100 x g for 5 minutes to pellet the cells. Cell pellets were resuspended in 700 µL of QIAzol lysis reagent (Qiagen). RNA was extracted using the miRNeasy mini kit as described.

2.9 DNA AND RNA QUALITY CHECK

DNA and RNA was quantified using the Nanodrop 8000 Spectrophotometer (Thermo Fisher Scientific, Loughborough, UK). Nucleic acids have an absorbance at 260 nm, proteins at 280 nm, and other contaminants e.g. phenol at 230 nm. The ratio of absorbance at 260/280 nm and 260/230 nm are the absorbance wavelengths used to assess the purity of DNA and RNA. The 260/280 and 260/230 absorbance ratios were used to evaluate the purity of the extracted DNA at 1.8 and 2, respectively. The

260/280 and 260/230 absorbance ratios used to evaluate the purity of extracted RNA were 2 and 2.2, respectively.

2.10 NON-DENATURING AGAROSE GEL ELECTROPHORESIS

Agarose gels used to visualise DNA throughout this thesis were cast at 1.5%. Molecular grade agarose (Scientific Laboratory Supplies) was dissolved in 0.5X TAE buffer by heating in a standard microwave oven for 2-5 minutes. Dissolved agarose was poured into an electrophoresis tray with comb and left to set for 30 minutes at room temperature. Once set, gels were placed in electrophoresis tanks containing 0.5X TAE running buffer. For each experiment, samples containing 6x Gel Loading Dye (New England Biolabs) were run on 1.5% non-denaturing agarose gel and stained with ethidium bromide (EtBr) (Thermo Fisher Scientific, Loughborough, UK). The gel was imaged using a UV transilluminator to determine sample integrity. Identified bands were gel extracted using the Zymoclean Gel DNA Recovery Kit (Cambridge Bioscience). Bands were exercised from the gel with sterile razor blades and added to separate sterile 1.5 mL tubes. Equal volumes of ADB buffer (supplied) was added to each gel section. Samples were incubated for 10 minutes at 55 ^{o}C to dissolve the gel. Each sample was loaded onto columns (supplied) and centrifuged at 12,000 x g for 30 seconds. The flow through was discarded. Each column was washed with 200 µL of DNA wash buffer (supplied) and centrifuged at 12,000 x g for 30 seconds. The wash step was repeated a second time. Each column was placed in a new 1.5 µL tube and 17 µL of nuclease free water was added. Samples were incubated for 1 minute at room temperature before centrifuging at 12,000 x g for 30 seconds. DNA was stored at -20 ^{o}C until sequencing.

2.11 WHOLE EXOME LIBRARY PREPARATION AND NEXT GENERATION SEQUENCING

DNA was isolated from nucleated blood cells using the Purelink genomic DNA kit (Invitrogen, Massachusetts, USA) according to manufacturer's instructions. DNA concentration and integrity was determined by the Nano-Drop 8000 Spectrophotometer (Thermo Fisher Scientific, Loughborough, UK). DNA was stored at -20 °C until use.

The TruSeq DNA Exome Library Prep Kit (Illumina, Cambridge, UK) was used to generate whole exome libraries. 150 bp PE sequencing was performed on a HiSeq 2500 (Illumina, Cambridge, UK). The adapter sequence can be found at https://support-docs.illumina.com/SHARE/AdapterSeq/Content/SHARE/AdapterSeq/TruSeq/UDIndexes.htm.

Direct sequencing for *CYP24A1* was performed using Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Loughborough, UK) and previously published primers (Table 2.3)[141]. Each PCR reaction consisted of 14 µL master mix (10 µL Phusion HF Master Mix with High Fidelity Buffer and 4 µL nuclease free water), 2 µL of primer pair (10 µM) and 100 ng of genomic DNA. A thermocycler was used for PCR cycling (Table 2.4).

Table 2.3: ***CYP24A1*** **forward and reverse primer oligonucleotides**[141]

Exon	Forward Primer (5'-3')	Reverse Primer (5'-3')
CYP24A1 Exon 1_1	AGGGCATGCTCTGTCTCC	AAGGCAGGAGGATGGGG
CYP24A1 Exon 1_2	CCCTCTTTGCTTCCTTTTCC	ATGTCGGGGAGGGTTTG
CYP24A1 Exon 2	GAGGAAGGAGGCGGGAG	CCGTCAGGCTCATCAGGTC
CYP24A1 Exon 3	GCTGGAGTATTTCTGCATCTCC	CCACCAATATCCCTATGTCCC
CYP24A1 Exon 4	ATGCGATGTAGCAAGACCTG	TGCCTGTTTACAAAAGAGTTGTC
CYP24A1 Exon 5	GGCATAGAATTGAGTCTTTAATAAC C	TGGGAATCACTGTGAAGTTCTG
CYP24A1 Exon 6	CCTCTTCCAGAACGAACATTG	TGAAGCTCCAGACACGGG
CYP24A1 Exon 7	TGCAAGAAGGAGTTTGGACTG	TGAATCCCAGTGAAATGAATG
CYP24A1 Exon 8	TTGCAGAATAAGGTGGTGGG	TAATTAGCTAGGGGAAGCCG
CYP24A1 Exon 9	AATCTGCATTCCCATTGACAC	CAAAGTCTAGGGAGATCTGGTG
CYP24A1 Exon 10-11	CAATTTTGCCATTCAAAGGTC	GCTCATCCCTCGTCATTCTC
CYP24A1 Exon 12_1	CCGGAAAGCAAACTTCAAAC	AACAAAATAATGCCCCAGTG
CYP24A1 Exon 12_2	GCTGGGAGTAATACTGACAATCC	TATTGCATGCATTTCTGTGC
CYP24A1 Exon 12_3	TTAGGATCTGTGGTGCAGGG	TTTGTGATATAGGGCTTGTAGGC

Table 2.4: PCR Cycles for DNA amplification

Temperature (°C)	Time	Cycles
98	30 sec	1
98	10 sec	35
60	30 sec	
72	30 sec	
72	10 mins	1
4	∞	∞

PCR products were sequenced using the same forward primers used for amplification. Samples were sent to Eurofins MWG Operon (www.eurofinsgenomics.eu) for direct sequencing. Sequences were aligned to the Human genome using the basic local alignment search tool (BLAST) provided by the National Centre for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/Blast.cgi). For direct sequencing analysis, FASTA reads were aligned with *CYP24A1* transcript variant 1 (NM_000782.5) using sequence alignment editor (BioEdit).

From all variants identified through WES, variants with <10 reads were excluded, plus any silent synonymous variants. The Human Genome (v38) was used as a control reference against the patient WES sequencing data. The variant allele frequency is a percentage that refers to the number of variant reads divided by the total reads. Allele frequency, indicating the relative frequency of an allele at a particular locus in a population, was used to identify any heterozygous (~0.5) or homozygous (~1) variants.

2.12 BIOINFORMATICS

Mutation taster (www.mutationtaster.org), ClinVar (www.ncbi.nlm.nih.gov/clinvar) and dbSNP (www.ncbi.nlm.nih.gov/projects/SNP) were used to determine SNV disease causing potential. RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) was used to determine *CYP24A1* mRNA structures. The RNAfold algorithm uses thermodynamic structure predictions based on the minimum free energy (MFE) generated by the nucleotide composition of the input sequence. RBPmap (www.rbpmap.technion.ac.il/) was used to determine whether the 5' and 3' UTR mutations impaired protein-RNA interaction. RBPmap employs an algorithm for mapping protein binding motifs on RNA transcripts whilst considering the clustering propensity of the motif plus the overall tendency of the regulatory region to be conserved. miRDB (http://mirdb.org/miRDB/index.html) was used to elucidate whether the 3' UTR mutations altered or introduced miRNA recognition elements. miRDB is a database for miRNA target prediction and functional annotation. All mRNA targets in miRDB were predicted by miRTarget, which was developed by analysing thousands of miRNA-mRNA functional interactions from next generation sequencing experiments.

2.13 DIGITAL PCR

Total RNA was quantified by density measurement after separation by agarose gel electrophoresis with ethidium bromide (Thermo Fisher Scientific, Loughborough, UK) staining. Equal amounts of RNA were reverse transcribed using the high-capacity RNA to cDNA kit (Thermo Fisher Scientific, Loughborough, UK). *CYP24A1* expression was validated in triplicate using a TaqMan gene expression assay (Thermo Fisher Scientific, Loughborough, UK). Digital PCR was performed on the QuantStudio 3D Digital PCR System using the GeneAmp PCR System 9700 (Thermo Fisher Scientific, Loughborough, UK). PCR chips were imaged on the QuantStudio 3D Instrument, which assesses raw data and calculates the concentration of the cDNA sequence targeted by FAM and VIC labelled probes by Poisson distribution. For in-depth analysis, the QuantStudio 3D Analysis Suite was used to report the data as copies per microliter.

2.14 ISOLATION OF MONONUCLEAR CELLS

Patient whole blood samples were diluted with equal volumes of PBS. Diluted samples were gently overlaid onto 3 mL of Ficoll-Paque PLUS (GE Life Sciences, Buckinghamshire, UK) to avoid mixing. Samples were centrifuged for 30 minutes at 400 x g (without break) before discarding the upper most layer. The mononuclear cell layer was removed and washed with 6 mL of PBS. Cells were resuspended by gentle pipetting and centrifuged at 400 x g for 15 minutes. The supernatant was discarded and a second PBS wash was performed, followed by centrifugation at 500 x g for 10 minutes. The remaining pellet was used for western blot analysis.

2.15 WESTERN BLOT

Mononuclear cells were resuspended in 200 µL M-PER lysis buffer (Thermo Fisher Scientific, Loughborough, UK) with complete protease inhibitor cocktail (Roche, Burgess Hill, UK) before incubating on ice for 30 minutes. Samples were pelleted by centrifugation at 13,000 rpm at 4 °C for 7 minutes. Protein concentrations were determined from the supernatants using the BCA protein assay system (Thermo Fisher Scientific, Loughborough, UK) according to manufacturer’s instructions. Samples were denatured with Laemmli buffer at 95 °C for 5 minutes. Proteins were then separated on a 4-12% gradient SDS-PAGE gel (Thermo Fisher Scientific, Loughborough, UK) and transferred onto immobilon PVDF (Millipore, Watford, UK) to blot. Protein transfer was conducted at 200 mA at 4 °C for 1 hour. The transferred membrane was blocked in 0.5% milk (Sigma Gillingham, UK) for 30 mins at ambient temperature. A monoclonal antibody for CYP24A1 (Sigma, #WH0001591M7) was used to probe membranes for 48 hours at 4 °C. Actin was included as a loading control in all samples. The membrane was washed 4 times for 15 minutes with TBST buffer at ambient temperature. IRDye labelled secondary antibodies at 1:10,000 were added for a 2 hour incubation period to detect primary antibodies. The TBST wash was repeated four times before the proteins were visualised using the Odyssey infrared system (LI-COR, Cambridge, UK).

2.16 ELISA

Plasma mitochondrial CYP24A1 expression was additionally analysed by enzyme-linked immunosorbent assay (ELISA) were performed in duplicate according to manufacturer's instructions (Cusabio, Houston, USA). To each well, 100 µL of standard or sample was added and incubated at 37 °C for 2 hours. The liquid was then removed from each well and 100 µL of biotinylated anti-human CYP24A1 antibody was added before incubating at 37 °C for 1 hour. Each well was washed 3 times with wash buffer before adding 100 µL streptavidin labelled horseradish peroxidase (HRP). Plates were incubated at 37 °C for 1 hour. Wells were washed five times with wash buffer and blotted to remove excess liquid. Once blotted dry, 90 µL of TMB substrate was added before incubating at 37 °C for 15 minutes. The reaction was terminated with 50 µL of stop solution. Sample absorbance was detected at 450 nm using a plate reader (Thermo Fisher Scientific, Loughborough, UK). The detection range of this CYP24A1 sandwich assay was 7.8 pg/mL-500 pg/mL and the sensitivity was 1.95 pg/mL. The inter/intra-assay CV was ≤10% and the mean assay recovery was 86 ± 6% across the analytical range. Sample absorbance was detected at 450 nm using a plate reader. Concentrations were calculated using a four-parameter logistic (4PL) curve ranging 0-500 pg/mL.

2.17 SHAPE REAGENT TREATMENT

SHAPE reagent, NMIA, was used for mRNA secondary structure analysis. In a chemical fume hood, 247 mg of 2-methylpyridine-3-carboxylic acid (Thermo Fisher Scientific, Loughborough, UK) was dissolved in 500 µL of anhydrous DMSO (Thermo Fisher Scientific, Loughborough, UK) in a sterile 2 mL tube and vortexed briefly at room temperature. In a separate sterile 2 mL tube, 324 mg of 1,1'-Carbonyldiimidazole (Thermo Fisher Scientific, Loughborough, UK) was dissolved in 500 µL of anhydrous DMSO. The second solution was slowly added to the 2-methylnicotinic acid solution over 5 minutes. The combined solution was vortexed at room temperature, occasionally opening the cap to allow gas evolution to escape. This solution was used as a 2 M stock solution containing a 1:1 ratio of NAI and imidazole as a by-product. The solution was aliquoted and kept frozen at -80 ^{o}C until use. The stock NMIA reagent was thawed at room temperature before use. Reaction buffer consisted of 895.5 mL molecular grade water, 40 mL of 1M HEPES, 100 mL of 1M KCL and 500 µL of 1M $MgCl_2$.

Patient whole blood was lysed with red cell lysis buffer for 10 mins before centrifugation at 500 x g for 5 minutes at room temperature. Supernatant was discarded and pellet used for SHAPE reagent treatment/ as a negative control.

Cells were resuspended in this 4.75 mL of 1X reaction buffer plus 250 µL of SHAPE reagent for 5 minutes to acylate unpaired nucleotides. Negative control cells from the same patient were resuspended in 5 mL of 1X reaction buffer minus SHAPE reagent. Samples were then centrifuged at 500 x g for 3 minutes. The supernatants were discarded and cell pellets were resuspended in molecular grade water. The cells were then centrifuged at 500 x g for 3 minutes. RNA was extracted from both treated and untreated cell pellets as previously stated in section 2.8.

2.18 DIMETHYL SULFATE SELECTIVE 2' HYDROXYL ACYLATION ANALYSED BY PRIMER EXTENSION LIGATION MEDIATED PCR (DMS/SHAPE-LMPCR) ANALYSIS OF *CYP24A1*

For secondary structure determination, the 3' UTR region (1,316 nt) of *CYP24A1* was divided into 3 smaller fragments for individual amplification and chemical probing. The three fragments of the 3' UTR are referred to as CE-R2, CE-R3 and CE-R4 throughout this study. Primers were designed for each 3'UTR fragment plus the full length 3' UTR region (full 3' UTR forward, full 3' UTR reverses, fragment CE-R2, fragment CE-R3 and fragment CE-R4). Additional primer sets containing fluorescent dye tags (VIC or NED) were designed for each 3'UTR fragment plus the full length 3'UTR region, to allow detection by capillary electrophoresis (CE) (Table 2.5).

Table 2.5: *CYP24A1* 3' UTR CE primers. Primers specific to the full length 3' UTR forward sequence, 3' UTR reverse sequence, CE-R2 (1,056 bp), CE-R3 (789 bp) and CE-R4 (459 bp). The primer positions were determined to ensure the full length of the 3;UTR sequence could be analysed.

Primer	**Sequence (5' – 3')**
3' UTR-Forward	ATGGTGGTATTTGCTAACATCATATC
3' UTR-Reverse	TCAAATATTTAAAAACTTTTATTTTTACTTTCAGAG
CE-2	ATATGCACAAATAAACCATCTGTAAACACAA
CE-3	GCACCACAGATCCTAAATCAAGTACTGCAA
CE-4	TCAATGCAGGAAGAACGCAATTTCATGGGAG

SHAPE treated and non-treated RNA extracted from patient samples (see section 2.17) underwent reverse transcription to obtain *CYP24A1* cDNA. RNA was first resuspended in 7 μL nuclease free water and combined with 0.5 μL 10 mM *CYP24A1* gene specific primer (IDT). Samples were heated at 65 °C for 5 minutes. Reverse transcription buffer, Rnase OUT (40 U/uL) (Thermo Fisher Scientific, Loughborough, UK) and SSIII (200 U/uL) (Invitrogen, Massachusetts, USA) were added to each sample before incubating at 50 °C for 60 minutes, then 75°C for 5 minutes to inactivate the SSII enzyme.

The reverse transcription was phenol chloroform extracted following manufacturer's instructions. cDNA was extracted by ethanol precipitation overnight at -80 °C. The cDNA pellet was resuspended in nuclease free water, 1 μL mly-HBLPCR5'ss DNA linker (50 μM) (New England Biolabs, Massachusetts, United States) and denatured at 95°C for 3 minutes. Samples were then cooled to 60 °C and combined with 10 μL ligation buffer (New England Biolabs, Massachusetts, United States) and 1 μL T4 ligase (New England Biolabs, Massachusetts, United States) before leaving at room temperature overnight to ligate.

Ligated samples were purified using MicroSpin S-200 HR columns (GE Healthcare Life Sciences) under manufacturer's instructions followed by ethanol precipitation at -80 °C for 2 hours. Ligated cDNA was dissolved in 15 μL nuclease free water. PCR reactions contained 1 μL of cDNA, 10 μM reverse transcription primer, 10 μM Mly-HBLPCR-RT-R and 7.5 μL of 2 x HIFI master mix (New England Biolabs, Massachusetts, United States). To test the optimal cycle number, the PCR reaction was initially set to 95 °C for 3 minutes followed by 98 °C for 20 seconds, 68 °C for 15 seconds and 72 °C for 40 seconds for 35 cycles. Test samples (5 μL) taken at 25, 30 and 35 cycles were analysed on agarose to determine the optimal cycle number.

Once the optimal cycle number was identified (30 cycles), a second PCR reaction was performed to incorporate VIC fluorescently labelled primers (Thermo Fisher Scientific, Loughborough, UK). To 13.6 µL ligated DNA, 0.7 µL of Mly-HBLPCR-RT-R, 0.7 µL VIC gene specific primer and 15 µL HIFI master mix was added. PCR conditions were repeated using the optimal cycle number determined in the previous step.

To remove the florescent primer, 0.5 µL of Exo1 (Thermo Fisher Scientific, Loughborough, UK) and 1 µL of rSAP (New England Biolabs, Massachusetts, United States) were combined with the PCR product and incubated at 37 ˚C for 15 minutes followed by 80 ˚C for 15 minutes. Samples were purified using S200 columns as manufacturer instructions followed by ethanol precipitation at -80 ˚C overnight.

Double stranded DNA was dissolved in 44 µL nuclease free water and combined with Mly restriction enzyme 10 U (New England Biolabs, Massachusetts, United States) and 5 µL 10 x NEB Buffer (New England Biolabs, Massachusetts, United States) followed by incubation at 37 ˚C for 2 hours. An additional 1 µL of Mly restriction enzyme was added before a second 2-hour incubation at 37 ˚C. Samples were purified using S200 columns as manufacturer instructions followed by ethanol precipitation at -80 ˚C overnight.

Dideoxy sequencing lanes were prepared prior to analysis. PCR reactions containing 0.5 µL of gene specific non-labelled full 3'UTR forward primer and reverse primers, template DNA (CE-R2, CE-R3 and CE-R4 with fluorescent NED tags), 10 µL of master mix red and 9.5 µL nuclease free water were amplified. Following agarose gel analysis, fragments were extracted and purified. Sequencing lane samples were fluorescently labelled using NED probes (Thermo Fisher Scientific, Loughborough, UK). To 5 µL of sample, 5 µM of NED primers were added plus 2 µL of 10 x React buffer (Thermo Fisher Scientific, Loughborough, UK). Samples were heated to 95 °C for 3 minutes and immediately placed on ice. Klenow DNA polymerase I 2 U/µL (New England Biolabs, Massachusetts, United States) and dideoxy mix A were combined with each sample and incubated at 37°C for 20 minutes to allow extension. dNTPs were added to samples before repeating the 20 minutes incubation at 37 °C. Finally, 34 µL of nuclease free water was added and samples were ethanol precipitated overnight a -80°C.

Both sequencing lane samples and SHAPE treated/control samples were combined with 10 µL Hi-Di formamide (Thermo Fisher Scientific, Loughborough, UK). Sequencing lane samples were further diluted 1:6 with formamide plus 0.2 µL Rox1 1 kb ladder (New England Biolabs, Massachusetts, United States). To each well, 5 µL of sequencing lane samples were added plus 5 µL of each SHAPE treated sample. The plate was denatured by incubating at 95 °C for 3 minutes.

Samples were size fractionated on 50 cm capillary array with POP7 matrix using a 3730XL DNA Analyzer (Applied Biosystems, Massachusetts, USA) with set parameters (voltage 15kV, T= 66 °C, injection time = 10 seconds). Fragments were assessed using the Peak Scanner x1.0 software (Applied Biosystems, Massachusetts, USA).

The difference in band intensity between NMIA treated and non-treated control samples was calculated. Dideoxy sequencing lanes were used to identify the nucleotide identity of each band. QuSHAPE software was used to analyse CE data. For each sample, NMIA treated RNA, DMSO control RNA and ladder sequencing data is aligned by QuSHAPE[183]. QuSHAPE software scales the NMIA to the DMSO negative control signal and subtracts the DMSO peak integrated values from the NMIA reagent peaks to normalise nucleotide-resolution reactivity for each RNA position using a box normalisation-based algorithm. Reactivity was normalised based on the 2/8% rule. The top 2% were regarded as outliers and removed from the average. The next 8% most intense nucleotide bands were averaged. All nucleotide intensities including outliers were divided by the 8% average to normalised SHAPE reactivity.

Gaussian integration was performed for all peaks in the NMIA and DMSO control channels to confirm our findings. Normalisation generated ranges 0-2 with 0 indicating no SHAPE reagent reactivity, ranging to highly reactive positions (Table 2.6). Nucleotide reactivity of >0.85 is suggestive of flexible regions that are highly reactive with the NMIA reagent. Nucleotide reactivity of 0 is suggestive of highly constrained regions. RNAthor software then utilises QuSHAPE data analysis results to display each nucleotide reactivity in a bar chart with each nucleotide reactivity.

Table 2.6 SHAPE reactivity corresponding to nucleotide normalisation value. Reactivity of each nucleotide corresponds with the level of interaction with SHAPE reagent NMIA and subsequent flexibility at each location within the RNA sequence.

Nucleotide Normalisation Value	**SHAPE Reactivity**
0	No reactivity
0-0.4	Unreactive
0.4-0.85	Moderately reactive
>0.85	Highly reactive

The RNAThor structure server website was used to determine the *in vivo* SHAPE-constrained RNA structure. Normalised and quantified reactivates were input as pseudo-energy constraints plus the RNA sequence. The RNA was then folded at 25 ^{o}C.

2.19 CRISPR CAS 9 MODIFICATION OF HEK293T CELL LINE

2.19.1 DIRECT SEQUENCING OF HEK293T CELLS

Genomic DNA was extracted from HEK293T cells using the PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific, Loughborough, UK) as described previously. Cells were screened for mutations in the 3' UTR of *CYP24A1*. Mutation screening was conducted on PCR-amplified DNA fragments of the 3' UTR as previously described (Table 2.3). Direct sequencing results for HEK293T *CYP24A1* 3' UTR were aligned to the Human genome using BLAST and FASTA reads with *CYP24A1* transcript variant 1 (NM_000782.5) using BioEdit as described previously.

2.19.2 CRISPR-CAS9 TRANSFECTION

Plasmid guide RNA (gRNA) and Cas 9 constructs based on the vector GE100002 and donor 100 bp ssDNA oligos (Origene, Maryland, USA) were designed to introduce specific SNVs into the 3' UTR sequence of *CYP24A1* (Table 2.7). The SNVs introduced corresponded with those observed in our patient cohort.

Table 2.7: gRNA/Cas 9 constructs and template ssDNA oligos (Origene, Maryland, USA) designed to target and modify specific regions of *CYP24A1* at two locations withing the 3'UTR of *CYP24A1* (c.1993 C>T and c.2658 C>G).

Target Mutation Location 5'–- 3'	c.1993 C>T	c.2658 C>G
gRNA sequence	CATTCTACAGGGTTCACTGC	TATGTGTGTGTGGGTTCTAA
Forward Primer Sequence	CCAACTTAGGGAAGCGGACTGAG TGCTGGGATCCAAGGCATTCTACA GGGTTCACTGCAGGTTTACACTTC ACCTGTGTCAGCACCATCTTCAGG TGCTTAGAATGGCCT	TGTGTGTGTGTCTGTGTGTGTGTG TGCGTGTATGTGTGTGTGGGTTCT AACGGTAATTTGCCTCAGTCATTT TTTTAATATTTGCAGTACTTGATTT AGG
Reverse Primer Sequence	AGGCCATTCTAAGCACCTGAAGAT GGTGCTGACACAGGTGAAGTGTAA ACCTGCAGTGAACCCTGTAGAATG CCTTGGATCCCAGCACTCAGTCCG CTTCCCTAAGTTGG	CCTAAATCAAGTACTGCAAATATT AAAAAAATGACTGAGGCAAATTAC CGTTAGAACCCACACACACATACA CGCACACACACACACAGACACAC ACACA

Single oligos were diluted in TE buffer to concentration of 0.3 μg/uL. To anneal forward and reverse oligos, 2 μg of each were combined with 50 μL of annealing buffer (10 nM Tris, 1 mM EDTA, 50 mM NaCl), heated at 95 °C for 2 minutes then allowed to cool to 25 °C over 1 hour.

HEK293T cells were seeded at 3.5 x 10^5 cells/well into 6 well plates and grown overnight to 70% confluency. In a sterile 1.5 mL tube, 125 μL Opti-MEM (Thermo Fisher Scientific, Loughborough, UK) was combined with 7.5 μL Lipofectamine 3000 (Thermo Fisher Scientific, Loughborough, UK) and vortexed. In a second sterile 1.5 mL tube, 125 μL Opti-MEM was combined with 1 μg gRNA/Cas9 and 1 μg DNA oligo. The Lipofectamine 3000 solution was added to the RNA/Cas 9 solution, mixed by pipetting and incubated at room temperature for 15 minutes before adding dropwise to cells. Cells were incubated at 48 hours at 37 °C in 5% CO_2. Control cells were treated with Lipofectamine 3000 minus gRNA/Cas9 vector or oligos. After 48 hours cells were split 1:10 with fresh medium.

2.19.3 CRISPR-CAS9 SINGLE CELL SELECTION

CRISPR treated cells and controls were seeded into a single well (A1) of a 96 well plate at 2 x 10^4 cell/well. Remaining wells in column 1 vertically below A1 were filled with 100 μL of fresh medium. From the cell suspension well (A1), 100 μL was serially diluted down column 1 using the same pipette tip. From the final well, 100 μL was discarded. The remaining empty 11 columns were filled with 100 μL fresh media. Serial dilution was performed horizontally across the plate by transferring 100 μL from column 1-12. From the final column, 100 μL was removed from each well and discarded. All wells were brought to a total volume of 200 μL with fresh media. Plates were incubated at 37 °C for 1 week. Wells were monitored daily and those that contained single cell colonies were identified. Once >50% confluent, single cell colonies were expanded into 24 well plates.

2.19.4 RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP)

The restriction enzyme TspGWI (EURx, Gdansk, Poland) was used to assess the outcome of CRISPR-Cas9 base editing. Genomic DNA was extracted from each single cell colony as previously described. Forward and reverse primers were used to amplify the *CYP24A1* region of interest containing both the TspGWI and CRISPR-Cas9 target sites (Table 2.3).

RFLP reactions consisted of 1 µg DNA, 5 µL 10x Buffer TspGWI (EURx, Gdansk, Poland) and 1 U TspGWI enzyme. Reactions were made up to 50 µL with nuclease free water. Samples were incubated at 70 ^{o}C for 2 hours. Reactions were inhibited by addition of 2.1 µL EDTA pH 8.0 [0.5 M]. Digested samples were analysed by non-denaturing agarose gel and sequenced as previously described to confirm SNV presence.

2.20 SINGLE MOLECULE FLUORESCENCE IN SITU HYBRIDISATION (smFISH) IMAGING

For the simultaneous detection of *CYP24A1* and a housekeeping gene, *RNA Polymerase II Subunit A (POLR2A),* this study used the Stellaris predesigned smFISH probes for *POLR2A* plus designed probes that accurately hybridise mRNA transcripts for *CYP24A1* (Table 2.8). *POLR2A* was selected as an appropriate housekeeping gene due to the similar low relative abundance to *CYP24A1 (*copy number of 20-100 per cell). The parameter design for *POLR2A* probes was against the coding sequence of NM_000937.4 (NCBI gene ID: 5430, nucleotides 387-6299). The inclusive probe set designed to detect *CYP24A1* variants was produced under the sequence NM_000782.5 (NCBI gene ID: 1591, nucleotides 1-643, 1842-1962). *POLR2A* was requested assigned to Quasar 570 dye and *CYP24A1* assigned to Quasar 670 dye to distinguish between the two mRNA transcripts.

Table 2.8: Oligonucleotide sequence for *CYP24A1* Stellaris probes (n=48)

***CYP24A1* Probe Sequence (5' to 3')**
aggtgtcaaggagggtaga
aaggctgacctctagggtc
aaagctgggagtcctcctg
gtctggggacctcttgaaa
cgacagcctcagagcattg
aggatgctcgacgctgcac
ggacaggcagcagaaagga
tgtgatggagcggggtgag
agtcggggcttaacgattc
ttggtacgaggtgctagtg
ctgatgggggagctcatgg
gtagatgtcaccagtctcg
ggctgaggggacgtgtacg
cgttctgagtctcgccac
tccagagaatctgcagcag
tgtcgtgctgtttcttgag
acccaacttcatgcggaaa
caggtgcaccgactcaaag
tccagcaggcatggcgag
tctcggtgcggtacagcg
tccacggtttgatctccag
tttgcggtagtcgcgatag
gcccataaaatcggccaag
ttcatcacagagctcatct
acaagtcttcaacgtggcc
gaccatttgttcagttcgc
cacgaggcagatactttca

gatgaagttcacagcttca
catcattgttttgatggcc
catcatcctcccaaacgtg
ttgtgcagctcgactggag
cagaccttggtgttgaggc
accggttgtcgatacaagc
ggctgctgagaatacttct
ctctgtgacagcagcatac
gactgtttgctgtcgtttc
cacttggggattacgggat
gcacctgattctcaggtaa
ttcctcaaatcttctgccc
caggcttttaaatacggca
agcctcatagattctttca
gtaaatggtacactcggcg
tgttgccttgtcaagagtc
gggtaaagcatattcaccc
tgtcgtatttgcggacaat
tgtagcatctcaacaggct
aaacgcgatggggagttcc
tgaggcgtattatcgctgg

HEK293T cells were grown on 18 mm round glass cover slips in 12 cell culture plates (Thermo Fisher Scientific, Loughborough, UK) until 80% confluent in conditions previously described. Medium was aspirated and cells were washed with 1 mL PBS. Cells were fixed with 1 mL fixation buffer and incubated at room temperature for 10 minutes. The PBS wash was repeated twice. Cells were permeabilised by immersing in 1 mL 90% (vol/vol) ethanol for >1 hour between 2-8 ^{o}C.

smFISH probes for *CYP24A1* and the RNA Polymerase II Subunit A (*POLR2A)* housekeeping gene (Stellaris, LGC Biosearch Technologies, Middlesex, UK) were allowed to reach room temperature before vortexing and centrifuging briefly. To prepare a 125 nM working probe solution, 1 µL of probe stock solution was added to 100 µL hybridisation buffer before vortexing and brief centrifugation.

The ethanol was aspirated from each coverslip and cells were washed with 1 mL wash buffer A (Stellaris, Surrey, UK) and incubated at room temperature for 5 minutes. A humidified chamber was created by lining a 150 mm culture plate with water saturated paper towel and a single layer of Parafilm. For each coverslip, 100 µL of the working probe solution was pipetted into the humidified chamber and the coverslip was placed cell side down onto the droplet. The humidified chamber was sealed with Parafilm and incubated at 37 ^{o}C for 4 hours protected from sunlight.

Coverslips were placed into fresh 12-well plates containing 1 mL of wash buffer A and incubated at 37 ^{o}C for 30 minutes. The wash buffer was aspirated and discarded. The coverslips were stained with 1 mL of DAPI nuclear stain (Wash Buffer A consisting of 5 ng/mL DAPI) to counterstain the nuclei. Plates were incubated at 37 ^{o}C for 30 minutes. DAPI was aspirated and discarded. Coverslips were washed with 1 mL of wash buffer B and incubated at room temperature for 5 minutes.

Coverslips were mounted with 15 µL Vectashield Mounting Medium (Thermo Fisher Scientific, Loughborough, UK, Loughborough, UK) onto a microscope slide cell side down. Excess mounting medium was wicked away and the coverslip perimeter was sealed with clear nail polish. Cells were imaged using an Axioplan 2 Zeiss, Elyra PS-1 microscope at 100x oil immersion objective with appropriate filter sets for the fluorophores. Cells were analysed following the publicly available method for mRNA counting using ImageJ[184]. This Radial-Symmetry-FISH (RS-FISH) software is a robust and rapid method for accurately detecting single molecule spots. Individual mRNA spots were located by taking a maximum projection z-stack containing all probe channels.

2.21 STATISTICAL ANALYSIS

Descriptive statistics, scatter plots, ROC and LOWESS curves were constructed and analysed by Statistical Package for the Social Science (SPSS) version 22.0.0.1 (IBM, New York, USA) and GraphPad Prism 7 (GraphPad, San Diego, CA, USA). Univariate and multivariable linear regression analyses and one-way ANOVA were used to estimate associations. LOWESS curve fitting was used to explore non-linear relationships between variables. Kruskal-Wallis independent analysis and Spearman's rho were used to establish associations in non-parametric variables. Frequency distribution histograms of the data were visually examined and checked for transcriptional and pre/post analytical errors before exclusion for statistical analysis. Confidence interval (CI) was established at 95% of the population. Circannual rhythm analysis was performed by population-mean cosinor analysis, based on cosinor-fitting equation $y = MESOR + Amplitude \times \cos(Frequency(x) + acrophase)$. Midline estimate statistic of rhythm (MESOR), defined as the rhythm-adjusted mean value. Acrophase is the difference (time) between MESOR and peak value in the cosine curve.

Throughout this thesis, for statistical tests e.g., t-tests, were considered statistically significant when $p = < 0.05$.

CHAPTER 3: THE RELATIONSHIP BETWEEN ACTIVE AND CATABOLIC VITAMIN D METABOLITE RATIOS ASSOCIATED WITH PTH REGULATION AND CALCIUM HOMEOSTASIS

3.1 INTRODUCTION

The vitamin D pathway is a dynamic system and its functional role in bone health and other diseases is the subject of intense research. Associations of vitamin D deficiency with a wide spectrum of disease states have drawn attention from the scientific community and increasing awareness of the general population. Despite vitamin D deficiency being a global public health concern, the approach through improving vitamin D status by supplementation, dietary intake and increased sunlight exposure, has resulted in mixed outcomes[185–187]. The contradictory evidence has prompted studies on the metabolites of vitamin D. The most abundant metabolite in circulation is 25OHD, which exists in two major forms: $25OHD_3$ (cholecalciferol) and $25OHD_2$ (ergocalciferol). Measurement of serum total 25OHD ($D_3 + D_2$) is the best indicator of vitamin D status; concentrations ≤30 nmol/L and between 30–50 nmol/L are defined as deficient and insufficient, respectively by the U.S Institute of Medicine (IOM)[102] and the Royal Osteoporosis Society (ROS)[188]. The metabolite $1,25(OH)_2D$ is synthesised by the hydroxylation of 25OHD through the actions of CYP27B1 produced in the renal tubules. The $1,25(OH)_2D$ metabolite is the most biologically active form of vitamin D and circulates in pmol/L concentration; it controls intestinal absorption of calcium and phosphate, stimulates osteoclast activity, and aids regulation of PTH. Although $1,25(OH)_2D$ is derived from 25OHD, there is no direct correlation in serum concentrations between the two vitamin D metabolites except in patients with chronic kidney disease (CKD)[189], where a greater association is observed, dependent upon the severity of the renal impairment. The lack of a direct relationship, despite their close proximity in the metabolic pathway, is due to the tight regulation by hydroxylation enzymes CYP27B1 and CYP24A1. *CYP24A1* produces 24-hydroxylase that converts 25OHD and 1,25(OH)$_2$D into 24,25(OH)$_2$D and 1,24,25(OH)$_2$D respectively. The transcription of the *CYP24A1* gene is stimulated by the phosphate-regulating hormone FGF23 and suppressed by PTH. Increased *CYP24A1* and FGF23 plus suppressed PTH results in an increase in serum 24,25(OH)$_2$D through 25OHD hydroxylation. Previous studies have described a positive concentration-dependent relationship between serum 24,25(OH)$_2$D and 25OHD[150] controlled by FGF23, PTH and subsequent CYP24A1 activity.

Prior work reported a patient, presenting with HCINF1, who was diagnosed with biallelic *CYP24A1* mutations resulting in the inability to produce 24,25(OH)$_2$D from 25OHD. This patient presented with elevated serum 1,25(OH)$_2$D and a persistent state of hypercalcaemia. The use of 25OHD:24,25(OH)$_2$D vitamin D metabolite ratio (VMR) can be a valuable tool in identifying such pathological conditions resulting from impaired CYP24A1 function[125,190]. The close correlation between 24,25(OH)$_2$D and 25OHD is less susceptible to season fluctuations making this ratio more reliable in assessing CYP24A1 function that VMRs of 1,25(OH)$_2$D, the functional metabolite of vitamin D. The use of VMR can provide an assessment of the vitamin D catabolic status and insight into CYP24A1 function in patients with suspected HCINF1; thus allowing appropriate vitamin D supplementation and CYP24A1 loss-of-function identification[191].

This chapter investigates the interpretation of serum 25OHD and 1,25(OH)$_2$D concentrations that incorporates 24,25(OH)$_2$D values. Using data from a large cohort of young healthy adults as the reference population, the intricate relationships between active and catabolic forms of vitamin D metabolites, and the influence on PTH was determined.

3.2 CLINICAL SAMPLES

The LC-MS/MS method for measuring 25OHD and 1,25$(OH)_2$D to perform VMR relative ratio was developed using samples from an MOD study investigating vitamin D status and fracture risk during basic training. The study received ethics approval from the UK Ministry of Defence Research Ethics Committee (MODREC 165/Gen/10 and 692/MoDREC/15 ClinicalTrials.gov Identifier[181]) and was conducted in accordance with the Declaration of Helsinki (2013). The characteristics of the subjects included in this chapter are shown below (Table 3.1). In total, 2,252 new British Army recruits at the start of phase one training volunteered for the study. Written informed consent was obtained from all study participants, and each participant was required to complete a detailed health questionnaire including medical history and the use of dietary supplements. All recruits undertook physical and cognitive testing, and a detailed medical examination prior to joining the army. The British Army entry requirements restrict individuals with chronic medical conditions; therefore, the study population represents a medically screened, disease-free, and physically fit population. In the analysis, individuals who reported the use of calcium and vitamin D supplements (including multivitamins and cod liver oil) and participants who reported injury and illness prior to recruitment were excluded. Additionally, participants with conditions such as being underweight, eating disorders, or those with a history of bone fracture were excluded. A total of 940 participants were included in the final statistical analyses. The majority of participants were from the Caucasian population (92.9%), with a minority from a diverse ethnicity (Asian 1.6%, Black 1.7%, Chinese 0.1%, mixed 3%, others 0.7%).

Table 3.1: Baseline characteristics of the subjects included in this chapter. *Data shown as the mean with ± SD in square brackets unless otherwise stated.

	Male	**Female**
No. of participants	652	288
Mean age, years [range]	21.7 [18–32]	22.1 [18–32]
Height, m [± SD]	1.77 [6.4]	1.66 [5.9]
body mass, kg [± SD]	75.9 [9.8]	64.7 [7.5]
Body mass index (BMI) [± SD]	24.1 [2.6]	23.4 [3.3]
Total body bone mineral density (BMD) (g/cm^2) [± SD]	1.24 [0.10]	1.16 [0.09]

Venous blood samples were obtained from the participants at the start of 14-week long basic military training. Sample collections were scheduled on a monthly basis to balance the seasonal variations because it is known that vitamin D status fluctuates during the year for individuals who live in the Northern hemisphere. Each recruitment intake comprised, on average (range), 86 (43–120) participants. Blood samples were collected into serum gel separator tube and EDTA plasma container (BD Vacutainer). Samples were centrifuged immediately after collection at 3,000 × g for 10 minutes. Plasma/serum layers were aliquoted into a separate polystyrene tube and stored at −20 °C until analysis. All samples were anonymised at the point of access.

<u>3.3 RESULTS</u>

Results from 940 healthy adult participants were included in the data analysis of this chapter. Statistical analyses on 25OHD and 24,25(OH)$_2$D concentrations was performed on the respective total values (D_3 + D_2), the distributions were untrimmed and no outliers were removed (Table 3.2). The 25OHD$_2$ metabolite was detectable in 57.8% of the subjects with a mean (range) of 4.2 nmol/L (0.6–29.1). The 24,25(OH)$_2$D$_2$ metabolite was detectable in 0.4% of the subjects with a mean (range) 1.5 nmol/L (1.2–1.8).

Table 3.2: Distribution of biochemical measurements performed in the study.

Profile	mean	SD	Min	2.5th Percentile	25th Percentile	Median	75th Percentile	97.5th Percentile	Max
25OHD nmol/L	62.4	29.8	6.9	18.1	39.8	59.2	81.0	130.9	222.5
24,25(OH)$_2$D nmol/L	5.4	3.3	0.5	1.0	2.9	4.9	7.4	13.0	29.6
1,25(OH)$_2$D pmol/L	138.8	39.6	32.3	71.9	111.0	135	161.0	229.7	380.0
25OHD: 24,25(OH)$_2$D (VMR)	13	4	2	7	10	12	15	25	39
1,25(OH)$_2$D: 24,25(OH)$_2$D (VMR)	38	33	5	9	18	28	45	132	300
Intact PTH pmol/L	3.7	1.2	1.0	1.9	2.9	3.5	4.3	6.8	11.4
aCa mmol/L	2.38	0.07	2.00	2.20	2.32	2.40	2.41	2.50	2.60

3.3.1 RELATIONSHIP BETWEEN 24,25$(OH)_2$D AND 25OHD

The mean concentration of 24,25$(OH)_2$D was on average 9.5-fold lower than 25OHD. Linear regression analysis showed a directly proportional relationship between 24,25$(OH)_2$D and 25OHD concentrations: [24,25$(OH)_2$D] = 0.0946 × [25OHD]–− 0.42; $r^2 = 0.7206$ (Figure 3.1A). Using this equation, serum 24,25$(OH)_2$D concentration of ≥4.3 nmol/L was deemed to be equivalent to the IOM vitamin D replete status (25OHD of 50 nmol/L), and 24,25$(OH)_2$D concentration of ≤2.4 nmol/L is equivalent to deficiency status (25OHD of ≤30 nmol/L).

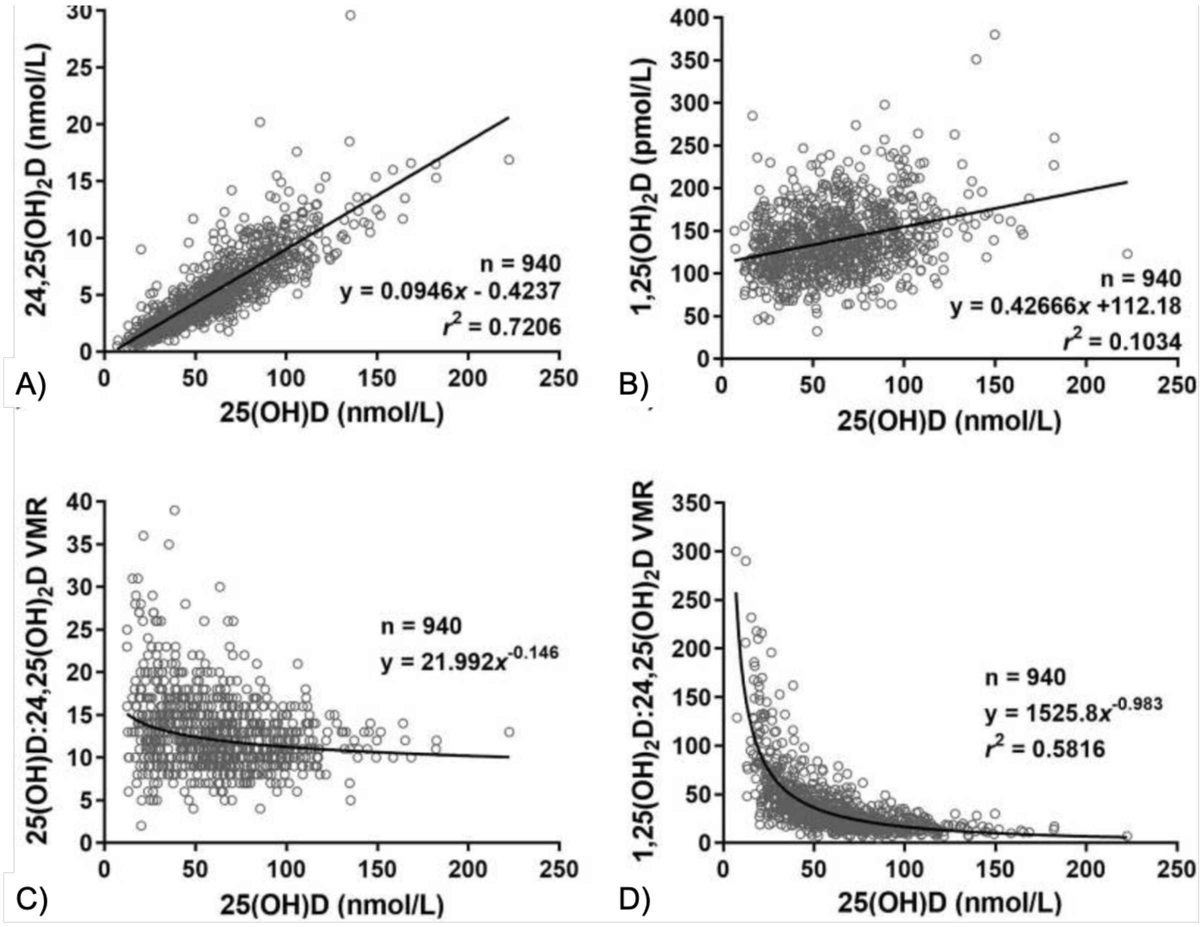


Figure 3.1: Non-parametric correlations of vitamin D metabolites against respective 25OHD concentrations[192] (**A**) 24,25$(OH)_2$D (**B**) 1,25$(OH)_2$D (**c**) 25OHD:24,25$(OH)_2$D VMR and (**D**) 1,25$(OH)_2$D:24,25$(OH)_2$D VMR, against respective 25OHD concentration. Solid lines in **A** and **B** represent the linear regression line. LOWESS fitted curve is plotted in **C** and **D** (99% point fit). The mean concentrations of 24,25$(OH)_2$D, 1,25$(OH)_2$D, 25OHD:24,25$(OH)_2$D VMR and 1,25$(OH)_2$D:24,25$(OH)_2$D VMR represent 8.7%, 222.4%, 20.8% and 60.9% of their respective 25OHD concentration. Assay lLoQ: 25OHD and 24,25$(OH)_2$D = 0.1 nmol/L, 1,25$(OH)_2$D = 12 pmol/L.

3.3.2 RELATIONSHIP BETWEEN 1,25(OH)$_2$D and 25OHD

Despite a direct enzymatic conversion of 25OHD to 1,25(OH)$_2$D, no strong correlation in serum concentrations was observed between these two vitamin D metabolites (Figure 3.1B). This finding is consistent with published studies[193,194]; 1,25(OH)$_2$D is able to inhibit the expression of CYP27B1 both directly and indirectly by suppressing PTH and stimulating FGF23 production. This negative feedback system provides an essential safeguard mechanism against hypercalcaemia, hence 1,25(OH)$_2$D concentration influenced the circulatory concentration of 25OHD through a complex relationship involving various factors e.g., PTH and FGF23 concentration.

3.3.3 VITAMIN D STATUS AND 25OHD:24,25(OH)$_2$D VMR

The 25OHD:24,25(OH)$_2$D VMR showed an indirect relationship with 25OHD (Figure 3.1C); LOWESS fitting showed a steady increase in 25OHD:24,25(OH)$_2$D VMR with the decline in 25OHD concentration. One-way ANOVA showed a significant increase in 25OHD:24,25(OH)$_2$D VMR ($p < 0.001$) at 25OHD below 40 nmol/L (Figure 3.2). The greatest increase in relative 25OHD:24,25(OH)$_2$D ratio was observed when 25OHD concentration decreased below ≤30 nmol/L. The decrease in relative production of serum 24,25(OH)$_2$D in response to the decline in 25OHD suggests down-regulation of CYP24A1 triggered by reduced 25OHD concentration.

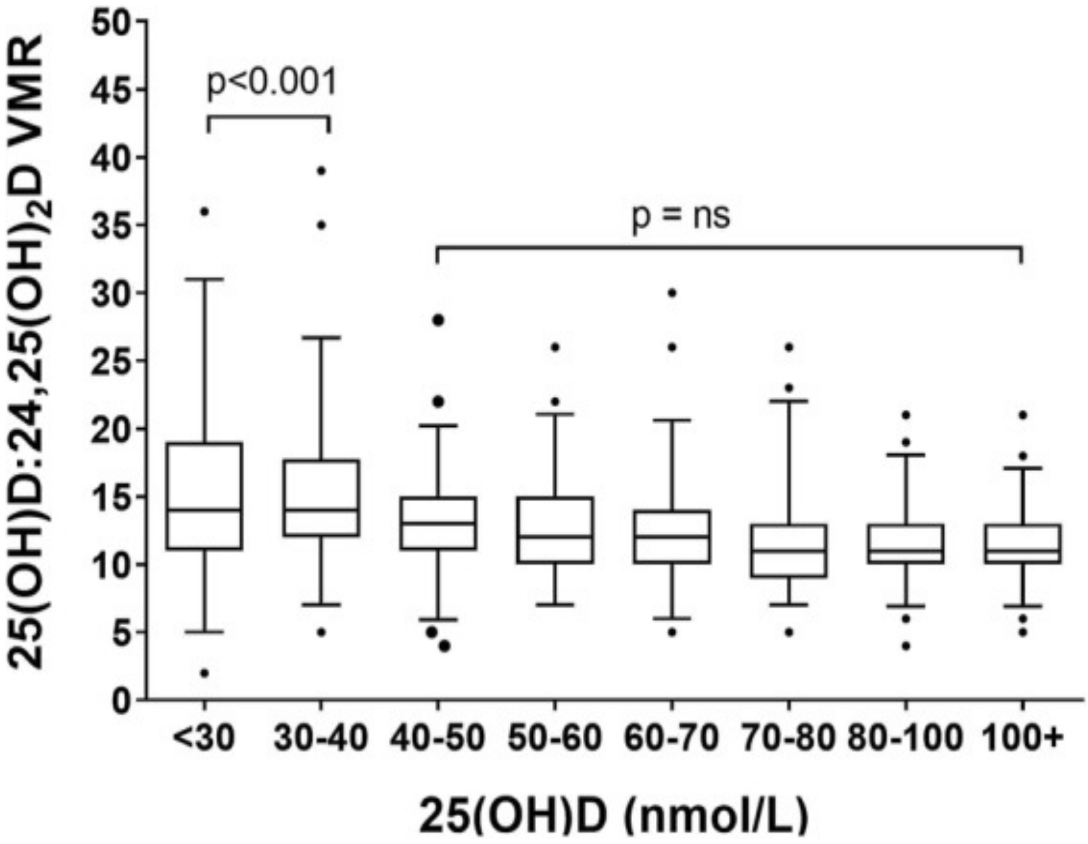


Figure 3.2: Distribution of 25OHD:24,25$(OH)_2$D VMR by 25OHD intervals[192]. Each interval contains an equal number of subjects. A significantly elevated ratio was observed in those with serum 25OHD ≤ 40 nmol/L ($p<0.001$). Box and whiskers represent the median, interquartile range and 95% population intervals.

3.3.4 VITAMIN D STATUS AND $1,25(OH)_2D:24,25(OH)_2D$ VMR

Vitamin D status, as defined by 25OHD concentration, revealed an exponential negative correlation ($r^2 = 0.582$) with $1,25(OH)_2D:24,25(OH)_2D$ VMR (Figure 3.1 D). Further analysis identified a significant increase in $1,25(OH)_2D:24,25(OH)_2D$ VMR at 25OHD ≤ 60 nmol/L (Figure 3.3).

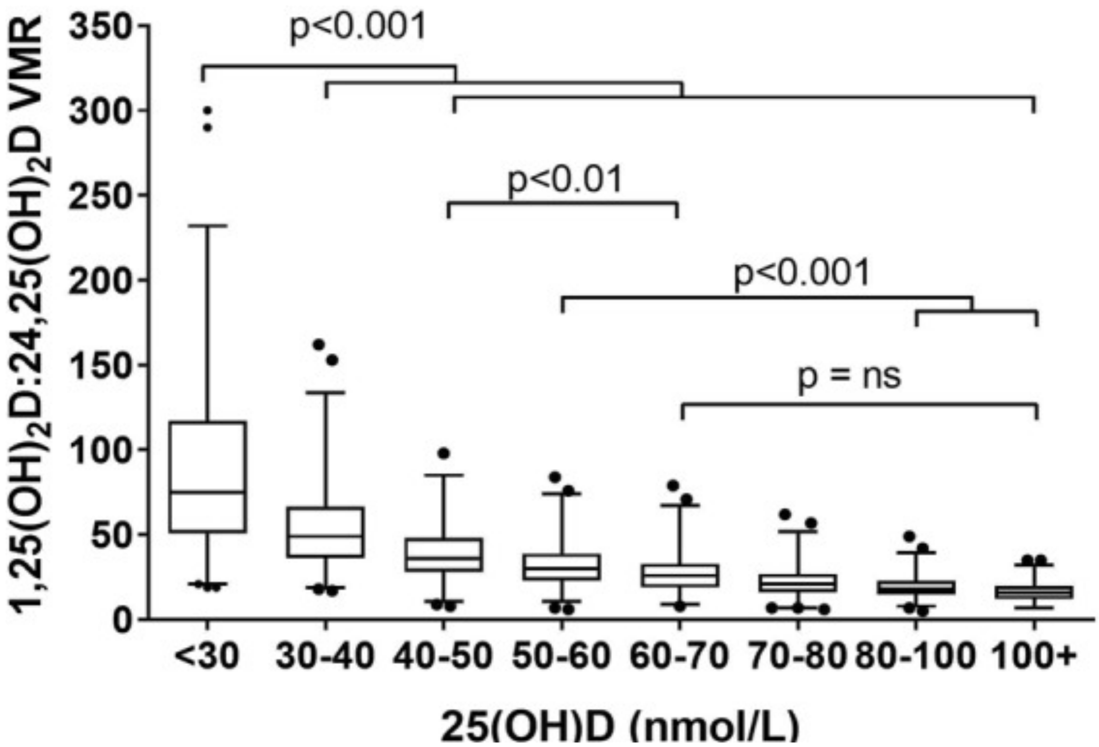


Figure 3.3: Distribution of 1,25$(OH)_2$D:24,25$(OH)_2$D VMR by 25OHD intervals[192]. The relationship between VMR and individual vitamin D metabolites in this figure demonstrates the exponential increase in 1,25$(OH)_2$D:24,25$(OH)_2$D VMR with the decrease in serum 25OHD. Box and whiskers represent the median, interquartile range and 95% population intervals. Each interval contains an equal number of subjects.

Using the Jacobson and Truax[195,196] method to determine the cut-off value for clinically significant change[150], 1,25$(OH)_2$D:24,25$(OH)_2$D VMR of ≥35 was estimated to be the predictive threshold value for vitamin D insufficiency, and ≥51 to be predictive threshold for vitamin D deficiency. The threshold values were determined from subject samples collected in the winter months (January to April) due to the seasonal variation of 25OHD. Receiver Operating Characteristic (ROC) curves generated from data collected between January to April produced area under the curve (AUC) values of 0.88 and 0.86, indicating the VMR cut-offs are excellent at discriminating individuals with vitamin D insufficiency and deficiency (Figure 3.4 A and B). The 1,25$(OH)_2$D:24,25$(OH)_2$D VMR at 35 and 51 achieved true positive rate (sensitivity) at 80% and 78%, respectively, and false positive rate (specificity) of 82% and 74%, respectively (Figure 3.4 A and B).

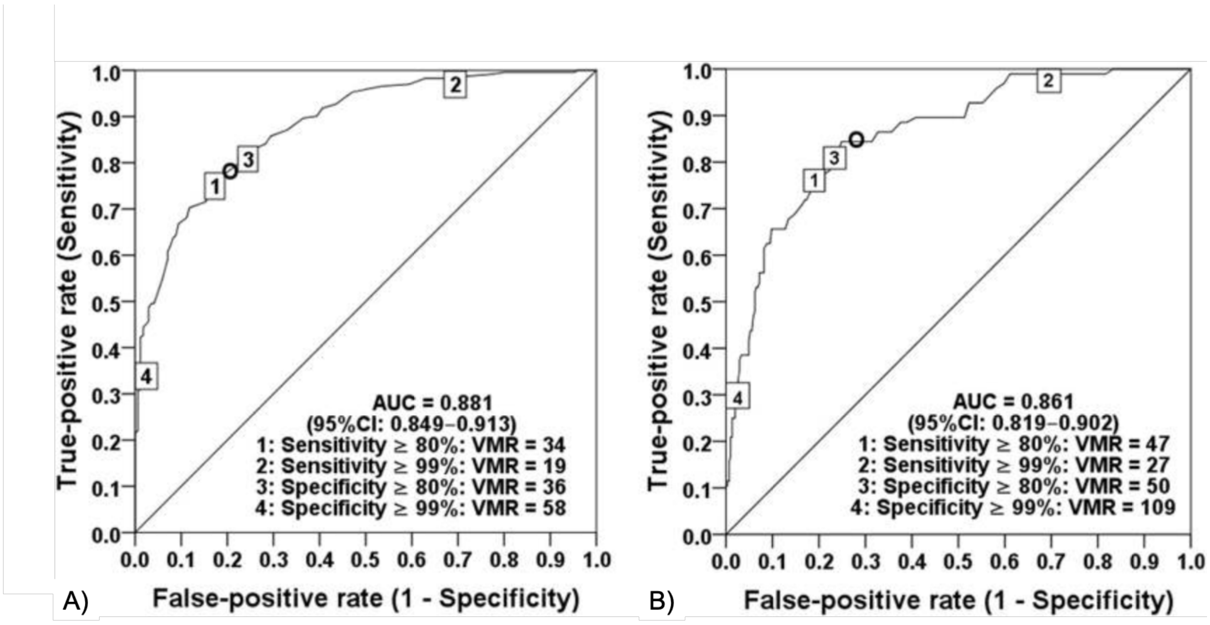


Figure 3.4: Diagnostic performance of 1,25(OH)$_2$D:24,25(OH)$_2$D VMR in the assessment of vitamin D status during winter months (Jan–April) (n = 402) [192]. Receiver Operating Characteristic (ROC) curve depicts diagnostic sensitivity and specificity levels. (**O**) represents decision threshold for (**A**) vitamin D replete (i.e. 25OHD ≥ 50 nmol/L), 1,25(OH)$_2$D:24,25(OH)$_2$D VMR threshold value of 35 (sensitivity = 80%, specificity = 78%), **(B)** vitamin D insufficiency (i.e. 25OHD ≥ 30 nmol/L), 1,25(OH)$_2$D:24,25(OH)$_2$D VMR threshold value of 51 (sensitivity = 82%, specificity = 74%). The diagonal lines represent the line of no discrimination.

3.3.5 PTH AND $1,25(OH)_2D:24,25(OH)_2D$ VMR

Circulating PTH is influenced by 25OHD and $1,25(OH)_2D$, and *vice versa*. This study hypothesised that PTH concentration changes with $1,25(OH)_2D:24,25(OH)_2D$ VMR and 25OHD. To test this hypothesis, median PTH concentrations were established from grid analysis based on groupings of $1,25(OH)_2D:24,25(OH)_2D$ VMR and 25OHD in ascending order (Table 3,3). A decrease in PTH concentration was observed from the high $1,25(OH)_2D:24,25(OH)_2D$ VMR (100+) and low 25OHD (<30 nmol/L) group, to the low $1,25(OH)_2D:24,25(OH)_2D$ VMR (<30) and high 25OHD (100+ nmol/L) group. Using Kruskal-Wallis independent non-parametric analysis to test the distribution of PTH across all groups, a highly significant ($p < 0.001$) change in PTH concentration was found across the $1,25(OH)_2D:24,25(OH)_2D$ VMR and 25OHD categories, hence the null hypothesis was rejected. Using Spearman's rank correlation coefficient (2-tailed) to assess monotonic functions between variables, significant positive correlations were evident between VMRs and PTH ($1,25(OH)_2D:24,25(OH)_2D$ VMR rho = 0.249, $p < 0.001$ and $25OHD:24,25(OH)_2D$ VMR rho = 0.134, $p < 0.001$); whereas vitamin D metabolites showed significant negative correlations with PTH (25OHD rho = −0.287, $p > 0.001$, $24,25(OH)_2D$ rho = −0.282, $p < 0.001$ and $1,25(OH)_2D$ rho = −0.87, $p < 0.001$). The statistical significance remains unchanged after adjustment for BMD and BMI as covariates (Table 3.3).

Table 3.3: Median (SEM) PTH concentrations in categories of increasing 1,25(OH)$_2$D:24,25(OH)$_2$D VMR and 25OHD. One-way ANOVA showed PTH concentrations decreased significantly ($p < 0.001$) from high 1,25(OH)$_2$D:24,25(OH)$_2$D VMR/low 25OHD to low 1,25(OH)$_2$D:24,25(OH)$_2$D VMR/high 25OHD. *Denotes significance at the $p < 0.05$ level.

Median PTH, pmol/L		**25OHD, nmol/L**			
		<30	**30–50**	**51–100**	**100+**
1,25(OH)$_2$D:24,25(OH)$_2$D VMR	100+	5.7 (0.2)*	5.4 (0.5)*	—	—
	51–100	4.3 (0.2)*	3.8 (0.2)	3.4 (0.3)	—
	30–50	4.0 (0.2)*	3.7 (0.1)	3.5 (0.1)	2.7 (0.2)*
	<30	3.8 (0.3)	3.9 (0.1)	3.3 (0.1)*	3.2 (0.1)*

3.3.6 CIRCANNUAL VARIATIONS IN VITAMIN D METABOLITES AND VMRs

Cosinor-fit curves (Figure 3.5) show significant circannual rhythm for 25OHD ($p < 0.001$), $24,25(OH)_2D$ ($p < 0.01$), 25OHD:$24,25(OH)_2D$ VMR ($p < 0.001$), $1,25(OH)_2D$:$24,25(OH)_2D$ VMR ($p < 0.001$) and PTH ($p < 0.05$). No significant rhythm was observed for $1,25(OH)_2D$ ($p = 3.125$). The rhythm observed for 25OHD is consistent with previous reports[197,198]. $24,25(OH)_2D$ showed a similar peak (July–Aug) and nadir (Jan–Mar) pattern as for 25OHD. 25OHD:$24,25(OH)_2D$ VMR and $1,25(OH)_2D$:$24,25(OH)_2D$ VMR exhibited patterns in the opposite direction, with peak (Mar–April) and nadir (Aug–Sept) suggesting that the production of $24,25(OH)_2D$ is relatively higher during summer/early autumn months. Acrophase, defined as the lag time between rhythm-adjusted mean and peak cycle value, was on average (SD) of 8.1(0.3) months for all vitamin D metabolites except for $1,25(OH)_2D$. A low amplitude, circasemiannual PTH secretory rhythm was observed, with an acrophase of 3.5 months.

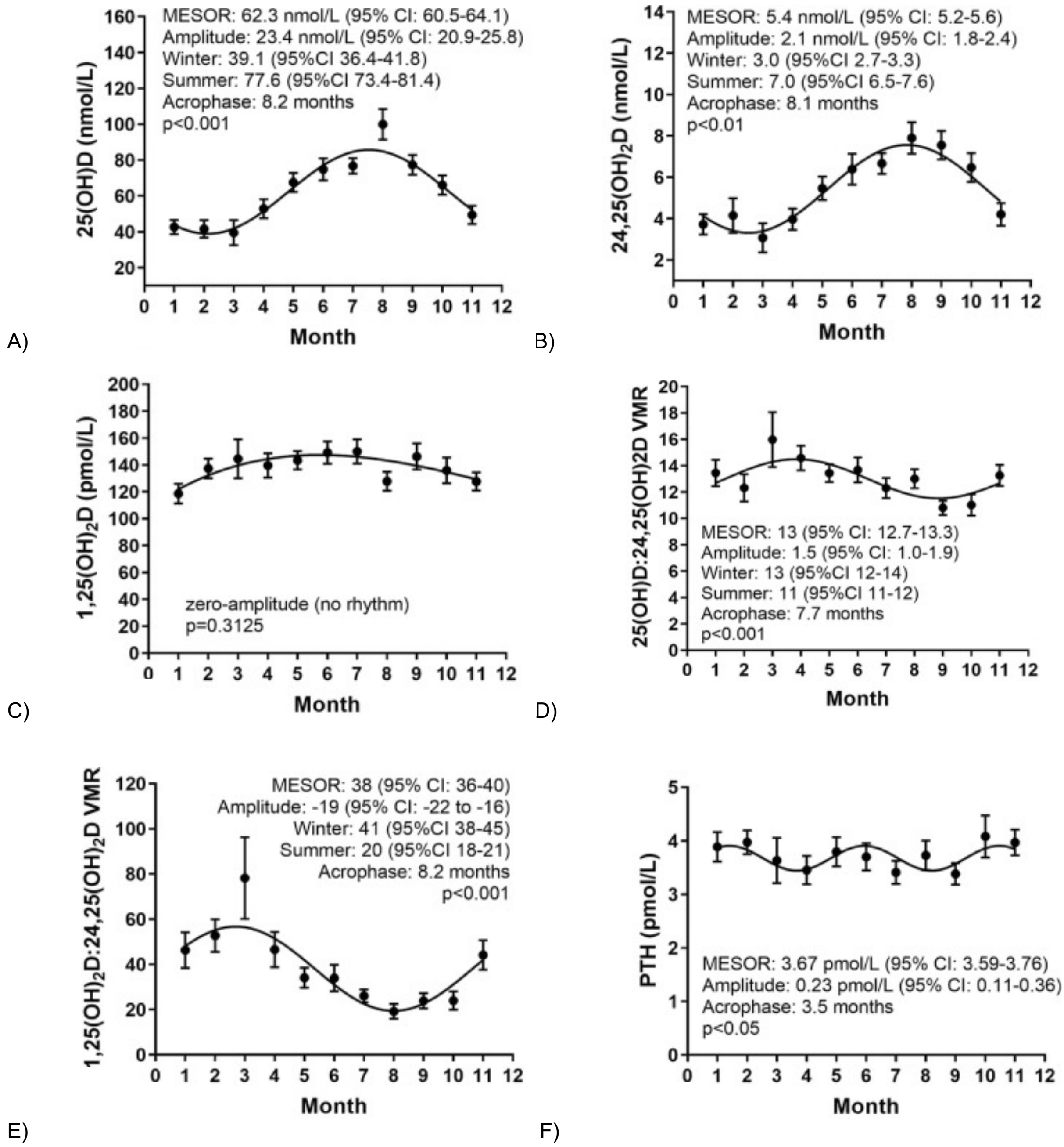


Figure 3.5: Cosinor-fit circannual rhythm for vitamin D metabolites[192] (**A**) 25OHD, (**B**) 24,25(OH)$_2$D, (**C**) 1,25(OH)$_2$D, (**D**) 25OHD:24,25(OH)$_2$D,I)1,25(OH)$_2$D:24,25(OH)$_2$D, (**F**) PTH. Error bars represent 95% CI.

3.4 DISCUSSION

This study demonstrates the relationship between serum concentrations of 25OHD and 1,25(OH)$_2$D when expressed as a relative ratio with serum 24,25(OH)$_2$D. This study provides evidence that the conversion of 25OHD to 1,25(OH)$_2$D is associated with the catabolism of 25OHD to 24,25(OH)$_2$D, which can be assessed by the measurement of serum 24,25(OH)$_2$D and its derived VMR.

3.4.1 RELATIONSHIP BETWEEN 1,25(OH)$_2$D:24,25(OH)$_2$D VMR and 25OHD

The inverse exponential correlation between 1,25(OH)$_2$D:24,25(OH)$_2$D VMR and 25OHD provides insight into the dynamics of vitamin D metabolites in healthy, young adults; when vitamin D status is sufficient, serum concentrations of 1,25(OH)$_2$D and 24,25(OH)$_2$D are maintained in relative proportion and showed no significant change beyond the sufficient threshold. In contrast, when vitamin D status is insufficient, a progressive and highly significant increase in 1,25(OH)$_2$D:24,25(OH)$_2$D VMR is evidence that the production of serum 1,25(OH)$_2$D is favoured over 24,25(OH)$_2$D as the availability of vitamin D precursors in circulation diminishes. Our data imply possible regulatory functions of the 24,25(OH)$_2$D pathway; in hypervitaminosis, the pathway is 'switched on' to allow excess 25OHD to be converted to 24,25(OH)$_2$D. The 24-hydroxylase pathway results in the formation of calcitroic acid for excretion. In hypovitaminosis, the 24,25(OH)$_2$D pathway is partially inactivated by decreased 25OHD concentration to conserve 25OHD and to maintain an adequate supply of substrate for conversion to 1,25(OH)$_2$D.

Although the biological activity of 24,25(OH)$_2$D is yet to be fully elucidated, its role in vitamin D catabolism appears certain. Low serum concentrations of 24,25(OH)$_2$D and elevated

$25OHD:24,25(OH)_2D$ VMR is useful in identifying patients with loss-of-function *CYP24A1* mutations[125,190,199]. In a previous publication, a case of biallelic *CYP24A1* mutation was described in a patient presenting with hypercalcaemia, elevated serum $1,25(OH)_2D$ concentration (293 pmol/L, reference range 43–144 pmol/L), and elevated $25OHD:24,25(OH)_2D$ VMR of 32[150]. At diagnosis, the patient's $1,25(OH)_2D:24,25(OH)_2D$ VMR was 212 (1.6 times the upper [97].5th percentile of 132), which was attributed to supplementation with vitamin D. One month after treatment for hypercalcaemia and cessation of vitamin D supplement, serum $1,25(OH)_2D$ was in the reference range, the $1,25(OH)_2D:24,25(OH)_2D$ VMR decreased to 130 (below the 97^{th} percentile), and the $25OHD:24,25(OH)_2D$ VMR remained elevated at 35[150].

3.4.2 RELATIONSHIP BETWEEN PTH AND VITAMIN D METABOLITES

A major finding of this current study was the novel link between the vitamin D metabolites and VMRs with the distribution of PTH. It is widely accepted that the PTH concentration is associated with 25OHD, but not with the active $1,25(OH)_2D$. This is due to the tight regulatory mechanisms, and the regulatory processes that take place via the VDR to activate intracellular transport of calcium and stimulate PTH secretion. Using the $1,25(OH)_2D:24,25(OH)_2D$ VMR and 25OHD model, this study shows that individuals with low 25OHD (≤50 nmol/L), normal $1,25(OH)_2D$ but high $1,25(OH)_2D:24,25(OH)_2D$ VMR (≥101) have significantly higher PTH concentration than those at the opposite end of the spectrum. An interpretation of our finding supports a biological role of $24,25(OH)_2D$ other than as a catabolic metabolite of vitamin D. Relative high production of $24,25(OH)_2D$ may reduce the bioactivity of 25OHD and $1,25(OH)_2D$, particularly extra-renal production of $1,25(OH)_2D$, to down-regulate the secretion of PTH, whilst maintaining $1,25(OH)_2D$ concentrations in the strict boundaries required for appropriate calcium homeostasis. Relatively low 24,25$(OH)_2$D could enhance the anabolic effects of vitamin D metabolism, by stimulating PTH production.

The biological action of 24,25$(OH)_2$D on the inhibition of PTH secretion was first reported in animal and *in vitro* models in the late 1970s[74,200]. More recently there is increasing evidence supporting physiological functions of 24,25$(OH)_2$D on bone and cartilage[201,202]. in promoting fracture healing, and protection against cartilage damage. The existence of a 24,25$(OH)_2$D-specific nuclear or membrane receptor has been reported[203], but its function has yet to be elucidated. Given that CYP24A1, the enzyme responsible for the production of 24,25$(OH)_2$D is present in most tissues with VDR, understanding the mechanisms controlling the production of 24,25$(OH)_2$D relative to other vitamin D metabolites may have significance beyond vitamin D catabolism, potentially shaping vitamin D supplementation strategies.

3.4.3 CIRCANNUAL VARIATIONS IN VITAMIN D METABOLITES AND VMRs

Mapping the circannual rhythms of vitamin D metabolites and VMRs is an important component of this study. Cosinor models are the only statistical model that allows identification of and adjustment for potential determinants of seasonal variation in 25OHD concentration. This model increases the accuracy of the assessment of vitamin D status in observational studies and can aid clinicians in identifying patients at risk of developing vitamin D insufficiency across seasons. Previously published reports on longitudinal studies describe the changes in serum 25OHD and 24,25$(OH)_2$D throughout a year in vitamin D supplemented or non-supplemented subjects[197,198,200]. In the VICtORy (Vitamin D and CardiOvascularRisk)[204] and VICtORy RECALL[197] randomised controlled studies performed using a group of postmenopausal women residing in the northeast of UK, the placebo group showed a two-fold increase in serum 25OHD in peak summer months (July–August), compared to the nadir in late winter months (January–March). The younger cohort of healthy individuals in this chapter showed similar trends; 24,25$(OH)_2$D had a propensity to fluctuate with 25OHD throughout the year, with changes between summer and winter months, as indicated by a lower 25OHD:24,25$(OH)_2$D VMR during January to March than during July to September. The cohort in the VICtORY study differs from the cohort in this chapter (postmenopausal women residing in the northeast vs healthy army recruits respectively) meaning that the outdoor exposure likely differs between the two groups. The data from the circannual rhythms presented in this chapter mirrors the previous VICtORY longitudinal studies regardless of differing study populations and outdoor exposure. While the cohorts mirror the circannual variations in vitamin D metabolites, this chapter represents young adults of Caucasian extraction (92.9%) and cannot be extrapolated to the wider population of mixed age and ethnicity without further, heterogenous cohort investigations. Serum 1,25$(OH)_2$D displayed no circannual rhythm and was in the reference range throughout the year. This study reported the circannual variation of 1,25$(OH)_2$D:24,25$(OH)_2$D VMR was dependent on 24,25$(OH)_2$D, with a peak-to-nadir difference of 19; such sharp change between seasons would inevitably create uncertainty when using 1,25$(OH)_2$D:24,25$(OH)_2$D VMR in diagnostic decision-making. In contrast, 25OHD:24,25$(OH)_2$D VMR is less susceptible to seasonal fluctuation, allowing the use of the VMR with fixed reference intervals irrespective of the time of the year.

3.4.4 CONCLUSION

One strength of this data is the chosen cohort; with participants attending blood sampling visits at strictly controlled time intervals and that the vitamin D metabolites were measured using gold-standard methodologies. The participants are well-defined, largely from a similar social-economic background, and exposed to the same level of fitness training, diet, and frequency of outdoor activities. Although males and females are not separated in this study, previous work has highlighted that neither vitamin D metabolites, PTH nor the relationships between are significantly affected by gender[205]. The relative homogeneity of the subjects of our study population in combination with our inclusion criteria allowed us to confidently form a reference population and identify important changes in analytes. The limitations are that our findings are observational and based on baseline sampling at the start of training. Additionally, this cohort represents young adults of Caucasian extraction (92.9%) and cannot be extrapolated to the wider population of mixed age and ethnicity. The predictive threshold values were therefore established based on the equivalent vitamin D status as described by IOM, and not based on the data generated in this study. DBP and free 25OHD were not analysed due to the ethnic homogeneity of our population (Caucasian 92.9%, Asian 1.6%, Black 1.7%, Chinese 0.1%, mixed 3%, others 0.7%) and factors that may influence DBP levels were not excluded (e.g., oral contraceptive use in female recruits).

In conclusion, the analysis in this chapter characterises the absolute and relative concentrations of the active and catabolic form of vitamin D metabolites in a well-defined young, healthy and physically fit population. The use of VMRs provides insight into the metabolic pathway for Vitamin D and the variations exhibited throughout the year. This study presents a three-dimensional model incorporating 1,25$(OH)_2$D, 24,25$(OH)_2$D and 25OHD measurements and report a strong correlation between metabolites that are linked with PTH. Such modelling could help establish vitamin D-adjusted PTH reference intervals, and contribute to the goal of a “Treat to target” approach to vitamin D supplementation. Additionally, measurement of 25OHD and 24,25$(OH)_2$D should be considered a part of the clinical workup in patients with hypercalcemia of otherwise unknown etiology as low serum concentrations of 24,25$(OH)_2$D combined with elevated 25OHD:24,25$(OH)_2$D could indicate underlying loss-of-function *CYP24A1* mutations.

CHAPTER 4: WHOLE EXOME SEQUENCING IN PATIENTS WITH SUSPECTED CMH

4.1 INTRODUCTION

LC-MS/MS can provide simultaneous analysis of the vitamin D metabolites 25OHD and $24,25(OH)_2D$. The ratios between 25OHD:$24,25(OH)_2D$ and $1,25(OH)_2D$:$24,25(OH)_2D$ are known as the VMR[192]. The interpretation of VMR allows initial insight into the presence of *CYP24A1* loss-of-function mutations that can lead to subsequent hypervitaminosis D or CMH. VMR ratios for the general population lie between 5-25, while in CMH patients for example, this can increase to >80[125,190]. The increase in VMR in HCINF1 patients has been shown to occur due to non-functional/partially functional CYP24A1 and can present as normal 25OHD with reduced metabolism to $24,25(OH)_2D$ and inappropriate or elevated $1,25(OH)_2D$ concentrations[206]. Serum $24,25(OH)_2D$ concentration is tightly correlated to 25OHD meaning the VMR is a good indicator of CYP24A1 enzyme activity[192]. LC-MS/MS measurement of vitamin D metabolites provides a rapid screening tool for the identification of potential *CYP24A1* abnormalities[190,192].

Although biochemical analysis is a useful tool in screening patients suspected of harbouring *CYP24A1* mutations, it does not provide a definitive diagnosis of underlying mutations. Biochemical analysis stemming from clinical presentation can be inconclusive in many cases due to an overlap in biochemical profiles in conditions with similarly presenting phenotypes e.g., Williams-Beuren Syndrome and *SLC34A1*[146,206–208]. To confirm the CMH diagnosis after initial LC-MS/MS screening, genetic sequencing is required to confirm the presence of *CYP24A1* mutations. Recent advances in high-throughput sequencing allows characterisation of somatic and germline variants at high resolution.

Whole exome sequencing (WES) allows high-throughput screening for a wide variety of disease associated variants to aid rapid diagnosis of rare conditions. The exome forms the protein coding region of DNA containing most known disease-associated variants. As the exome forms just ~2% of the entire DNA sequence, WES sequencing can easily screen the entire exome in a single test, allowing rapid interpretation of sequencing results[209]. WES is therefore considered an efficient tool in identifying potential disease-causing mutations, especially in cases with ambiguous phenotypes.

Creating a WES ‘universal panel’ enables analysis and evaluation of all gene-disease associations in the exome from a single sample, which may have been missed with targeted genetic sequencing[210]. Sequencing the entire exome in a single test reduces the potential need for multiple tests to reach a diagnosis. Selecting specific genes for targeted sequencing of patients whose phenotype is non-conclusive can be difficult. WES allows the analysis of all protein coding genes, permitting assessment of all disease associated genes in the exome that may contain disease causing variants[210]. The WES approach is useful in conditions with significant genetic heterogeneity, allowing a cost effective, time saving and efficient diagnosis. WES also provides insight into novel genes that have not yet been associated with disease.

While there is much research on the clinical biochemistry of metabolic bone disease, research linking other areas of biology that may cause diseases to the biochemical analysis is lacking. Functional genomics aids our understanding of the full biological process of genotype to phenotype. Investigating all aspects of disease (genome > epigenome >transcriptome > proteome > metabolome) is vital in understanding homeostasis and disease pathogenesis.

In this study WES sequencing was performed on genomic DNA from patients (n=9) with a hypercalcaemic phenotype and biochemistry consistent with *CYP24A1* loss-of-function mutations. Alongside *CYP24A1* this research sought to investigate multiple genes associated with bone remodelling, calcium handling and vitamin D metabolism: Sequestosome-1 (*SQSTM1),* TNF Receptor Superfamily Member 11b *(TNFRSF11),* Dendrocyte Expressed Seven Transmembrane Protein *(DCSTAMP),* Nucleoporin 205 *(NUP205),* Optineurin *(OPTN),* Promyelocytic leukemia protein *(PML), CSF1,* Ras And Rab Interactor 3 *(RIN3), VDR, CYP27B1,* Valosin Containing Protein *(VCP), CYP2R1, CASR,* Phosphate regulating endopeptidase homolog (*PHEX),* Alkaline Phosphatase, Biomineralization Associated *(ALPL*). This 'bone panel' of genes of interest could be utilised in further studies to screen patients with unidentified phenotype and biochemical analysis relating to abnormal calcium handling and CMH. Our data provides a comprehensive view of genetic alteration in patients with hypervitaminosis D.

4.2 CLINICAL SAMPLES

The University of East Anglia (UEA) Faculty of Medicine and Health Sciences Research Ethics Committee approved the collection and study of Human samples for non-clinical procedures investigating *CYP24A1* abnormalities (Reference: 2018–19 - 100). 147 patient serum samples were collected as part of routine requests for 25OHD LC-MS/MS analysis from the Department of Laboratory Medicine at the Norfolk and Norwich University Hospital between June 2016 and June 2017. Patients were referred from the metabolic or stone former clinics. Blood samples were collected into serum gel separator tubes (BD Vacutainer) and centrifuged immediately. The serum layer was the aliquoted and stored at -20 °C until analysis.

Whole blood from the Norfolk and Norwich University Hospital metabolic and stone former clinics for genetic analysis was obtained from patients identified with inappropriate 1,25(OH)$_2$D and/or VMR plus clinical presentation of nephrolithiasis and/or hypercalciuria serum. All adults or infant parents/guardians provided written informed consent to donate samples to this study.

Negative control whole blood samples were collected at the Norfolk and Norwich University Hospital blood typing service (n=10). Exclusion criteria for control samples were those with a vitamin D, calcium or other metabolic disorder clinical history. Control samples were collected using the UK NHS Research Ethics Committee decision toolkit (http://www.hra-decisiontools.org.uk/ethics/).

4.3 RESULTS

This WES study investigated protein coding variants in genes related to bone remodelling and the metabolism of vitamin D in patients with hypervitaminosis D due to potential non-functional *CYP24A1*. An initial cohort of patients within the NNUH metabolic bone clinic (n=147) underwent LC-MS/MS biochemical analysis of vitamin D metabolites to determine their VMR, as previously described. A total of 9 patients from this cohort showed biochemistry consistent with non-functional/partially functional *CYP24A1* consistent with CMH and underwent WES to confirm the biochemical findings. The percentage of patients in this cohort with biochemical findings suggestive of CMH (6%) mirrored those estimated in the general population (4-20%[126]), likely due to patients originating from the metabolic bone clinic with disease affecting their calcium handling. The biochemical profiles of the 9 patients suspected of having CMH included elevated 1,25$(OH)_2$D (n = 8), elevated 25OHD:24,25$(OH)_2$D (n = 5), elevated 1,25$(OH)_2$D:24,25$(OH)_2$D (n = 5), and/or hypercalcemia (n = 3), indicating potential *CYP24A1* loss-of-function mutations (Table 4.1).

Table 4.1: Biochemical analysis of WES patient cohort. The values in red indicate that the analyte measured is outside of the reference range (shown in brackets).

Patient	Age	Total 25OHD (50-120 nmol/L)	$1,25(OH)_2D$ (55-139 pmol/L)	Total $24,25(OH)_2D$ (1.1-13.5 nmol/L)	Total 25OHD:$24,25(OH)_2D$ Relative Ratio (7-23)	$1,25(OH)_2D$:$24,25(OH)_2D$ Relative Ratio (11-62)	Adjusted Calcium (2.1-2.6 mmol/L)	Phosphate (0.8-1.5 mmol/L) (Adult)
1	33	69	**243**	2.7	**26**	**90**	**2.68**	1.27
2	58	74	**168**	2	**37**	**84**	2.58	0.98
3	53	55	**169**	2.5	**22**	**68**	-	-
4	24	78	**158**	5.3	15	30	2.45	1.41
5	39	80	**153**	9.2	9	17	2.43	1.11
6	67	69	**203**	7.4	**27**	**9**	**2.62**	1.04
7	38	61	**180**	4.5	14	40	2.37	1.23
8	1	106	**301**	5.6	19	54	**2.75**	**2.06**
9	33	83	138	2.1	**40**	**66**	-	-

As WES provides genetic information for the exome, this study established a focus list of genes associated with bone remodelling and calcium handling, which could present phenotypes similar to *CYP24A1* loss-of-function. This WES ‘bone panel’ containing genes of interest was our focus for this patient cohort (Table 4.2).

Table 4.2: WES bone panel genes of interest associated with bone remodelling, calcium handling and vitamin D metabolism.

Gene	Location	Role in Bone Biology	Disease Association
SQSTM1	ENSG00000161011 Chromosome 5 179,233,388-179,265,078 forward strand	Bone remodelling through promotion of osteoclast formation. Additional roles include promoting autophagy and apoptosis in immune response and inflammation	Paget's Disease of Bone (PDB)
TNFRSF11A	ENSG00000141655 Chromosome 18 59,992,520-60,058,516 forward strand	Bone remodelling through regulating osteoclast formation and activity.	PDB, Osteopetrosis, Familial Expansile Osteolysis, Expansile Skeletal Hyperphosphatasia, Autosomal Recessive Osteopetrosis
DCSTAMP	ENSG00000164935 Chromosome 8 104,339,087-104,356,689 forward strand	Bone remodelling through regulating osteoclast formation and activity. Additional roles include immunological functions and myeloid differentiation.	PDB, Osteopetrosis
NUP205	ENSG00000155561 Chromosome 7 135,557,919-135,648,757 forward strand	Nuclear pore complex assembly and maintenance. Allowing active	PDB, Nephrotic Syndrome

		transport of proteins, RNA and ribonucleoproteins between cytoplasm and nucleus of cells.	
OPTN	ENSG00000123240 Chromosome 10 13,099,449-13,138,308 forward strand	Regulation of bone metabolism by negatively regulating osteoclast differentiation	PDB, Glaucoma
PML	ENSG00000140464 Chromosome 15 73,994,673-74,047,812 forward strand	Role in bone metabolism by regulation of osteoclast development. Additionally regulates cellular senescence, apoptosis, metabolism and angiogenesis.	PDB, Acute Promyelocytic Leukaemia
CYP24A1	ENSG00000019186 Chromosome 20 54,153,449-54,173,973 reverse strand	Vitamin D metabolism. Facilitating the metabolism of vitamin D metabolites to prevent toxicity e.g., 25OHD and 1,25(OH)2D to	Idiopathic Infantile Hypercalcemia (IIH)/ Infantile Hypercalcemia Type 1 (HCINF1)

		24,25(OH)2D to 1,24,25(OH)2D	
CSF1	ENSG00000184371 Chromosome 1 109,910,242-109,930,992 forward strand	Promotes mononuclear cell proliferation, differentiation plus the migration of mature osteoclasts	Adult-Onset Leukoencephalopathy with Axonal Spheroids and Pigmented Glia
RIN3	ENSG00000100599 Chromosome 14 92,513,774-92,688,994 forward strand	Bone metabolism through regulation of osteoclasts	PDB
VDR	ENSG00000111424 Chromosome 12 47,841,537-47,943,048 reverse strand	Vitamin D metabolism through binding and transporting vitamin D metabolites e.g., 1,25(OH)2D allowing regulation of calcium and phosphate homeostasis	Vitamin D Dependent Rickets, Alopecia Areata, Intervertebral Disc Disease, Kidney Stones
CYP27B1	ENSG00000111012 Chromosome 12 57,762,334-57,768,986 reverse strand	Vitamin D metabolism. Facilitating the metabolism of vitamin D metabolites e.g., 25OHD to 1,25(OH)2D.	Vitamin D Dependent Rickets
VCP	ENSG00000165280 Chromosome 9 35,056,064-35,073,249 reverse strand	Regulation of calcium homeostasis through mitochondria-	PDB, Amyotrophic Lateral Sclerosis

		associated endoplasmic reticulum membranes	
CYP2R1	ENSG00000186104 Chromosome 11 14,877,440-14,892,252 reverse strand	Vitamin D metabolism. Facilitating the metabolism of vitamin D into its active form 25OHD.	Vitamin D Dependent Rickets
CASR	ENSG00000036828 Chromosome 3 122,183,683-122,291,629 forward strand	Calcium homeostasis through regulation of PTH	Autosomal Dominant Hypocalcemia, Familial Isolated Hyperparathyroidism, Kidney Stones
PHEX	ENSG00000102174 Chromosome X 22,032,441-22,251,310 forward strand	Phosphate Homeostasis and bone remodelling through regulation of FGF23.	Hereditary Hypophosphatemic Rickets
ALPL	ENSG00000162551 Chromosome 1 21,509,372-21,578,412 forward strand	Bone mineralisation through the enzyme tissue-nonspecific alkaline phosphatase (TNSALP)	Hypophosphatasia

Of the 304 total WES identified variants in the selected bone panel genes, 37 were suggestive of germline mutations and were either heterozygous or homozygous (Table 4.3 and 4.4). In our cohort (n=9), no germline mutations were identified in *CYP24A1* (Table 4.3 and 4.4). *DCSTAMP* was identified as the most common heterozygous mutation appearing across 3 of the 9 patients analysed, while *CASR* and *SQSTM1* were the joint most common homozygous mutations observed (n=3). All 6 germline mutations in *CASR* and *SQSTM1* were present in a single patient (Table 4.3 and 4.4). Of the identified heterozygous variants, 7 of 22 were supported by >100 total reads (Table 3.4). Of the 15 homozygous variants, 6 were supported by >100 total reads (Table 4.4). Identified WES heterozygous and homozygous variants were identified in numerous locations within the gene with varying consequences (Table 4.5).

Table 4.3: Heterozygous mutations identified in the bone panel in our patient cohort (n=9). The Human Genome (v38) was used as a control reference against the patient WES sequencing data. Allele frequency represents the relative frequency of an allele at a particular locus in a population. Allele frequency ~0.5 was deemed a likely heterozygous germline mutation. Variants with total reads <10 were excluded from analysis.

Patient	Gene	NCBI Locus	Mutation Consequence	Reference [Human Genome Sequence]	WES Identified Alteration	Allele Freq	Total Reads	Genomic Location	Clinical Significance
1	*ALPL*	NC_000001.11	Non-coding transcript exon	C	T	0.55	91	21577713	Benign/Likely [211]
4	*CASR*	NC_000003.12	3' UTR	TAAA AAAA AGAAG AGC	TAAA AAAA AAGA AGAGC	0.45	11	122285371 - 122285386	Benign[212]
4			Missense	G	T	0.49	322	122284910	Benign[213]
1	*CSF1*	NC_000001.11	Missense	T	C	0.48	282	109923844	Not Reported[214]
1			Synonymous	C	A	0.50	484	109923716	Not Reported[215]
1			Splice region/ synonymous	C	A	0.52	29	109924188	Not Reported[216]
5	*DCSTAMP*	NC_000008.11	Missense	G	C	0.46	140	104349126	Not Reported[217]
4			Synonymous	G	A	0.47	103	104348825	Not Reported[218]
5			Synonymous	C	T	0.49	97	104348867	Not Reported[219]
9			Intron	C	T	0.59	81	104355036	Not Reported[220]
7	*NUP205*	NC_000007.14	Synonymous	C	T	0.51	57	135648424	Likely Benign[221]
1			Intron	T	C	0.65	20	135577785	Not Reported[222]
2	*PHEX*	NC_000023.11	5' UTR	A	G	0.42	33	22032916	Benign[223]
4	*PML*	NC_000015.10	Missense	G	A	0.40	10	74033411	Not Reported[224]
1			Missense	T	C	0.41	49	74044292	Not Reported[225]
9			Synonymous	C	T	0.52	138	74023017	Benign[226]
5	*RIN3*	NC_000014.9	Synonymous	C	T	0.41	39	92651853	Not Reported[227]
3			Missense	A	G	0.43	317	92651693	PDB[228]
9			Missense	C	T	0.59	32	92651884	Not Reported[229]
6	*SQSTM1*	NC_000005.10	Missense	T	C	0.43	28	179833654	Benign/Likely[230]
5	*TNFRSF11A*	NC_000018.10	3' UTR	A	G	0.49	83	62384811	Not Reported[231]
1	*VDR*	NC_000012.12	Synonymous	A	G	0.58	71	47844974	Benign[232]

Table 4.4: Homozygous mutations identified in the bone panel in our patient cohort (n=9). The Human Genome (v38) was used as a reference against the patient WES sequencing data. Allele frequency represents the relative frequency of an allele at a particular locus in a population. Allele frequency ~1 was deemed a likely homozygous germline mutation. Variants with total reads <10 were excluded from analysis.

Patient	Gene	NCBI Locus	Mutation Consequence	Reference [Human Genome Sequence]	WES Identified Alteration	Allele Freq	Total Reads	Genomic Location	Clinical Significance
1	*ALPL*	NC_000001.11	Synonymous	T	C	1.00	16	21563142	Likely Benign[233]
1	*CASR*	NC_000003.12	Synonymous	G	C	0.99	133	122284198	Benign[234]
1			Missense	G	C	1.00	379	122284985	Benign[235]
1			3' UTR	A	T	1.00	105	122285251	Benign[236]
1	*CSF1*	NC_000001.11	Missense	T	C	1.00	251	109924087	Benign[237]
1	*CYP2R1*	NC_000011.10	Synonymous	G	A	1.00	10	14892029	Not Reported[238]
7	*NUP205*	NC_000007.14	Synonymous	C	T	0.88	17	135619467	Not reported[239]
1			Missense	G	C	1.00	43	135619525	Not Reported[240]
5	*PHEX*	NC_000023.11	5' UTR	C	T	1.00	39	22032973	Benign[241]
1	*RIN3*	NC_000014.9	Synonymous	G	A	1.00	187	92652324	Nor Reported[242]
2			Inframe deletion	TGGC GGCG GCGG CGGC GGGA	TGGC GGCG GCGG CGGGA	1.00	27	92688192-92688211	Not reported[243]
1	*SQSTM1*	NC_000005.10	Synonymous	C	T	1.00	25	179833153	Benign[243]
1			Downstream Gene variant	T	C	1.00	99	179837731	Benign[244]
1			Downstream gene variant	G	T	1.00	26	179837915	Benign[245]
1	*TNFRSF11A*	NC_000018.10	Intron	A	G	1.00	237	62368850	Benign[246]

Table 4.5: Mutation type and consequence of variants identified in WES analysis[247].

Mutation Type	Location/Consequence
3' UTR variant	Variant located in the 3' UTR Altered regulation of mRNA processes e.g. mRNA localisation, stability and translation
5'UTR variant	Variant located in the 5'UTR Altered ribosome recruitment and translation efficiency
Downstream gene variant	Variant located 3' of the gene Non coding variant with unknown consequence
Frameshift variant	Variant that disrupts the translational reading frame
Inframe deletion	Non synonymous variant that deletes bases from the coding sequence
Intron variant	Variant located in the non coding region (intron) Impact alternative splicing by interfering with splice site recognition
Missense variant	Variant changes one or multiple bases resulting in a different amino acid
Splice acceptor variant	Splice variant in the 2-base region at the 3' end of an intron Activation of 3′ (acceptor) splice site
Splice region variant	Variant in the splice site region within 3-8 bases of an intron or 1-3 bases of an exon Disruption of RNA splicing resulting in the loss of exons or the inclusion of introns and an altered protein-coding sequence
Stop gained	Variant that changes at least one base of a codon causing a premature stop codon and shortened transcript
Stop lost	Variant that changes at least one base of a terminator codon causing a lengthened transcript
Synonymous variant	Variant causing no change to the amino acid sequence
Upstream gene variant	Variant located 5' of the gene Non coding variant with unknown consequence

When analysing the number of mutations at different allele frequencies, most mutations had an allele frequency of <20% (Figure 4.1). Only 6.7% of mutations identified with an allele frequency of <20% had total number of supporting reads greater than 100. A low allele frequency of <20% plus low read coverage is suggestive of a false positive mutation, while a low allele frequency plus high read coverage is suggestive of a true somatic mutation. Currently no standardisation of threshold is universally accepted and frequency/read cut off values are down to individual lab discretion[248–250]. While most somatic mutations identified were poorly supported by minimal read count, Patient 2 was identified with a missense variant (g.122261785 A>C) in CASR at an allele frequency of 13% with 225 supporting reads. Patient 3 was also identified with a missense variant (g.122284824 A>C) in CASR at an allele frequency of 6% with 712 supporting reads. CASR variants are associated with hypercalcemia and stone formation[251,252] which supports the biochemical profile of elevated 1,25$(OH)_2$D and abnormal VMR observed in Patient 2 and 3 (Table 4.1).

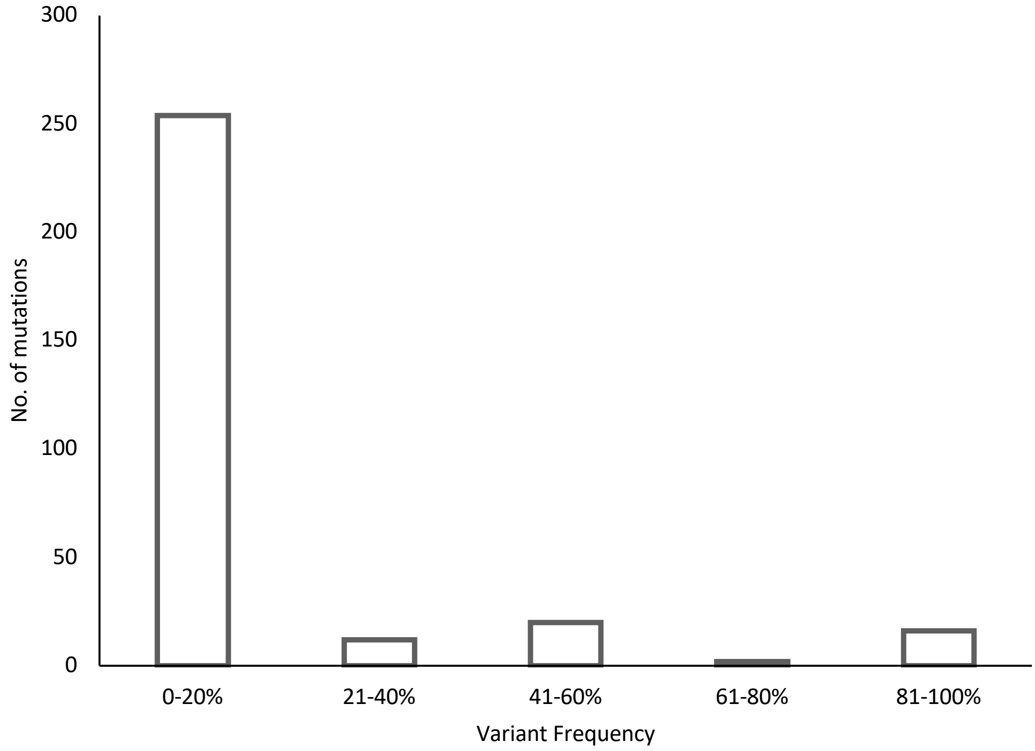


Figure 4.1: Number of WES mutations identified in each allele frequency bracket. Most variants demonstrated an allele frequency of <20%. Each gene within the bone panel presented with variants of apparent <20% frequency: ALPL (6%), CASR (4%), CSF1 (3%), CYP24A1 (4%), CYP27B1 (6%), CYP2R1 (5%), DCSTAMP (2%), NUP205 (11%), OPTN (9%), PHEX (3%), PML (12%), RIN3 (16%), SQSTM1 (6%), TNFRSF11A (0.5%) and VCP (15%), Only 6.7% of variants with <20% frequency are supported by a total read count >100. Mutations were deemed heterozygous with an allele frequency of 40-60% (Table 4.3) or homozygous (Table 4.4) with an allele frequency >85%.

When analysing the mean mutated allele frequency for all somatic and germline variants in the bone panel, each gene of interest contained variants in at least two patients in our cohort (Figure 4.2). Every patient harboured a variant in *RIN3* and *OPTN* including missense variants, in frame deletions, splice region variants, 3' UTR variants and/or synonymous variants, which are associated with Paget's Disease of Bone (PDB). Additionally, each patient harboured either missense, splice region, 3' UTR, synonymous or frame shift variants in *ALPL*, associated with the development of hypophosphatasia (Figure 4.2). The highest mutant allele frequency was observed in *SQSTM1*, *RIN3*, *CSF1* and *CASR*, which are associated with PDB. Patient 6 had variants in every bone panel gene of interest, except *TNRFSF11A*, the majority of which are somatic mutations (Figure 4.2).

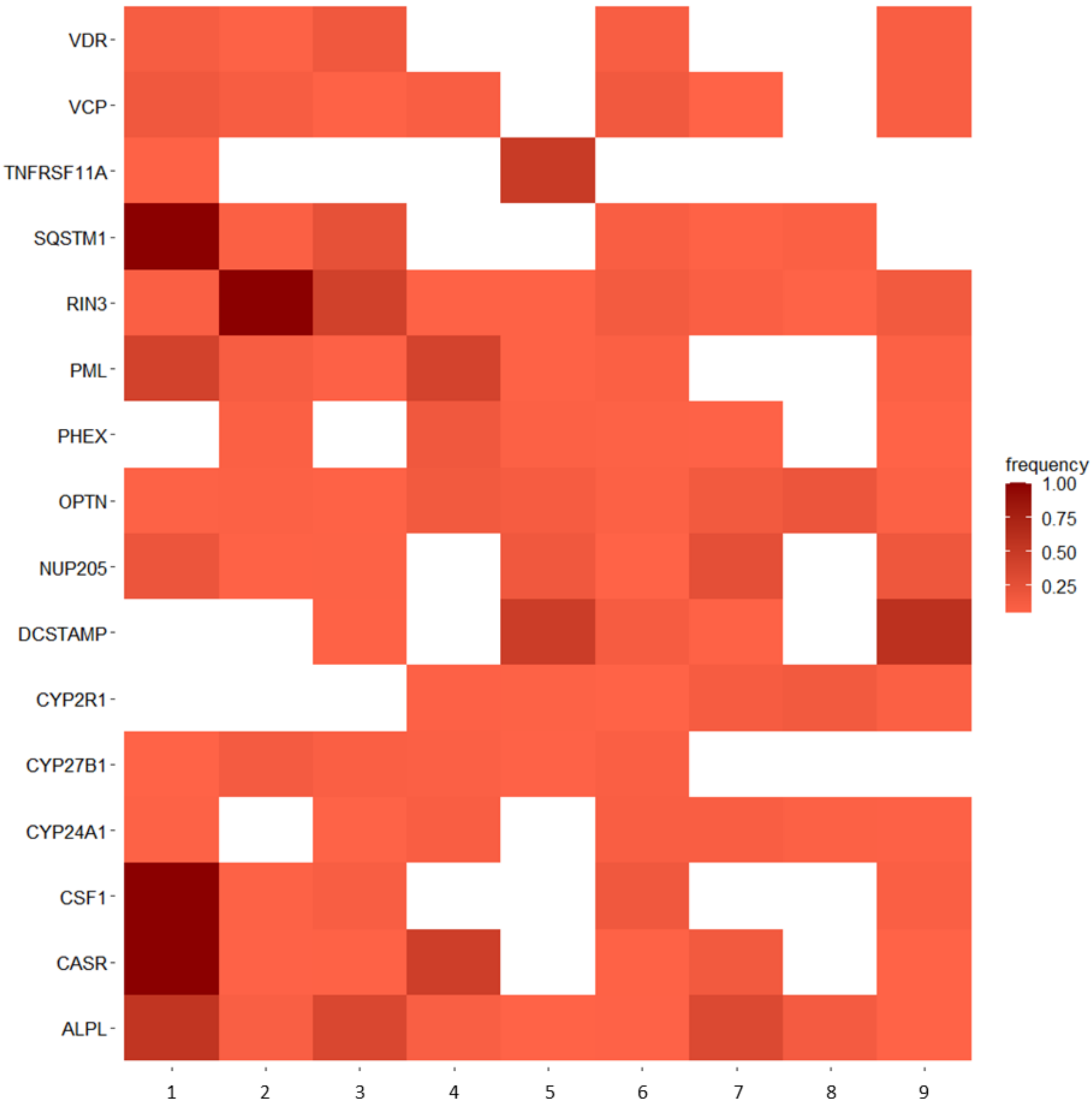


Figure 4.2: WES results for our bone panel in 9 patient samples. This heatmap depicts the allele frequency of mutations within each gene, darkest indicating highest variant frequency to white, indicating no mutations. Only coding germline and non-silent somatic mutations were included with >10 total reads.

This WES data indicates that there was a wide range of mutation types within the bone panel in this patient cohort (Figure 4.3). In this data set, 57% of identified variants were identified as missense (Figure 4.4). All patients in the cohort, except patient 6, contain missense mutations in *OPTN,* which is associated with PDB. Patient 6 is shown to harbour a codon substitution in *OPTN*, which appears synonymous meaning that the encoded amino acid remains unchanged.

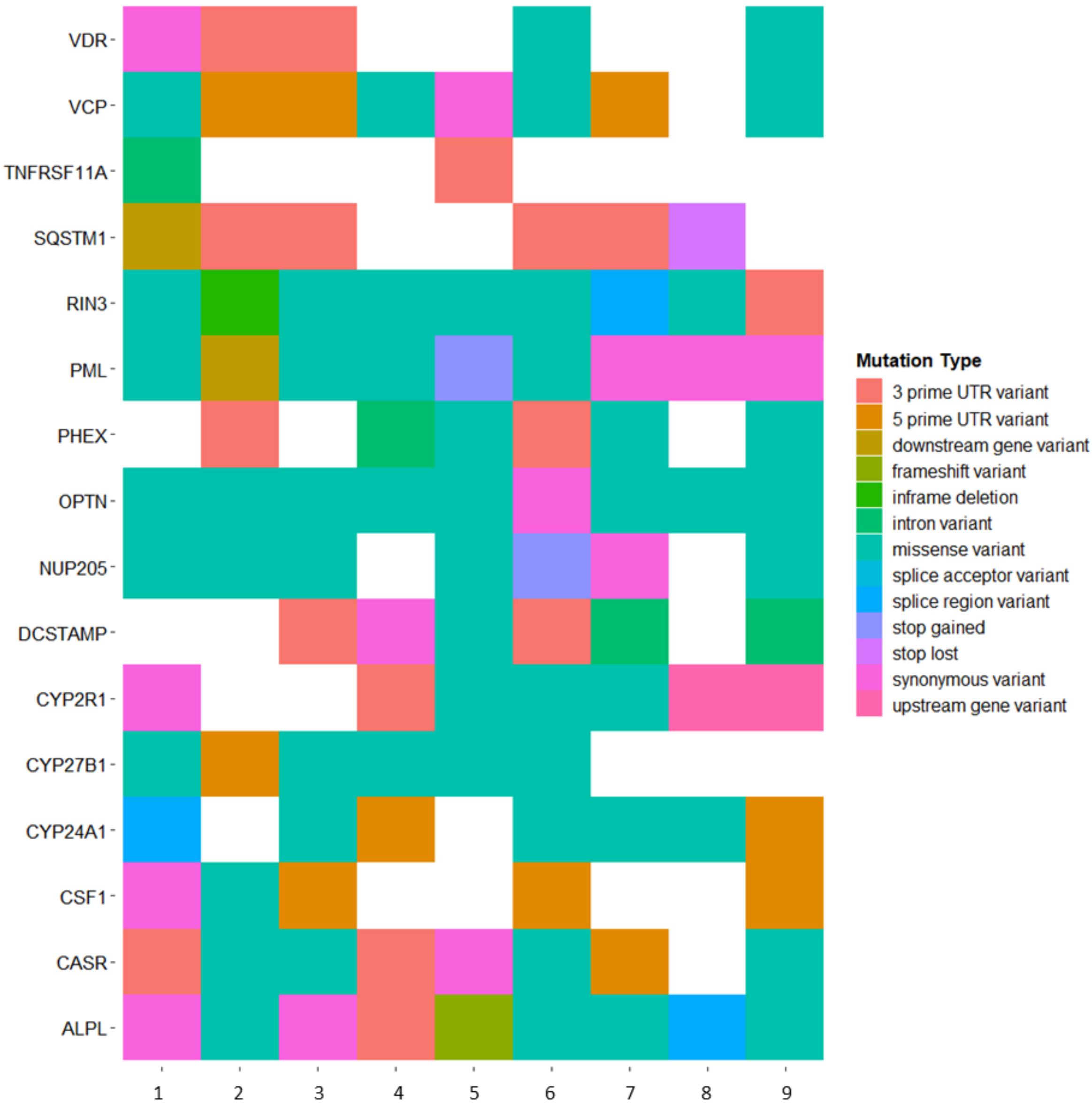


Figure 4.3: Different mutation types identified by WES in our patient cohort for the bone panel genes of interest. Only coding germline and non-silent somatic mutations were included with >10 total reads. Of 304 variants identified, 57% were missense variants, which are single base pair substitutions that alter the amino acid produced at this location. All patients, excluding patient 6, harboured missense variants in the PDB associated gene *OPTN*.

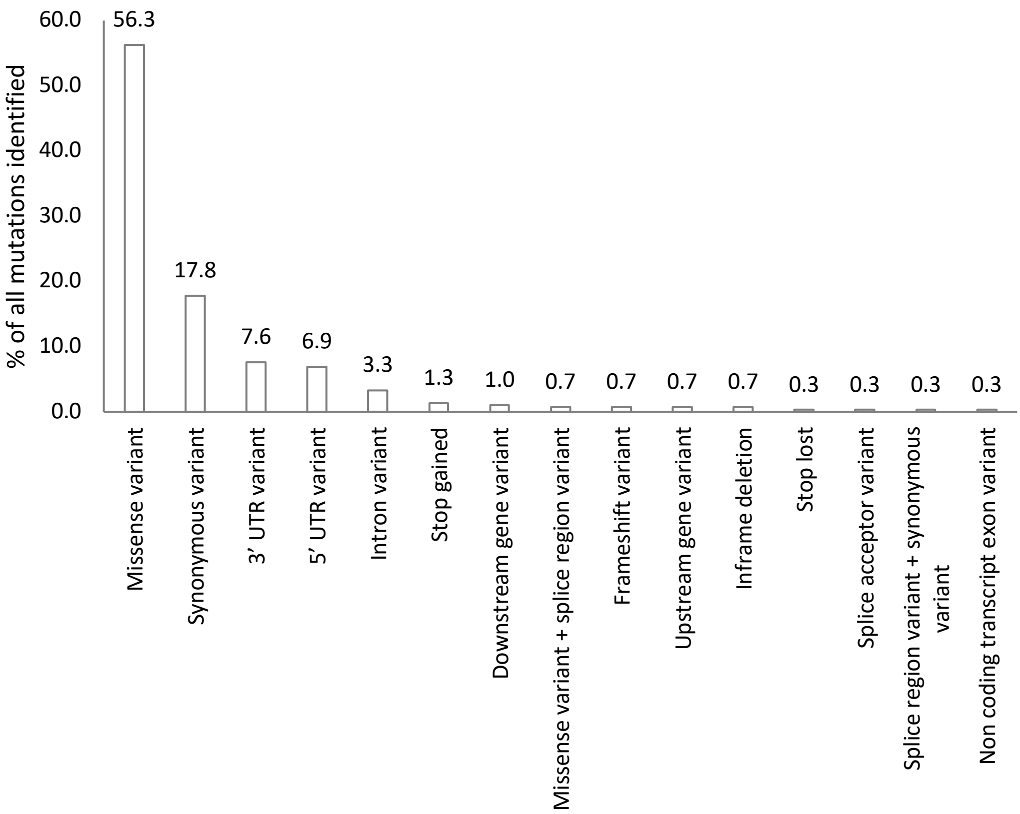


Figure 4.4: Percentage of each WES mutation type identified in the bone panel. Of all variants, 57% were identified as missense with the second most common variant being synonymous variants at 17.8%.

While the biochemical analysis of this patient cohort was indicative of potential underlying *CYP24A1* hypomorphic mutations, e.g., inappriately elevated 1,25$(OH)_2$D, abnormal VMR and/or elevated calcium, no germline mutations were identified in the *CYP24A1* protein coding region. WES confirmed that 7 patients had somatic *CYP24A1* mutations including missense (44%), 5' UTR (22%) and splice region variants (11%) (Figure 4.5). All *CYP24A1* variants identified were SNVs bar one exception in patient 6, who harboured a *CYP24A1* in frame deletion (Table 4.6).

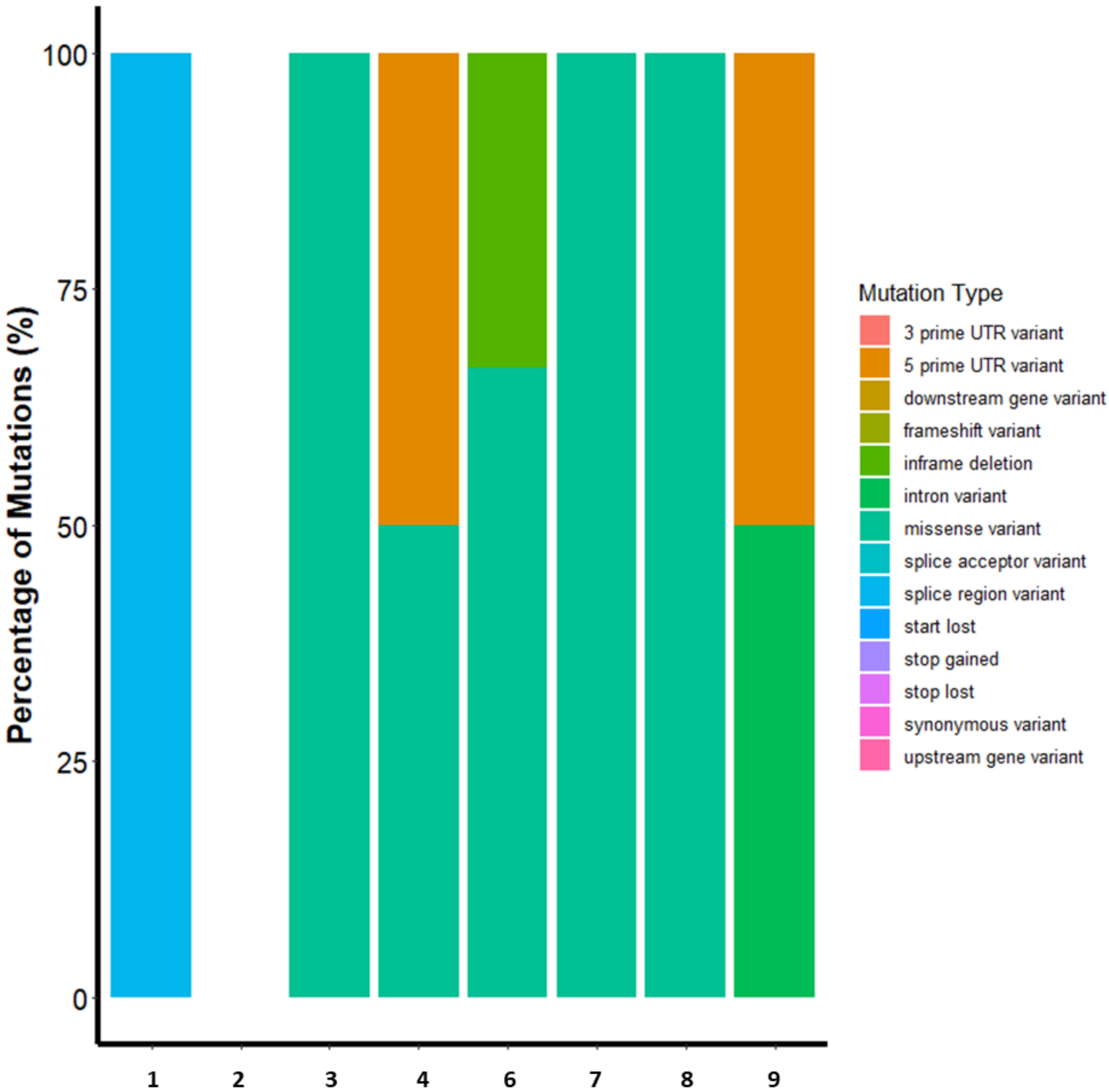


Figure 4.5: *CYP24A1* somatic mutation type identified in our patient cohort (n = 9). All patients except patient 2 harboured somatic *CYP24A1* patients with the majority being missense variants.

Table 4.6: *CYP24A1* variant location and type identified in WES analysis (n = 9). All mutations, except one in frame deletion in patient 6, were SNVs.

Patient	***CYP24A1* Variant**	**Genomic Location**	***Variant Type***
4	T>G	54173564	Missense variant
	A>T	54173607	5' UTR variant
6	A>C	54157464	Missense variant
	CCTTCTTTG > CCTTTG	54172935	Inframe deletion
	A>G	54173573	Missense variant
7	G>C	54172951	Missense variant
8	A>G	54154632	3' UTR variant
	C>A	54162814	Missense variant
	T>C	54173072	Missense variant
	T>G	54173567	Missense variant
	G>T	54157596	Intron variant
	A>C	54173593	5' UTR variant
1	G>A	54173315	Splice region variant, Intron variant
3	A>T	54172909	Missense variant, Splice region variant
2	C>T	54158989	Synonymous variant

4.4 DISCUSSION

WES has been established as an effective and high throughput detection tool used to identify variants when analysing Mendelian disorders[209,253–256]. WES analysis is a steppingstone to the increase in personalised medicine due to the potential for extensive genetic profiles being utilised in personalised pharmacogenetics and subsequently further therapy[210]. This work demonstrates the strengths of implementing functional genomics and applying new “omic” techniques to established endocrine diseases.

4.4.1 BONE PANEL GERMLINE AND SOMATIC MUTATIONS IDENTIFIED BY WES

The data presented here identified many variants in genes associated with bone remodelling and calcium handling in our patient cohort. While some variants were suggestive of germline mutations, either homozygous or heterozygous, most variants identified were somatic and had a frequency of <20% meaning that they were likely not inherited. Somatic mutations are acquired by non-germline cells and can occur during the aging process either spontaneously from RNA repair mechanism errors, in response to stress or environmental factors e.g., UV radiation[257]. Somatic mutations can lead to varying disease development such as cancer. While many somatic mutations were observed in this cohort, only 6.7% were supported by >100 total reads. Total reads represent the depth of coverage. Total reads plus the variant allele frequency are important considerations when interpreting sequencing results to reduce false positives, however there is currently no standardised threshold for total read to allele frequency ratios and interpretations falls largely to the laboratories discretion. While most somatic mutations identified were poorly supported by minimal read count and therefore considered inconclusive results, patient 2 was identified with a missense variant (g.122261785 A>C) in *CASR* at an allele frequency of 13% with 225 supporting reads. While the *CASR* missense variant in patient 2 has not been previously reported, a SNV in close proximity (g.122261786 G>C) has been reported in association with familial hypocalciuric hypercalcemia[258]. Patient 3 was also identified with a missense variant (g.122284824 A>C) in *CASR* at an allele frequency of 6% with 712 supporting reads. The *CASR* missense variant identified in patient 3 has previously been reported in relation to familial hypocalciuric hypercalcemia. Due to the high number of supporting reads in the *CASR* missense variants identified in patient 3 and 8, there is greater confidence them being a true somatic variant rather than false positives.

Homozygous mutations were identified in *ALPL, CASR, CSF1, CYP2R1, NUP20, PHEX, RIN3, SQSTM1* and *TNFRSF11A* while heterozygous mutations were identified in *ALPL, CASR, CSF1, NUP20, PHEX, PML, RIN3, SQSTM1, TNFRSF11A* and *VDR*. No *CYP24A1* germline mutations were identified. The mutations observed in *CYP24A1* had a variant allele frequency of <20% indicating somatic mutations. The average supporting total reads to *CYP24A1* somatic mutations was 27. Due to the low variant allele frequency plus low total read count, the *CYP24A1* somatic mutations identified are inconclusive and would need improved sequencing depth to either confirm the mutation or rule as a false positive in this data. Somatic mutations have not previously been reported in association with HCINF1 and hypervitaminosis D conditions.

4.4.2 LIMIATATIONS OF WES

The abundance of somatic mutations in this cohort compared to germline mutations was unexpected, suggesting that potential false positives may be affecting the WES results. It is worth noting that this sequencing data lacked an evaluation of true or false positives and negatives as an appropriate control cohort of patients was not available for this study.

Conventional NGS error thresholds suggest that variant frequencies below 2% are subject to high risks of false positives regardless of coverage depth[259]. There is currently no universal consensus on the minimum coverage required in clinical research for NGS meaning there is variation between laboratory thresholds[248–250]. This highlights the need for universal standardised coverage depth and parameters to be established for diagnostic NGS, which considers assay specific errors.

A study in 2015 compared the sequencing results of WES to Sanger sequencing for direct confirmation of any variants identified by WES[164]. A significant proportion of false positive SNVs (91%) was observed in variants exclusively reported by WES, along with a small proportion of false positive SNVs in whole genome sequencing (25%). In this study, only 27.2% of variants that were fully exclusive to WES were reported in the 1000 Genome database, in contrast to 84.7% of whole genome sequencing (WGS) exclusive variants identified in the same sample set[164]. The majority of WES false positive SNVs identified had a Phred CADD of >10, which may lead to futile further investigations. Phred CADD scores range from 1-99 for each variants relative to all possible substitutions within the human reference genome. A score of >10 suggest these variants to be 10% most deleterious substitutions within the human genome, while >20 suggests the 1% most deleterious variants. This study highlighted that 3% of coding SNVs that were identified in WGS were missed by WES.

Additionally, 44% of indel variants identified by WES were false positive with most copy number variants (CNVs) extending outside the target region, highlighting WES lack of reliability in identifying CNVs[164].

A study in 2020 sought to evaluate different WES methodologies when paired with short read sequencing, including TrueSeq DNA Exome (Illumina, Cambridge, UK), which was utilised in this thesis[260]. Many drawbacks to the TrueSeq DNA Exome solution were identified, which could explain the potential for increased false positives found in the data presented in this thesis. Of all methodologies evaluated, TrueSeq DNA Exome showed the greatest production of duplicate reads indicating that there could be limited complexity produced using this method. Greater DNA starting material should provide less duplication due to an increased number of unique molecules being present in the sample. While the TrueSeq DNA Exome requires 100 ng starting genomic DNA, which should be adequate to improve library complexity, it was shown in this study to perform the worst overall of the WES methodologies assessed. It has also been reported that tight fragment distribution due to the sonication required in the TrueSeq DNA Exome protocol, rather than mechanical fragmentation, could lead to increase in duplicates regardless of the starting quantity of genomic DNA required. The TrueSeq DNA Exome method was shown to have a higher GC bias of all the exome sequencing kits compared[260]. GC bias is a common NGS issue that can arise during PCR amplification or reduced capture probe hybridisation efficiency[261]. Greater on-target rates indicate high probe specificity, excellent quality probes and efficient hybridisation. TrueSeq DNA Exome was also shown to have a major weakness due to a high percentage of off target reads with many mapping >200 bp from the target enrichment site[262].

While vast amounts of genetic information can be obtained from WES, there is a tendency to halt further genetic investigations once an inconclusive or negative WES result is obtained such as the results observed in this study. As WES focuses on the exome genetic information, this could lead to many patients going undiagnosed due WES not detecting or confirming rare variants, especially in the UTR. Additional genetic testing, such as direct sequencing, has been shown to diagnose rare genetic variants that have been missed by WES alone[263].

Due to the frequency of somatic mutations with minimal supporting reads identified in this study, these mutations would need to be confirmed through direct sequencing e.g., Sanger sequencing. Previous studies have used Sanger sequencing to confirm WES results with variant frequencies <5%[264]. Sanger sequencing is also able to investigate any variants outside of the target genes coding region. While WES possesses many advantages in identifying disease associated variants in the coding region, less is known about the potential disease associated variants in the non-coding regions. Variants in the non-coding region of DNA can alter gene activity and have been shown to cause structural changes affecting gene function.

4.4.3 CONCLUSION

This work demonstrates that when insufficient coverage is reported in WES data, there is increased risk of false positive results. When specific genes are investigated due to correlations with clinical presentation, the coverage of WES analysis performed on the genes of interest should be evaluated before a conclusion is reached. When investigating patients with complex phenotypes and unknown genetic variants, it is important to consider that including more genes is not always better. WES lack of full coverage across the genome in comparison to targeted gene sequencing should be considered. Genes of interest with poor coverage by WES may need to be re-sequenced by targeted sequencing, which increases coverage of clinical targets. Future standardisation of WES coverage and quality measures could enable WES to become more reliable and efficient in diagnosis of rare genetic disease without the need for targeted confirmation.

This study indicates that in our suspected CMH patient cohort, no germline *CYP24A1* mutations were identified. The WES technology used lacks in depth sequencing of the non-coding regions and UTR of each gene, focusing on the exons/protein coding regions. The lack of protein coding region *CYP24A1* variants in patients, with a phenotype and biochemical profile suggestive of CMH, has been similarly reported in previous studys[125,148]. If WES sequencing is unable to detect potential disease causing hypomorphic *CYP24A1* variants within the non-coding region, this could potentially leave patients undiagnosed or misdiagnosed, which would affect subsequent patient care and treatment decisions. While the WES data presented in this study identified multiple abnormalities in our patient cohort, no germline mutations were identified in *CYP24A1*, which is not what the patient biochemistry suggested. One notable limitation of this study to be taken into consideration is the sample size of the cohort used. Future studies would benefit from performing power calculations prior to performing the WES analysis and determining an appropriate number of patients. Statistical power measures the likelihood that a statistical significance found in the sample if the effect exists in the general population. The cohort in this chapter was limited by the rare patient cohort that was available and was unlikely to reach the minimum number for significant power. Future WES analysis similar to the work performed in this study would benefit from performing power analysis, which is typically used to determine the required cohort size in a study, prior to completing the analysis. Additionally, *CYP24A1* somatic mutations have not previously been reported in association with HCINF1 and hypervitaminosis D conditions. Future studies may benefit from investigating the non-coding region via direct sequencing to confirm and potentially uncover any disease-causing variants overlooked by WES.

CHAPTER 5: mRNA STRUCTURAL ELEMENTS IN THE 3' UTR DICTATE CYP24A1 INTRACELLULAR ACTIVITY

5.1 INTRODUCTION

RNAs fold into complex structures that are critical for their function and regulation including post-transcriptional modification, localisation, translation and degradation. RNA structure and potential RNA misfolding has scarcely been studied in a clinical setting[265]. Hypomorphic mutations in the *CYP24A1* protein coding region causing persistently elevated active vitamin D metabolites have been observed in some cases of CMH and adult-onset nephrolithiasis. It is unclear why some cases present with superficial *CMH* but do not exhibit *CYP24A1* mutations. Here a combination of biochemical profiling, next generation sequencing, bioinformatics and proteomic approaches were used to examine *CYP24A1* in a patient cohort with superficial CMH.

Vitamin D plays a key role in classical calciotropic processes including calcium and bone metabolism and is postulated to contribute to non-classical disorders including cancer, diabetes and multiple sclerosis[60]. Sterol vitamin D is obtained from the diet and in the skin after photochemical conversion of 7-dehydrocholesterol undergoes transport to the liver where it is hydroxylated by CYP2R1 to form 25OHD. A second hydroxylation in the kidney by CYP27B1 generates the active systemic metabolite 1,25$(OH)_2$D that is essential for calcium homeostasis. Vitamin D metabolism is co-regulated by the activity of CYP27B1 and CYP24A1. CYP24A1 converts the precursor 25OHD into 24,25$(OH)_2$D, which is considered a catabolic product to prevent vitamin D toxicity. CYP24A1 also converts 1,25$(OH)_2$D into 1,24,25$(OH)_3$D. Both 24,25$(OH)_2$D and 1,24,25$(OH)_3$D are subject to hydroxylation and excretion.

Activity of CYP27B1 and CYP24A1 is predominantly controlled by the serum concentration of 1,25$(OH)_2$D, calcium, PTH and FGF23[266–270]. CYP27B1 is suppressed by 1,25$(OH)_2$D and FGF23 and induced by PTH. Inversely, CYP24A1 is induced by 1,25$(OH)_2$D and FGF23 and suppressed by PTH. Varying expression of CYP24A1 and CYP27B1 maintains 1,25$(OH)_2$D concentrations and subsequent calcium and phosphorous homeostasis.

As 24,25$(OH)_2$D is produced from the catabolism of 25OHD by CYP24A1, the measurement of 24,25$(OH)_2$D in serum is useful in detecting loss-of-function mutations in *CYP24A1*. A VMR such as 25OHD:24,25$(OH)_2$D serves as an indicator of vitamin D catabolic status[150]. Previous studies have shown that VMR analysis can act as an unambiguous marker of vitamin D catabolism in the diagnosis of patients with CMH[125,190,192].

Vitamin D toxicity and/or sensitivity is a presenting feature of CMH as hypomorphic mutations in *CYP24A1* cause over accumulation of serum 1,25$(OH)_2$D metabolites[127]. Infant presentation includes vomiting, failure to thrive, colic and in rare cases neonatal death. Adult presentation can include flu-like symptoms, hypercalciuria and renal stone formation. In some female patients this is triggered by pregnancy itself[128]. Several studies have failed to observe *CYP24A1* mutations despite patients presenting with apparent CMH[148].

This study used the VMR to screen patients with suspected CMH. A separate cohort to those in previous chapters consisting of 47 patients with clinical presentations suggestive of *CYP24A1* hypermorphic mutations underwent *CYP24A1* direct sequencing. Sanger sequencing identified 6 out of 47 (13%) patients who lacked mutations in the *CYP24A1* protein coding region but exhibited SNVs in the 3' UTR. Given that the 3' UTR is of significant regulatory importance and that RNAs fold into complex structures that are critical for their function and regulation including post-transcriptional modification, localisation, translation and degradation[271–276], this study hypothesised that SNVs in the 3' UTR affect *CYP24A1* mRNA structure leading to the extremely heterogeneous phenotypes observed in CMH.

5.2 CLINICAL SAMPLES

The University of East Anglia (UEA) Faculty of Medicine and Health Sciences Research Ethics Committee approved the collection and study of Human samples for non-clinical procedures investigating *CYP24A1* abnormalities (Reference: 2018–19 - 100). Forty-seven patient serum samples were collected as part of routine requests for 25OHD LC-MS/MS analysis from the Department of Laboratory Medicine at the Norfolk and Norwich University Hospital between June 2016 and June 2017. Patients were referred from the metabolic or stone former clinics. Blood samples were collected into serum gel separator tubes (BD Vacutainer) and centrifuged immediately. The serum layer was the aliquoted and stored at -20 °C until analysis.
Whole blood from the Norfolk and Norwich University Hospital metabolic and stone former clinics for genetic analysis was obtained from patients identified with inappropriate $1,25(OH)_2D$ and/or VMR plus clinical presentation of nephrolithiasis and/or hypercalciuria serum. Genomic DNA was obtained from Croydon Hospital from an infant presenting with nephrocalcinosis and Williams-Beuren syndrome (Patient 5) and from the Glasgow Children's Hospital Renal Unit from an infant presenting with nephrocalcinosis and polyurea (Patient 6). All adults or infant parents/guardians provided written informed consent to donate samples to this study.

Negative control whole blood samples were collected at the Norfolk and Norwich University Hospital blood typing service (n=10). Exclusion criteria for control samples were those with a vitamin D, calcium or other metabolic disorder clinical history. Control samples were collected using the UK NHS Research Ethics Committee decision toolkit (http://www.hra-decisiontools.org.uk/ethics/).

5.3 RESULTS

5.3.1 IDENTIFICATION OF PATIENTS WITH SUSPECTED *CYP24A1* MUTATIONS

Biochemical tests investigating VMR by LC-MS/MS plus 1,25$(OH)_2$D, calcium and phosphate concentrations by immunoassay was performed on an initial cohort of 47 patients originating from a metabolic bone clinic who presented with complications associated with hypercalcemia. The 47 patients in this cohort are independent of the patients studied in the previous WES study discussed in this thesis. Six out of 47 (13%) patients were selected for further study based on their biochemical profile indicating superficial CMH (Table 19). All adult patients (n=4) demonstrated recurrent stone formation and both infants presented with hypercalcinosis. Except for one infant (Patient 5) who was marginally above the reference range, 25OHD was in the reference range (Table 5.1). Except for one adult (Patient 1) who was in the reference range, 1,25$(OH)_2$D was markedly elevated (Table 5.1). Except for one adult (Patient 1) all the adults showed a 25OHD:24,25$(OH)_2$D VMR in the lower 25th percentile indicating abnormal 25OHD metabolism (Table 5.1). In both infants plus the adult exception case (Patient 1) the 25OHD:24,25$(OH)_2$D VMR was in the upper 75th percentile or above the upper limit indicating abnormal 25OHD metabolism (Table 5.1). Here a markedly elevated 1,25$(OH)_2$D plus low normal 25OHD:24,25$(OH)_2$D VMR was associated with hypercalciuria and nephrolithiasis (Table 5.1). Markedly elevated 1,25$(OH)_2$D plus high normal and elevated 25OHD:24,25$(OH)_2$D VMR was associated with hypercalcemia and nephrocalcinosis (Table 5.1).

Table 5.1: Serum biochemistry on patients with suspected CMH. A low 25OHD:24,25(OH)2D VMR is associated with hypercalciuria and nephrolithiasis. A high 25OHD:24,25(OH)2D VMR is associated with hypercalcemia and nephrocalcinosis. The values in red indicate that the analyte measured is outside of the reference range (shown in brackets).

Patient	**Age**	**Total 25OHD (50-120 nmol/L)**	**1,25(OH)$_2$D (55-139 pmol/L)**	**Total 24,25(OH)$_2$D (1.1-13.5 nmol/L)**	**Total 25OHD: 24,25(OH)$_2$D Relative Ratio (7-23)**	**1,25(OH)$_2$D: 24,25(OH)$_2$D Relative Ratio (11-62)**	**Adjusted Calcium (2.1-2.6 mmol/L)**	**Phosphate (0.8-1.5 mmol/L) (Adult)**
1	33	104	83	3.3	**32**	25	**3.27**	1.04
2	28	91	**262**	10.6	9	25	2.33	1.29
3	33	97	**177**	7	14	25	2.32	0.93
4	55	73	**171**	9.4	8	18	2.44	0.93
5	<1	**122**	**616**	3.5	**35**	**176**	**3.41**	**1.98**
6	<1	108	**175**	5.8	19	30	**3.2**	Not tested

5.3.2 IDENTIFICATION OF *CYP24A1* 3' UTR SNVs

Six patients were screened using whole exome sequencing plus *CYP24A1* direct sequencing. There was a lack of pathogenic mutations in the *CYP24A1* protein coding region except for Patient 6, but SNVs were detected in the 3' UTR for all patients including Patient 6 (Table 5.2). Five SNVs were detected that reside in the 3' UTR (c.1993C>T; c.2083T>C; c.2512T>A; c.2658C>G; c.2691G>A) across these six individuals (Table 5.2). One infant (Patient 6) harboured two hypomorphic mutations in the protein coding region, as observed in classical *CMH*, as well as a SNV in the 3' UTR (Table 5.2). Initial investigations into whether these SNVs introduced *de novo* and /or mutated endogenous miRNA recognition elements showed no causative abnormal RNA silencing.

Table 5.2: Direct sequencing results for 6 patients with 3' UTR mutations in *CYP24A1.*

Patient	SNV	Genotype	Location
1	c.2083T>C	Homozygous	3' UTR
2	c.1993C>T	Homozygous	3' UTR
	c.2658C>G	Homozygous	3' UTR
3	c.2083T>C	Homozygous	3' UTR
	c.2512T>A	Homozygous	3' UTR
4	c.2658C>G	Homozygous	3' UTR
5	c.2691G>A	Homozygous	3' UTR
6	c.368insC	Homozygous	Coding Region
	c.1144insT	Homozygous	Coding Region
	c.2083T>C	Homozygous	3' UTR

5.3.3 3' UTR SNVs ASSOCIATED WITH *CYP24A1* DO NOT INDUCE *de novo* miRNA RECOGNITION ELEMENTS

3' UTR SNVs in *CYP24A1* are causative for RNA misfolding *in silico*. All six genotypes were uploaded to RNAfold to produce graphical outputs visualising predicted RNA structure abnormalities (Figure 5.1-5.6). Mountain plot representation of the MFE structure and thermodynamic ensemble of RNA structures were produced, plus the centroid structure for each genotype (Figure 5.7-5.13, coloured graphs). The entropy for each nucleotide was also generated (Figure 5.7-5.13, black graphs). mRNA structures for each patient are visibly abnormal when compared to the wild type *CYP24A1* mRNA structure.

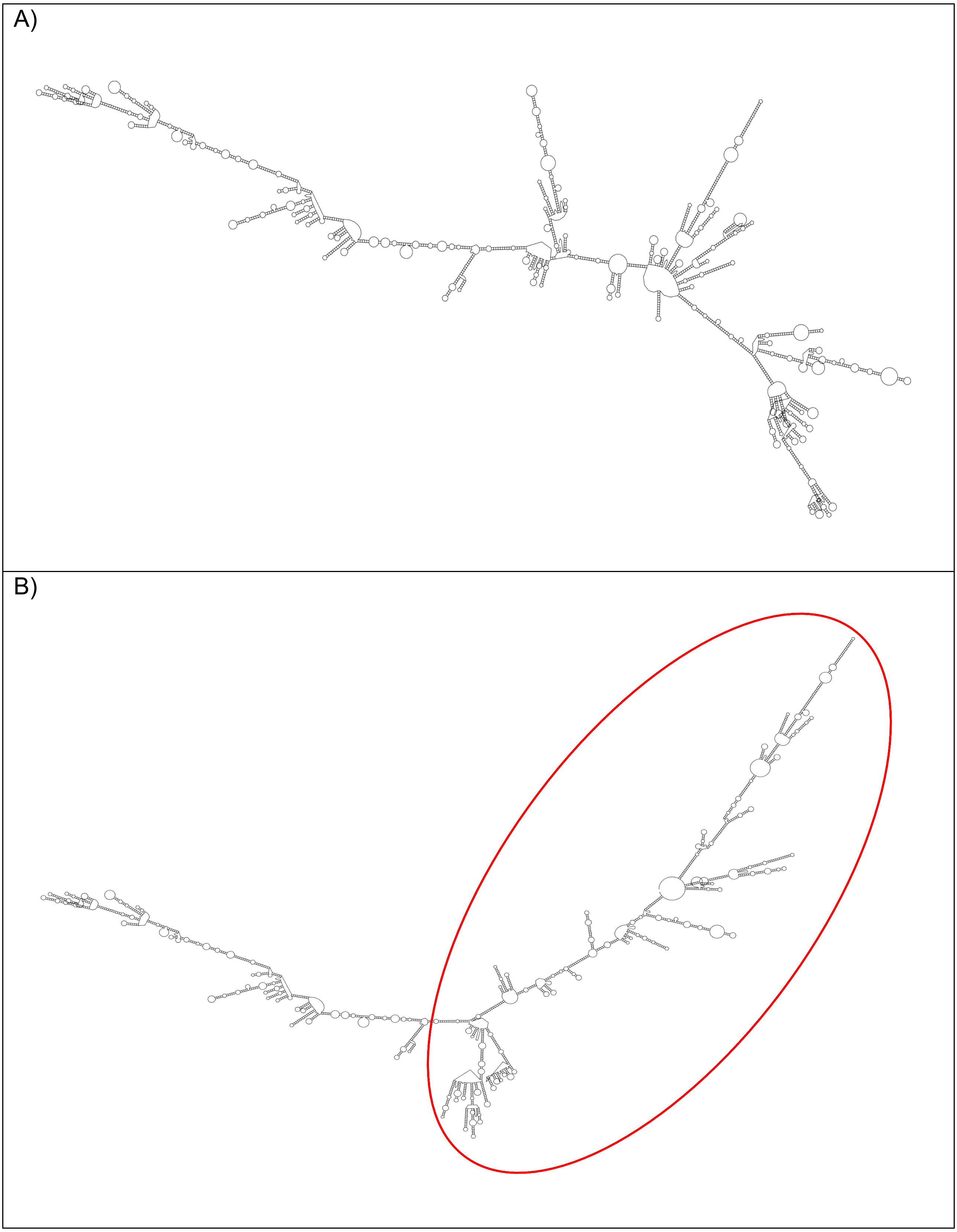


Figure 5.1: Predicted MFE-based wild type (A) and patient 1 c.2083T>C (B) *CYP24A1* mRNA structure. Structure generated by RNAFold software using NCBI Reference transcript NM_000782.5. The red circle highlights the differing structure compared to the wildtype.

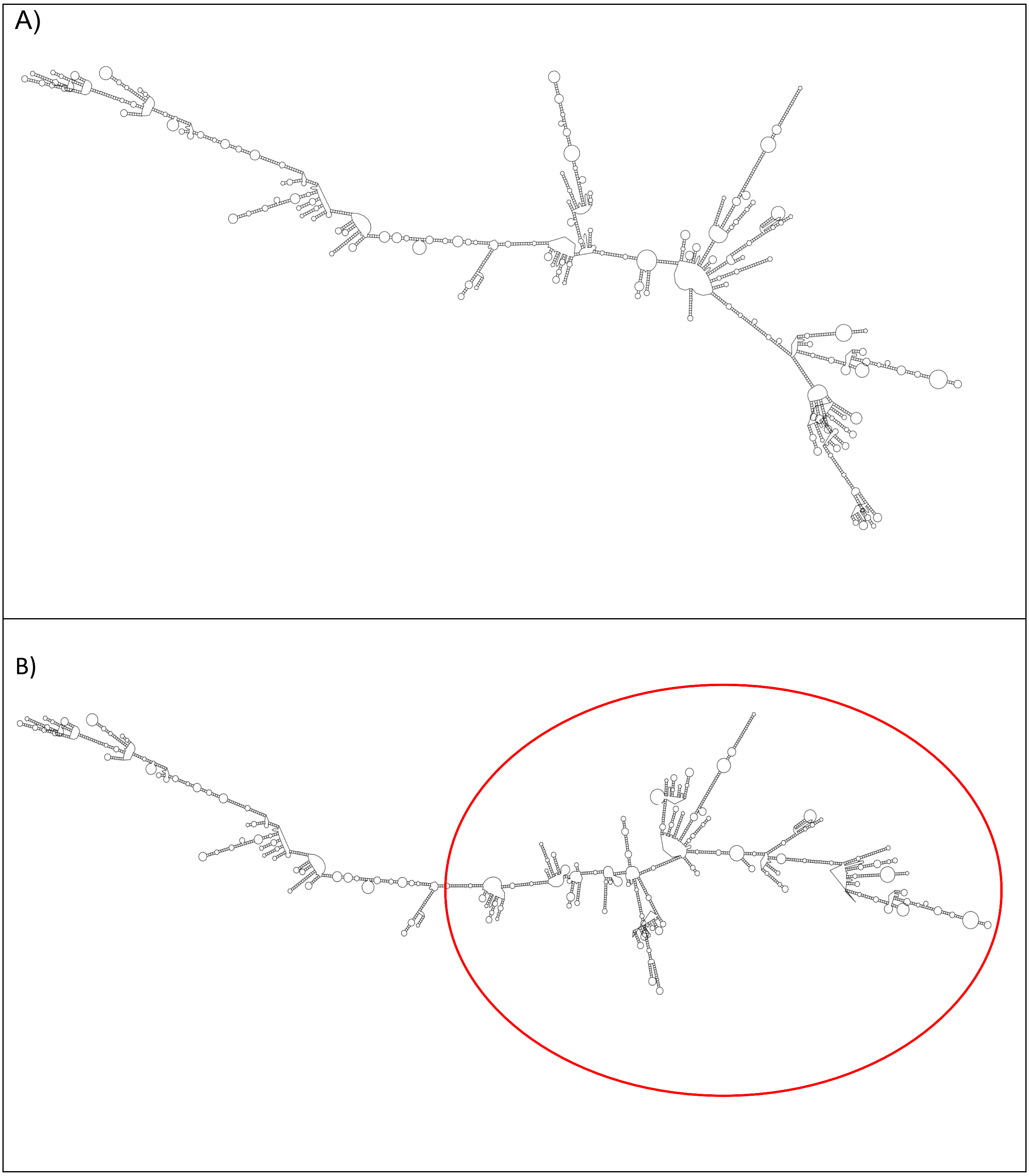


Figure 5.2: Predicted MFE-based wild type (A) and patient 2 c.1993C>T and c.2658C>G (B) *CYP24A1* mRNA structure. Structure generated by RNAFold software using NCBI Reference transcript NM_000782.5.

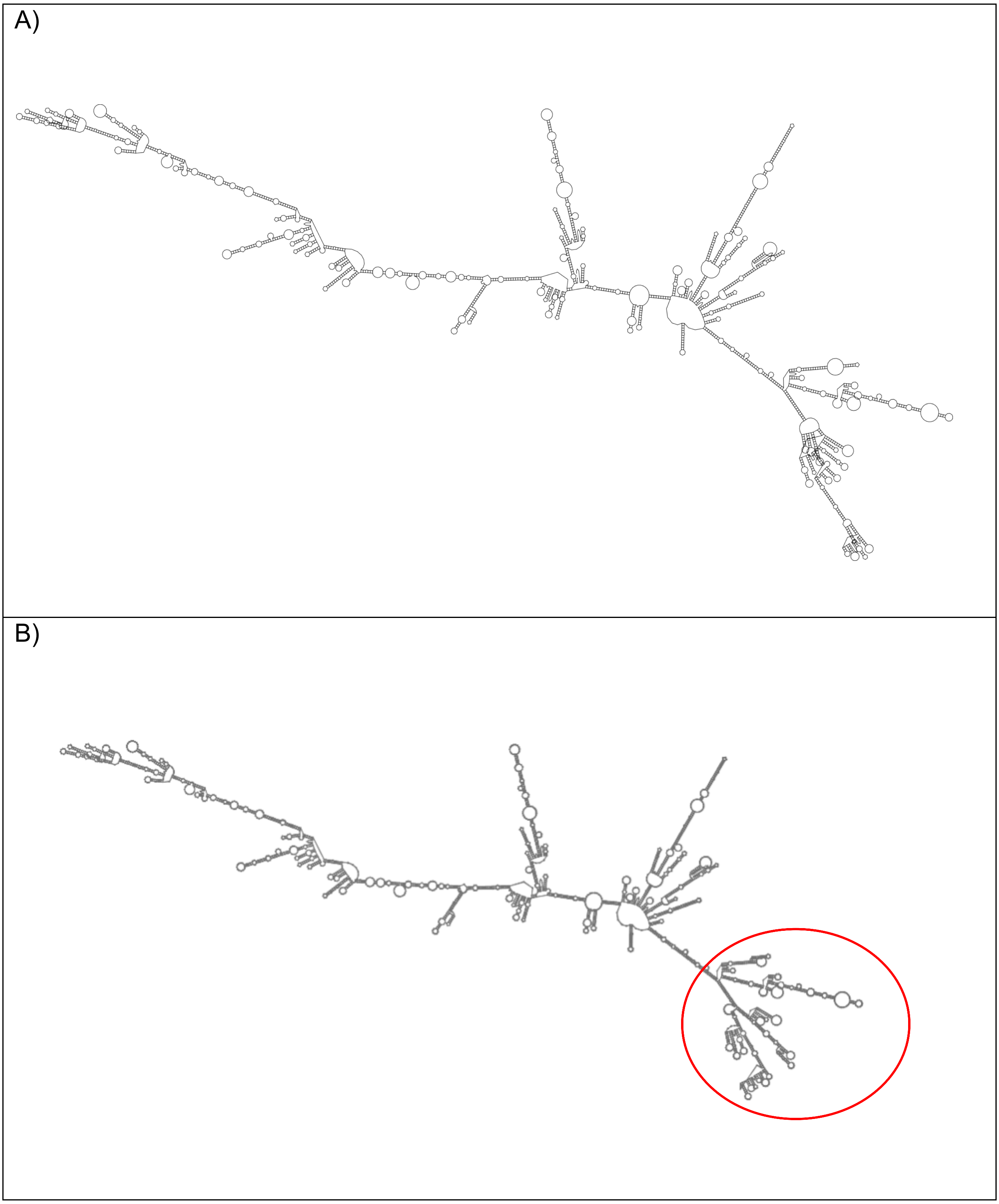


Figure 5.3: Predicted MFE-based wild type (A) and patient 3 c.2083T>C and c.2512T>A (B) *CYP24A1* mRNA structure. Structure generated by RNAFold software using NCBI Reference transcript NM_000782.5.

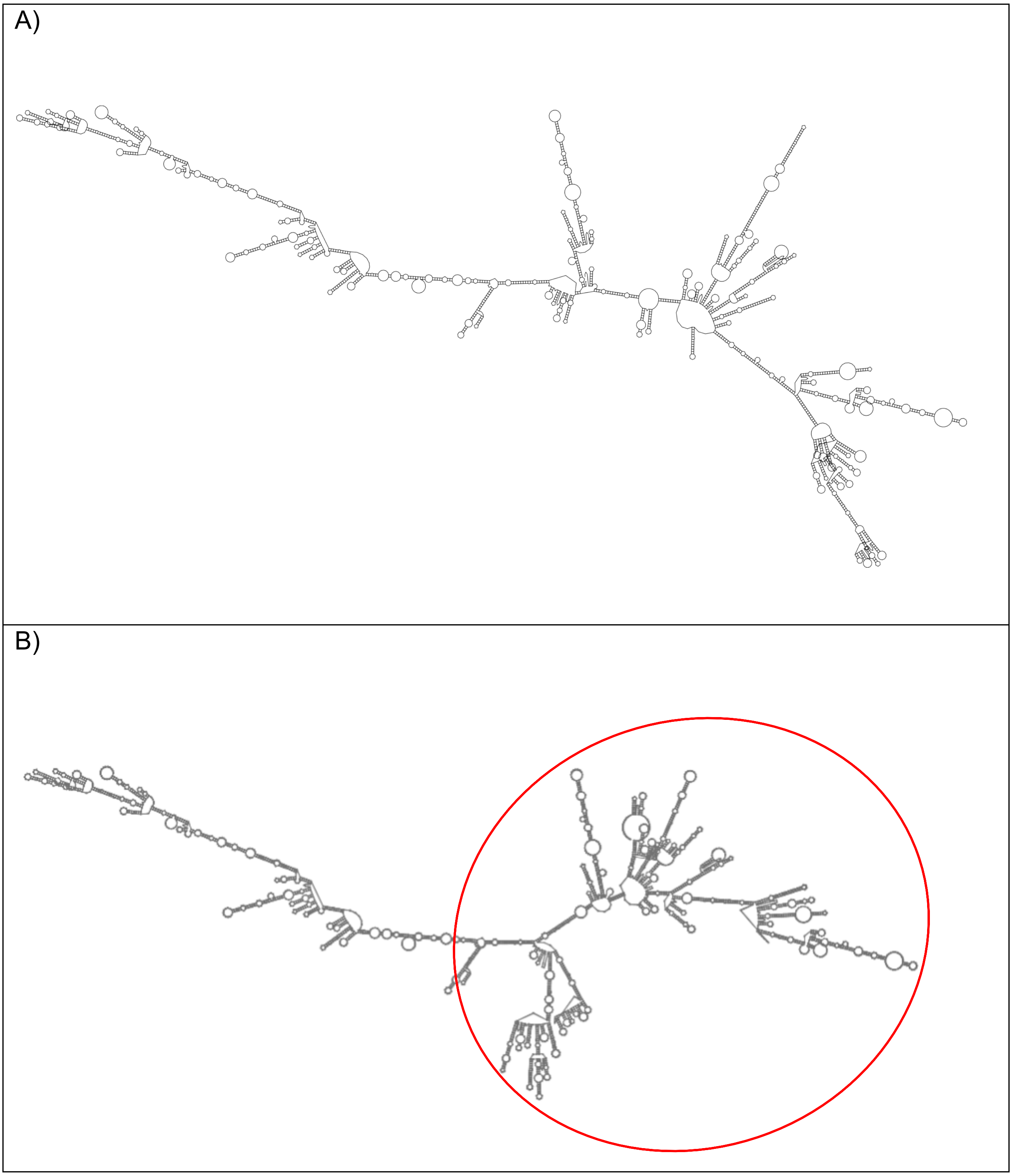


Figure 5.4: Predicted MFE-based wild type (A) and patient 4 c.2658C>G. (B) *CYP24A1* mRNA structure. Structure generated by RNAFold software using NCBI Reference transcript NM_000782.5.

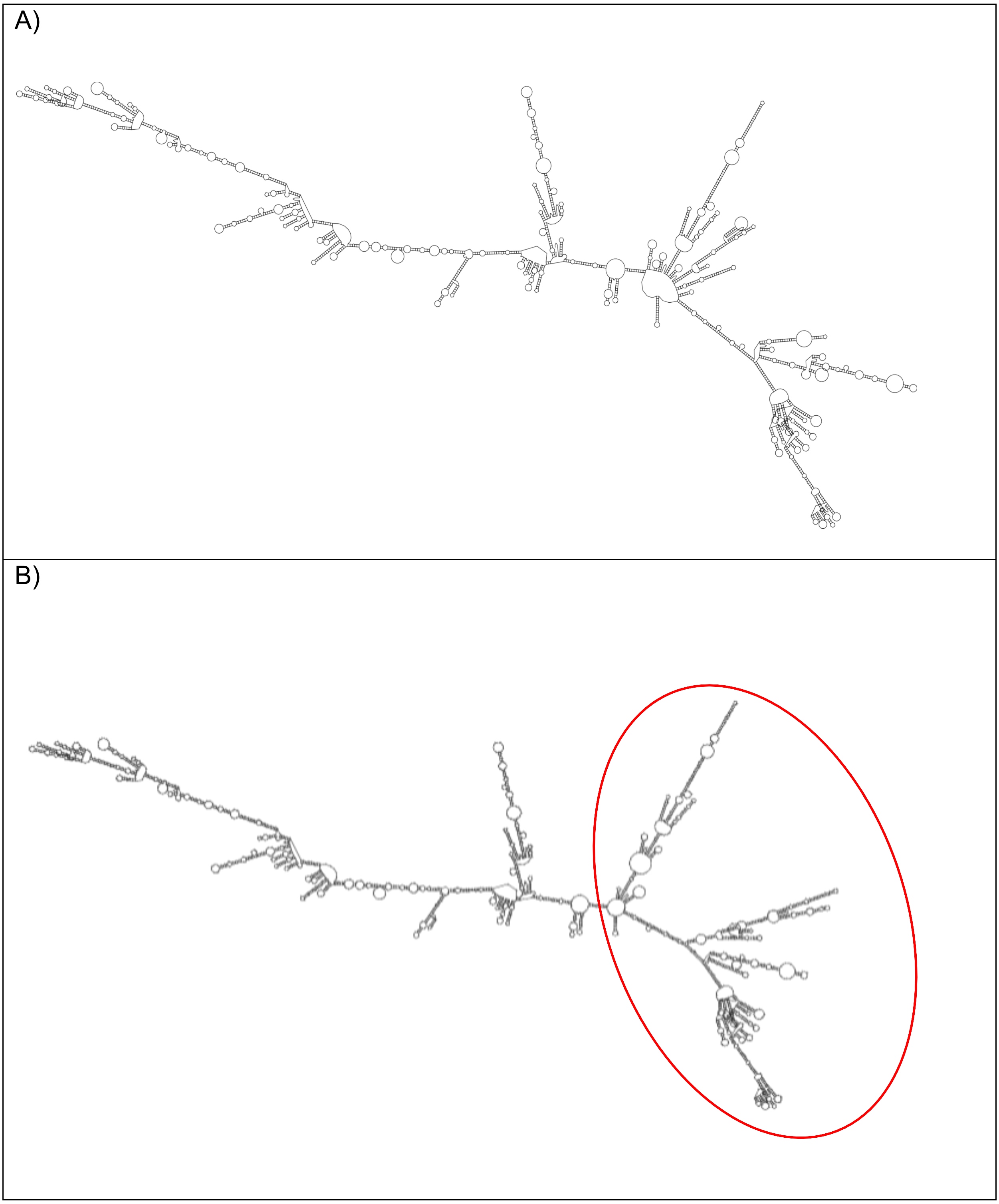


Figure 5.5: Predicted MFE-based wild type (A) and patient 5 c.2691G>A. (B) *CYP24A1* mRNA structure. Structure generated by RNAFold software using NCBI Reference transcript NM_000782.5.

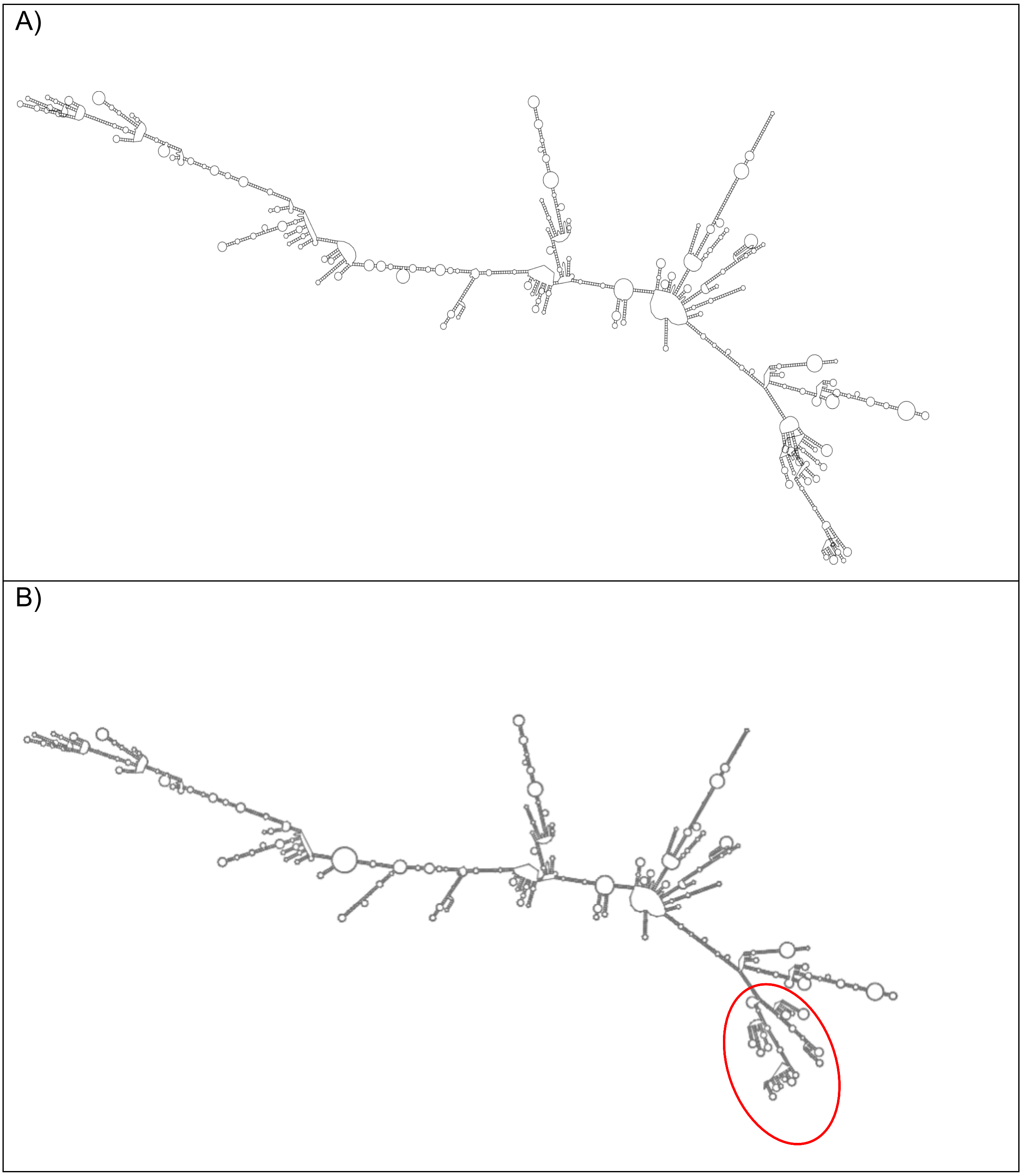


Figure 5.6: Predicted MFE-based wild type (A) and patient 6 c.368insC, c.1144insT, c.2083T>C. (B) *CYP24A1* mRNA structure. Structure generated by RNAFold software using NCBI Reference transcript NM_000782.5.

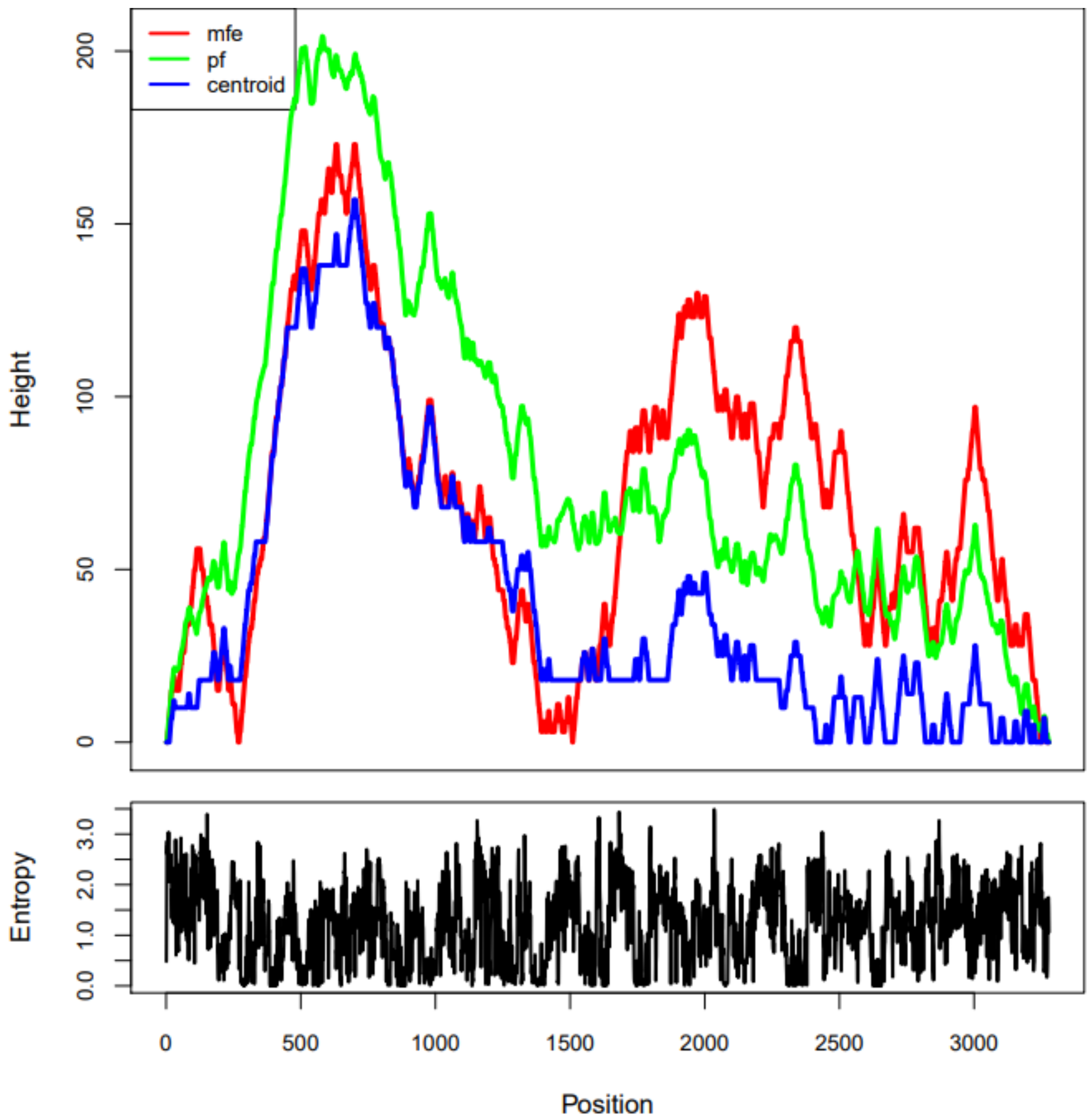


Figure 5.7: Mountain plot representation of the MFE based RNA structures of wildtype *CYP24A1*. RNA structure thermodynamic ensemble plus the centroid structure. Mountain plots represent structure in a plot of height versus position where the height m(K) is given by the number of base pairs enclosing the base at position k i.e., loops correspond to plateaus, hairpin loops are peaks, helices to slopes. RNAfold provides three dynamic programming algorithms including (i) minimum free energy (MFE) in red, which generates a single optimal structure using thermodynamic predictions based on the MFE generated by the nucleotide composition of the input sequence. (ii) Partition function in green that calculates base pair probabilities in the thermodynamic ensemble. (iii) Suboptimal folding that generates all suboptimal structures within a given energy range of the optimal energy in blue. The entropy for each position is also presented.

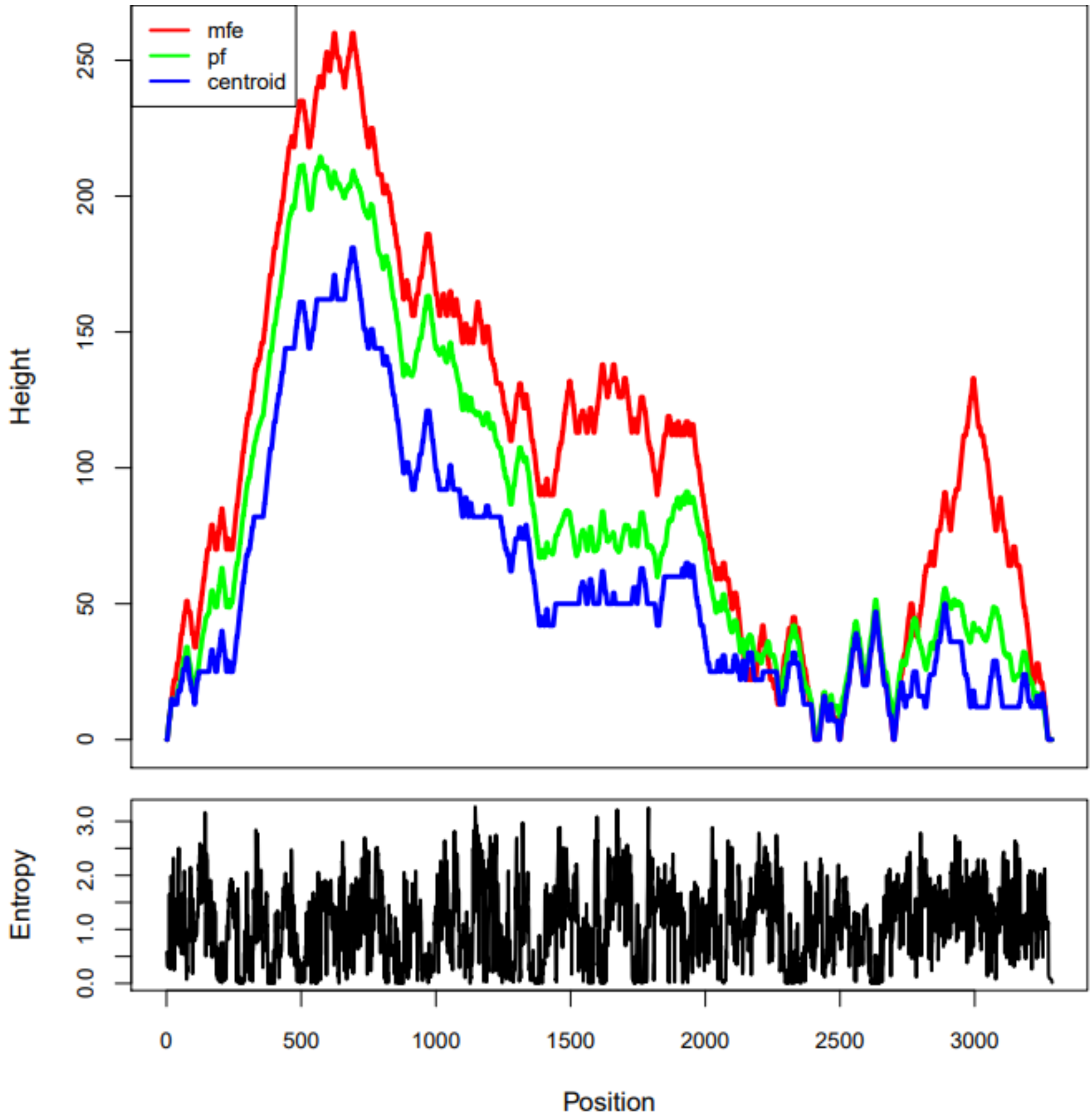


Figure 5.8. Mountain plot representation of the MFE based RNA structures from Figure 5.1B. MFE based of patient 1 genotype (c.2083T>C). RNA structure thermodynamic ensemble plus the centroid structure. Mountain plots represent structure in a plot of height versus position where the height m(K) is given by the number of base pairs enclosing the base at position k i.e., loops correspond to plateaus, hairpin loops are peaks, helices to slopes. RNAfold provides three dynamic programming algorithms including (i) minimum free energy (MFE) in red, which generates a single optimal structure using thermodynamic predictions based on the MFE generated by the nucleotide composition of the input sequence. (ii) Partition function in green that calculates base pair probabilities in the thermodynamic ensemble. (iii) Suboptimal folding that generates all suboptimal structures within a given energy range of the optimal energy in blue. The entropy for each position is also presented.

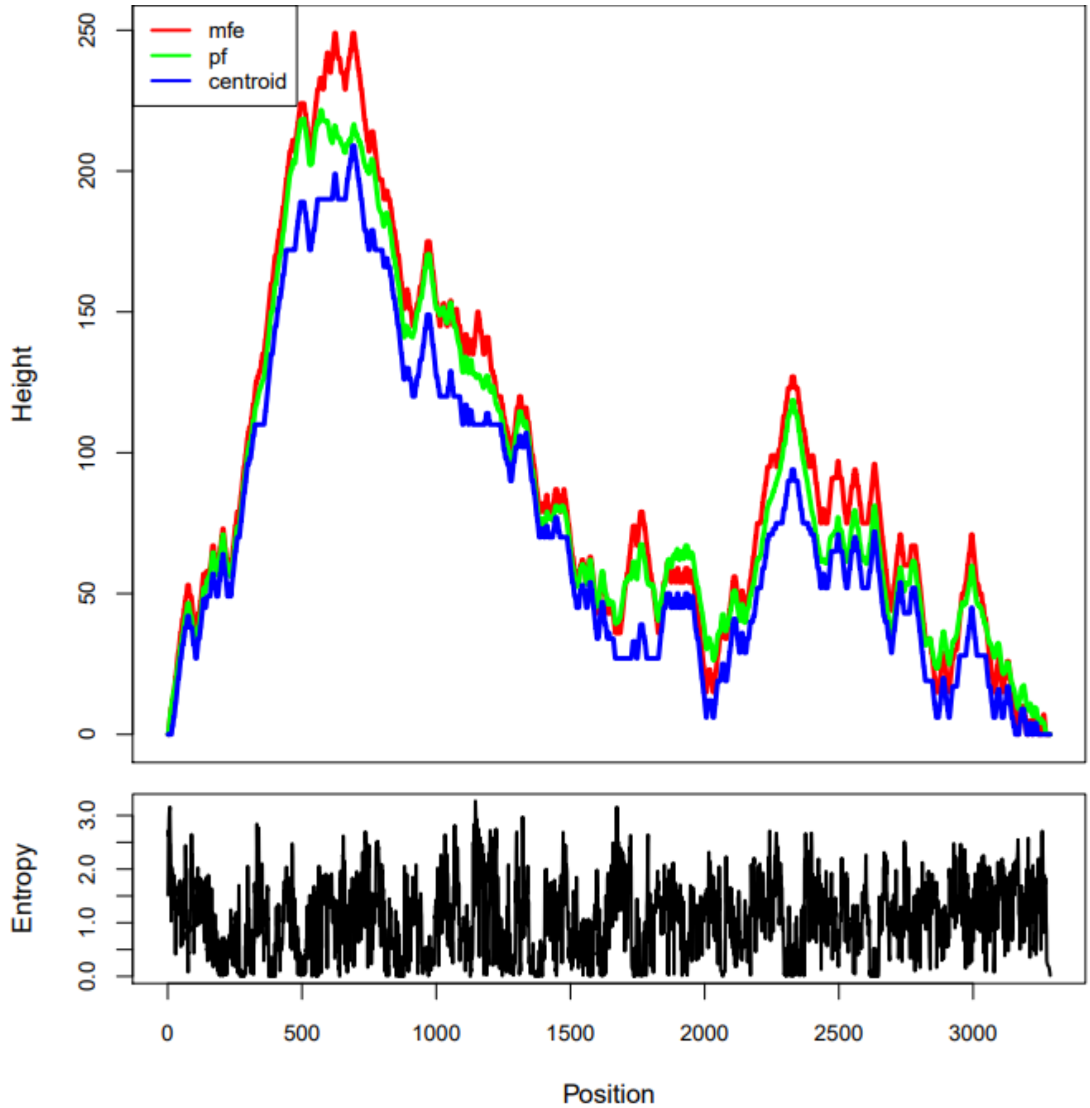


Figure 5.9. Mountain plot representation of the MFE based RNA structures from Figure 5.2B. MFE based of patient 2 genotype (c.1993C>T; c.2658C>G). RNA structure thermodynamic ensemble plus the centroid structure. Mountain plots represent structure in a plot of height versus position where the height m(K) is given by the number of base pairs enclosing the base at position k i.e., loops correspond to plateaus, hairpin loops are peaks, helices to slopes. RNAfold provides three dynamic programming algorithms including (i) minimum free energy (MFE) in red, which generates a single optimal structure using thermodynamic predictions based on the MFE generated by the nucleotide composition of the input sequence. (ii) Partition function in green that calculates base pair probabilities in the thermodynamic ensemble. (iii) Suboptimal folding that generates all suboptimal structures within a given energy range of the optimal energy in blue. The entropy for each position is also presented.

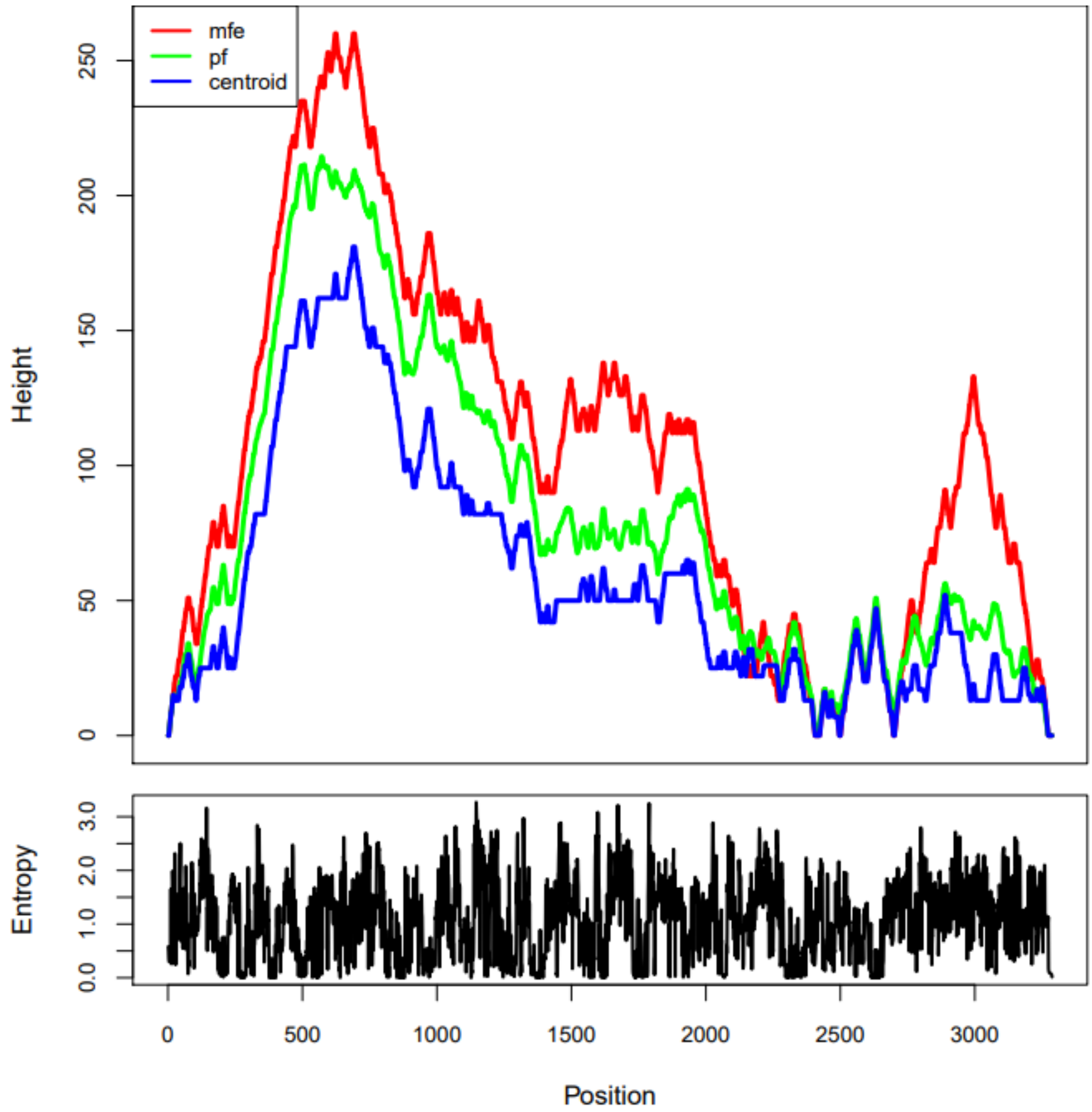


Figure 5.10. Mountain plot representation of the MFE based RNA structures from Figure 5.3B. MFE based of patient 3 genotype (c.2083T>C; c.2512T>A). RNA structure thermodynamic ensemble plus the centroid structure. Mountain plots represent structure in a plot of height versus position where the height m(K) is given by the number of base pairs enclosing the base at position k i.e., loops correspond to plateaus, hairpin loops are peaks, helices to slopes. RNAfold provides three dynamic programming algorithms including (i) minimum free energy (MFE) in red, which generates a single optimal structure using thermodynamic predictions based on the MFE generated by the nucleotide composition of the input sequence. (ii) Partition function in green that calculates base pair probabilities in the thermodynamic ensemble. (iii) Suboptimal folding that generates all suboptimal structures within a given energy range of the optimal energy in blue. The entropy for each position is also presented.

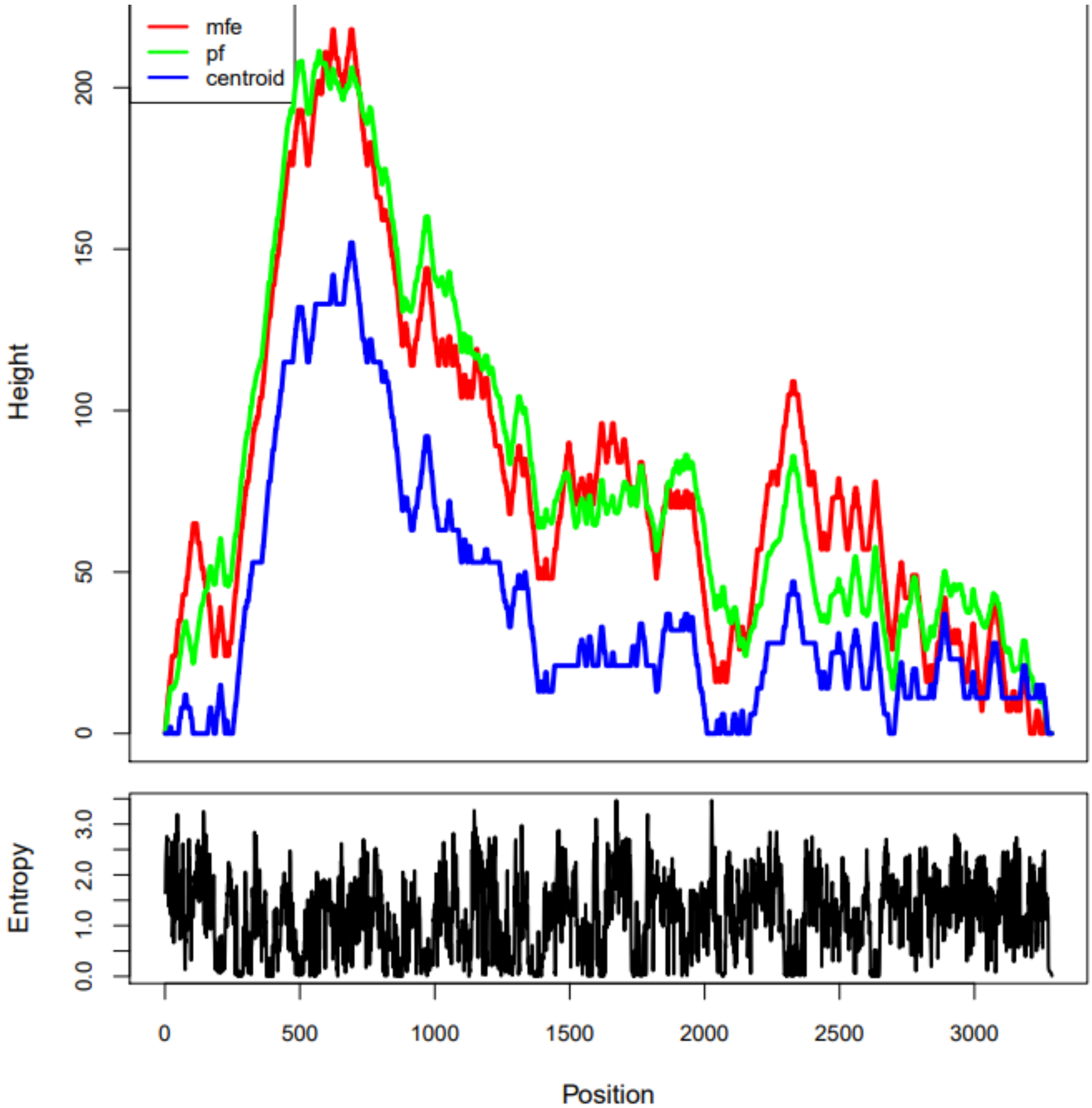


Figure 5.11. Mountain plot representation of the MFE based RNA structures from Figure 5.4B. MFE based of patient 4 genotype (c.2658C>G). RNA structure thermodynamic ensemble plus the centroid structure. Mountain plots represent structure in a plot of height versus position where the height m(K) is given by the number of base pairs enclosing the base at position k i.e., loops correspond to plateaus, hairpin loops are peaks, helices to slopes. RNAfold provides three dynamic programming algorithms including (i) minimum free energy (MFE) in red, which generates a single optimal structure using thermodynamic predictions based on the MFE generated by the nucleotide composition of the input sequence. (ii) Partition function in green that calculates base pair probabilities in the thermodynamic ensemble. (iii) Suboptimal folding that generates all suboptimal structures within a given energy range of the optimal energy in blue. The entropy for each position is also presented.

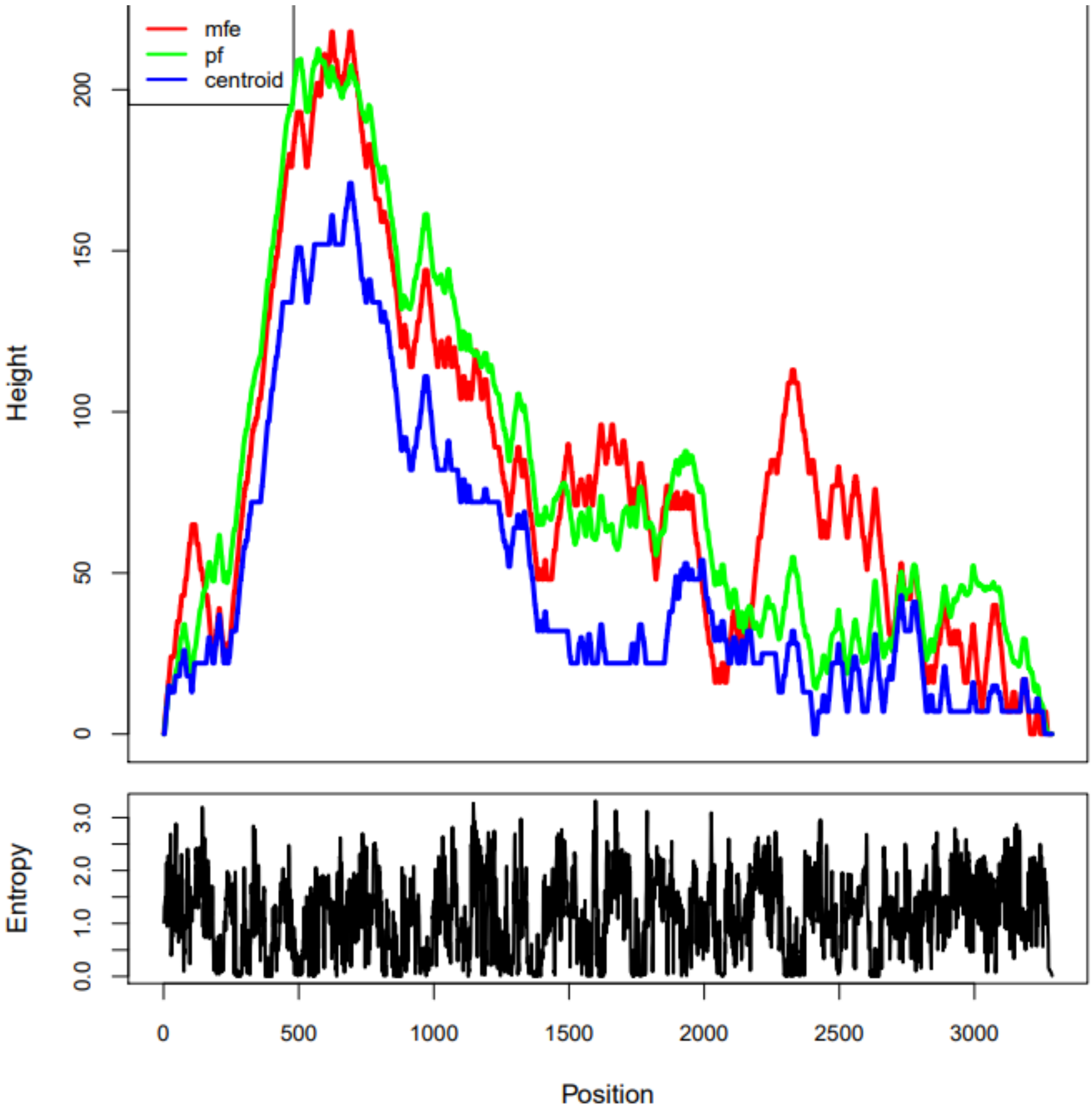


Figure 5.12. Mountain plot representation of the MFE based RNA structures from Figure 5.5B. MFE based of patient 5 genotype (c.2691G>A). RNA structure thermodynamic ensemble plus the centroid structure. Mountain plots represent structure in a plot of height versus position where the height m(K) is given by the number of base pairs enclosing the base at position k i.e., loops correspond to plateaus, hairpin loops are peaks, helices to slopes. RNAfold provides three dynamic programming algorithms including (i) minimum free energy (MFE) in red, which generates a single optimal structure using thermodynamic predictions based on the MFE generated by the nucleotide composition of the input sequence. (ii) Partition function in green that calculates base pair probabilities in the thermodynamic ensemble. (iii) Suboptimal folding that generates all suboptimal structures within a given energy range of the optimal energy in blue. The entropy for each position is also presented.

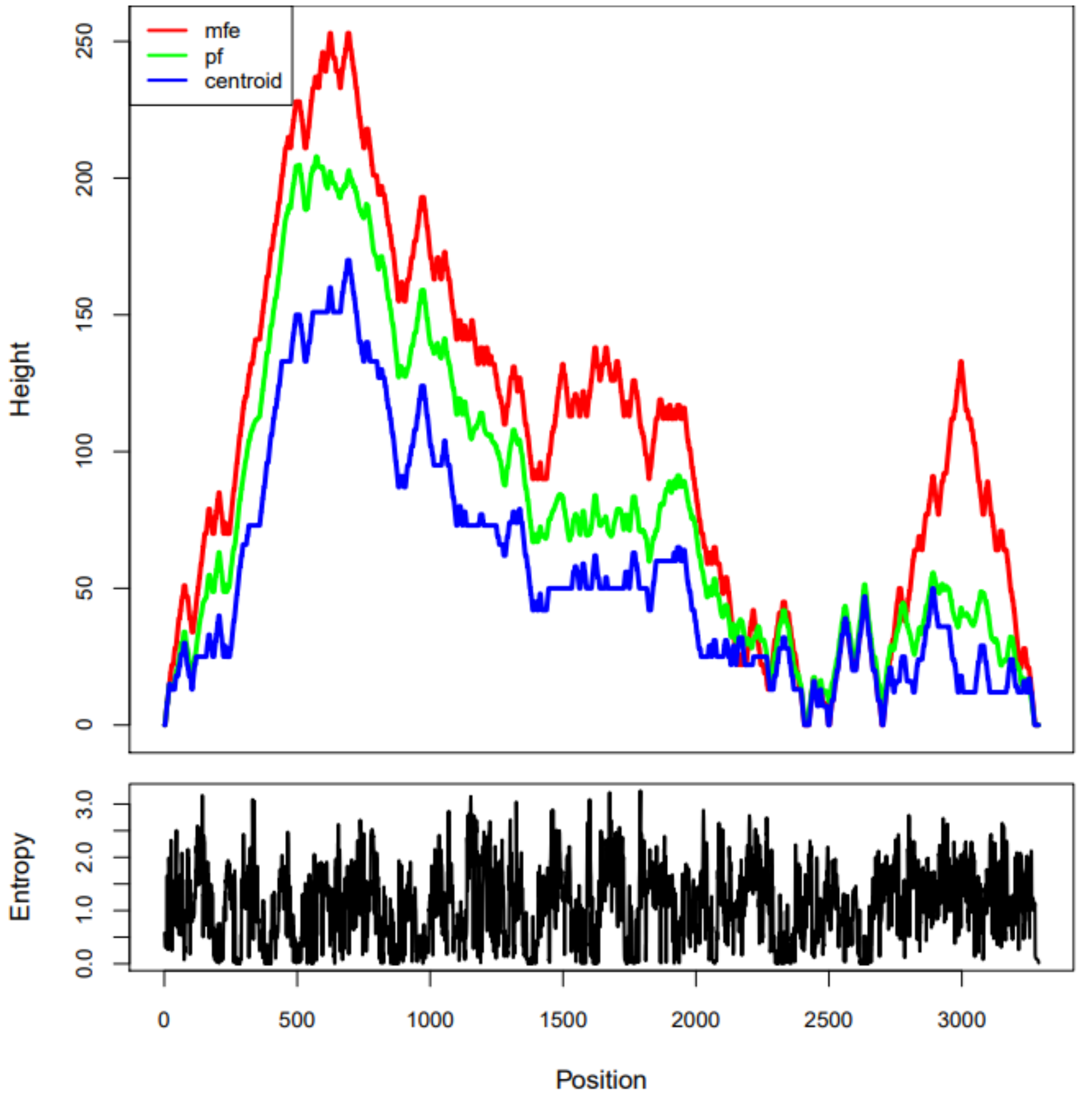


Figure 5.13. Mountain plot representation of the MFE based RNA structures from Figure 5.6B. MFE based of patient 1 genotype (c.368insC, c.1144insT, c.2083T>C). RNA structure thermodynamic ensemble plus the centroid structure. Mountain plots represent structure in a plot of height versus position where the height m(K) is given by the number of base pairs enclosing the base at position k i.e., loops correspond to plateaus, hairpin loops are peaks, helices to slopes. RNAfold provides three dynamic programming algorithms including (i) minimum free energy (MFE) in red, which generates a single optimal structure using thermodynamic predictions based on the MFE generated by the nucleotide composition of the input sequence. (ii) Partition function in green that calculates base pair probabilities in the thermodynamic ensemble. (iii) Suboptimal folding that generates all suboptimal structures within a given energy range of the optimal energy in blue. The entropy for each position is also presented.

.

5.3.4 mRNA STRUCTURAL ELEMENTS DICTATE CYP24A1 INTRACELLULAR ACTIVITY

Genotype to phenotype includes abundant regulatory stages so this study sought to identify whether RNA misfolding was associated with a transcriptional or translational error. Digital PCR was used to compare *CYP24A1* gene expression between controls and three available *CYP24A1* patients. There was no significant difference in *CYP24A1* gene expression between controls and patients (Figure 4.14) demonstrating that *CYP24A1* mRNA is transcribed similarly between controls and patients with 3' UTR mutations. The effect of RNA misfolding in association with translation was then determined. Western blot analysis was performed and a significant increase and/or over accumulation of intracellular CYP24A1 protein in patients was observed (Figure 4.15 A and B). This observation was supported by ELISA analysis that confirmed significant increase and/or over accumulation of CYP24A1 in patients with 3' UTR abnormalities (Figure 4.16).

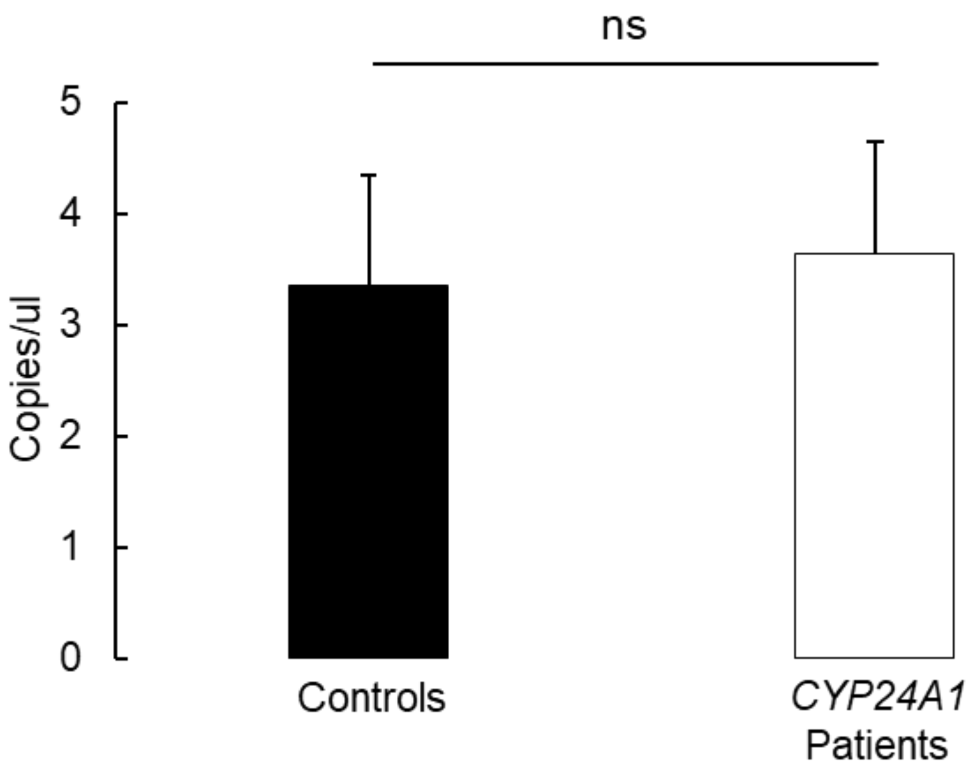
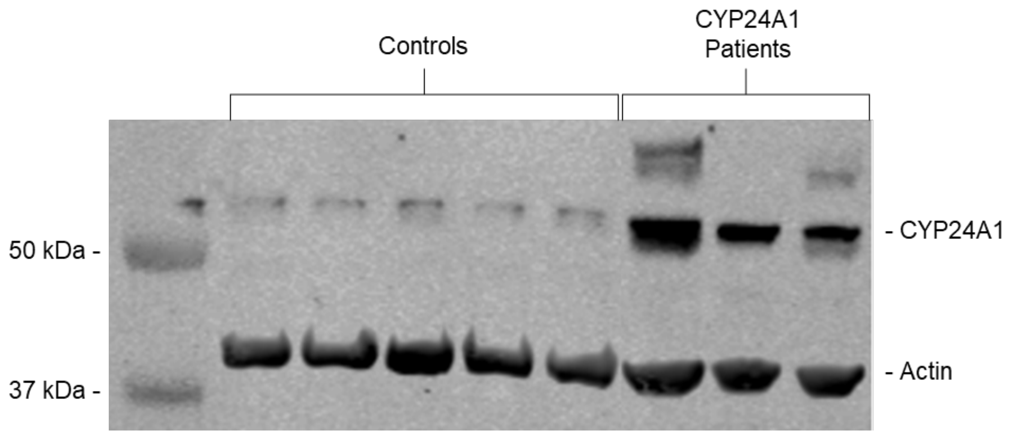


Figure 5.14: Digital PCR analysis for *CYP24A1* gene expression. No significant difference (p = 0.07) between controls (n=5) and patients (n=3) was observed . Data is reported as copies/µl as calculated by Poisson distribution. Error bars represent standard deviation. Statistical significance was calculated using an unpaired *t* test with Prism 6 (GraphPad). Three patients were available for *ex vivo* studies.

A)



B)

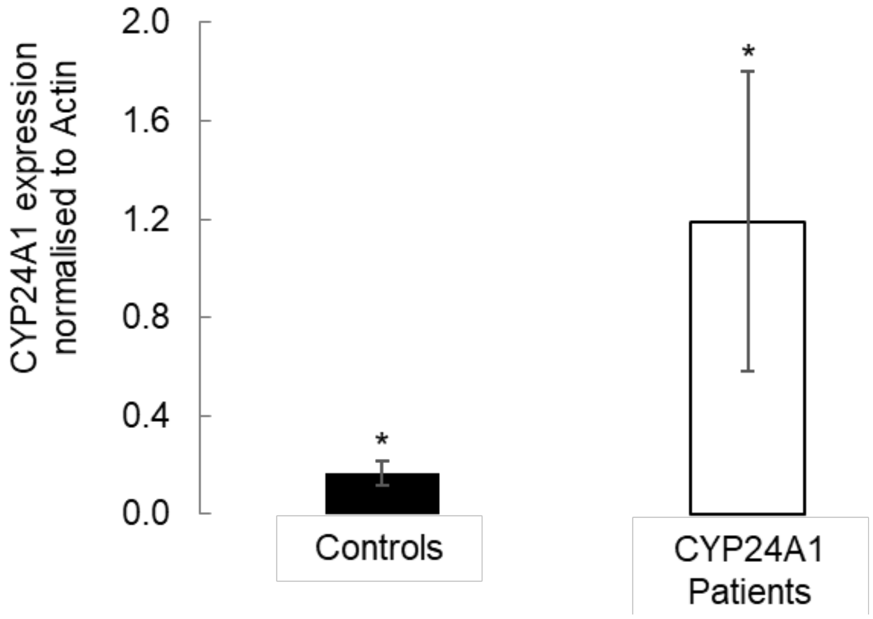


Figure 5.15: Western blot analysis of CYP24A1 expression A) Western blot analysis of PBMCs using a mouse anti-CYP24A1 antibody shows an increase and/or over accumulation of CYP24A1 protein in patients (n=3) when compared to healthy controls (n=5). Actin was used as a loading control. **B)** Density analysis of the western blot data to show increase and/or over accumulation of CYP24A1 protein in patients when compared to controls. Error bars represent standard deviation. Statistical significance was calculated using an unpaired *t* test (*p = 0.023). Three patients were available for *ex vivo* studies.

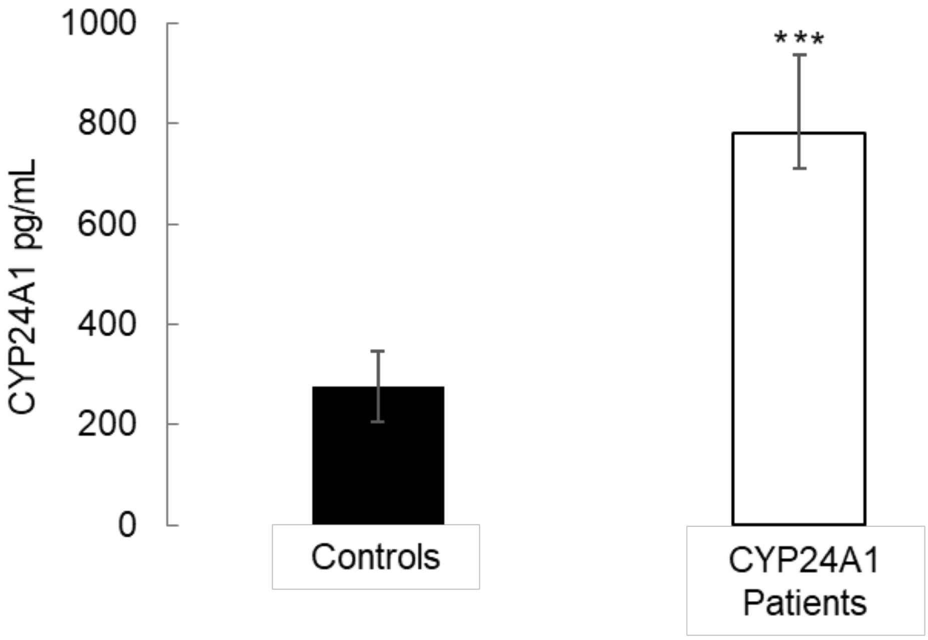


Figure 5.16: ELISA analysis of CYP24A1 expression. ELISA analysis of plasma samples confirmed increase and/or over accumulation of CYP24A1 in patient samples (n=3) when compared to healthy controls (n=5). Error bars represent standard deviation. Statistical significance was calculated using an unpaired *t* test (****p* = 0.008). Three patients were available for *ex vivo* studies.

5.4 DISCUSSION

SNVs in the *CYP24A1* protein coding region are associated with an inability to catabolise the active vitamin D metabolites 25OHD and 1,25 $(OH)_2D$[127,148,126,132–135,277–279]. CMH produces an abnormal calcium handling phenotype observed as a lack of 1,25$(OH)_2D$ metabolite clearance. Some previous studies have failed to observe SNVs in the *CYP24A1* protein coding region in patients presenting with apparent CMH[141,148]. This study shows SNVs in the non-coding 3' UTR region are predicted to cause RNA misfolding leading to an apparent inactive CYP24A1 protein.

Schlingmann et al produced the primary evidence for hypomorphic mutations in the *CYP24A1* protein coding region associated with an HCINF1 cohort. That study outlined the importance of CYP24A1 in the hydroxylation of 1,25$(OH)_2D$ in Humans[127]. The study suggested that the varied SNV locations observed were causative of the extremely heterogeneous phenotype. Subsequent studies have identified patients with *CMH* and associated *CYP24A1* inactivating mutations[132–135].

Several studies have reported that *CYP24A1* protein coding region SNVs are not present in all patients presenting with apparent CMH. Following identification of a proband with *CYP24A1* protein coding region SNVs, Dauber et al. recruited 27 patients with phenotypes associated with HCINF1. All 27 patients lacked protein coding region mutations despite displaying an analogous phenotype to the proband, demonstrating the heterogeneity of HCINF1 plus the existence of unexplained disease mechanisms[148].

5.4.1 3' UTR SNVs ALTER mRNA STRUCTURE OF *CYP24A1*

To investigate causative hypercalcemic phenotypes in our patient cohort, biochemical analysis was combined with *CYP24A1* mutational analysis that included the 5' and 3' UTR non-coding regions. Whole exome and direct sequencing identified SNVs in the 3' UTR in all six patients. The 3' UTR has significant regulatory importance including influencing mRNA secondary structure. Both RNA sequence and RNA structure are critical for proper post-transcriptional modulation, processing, localisation, translation and degradation. The diverse architecture of mRNA secondary structure is an added layer of gene regulation ensuring that the correct genes are expressed in the correct cells at the correct time[280].

For mRNA secondary structure prediction, RNAfold provides three dynamic programming algorithms including (i) minimum free energy (MFE), which generates a single optimal structure using thermodynamic predictions based on the MFE generated by the nucleotide composition of the input sequence. (ii) Partition function that calculates base pair probabilities in the thermodynamic ensemble. (iii) Suboptimal folding that generates all suboptimal structures within a given energy range of the optimal energy. For quantifying mRNA secondary structure comparison (to the control/wild type), the package contains several measures of distance (i.e. dissimilarities) using either string alignment or tree-editing. RNAfold performance was extensively tested and validated by comparing MFE predictions between RNAfold 1.8.5, RNAfold 2.1.8, UNAfold 3.8 and RNAstructure 5.7 including accuracy; sensitivity, positive predictive value, Phi coefficient and F-measure. The test set was based on 1,919 non-multimer sequence/structure pairs obtained from the RNAstrand database (all without pseudoknots in the reference structure). Both versions of

RNAfold were run with -d2 option whereas UNAfold and RNAstructure were run with default options[281].

The *in silico* mRNA structure outputs indicate that the observed 3' UTR SNVs in *CYP24A1* are predicted to affect the mRNA structure and likely result in enzymatic inactivation. Structural changes triggered by the presence of 3' UTR variants may induce translational heterogeneity or impair translation completely by destabilisation of mRNA, perturbed mRNA trafficking and/or reduced efficacy of ribosome scanning. Structural alterations caused by 3' UTR variants may impair production of functional CYP24A1 protein resulting in elevated 1,25$(OH)_2$D concentration with a low/high VMR.

5.4.2 mRNA STRUCTURAL ELEMENTS DICTATE CYP24A1 INTRACELLULAR ACTIVITY

No miRNA recognition element (MRE) abnormalities were found using miRDB. While no MRE was observed, that there is a possibility that the 3' UTR variants have removed pre-existing MREs preventing miRNA binding. Digital PCR analysis was performed to investigate potential mechanistic effects that mRNA structural changes had on transcription of *CYP24A1*. No significant effect on *CYP24A1* transcription was observed in patients with 3' UTR variants. Although no significant difference was observed between the transcription of *CYP24A1* in CMH patients and controls, the comparison was close to significant (0.07). The sample size was limited in this chapter (5 controls, 3 CMH patients) meaning that both the transcriptional and further translational comparison could benefit from a greater sample cohort in future work to support these findings by increasing the power of the study and reducing the margin of error.

The lack of transcriptional abnormalities in our analyses observed between our patient cohort and controls suggested that a translational impairment was the more likely mechanism underlying abnormal CYP24A1 enzyme function giving rise to the hypercalcaemic phenotype. Western blot and ELISA analysis was performed to investigate potential mechanistic effects that mRNA structural changes had on translation of CYP24A1. A significant increase/over accumulation in CYP24A1 was found in CMH patients suggesting a translational impairment. This study used PBMCs for the western blot analysis of CMH and control patients. PBMC represents a broad spectrum of cell types with distinct differences in cellular function that may vary between individual patients[282–284]. Bittersohl et al reported PBMCs have an average composition of T Cells (60%), Monocytes/Macrophages (15%), Natural Killer Cells (15%) and B Cells (10%)[285]. This homogeneity in PBMC samples could effect the comparison when comparing phenotypes between patients if CYP24A1 displays varied expression different cell types. PBMC may add another level of variability that interdicts their use for this purpose. Future investigations would benefit from isolating different cells within the PBMC population to compare CYP24A1 expression in patients. PBMC separation can be performed by flow cytometry and fluorescence activated cell sorting (FACS), commonly used for its ability to isolate the required cell populations allow the enrichment of even low abundant subpopulations with high purity. Additionally, magnetic activated cell sorting (MACS) can be used to isolate cells of interest using specialized magnets. MACS is typically advantages for bulk analysis due to the ease-of-use of this technique, while FACS provides higher purity of sample[286].

An increased and/or over accumulated but ineffective CYP24A1 protein in patients with mRNA structure altering 3' UTR mutations was observed. Though not explored in this chapter, there are three possible hypotheses for this observation, (i)

Upregulation is the expected response to elevated serum calcium but mRNA structural elements signal for an as yet unidentified post-translational modification that fails to clear elevated calcium. (ii) 3' UTR variants in *CYP24A1* have altered or removed pre-existing MREs preventing miRNA binding, causing upregulation of protein. (iii) mRNAs are trafficked from the nucleus to regions where the subsequent protein will be required more rapidly. mRNA localisation to specific regions within a cell provide regulation of protein expression but mRNA misfolding interferes with this trafficking process[287]. Given that CYP24A1 is functional in the inner mitochondrial membrane in 1,25$(OH)_2$D target cells and that some RNAs are known to act as structural components to the mitochondrial membrane[288], it is possible that mRNA structural abnormalities 'anchor' translational machinery and prevent the protein from proper localisation. Improper localisation affecting translational machinery could explain the increased CYP24A1 expression with little effect on mRNA transcription as observed in this CMH cohort.

5.4.3 CONCLUSION

This study highlights the importance of RNA misfolding in clinical phenotypes, although this chapter has not determined the fundamental mechanism describing misfolding to phenotype, an association between misfolding and phenotype is presented. This data provides some evidence on how RNA structural abnormalities caused by sequence dependent structural elements affect biological function. This model system is unique in the study of the role of RNA misfolding in Human disease as this thesis identified a cohort of patients with *CMH* associated phenotypes harbouring non-coding genetic abnormalities. The application of *in vivo* or probe-based mRNA structure technologies with appropriate sequencing depth should elucidate the current data more thoroughly. These techniques will provide accurate characterisation of mRNA structure-function relationships and are a key advancement in RNA biology beyond *in silico* analyses[177,289,290].

CHAPTER 6: PROBE BASED RNA STRUCTURE ANALYSIS CONFIRMS 3' UTR VARIANTS IN *CYP24A1* INDUCE mRNA STRUCTURE ALTERATIONS

6.1 INTRODUCTION

RNA secondary structure provides an added layer of regulation for gene expression depending on cell type and required gene function[280]. The dynamic structure of RNA affects post-transcriptional regulation including RNA maturation, translation and degradation. The intricate 3D conformation of RNA secondary structure has a central role in the biological function of RNA molecules, yet these structures have remained difficult to accurately predict. Understanding RNA structure using probe-based techniques would provide functional insight into the mechanism of action of *CYP24A1* in patients with predicted *in silico* mRNA structural alterations, plus confirming *in silico* findings.

Many biological components that influence the folding and function of RNA *in vivo*, including ligands, proteins and crowding interactions, are absent from *in vitro* structure preditions[280,291,292]. Recent work utilising high throughput methods of RNA structure probing in living cells have revealed significant variances between probe based RNA structure determination and *in silico/vitro* structure predictions[289,291,293]. Current research into probe-based structure predictions aim to incorporate the influence from biological functions lost during *in vitro* experimentation.

Determination of RNA secondary structure has previously relied on extensive resources and techniques including X-ray crystallography, NMR and gel electrophoresis[294]. Secondary structure prediction tools using thermodynamic parameters are useful in predicting the stability of RNA helices and hairpins however large-scale structures are difficult to predict[295]. Recent advances in determination of RNA secondary structure using chemicals that penetrate the cell and modify RNA have allowed probe-based profiling of RNA secondary structure[177,280,290,291,295,296]. Using DMS to alkylate the adenine and cytosine single stranded regions of RNA, followed by NGS, allowed the first probe-based genome-wide RNA secondary structure determination in plants[177]. This structure prediction technology utilises SHAPE technology to assess local nucleotide flexibility of the total RNA transcript[295]. By combining SHAPE technology and NGS, the entire RNA transcript can be chemically modified and the secondary structure determined from a single experiment. RNA chemical structure probing has been utilised to successfully study RNA folding in single gene and genome wide parameters[177].

One main advantage of SHAPE reagents is their reactivity with each nucleotide, meaning that they can sensitively assess the structural context of each base. Single stranded nucleotides are reactive to SHAPE reagents (small ligands such as NMIA that preferentially bind to the oxygen of 2′-hydroxyl group of RNA nucleotides) when they are not constrained. SHAPE reactivity is determined by the ability of each nucleotide to bind SHAPE reagents. SHAPE analysis can detect these flexible regions in longer segments of RNA that are not identified using computational bioinformatics[297]. Computational algorithms lack the ability to accurately identify higher order structures such as intermolecular interactions or pseudoknots (RNA structures composted of two helices connected by single stranded regions), whereas SHAPE reagents can be used to assess the structure along the entire length of RNA[297].

Previous knowledge of RNA structure formation has arisen from *in vitro* studies[298]. Prior to current methods of structure determination including chemical probing, little was known about how RNA is structured and its subsequent function *in vivo*. Many recent genome wide studies have revealed that RNA structures differ between chemical probing and *in vitro* conditions[280,291,292]. In 2014, datasets produced genome wide data containing as yet undiscovered *in vivo* information[291,291]. One drawback to fully understanding the data drawn from *in silico* methodology is the assumption that RNA can fold into a single structure only, rather than averaging probing data across all structures produced in a single experiment[295,296,299–301]. The advantages of utilising SHAPE technology include simple sample preparation, reducing the time consumption of other selective RNA enrichment steps or high abundance rRNA depletion.

Previous DMS/SHAPE RT studies have successfully determined the structures of high abundance RNAs using probe based analysis[302–305]. Low abundance RNAs are essential to many biological functions including protein synthesis however, they remain difficult to assess using chemical probing RT alone, meaning little is known about their structural alteration significance. To investigate low abundance RNAs, newly developed methods allowing 10^5-fold increase in sensitivity have been reported[280]. Using LMPCR-based DMS/SHAPE chemical probing (DMS/SHAPE-LMPCR), studies have presented the first report of low abundance RNA structures in plants[280].

The 3' UTR has significant regulatory importance including influencing mRNA secondary structure. Both RNA sequence and RNA structure are critical for proper post-transcriptional modulation, processing, localisation, translation and degradation[280]. *In silico* mRNA structure outputs presented previously in chapter five of this thesis indicate that the observed 3' UTR SNVs in *CYP24A1* affect the mRNA structure and likely result in enzymatic inactivation. Structural changes triggered by the presence of 3' UTR variants could induce translational heterogeneity or impair translation completely by destabilisation of mRNA, perturbed mRNA trafficking and/or reduced efficacy of ribosome scanning[306,307,308]. Structural alterations caused by 3' UTR variants may impair production of functional *CYP24A1* protein resulting in elevated 1,25$(OH)_2D$ concentration with a low/high VMR.

This thesis previously identified that *CYP24A1* is in low abundance in human whole blood by dPCR analysis (3.5 copies/uL). The aim of this chapter is to utilise newly developed probe based DMS/SHAPE LMPCR methodology to identify the structure of *CYP24A1* in patients with 3' UTR mutations and confirm *in silico* structural change predictions. Probe-based structural confirmation may shed light on potential greater *CYP24A1* interindividual variations at both mRNA and protein level.

6.2 CLINICAL SAMPLES

The University of East Anglia (UEA) Faculty of Medicine and Health Sciences Research Ethics Committee approved the collection and study of Human samples for non-clinical procedures investigating *CYP24A1* abnormalities (Reference: 2018–19 - 100). Forty-seven patient serum samples were collected as part of routine requests for 25OHD LC-MS/MS analysis from the Department of Laboratory Medicine at the Norfolk and Norwich University Hospital between June 2016 and June 2017. Patients were referred from the metabolic or stone former clinics. Blood samples were collected into serum gel separator tubes (BD Vacutainer) and centrifuged immediately. The serum layer was the aliquoted and stored at -20 °C until analysis.

Whole blood from the Norfolk and Norwich University Hospital metabolic and stone former clinics for genetic analysis was obtained from Patient 2 Chapter 5 who identified with inappropriate 1,25$(OH)_2$D and VMR plus clinical presentation of nephrolithiasis and/or hypercalciuria serum.

The negative control whole blood sample used in this chapter was collected at the Norfolk and Norwich University Hospital blood typing service. Exclusion criteria for control samples were those with a vitamin D, calcium or other metabolic disorder clinical history. Control samples were collected using the UK NHS Research Ethics Committee decision toolkit (http://www.hra-decisiontools.org.uk/ethics/).

6.3 RESULTS

6.3.1 DMS/SHAPE LMPCR

This study investigated the structure of the *CYP24A1* 3' UTR region from RNA extracted from whole blood of a patient with previously identified *CYP24A1* 3'UTR variants (Patient 2, chapter five). This 3'UTR region was selected due to the presence of identified SNVs presented previously in chapter 5 of this thesis that were predicted to alter mRNA secondary structure *in silico* (Table 5.2). The aim of this study was to chemically probe the 3' UTR region with SHAPE reagent using DMS/SHAPE LMPCR methodology to compare *in silico* predictions to probe based resolved structure. Improved sensitivity was achieved by using a DNA adaptor that was ligated to the 3' end of complementary DNA, followed by PCR amplification using *CYP24A1* specific primers (Table 2.5). For DMS/SHAPE LMPCR structural determination of the *CYP24A1* 3' UTR, 5 ug of total CMH patient RNA (Patient 2, chapter 5) was isolated from whole blood. RNA was treated with either NMIA (SHAPE reagent) as modified samples or DMSO negative controls, as previously described[280]. As the SNVs in

CYP24A1 of patient 2 were identified in the 3' UTR, the structure determination solely focused on the resolution of this region of the *CYP24A1* gene.

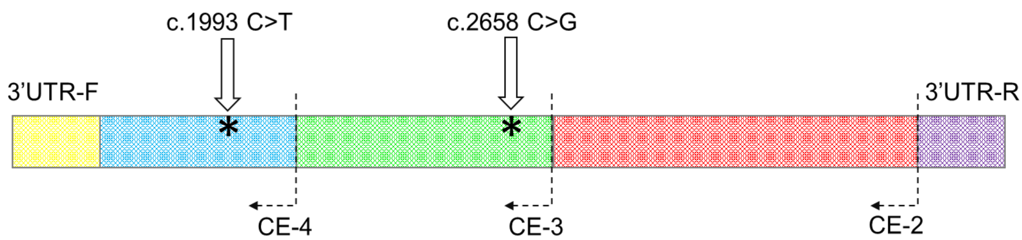
For secondary structure determination, the 3' UTR region (1,316 nt) of *CYP24A1* was divided into 3 smaller fragments for individual amplification and chemical probing. The three fragments of the 3' UTR are referred to as CE-R2, CE-R3 and CE-R4 throughout this study. The previously identified SNV 3' UTR locations in patient 2 were present within different fragment regions of the 3' UTR (Figure 6.1 A, B).

5' -

ATGGTGGTATTTGCTAACATCATATCCAACTTAGGGAAGCGGACTGAGTGCTGGGATCCAAGGC
ATTCTACAGGGTTCACTGCTGGTTTACACTTCACCTGTGTCAGCACCATCTTCAGGTGCTTAGAAT
GGCCTGGGAGCCTGTTCTGTCTTGCATCTTCCATGACATGAAAGGGAGGCTGGCACTTGTCAGTC
AGGTAGAGGTTACAAACCGTTTCAGGCCCTGCCTACCACATTCACTGTTTGAATCTTTAATTCCCA
AGAATAAGTTTACATTTCACAATGAATGACCTACAACAGCTAAATTTTCTGGGGCTGGGAGTAATA
CTGACAATCCATTTACTGTAGCTCTGCTTAATGTACTACTTAGGAAAATGTCCCTGCTTAATAATGT
AAGCCAAGCTAAATGATGGTTAAAGTTATCAGGC**CTCCCATGAAATTGCGTTCTTCCTGCATTGA**
AATAAAAACATTATTGGGAAACTAGAGAACACCTCTATTTTTAAAAGGACTTTAACGAAGTCAAACA
ACTTATAAGACTAGTGATTCACTGGGGCATTATTTTGTTAGAGGACCTTAAAATTGTTTATTTTTTAA
ATGTGATTCCTTTATGGCATTAGGGTAAAGATGAAGCAATAATTTTTAAATTGTGTATGTGCATATG
AAGCACAGACATGCATGTGTGTGTGTGTCTGTGTGTGTGTGTGCGTGTATGTGTGTGTGGGTTCT
AATGGTAATTTGCCTCAGTCATTTTTTTAATAT**TTGCAGTACTTGATTTAGGATCTGTGGTGC**AGG
GCAATGTTTCAAAGTTTAGTCACAGCTTAAAAACATTCAGTGTGACTTTAATATTATAAAATGATTTC
CCATGCCATAATTTTTCTGTCTATTAAATGGGACAAGTGTAAAGCATGCAAAAGTTAGAGATCTGTT
ATATAACATTTGTTTTGTGATTTGAACTCCTAGGAAAAATATGATTTCATAAATGTAAAATGCACAG
AAATGCATGCAATACTTATAAGACTTAAAAA**TTGTGTTTACAGATGGTTTATTTGTGCATAT**TTTTA
CTACTGCTTTTCCTAAATGCATACTGTATATAATTCTGTGTATTTGATAAATATTTCTTCCTACATTA
TATTTTTAGAATATTTCAGAAATATACATTTATGTCTTTATATTGTAATAAATATGTACATATCTAGGT
ATATGCTTTCTCTCTGCTGTGAAATTATTTTTAGAATTATAAATTCACGTCTTGTCAGATTTCATCTG
TATACCTTCAAATT**CTCTGAAAGTAAAAATAAAAGTTTTAAATATTTGA**

- 3'

A)



B)

Figure 6.1: *CYP24A1* 3' UTR sequence plus CE primer location and position of SNVs (Patient 2). **A)** Primers specific to 3' UTR 5' to 3' forward sequence (yellow), 3' to 5' reverse sequence (purple), primer CE-R2 1056 nt (red), primer CE-3 789 nt (green) and primer CE-4 459 nt (blue). **B)** The three primers (CE-2 (red), CE-3 (green) and CE-4 (blue)) cover the region containing the 3' UTR variants identified in Patient 2 (asterisks). Patient 2 SNV locations are present in fragment CE-R3 (c. 2658 C>G) and fragment CE-R4 (c. 1993 C>T).

For efficient nucleic acid probing, a ‘two capillary’ approach for resolving primer extension reactions in a single experiment is required[280]. One capillary contained the NMIA treated reaction plus the sequencing lane. The second capillary contained the corresponding DMSO negative control reaction for the same RNA sample, plus the same sequencing lane. The sequencing lane provides alignment to the correct RNA sequence.

PCR product for each 3’ UTR section CE-R2, CE-R3 and CE-R4 were run on agarose gel (n=10) to determine the correct respective fragment amplification (Figure 6.2 A, B, C).

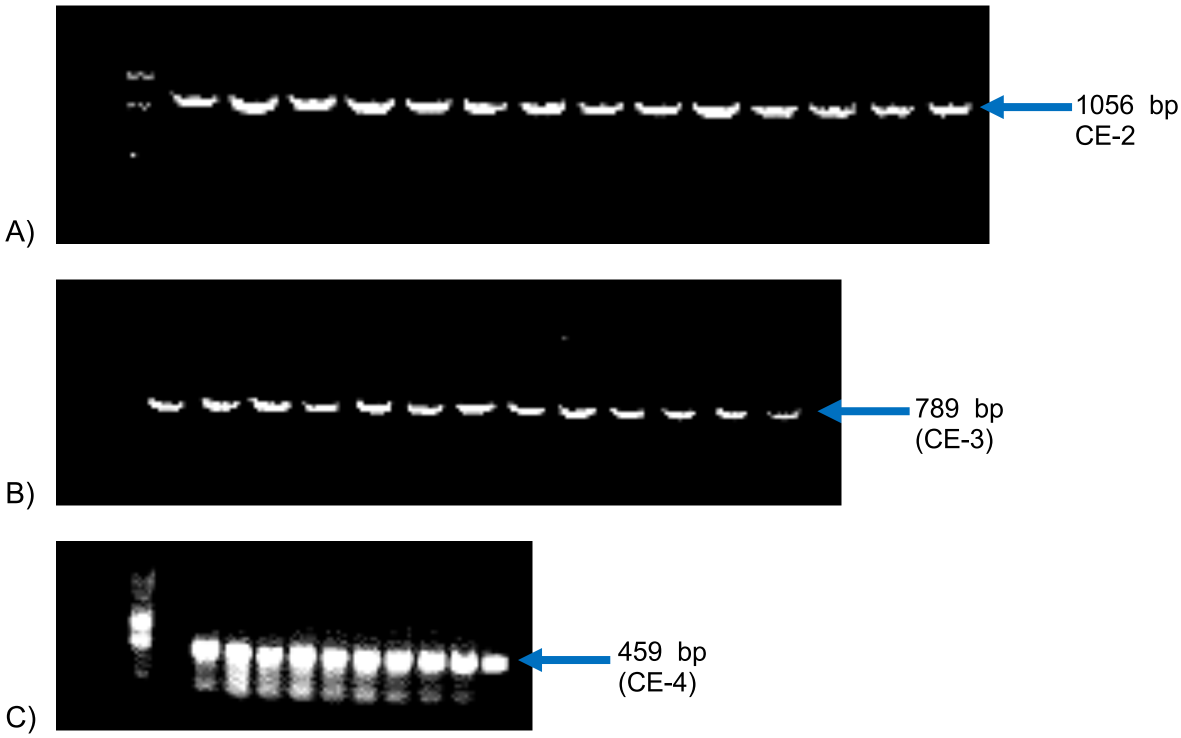


Figure 6.2: Agarose gel separation of each *CYP24A1* 3' UTR fragment A) Primer CE-2 n=14 individual amplification reaction replicates (1056 bp), **B)** Primer CE-3 n=13 individual amplification reaction replicates (789 bp) and **C)** Primer CE-4 n=10 individual amplification reaction replicates (459 bp). Each fragment of the 3' UTR (CE-2, CE-3 and CE-4) was separated on agarose gel in a total of 10-14 lanes.

Once the sequencing lane was prepared, 5 µg of patient 2 and control patient's NMIA treated and corresponding DMSO negative control RNA were reverse transcribed with non-labelled CE-2, CE-3 and CE-4 primers as before. Samples were labelled using corresponding VIC fluorescently tagged CE-2, CE-3 and CE-4 primers in a further PCR reaction in preparation for CE analysis (Figure 6.3). Patient 2 samples, control patient samples, sequencing lane plus 2 x Rox1 1 kb ladder were loaded onto 96 well plates and analysed by CE.

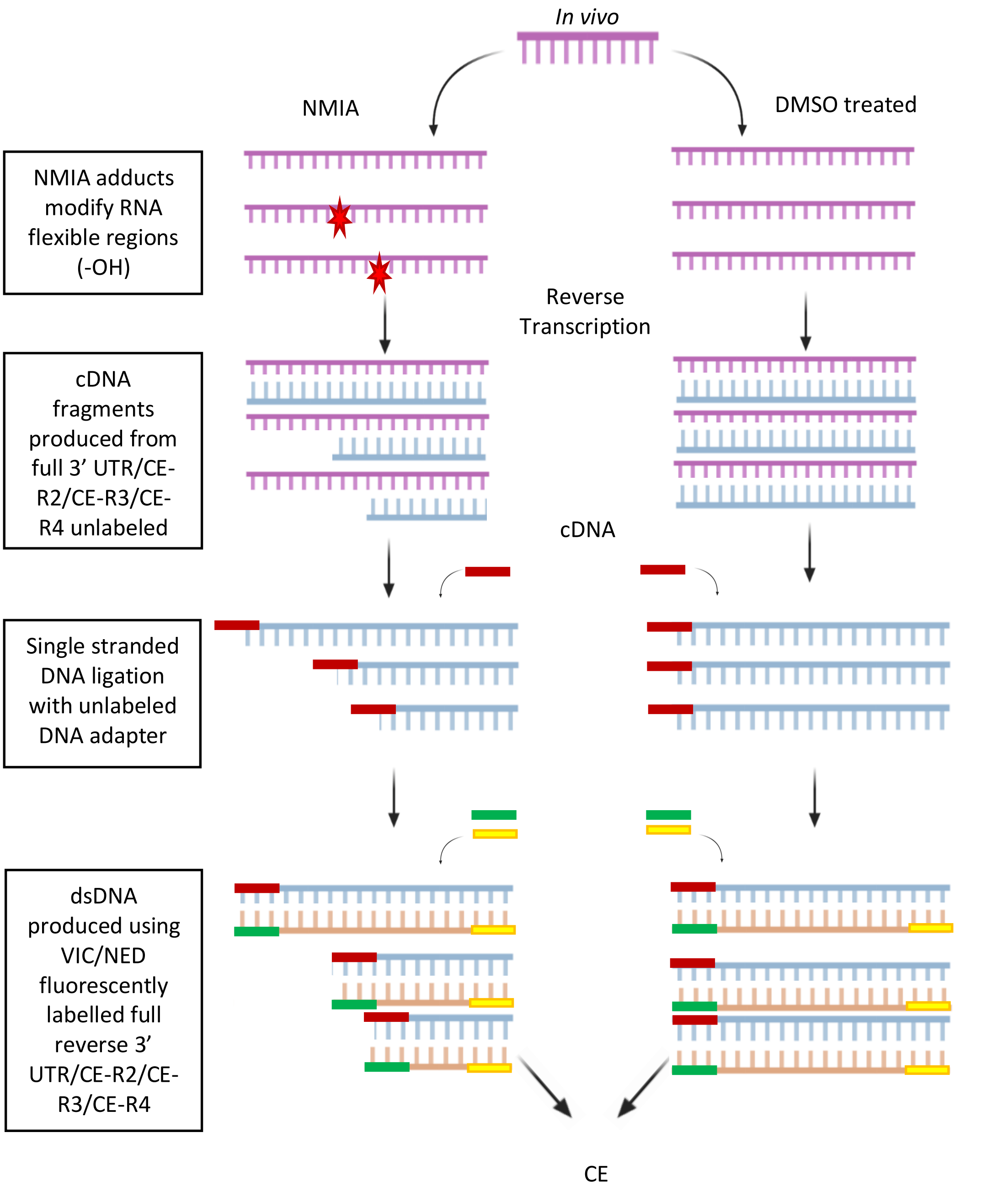


Figure 6.3: DMS/SHAPE-LMPCR workflow. Total RNA (pink) is treated with SHAPE reagent (NMIA) or DMSO as a negative control. The SHAPE reagent NMIA modifies flexible regions along the entire RNA transcript by introducing an -OH adduct at unpaired nucleotide regions (red star). Reverse transcription using unlabelled full length 3' UTR, CE-R2, CE-R3 or CE-R4 primers produced cDNA fragments (blue). Reverse transcription is halted at adduct locations in NMIA treated samples resulting in cDNA fragments of varying length. The unlabelled cDNA is ligated by single stranded DNA ligation with a DNA adapter (red). Once ligated, fluorescently labelled primers (VIC labelled patient samples (yellow)/NED labelled sequencing lane (green)) are used for 3' UTR full length, CE-R2, CE-R3 or CE-

R4 PCR amplification of ligated cDNA fragments. Fluorescently labelled primers identifying the length of each fragment indicating where transcription was halted are then detected by CE analysis[280].

6.3.2 SHAPE NMIA probing of *CYP24A1 ex vivo*

QuSHAPE software was used to analyse CE data. For each sample, NMIA treated RNA, DMSO control RNA and ladder sequencing data was aligned by QuSHAPE[183]. The online tool performs base calling, which classifies all of the peaks as either specific (produced by ddNTP-paired nucleotides) or nonspecific (background peaks corresponding to nucleotides of the other three bases). The sequencing lane peaks are then aligned to the RNA sequence of the 3'UTR. QuSHAPE matches each nucleotide peak of the sequencing lane to the corresponding NMIA treated and DMSO control RNA traces. Traces are aligned and linked by vertical arrows connecting base-calling results between the ladder, sequencing lane and NMIA/DMSO samples (Figure 6.4). Successful sequence alignment is represented by vertical arrows indicating matching correspondent NMIA treated RNA, DMSO control RNA, sequencing lane and ladder peaks. The sequencing data that was generated by CE for CMH patient 2 in this study indicated insufficient detection of the *CYP24A1* 3' UTR region, shown by poor peak identification and subsequent QuSHAPE alignment (Figure 6.4). The QuSHAPE software provides example data as a guide for users. The clear peaks produced in the traces generated from CE data in the example provided, highlights the lack of sufficient signal produced in this chapter of the *CYP24A1* 3'UTR (Figure 6.5)[183].

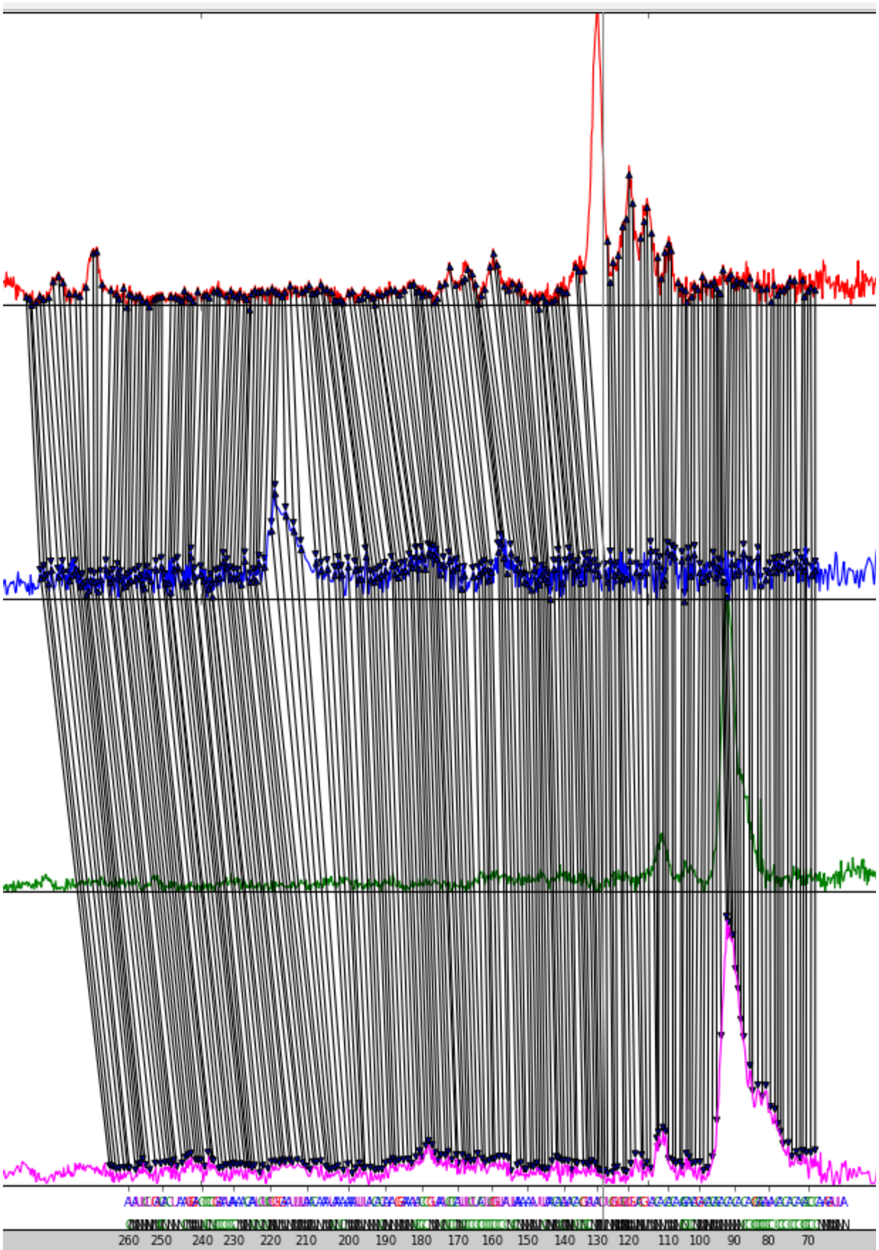


Figure 6.4: QuSHAPE alignment for *CYP24A1* (Patient 2, chapter 5). Four electropherogram sequence traces for NMIA treated sample (red), DMSO control sample (blue), sequencing lane (green) and ladder (pink). Peaks classified as specific are labelled G (green) at the bottom of the window, while peaks classified as nonspecific are labelled N (black). The optimally aligned mRNA nucleotide sequence is also displayed beneath the ladder trace at the bottom of the window. Poor detection shown by unclear peaks in each trace, resulted in inadequate alignment of the sequencing data and determination of specific/non-specific peaks.

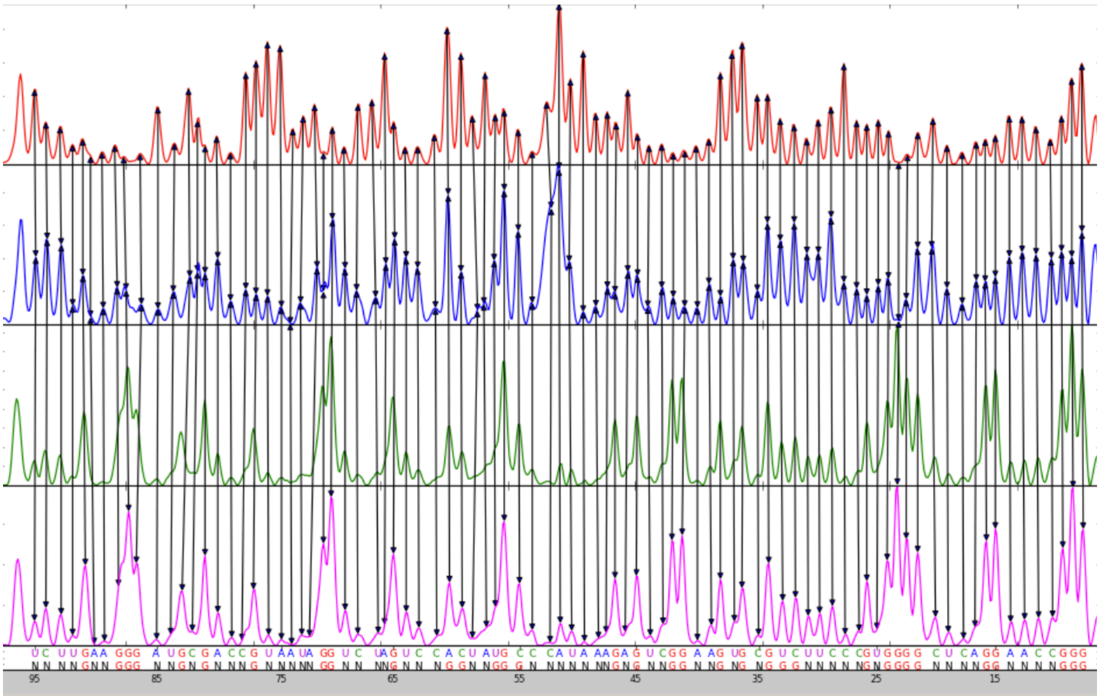


Figure 6.5: QuSHAPE alignment for 'Test Data' available from the QuSHAPE software[183]. Four electropherogram sequence traces for NMIA treated sample (red), DMSO control sample (blue), sequencing lane (green) and ladder (pink). Peaks classified as specific (produced by ddNTP-paired nucleotides) are labelled G (red) at the bottom of the window, while peaks classified as nonspecific (background peaks corresponding to nucleotides of the other three bases) are labelled N (black). The optimally aligned RNA nucleotide sequence is also displayed beneath the ladder trace at the bottom of the window. This example data demonstrates how optimal CE data is processed by QuSHAPE, producing clear signal peaks in each trace that can be aligned between treated, non-treated and sequencing lane samples.

Initial QuSHAPE analysis suggested poor alignment requiring optimisation of the sample preparation (Figure 5.4). Gaussian integration was performed for all peaks in the NMIA and DMSO control channels to confirm our findings. Nucleotide reactivity of >0.85 is suggestive of flexible regions that are highly reactive with the NMIA reagent, 0 is suggestive of highly constrained regions (Figure 6.6). Although SHAPE reactivity could be estimated for each nucleotide, no negative SHAPE reactivities should be produced (Figure 6.6). The negative values produced in this figure support our lack of confidence in the QuSHAPE alignment produced in Figure 6.4.

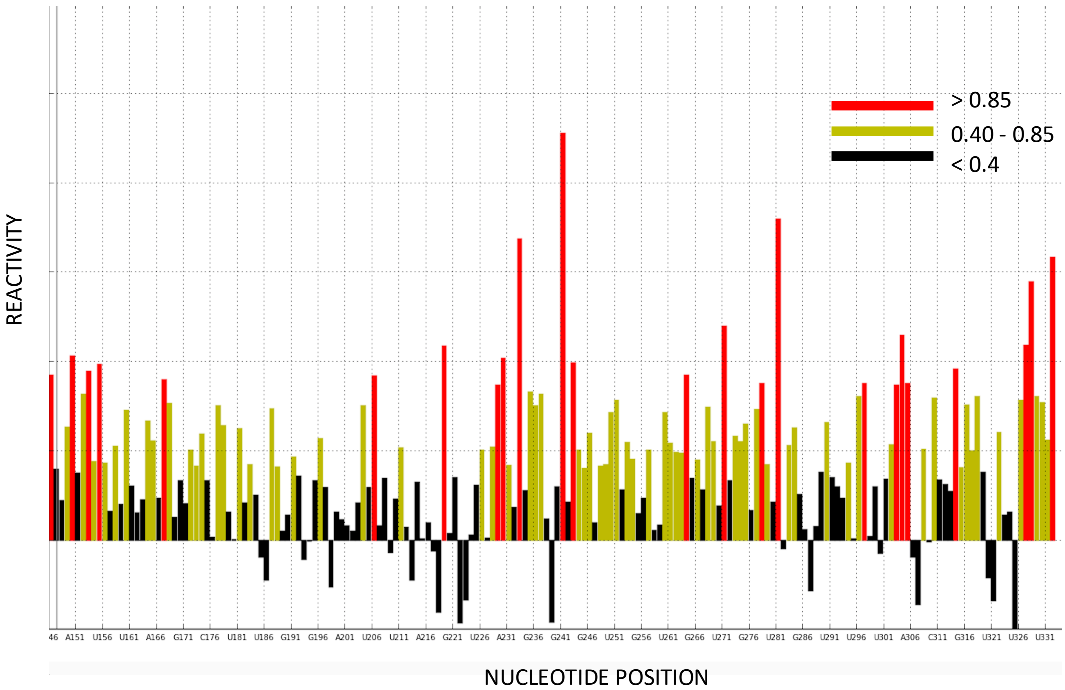


Figure 6.6: Reactivity of CMH patient 2 generated from QuSHAPE analysis. Each nucleotide position flexibility corresponds to individual SHAPE reactivity of 0.4-0.85 (red), >0.85 (green) or <0.4 (black). Nucleotide positions with 0 reactivity signify highly structured positions, therefore it is important to differentiate from positions that produced no data. If the SHAPE reaction and QuSHAPE analysis was optimal and reagent/background traces are successfully scales, no negative SHAPE reactivities should be produced in this figure. The negative values produced in this figure support our lack of confidence in the QuSHAPE alignment produced in Figure 6.4.

RNAthor software is used to visualise the 2D RNA structure from the DMS/SHAPE LMPCR CE reactivity data. The structure produced could then be compared to control patient samples to determine if mutations identified in the *CYP24A1* mutated patient caused an mRNA structural change, as predicted *in silico*. The QuSHAPE software and reactivity calculations in this experiment showed poor alignment meaning if a secondary structure was produced by RNAthor, it would likely be a false positive structure and not a true representation of the *CYP24A1* mRNA secondary structure. This data indicates that further method optimisation is required, therefore no secondary structure is presented for CMH patient 2 in this thesis.

6.4 DISCUSSION

Previous work in this thesis suggests that varying mutations in the 3’ UTR region can alter the mRNA secondary structure of *CYP24A1* (*in silico)* and lead to heterogeneous clinical presentations in CMH patients. The study in this chapter sought to demonstrate that probe-based structure determination would similarly show structural alterations to those observed in *in silico* predictions. The aim of this work was to determine the structure of the 1,316 nt section of the *CYP24A1* 3’ UTR region in the CMH cohort (Patient 2, chapter five). Changes in the 3’ UTR confirmed by SHAPE chemical modification would further support the cause of the hypervitaminosis D phenotype displayed.

6.4.1 DMS/SHAPE LMPCR OPTIMISATION REQUIRED FOR *EX VIVO* *CYP24A1* STRUCTURE DETERMINATION

Due to the relatively low abundance of *CYP24A1* in human whole blood, traditional SHAPE-RT probe-based structure determination technology would be hindered by low detection of the RNA transcript patient samples. DMS/SHAPE LMPCR has been shown to enhance the ability to accurately visualise the structure of RNA unlike the reduced sensitivity of SHAPE-RT alone. While DMS/SHAPE LMPCR can increase detectability of low abundance transcripts, the method described in this thesis did not effectively resolve the *CYP24A1* structure in this study. Low level signal was achieved for *CYP24A1* 3’ UTR in this experiment however the result was not deemed to have produced a clear signal to accurately visualise the mRNA structure using RNAthor software. Due to the poor signal achieved and subsequent inability of QuSHAPE software to align the CE sequencing data, correct structure determination for *CYP24A1* could not be visualised.

Integration of chemical probing methods like DMS/SHAPE LMPCR profiling alongside mutagenesis and phenotype analysis can allow better understanding of the role that RNA secondary structure holds on biological function. This research has identified that *CYP24A1* transcript abundance remains too low in RNA samples extracted from patient whole blood. As previously discussed in chapter 5, a homogenous sample of whole blood was used for this analysis. Future work may benefit from isolating different individual cell types before SHAPE analysis is performed in order to effectively compare between patient samples. Improved sample preparation to increase *CYP24A1* transcript in the starting material could resolve this in future study. To determine the *in vivo or ex vivo CYP24A1* structure in future work research should investigate using samples that express a greater abundance of *CYP24A1*. Previously reported work indicates that *CYP24A1* is expressed in all target cells containing the VDR including the distal convoluted tubule in the kidney, bone cells and intestinal tissues[309]. Subsequent studies found that *CYP24A1* expression is greatest in the endometrium, and urinary bladder[310]. While *CYP24A1* expression is greater in these tissues than the whole blood used in this chapter, human samples of an invasive source, including biopsies, were beyond the ethics approval of this thesis and would need to be explored in further research e.g., kidney biopsies. In order to improve the CE signal generated in this study and allow visualisation of the mRNA secondary structure alternative methods to stimulate *CYP24A1* expression prior to SHAPE, e.g., vitamin D metabolite or 1,25$(OH)_2$D stimulation in cell culture, may elevate *CYP24A1* abundance to an expression suitable for SHAPE analysis. While *CYP24A1* stimulation could be possible in cell lines for the structural determination of wildtype *CYP24A1*, there are no commercially available cell lines containing *CYP24A1* 3'UTR mutations mimicking conditions such as CMH seen in our patient cohort. Development of such cell lines could allow initial in vitro *CYP24A1* investigations prior to in vivo structure determination.

6.4.2 CONCLUSION

Once optimised for *CYP24A1*, further research utilising DMS/SHAPE LMPCR methodology could be extended to determine the structure of the entire *CYP24A1* gene beyond the 3' UTR region focused on in this study. This work provides a basis for *CYP24A1* probe-based structural determination, that may be built upon by further method development and sample optimisation to achieve increased sequencing detection. Additionally, novel gene mechanisms could be explored in Humans using sensitive protocols such as DMS/SHAPE LMPCR profiling. This provides the opportunity to determine how mutations in the mRNA affect the secondary structure and function, potentially shedding light on the phenotypic variability in conditions such as CMH and adult-onset nephrolithiasis.

CHAPTER 7: INTRODUCTION OF SNVs INTO HUMAN EMBRYONIC KIDNEY CELLS USING CRISPR-CAS 9

7.1 INTRODUCTION

Many research questions surrounding CMH diagnosis and mechanism of action remain unanswered. Precise understanding of prevalence, genotype-phenotype correlation and therapy require further investigation. Due to the range of reports with varying inheritance patterns, the inheritance of disease in heterozygous patients remains inconclusive[125,127,142,147]. Several reports have presented case studies exhibiting phenotype of hypercalcaemic biochemical findings while others present normal serum calcium[125]. Previous findings have indicated the potential for as yet unknown alternate metabolism pathways present in the absence of functional *CYP24A1* to prevent elevated 1,25$(OH)_2$D and subsequent hypercalcemia[120], or that CYP24A1 remains partially functional to prevent elevated 1,25$(OH)_2$D and subsequent hypercalcemia. It is unknown whether CMH patients with non-functional or partially functional CYP24A1 have impaired fracture healing from an inability or reduced ability to metabolise 25OHD into 24,25$(OH)_2$D[75]. Due to the rarity of cases and with many unanswered questions relating to *CYP24A1* and CMH, a robust disease model would support investigations in this field.

By genetically engineering biological systems to model different disease states, huge potential is opened for medicine and biotechnology. Genome editing is key to understanding gene function analysis in human health. Mouse genetic engineering was revolutionised in the 1980s when scientists first selectively knocked out genes of interest to allude to their role in development and physiology[311]. While the generation of the first mouse models revolutionised gene targeting technology, the process was slow and expensive with minimal genetic alterations taking 2-3 years and costing hundreds of thousands of dollars[312]. Alternative technology has since been developed that provides a faster and cost-effective method for genetic engineering. Previous non-specific nucleases fused to sequence-specific DNA binding domains for genome editing includes zinc-finger nuclease (ZFN) editing and transcription activator-like effector nucleases (TALEN) to introduce double stranded breaks (DSBs) into DNA[313]. DSBs can disrupt gene function through non-homologous end joining (NHEJ) forming premature stop codons, or homologous directed repair (HDR) that repairs target gene regions with DNA templates provided[313]. While TALEN and ZFN provide modification technology that can be applied to a wider range of models e.g., zebrafish, fruit flies, livestock etc, the technology remained expensive and there was difficulty in replacing large fragments of DNA sequence, and it remained difficult to construct zinc finger domains that bind to selective target nucleotide stretches with high affinity[314].

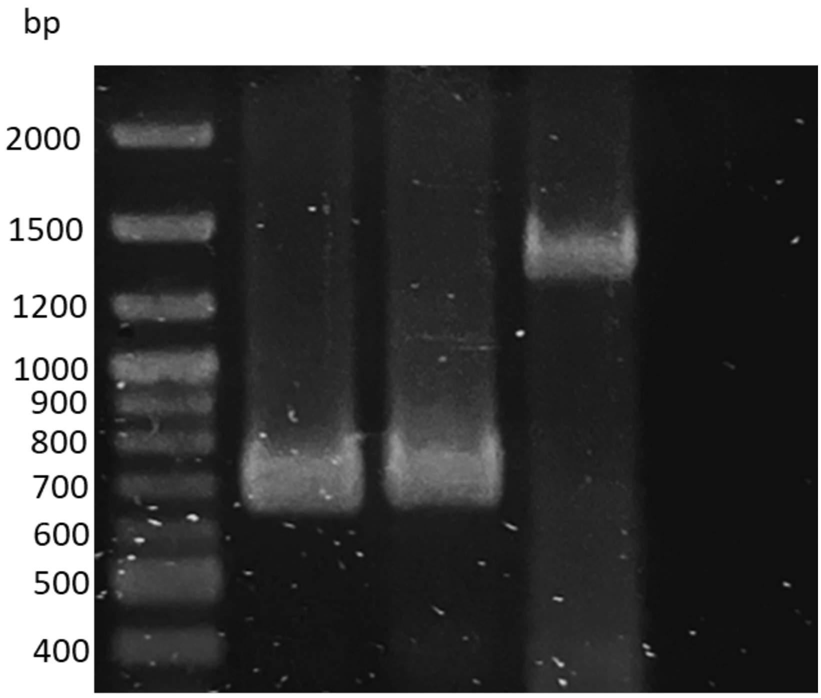
Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas 9) is a widely used RNA-based gene editing technology that utilises bacterial defence mechanisms, which identify and edit foreign DNA[313, 314]. In contrast to TALEN and ZFN, Cas9 is nuclease guided by small RNAs that target DNA with high specificity and efficiency, while being easy to design and suitable for high throughput multiplexed gene editing in a range of organism and cell types[313, 314]. Cas9 technology introduces DSB at target loci followed by DNA repair by NHEJ in the absence of DNA template or HDR. NHEJ is often error prone and results in frequent insertion or deletion mutations (indels) that can induce frameshift that disrupt target genes[313, 314]. In contrast, HDR uses a homologous DNA template, which is complementary to the genomic DNA target. HDR is more precise than NHEJ due to the exogenous DNA donor[313]. Through HDR, precise mutations can be introduced into cells by designing homologous DNA templates containing desired alteration e.g., SNVs. The template DNA containing the desired alteration can be used to introduce a new sequence or correct an existing mutation via HDR repair[313]. The development of CRISPR Cas 9 plus NHEJ/HDR has increased the suitability of gene editing in clinical research through providing technology to alter a wide variety of cells to mimic or correct genetic disease[315]. Recent *in vitro* clinical research using genome editing has included cardiovascular disease, metabolic disease, immune system defects and viral infections[316–324]. Through utilising CRISPR Cas 9 technology, the future for developing cell models of disease to aid understanding of progression and pathogenesis is expanding.

A major limiting factor in *CYP24A1* research in this thesis, plus the wider research community, is the rarity of patients that harbour, and are identified with, *CYP24A1* hypomorphic mutations. The development of a robust models to aid further *in vitro* investigations into the mechanism of disease in CMH patients would provide availability to better our understanding of the condition. This study sought to harness CRISPR-Cas9 technology to introduce a SNV into the 3' UTR of *CYP24A1* of HEK293T cells to mimic the identified loss-of-function variants in our patient cohort.

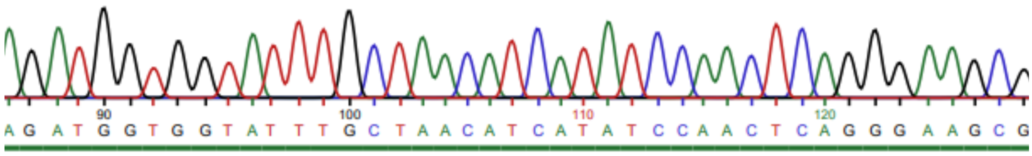
7.2 RESULTS

7.2.1 HEK293T CELL LINE SUITABILITY

HEK293T cells are widely used for CRISPR-Cas9 modification due to their ease of transfection, rapid growth rate, robustness and expression of *CYP24A1*. To first establish the suitability of HEK293T cells as a basis for our CMH model, the 3' UTR sequence of *CYP24A1* in HEK293T cells was analysed to ensure no somatic mutations had occurred over time. It is common for somatic mutations to occur over years of multiplying commonly used cell lines, including HEK293T. To investigate the sequence of HEK293T cells, Sanger sequencing was performed as previously described (Figure 7.1 A and B).



A)



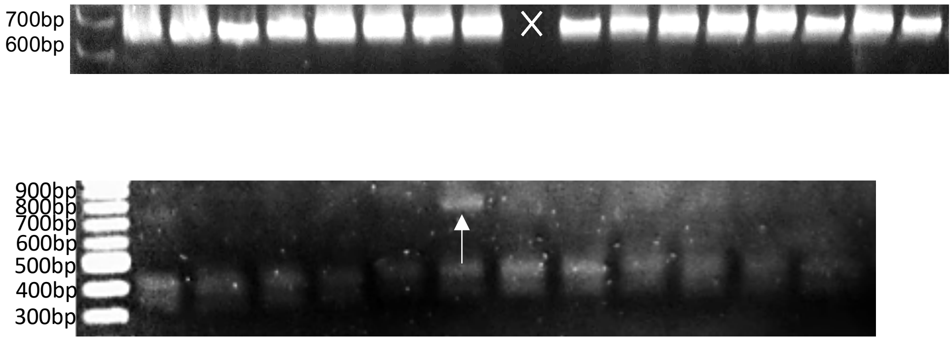
B)

Figure 7.1: HEK293T Sanger sequencing results. **A)** Gel electrophoresis of *CYP24A1* 3’ UTR fragments 700bp, 700bp, 1500bp. **B)** Human Genome sequence of *CYP24A1* 3’ UTR[325] (top line) compared to the 3’ UTR sequence identified by Sanger sequencing in HEK293T cells (bottom line). **C)** The sequence chromatogram of HEK293T cells indicates that direct sequencing of the 3’ UTR of *CYP24A1.* Successful sequencing revealed no alterations in comparison to the wildtype sequence (Human Genome Project).

7.2.2 CRISPR CAS 9 TRANSFECTION

Once the sequence of HEK293T cells was confirmed, CRISPR Cas 9 was used to introduce a 3' UTR SNV into *CYP24A1* to mimic patients with CMH. CRISPR Cas 9 was designed and purchased from Origene (Rockville, USA) to introduce SNVs in two separate locations of the *CYP24A1* 3'UTR simultaneously (Table 2.7). The location of transfection targets mimicked Patient 2 of our cohort in chapter 5 of this thesis, c. 1993 C>T and c. 2658 C>G (Table 5.2).

Transfection was performed over 48 hours before splitting and performing an array serial dilution to separate single colonies. Successfully isolated single colonies (n = 44) were expanded into individual cell lines. Restriction Fragment Length Polymorphism (RFLP) was used to assess the success of CRISPR Cas 9 modification in each colony and identify which contained the desired c.1993 C>T variant. The SNV c.1993 C>T removes a restriction enzyme cut site (TspGWI enzyme). By extracting the DNA from each colony and PCR amplifying the region surrounding the digest site, then incubating with TspGWI, the success of SNV introduction in each colony can be determined (Figure 7.2 A and B). Post enzyme digestion, if the PCR amplified region (~700 bp) surrounding the SNV and enzyme cut site is halved (~350 bp) this would indicate that the cut site was not removed and CRISPR Cas 9 modification has been unsuccessful.



A)

B)

Figure 7.2: RFLP assessment of CRISPR modified HEK293T cell colonies. **A)** DNA fragment of CYP24A1 3' UTR amplified from wildtype HEK293T cells prior to CRISPR Cas 9 transfection by PCR (700bp). CYP24A1 was successfully amplified (700bp) in all samples bar one which failed PCR amplification'('X'). **B)** Post TspGW1 enzyme digestion showing that all CRIPSR Cas9 transfected HEK293T cell colonies amplified by PCR presented with bands at 350 bp consistent with enzyme digestion, apart from colony one (white arrow). The single colony that was not digested by TspGW1 was partially resistant to the enzyme activity with a DNA fragment remaining at 700 bp (white arrow) indicating that the desired 3'UTR mutations have been successfully transfected into this colony.

One colony was identified by RFLP as likely containing the desired SNV c.1993 C>T generated using annealed oligonucleotides. The CRISPR Cas 9 percentage efficiency of the first transfections was 2.3%. This cell colony was expanded and used for the second CRISPR Cas 9 transfection using gRNA and annealed oligonucleotides for SNV c.2653 C>G (Table 2.7). The transfection was repeated as to introduce the second SNV within the 3' UTR of *CYP24A1* into the proposed mutated cell line.

To confirm the presence of both mutations, Sanger sequencing was performed on the newly isolated single cell colonies (n=36) to confirm the exact sequence of the newly established cell lines. Of the total cell lines produced, one cell line was confirmed to have contained a deletion within the 3' UTR of *CYP24A1* (Figure 7.3 A and B). The CRISPR Cas 9 percentage efficiency of the second transfections was 2.8%. The sequencing was repeated a second time to confirm the findings.

A)

B)

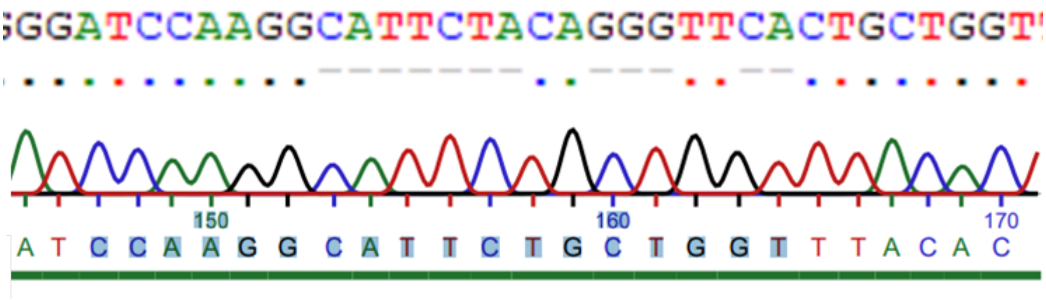
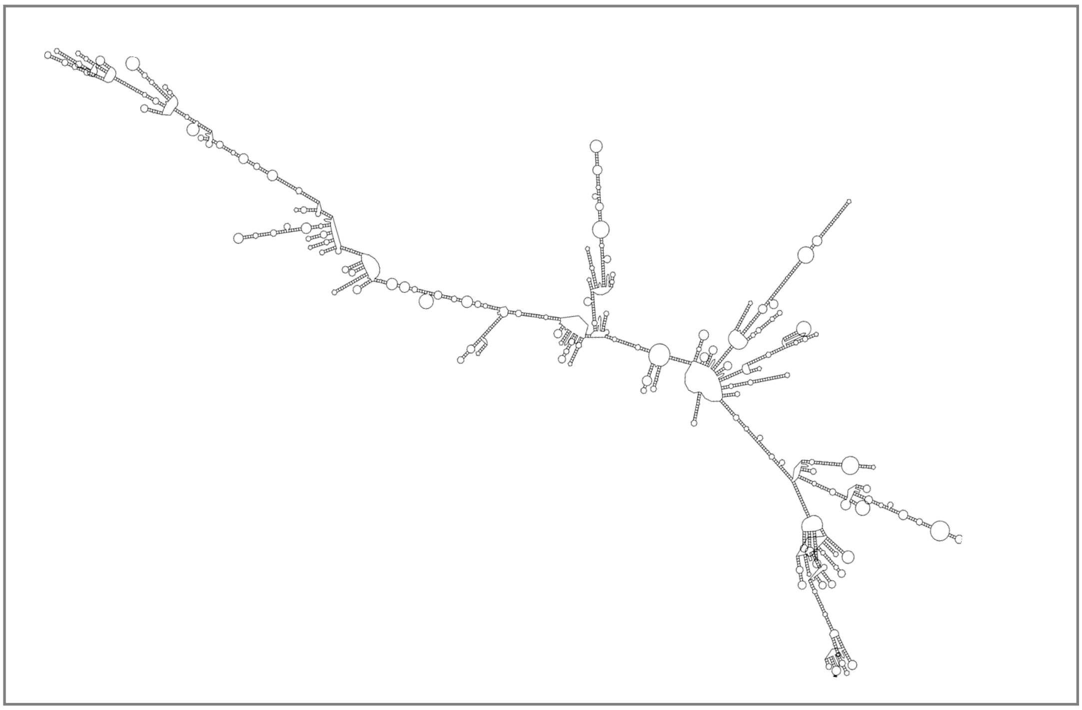


Figure 7.3: *CYP24A1* 3' UTR sequencing revealing variants at positions c.2026_2032del, c.2035_2037del and 2040-2041 del, within the CRISPR modified cell line. **A)** Sequence of wild type *CYP24A1* (top row) compared to CRISPR modified cell line sequence (bottom dot and dash plot), with dashes indicating inconsistency with the wild type sequence**. B)** Chromatogram of CRISPR-Cas9 modified sequence at the region of genome modification.

7.2.3 INVESTIGATIONS INTO THE CRISPR-CAS9 MODIFIED CELL LINE

The effect of the 3' UTR modification was assessed to understand the effect on *CYP24A1* mRNA structure, function and vitamin D metabolism. Using RNAfold, as previously described, the predicted *in silico* structure was determined and compared to the wild type *CYP24A1* sequence (Figure 7.4 A and B). The mountain plot for the CRISPR modified sequence in comparison to the wild type was also determined using RNAfold software (Figure 7.5 A and B).

A)



B)

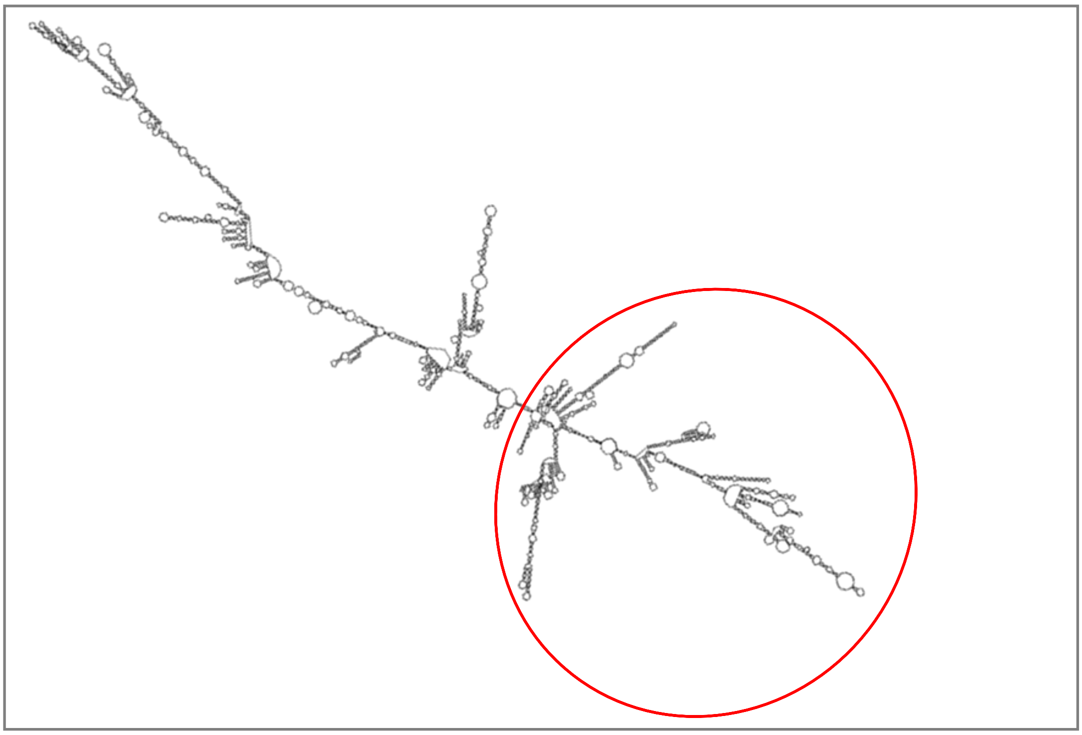
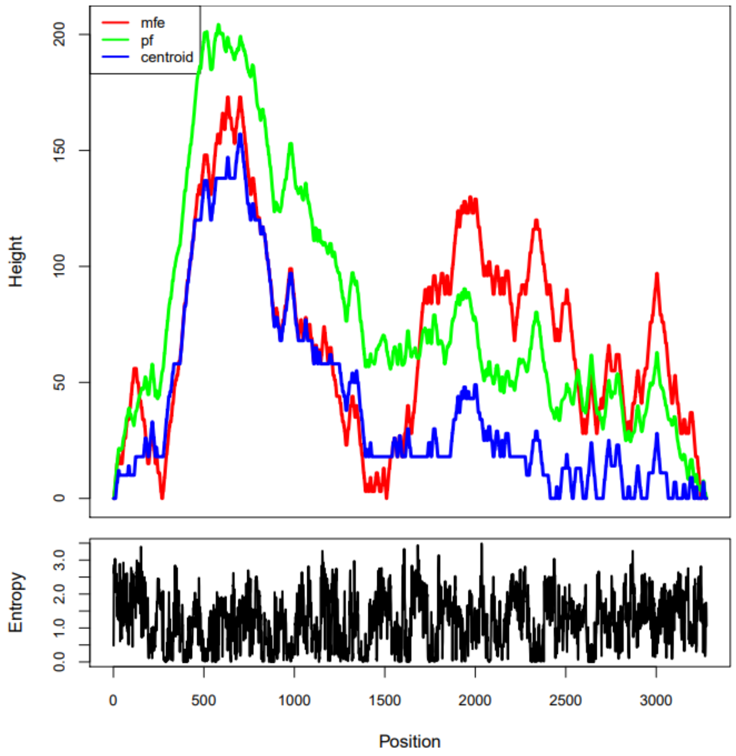
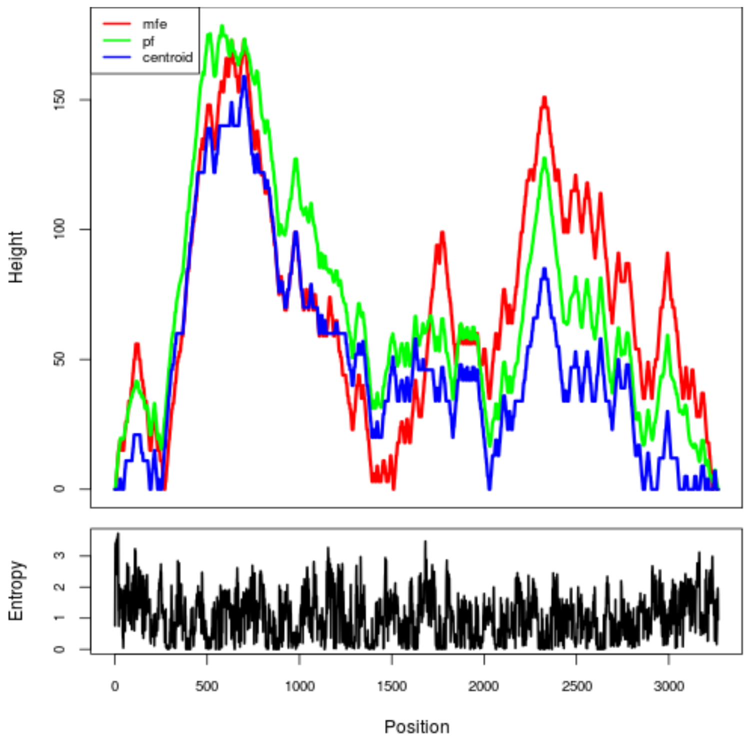


Figure 7.4: MFE based RNA structures of wild type and CRISPR-Cas9 modified HEK293T cells per genotype. A) Wild type *CYP24A1* transcript. **B)** CRISPR Cas 9 Modified HEK293T *CYP24A1* (c.2026_2032del, c.2035_2037del and 2040-2041 del).



A)



B)

Figure 7.5: Mountain plot representation of the MFE based RNA structures from Figure 7.4. RNA structure thermodynamic ensemble plus the centroid structure. The minimum free energy (mfe) (red), thermodynamic ensemble (green) and centroid structure (blue) is shown for both the wildtype HEK293T and mutant cell line CYP24A1 structure. Mountain plots represent structure in a plot of height versus

position where the height m(K) is given by the number of base pairs enclosing the base at position k i.e., loops correspond to plateaus, hairpin loops are peaks, helices to slopes. Entropy for each position is also presented. CRISPR Cas 9 Modified HEK293T cell genotypes are visibly different when compared to the wild type structure. **a** Wild type *CYP24a1*. **b** CRISPR Cas 9 Modified HEK293T *CYP24A1* (c.2026_2032del, c.2035_2037del and 2040-2041 del).

Due to the structural alteration in the CRISPR Cas 9 modified cell line, the effect on *CYP24A1* function and vitamin D metabolism was compared to wild type HEK293T cells. Each cell line was treated with 10 nM 1,25$(OH)_2$D and 200 nM of 25OHD over a 48-hour period as previously described[326–330]. Cells and media were collected at intervals (1, 4, 12, 24, and 48 hours) over the time course to assess the metabolism of the vitamin D metabolites. LC-MS/MS analysis of 25OHD, 24,25$(OH)_2$D and 1,25$(OH)_2$D plus immunoassay analysis of 1,25$(OH)_2$D was performed, as previously described, on media taken at each timepoints from both cell lines. Comparing the 25OHD to 24,25$(OH)_2$D concentration over 48 hour period indicated a reduced conversion rate in CRISPR modified HEK293T cells in comparison to wild type HEK293T cells (Figure 7.6). The 1,25$(OH)_2$D metabolism was also reduced in CRISPR-Cas9 HEK293T cells (Figure 7.7 A and B). While method comparison between LC-MS/MS and immunoassay analysis of 1,25$(OH)_2$D in HEK293T cells showed similar profiles over the time course, method comparison profiles of CMH cell line differed between LC-MS/MS and immunoassay (Figure 7.7 A and B).

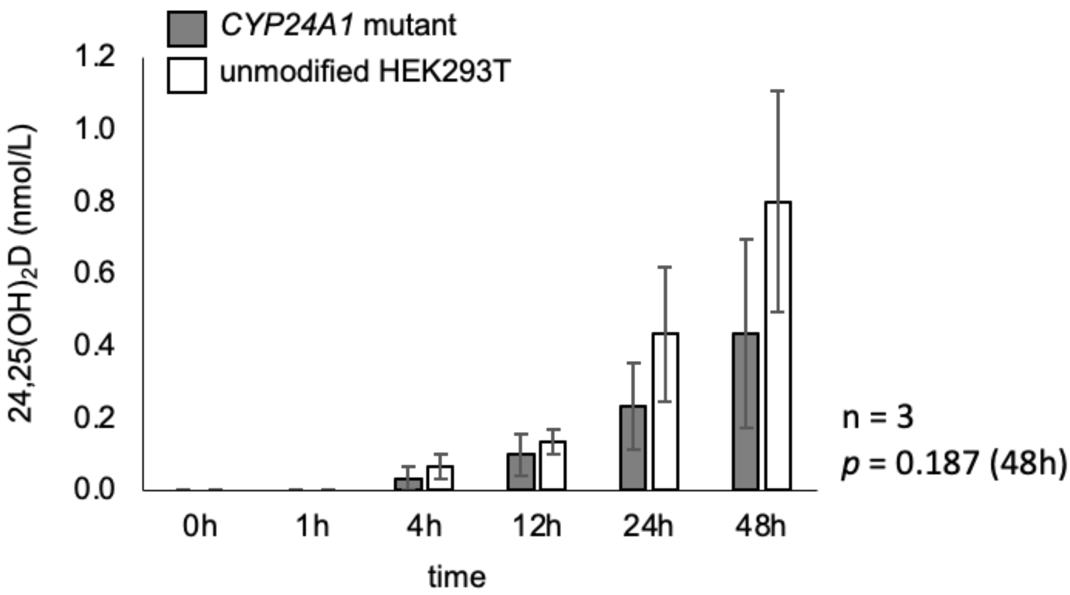


Figure 7.6: HEK293T vs CRISPR Cas 9 model cell line 25OHD metabolism analysis. Each cell line was treated with 200 nM 25OHD over 48h. LC-MS/MS analysis of 24,25$(OH)_2$D in culture medium indicated 25OHD metabolism activity by CYP24A1 in each cell line. Results are displayed as average 24,25$(OH)_2$D concentration at each timepoint measured plus SDs (n=3). Significant difference in 24,25$(OH)_2$D was observed in the mutant cell line after 48 hours (p = 0.187).

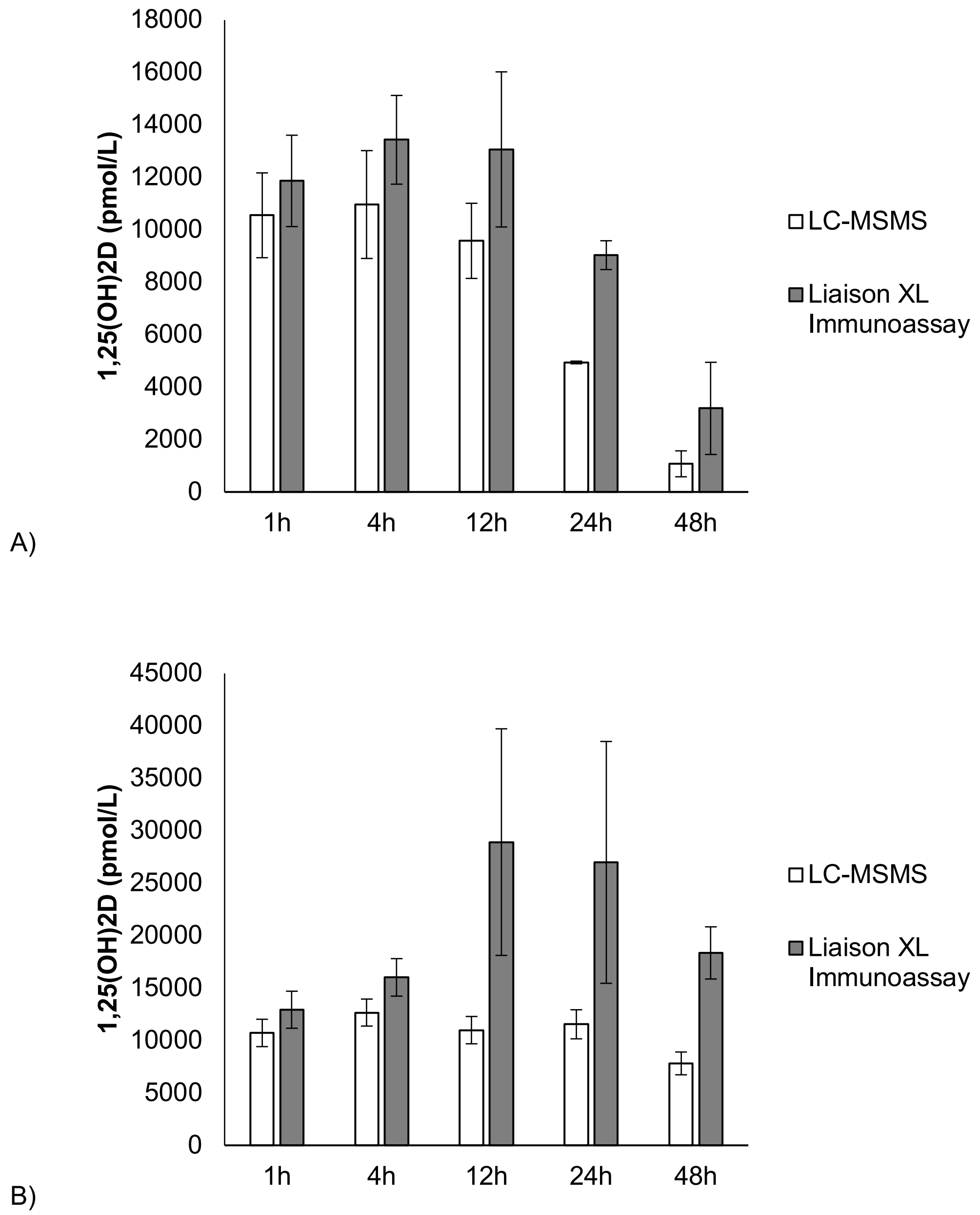


Figure 7.7: Average 1,25(OH)$_2$D metabolism analysis by LC-MS/MS and immunoassay **A)** wildtype HEK293T cells (n=3) and **B)** CRISPR Cas 9 modified HEK293T cells (n=3). Error bars indicate standard error at each timepoint. Significant difference in 1,25(OH)$_2$D catabolism was observed in the mutant cell line after 24 hours (p = 0.04).

7.3 DISCUSSION

This study presents the generation of a new model for *CYP24A1* loss-of-function mutations mimicking conditions such as CMH. Due to the rarity of this condition, the availability of a disease model would enable research in this area. Much is still unknown about the pathogenesis of *CYP24A1* 3' UTR mRNA structure altering SNVs such as those found in the patient cohort in this thesis.

The development of *CYP24A1* knockout mice have broadened our understanding of *CYP24A1* function effects on many related biological processes e.g., bone development[12,131,331]. The *CYP24A1* knockout mice *in vivo* model lacks exon 9 and 10 of *CYP24A1*, which encodes the heme binding domain[332]. The inactivation of *CYP24A1* in mice lead to hypercalcemia secondary to hypervitaminosis D due to an inability to metabolise 1,25$(OH)_2$D[131]. This model correlated with patient studies as gestating knockout mice showed an inability to regulate increased 1,25$(OH)_2$D concentrations, triggered by pregnancy, with their serum concentrations becoming fatal concentrations[131]. This model has also aided subsequent therapeutic understand of conditions such as CMH through Ketoconazole treatment to counteract elevated 1,25$(OH)_2$D in patients with hypercalcemia[333]. The mouse model additionally indicated a potential unknown alternate pathway in regulating vitamin D homeostasis as half the mutant population were unaffected by the *CYP24A1* knockout as 1,25$(OH)_2$D catabolism and calcium homeostasis were unaffected[120].

The knockout model has provided information on the biological activity and importance of metabolites such as 24,25$(OH)_2$D and the effect that *CYP24A1* loss-of-function has on processes such as bone development, fracture repair and progression of diseases associated with hypercalcemia. While the importance of this model is clear and can be useful in understanding more about patients with hypomorphic *CYP24A1* mutations, no models are available containing 3’ UTR *CYP24A1* SNVs leading to reduced or partially functional CYP24A1. The availability of a robust *in vitro* model of CMH could aid primary investigations into the interactions, localisation, transcription and translation of *CYP24A1* in these patients.

7.3.1 CMH CELL LINE MODEL GENERATED USING CRISPR CAS 9

In this work, CRISPR-Cas9 was used to successfully construct a CMH cell line model through introducing deletions into the 3’ UTR of *CYP24A1* (c.2026_2032del, c.2035_2037del and 2040-2041 del) in the HEK293T cell line. HEK293T cells are widely used for CRISPR-Cas9 modification due to their ease of transfection, rapid growth rate, robustness and expression of *CYP24A1*. CYP24A1 catabolism of 1,25$(OH)_3$D primarily occurs within the kidney meaning HEK293T cells would provide a clinically relevant model for CMH investigations. While successful transfection was achieved in this study, future work may benefit from screening *CYP24A1* expression across multiple immortalised cell lines to determine the greatest clinical relevance and *CYP24A1* expression. Previous reports have suggested that CYP24A1 expression is greatest in lung cancer cell lines e.g. HeLa cells[334]. Additionally, while the use of RFLP is a useful tool in highlighting mutations that affect restriction cut sites, direct sequencing was ultimately required to confirm the presence of alterations within the 3’ UTR.

CRISPR-Cas9 technology is prone to target effects, which can be introduced during gene editing experiments[335]. While the CMH model created contains 3’ UTR alterations within the *CYP24A1*, further investigations into the effect that the introduced variants had on *CYP24A1* mRNA structure showed a clear alteration in mRNA secondary folding, similar to what was observed previously in patients with *CYP24A1* 3' UTR SNVs in chapter 5. This abnormal mRNA folding indicates that the desired structural alteration effect had been achieved in this newly generated CMH cell line. Additionally, the development of this CMH cell line was limited by the control cell lines used during development. Typically, a transfection control, e.g. fluorescent tag, is implemented to allow visual confirmation of successful transfection. No fluorescent detection cassette was included to identify the successfully edited cells and confirm the right clones as the introduced mutations were in the 3'UTR. Confirmation was instead performed by single cell isolation and sequencing. While transfection controls allow transfection efficiency to be assessed easier than single cell selection in this chapter, this control does not confirm that the DNA cut was successful. Extensively validated positive editing controls which have been shown to have high editing efficiencies for common target genes e.g. TRAC, RELA, and CDC42BPB when transfection conditions are optimal. Although HEK293T cells were transfected under identical conditions minus the CRISPR Cas 9 vector as a negative control, this could have been replaced with a scramble guide RNAs, which are RNAs that do not have a complementary sequence in the genome to direct a Cas nuclease to cut the genome[336]. Addition of a fluorescent detection cassette and/or scramble guide RNAs in future CRISPR Cas 9 experiments would provide a better indication of successful gene editing.

7.3.2 FUNCTIONALITY OF CYP24A1 IS IMPARED IN THE CMH CELL LINE MODEL

Once the CMH cell line was established, the influence on *CYP24A1* function and vitamin D metabolism was assessed. LC-MS/MS analysis showed that when CMH cells were treated with the vitamin D metabolite 25OHD, there was a decreased rate of conversion to 24,25$(OH)_2$D over 48 hours in comparison to the wild type HEK293T cell line. This decreased metabolism rate corresponds with the phenotype of patients with CMH with a biochemical diagnosis of reduced to undetectable 24,25$(OH)_2$D circulating concentration. Similarly, when CMH cells were treated with the vitamin D metabolite 1,25$(OH)_2$D the concentration of 1,25$(OH)_2$D remained elevated throughout the 48-hour time course in comparison to wild type HEK293T cells, which gradually metabolised 1,25$(OH)_2$D over time.

Method comparison between LC-MS/MS and immunoassay (Liaison XL) measurement of 1,25$(OH)_2$D indicated that the immunoassay produced increased concentrations compared to LC-MS/MS, which is consistent with previous findings[337,338]. This corresponds with the fact that immunoassays can produce falsely elevated 1,25$(OH)_2$D results due to multiple factors e.g., cross reactivity with metabolites such as 1,25$(OH)_2D_2$[339]. The inflated 1,25$(OH)_2$D observed in the immunoassay data in this chapter supports LC-MS/MS as the superior assay for vitamin D metabolite analysis due to its specificity. Whilst both immunoassay and LC-MS/MS 1,25$(OH)_2$D concentration in HEK293T cell lines showed similar profiles with gradually depleting 1,25$(OH)_2$D concentrations, the CMH cell line showed an increase in 1,25$(OH)_2$D production at 12 hours in immunoassay results, which was not observed by LC-MS/MS analysis. This may be the an example of cross reactivity by an unidentified vitamin D metabolite that is falsely elevating the immunoassay detection of 1,25$(OH)_2$D.

Further investigations are required to investigate both the metabolite clearance and CYP24A1 protein expression under differing conditions to conclude the CMH model suitability beyond the initial investigations presented in this study. Firstly, analysis of CYP24A1 activity under physiological concentrations of 25OHD and 1,25$(OH)_2$D, which are lower than what was used to stimulate the cell lines in this chapter (50-120 nmol/L and 55-139 pmol/L respectively) would allow comparison of CYP24A1 activity in the model cell line compared to patients with 3'UTR variants. Additionally, the VMR could be assessed in the CRISPR modified cell line to aid the evaluation of CYP24A1 function. As the LC-MS/MS analysis implemented in this study cannot differentiate 1,24,25$(OH)_2$D, this was not measured. Implementing 1,24,25$(OH)_2$D analysis could indicate the efficiency of CYP24A1 hydroxylation of 1,25$(OH)_2$D or lack of. Previous work investigating the activity of CYP24A1 function utilised radioactivity and photodiode-array detectors to demonstrate variants that resulted in complete loss-of-function to small levels of activity retention[127]. Similar techniques could be applied to assessing the CYP24A1 activity in this mutant cell line to compare functional differences between wildtype, protein coding and 3'UTR variants in the future *in vitro*.

7.3.3 CONCLUSION

In vitro disease modelling can be harnessed to aid better understanding of genetic diseases, leading to the development of future improvements in therapy. Immortalised cell lines are frequently used for modelling disease due to their ease of manipulation and transfection. The CMH cell line that has been established in this study can aid future investigations into the localisation and interactions of *CYP24A1* and the effect that 3' UTR variants have CMH disease mechanism. This model provides a robust basis to assess the hypothesis previously highlighted in this thesis: i) CYP24A1 upregulation is the expected homeostatic response to increased serum calcium, hence increased protein on a western blot, while mRNA structural elements signal for an unknown and detrimental post-translational modification hindering protein function. (ii) 3' UTR variants in *CYP24A1* have altered or removed pre-existing MREs preventing miRNA binding and causing protein upregulation. (iii) mRNAs are trafficked from the nucleus to regions where the subsequent protein will be required more rapidly; mRNA localisation to specific regions within a cell provide regulation of protein expression but mRNA misfolding interferes with this trafficking process[340]. Given that CYP24A1 is functional in the inner mitochondrial membrane in 1,25$(OH)_2$D target cells and that some RNAs, particularly long non-coding RNAs, are known to act as structural components to the mitochondrial membrane[288], it is possible that mRNA structural abnormalities ‘anchor’ translational machinery and prevent the protein from proper localisation. Improper localisation affecting translational machinery could go some way to describe increased CYP24A1 expression with little effect on mRNA transcription as was observed in the CMH cohort described in chapter 5.

CHAPTER 8: A METHOD FOR SINGLE MOLECULE mRNA *CYP24A1* DETECTION IN *IN VITRO* HUMAN CELL LINES AND *IN VIVO* SAMPLES

8.1 INTRODUCTION

Current methods for gene expression quantification, e.g., qPCR, are limited to providing quantification of gene abundance and lack information on transcript localisation[341,342]. Single molecule fluorescence in situ hybridisation (smFISH) provides accurate staining of individual RNA molecules in single cells, allowing effective analysis of gene expression[343–346]. smFISH can distinguish sub-cellular localisation of different RNA molecules, offering analysis of cellular variation in and between cell types[343].

Nucleic acid in situ labelling has been extensively developed and adapted since first discovery in the 1960s[347]. Initial in situ hybridisation protocols required autoradiography probes for visualisation, which was limited by low sensitivity and availability of sequence specific probes[347]. Fluorescence in situ detection of DNA followed, requiring indirect immunofluorescence labelling of polytene chromosomes in *Drosophila melanogaster* using labelled antibodies[348]. Bauman et al. later developed a direct fluorescence method for DNA without requiring labelled antibodies by using RNA based probes[349]. While this development of DNA imaging required RNA probes, RNA-FISH was first developed by Singer & Ward for the detection of actin mRNA in chicken skeletal muscle cells[350]. Initial RNA-FISH provided relative quantification of gene expression, while lacking a clear signal from individual fixed fluorophores, hindering quantification of distinct RNA transcripts.

The recent development of smFISH provided direct labelling using multiple single probes to resolve individual transcripts, which increased the sensitivity over previous in situ hybridisation methods[351,352]. Recent developments utilise 48 fluorescently labelled DNA oligonucleotides, which hybridise separate regions of target RNA transcripts[343,353]. The varying probe length, plus the probe abundance, produces high sensitivity, while reducing off target binding[343,]. smFISH has allowed visualisation of transcription location in cells plus localisation of long non-coding RNA. smFISH methodology was previously reported in fixed *Arabidopsis* root cells[343]. This work demonstrated the expression of the low abundance housekeeping gene, *Protein Phosphate1A* (PP2A), using a unique imaging pipeline for automated cellular transcript counting[343,]. This straightforward methodology for use in plant research allowed researchers to study gene expression of individual cells at single molecular resolution with relative ease, setting the outlook for smFISH in human research.

This work sought to establish a method built upon the plant-based smFISH protocol[354] for use in human cell lines. Many known RNA localisation elements are within the 3'UTR, meaning 3'UTR variants observed in CMH patients could alter *CYP24A1* RNA localisation leading to disease progression. Deciphering the localisation of RNA transcripts could provide valuable insight into the mechanism of disease.

The localisation of *CYP24A1* was investigated *in vitro* in HEK293T cells to assess the mechanism of transcription. This work presents a novel method for visualising single RNA transcripts *in vivo* from patient whole blood samples. This important advancement is valuable in relation to understanding patient disease and vitamin D abnormalities caused by novel *CYP24A1* mutations, which affect RNA transcription and protein expression. While this research focuses on *CYP24A1* to shed light on vitamin D metabolism, this method could be adapted to investigate a range of RNA in patient samples by designing probes for any RNA of interest.

8.2 CLINICAL SAMPLES

Negative control whole blood samples used for the *ex vivo* investigations in this chapter were collected at the Norfolk and Norwich University Hospital blood typing service (n=5). Exclusion criteria for control samples were those with a vitamin D, calcium or other metabolic disorder clinical history. Control samples were collected using the UK NHS Research Ethics Committee decision toolkit (http://www.hra-decisiontools.org.uk/ethics/).

8.3 RESULTS

8.3.1 *CYP24A1* smFISH IN HEK293T CELLS

Cells were analysed following the publicly available method for mRNA counting using ImageJ[184]. This Radial-Symmetry-FISH (RS-FISH) software is a robust and rapid method for accurately detecting single molecule spots. Individual mRNA spots were located by taking a maximum projection z-stack containing all probe channels. Culturing cells to 70% confluency before fixing and hybridizing with *POLR2A* probes provided clear single mRNA visualisation in HEK293T cells with minimal background observed under these conditions (Figure 8.1 A-D). The even cell-to-cell distribution of the housekeeping gene *POLR2A* provided confidence in the conditions used to stain cells. The clarity of each image allowed visualisation of potential sites of transcription due to intense clustering of single transcript signals, indicated by white arrows (Figure 8.1 B and D) plus cells underdoing mitosis (Figure 8.1 C). Due to the absence of a denaturing step in the smFISH method, the DNA binding sites are retained in the duplex structure and are unable to be hybridised. Therefore, any probe spots identified are from true mRNA transcripts and not DNA false positives.

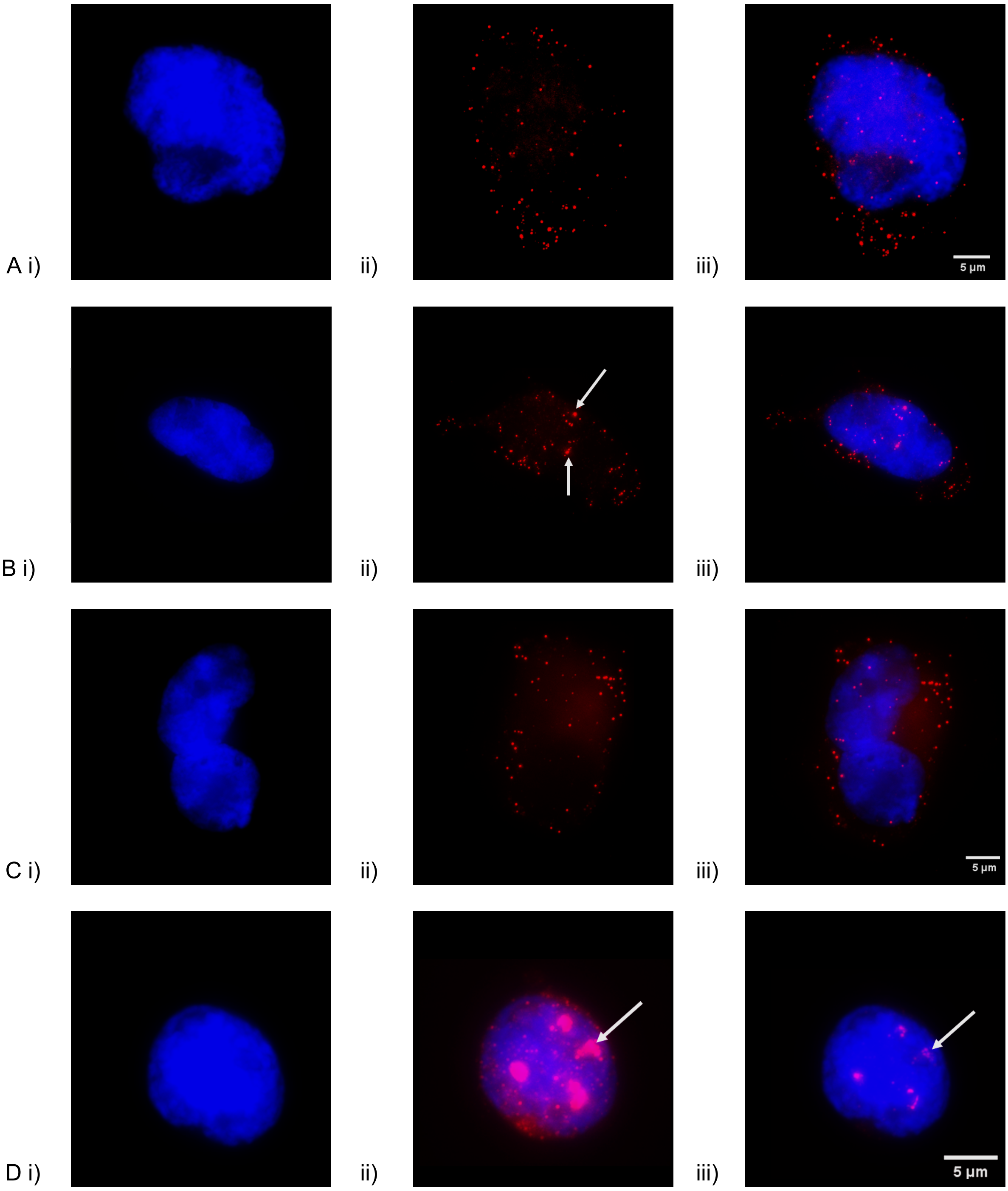


Figure 7.1: HEK293T cell expression of *POLR2A* using smFISH probes. Four individual HEK293T cells showing DAPI nuclear stain (blue) **(i)**, individual housekeeper *POLR2A* mRNA transcripts (red) **(ii)** and combined nuclear and POLR2A fluorescence **(iii)**. **Ai-iii)** even distribution of POLR2A probes observed throughout the nucleus and cytoplasm of the cell, **Bi-iii)** potential sites of POLR2A RNA transcription signalled by florescent cluster (white arrow), **Ci-iii)** HEK293T cells undergoing mitosis with even POLR2A probe distribution between each developing cell, **Di-iii)** POLR2A mRNA transcription occurring in nucleolus of cell indicated by florescent clusters (white arrow).

After confirming that the housekeeping gene *POLR2A* could be visualised in HEK293T cells using the smFISH method (Figure 8.1), the protocol was repeated for simultaneous detection of *POLR2A* and the gene of interest *CYP24A1*, plus MitoTracker Green for immunofluorescence mitochondrial staining. Initial experiments indicated that *CYP24A1* expression in HEK293T cells were below the detectable limit of the target Stellaris probes (Figure 8.2).

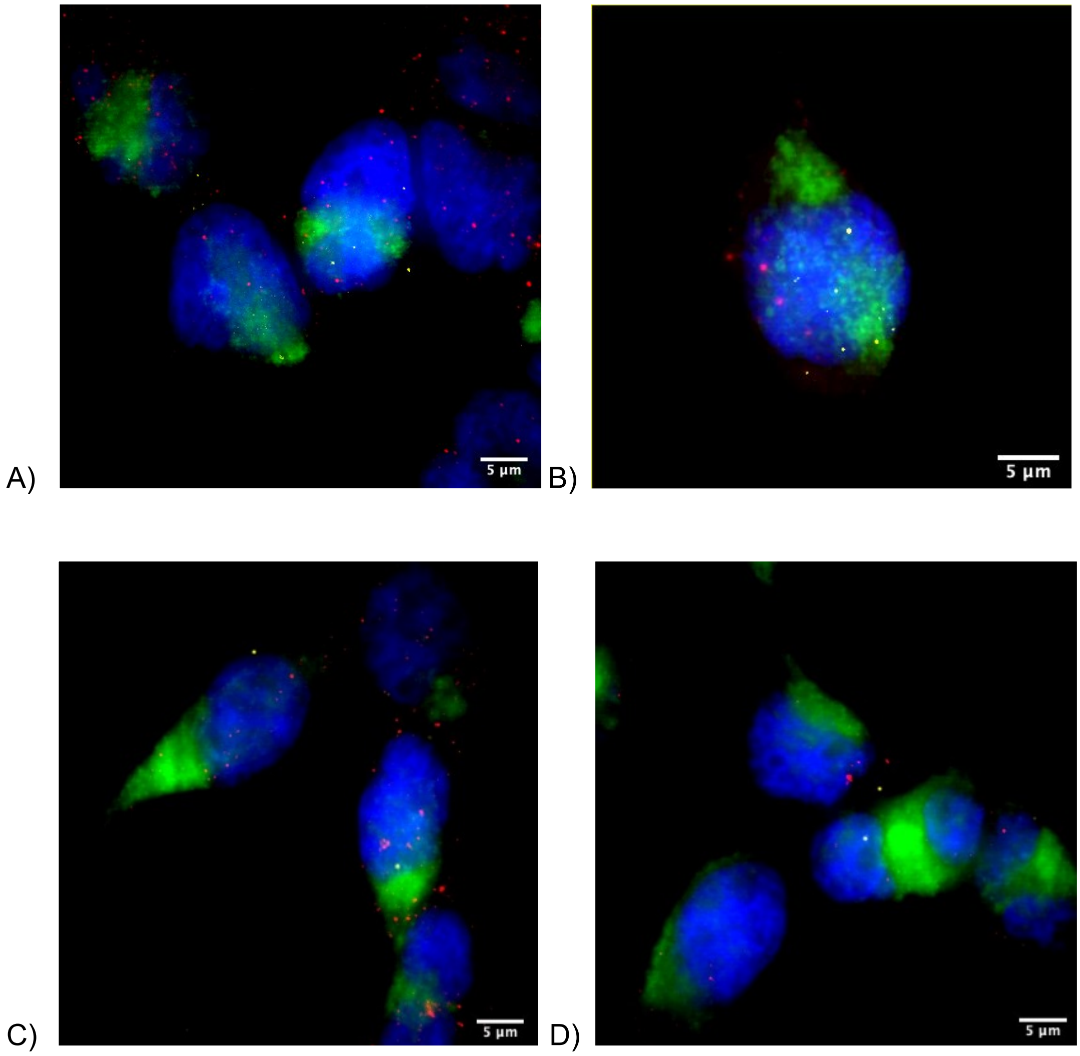


Figure 8.2: Identification of *CYP24A1*, *POLR2A* individual mRNA transcripts plus mitochondrial stain in HEK293T cells. A-D) Randomly selected HEK293T cells showing individual housekeeper *POLR2A* mRNA transcripts (red), *CYP24A1* mRNA transcripts (yellow), MitoTracker mitochondrial stain (green) and DAPI nuclear stain (blue). Minimal *CYP24A1* mRNA signal was observed and any CYP24A1 fluorescence observed was masked by MitoTracker staining.

While the immunofluorescence staining using MitoTracker Green provided insight into the mitochondrial location, the smearing and background fluorescence interfered with the smFISH visualisation of *CYP24A1* (Figure 8.2 A-D). As the initial focus of this work was to visualise *CYP24A1* mRNA transcripts, the MitoTracker Green stain was removed in further experiments to focus solely on increasing *CYP24A1* expression and visualisation.

Previous studies have reported that increased *CYP24A1* mRNA expression detected by qPCR was observed by treating cells with the vitamin D metabolite 1,25$(OH)_2$D, which stimulates CYP24A1 activity[326–330]. While this stimulation has predominantly been analysed by qPCR, this study sought to visualise 1,25$(OH)_2$D stimulation of *CYP24A1* RNA transcripts through increase *CYP24A1* smFISH probe identification. After stimulating the HEK293T cells with 10nm 1,25$(OH)_2$D for 12 hours prior to fixing, a significant increase in *CYP24A1* mRNA expression was observed (Figure 8.3 A-C). By comparing the average *CYP24A1* expression in treated and non-treated HEK293T cells, a significant difference after 1,25$(OH)_2D_3$ stimulus was observed, $p=0.0085$ (Figure 8.4).

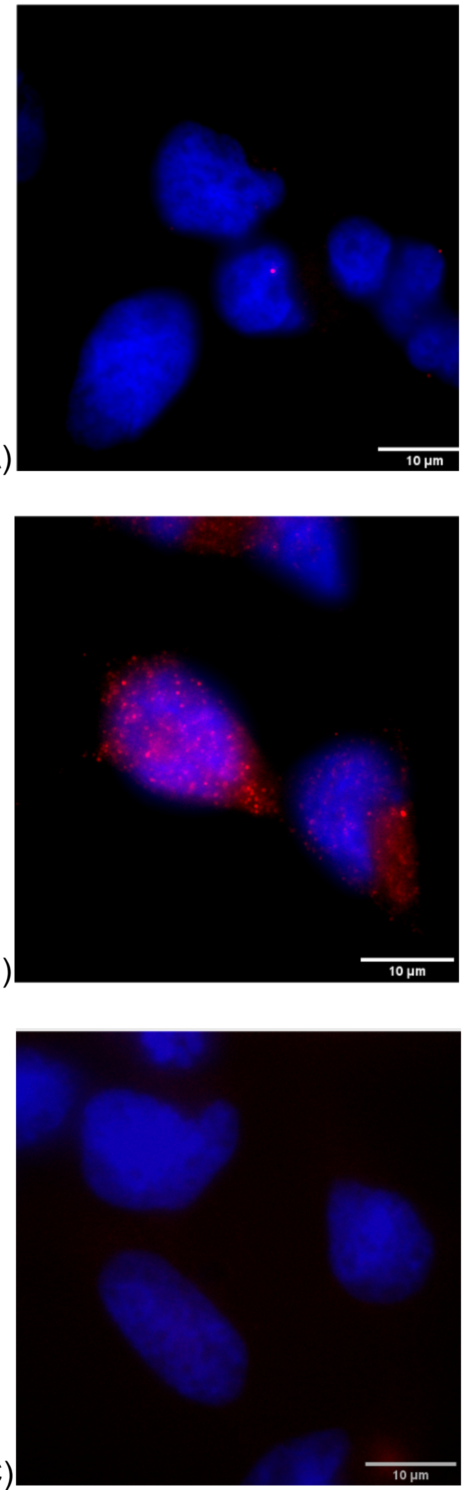


Figure 8.3: HEK293T divided into three groups to assess 1,25(OH)$_2$D stimulation on *CYP24A1* mRNA expression (red). Examples of each condition are shown minus the housekeeping gene: **A)** Unstimulated HEK293T cells show minimal *CYP24A1* expression. **B)** 10nM of 1,25(OH)$_2$D 12h significantly increased *CYP24A1* expression. **C)** No probe control grown and stained under identical conditions as **A** and **B**. DAPI stain (blue) indicates the nucleus of each cell.

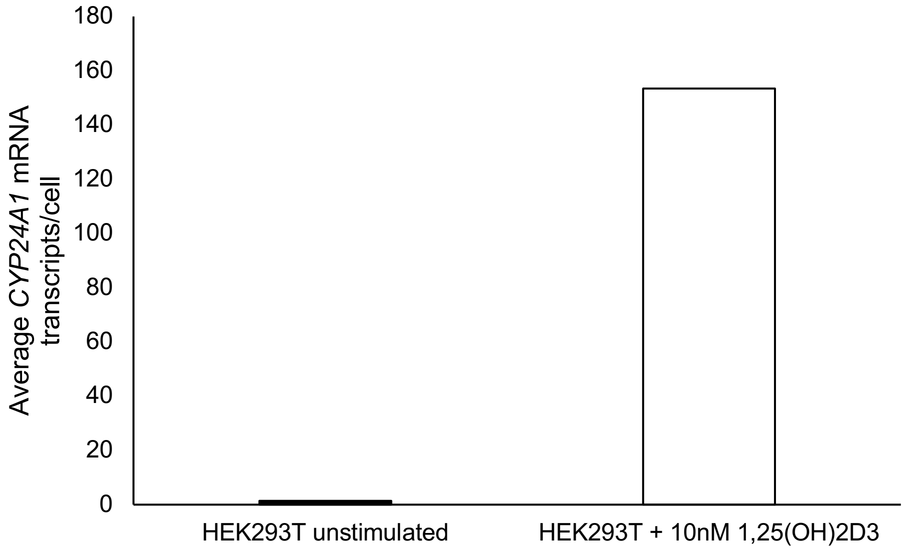


Figure 8.4: Comparison of the average *CYP24A1* mRNA transcripts/cell observed in wildtype HEK293T cells and HEK239T cells stimulated with 10nM of 1,25(OH)$_2$D (p = 0.0085).

Analysis of the 1,25$(OH)_2$D treated HEK293T cells revealed that there was also an increase in *CYP24A1* mRNA expression in comparison to *POLR2A* (Figure 8.4). Once stimulated, the smFISH method also allows visualisation of sites of *CYP24A1* transcription through probe clustering, which increases the fluorescence intensity at transcription sites. By stimulating *CYP24A1* with 1,25$(OH)_2$D potential sites of mRNA transcription in HEK293T cells were observed (Figure 8.5).

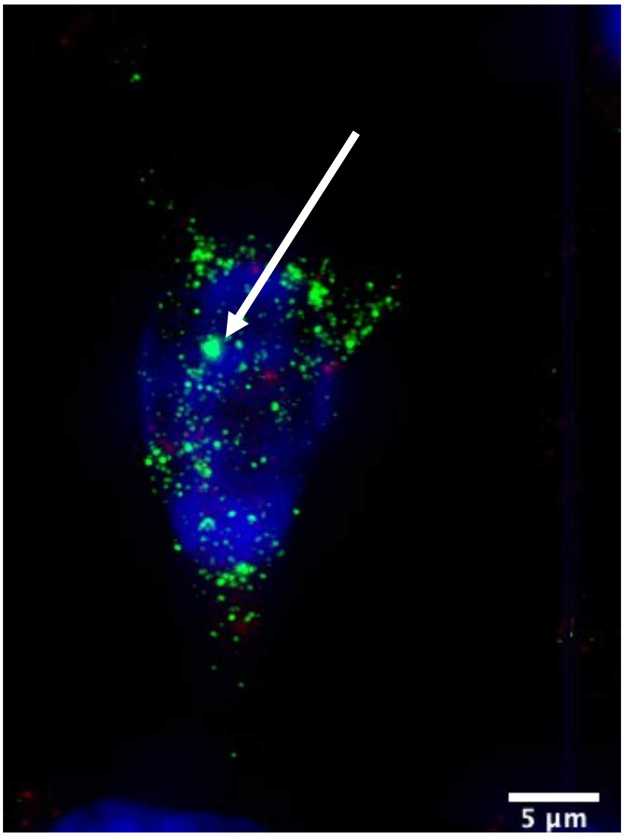
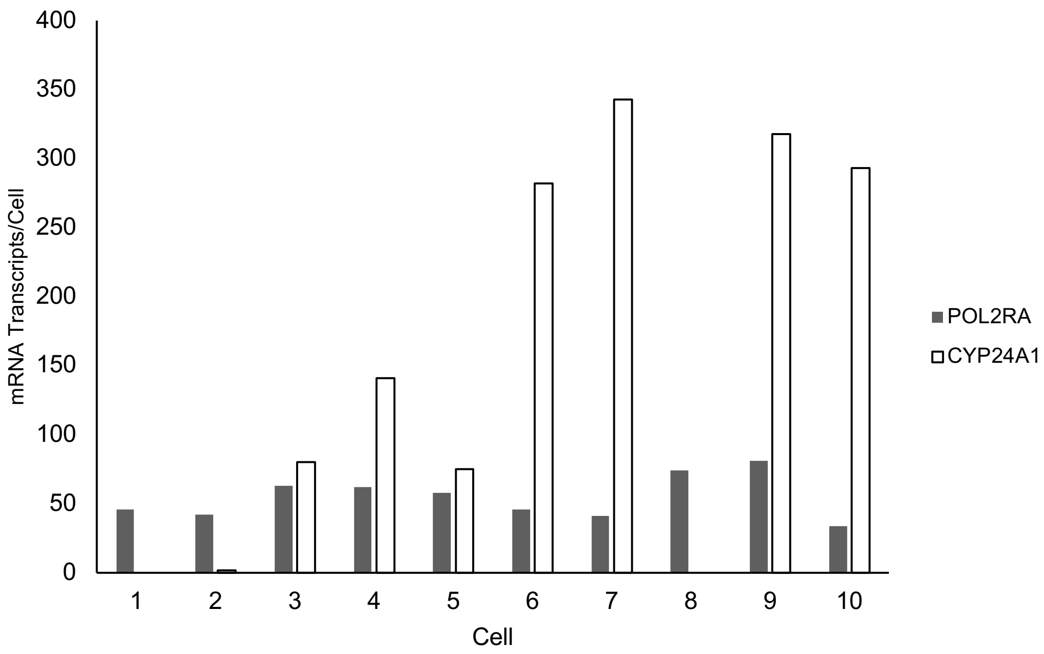
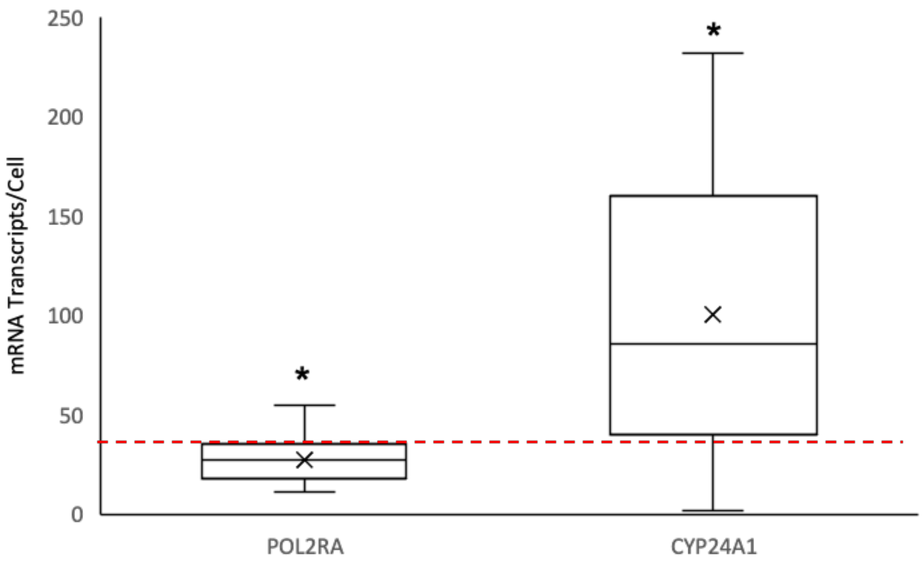


Figure 8.5: HEK293T cell stimulated with 10 nM 1,25(OH)$_2$D. *CYP24A1* mRNA transcripts (green) are more abundant than *POLR2A* (red) after 1,25(OH)$_2$D stimulation is introduced. Sites of *CYP24A1* transcription where multiple CYP24A1 mRNA transcripts are grouped in close proximity are shown by green fluorescent clustering (white arrow). DAPI stain (blue) indicates the nucleus of each cell.

When analysing the disparity in *CYP24A1* expression between cells, a significant increase in cell-to-cell variation between *CYP24A1* mRNA transcripts when compared to the housekeeping *POLR2A* probe frequency was observed, p = 0.005 (Figure 8.6 A and B). The median value for *POLR2A* falls outside of the interquartile range (IQR) of *CYP24A1* transcripts per cell count, plus the IQR of both groups do not overlap, emphasising the difference between the abundance of the two genes (Figure 8.6 B). The consistency of *POLR2A* confirmed the suitability of this transcript as the housekeeping gene.



A)



B)

Figure 8.6: The frequency of *CYP24A1* and *POLR2A* transcripts per cell A) The frequency of *CYP24A1* and *POLR2A* transcripts across different, randomly selected 1,25(OH)$_2$D stimulated HEK293T cell (n=10). **B)** A box and whisker plot displaying the significant difference in *CYP24A1* and *POLR2A* probe frequency between cells (p= 0.005). The red line shows a clear separation between the interquartile range of each group.

To further investigate the variability observed in *CYP24A1* expression between cells, this study sought to compare the sub-cellular location of *CYP24A1* transcripts. When comparing the frequency of each probe located in the cytoplasm and the nucleus of each cell, a significant difference between *POLR2A* and *CYP24A1* was observed (Figure 8.7). A larger range of transcripts per cell was observed in *CYP24A1* vs *POLR2A* with *CYP24A1* nucleus showing the greatest diversity of all probes and sub-cellular locations (Figure 8.7).

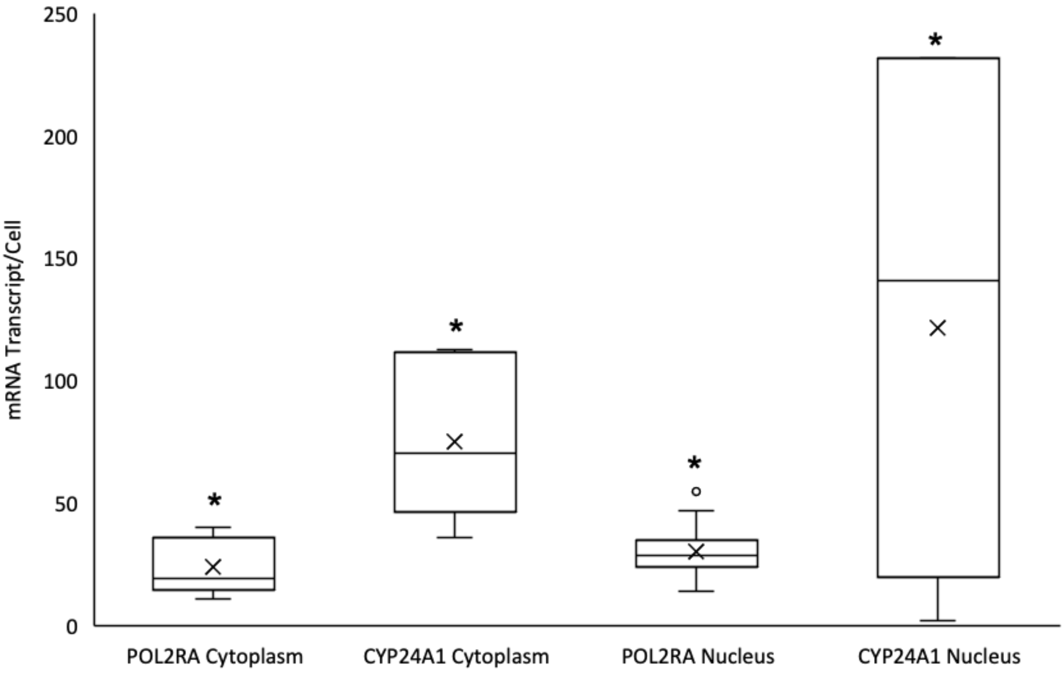


Figure 8.7: *CYP24A1* and *POLR2A* mRNA transcripts/cell in the cytoplasm and nucleus. A significant difference between frequency of *CYP24A1* and *POLR2A* was observed in the cytoplasm (*p=0.03) and the nucleus (**p=0.04) in 1,25$(OH)_2$D stimulated HEK293T cells (n=10).

Our results indicate substantial variation in *CYP24A1* probe frequency plus significant location variation in the cells. This study reports *CYP24A1* accumulation in some cells that have minimal nucleus expression (Figure 8.8 Ai, ii), *CYP24A1* low abundance or absent from nucleus and cytoplasm (Figure 8.8 Bi, ii) plus multiple sites of *CYP24A1* expression in the nucleus in cells with minimal probe detection in the cytoplasm (Figure 8.8 Ci, ii). This could be an indication of the effect cell cycle is playing on the expression and localisation of *CYP24A1*.

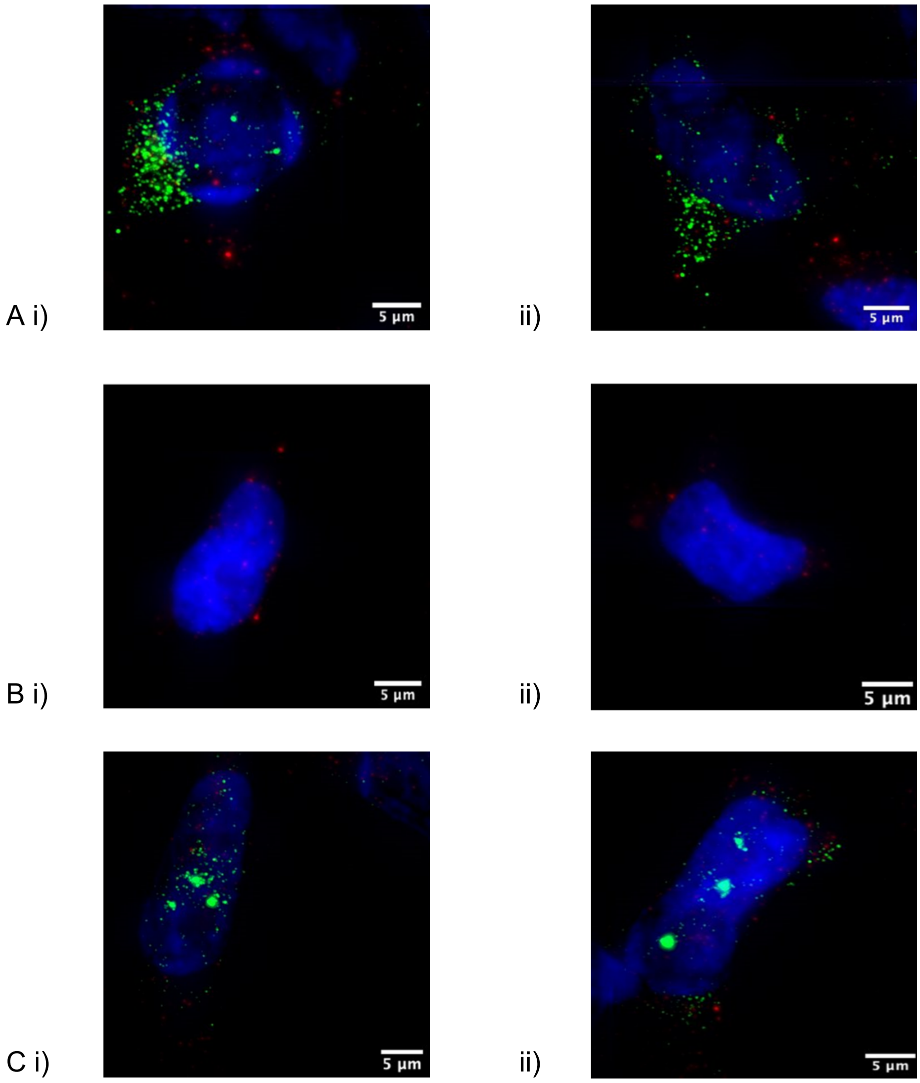
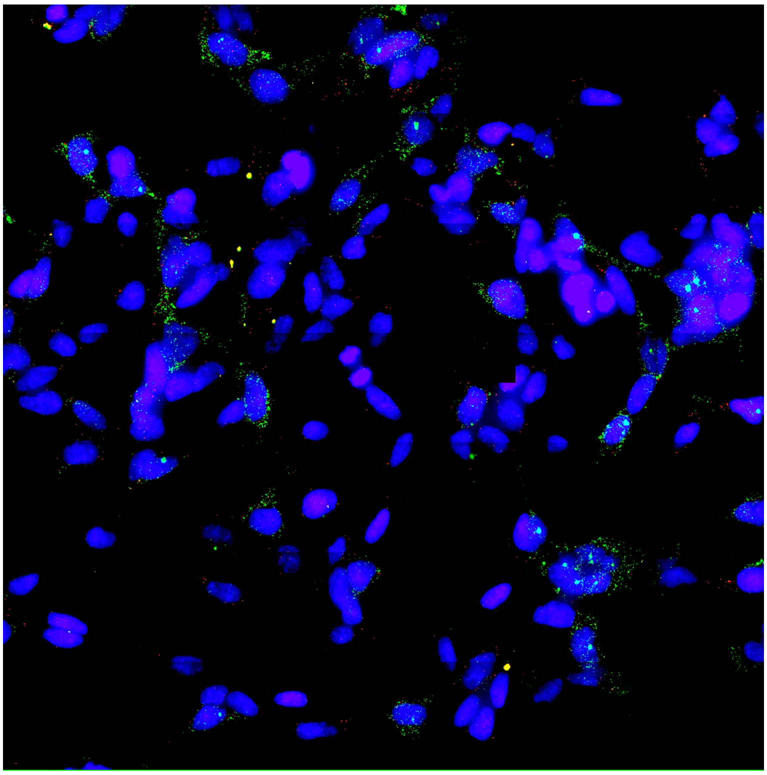


Figure 8.8: HEK293T cells with varying localisation and expression of *CYP24A1*. The probes indicate individual *POLR2A* mRNA transcripts (red), CYP24A1 mRNA transcripts (green) and DAPI stain (blue) in all images. **Ai and ii)** Accumulation of *CYP24A1* mRNA transcripts in the cytoplasm of the cell, **Bi and ii)** Cells containing no *CYP24A1* mRNA transcripts, absent of *CYP24A1*, **Ci and ii)** Intense sites of fluorescence observed in the nucleus indicating *CYP24A1* transcription, minimal probes elsewhere in the cell.

8.3.2 *CYP24A1* LOCALISATION IN CRISPR MODIFIED CELLS VS HEK293T

To investigate the effect that 3' UTR SNVs have on the localisation of *CYP24A1* mRNA, *CYP24A1* mRNA transcript locations were compared in wildtype HEK293T cells vs CRISPR-Cas9 genetically modified HEK293T cells. The CRISPR-Cas9 modified cells contain a SNV in the 3' UTR of *CYP24A1*, which has been predicted to modify the mRNA secondary structure. Both wild type and CRISPR modified HEK293T cells were treated with 10 nM 1,25$(OH)_2$D to stimulate *CYP24A1* expression before fixing and performing the Stellaris smFISH protocol as previously described. A similar diverse cell-to-cell frequency of *CYP24A1* was observed in CRISPR modified cells to the wild type (Figure 8.9 A and B).

A) 

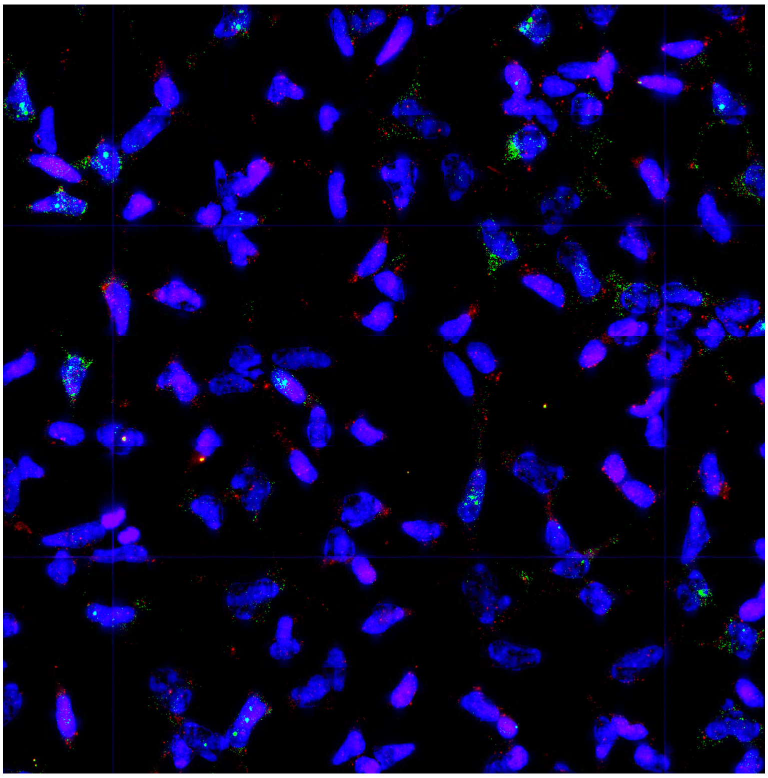
B) 

Figure 8.9: Cell lines stimulated with 10 nM 1,25(OH)$_2$D. A) CRISPR-Cas9 modified HEK293T cells, **B)** Wild type HEK293T cells. The fluorescent probes indicate individual housekeeping *POLR2A* mRNA transcripts (red), *CYP24A1* mRNA transcripts (green) and DAPI stain (blue) in both images.

Analysis of the frequency of *CYP24A1* probes per cell in the CRISPR modified cell line revealed a similar variance that was observed in the wild type HEK293T cells. In the CRISPR cell line, 15% of cells contained <10 transcripts per cell ranging to 10% of cells containing >250 cells. This varies significantly in comparison to the *POLR2A* frequency, which was relatively consistent in comparison at 0-50 across all analysed cells. Every cell that was analysed contained a minimum of 2 and a maximum of 35 *POLR2A* cells, while some cells were found to contain no detectable *CYP24A1* transcripts to a maximum of 317 (Figure 8.10).

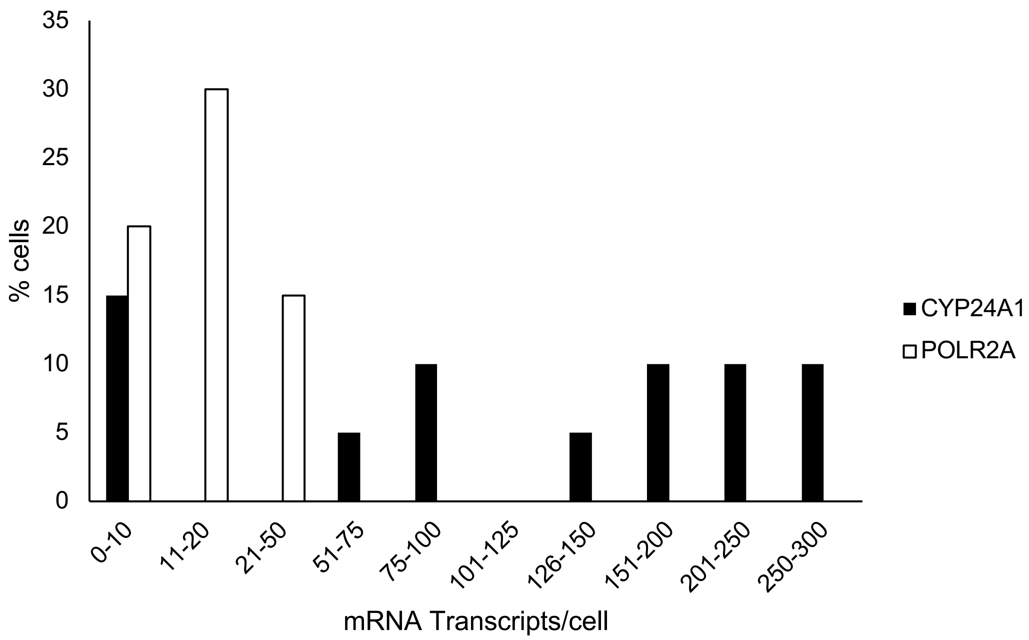
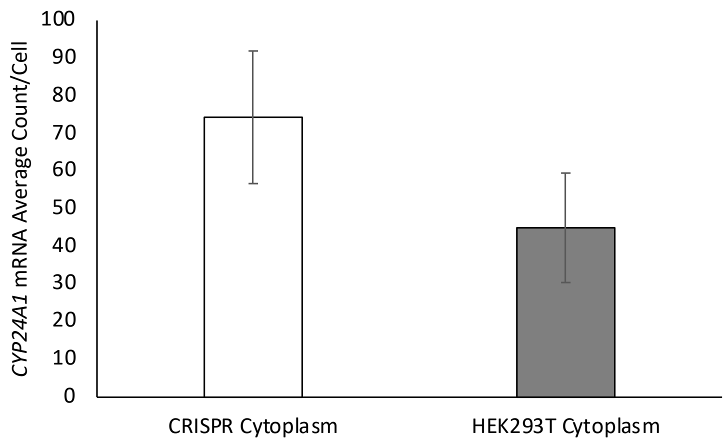
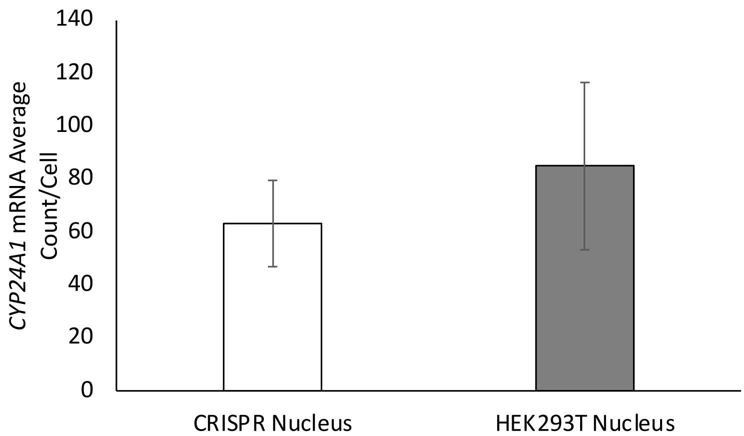


Figure 8.10: Percentage of CRISPR modified cells containing different frequencies of *CYP24A1* and *POLR2A* transcripts. *POLR2A* was only observed in cells at a frequency of 0-34, *CYP24A1* frequency ranged from 0-314.

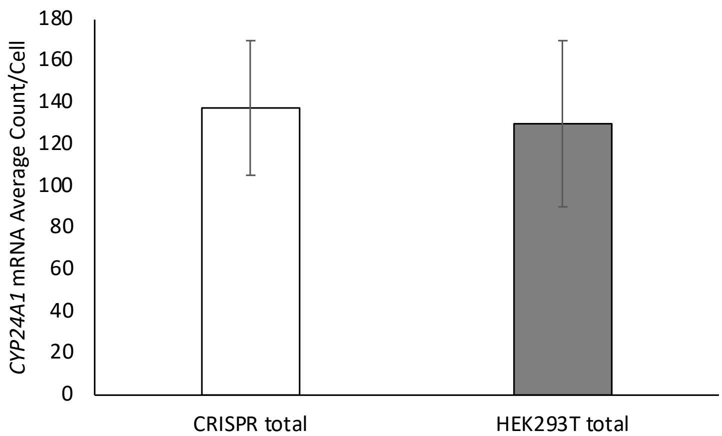
Initial comparison between CRISPR modified HEK293T cells and the wildtype appeared to show an increase in *CYP24A1* probes in the modified cell line (Figure 8.11 C). Further analysis of *CYP24A1* mRNA transcript frequency concluded that when comparing the total cellular probe frequency, plus the frequency in different sub-cellular locations, no significant difference between the CRISPR modified cell line and the wild type HEK293T cells was observed (Figure 8.11 A, B and C).



A)



B)

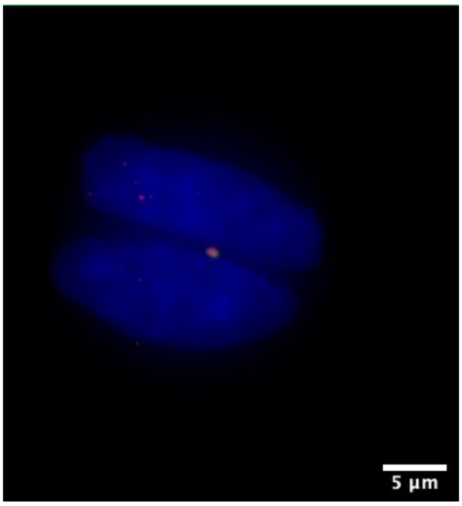


C)

Figure 8.11: Comparison of the average *CYP24A1* mRNA transcript counts per cell in two different sub-cellular locations of CRISPR Cas 9 cells (n=13) and HEK293T cells (n=10). A) cytoplasm, **B)** nucleus, plus **C)** the total *CYP24A1* mRNA count in CRISPR modified cells to HEK293T wild type. No significant difference was observed between cell type in any location, p=0.5, 0.4 and 0.8 respectively.

Similarly, to the previous observations of the wild type HEK293T cells, this study presents CRISPR modified cells with minimal to no *CYP24A1* probe signal (Figure 8.12 A). Interestingly in the modified line, this work presents cells seemingly both in the same stage of cell cycle with varying *CYP24A1* accumulation (Figure 8.12 A, B and 8.13).

When comparing cells in similar stages of the cell cycle, cells that contain potential sites of *CYP24A1* transcription have an apparent increased expression in *CYP24A1* cytoplasmic mRNA (Figure 8.12 A, B and 8.13).

A) 

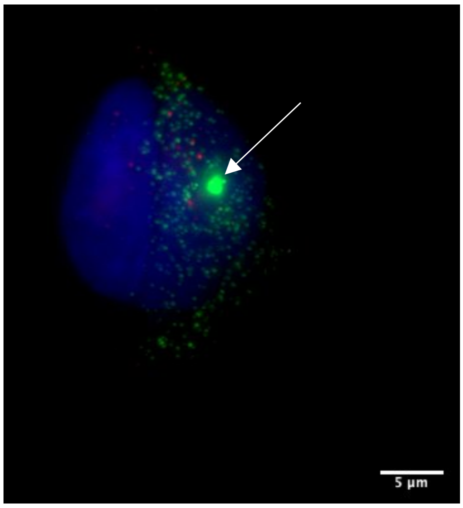
B) 

Figure 8.12: Two sets of CRISPR modified cells undergoing mitosis. A) Cells undergoing mitosis with zero *CYP24A1* mRNA transcripts **B)** Cells undergoing mitosis with different *CYP24A1* expression. A strong *CYP24A1* probe signal and *CYP24A1* transcription sight (white arrow) is shown in one, while minimal *CYP24A1* signal is observed in its pair. The cell that contains a potential site of *CYP24A1* transcription (white arrow) also expresses more *CYP24A1* in the cytoplasm. The probes indicate individual *POLR2A* mRNA transcripts (red), CYP24A1 mRNA transcripts (green) and DAPI stain (blue) in both images.

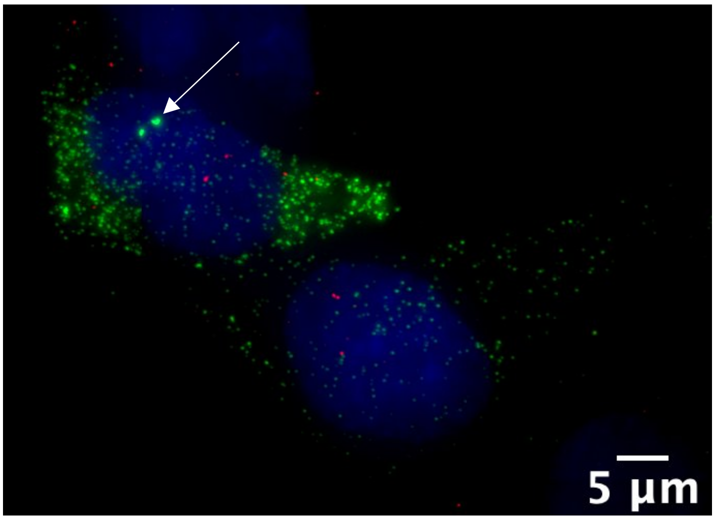


Figure 8.13: Two CRISPR modified HEK293T cells in close proximity presenting with varying *CYP24A1* mRNA transcript intensities in the cytoplasm. The cell that contains a potential site of *CYP24A1* transcription (white arrow) also expresses more *CYP24A1* in the cytoplasm. The probes indicate individual POLR2A mRNA transcripts (red), CYP24A1 mRNA transcripts (green) and DAPI stain (blue) in both images.

8.3.3 *CYP24A1* smFISH IN PATIENT SAMPLES

While cell line *in vitro* studies have the potential to unveil novel findings about cell biology, *ex vivo* studies would allow important insight into the effect of *CYP24A1* 3' UTR mutations on mRNA transcript location. To further adapt the Stellaris protocol, which successfully visualised *CYP24A1* in human cell lines, this study sought to detect *CYP24A1* mRNA transcripts in human control blood samples. Peripheral blood mononuclear cells (PBMCs) extracted from a single control patient whole blood via Ficoll-Plaque separation were seeded onto glass cover slips before performing smFISH using both *CYP24A1* and housekeeping *POLR2A* probes as before (Figure 8.14).

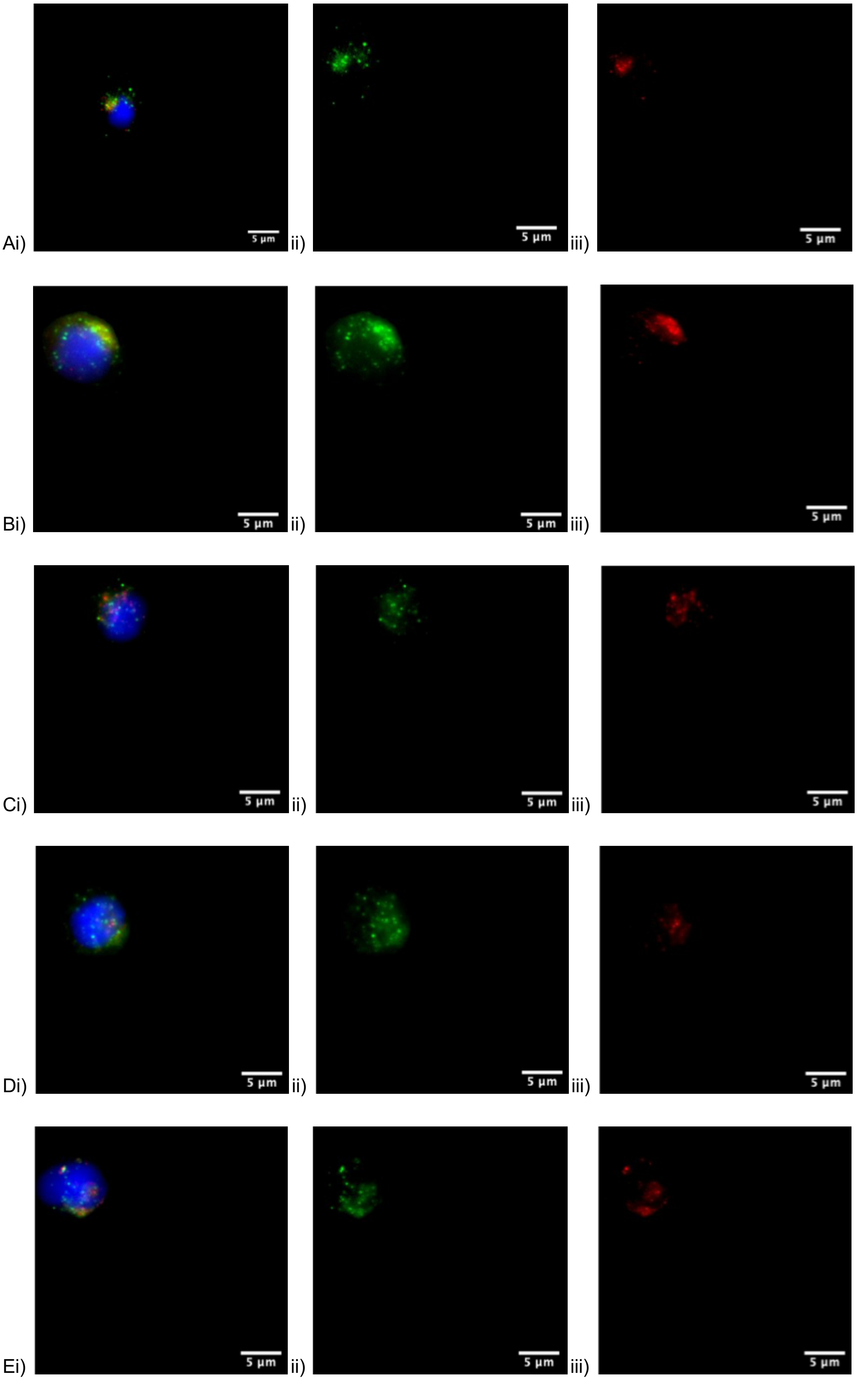

Ai)
5 µm
ii)
5 µm
iii)
5 µm
Bi)
5 µm
ii)
5 µm
iii)
5 µm
Ci)
5 µm
ii)
5 µm
iii)
5 µm
Di)
5 µm
ii)
5 µm
iii)
5 µm
Ei)
5 µm
ii)
5 µm
iii)
5 µm

Figure 8.14: Human PBMCs extracted from a single control patient whole blood. PBMCs were not treated with 1,25$(OH)_2$D stimulation prior to fixing to assess the naturally occurring *CYP24A1* expression *ex vivo*. **Ai-Ei)** Five separate cells from a single control patient sample with combined fluorescent imaging of DAPI (blue), *POLR2A* (red) and *CYP24A1* (green) expression combined. **Aii-Eii)** *CYP24A1* expression only **Aiii-Eiii)** *POL2RA* expression only.

The adapted Stellaris smFISH method for the detection smRNA in whole blood samples was successful in detecting *CYP24A1* and *POLR2A* probes. The initial method development images indicate that there is sufficient *CYP24A1* labelling in individual PBMC cell without the need for 1,25(OH)$_2$D stimulation. Less cell to cell variation was observed in *CYP24A1* in human PBMCs than *in vitro* cell lines. In comparison to HEK293T cells, the PBMCs have a smaller cell volume with a reduced cytoplasm area. *CYP24A1* was observed both in the nucleus and cytoplasm of PBMCs however, unlike 1,25(OH)$_2$D stimulated HEK293T cells, no obvious sites of transcription were identified. The lack of *CYP24A1* transcription sites in PBMCs may be due to the lack of 1,25(OH)$_2$D stimulation, which was shown in this study to elevate *CYP24A1* transcription *in vitro*.

8.4 DISCUSSION

smFISH provides single cell observation of *CYP24A1* individual mRNA transcripts in *in vitro* cell culture plus *ex vivo* human blood. Previous indications of *CYP24A1* mRNA abundance by qPCR are limited to solely quantifying the frequency of mRNA transcripts. By performing smFISH this work indicated not only the abundance of *CYP24A1* in individual cells but also the exact cellular location of individual mRNA transcripts.

8.4.1 *CYP24A1* LOCALISATION IN HEK293T CELLS

This study observed significant cell-to-cell variability in the expression and localisation of *CYP24A1* in the HEK293T cell line. The variability in cell-to-cell expression could be due to the progression of the cell cycle, which has been shown to effect mRNA transcription[355]. Prior to cell division, mRNA is increased in preparation for the reduced mRNA production. During early S phase when DNA is synthesised, mRNA abundance is minimal. The mRNA frequency increases gradually through the late S phase and into the G2 growth phase where the cell prepares for mitosis. High abundance of mRNA transcription is observed during the G2 and mitosis phase before drastically reducing as the cell reaches G1 growth phase and enters the next cell cycle[355,356]. These 'bursts' of RNA transcription occur at different stages throughout the cell cycle depending on the mRNAs abundance and function. Changes in mRNA abundance are also dictated by non-uniform degradation and transcription rates, which vary between genes. It has also been reported that regulation in mRNA transcription can alter throughout the cell cycle in response to transcription factor abundance or metabolic stimuli[357]. Gene expression noise relates to expression variation in isogenic cells maintained under identical conditions[358]. Intrinsic gene expression noise is an innate consequence of biochemical interaction fluctuations during transfection and translation e.g., promoter activation leading to 'bursts' in mRNA transcription[358]. Extrinsic gene expression noise occurs from upstream cellular variation, which can increase gene expression or cause mRNA degradation. Extrinsic noise is the main cause of cell variability[358]. While our experiments do not contain any cell cycle markers to identify the stage of division each cell is undergoing, it could explain why such variation is observed between the *CYP24A1* expression. Not all cells that appear to be in the same stage of the cell cycle express similar *CYP24A1* abundance (Figure 8.12 and 8.13). By introducing a cell cycle marker in future experiments, e.g. CDT1, SLBP, cytoplasmic Cyclin B1 and nuclear Cyclin B1 for G1, S, G2 and M phase determination respectively, this could shed light onto this abnormal presentation and aid our understanding of the effect cell cycle has on *CYP24A1* transcription[359].

The complex transcriptional process of cells is heavily influenced by kinetic parameters influencing the cell cycle. The significant cell-to-cell mRNA variability observed in this study may be missed by routine mRNA quantification methods such as qPCR that analyses mRNA abundance in samples as a whole. This extensive cell-to-cell variation was observed in human *CYP24A1* in this study. The ability to visualise RNA sub-cellular localisation in the nucleus or cytoplasm of cells also provides insight into CYP24A1 activity. Previous studies have reported that retention in either the nucleus or cytoplasm of cells can affect gene expression and allows regulation of transcriptional 'bursts'[356]. smFISH is therefore a compelling tool in investigating the mRNA life cycle of different genes of interest.

8.4.2 *CYP24A1* LOCALISATION IN THE CMH CELL MODEL

When comparing the CRISPR modified HEK293T cells to the wild type HEK293T cells, this work initially indicated an apparent increase in *CYP24A1* abundance. Once analysing the probe frequency between the two cell lines it was shown that there was no significant difference in *CYP24A1* mRNA transcription. The lack of significant variability in *CYP24A1* abundance between the two cell lines suggests that the 3' UTR SNV in the CRISPR modified cell line likely has no effect on *CYP24A1* mRNA transcription rate. While an increase in *CYP24A1* mRNA was observed in the cytoplasm and a decrease in nuclear *CYP24A1* in the CRISPR modified cell line, this was shown to be not significantly different to the HEK293T wild type. The localisation of *CYP24A1* transcripts do not appear to be altered due to the presents of the 3' UTR SNV. It is worth noting that although the nucleus and cytoplasm probe count can be estimated, due to the Z-stack style of imaging, a small proportion of the cytoplasm signal will be contained in the nucleus count and lost from the cytoplasm count.

8.4.3 CONCLUSION

While previous studies have investigated *CYP24A1* mRNA abundance in patient samples via qPCR techniques[326–330], this chapter presents visualisation of individual mRNA transcripts in single cells *in vivo* using the adapted smFISH technique. This unique technique for analysis of mRNA transcripts in human samples can be used to analyse patient with *CYP24A1* 3' UTR mutations in comparison to control patients plus the CRISPR modified *CYP24A1* 3' UTR model that has been established in this thesis. The application of *in vivo CYP24A1* localisation could further our understanding of how abnormal mRNA folding from 3' UTR SNVs can lead to the phenotype observed in our patient cohort. Human *in vivo* studies using smFISH may allow accurate visualisation of mRNA abundance and localisation relating to structure-function relationships, which is a major advancement on current qPCR techniques in RNA biology.

By adapting the Stellaris smFISH method for the use on human PBMCs this work highlights the potential to address important CMH disease pathogenesis questions *in vivo*. The *CYP24A1* visualisation could be utilised on a patient-by-patient basis to assess the effect that different *CYP24A1* mutations (both protein coding and non-coding) have on mRNA stability, expression, localisation and transcriptional 'burst' rates in the future. As previously discussed in chapter 5, a small heterogenous sample of PBMCs was used for this analysis. Future work would benefit from isolating different individual cell types in the PBMC pool before smFISH analysis is performed to effectively compare between patient samples as the PBMCs investigated in this chapter are of unknown classification. This novel technology applied to *in vivo* human samples opens up new lines of investigation into the effect of 3' UTR mutations on mammalian transcription. *In vivo* smFISH technology can be utilised in future work to improve our understanding of transcription beyond plant RNA biology.

CHAPTER 9: DISCUSSION

With much still to uncover in the understanding of *CYP24A1* loss-of-function mutations, this thesis aimed to investigate the disease pathogenesis of *CYP24A1* hypomorphic variants in patients with non-classical CMH. This work has shed light on the CMH patient cohort from phenotype to genotype utilising biochemical, genomic and proteomic analysis, setting a precedent for further research into this rare disease mechanism. This thesis demonstrates the strengths of implementing functional genomics and applying new "omic" techniques to established endocrine diseases.

9.1 LC-MS/MS ANALYSIS UNCOVERS THE RELATIONSHIP BETWEEN VITAMIN D METABOLITE RATIOS ASSOCIATED WITH PTH REGULATION AND CALCIUM HOMEOSTASIS

The implementation of high-throughput LC-MS/MS technology with excellent sensitivity has improved clinical research over traditional immunoassay methods for vitamin D metabolism analysis. Previous immunoassay methods for 25OHD measurement had poor standardisation and increased cross reactivity resulting in over estimation of metabolites in patient samples. Immunoassay methods can therefore lead to incorrect diagnosis and inappropriate therapy[360]. Although more technically demanding compared to fully automated immunoassays, improved vitamin D metabolite analysis performed by LC-MS/MS removes the cross reactivity encountered with immunoassay measurement of 25OHD by differentiating between $24,25(OH)_2D$ and 25OHD in both the D_3 and D_2 forms. Differentiation is achieved simultaneously from a single sample, allowing for the rapid assessment of patient's vitamin D catabolic status[150]. LC-MS/MS is required for the quantitative evaluation of $24,25(OH)_2D$ due to the low concentrations present in human serum[361,362]. LC-MS/MS is currently the only available method providing enough sensitivity to accurately measure 24,25$(OH)_2$D and is therefore the gold standard technique for simultaneous detection and differentiation of 24,25$(OH)_2$D from alternate vitamin D metabolites.

Initial work in chapter 3 of this thesis used LC-MS/MS technology to explore the relationship between serum concentrations of vitamin D metabolites, 25OHD and 1,25$(OH)_2$D, when expressed as a relative ratio with serum 24,25$(OH)_2$D. The respective ratio between vitamin D metabolites provides a mechanistic view of the physiological response to vitamin D deficiency and toxicity. The accurate and specific LC-MS/MS analysis of vitamin D metabolites provides evidence that the conversion of 25OHD and 1,25$(OH)_2$D is linked to the catabolism of 25OHD and 24,25$(OH)_2$D. The inverse exponential correlation between 1,25$(OH)_2$D:24,25$(OH)_2$D VMR and 25OHD highlights the dynamic vitamin D metabolism process. Metabolite investigations can therefore be performed on a patient-by-patient basis by measuring serum 24,25$(OH)_2$D and calculating subsequent VMR values. This work identified that under sufficient vitamin D status, 1,25$(OH)_2$D and 24,25$(OH)_2$D are maintained in relative proportion. In vitamin D insufficiency, 1,25$(OH)_2$D:24,25$(OH)_2$D VMR progressively and significantly increases due to serum 1,25$(OH)_2$D production favouritism over CYP24A1 activity (Figure 9.1). The reduced CYP24A1 activity results in decreased production of 24,25$(OH)_2$D in the absence of sufficient 25OHD. Conversely, this research suggests that CYP24A1 conversion of 25OHD to 24,25$(OH)_2$D is favoured in hypervitaminosis conditions, where 1,25$(OH)_2$D is elevated. This increased conversion results in increased 25OHD:24,25$(OH)_2$D VMR due to elevated concentrations of 24,25$(OH)_2$D produced (Figure 9.1). Abnormal VMR values alongside absent or low concentrations of 24,25$(OH)_2$D serves as a key identifier of loss-of-function *CYP24A1* mutations in patients with hypervitaminosis D, a finding that has been supported in multiple publications[125,190,199] (Figure 9.1). The parallel analysis of different vitamin D metabolites and their respective VMR can allow differentiation in hypercalcaemic patients that harbour *CYP24A1* hypomorphic mutations plus similar conditions e.g., Williams-Beuren syndrome or *SLC34A1* loss-of-function variants[146,207,363,364,208]. LC-MS/MS analysis described here can therefore guide genetic testing of patients with inherited hypercalcaemia by narrowing the causes for genetic investigations.

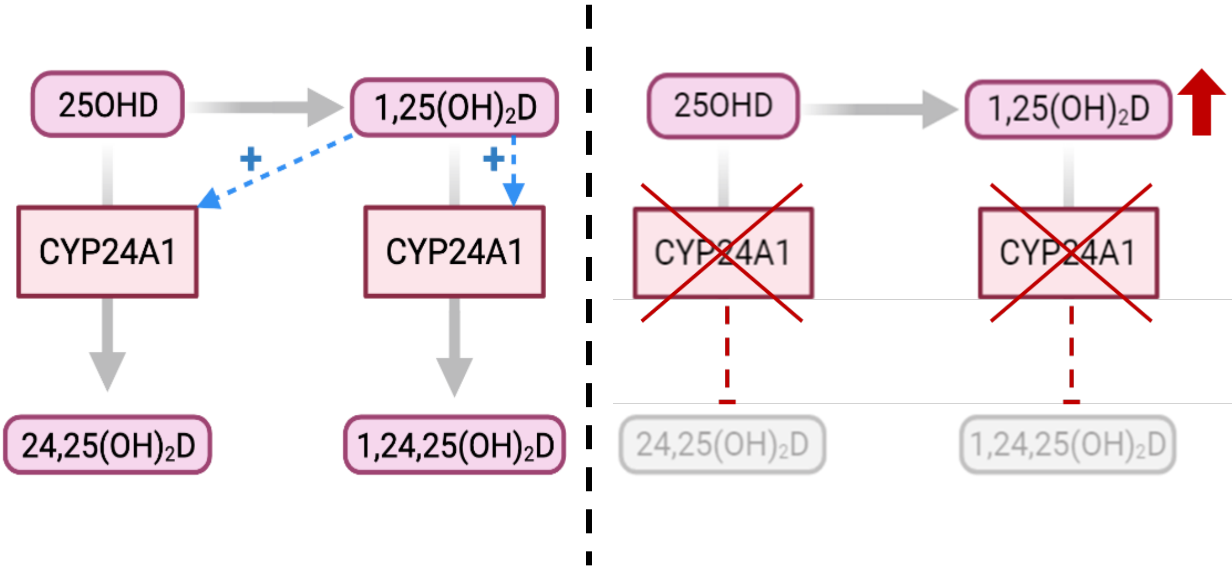


Figure 9.1: CYP24A1 regulation of 1,25(OH)$_2$D concentration. Increased 1,25(OH)$_2$D triggers CYP24A1 activity. Blue arrows indicating increased functional CYP24A1 stimulation from 1,25(OH)$_2$D. CYP24A1 hydroxylates 1,25(OH)$_2$D into 1,24,25(OH)$_2$D and 25OHD into 24,25(OH)$_2$D to prevent hypervitaminosis D. Loss-of-function mutations in *CYP24A1* result in an inability to respond to elevated 1,25(OH)$_2$D concentrations. Reduced CYP24A1 activity causing 1,25(OH)$_2$D over accumulation, abnormal VMR and can lead to vitamin D toxicity and hypercalcemia.

The work in chapter 3 also presents evidence of the relationship between vitamin D metabolites 25OHD, 1,25$(OH)_2$D, 24,25$(OH)_2$D and the distribution of PTH, which was published in Scientific Reports in 2019[192]. It is widely accepted in the literature that the PTH concentration is associated with 25OHD, but not with the active 1,25$(OH)_2$D. This is due to the tight regulatory mechanisms, and the regulatory processes that take place via the VDR to activate intracellular transport of calcium and stimulate PTH secretion. Using the 1,25$(OH)_2$D:24,25$(OH)_2$D VMR and 25OHD model, this study shows that individuals with low 25OHD (≤50 nmol/L), normal 1,25$(OH)_2$D but high 1,25$(OH)_2$D:24,25$(OH)_2$D VMR (≥101) have significantly higher PTH concentration than those at the opposite end of the spectrum. Relatively low 24,25$(OH)_2$D could enhance the anabolic effects of vitamin D metabolism, by stimulating PTH production. The three-dimensional model depicted in this study provides insight into the mechanism of the vitamin D-PTH endocrine system and the strong correlation between metabolites that are linked with PTH in human physiology.

CYP24A1 mutations are associated with PTH independent inappropriately elevated 1,25$(OH)_2$D concentrations. Recent publications have investigated the ratio of PTH to vitamin D metabolites in relation to non-canonical *CYP24A1* phenotypes[365,366]. The 1,25$(OH)_2$D:PTH ratio has been shown to increase in patients with CYP24A1 loss-of-function mutations due to lack of 1,25$(OH)_2$D catabolism. Elevated 1,25$(OH)_2$D:PTH is consistent with PTH independent production of 1,25$(OH)_2$D, aiding the differentiation between *CYP24A1* mutations and primary hyperparathyroidism[365,366].

While the work in chapter 3 was limited to only healthy, young adults, the identified relationship between vitamin D metabolites emphasises that the measurement of 1,25$(OH)_2$D, 25OHD, 24,25$(OH)_2$D, PTH and comparative ratios should be considered a part of the clinical workup in patients with hypercalcemia of otherwise unknown etiology. The data presented in this thesis highlights the ability to identify potential *CYP24A1* loss-of-function mutations in patients by using this sensitive LC-MS/MS analysis of serum vitamin D metabolites, PTH and subsequent VMR values. CYP24A1 activity also catalyses vitamin D metabolites through the C23-hydroxylation pathway producing 25OHD,26,23-lactone and 23,24,25$(OH)_3$D. Although not measured in this thesis, reduced 1,24,25$(OH)_2$D, 25OHD,26,23-lactone and 23,24,25$(OH)_3$D have been linked to *CYP24A1* loss-of-function[208]. Future analysis of patients with suspected *CYP24A1* variants may benefit from further vitamin D metabolite analysis beyond those listed in this study (Figure 9.2).

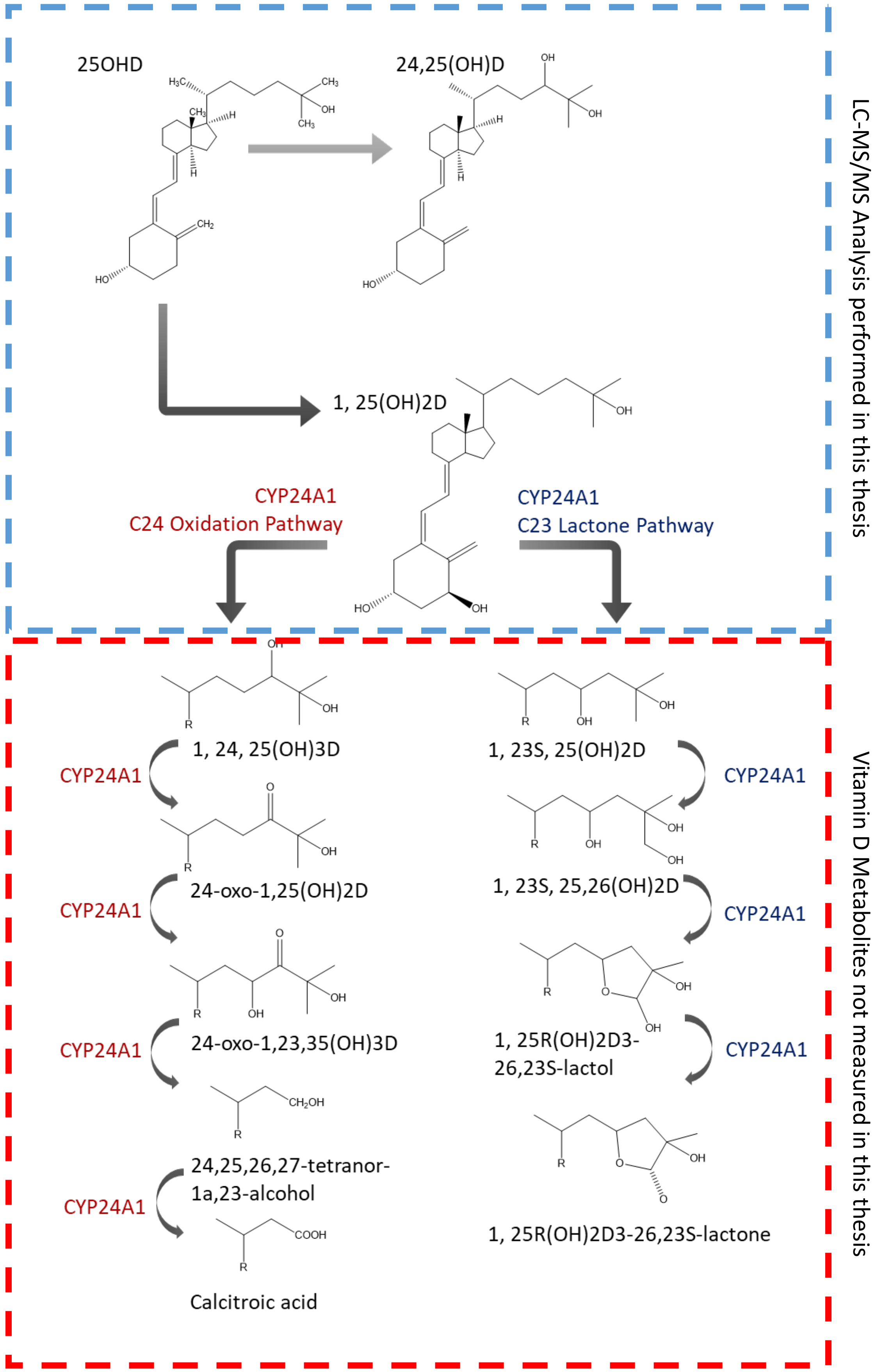

25OHD
24,25(OH)D
1, 25(OH)2D
CYP24A1
C24 Oxidation Pathway
CYP24A1
C23 Lactone Pathway
LC-MS/MS Analysis performed in this thesis
1, 24, 25(OH)3D
CYP24A1
24-oxo-1,25(OH)2D
CYP24A1
24-oxo-1,23,35(OH)3D
CYP24A1
24,25,26,27-tetranor-1a,23-alcohol
CYP24A1
Calcitroic acid
1, 23S, 25(OH)2D
CYP24A1
1, 23S, 25,26(OH)2D
CYP24A1
1, 25R(OH)2D3-26,23S-lactol
CYP24A1
1, 25R(OH)2D3-26,23S-lactone
Vitamin D Metabolites not measured in this thesis

Figure 9.2: CYP24A1 hydroxylation pathways analysed in this thesis. The blue box signifies the vitamin D metabolites measured by LC-MS/MS in this thesis. The red box indicated the remaining CYP24A1 enzymatic hydroxylation pathway of 1,25$(OH)_2D$ via C-24 oxidation (in red) and C-23 hydroxylation (in blue). Water soluble end products from each pathway (calcitroic acid and 1,25$(OH)_2D_3$-26,23-lactone) are excreted in bile to prevent hypervitaminosis D. LC-MS/MS detection methods for these remaining metabolites could be optimised in future experiments to expand the biochemical profile of patients with suspected CMH.

9.2 WHOLE EXOME SEQUENCING IN PATIENTS WITH SUSPECTED CMH UNCOVERED NO PROTEIN CODING MUTATIONS IN *CYP24A1*

Once the LC-MS/MS method was identified as a suitable method for analysis of vitamin D metabolites, plus reference thresholds confirmed for VMR values within a healthy population, this thesis sought to utilise this method to analyse patients with phenotypes consistent with CYP24A1 abnormal calcium handling. In chapter 4, the initial LC-MS/MS biochemical analysis of 147 patients, who presented with abnormal calcium handling, revealed that 9 patients had elevated 1,25$(OH)_2$D, of which 3 had elevated 25OHD:24,25$(OH)_2$D VMR and 4 had elevated 1,25$(OH)_2$D:24,25$(OH)_2$D. Analysing the vitamin D metabolites alone in this patient cohort would have revealed that the 25OHD and 24,25$(OH)_2$D were within the normal range and only 1,25$(OH)_2$D was elevated in all 9 patients. The production of 24,25$(OH)_2$D could be misleading as CYP24A1 appears to remain partially functional in this cohort. Only once the VMR analysis is performed did abnormalities arise in some patients, which could have been overlooked when analysing the vitamin D metabolites individually. Due to the tight relationship between 25OHD and 24,25$(OH)_2$D, the interpretation requires the combination of vitamin D metabolites individually plus the VMR to reveal true clinical relevance, as seen in this cohort.

The abnormal calcium handling phenotype in the cohort described in chapter 4, plus the abnormal biochemical results identified in 9 of these patients through LC-MS/MS analysis, were suggestive of potential loss-of-function mutations in *CYP24A1*. While LC-MS/MS biochemical analysis could suggest causative *CYP24A1* loss-of-function mutations in this cohort, loss-of-function *CYP24A1* mutations have been reported as extremely heterophenotypic e.g. adult onset nephrolithiasis, HCINF1, hypercalcemia and/or hypervitaminosis D. While LC-MS/MS evaluation can prove an important indicator of *CYP24A1* hypomorphic mutations, it is not conclusive. Genomic investigations in these 9 patients aimed to gain a boarder picture of each patient's genetic makeup by employing WES analysis, rather than a targeted gene approach, due to its extensive genetic profile output.

WES analysis performed on the 9 patients with abnormal calcium handling phenotypes, elevated 1,25$(OH)_2$D and abnormal VMR, revealed multiple germline mutations in genes that aid calcium handling and bone metabolism. While potential germline disease causing mutations were identified in this bone panel, no *CYP24A1* germline mutations were observed in this cohort. Most variants identified by WES within the bone panel investigated were somatic mutations, including 15 *CYP24A1* somatic mutations in 7 patients that have not been previously reported in association with CMH and/or hypervitaminosis conditions. The high abundance of somatic mutations within this cohort could be suggestive of potential false positives, due to the low frequencies and read counts, which may influence the interpretation of WES data. Conventional NGS error thresholds suggest that variant frequencies below 2%, of which there are many in this WES dataset, are subject to high risks of false positives regardless of coverage depth[259]. There is currently no universal consensus on the minimum coverage required in clinical research for NGS meaning there is variation between laboratory thresholds[248–250].

Results from the WES study in this thesis support the need for universal thresholds that would negate the issue of a high proportion of potentially false positive somatic mutations being reported. The high proportion of potential false positive WES results that would need to be confirmed by additional targeted sequencing reduces the advantage of WES analysis as an initial assessment due to the requirement of repeat sequencing to confirm results, e.g., via Sanger sequencing[264]. Although WES analysis provides a vast amount of genetic information from one data set, when inconclusive or negative WES findings are observed there is a tendency to halt further genetic investigations, which could leave patients undiagnosed or miss diagnosed. For example, inconclusive somatic *CYP24A1* variants were observed in chapter 4, which could result in these patients being deemed as not having *CYP24A1* hypomorphic germline mutations as the cause of their hypercalcaemic phenotype. Although alternate disease-causing variants other than loss-of-function *CYP24A1* mutations could be the cause of the hypervitaminosis phenotype observed in this cohort, WES is unable to effectively detect variants outside of the exome. The genomic portions not analysed by WES could hold further information into each patient's clinical presentation. Many patients could go undiagnosed as disease causing variants outside the coding region could be missed e.g., the UTRs. While WES analysis in this thesis did not display any germline *CYP24A1* mutations in the patient cohort in chapter 4, the understanding that large non-coding regions of the *CYP24A1* gene were not analysed suggested that further studies may benefit from sequencing the non-coding region. Analysis into the non-coding regions could potentially uncover variants missed by WES in patients that lack coding region variants.

While Sanger sequencing could provide added information on the non-coding regions of targeted genes, specific genes would need to be investigated. Targeted approach would remove the advantage of WES analysis that provides information on the whole exome in a single experiment however, would provide greater information on all coding and non-coding regions of individual genes. Although WES is increasingly used in clinical investigations, little is still known about the potential disease associated variants in the non-coding regions of genes. Variants in the non-coding region of DNA can alter gene activity and can cause structural changes affecting gene function.

9.3 mRNA STRUCTURAL ELEMENTS IN THE 3' UTR DICTATE CYP24A1 INTRACELLULAR ACTIVITY

In chapter 5 of this thesis, a second cohort of patients presenting with abnormal calcium handling phenotypes were investigated. Initial identification through LC-MS/MS vitamin D analysis and VMR determination indicated possible *CYP24A1* hypomorphic mutations (inappropriately elevated 1,25(OH)$_2$D with normal 25OHD concentrations and/or abnormal VMR). To investigate potential CMH phenotypes in this patient cohort, chapter 5 sought to further the *CYP24A1* mutational analysis to include the 5' and 3' UTR non-coding regions by implementing targeted Sanger sequencing. The decision to investigate the non-coding region of *CYP24A1* further in this cohort was supported by previous publications reporting an absence of disease causing mutations in the protein coding region of *CYP24A1* despite patients displaying an analogous phenotype (e.g., hypercalcemia or HCINF1) to those with protein coding loss-of-function *CYP24A1* mutations[125,148]. This absence of protein coding variants is consistent with the lack of germline *CYP24A1* variants identified by WES sequencing previously in this thesis.

Sanger sequencing of *CYP24A1* in chapter 5 revealed that each patient harboured *CYP24A1* SNV within the non-coding region. As discussed, the identified *CYP24A1* 3' UTR mutations in this cohort would have potentially been missed by broad WES analysis that focuses on the protein coding region alone. The varied 3' UTR SNV locations observed in this cohort may be indicative of the heterogeneous phenotype observed in patients with CMH[136]. This variation in variant location supports the diverse phenotype and biochemical analysis observed in previous patient cohorts plus the potential existence of unexplained disease mechanisms (Figure 9.3)[136, 148].

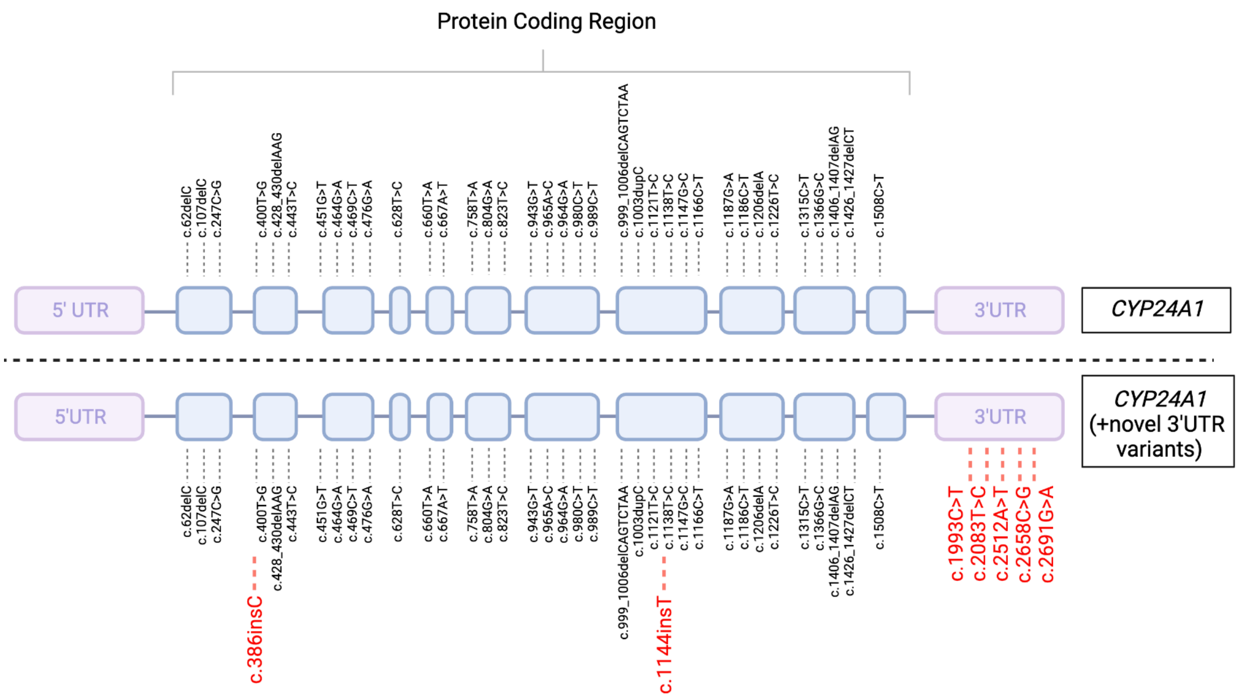


Figure 9.3: Linear maps of *CYP24A1* and the location of the pathogenic mutations previously identified[363] vs location of novel 3'UTR variants identified in this thesis. Each rectangle represents a single exon, whereas the horizontal lines represent intronic regions. The top panel depicts the *CYP24A1* gene annotated with previously identified protein coding mutations (black) [363]. The bottom panel shows the *CYP24A1* gene with the addition of *CYP24A1* 3' UTR SNVs and protein coding mutations (red) identified my direct sequencing in chapter 5. Five SNVs were detected that reside in the 3' UTR (c.1993C>T; c.2083T>C; c.2512T>A; c.2658C>G; c.2691G>A) across six individuals.

This study reports that the SNVs located in the *CYP24A1* 3' UTR of each patient in chapter 5 are predicted to lead to *CYP24A1* mRNA misfolding, as determined by RNAFold software. RNA structure flexibility caused by 3' UTR structural motifs is associate with the dynamics of biological function as mRNA secondary structure is significantly influenced and regulated by the 3' UTR[306–308]. The specific sequence and structure of mRNA is essential for post-transcriptional modulation processing, localisation, translation and degradation[306–308]. The diverse architecture of mRNA secondary structure is therefore an added layer of gene regulation[280]. Prior to this study, the effect of *CYP24A1* loss-of-function mutations affecting the mRNA secondary structure has not been reported, highlighting the impact of these findings. The work from this thesis presents a new area for investigation into mRNA misfolding in CMH disease pathogenesis that prior to this study is yet to be explored. Additionally, previous studies have reported 3' UTR polymorphism in the other genes associated with a spectrum of human disease[174,175,176]. This highlights the impact that this work could have on a wider field of clinical research beyond structural determination of *CYP24A1* mRNA alone.

The *in silico* mRNA structure outputs indicate the 3' UTR SNVs in *CYP24A1* affect the mRNA structure and could result in enzymatic inactivation. The structural changes triggered by 3' UTR variants may destabilise mRNA[306], disrupt mRNA trafficking[306,307], reduce ribosome scanning efficiency[308] and/or cause translational heterogeneity or impaired translation[306]. Impaired production of functional CYP24A1 due to structural alterations arising from 3' UTR variants could explain the elevated 1,25$(OH)_2$D concentration and abnormal VMR observed by LC-MS/MS analysis in the patient cohort. Lack of functional CYP24A1 would reduce 1,25$(OH)_2$D and 25OHD conversion into 1,24,25$(OH)_2$D and 24,25$(OH)_2$D respectively.

Digital PCR, western blot and ELISA analysis was performed to investigate potential mechanistic effects that mRNA structural changes had on transcription and translation of CYP24A1. Digital PCR results indicated no significant effect on *CYP24A1* transcription in patients with 3' UTR variants. In contrast, western blot and ELISA analysis suggest that the RNA structural abnormalities contribute to an over accumulation of an apparently less active or inactive CYP24A1 protein. As transcription appears unaffected, while a translational impairment in this cohort is observed, this work provides novel insight into what stage the CYP24A1 enzyme function was compromised. This thesis aimed to address and explore the cause of the observed CYP24A1 functional alteration in CMH patients further.

Investigations into the consequence of 3' UTR mRNA structural alterations led to three potential hypotheses that needed to be addressed; i) CYP24A1 upregulation is the expected homeostatic response to increased serum calcium, hence increased protein on a western blot, while mRNA structural elements signal for an unknown and detrimental post-translational modification hindering protein function. (ii) 3' UTR variants in *CYP24A1* have altered or removed pre-existing MREs preventing miRNA binding and causing protein upregulation. (iii) mRNAs are trafficked from the nucleus to regions where the subsequent protein will be required more rapidly; mRNA localisation to specific regions within a cell provide regulation of protein expression but mRNA misfolding interferes with this trafficking process[340]. Given that CYP24A1 is functional in the inner mitochondrial membrane in 1,25(OH)$_2$D target cells and that some RNAs, particularly long non-coding RNAs, are known to act as structural components to the mitochondrial membrane[288], it is possible that mRNA structural abnormalities 'anchor' translational machinery and prevent the protein from proper localisation. Improper localisation affecting translational machinery could go some way to describe increased CYP24A1 expression with little effect on mRNA transcription as was observed in this CMH cohort.

The work from this thesis emphasises the role of RNA misfolding in Human disease. Although the structural alterations were predicted *in silico* to differ from the wildtype *CYP24A1* structure, this work sought to further the understanding of said RNA structural alterations by determining the *in vivo* mRNA *CYP24A1* structure in the CMH patient cohort. Recent work has developed *in vivo* mRNA structure technologies with appropriate sequencing depth that can be used to elucidate the *in silico* structure predictions[280]. By investigating the effect of 3' UTR variants on mRNA structure using such *ex vivo* techniques, this thesis pursued accurate characterisation mRNA structure-function relationships in CMH patients, which would be a key advancement in *CYP24A1* research, plus wider investigations in mammalian RNA biology beyond standard *in silico* prediction analysis.

9.4 PROBE BASED RNA STRUCTURE ANALYSIS REQUIRES FURTHER OPTIMISATION TO DETERMINE 3' UTR VARIANTS IN *CYP24A1*

Chapter 6 of this thesis utilised novel DMS/SHAPE-LMPCR to visualise the *in vivo* structure of *CYP24A1* beyond the capabilities of previously published, less sensitive SHAPE-RT methods[280]. SHAPE reagents, like NMIA use in this thesis, react with each nucleotide along the RNA sequence allowing sensitive assessment of the structural context of each individual base. SHAPE reactivity is increased in regions where nucleotides are non-constrained/single stranded nucleotide regions. While bioinformatic predictions, such as RNAFold used in chapter 5 of this thesis, provide key insight into *in silico* predictions of CYP24A1 structural alteration, they cannot identify higher order structures e.g., intermolecular interactions and pseudoknots[297]. *In silico* methodology assumes that RNA can fold into a single structure only, rather than averaging probing data across all structures produced in a single experiment.

Previous studies have indicated that there is potential poor alignment between structures determined by chemical probing and those predicted by *in silico* software e.g., RNAFold[177]. Ding et al demonstrated poor comparability between *in vivo* DMS-restrained structure and unrestrained *in silico* predictions of the same RNA[177]. This poor comparability suggests that RNA structures cannot be predicted solely using RNA sequence and thermodynamic parameters. Although RNAFold is a useful prediction tool for RNA secondary structures, confirming the true RNA structure requires methods such as probe-based chemical modification techniques. By integrating chemical probing methods like DMS/SHAPE-LMPCR profiling alongside mutagenesis and phenotype analysis in CMH patient cohorts that harbour predicted 3' UTR structural altering SNVs, confirmation and better understanding of the role that RNA secondary structure holds on biological function could be unveiled. This study aimed to utilise DMS/SHAPE LMPCR methodology to visualise the low abundance transcript *CYP24A1* in human whole blood.

The novel simultaneous DMS/SHAPE LMPCR methodology provides increased sensitivity and selectivity in contrast to chemical probing alone, allowing low abundance RNAs to be detected from routine RNA extraction[280]. While *CYP24A1* was detectible by digital PCR in patient whole blood samples (chapter 5), the probe-based structure analysis by DMS/SHAPE-LMPCR identified that *CYP24A1* transcript abundance remained too low in RNA samples extracted from patient whole blood. Due to the low abundance of *CYP24A1* expression in samples utilised in this thesis for SHAPE structure determination, the structural analysis of *CYP24A1* in chapter 6 is not likely to produce a true mRNA secondary structure for *CYP24A1* due to low signal detected by CE resulting in poor alignment to the sequencing lane.

Although DMS/SHAPE-LMPCR provides improved detection of low abundance transcripts compared to traditional SHAPE-RT, further optimisation is required for visualisation of *CYP24A1* in human samples. Future probe-based *CYP24A1* DMS/SHAPE-LMPCR structure work would benefit from trialling different sample sources that express a greater abundance of *CYP24A1* e.g., skin tissue or kidney biopsies. Human samples of an invasive source, including biopsies, were beyond the ethics approval of this thesis and would need to be explored in further research. Alternatively, methods to stimulate *CYP24A1* expression prior to SHAPE treatment, e.g., vitamin D metabolite 1,25$(OH)_2$D stimulation provides potential to elevate *CYP24A1* abundance to an expression level suitable for SHAPE analysis and temporarily negate the need to source invasive patient tissue samples.

Once SHAPE analysis of *CYP24A1* in human samples is optimised, further research utilising the methodology described in this thesis could be extended to visualise the entire *CYP24A1* gene beyond the 3' UTR region focused on in this study. Optimisation would open up the opportunity to determine how mutations in in the 3' UTR affect the secondary structure and function of mRNA as well as the potential to shed light on the phenotypic variability in conditions affecting vitamin D metabolism such as CMH.

9.5 GENERATION OF A CMH MODLE CELL LINE USING CRISPR CAS 9

The limited availability of patient samples with rare non-canonical *CYP24A1* mutations can hinder the progression of CMH research. Only one CMH patient sample from chapter 5 was available to progress to *in vivo* SHAPE analysis after initial LC-MS/MS analysis and DNA sequencing. This limited the SHAPE method development due to the rarity of samples available for method development. The availability of a disease model that mimics CMH pathogenesis for initial *in vitro* investigations would greatly aid CMH research. The pathogenesis of the *CYP24A1* 3' UTR mRNA structure altering SNVs identified in this research remains unanswered, meaning a robust *in vivo* model of CMH would assist primary investigations into the interactions, localisation, transcription and translation of CYP24A1 in such patients.

While *CYP24A1* stimulation with 1,25$(OH)_2$D could be possible in cell lines for the *in vitro* structural determination of wildtype *CYP24A1* using DMS/SHAPE-LMPCR technology, there are no commercially available cell lines containing 3' UTR *CYP24A1* mutations mimicking conditions such as CMH seen in our patient cohort. As previously discussed, while the importance of the *CYP24A1* knockout mouse model that is currently available can be useful in further understanding patients with complete loss-of-function *CYP24A1* mutations, the CMH patients studied in this thesis appear to have only reduced function *CYP24A1* hypomorphic mutations. Currently no models are commercially available containing 3' UTR *CYP24A1* variants leading to mRNA structural alterations causing partially functional *CYP24A1* as seen in the CMH cohort presented in this thesis.

The work in chapter 7 presents the generation of a novel cell model for *CYP24A1* 3' UTR variants, that mimics conditions such as CMH observed in this thesis, generated in HEK293T cells using CRIPSR Cas9 technology. Immortalised cell lines such as HEK293T cells are frequently used for modelling disease due to their ease of manipulation and transfection. Additionally, HEK293T cells express *CYP24A1* making them a possible candidate for CMH cell line development. The newly generated HEK293T cell line was successfully transfected with 3' UTR deletions within *CYP24A1* (c. c.2026_2032del, c.2035_2037del and 2040-2041 del). Direct sequencing confirmed the location of the introduced modification. The deletion introduced by CRISPR-Cas9 in the HEK293T cell line was present in the 3' UTR region. RNAFold *in silico* prediction was initially utilised to demonstrate that the deletions introduced by CRISPR-Cas9 altered *CYP24A1* mRNA secondary structure folding, similar to what was observed previously in this thesis in CMH patients with *CYP24A1* 3' UTR SNVs. RNAFold results indicate that the desired structural alteration effect had been achieved by CRISPR-Cas9 gene modification.

To elucidate that the developed CMH cell line in chapter 7 influenced CYP24A1 function, similarly to the vitamin D metabolism effects observed in CMH patients, this work used LC-MS/MS biochemical analysis technology to investigate vitamin D metabolites in cell culture medium. The LC-MS/MS analysis revealed that after 25OHD treatment, there was a decrease in metabolism to $24,25(OH)_2D$ over 48 hours in the CMH cell line compared to wild type HEK293T cell line. This decreased metabolism rate corresponds with the phenotype of patients with CMH with biochemical analysis indicating reduced to undetectable $24,25(OH)_2D$ concentration due to partially functional/non-functional CYP24A1. Additionally, the CMH cell line had reduced $1,25(OH)_2D$ metabolism over 48 hours in comparison to the wild type HEK293T cell line, which metabolised $1,25(OH)_2D$ at a faster rate over the 48 hour time course under the same conditions. This lack of efficient $1,25(OH)_2D$ metabolism by the CMH cell line is consistent with the elevated $1,25(OH)_2D$ concentration observed in CMH patients resulting in the hypercalcemia/hypervitaminosis D phenotype, due to partially functional/non-functional CYP24A1. Further LC-MS/MS analysis using the CMH cell line could include $1,24,25(OH)_2D$. As the LC-MS/MS analysis implemented in this study cannot differentiate $1,24,25(OH)_2D$ from other vitamin D metabolites, this was not measured. Implementing $1,24,25(OH)_2D$ analysis could indicate the efficiency of CYP24A1 hydroxylation of $1,25(OH)_2D$ or lack of. The immunoassay quantifying $1,25(OH)_2D$ consistently overestimated the concentration of $1,25(OH)_2D$ in comparison to LC-MS/MS, which suggests possible cross reactivity of metabolites in the immunoassay not observed in LC-MS/MS measurement e.g.

$1{,}25(OH)_2D_2$. In the CMH cell line, the immunoassay detected a significant increase in $1{,}25(OH)_2D$ part way through the time course (12 hours) that was not observed by LC-MS/MS. This increase may be an example of the cross reactivity with is falsely elevating the immunoassay detection of $1{,}25(OH)_2D$.

In vitro disease modelling can be harnessed to aid better understanding of genetic diseases, leading to the development of future improvements of novel therapeutic techniques. The CMH cell line that has been established in this thesis can aid future research looking into the localisation and interactions of *CYP24A1* and the effect that 3' UTRs have on vitamin D metabolism and calcium handling. The work from this thesis has led to the generation of a robust CMH model with initial evaluation of vitamin D metabolites indicating suitability for modelling the pathogenesis of CMH. As previously discussed, one hypothesis for the CMH phenotype observed in this thesis is that *CYP24A1* mRNA is trafficked to intracellular regions where the resultant protein is required more rapidly[287]. Given that CYP24A1 is metabolically active in the inner mitochondrial membrane in $1{,}25(OH)_2D$ target cells and that some RNAs are known to act as structural components to the mitochondrial membrane[288], it is possible that mRNA structural abnormalities "anchor" translational machinery and prevent the protein from proper localisation. Improper localisation affecting translational machinery could explain the increased CYP24A1 expression with little effect on mRNA transcription as observed in this CMH cohort. The CMH cell model was therefore use in this thesis to begin investigations into mRNA localisation in wildtype cells and the mutant cell line containing *CYP24A1* mRNA structural alterations.

9.6 A METHOD FOR SINGLE MOLECULE mRNA *CYP24A1* DETECTION IN *IN VITRO* HUMAN CELL LINES

Commonly used techniques used to quantify RNA abundancy include qPCR and digital PCR methods. In addition to mRNA abundance, the development of novel smFISH analysis provides exact localisation of individual mRNA transcripts within individual cells, which is beyond the capabilities of qPCR analysis[343]. To the authors knowledge, at the time of reporting this thesis present novel single cell observation of *CYP24A1* individual mRNA transcripts *in vitro* plus *in vivo* in human blood using smFISH technology. This research presents not only the *CYP24A1* mRNA abundance, but additionally the exact cellular localisation within the *CYP24A1* mutant cell line, wildtype HEK293T cell line, plus human PBMCs.

The results in chapter 8 revealed a significant cell-to-cell variability in the expression and localisation of *CYP24A1* in HEK293T cells. It could be hypothesised that the cell-to-cell variability was observed due to the progression of the cell cycle, which has been shown to effect mRNA transcription rates. The complex transcriptional process of cells is heavily influenced by kinetic parameters influencing the cell cycle. The cells use for smFISH analysis may have been fixed at different stages of the cell cycle effecting the *CYP24A1* expression explaining the variation observed. This finding warrants further investigations into the expression of *CYP24A1* fluctuations throughout the cell cycle in different cell types under different conditions to further the understanding of *CYP24A1* transcription throughout the cell life cycle not only in CMH cells but additionally in the wildtype. Future investigations benefiting from a cell cycle marker could also determine why some cells observed in this work appear to be in the same stage of the cell cycle, however express different *CYP24A1* abundance and localisation and shed light onto this abnormal presentation, aiding the understanding of the effect that the cell cycle has on *CYP24A1* transcription.

The variability in *CYP24A1* localisation observed in the wildtype HEK293T cells would have been missed by routine mRNA quantification methods such as qPCR that analyse mRNA abundance in samples. The ability to differentiate *CYP24A1* sub-cellular localisation between nucleus and cytoplasm by using smFISH allows insight into *CYP24A1* activity. Previous studies have reported that retention in either the nucleus or cytoplasm of cells can affect gene expression and allows regulation of transcriptional 'bursts'[356]. smFISH is therefore a compelling tool in investigating the mRNA life cycle of different genes of interest.

Using the CMH cell line generated in this thesis compared to wild type HEK293T cell lines, smFISH analysis did not show a significant different in *CYP24A1* abundance. As no significant *CYP24A1* transcript variability was observed between CMH and HEK283T cells lines, the 3' UTR structure altering deletion introduced by CRISPR Cas 9 likely has little to no effect on *CYP24A1* mRNA transcription rate. This finding supports the lack of mRNA transcription variability observed in patient samples analysed by digital PCR in chapter 5. While smFISH analysis does highlight an increase in *CYP24A1* mRNA in the cytoplasm and a decrease in nuclear *CYP24A1* within the CMH cell line, this was deemed not significantly different to the HEK293T wildtype. This suggest that the 3' UTR structure altering deletion likely has little to no effect on *CYP24A1* localisation within the cell. Although no localisation or abundance differentiation was observed between CMH cell line and HEK293T cells, this work has allowed insight into the localisation of *CYP24A1* within cells, which has not previously been reported, regardless of *CYP24A1* 3' UTR variants.

9.7 A METHOD FOR SINGLE MOLECULE mRNA *CYP24A1* DETECTION IN *IN VIVO* SAMPLES

In addition to performing *in vitro* smFISH analysis, this thesis sought to develop *an in vivo* method to observe *CYP24A1* mRNA transcripts in human samples. This work presents novel visualisation of human *in vivo* mRNA transcripts using smFISH by observing *CYP24A1* in human PBMCs. Due to the unavailability of CMH patient whole blood, the initial *in vivo* investigations performed in this thesis used control patient blood to obtain PBMCs for analysis. This unique technique for mRNA transcript visualisation in human samples could be further optimised and used to analyse patients with *CYP24A1* 3' UTR SNVs in comparison to control patients. The application of *in vivo CYP24A1* visualisation could further our understanding of how abnormal mRNA folding from 3' UTR SNVs can lead to the phenotype observed in our patient cohort. By adapting the Stellaris smFISH method for the use on human PBMCs the thesis presents the potential to address important CMH disease pathogenesis questions *in vivo*. The *CYP24A1* visualisation could be utilised on a patient-by-patient basis to assess the effect that different *CYP24A1* mutations (both protein coding and non-coding) have on mRNA stability, expression, localisation and transcriptional 'burst' rates in the future. Human *in vivo* studies using smFISH will allow accurate visualisation of mRNA abundance and localisation relating to structure-function relationships, which is a major advancement on current qPCR techniques in RNA biology. In addition to 3' UTR effects, many studies investigating mammalian transcription could benefit from *in vivo* smFISH insight to improve our understanding of transcription beyond that of plant RNA biology.

9.8 FUTURE INVESTIGATIONS INTO UNDERLYING MECHANISM OF STRUCTURE ALTERING *CYP24A1* 3'UTR VARIANTS

Though this thesis could not describe the underlying mechanism linking mRNA structure with partially functional protein activity here, the generated CMH cell line model will allow further *in vitro* research into the pathogenesis of non-canonical *CYP24A1* mutations. Our work presents a pathway for employing ribosome frameshift profiling and/or protein sequencing to determine the core C-term region required for catabolising 1,25$(OH)_2$D, which only differs from 25OHD by the OH group. Some mRNAs carry specific structural elements in their 3' ends that cause ribosomes to slip and then readjust the reading frame[367]. The frameshift results from a change in the reading frame by one or more bases in either the 5′ (−1) or 3′ (+1) directions during translation[367]. Additionally, as the 3'UTR has been linked to mRNA stabilisation, investigating the stability of CYP24A1 in the newly generated cell line could indicate any effect on mRNA degradation that may lead to the elevated protein accumulation observed in this thesis. As 3'UTR variants identified in chapter 5 appeared to have no significant effect on transcription, while an apparent increase in translation was observed, this could be due to a decrease in mRNA degradation and should be explored in further work using the CMH cell line. Along with mRNA stability, the work presented in this thesis highlighted a variability in mRNA secondary structure between patients with differing 3'UTR mutations. This suggest that investigations into the correlation between 3'UTR variant location, mRNA structure, CYP24A1 function and ultimately the phenotype should be explored in future analysis. By investigating the correlation between variant location and phenotype, future work could determine if the 3'UTR or protein coding mutation location is causative of the phenotypic heterogeneity reported in CYP24A1 patients. Previous work investigating the activity of CYP24A1 function utilised radioactivity and photodiode-array detectors to demonstrate that differing mutations resulted in variable activity from complete loss-of-function to small levels of activity retention[127]. Similar techniques could be applied to assessing the CYP24A1 activity in this mutant cell line to compare functional differences between wildtype, protein coding and 3'UTR variants in the future *in vitro,* shedding light on the consequence of differing variant location.

As previously discussed in this thesis, ketoconazole has been identified as a promising treatment for patients with CYP24A1 abnormalities due to into CYP27B1 inhibiting properties[129,138,149,333]. While case reports have indicated success in patients with pathogenetic CYP24A1 mutations, treatment of patients with rare 3'UTR variants resulting in partially functional CYP24A1 have not been investigated. The CMH cell line generated in this thesis provides the potential for large scale *in vitro* studies using treatments like ketoconazole to investigate appropriate therapy for patients with non-canonical CYP24A1 mutations.

9.9 CONCLUSION AND MAIN CLINICAL FINDINGS

This thesis presents investigations into CMH using biochemical, genomics and proteomic techniques. Using LC-MS/MS technology, biochemical analysis was used to identify patients with suspected *CYP24A1* hypomorphic mutations leading to hypervitaminosis D and abnormal calcium handling complications. By implementing both WES and Sanger sequencing DNA analysis techniques, this research highlights the need for careful genetic investigation of patients with CMH. WES analysis may lead to missed hypomorphic variants if present in non-coding regions of *CYP24A1* resulting in potential misdiagnosis and improper treatment. This thesis presents identification of novel 3' UTR SNVs in a small cohort of CMH patients that were shown to result in mRNA secondary structure abnormalities and increased partially-functional CYP24A1 expression. One of the main clinical benefits of this work is the evidence that not all patients presenting with CMH have protein coding mutations and that further sequencing of the non-coding regions may reveal pathogenic variants in patients previously misdiagnosed. This work present initial investigations into *in vivo* determination of mRNA structural changes in patients with 3'UTR variants, providing a basis for further *in vivo* SHAPE secondary structure optimisation that would confirm *in silico* structure predictions in human samples.

An important development in this research was the generation of a CMH cell line model using CRISPR-Cas9 that mimics patients with 3' UTR variants causing mRNA structural alterations. The generation of this cell line can be utilised for *in vitro* investigations and potential treatments for CMH patients, alleviating some requirement for rare CMH patient samples and accelerating our understanding of CMH disease pathogenesis. Initial *in vitro* experiments with the CMH cell line using smFISH indicate the exact location of individual *CYP24A1* transcripts *in vitro,* providing novel observations of *CYP24A1* localisation in addition to mRNA abundance. Following from the *in vitro* smFISH analysis, this thesis presents a novel method for *in vivo* smFISH analysis in patient samples that can be utilised in future work to further understand the localisation of *CYP24A1* in CMH patient cohorts, when sample collection is permitted.

The findings from this thesis present innovative insights into novel *CYP24A1* hypomorphic variants and provides a foundation for further *in vitro* and *in vivo* investigations into CMH disease pathogenesis. The complexity of novel techniques used in this research plus the development of techniques and cell models produced that will not only aid research in this area but beyond, are a testament to the impact of this study. The newly developed CRISPR-Cas 9 mutated HEK293T cell line could provide the foundation for large scale *in vitro* studies and will continue to support our understanding of vitamin D metabolism in patients with novel 3' UTR mutations. The findings of this research provide a framework that can be used to better understand the molecular basis of pathogenesis in patients lacking protein coding region abnormalities.

REFERENCES

1. Darja Obradovic Wagner & Per Aspenberg. Where Did Bone Come From? *Acta Orthopaedica* **82**, 393–398 (2011).
2. Lee, M. S. Y., Soubrier, J. & Edgecombe, G. D. Rates of Phenotypic and Genomic Evolution during the Cambrian Explosion. *Curr. Biol.* **23**, 1889–1895 (2013).
3. Chemical and Physiological Change from Calcium Carbonate to Calcium Phosphate in Skeletal Structures. *Insights Biomed. Res.* **5**, (2021).
4. Benton, M., Donoghue, P. & Asher, R. Calibrating and constraining molecular clocks. in *The timetree of Life* (eds. null, H., B., S. & Kumar, S.) 35–86 (Oxford University Press, 2009).
5. Ito, Y. & Miyazono, K. RUNX transcription factors as key targets of TGF-beta superfamily signaling. *Curr. Opin. Genet. Dev.* **13**, 43–47 (2003).
6. Biology of Bone Tissue: Structure, Function, and Factors That Influence Bone Cells. https://www.hindawi.com/journals/bmri/2015/421746/.
7. Long, F. & Ornitz, D., M. Development of the Endochondral Skeleton.
8. Kenkre, J. & Bassett, J. The bone remodelling cycle. *Ann. Clin. Biochem.* **55**, 308–327 (2018).
9. Young, B., Woodford, P. & O'Dowd, G. *Wheater's Functional Histology E-Book: A Text and Colour Atlas*. (Elsevier Health Sciences, 2013).
10. Mackie, E. J., Tatarczuch, L. & Mirams, M. The skeleton: a multi-functional complex organ: the growth plate chondrocyte and endochondral ossification. *J. Endocrinol.* **211**, 109–121 (2011).
11. Yang, G. *et al.* Osteogenic fate of hypertrophic chondrocytes. *Cell Res.* **24**, 1266–1269 (2014).
12. Kronenberg, H. M. Developmental regulation of the growth plate. *Nature* **423**, 332–336 (2003).

13. Xiong, J. *et al.* Matrix-embedded cells control osteoclast formation. *Nat. Med.* **17**, 1235–1241 (2011).

14. Murshed, M., Harmey, D., Millán, J. L., McKee, M. D. & Karsenty, G. Unique coexpression in osteoblasts of broadly expressed genes accounts for the spatial restriction of ECM mineralization to bone. *Genes Dev.* **19**, 1093–1104 (2005).

15. Long, F. Building strong bones: molecular regulation of the osteoblast lineage. *Nat. Rev. Mol. Cell Biol.* **13**, 27–38 (2011).

16. Ducy, P., Zhang, R., Geoffroy, V., Ridall, A. L. & Karsenty, G. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. *Cell* **89**, 747–754 (1997).

17. Daoussis, D. & Andonopoulos, A. P. The emerging role of Dickkopf-1 in bone biology: is it the main switch controlling bone and joint remodeling? *Semin. Arthritis Rheum.* **41**, 170–177 (2011).

18. Marie, P. J. Transcription factors controlling osteoblastogenesis. *Arch. Biochem. Biophys.* **473**, 98–105 (2008).

19. Hartmann, C. A Wnt canon orchestrating osteoblastogenesis. *Trends Cell Biol.* **16**, 151–158 (2006).

20. Boyle, W. J., Simonet, W. S. & Lacey, D. L. Osteoclast differentiation and activation. *Nature* **423**, 337–342 (2003).

21. Fp, R. Osteoclast biology and bone resorption. *Primer Metab. Bone Dis. Disord. Miner. Metab.* (2006).

22. Udagawa, N. *et al.* Osteoprotegerin produced by osteoblasts is an important regulator in osteoclast development and function. *Endocrinology* **141**, 3478–3484 (2000).

23. Takayanagi, H. New developments in osteoimmunology. *Nat. Rev. Rheumatol.* **8**, 684–689 (2012).

24. Franz-Odendaal, T. A., Hall, B. K. & Witten, P. E. Buried alive: how osteoblasts become osteocytes. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **235**, 176–190 (2006).

25. Bonewald, L. F. & Johnson, M. L. Osteocytes, mechanosensing and Wnt signaling. *Bone* **42**, 606–615 (2008).

26. Dallas, S. L., Prideaux, M. & Bonewald, L. F. The osteocyte: an endocrine cell ... and more. *Endocr. Rev.* **34**, 658–690 (2013).

27. Chen, H., Senda, T. & Kubo, K. The osteocyte plays multiple roles in bone remodeling and mineral homeostasis. *Med. Mol. Morphol.* **48**, 61–68 (2015).

28. Feng, J. Q. *et al.* Loss of DMP1 causes rickets and osteomalacia and identifies a role for osteocytes in mineral metabolism. *Nat. Genet.* **38**, 1310–1315 (2006).

29. Quarles, L. D. Role of FGF23 in vitamin D and phosphate metabolism: implications in chronic kidney disease. *Exp. Cell Res.* **318**, 1040–1048 (2012).

30. Moester, M. J. C., Papapoulos, S. E., Löwik, C. W. G. M. & van Bezooijen, R. L. Sclerostin: current knowledge and future perspectives. *Calcif. Tissue Int.* **87**, 99–107 (2010).

31. Osterhoff, G. *et al.* Bone mechanical properties and changes with osteoporosis. *Injury* **47**, S11–S20 (2016).

32. Breeland, G., Sinkler, M. A. & Menezes, R. G. Embryology, Bone Ossification. in *StatPearls* (StatPearls Publishing, 2020).

33. Karsenty, G. & Wagner, E. F. Reaching a genetic and molecular understanding of skeletal development. *Dev. Cell* **2**, 389–406 (2002).

34. Winkler, D. G. *et al.* Osteocyte control of bone formation via sclerostin, a novel BMP antagonist. *EMBO J.* **22**, 6267–6276 (2003).

35. Guntur, A. R. & Rosen, C. J. Bone as an Endocrine Organ. *Endocr. Pract. Off. J. Am. Coll. Endocrinol. Am. Assoc. Clin. Endocrinol.* **18**, 758–762 (2012).

36. Zhou, R. *et al.* Endocrine role of bone in the regulation of energy metabolism. *Bone Res.* **9**, 25 (2021).

37. Andrukhova, O., Streicher, C., Zeitz, U. & Erben, R. G. Fgf23 and parathyroid hormone signaling interact in kidney and bone. *Mol. Cell. Endocrinol.* **436**, 224–239 (2016).

38. Cai, X., Xing, J., Long, C. L., Peng, Q. & Humphrey, M. B. DOK3 Modulates Bone Remodeling by Negatively Regulating Osteoclastogenesis and Positively Regulating Osteoblastogenesis. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **32**, 2207–2218 (2017).

39. Kalbasi Anaraki, P. *et al.* Urokinase receptor mediates osteoclastogenesis via M-CSF release from osteoblasts and the c-Fms/PI3K/Akt/NF-κB pathway in osteoclasts. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **30**, 379–388 (2015).

40. Matsuoka, K., Park, K.-A., Ito, M., Ikeda, K. & Takeshita, S. Osteoclast-derived complement component 3a stimulates osteoblast differentiation. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **29**, 1522–1530 (2014).

41. Shimada, T. *et al.* FGF-23 is a potent regulator of vitamin D metabolism and phosphate homeostasis. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **19**, 429–435 (2004).

42. Fraser, W. D. Hyperparathyroidism. *Lancet Lond. Engl.* **374**, 145–158 (2009).

43. Institute of Medicine (US) Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. *Dietary Reference Intakes for Calcium and Vitamin D*. (National Academies Press (US), 2011).

44. Marcadet, L., Bouredji, Z., Argaw, A. & Frenette, J. The Roles of RANK/RANKL/OPG in Cardiac, Skeletal, and Smooth Muscles in Health and Disease. *Front. Cell Dev. Biol.* **10**, (2022).

45. Kim, N., Odgren, P. R., Kim, D. K., Marks, S. C. & Choi, Y. Diverse roles of the tumor necrosis factor family member TRANCE in skeletal physiology revealed by TRANCE deficiency and partial rescue by a lymphocyte-expressed TRANCE transgene. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 10905–10910 (2000).

46. Li, J. *et al.* RANK is the intrinsic hematopoietic cell surface receptor that controls osteoclastogenesis and regulation of bone mass and calcium metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1566–1571 (2000).

47. Dougall, W. C. *et al.* RANK is essential for osteoclast and lymph node development. *Genes Dev.* **13**, 2412–2424 (1999).

48. Bucay, N. *et al.* osteoprotegerin-deficient mice develop early onset osteoporosis and arterial calcification. *Genes Dev.* **12**, 1260–1268 (1998).

49. Hughes, A. E. *et al.* Mutations in TNFRSF11A, affecting the signal peptide of RANK, cause familial expansile osteolysis. *Nat. Genet.* **24**, 45–48 (2000).

50. Whyte, M. P. & Hughes, A. E. Expansile Skeletal Hyperphosphatasia Is Caused by a 15-Base Pair Tandem Duplication in TNFRSF11A Encoding RANK and Is Allelic to Familial Expansile Osteolysis. *J. Bone Miner. Res.* **17**, 26–29 (2002).

51. Cundy, T. *et al.* A mutation in the gene TNFRSF11B encoding osteoprotegerin causes an idiopathic hyperphosphatasia phenotype. *Hum. Mol. Genet.* **11**, 2119–2127 (2002).

52. Langdahl, B. L., Carstens, M., Stenkjaer, L. & Eriksen, E. F. Polymorphisms in the Osteoprotegerin Gene Are Associated With Osteoporotic Fractures. *J. Bone Miner. Res.* **17**, 1245–1255 (2002).

53. Santa Maria, C. *et al.* Interplay between CaSR and PTH1R signaling in skeletal development and osteoanabolism. *Semin. Cell Dev. Biol.* **49**, 11–23 (2016).

54. Shaker, J. L. & Deftos, L. Calcium and Phosphate Homeostasis. in *Endotext* (eds. Feingold, K. R. et al.) (MDText.com, Inc., 2000).

55. Editorial: Paget's disease and calcitonin. *Br. Med. J.* **3**, 505–506 (1975).

56. Rodríguez-Ortiz, M. E. & Rodríguez, M. FGF23 as a calciotropic hormone. *F1000Research* **4**, (2015).

57. Kolb, F. O. Clinical Essentials of Calcium and Skeletal Disorders. *Ann. Intern. Med.* **131**, 635 (1999).

58. Brown EM. Physiology of Calcium Homeostasis. in *The Parathyroids, Second Edition*.

59. Bruder, J., Guise TA, & Mundy GR. Mineral Metabolism. in *Endocrinology & Metabolism, Fourth Edition* vol. 4 (2001).

60. Fraser, W. D. & Milan, A. M. Vitamin D Assays: Past and Present Debates, Difficulties, and Developments. *Calcif. Tissue Int.* **92**, 118–127 (2013).

61. Chen, P. *et al.* Meta-analysis of vitamin D, calcium and the prevention of breast cancer. *Breast Cancer Res. Treat.* **121**, 469–477 (2010).

62. Mattila, C. *et al.* Serum 25-Hydroxyvitamin D Concentration and Subsequent Risk of Type 2 Diabetes. *Diabetes Care* **30**, 2569–2570 (2007).

63. Baseline Serum 25-Hydroxy Vitamin D Is Predictive of Future Glycemic Status and Insulin Resistance | Diabetes | American Diabetes Association. https://diabetesjournals.org/diabetes/article/57/10/2619/13355/Baseline-Serum-25-Hydroxy-Vitamin-D-Is-Predictive.

64. Pittas, A. G. *et al.* Vitamin D and Calcium Intake in Relation to Type 2 Diabetes in Women. *Diabetes Care* **29**, 650–656 (2006).

65. Brown, A. J. & Slatopolsky, E. Vitamin D analogs: Therapeutic applications and mechanisms for selectivity. *Mol. Aspects Med.* **29**, 433–452 (2008).

66. Nnoaham, K. E. & Clarke, A. Low serum vitamin D levels and tuberculosis: a systematic review and meta-analysis. *Int. J. Epidemiol.* **37**, 113–119 (2008).

67. Chowdhury, R. *et al.* Vitamin D and risk of cause specific death: systematic review and meta-analysis of observational cohort and randomised intervention studies. *BMJ* **348**, g1903 (2014).

68. Gaksch, M. *et al.* Vitamin D and mortality: Individual participant data meta-analysis of standardized 25-hydroxyvitamin D in 26916 individuals from a European consortium. *PLoS ONE* **12**, e0170791 (2017).

69. Spiro, A. & Buttriss, J. L. Vitamin D: An overview of vitamin D status and intake in Europe. *Nutr. Bull. Bnf* **39**, 322–350 (2014).

70. Bikle, D. D., Siiteri, P. K., Ryzen, E., Haddad, J. G. & Gee, E. Serum Protein Binding of 1,25-Dihydroxyvitamin D: A Reevaluation by Direct Measurement of Free Metabolite Levels. *J. Clin. Endocrinol. Metab.* **61**, 969–975 (1985).

71. Bikle, D. D. Vitamin D Metabolism, Mechanism of Action, and Clinical Applications. *Chem. Biol.* **21**, 319–329 (2014).

72. Plachot, J. J. *et al.* In vitro action of 1,25-dihydroxycholecalciferol and 24,25-dihydroxycholecalciferol on matrix organization and mineral distribution in rabbit growth plate. *Metab. Bone Dis. Relat. Res.* **4**, 135–142 (1982).

73. Petkovich, M. & Jones, G. CYP24A1 and kidney disease. *Curr. Opin. Nephrol. Hypertens.* **20**, 337–344 (2011).

74. Henry, H. L., Taylor, A. N. & Norman, A. W. Response of Chick Parathyroid Glands to the Vitamin D Metabolites, 1,25-Dihydroxycholecalciferol and 24,25-Dihydroxycholecalciferol. *J. Nutr.* **107**, 1918–1926 (1977).

75. Seo, E. G., Einhorn, T. A. & Norman, A. W. 24R,25-dihydroxyvitamin D3: an essential vitamin D3 metabolite for both normal bone integrity and healing of tibial fracture in chicks. *Endocrinology* **138**, 3864–3872 (1997).

76. Bosworth, C. *et al.* The serum 24,25-dihydroxyvitamin D concentration, a marker of vitamin D catabolism, is reduced in chronic kidney disease. *Kidney Int.* **82**, 693–700 (2012).

77. Maestro, M. A., Molnár, F., Mouriño, A. & Carlberg, C. Vitamin D receptor 2016: novel ligands and structural insights. *Expert Opin. Ther. Pat.* **26**, 1291–1306 (2016).

78. Plum, L. A. & DeLuca, H. F. Vitamin D, disease and therapeutic opportunities. *Nat. Rev. Drug Discov.* **9**, 941–955 (2010).

79. Sirajudeen, S., Shah, I. & Al Menhali, A. A Narrative Role of Vitamin D and Its Receptor: With Current Evidence on the Gastric Tissues. *Int. J. Mol. Sci.* **20**, 3832 (2019).

80. Makin, G., Lohnes, D., Byford, V., Ray, R. & Jones, G. Target cell metabolism of 1,25-dihydroxyvitamin D3 to calcitroic acid. Evidence for a pathway in kidney and bone involving 24-oxidation. *Biochem. J.* **262**, 173–180 (1989).

81. Bhattacharjee, S., Renganaath, K., Mehrotra, R. & Mehrotra, S. Combinatorial Control of Gene Expression. *BioMed Res. Int.* **2013**, 407263 (2013).

82. Uitterlinden, A. G., Fang, Y., Van Meurs, J. B. J., Pols, H. A. P. & Van Leeuwen, J. P. T. M. Genetics and biology of vitamin D receptor polymorphisms. *Gene* **338**, 143–156 (2004).

83. Huang, J. & Xie, Z.-F. Polymorphisms in the vitamin D receptor gene and multiple sclerosis risk: a meta-analysis of case-control studies. *J. Neurol. Sci.* **313**, 79–85 (2012).

84. Menegaz, D. *et al.* Vitamin D Receptor (VDR) Regulation of Voltage-Gated Chloride Channels by Ligands Preferring a VDR-Alternative Pocket (VDR-AP). *Mol. Endocrinol.* **25**, 1289–1300 (2011).

85. Hii, C. S. & Ferrante, A. The Non-Genomic Actions of Vitamin D. *Nutrients* **8**, 135 (2016).

86. Nemere, I. *et al.* Ribozyme knockdown functionally links a 1,25(OH)2D3 membrane binding protein (1,25D3-MARRS) and phosphate uptake in intestinal cells. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7392–7397 (2004).

87. Martineau, C. *et al.* Optimal bone fracture repair requires 24R,25-dihydroxyvitamin D3 and its effector molecule FAM57B2. *J. Clin. Invest.* **128**, 3546–3557.

88. Bouillon, R. Comparative analysis of nutritional guidelines for vitamin D. *Nat. Rev. Endocrinol.* **13**, 466–479 (2017).

89. Pilz, S. *et al.* Rationale and Plan for Vitamin D Food Fortification: A Review and Guidance Paper. *Front. Endocrinol.* **9**, 373 (2018).

90. US Preventive Services Task Force. Vitamin D, Calcium, or Combined Supplementation for the Primary Prevention of Fractures in Community-Dwelling Adults: US Preventive Services Task Force Recommendation Statement. *JAMA* **319**, 1592–1599 (2018).

91. Brouwer-Brolsma, E. M. *et al.* Vitamin D: do we get enough? *Osteoporos. Int.* **24**, 1567–1577 (2013).

92. Pilz, S. *et al.* Vitamin D: Current Guidelines and Future Outlook. *Anticancer Res.* **38**, 1145–1151 (2018).

93. Cashman, K. D. Vitamin D Requirements for the Future—Lessons Learned and Charting a Path Forward. *Nutrients* **10**, 533 (2018).

94. Autier, P., Boniol, M., Pizot, C. & Mullie, P. Vitamin D status and ill health: a systematic review. *Lancet Diabetes Endocrinol.* **2**, 76–89 (2014).

95. Swart, K. M. *et al.* Effects of vitamin D supplementation on markers for cardiovascular disease and type 2 diabetes: an individual participant data meta-analysis of randomized controlled trials. *Am. J. Clin. Nutr.* **107**, 1043–1053 (2018).

96. Autier, P. *et al.* Effect of vitamin D supplementation on non-skeletal disorders: a systematic review of meta-analyses and randomised trials. *Lancet Diabetes Endocrinol.* **5**, 986–1004 (2017).

97. Rejnmark, L. *et al.* Non-skeletal health effects of vitamin D supplementation: A systematic review on findings from meta-analyses summarizing trial data. *PLoS ONE* **12**, e0180512 (2017).

98. Bolland, M. J., Avenell, A. & Grey, A. Should adults take vitamin D supplements to prevent disease? *BMJ* **355**, i6201 (2016).

99. Ebeling, P. R. *et al.* MANAGEMENT OF ENDOCRINE DISEASE: Therapeutics of vitamin D. *Eur. J. Endocrinol.* **179**, R239–R259 (2018).

100. Holick, M. F. *et al.* Guidelines for Preventing and Treating Vitamin D Deficiency and Insufficiency Revisited. *J. Clin. Endocrinol. Metab.* **97**, 1153–1158 (2012).

101. Holick, M. F. *et al.* Evaluation, Treatment, and Prevention of Vitamin D Deficiency: an Endocrine Society Clinical Practice Guideline. *J. Clin. Endocrinol. Metab.* **96**, 1911–1930 (2011).

102. Ross, A. C. *et al.* The 2011 Report on Dietary Reference Intakes for Calcium and Vitamin D from the Institute of Medicine: What Clinicians Need to Know. *J. Clin. Endocrinol. Metab.* **96**, 53–58 (2011).

103. Rosen, C. J. *et al.* IOM Committee Members Respond to Endocrine Society Vitamin D Guideline. *J. Clin. Endocrinol. Metab.* **97**, 1146–1152 (2012).

104. Reid, I. R. Calcium and vitamin D: To supplement or not? *Cleve. Clin. J. Med.* **85**, 693–698 (2018).

105. Pludowski, P. *et al.* Vitamin D supplementation guidelines. *J. Steroid Biochem. Mol. Biol.* **175**, 125–135 (2018).

106. Fogelholm, M. New Nordic Nutrition Recommendations are here. *Food Nutr. Res.* **57**, (2013).

107. Hyppönen, E. & Power, C. Hypovitaminosis D in British adults at age 45 y: nationwide cohort study of dietary and lifestyle predictors. *Am. J. Clin. Nutr.* **85**, 860–868 (2007).

108. Vitamin D Standardization Program (VDSP). https://ods.od.nih.gov/Research/vdsp.aspx.

109. Jones, G. Pharmacokinetics of vitamin D toxicity. *Am. J. Clin. Nutr.* **88**, 582S-586S (2008).

110. Brett, N. R., Gharibeh, N. & Weiler, H. A. Effect of Vitamin D Supplementation, Food Fortification, or Bolus Injection on Vitamin D Status in Children Aged 2-18 Years: A Meta-Analysis. *Adv. Nutr. Bethesda Md* **9**, 454–464 (2018).

111. Vieth, R. Vitamin D supplementation, 25-hydroxyvitamin D concentrations, and safety. *Am. J. Clin. Nutr.* **69**, 842–856 (1999).

112. Klingenberg, M. Pigments of rat liver microsomes. *Arch. Biochem. Biophys.* **75**, 376–386 (1958).

113. Cooper, D. Y., Estabrook, R. W. & Rosenthal, O. The stoichiometry of C21 hydroxylation of steroids by adrenocortical microsomes. *J. Biol. Chem.* **238**, 1320–1323 (1963).

114. Cooper, D. Y., Levin, S., Narasimhulu, S. & Rosenthal, O. PHOTOCHEMICAL ACTION SPECTRUM OF THE TERMINAL OXIDASE OF MIXED FUNCTION OXIDASE SYSTEMS. *Science* **147**, 400–402 (1965).

115. Ingelman-Sundberg, M. The human genome project and novel aspects of cytochrome P450 research. *Toxicol. Appl. Pharmacol.* **207**, 52–56 (2005).

116. Nebert, D. W., Wikvall, K. & Miller, W. L. Human cytochromes P450 in health and disease. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, (2013).

117. Engstrom, G. W., Reinhardt, T. A. & Horst, R. L. 25-Hydroxyvitamin D3-23-hydroxylase, a renal enzyme in several animal species. *Arch. Biochem. Biophys.* **250**, 86–93 (1986).

118. Pedersen, J. I., Hagenfeldt, Y. & Björkhem, I. Assay and properties of 25-hydroxyvitamin D3 23-hydroxylase. Evidence that 23,25-dihydroxyvitamin D3 is a major metabolite in 1,25-dihydroxyvitamin D3-treated or fasted guinea pigs. *Biochem. J.* **250**, 527–532 (1988).

119. Simboli-Campbell, M. & Jones, G. Dietary Phosphate Deprivation Increases Renal Synthesis and Decreases Renal Catabolism of 1,25-Dihydroxycholecalciferol in Guinea Pigs. *J. Nutr.* **121**, 1635–1642 (1991).

120. St-Arnaud, R. Targeted inactivation of vitamin D hydroxylases in mice. *Bone* **25**, 127–129 (1999).

121. Hypercalcemia - Endocrine and Metabolic Disorders. *MSD Manual Professional Edition* https://www.msdmanuals.com/en-gb/professional/endocrine-and-metabolic-disorders/electrolyte-disorders/hypercalcemia.

122. Lightwood, R. & Stapleton, T. Idiopathic hypercalcaemia in infants. *Lancet Lond. Engl.* **265**, 255–256 (1953).

123. Rhaney, K. & Mitchell, R. G. IDIOPATHIC HYPERCALCÆMIA OF INFANTS. *The Lancet* **267**, 1028–1033 (1956).

124. Group, B. M. J. P. Hypercalcaemia in Infants and Vitamin D. *Br Med J* **2**, 149–149 (1956).

125. Molin, A. *et al. CYP24A1* Mutations in a Cohort of Hypercalcemic Patients: Evidence for a Recessive Trait. *J. Clin. Endocrinol. Metab.* **100**, 1343–1352 (2015).

126. Nesterova, G. *et al.* 1,25-(OH)2D-24 Hydroxylase (CYP24A1) Deficiency as a Cause of Nephrolithiasis. *Clin. J. Am. Soc. Nephrol.* **8**, 649–657 (2013).

127. Schlingmann, K. P. *et al.* Mutations in CYP24A1 and Idiopathic Infantile Hypercalcemia. *N. Engl. J. Med.* **365**, 410–421 (2011).

128. Schoenmakers, I. *et al.* Gestational hypercalcemia: Prevalence and biochemical profile. *J. Steroid Biochem. Mol. Biol.* **199**, 105611 (2020).

129. Pilz, S. *et al.* Hypercalcemia in Pregnancy Due to CYP24A1 Mutations: Case Report and Review of the Literature. *Nutrients* **14**, 2518 (2022).

130. Masuda, S. *et al.* Altered Pharmacokinetics of 1α,25-Dihydroxyvitamin D3 and 25-Hydroxyvitamin D3 in the Blood and Tissues of the 25-Hydroxyvitamin D-24-Hydroxylase (Cyp24a1) Null Mouse. *Endocrinology* **146**, 825–834 (2005).

131. St-Arnaud, R. *et al.* Deficient Mineralization of Intramembranous Bone in Vitamin D-24-Hydroxylase-Ablated Mice Is Due to Elevated 1,25-Dihydroxyvitamin D and Not to the Absence of 24,25-Dihydroxyvitamin D*. *Endocrinology* **141**, 2658–2666 (2000).

132. Figueres, M.-L. *et al.* Kidney Function and Influence of Sunlight Exposure in Patients With Impaired 24-Hydroxylation of Vitamin D Due to CYP24A1 Mutations. *Am. J. Kidney Dis.* **65**, 122–126 (2015).

133. Jacobs, T. P. *et al.* A Lifetime of Hypercalcemia and Hypercalciuria, Finally Explained. *J. Clin. Endocrinol. Metab.* **99**, 708–712 (2014).

134. Dinour, D. *et al.* Maternal and infantile hypercalcemia caused by vitamin-D-hydroxylase mutations and vitamin D intake. *Pediatr. Nephrol.* **30**, 145–152 (2015).

135. Colussi, G. *et al.* Chronic hypercalcaemia from inactivating mutations of vitamin D 24-hydroxylase (CYP24A1): implications for mineral metabolism changes in chronic renal failure. *Nephrol. Dial. Transplant.* **29**, 636–643 (2014).

136. Schlingmann, K. P., Cassar, W. & Konrad, M. Juvenile onset IIH and CYP24A1 mutations. *Bone Rep.* **9**, 42–46 (2018).

137. McTaggart, S. J., Craig, J., MacMillan, J. & Burke, J. R. Familial occurrence of idiopathic infantile hypercalcemia. *Pediatr. Nephrol.* **13**, 668–671 (1999).

138. Nguyen, M. *et al.* Infantile Hypercalcemia and Hypercalciuria: New Insights into a Vitamin D-Dependent Mechanism and Response to Ketoconazole Treatment. *J. Pediatr.* **157**, 296–302 (2010).

139. Shah, A. D. *et al.* Maternal Hypercalcemia Due to Failure of 1,25-Dihydroxyvitamin-D3 Catabolism in a Patient With CYP24A1 Mutations. *J. Clin. Endocrinol. Metab.* **100**, 2832–2836 (2015).

140. Jobst-Schwan, T. *et al.* Discordant Clinical Course of Vitamin-D-Hydroxylase (CYP24A1) Associated Hypercalcemia in Two Adult Brothers With Nephrocalcinosis. *Kidney Blood Press. Res.* **40**, 443–451 (2015).

141. Sayers, J., Hynes, A., Rice, S., Hogg, P. & Sayer, J. Searching for CYP24A1 mutations in cohorts of patients with calcium nephrolithiasis. *OA Nephrol.* **1**, (2013).

142. Dinour, D. *et al.* Loss-of-Function Mutations of CYP24A1, the Vitamin D 24-Hydroxylase Gene, Cause Long-standing Hypercalciuric Nephrolithiasis and Nephrocalcinosis. *J. Urol.* **190**, 552–557 (2013).

143. Meusburger, E. *et al.* Medullary nephrocalcinosis in an adult patient with idiopathic infantile hypercalcaemia and a novel CYP24A1 mutation. *Clin. Kidney J.* **6**, 211–215 (2013).

144. Castanet, M., Mallet, E. & Kottler, M.-L. Lightwood Syndrome Revisited with a Novel Mutation in CYP24 and Vitamin D Supplement Recommendations. *J. Pediatr.* **163**, 1208–1210 (2013).

145. O'Keeffe, D. T. *et al.* Clinical and biochemical phenotypes of adults with monoallelic and biallelic CYP24A1 mutations: evidence of gene dose effect. *Osteoporos. Int.* **27**, 3121–3125 (2016).

146. Schlingmann, K. P. *et al.* Autosomal-Recessive Mutations in SLC34A1 Encoding Sodium-Phosphate Cotransporter 2A Cause Idiopathic Infantile Hypercalcemia. *J. Am. Soc. Nephrol. JASN* **27**, 604–614 (2016).

147. Tebben, P. J. *et al.* Hypercalcemia, hypercalciuria, and elevated calcitriol concentrations with autosomal dominant transmission due to CYP24A1 mutations: effects of ketoconazole therapy. *J. Clin. Endocrinol. Metab.* **97**, E423-427 (2012).

148. Dauber, A. *et al.* Genetic Defect in CYP24A1, the Vitamin D 24-Hydroxylase Gene, in a Patient with Severe Infantile Hypercalcemia. *J. Clin. Endocrinol. Metab.* **97**, 268–274 (2012).

149. Davidson Peiris, E. & Wusirika, R. A Case Report of Compound Heterozygous CYP24A1 Mutations Leading to Nephrolithiasis Successfully Treated with Ketoconazole. *Case Rep. Nephrol. Dial.* **7**, 167–171 (2017).

150. Tang, J. C. Y. *et al.* Reference intervals for serum 24,25-dihydroxyvitamin D and the ratio with 25-hydroxyvitamin D established using a newly developed LC–MS/MS method. *J. Nutr. Biochem.* **46**, 21–29 (2017).

151. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).

152. Strachan, T. & Read, A. *Human Molecular Genetics*. (Garland Science, Taylor & Francis Group, 2012).

153. Adema, G. J. & Baas, P. D. A novel calcitonin-encoding mRNA is produced by alternative processing of calcitonin/calcitonin gene-related peptide-I pre-mRNA. *J. Biol. Chem.* **267**, 7943–7948 (1992).

154. Corbett, A. H. Post-transcriptional Regulation of Gene Expression and Human Disease. *Curr. Opin. Cell Biol.* **52**, 96–104 (2018).

155. Brief Introduction of Post-Translational Modifications (PTMs) – Creative Proteomics Blog. https://www.creative-proteomics.com/blog/index.php/brief-introduction-of-post-translational-modifications-ptms/.

156. Miyamura, Y. *et al.* Mutations of the RNA-Specific Adenosine Deaminase Gene (DSRAD) Are Involved in Dyschromatosis Symmetrica Hereditaria. *Am. J. Hum. Genet.* **73**, 693–699 (2003).

157. Bykhovskaya, Y., Casas, K., Mengesha, E., Inbal, A. & Fischel-Ghodsian, N. Missense Mutation in Pseudouridine Synthase 1 (PUS1) Causes Mitochondrial Myopathy and Sideroblastic Anemia (MLASA). *Am. J. Hum. Genet.* **74**, 1303–1308 (2004).

158. Rice, G. I. *et al.* Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nat. Genet.* **44**, 1243–1248 (2012).

159. Destefanis, E. *et al.* A mark of disease: how mRNA modifications shape genetic and acquired pathologies. *RNA* **27**, 367–389 (2021).

160. Gibbs, R. A. *et al.* The International HapMap Project. *Nature* **426**, 789–796 (2003).

161. Conne, B., Stutz, A. & Vassalli, J.-D. The 3′ untranslated region of messenger RNA: A molecular 'hotspot' for pathology? *Nat. Med.* **6**, 637–641 (2000).

162. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

163. Bicknell, A. A., Cenik, C., Chua, H. N., Roth, F. P. & Moore, M. J. Introns in UTRs: why we should stop ignoring them. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **34**, 1025–1034 (2012).

164. Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5473–5478 (2015).

165. Human Genetic Disorders: Studying Single-Gene (Mendelian) Diseases | Learn Science at Scitable. http://www.nature.com/scitable/topicpage/rare-genetic-disorders-learning-about-genetic-disease-979.

166. Collins, A. L., Lunt, P. W., Garrett, C. & Dennis, N. R. Holoprosencephaly: a family showing dominant inheritance and variable expression. *J. Med. Genet.* **30**, 36–40 (1993).

167. Prasun, P., Pradhan, M. & Agarwal, S. One gene, many phenotypes. *J. Postgrad. Med.* **53**, 257 (2007).

168. Valkov, N. & Das, S. Y RNAs: Biogenesis, Function and Implications for the Cardiovascular System. *Adv. Exp. Med. Biol.* **1229**, 327–342 (2020).

169. Determination of in vivo RNA structure in low-abundance transcripts | Nature Communications. https://www.nature.com/articles/ncomms3971.

170. Bernat, V. & Disney, M. D. RNA structures as mediators of neurological diseases and as drug targets. *Neuron* **87**, 28–46 (2015).

171. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and Disease. *Cell* **136**, 777–793 (2009).

172. Nadeau, J. H. Modifier genes in mice and humans. *Nat. Rev. Genet.* **2**, 165–174 (2001).

173. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* **19**, 158–174 (2018).

174. Wang, J., Pitarque, M. & Ingelman-Sundberg, M. 3'-UTR polymorphism in the human CYP2A6 gene affects mRNA stability and enzyme expression. *Biochem. Biophys. Res. Commun.* **340**, 491–497 (2006).

175. Beffagna, G. *et al.* Regulatory mutations in transforming growth factor-β3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. *Cardiovasc. Res.* **65**, 366–373 (2005).

176. Reamon-Buettner, S. M., Cho, S.-H. & Borlak, J. Mutations in the 3'-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD). *BMC Med. Genet.* **8**, 38 (2007).

177. Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).

178. Deng, H. *et al.* Rice In Vivo RNA Structurome Reveals RNA Secondary Structure Conservation and Divergence in Plants. *Mol. Plant* **11**, 607–622 (2018).

179. Chakravarti, A. . . .to a future of genetic medicine. *Nature* **409**, 822–823 (2001).

180. Shatoff, E. & Bundschuh, R. Single nucleotide polymorphisms affect RNA-protein interactions at a distance through modulation of RNA secondary structures. *PLoS Comput. Biol.* **16**, e1007852 (2020).

181. Bangor University. *Vitamin D Status and in Vivo Immune Function*. https://clinicaltrials.gov/ct2/show/NCT02416895 (2017).

182. Tang, J. Methods for the Measurement of Vitamin D Metabolites and Studies on Their Relationships in Health and Disease. (University of East Anglia, 2019).

183. QuShape – The Weeks Lab. https://weekslab.com/software/qushape/.

184. RS-FISH documentation. (2022).

185. Bolland, M. J., Grey, A. & Avenell, A. Effects of vitamin D supplementation on musculoskeletal health: a systematic review, meta-analysis, and trial sequential analysis. *Lancet Diabetes Endocrinol.* **6**, 847–858 (2018).

186. Martineau, A. R. *et al.* Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data. *BMJ* **356**, i6583 (2017).

187. Khaw, K.-T. *et al.* Effect of monthly high-dose vitamin D supplementation on falls and non-vertebral fractures: secondary and post-hoc outcomes from the randomised, double-blind, placebo-controlled ViDA trial. *Lancet Diabetes Endocrinol.* **5**, 438–447 (2017).

188. Francis, R. M. *et al.* National Osteoporosis Society practical clinical guideline on vitamin D and bone health. *Maturitas* **80**, 119–121 (2015).

189. Levin, A. *et al.* Prevalence of abnormal serum vitamin D, PTH, calcium, and phosphorus in patients with chronic kidney disease: Results of the study to evaluate early kidney disease. *Kidney Int.* **71**, 31–38 (2007).

190. Kaufmann, M. *et al.* Clinical utility of simultaneous quantitation of 25-hydroxyvitamin D and 24,25-dihydroxyvitamin D by LC-MS/MS involving derivatization with DMEQ-TAD. *J. Clin. Endocrinol. Metab.* **99**, 2567–2574 (2014).

191. Binkley, N. *et al.* Can vitamin D metabolite measurements facilitate a "treat-to-target" paradigm to guide vitamin D supplementation? *Osteoporos. Int.* **26**, 1655–1660 (2015).

192. Tang, J. C. Y., Jackson, S., Walsh, N. P., Greeves, J. & Fraser, W. D. The dynamic relationships between the active and catabolic vitamin D metabolites, their ratios, and associations with PTH. *Sci. Rep.* **9**, 6974 (2019).

193. Armbrecht, H. J., Boltz, M. A. & Hodam, T. L. PTH increases renal 25(OH)D3-1alpha - hydroxylase (CYP1alpha) mRNA but not renal 1,25(OH)2D3 production in adult rats. *Am. J. Physiol. Renal Physiol.* **284**, F1032-1036 (2003).

194. Collins, M. T. *et al.* Fibroblast growth factor-23 is regulated by 1alpha,25-dihydroxyvitamin D. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **20**, 1944–1950 (2005).

195. Jacobson, N. S. & Truax, P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J. Consult. Clin. Psychol.* **59**, 12–19 (1991).

196. Jacobson, N. S., Roberts, L. J., Berns, S. B. & McGlinchey, J. B. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J. Consult. Clin. Psychol.* **67**, 300–307 (1999).

197. Macdonald, H. M. *et al.* Longevity of daily oral vitamin D3 supplementation: differences in 25OHD and 24,25(OH)2D observed 2 years after cessation of a 1-year randomised controlled trial (VICtORy RECALL). *Osteoporos. Int. J. Establ. Result Coop. Eur. Found. Osteoporos. Natl. Osteoporos. Found. USA* **28**, 3361–3372 (2017).

198. Macdonald, H. M. *et al.* 25-Hydroxyvitamin D Threshold for the Effects of Vitamin D Supplements on Bone Density: Secondary Analysis of a Randomized Controlled Trial. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **33**, 1464–1469 (2018).

199. Ketha, H., Kumar, R. & Singh, R. J. LC-MS/MS for Identifying Patients with CYP24A1 Mutations. *Clin. Chem.* **62**, 236–242 (2016).

200. Chertow, B. S., Baker, G. R., Henry, H. L. & Norman, A. W. Effects of vitamin D metabolites on bovine parathyroid hormone release in vitro. *Am. J. Physiol.-Endocrinol. Metab.* **238**, E384–E388 (1980).

201. Boyan, B. D. *et al.* 24R,25-Dihydroxyvitamin D3 Protects against Articular Cartilage Damage following Anterior Cruciate Ligament Transection in Male Rats. *PLOS ONE* **11**, e0161782 (2016).

202. Gal-Moscovici, A., Gal, M. & Popovtzer, M. M. Treatment of osteoporotic ovariectomized rats with 24,25(OH)2D3. *Eur. J. Clin. Invest.* **35**, 375–379 (2005).

203. St-Arnaud, R. CYP24A1-deficient mice as a tool to uncover a biological activity for vitamin D metabolites hydroxylated at position 24. *J. Steroid Biochem. Mol. Biol.* **121**, 254–256 (2010).

204. Tripkovic, L. *et al.* Comparison of vitamin D2 and vitamin D3 supplementation in raising serum 25-hydroxyvitamin D status: a systematic review and meta-analysis. *Am. J. Clin. Nutr.* **95**, 1357–1364 (2012).

205. Arabi, A., Baddoura, R., El-Rassi, R. & El-Hajj Fuleihan, G. Age but not gender modulates the relationship between PTH and vitamin D. *Bone* **47**, 408–412 (2010).

206. Kaufmann, M. *et al.* Improved Screening Test for Idiopathic Infantile Hypercalcemia Confirms Residual Levels of Serum 24,25-(OH)2D3 in Affected Patients. *J. Bone Miner. Res.* **32**, 1589–1596 (2017).

207. Herrmann, M. Towards a personalized assessment of vitamin D status. *Clin. Chem. Lab. Med. CCLM* **58**, 149–151 (2020).

208. Kaufmann, M. *et al.* Differential diagnosis of vitamin D–related hypercalcemia using serum vitamin D metabolite profiling. *J. Bone Miner. Res.* **36**, 1340–1350 (2021).

209. Goh, G. & Choi, M. Application of Whole Exome Sequencing to Identify Disease-Causing Variants in Inherited Human Diseases. *Genomics Inform.* **10**, 214–219 (2012).

210. Seaby, E. G., Pengelly, R. J. & Ennis, S. Exome sequencing explained: a practical guide to its clinical application. *Brief. Funct. Genomics* **15**, 374–384 (2016).

211. VCV000295559.6 - ClinVar - NCBI. https://www.ncbi.nlm.nih.gov/clinvar/variation/295559/?new_evidence=false.

212. VCV000342809.3 - ClinVar - NCBI. https://www.ncbi.nlm.nih.gov/clinvar/variation/342809/?new_evidence=false.

213. VCV000008349.23 - ClinVar - NCBI. https://www.ncbi.nlm.nih.gov/clinvar/variation/8349/?new_evidence=true.

214. rs1058885 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs1058885.

215. rs333970 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs333970.

216. rs2229166 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs2229166.

217. rs67114147 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs67114147.

218. rs759686664 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs759686664.

219. rs3088027 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs3088027.

220. rs62620995 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs62620995.

221. VCV001548855.3 - ClinVar - NCBI.

https://www.ncbi.nlm.nih.gov/clinvar/variation/1548855/?new_evidence=false.

222. rs10252250 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs10252250.

223. rs178710 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs178710.

224. rs771198810 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs771198810.

225. rs5742915 RefSNP Report - dbSNP - NCBI.

https://www.ncbi.nlm.nih.gov/snp/rs5742915.

226. VCV000774496.3 - ClinVar - NCBI.

https://www.ncbi.nlm.nih.gov/clinvar/variation/774496/?new_evidence=false.

227. rs3814830 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs3814830.

228. Vallet, M. *et al.* Targeted sequencing of the Paget's disease associated 14q32 locus identifies several missense coding variants in RIN3 that predispose to Paget's disease of bone. *Hum. Mol. Genet.* **24**, 3286–3295 (2015).

229. rs117068593 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs117068593.

230. VCV000586682.7 - ClinVar - NCBI. https://www.ncbi.nlm.nih.gov/clinvar/variation/586682/?new_evidence=false.

231. rs1911573006 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs1911573006.

232. rs731236 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs731236.

233. rs1780316 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs1780316.

234. rs2036400 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs2036400.

235. rs1801726 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs1801726.

236. rs4677948 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs4677948.

237. rs333971 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs333971.

238. rs12794714 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs12794714.

239. rs11550518 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs11550518.

240. rs7810767 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs7810767.

241. rs5951494 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs5951494.

242. rs3742716 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs3742716.

243. rs4935 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs4935.

244. rs10277 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs10277.

245. rs1065154 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs1065154.

246. rs8092336 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs8092336.

247. Open Targets Platform. https://platform.opentargets.org/.

248. Bacher, U. *et al.* Challenges in the introduction of next-generation sequencing (NGS) for diagnostics of myeloid malignancies into clinical routine use. *Blood Cancer J.* **8**, 1–10 (2018).

249. Salk, J. J., Schmitt, M. W. & Loeb, L. A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat. Rev. Genet.* **19**, 269–285 (2018).

250. Jennings, L. J. *et al.* Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J. Mol. Diagn.* **19**, 341–365 (2017).

251. Scillitani, A. *et al.* Primary Hyperparathyroidism and the Presence of Kidney Stones Are Associated with Different Haplotypes of the Calcium-Sensing Receptor. *J. Clin. Endocrinol. Metab.* **92**, 277–283 (2007).

252. Vezzoli, G. *et al.* R990G polymorphism of calcium-sensing receptor does produce a gain-of-function and predispose to primary hypercalciuria. *Kidney Int.* **71**, 1155–1162 (2007).

253. Wang, Z., Liu, X., Yang, B.-Z. & Gelernter, J. The Role and Challenges of Exome Sequencing in Studies of Human Diseases. *Front. Genet.* **4**, (2013).

254. Linderman, M. D. *et al.* Analytical validation of whole exome and whole genome sequencing for clinical applications. *BMC Med. Genomics* **7**, 20 (2014).

255. Strom, S. P. *et al.* Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **16**, 510–515 (2014).

256. Yi, M. *et al.* Performance comparison of SNP detection tools with illumina exome sequencing data--an assessment using both family pedigree information and sample-matched SNP array data. *Nucleic Acids Res.* **42**, e101 (2014).

257. Miles, B. & Tadi, P. Genetics, Somatic Mutation. in *StatPearls* (StatPearls Publishing, 2022).

258. rs1060502854 RefSNP Report - dbSNP - NCBI. https://www.ncbi.nlm.nih.gov/snp/rs1060502854.

259. Petrackova, A. *et al.* Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics. *Front. Oncol.* **9**, (2019).

260. Díaz-de Usera, A. *et al.* Evaluation of Whole-Exome Enrichment Solutions: Lessons from the High-End of the Short-Read Sequencing Scale. *J. Clin. Med.* **9**, 3656 (2020).

261. Browne, P. D. *et al.* GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms. *GigaScience* **9**, giaa008 (2020).

262. Guo, Y. *et al.* Exome sequencing generates high quality data in non-target regions. *BMC Genomics* **13**, 194 (2012).

263. Tiwari, A. *et al.* Identification of Novel and Recurrent Disease-Causing Mutations in Retinal Dystrophies Using Whole Exome Sequencing (WES): Benefits and Limitations. *PLOS ONE* **11**, e0158692 (2016).

264. Yan, Y. H. *et al.* Confirming putative variants at ≤ 5% allele frequency using allele enrichment and Sanger sequencing. *Sci. Rep.* **11**, 11640 (2021).

265. Kim, Y.-K. RNA therapy: rich history, various applications and unlimited future prospects. *Exp. Mol. Med.* **54**, 455–465 (2022).

266. Garabedian, M., Holick, M. F., Deluca, H. F. & Boyle, I. T. Control of 25-Hydroxycholecalciferol Metabolism by Parathyroid Glands. *Proc. Natl. Acad. Sci.* **69**, 1673–1676 (1972).

267. DeLuca, H. F. Parathyroid Hormone as a Trophic Hormone for 1,25-Dihydroxyvitamin D3, the Metabolically Active Form of Vitamin D. *N. Engl. J. Med.* **287**, 250–251 (1972).

268. Tanaka, Y. & DeLuca, H. F. Rat renal 25-hydroxyvitamin D3 1- and 24-hydroxylases: their in vivo regulation. *Am. J. Physiol.-Endocrinol. Metab.* **246**, E168–E173 (1984).

269. Shimada, T. *et al.* Targeted ablation of Fgf23 demonstrates an essential physiological role of FGF23 in phosphate and vitamin D metabolism. *J. Clin. Invest.* **113**, 561–568 (2004).

270. Shimada, T. *et al.* FGF-23 transgenic mice demonstrate hypophosphatemic rickets with reduced expression of sodium phosphate cotransporter type IIa. *Biochem. Biophys. Res. Commun.* **314**, 409–414 (2004).

271. Sharp, P. A. The Centrality of RNA. *Cell* **136**, 577–580 (2009).

272. Pan, T. & Sosnick, T. Rna Folding During Transcription. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 161–175 (2006).

273. Warf, M. B. & Berglund, J. A. The role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* **35**, 169–178 (2010).

274. Martin, K. C. & Ephrussi, A. mRNA Localization: Gene Expression in the Spatial Dimension. *Cell* **136**, 719–730 (2009).

275. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13–37 (2005).

276. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **8**, 113–126 (2007).

277. Schlingmann, K. P., Cassar, W. & Konrad, M. Juvenile onset IIH and CYP24A1 mutations. *Bone Rep.* **9**, 42–46 (2018).

278. Griffin, T. P. *et al.* Biallelic CYP24A1 variants presenting in pregnancy: clinical/biochemical phenotypes. *Endocr. Connect.* **1**, (2020).

279. Hedberg, F., Pilo, C., Wikner, J., Törring, O. & Calissendorff, J. Three Sisters With Heterozygous Gene Variants of CYP24A1: Maternal Hypercalcemia, New-Onset Hypertension, and Neonatal Hypoglycemia. *J. Endocr. Soc.* **3**, 387–396 (2018).

280. Kwok, C. K., Ding, Y., Tang, Y., Assmann, S. M. & Bevilacqua, P. C. Determination of in vivo RNA structure in low-abundance transcripts. *Nat. Commun.* **4**, 2971 (2013).

281. RNAfold web server. http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi.

282. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

283. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

284. Wong, L. *et al.* Limits of Peripheral Blood Mononuclear Cells for Gene Expression-Based Biomarkers in Juvenile Idiopathic Arthritis. *Sci. Rep.* **6**, 29477 (2016).

285. Chapter 9. Intracellular concentrations of immunosuppressants | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/B9780128008850000096?token=E1D125258388FC97BFB6A3236884526E3983E0F4B8ACF18D95BCF76835B2EADA99B4F26221E0D4718FBB01293DE094BD&originRegion=eu-west-1&originCreation=20230103124645 doi:10.1016/B978-0-12-800885-0.00009-6.

286. Maes, E., Cools, N., Willems, H. & Baggerman, G. FACS-Based Proteomics Enables Profiling of Proteins in Rare Cell Populations. *Int. J. Mol. Sci.* **21**, 6557 (2020).

287. Monosomes actively translate synaptic mRNAs in neuronal processes. https://www.science.org/doi/10.1126/science.aay4991 doi:10.1126/science.aay4991.

288. Vendramin, R. *et al.* SAMMSON fosters cancer cell fitness by concertedly enhancing mitochondrial and cytosolic translation. *Nat. Struct. Mol. Biol.* **25**, 1035–1046 (2018).

289. Tang, Y. *et al.* StructureFold: genome-wide RNA secondary structure mapping and reconstruction in vivo. *Bioinformatics* **31**, 2668–2675 (2015).

290. Ding, Y., Kwok, C. K., Tang, Y., Bevilacqua, P. C. & Assmann, S. M. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protoc.* **10**, 1050–1066 (2015).

291. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2014).

292. Tyrrell, J., McGinnis, J. L., Weeks, K. M. & Pielak, G. J. The Cellular Environment Stabilizes Adenine Riboswitch RNA Structure. *Biochemistry* **52**, 8777–8785 (2013).

293. Talkish, J., May, G., Lin, Y., Woolford, J. L. & McManus, C. J. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* **20**, 713–720 (2014).

294. Bevilacqua, P. C., Ritchey, L. E., Su, Z. & Assmann, S. M. Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.* **50**, 235–266 (2016).

295. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci.* **106**, 97–102 (2009).

296. Cordero, P., Kladwang, W., VanLang, C. C. & Das, R. Quantitative DMS mapping for automated RNA secondary structure inference. *Biochemistry* **51**, 7037–7039 (2012).

297. Jayaraman, D. & Kenyon, J. C. New windows into retroviral RNA structures. *Retrovirology* **15**, 11 (2018).

298. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Y. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* **12**, 641–655 (2011).

299. Ouyang, Z., Snyder, M. P. & Chang, H. Y. SeqFold: Genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.* **23**, 377–387 (2013).

300. Wu, Y. *et al.* Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Res.* **43**, 7247–7259 (2015).

301. Hajdin, C. E. *et al.* Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5498–5503 (2013).

302. Senecoff, J. F. In vivo analysis of plant RNA structure: soybean 18S ribosomal and ribulose-l,5-bisphosphate carboxylase small subunit RNAs. 16.

303. Zaug, A. J. & Cech, T. R. Analysis of the structure of Tetrahymena nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA* **1**, 363–374 (1995).

304. PII: S8756-3282(99)00118-0 | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S8756328299001180?token=EEEFD9605F628090E9C1EC8A8C753BFFD64A0A4DD89FDBE0C2CAE8B12181D24B082719745038C3C0

D7CB7D53490FD6E8&originRegion=eu-west-1&originCreation=20220305211130 doi:10.1016/S8756-3282(99)00118-0.

305. Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9**, 18–20 (2013).

306. Ainger, K. *et al.* Transport and Localization Elements in Myelin Basic Protein mRNA. *J. Cell Biol.* **138**, 1077–1087 (1997).

307. An, J. J. *et al.* Distinct Role of Long 3'UTR BDNF mRNA in Spine Morphology and Synaptic Plasticity in Hippocampal Neurons. *Cell* **134**, 175–187 (2008).

308. Steri, M., Idda, M. L., Whalen, M. B. & Orrù, V. Genetic Variants in mRNA Untranslated Regions. *Wiley Interdiscip. Rev. RNA* **9**, 1474 (2018).

309. Jones, G., Prosser, D. E. & Kaufmann, M. 25-Hydroxyvitamin D-24-hydroxylase (CYP24A1): Its important role in the degradation of vitamin D. *Arch. Biochem. Biophys.* **523**, 9–18 (2012).

310. Fagerberg, L. *et al.* Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. *Mol. Cell. Proteomics MCP* **13**, 397–406 (2014).

311. Capecci, M., R., Evans, M., J. & Smithies, O. Gene Modification in Mice. *NobelPrize.org* https://www.nobelprize.org/prizes/medicine/2007/advanced-information/.

312. Making a Mightier Knockout Mouse. *Genome.gov* https://www.genome.gov/27532436/2009-news-feature-making-a-mightier-knockout-mouse.

313. Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).

314. Gupta, R. M. & Musunuru, K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J. Clin. Invest.* **124**, 4154–4161 (2014).

315. Wu, Y. *et al.* Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9. *Cell Stem Cell* **13**, 659–662 (2013).

316. Huang, N. *et al.* Induction of apoptosis in imatinib sensitive and resistant chronic myeloid leukemia cells by efficient disruption of bcr-abl oncogene with zinc finger nucleases. *J. Exp. Clin. Cancer Res. CR* **37**, 62 (2018).

317. Abrahimi, P. *et al.* Efficient gene disruption in cultured primary human endothelial cells by CRISPR/Cas9. *Circ. Res.* **117**, 121–128 (2015).

318. Bao, D. *et al.* Preliminary Characterization of a Leptin Receptor Knockout Rat Created by CRISPR/Cas9 System. *Sci. Rep.* **5**, 15942 (2015).

319. Wang, X. *et al.* Characterization of novel cytochrome P450 2E1 knockout rat model generated by CRISPR/Cas9. *Biochem. Pharmacol.* **105**, 80–90 (2016).

320. Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.* **25**, 1298–1306 (2007).

321. Flisikowska, T. *et al.* Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. *PloS One* **6**, e21045 (2011).

322. Ebina, H., Misawa, N., Kanemura, Y. & Koyanagi, Y. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci. Rep.* **3**, 2510 (2013).

323. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).

324. Hu, W. *et al.* RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11461–11466 (2014).

325. Homo sapiens cytochrome P450 family 24 subfamily A member 1 (CYP24A1), transcript variant 1, mRNA; nuclear gene for mitochondrial product. (2022).

326. Höbaus, J. *et al.* Impact of CYP24A1 overexpression on growth of colorectal tumour xenografts in mice fed with vitamin D and soy. *Int. J. Cancer* **138**, 440–450 (2016).

327. Susa, T., Iizuka, M., Okinaga, H., Tamamori-Adachi, M. & Okazaki, T. Without 1α-hydroxylation, the gene expression profile of 25(OH)D3 treatment overlaps deeply with that of 1,25(OH)2D3 in prostate cancer cells. *Sci. Rep.* **8**, 9024 (2018).

328. Chun, R. F. *et al.* Vitamin D-Binding Protein Directs Monocyte Responses to 25-Hydroxy- and 1,25-Dihydroxyvitamin D. *J. Clin. Endocrinol. Metab.* **95**, 3368–3376 (2010).

329. Berkowska, K. *et al.* Investigating the Role of VDR and Megalin in Semi-Selectivity of Side-Chain Modified 19-nor Analogs of Vitamin D. *Int. J. Mol. Sci.* **20**, E4183 (2019).

330. Wu, S., Ren, S., Nguyen, L., Adams, J. S. & Hewison, M. Splice variants of the CYP27b1 gene and the regulation of 1,25-dihydroxyvitamin D3 production. *Endocrinology* **148**, 3410–3418 (2007).

331. Zelzer, E. & Olsen, B. R. The genetic basis for skeletal diseases. *Nature* **423**, 343–348 (2003).

332. Ohyama, Y. *et al.* Structural characterization of the gene encoding rat 25-hydroxyvitamin D3 24-hydroxylase. *Biochemistry* **32**, 76–82 (1993).

333. Reinhardt, T. A. & Horst, R. L. Ketoconazole inhibits self-induced metabolism of 1,25-dihydroxyvitamin D3 and amplifies 1,25-dihydroxyvitamin D3 receptor up-regulation in rat osteosarcoma cells. *Arch. Biochem. Biophys.* **272**, 459–465 (1989).

334. Cell line - CYP24A1 - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000019186-CYP24A1/cell+line#lung_cancer.

335. Naeem, M., Majeed, S., Hoque, M. Z. & Ahmad, I. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. *Cells* **9**, 1608 (2020).

336. Ensure Proper Controls in Your CRISPR Experiments. *Synthego* https://www.synthego.com/blog/controls-crispr-experiments.

337. Zittermann, A. *et al.* Measurement of Circulating 1,25-Dihydroxyvitamin D: Comparison of an Automated Method with a Liquid Chromatography Tandem Mass Spectrometry Method. *Int. J. Anal. Chem.* **2016**, e8501435 (2016).

338. Higgins, V. *et al.* Pediatric reference intervals for 1,25-dihydroxyvitamin D using the DiaSorin LIAISON XL assay in the healthy CALIPER cohort. *Clin. Chem. Lab. Med. CCLM* **56**, 964–972 (2018).

339. Spanaus, K. & von Eckardstein, A. Evaluation of two fully automated immunoassay based tests for the measurement of 1α,25-dihydroxyvitamin D in human serum and comparison with LC-MS/MS. *Clin. Chem. Lab. Med.* **55**, 1305–1314 (2017).

340. Biever, A. *et al.* Monosomes actively translate synaptic mRNAs in neuronal processes. *Science* **367**, 4991 (2020).

341. Lécuyer, E. *et al.* Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174–187 (2007).

342. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA Synthesis in Mammalian Cells. *PLOS Biol.* **4**, e309 (2006).

343. Duncan, S., Olsson, T. S. G., Hartley, M., Dean, C. & Rosa, S. A method for detecting single mRNA molecules in Arabidopsis thaliana. *Plant Methods* **12**, 13 (2016).

344. Neuert, G. *et al.* Systematic identification of signal-activated stochastic gene regulation. *Science* **339**, 584–587 (2013).

345. Yang, B. *et al.* Single-cell phenotyping within transparent intact tissue through whole-body clearing. *Cell* **158**, 945–958 (2014).

346. Ji, N. *et al.* Feedback Control of Gene Expression Variability in the Caenorhabditis elegans Wnt Pathway. *Cell* **155**, 869–880 (2013).

347. Gall, J. G. & Pardue, M. L. Formation and detection of rna-dna hybrid molecules in cytological preparations*. *Proc. Natl. Acad. Sci.* **63**, 378–383 (1969).

348. Rudkin, G. T. & Stollar, B. D. High resolution detection of DNA–RNA hybrids in situ by indirect immunofluorescence. *Nature* **265**, 472–473 (1977).

349. Bauman, J. G. J., Wiegant, J., Borst, P. & van Duijn, P. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Exp. Cell Res.* **128**, 485–490 (1980).

350. Singer, R. H. & Ward, D. C. Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. *Proc. Natl. Acad. Sci.* **79**, 7331–7335 (1982).

351. Pontes, O., Costa-Nunes, P., Vithayathil, P. & Pikaard, C. S. RNA polymerase V functions in Arabidopsis interphase heterochromatin organization independently of the 24-nt siRNA-directed DNA methylation pathway. *Mol. Plant* **2**, 700–710 (2009).

352. Pontvianne, F. *et al.* Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes Dev.* **27**, 1545–1550 (2013).

353. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).

354. Duncan, S., Olsson, T. S. G., Hartley, M., Dean, C. & Rosa, S. Single Molecule RNA FISH in Arabidopsis Root Cells. *Bio-Protoc.* **7**, e2240 (2017).

355. Perez-Carrasco, R., Beentjes, C. & Grima, R. Effects of cell cycle variability on lineage and population measurements of messenger RNA abundance. *J. R. Soc. Interface* **17**, 20200360 (2020).

356. Janes, K. A. Cell-to-cell transcript variability—seeing signal in the noise. *Cell* **163**, 1566–1568 (2015).

357. Hayles, B., Yellaboina, S. & Wang, D. Comparing Transcription Rate and mRNA Abundance as Parameters for Biochemical Pathway and Network Analysis. *PLOS ONE* **5**, e9908 (2010).

358. Wu, S. *et al.* Independent regulation of gene expression level and noise by histone modifications. *PLoS Comput. Biol.* **13**, e1005585 (2017).

359. Eastman, A. E. & Guo, S. The palette of techniques for cell cycle analysis. *FEBS Lett.* **594**, 2084–2098 (2020).

360. Ghazal, K., Brabant, S., Prie, D. & Piketty, M.-L. Hormone Immunoassay Interference: A 2021 Update. *Ann. Lab. Med.* **42**, 3–23 (2022).

361. Wagner, D. *et al.* The ratio of serum 24,25-dihydroxyvitamin D3 to 25-hydroxyvitamin D3 is predictive of 25-hydroxyvitamin D3 response to vitamin D3 supplementation. *J. Steroid Biochem. Mol. Biol.* **126**, 72–77 (2011).

362. Tai, S. S.-C. & Nelson, M. A. Candidate Reference Measurement Procedure for the Determination of (24R),25-Dihydroxyvitamin D3 in Human Serum Using Isotope-Dilution Liquid Chromatography–Tandem Mass Spectrometry. *Anal. Chem.* **87**, 7964–7970 (2015).

363. Paolis, E. D., Scaglione, G. L., Bonis, M. D., Minucci, A. & Capoluongo, E. CYP24A1 and SLC34A1 genetic defects associated with idiopathic infantile hypercalcemia: from genotype to phenotype. *Clin. Chem. Lab. Med. CCLM* **57**, 1650–1667 (2019).

364. Differential diagnosis of vitamin D–related hypercalcemia using serum vitamin D metabolite profiling - Kaufmann - 2021 - Journal of Bone and Mineral Research - Wiley Online Library. https://asbmr.onlinelibrary.wiley.com/doi/10.1002/jbmr.4306.

365. Lenherr-Taube, N. *et al.* Mild Idiopathic Infantile Hypercalcemia—Part 1: Biochemical and Genetic Findings. *J. Clin. Endocrinol. Metab.* **106**, 2915–2937 (2021).

366. Lenherr-Taube, N. *et al.* Mild Idiopathic Infantile Hypercalcemia—Part 2: A Longitudinal Observational Study. *J. Clin. Endocrinol. Metab.* **106**, 2938–2948 (2021).

367. Rampersad, S. & Tennant, P. Chapter 3 - Replication and Expression Strategies of Viruses. in *Viruses* 55–82 (Academic Press, 2018). doi:10.1016/B978-0-12-811257-1.00003-6.