Methods for analysis of sex-specific methylation in the CpG island of ALS-linked gene *UBQLN2*

Sonalani Simran Shandil

A thesis submitted in partial fulfilment of the requirements for

the degree of Master of Science in Biomedical Science

the University of Auckland, 2023.

Abstract

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disorder with a devastating clinical phenotype including the loss of motor control, muscle atrophy, and paralysis. While the aetiology of ALS is largely unknown, significant progress has been made in understanding its underlying mechanisms. The identification of novel mutations in *UBQLN2,* responsible for X-linked dominant ALS, has provided a new avenue of research into the role of *UBQLN2* in ALS pathology as well as sex-specific differences in disease manifestation. X-chromosome inactivation (XCI) is a form of gene dose compensation in females that has been found to provide some protection against *UBQLN2*-linked ALS in females, but mechanisms driving this protection require further investigation.

To address this research gap, this study presents an analysis of sex-specific methylation patterns across the CpG island of *UBQLN2,* and methods to validate these patterns in patient-derived tissue. Publicly available data revealed elevated methylation in females compared to males, suggesting that epigenetic modification attributed to XCI is involved in the sex-specific phenotypic variation of *UBQLN2*-linked ALS. Patient-derived dermal fibroblast cells were used to optimise pipelines for DNA methylation analysis, and despite challenges associated with bisulfite sequencing, successful conversion and Sanger sequencing was achieved for three *UBQLN2*-heterozygous female patients.

This study establishes a foundation for investigating female protection in X-linked diseases such as *UBQLN2*-linked ALS by assessing epigenetic differences associated with XCI.

Acknowledgements

I have stared at this blank page, with gratitude in my heart, wondering who to possibly thank first. My journey in completing my master's has been a testament to my determination and wouldn't be complete without the help of so many people.

I am deeply indebted to my supervisor, Dr. Emma Scotter, for her unwavering support, patience, expertise and guidance. Emma, your encouragement and the loving environment you've curated in the Scotter Lab has been instrumental in the shaping of not only this research, but also me as a scientist. I am extremely fortunate to have started my research career with a dedicated mentor that challenged me to my full potential.

Conducting the research for this thesis could not be possible without the members of the Scotter Lab family: David, Kyrah, Maize, Miran, Molly, Serey and honourable Scotter Lab member, Jaime. Words cannot express how extremely grateful I am to have met and experienced life with you all over the last year. Thank you for all the advice, laughs and drinks we have shared – it has been a pleasure. The shoulders you have extended in times of need have been an unwavering source of support, that has helped me navigate the challenges of life with greater resilience and courage.

I would like to extend my gratitude to the donors and their families, without whom this research would not be possible. To all the researchers and scholars whose work has been referenced in this thesis – their contributions to the field have been invaluable in shaping this research.

Finally, my family – while my resilience may waver, your support for me never does. Thank you for cushioning me with love when I need it the most.

I am humbled and thankful for the opportunity to have worked on this research.

Table of Contents

List of Figures

List of Tables

Abbreviations

AR	Androgen Receptor	**HUMARA**	Human androgen receptor assay
ALS	Amyotrophic Lateral Sclerosis	**LC3**	Light chain 3
C9orf72	Chromosome 9 open reading frame 72	**LLPS**	Liquid-liquid phase separation
CDK16	Cyclin Dependant Kinase 16	**MND**	Motor neuron disease
DAPK1	Death associated protein kinase 1	**NCBI**	National Centre for Biotechnology Information
DEPC	Diethyl pyrocarbonate	**PCR**	Polymerase chain reaction
DMD	Duchenne muscular dystrophy	**PolyGR**	Poly- glycine-arginine
DMEM	Dulbecco's Modified Eagle Medium	**qPCR**	Quantitative PCR
DRP	Dipeptide repeat proteins	**RNA-seq**	RNA-Sequencing
ERAD	Endoplasmic reticulum–associated protein degradation	**sALS**	Sporadic ALS
EWAS	Epigenome-wide association studies	**SOD1**	Superoxide dismutase 1
fALS	Familial ALS	**STI-1**	Stress-induced protein 1
FBS	Fetal bovine serum	**TAR**	Trans-activation response
FMR1	Fragile X messenger ribonucleoprotein 1	**TARDBP**	Trans-activation response DNA-binding protein
FMRP	Fragile X messenger ribonucleoprotein	**TCGA**	The Cancer Genome Atlas
FTD	Frontotemporal dementia	**TDP-43**	TAR DNA-binding protein 43
FUS	Fused in sarcoma	**UBA**	Ubiquitin-associated
FXS	Fragile X-syndrome	**UBL**	Ubiquitin-like
GEO	Gene Expression Omnibus	**UCSC**	University of California, Santa Cruz
HRE	Hexanucleotide repeat expansion	**UPS**	Ubiquitin proteasome system
Hsp70	Heat-shock protein 70	**XCI**	X-chromosome inactivation

1 INTRODUCTION

Motor neuron disease (MND) encompasses a range of debilitating movement disorders that significantly impacts quality of life. Characterised by the progressive loss of upper and lower motor neurons originating in the motor cortex, brainstem or anterior horn of the spinal cord, MND is commonly associated with dampened motor control, cognitive deficits, and behavioural changes[1]. Different classifications of MND are based on the differential involvement of upper and/or lower motor neurons[2]. Amyotrophic Lateral Sclerosis (ALS) is a form of MND characterised by progressive neurodegeneration that causes muscle weakness, atrophy, aphasia and difficulty breathing. Approximately 10% of ALS cases display a family history, classified as familial ALS (fALS) while the majority (90%) of ALS cases are sporadic in origin (sALS)[3].

Despite significant progress in understanding the underlying mechanisms of ALS pathogenesis aetiology remains unknown due to its complex nature[4]. ALS research mainly focuses on disease causing mutations in ALS-linked genes to elucidate aberrant pathways that cause disease progression. The discovery of novel mutations in *UBQLN2* located on the X-chromosome, and its contribution to X-linked dominant ALS[5], has prompted investigations into the role of *UBQLN2* in ALS pathology[6] and differences in disease manifestation between males and females[5,7]. Although studies have demonstrated that X-chromosome inactivation provides some protection against *UBQLN2*-linked ALS in females[5–7], further investigation of the mechanisms driving protection is necessary.

This review will outline the genetic contribution of ALS, specifically focusing on *UBQLN2*, the epigenetics of X-chromosome inactivation and the current methodologies and models used to understand X-chromosome inactivation in disease.

1.1 AMYOTROPHIC LATERAL SCLEROSIS

1.1.1 ALS EPIDEMIOLOGY

Amyotrophic lateral sclerosis (ALS) is the most common form of MND with an incidence of 0.6 to 3.8 per 100,000 people per year worldwide[8]; this incidence is also reflected in New Zealand where 2.8 in 100,000 people are diagnosed with ALS per year[9]. The mean age of ALS onset can vary from 59-63 years in sporadic ALS cases (sALS) and 40-60 years for familial ALS (fALS), with peak incidence in patients between 70-79 years of age[10]. The burden of disease is expected to increase with improved diagnostic criteria and reporting of the disease in addition to growth of the aging population[11–13]. Sporadic ALS is relatively more common than fALS, making up 90-95% of all ALS cases[14], early observations of familial motor pathologies in the Farr family by Sir William Osler in 1880[15] and subsequent cases of fALS have aided in the identification of more than 35 ALS causing genes present in both sALS and fALS[16]. While ALS is rare in patients under the age of 40, the risk of ALS can increase with the presence of genetic risk factors and exposure to certain lifestyle and environmental elements[11].

Geographical variability in ALS incidence is evidenced by population-based studies. A study conducted was Retchman et al[17] assessing racial and ethnic differences in ALS cases within the United States. In this study, of the 5,883 ALS cases identified 74.8% were white while African-American and Asian populations made up 9.3% and 3.6%, respectively, leading to the interpretation of a lower ALS incidence in non-white individuals[17]. When compared to Asian population-based studies, the reported incidence of ALS in European populations is higher, ranging from 2.1 to 3.8 people per 100,000 per year[8,18–20] relative to an incidence of 1.2 per 100,000 people per year in a South Korean study[21] and approximately 0.3-0.8 per 100,000 people per year in Beijing[22], Hong Kong[23,24] and Taiwan[25]. However, ascertainment bias,

differing study designs, and data collection methods may confound comparisons between studies[8]. A meta-analysis of 44 different population-based ALS studies supported these trends in the geographical incidence of ALS, with a standardised incidence of 1.89 per 100,000 people per year in North European populations, 0.83 per 100,000 people per year in East Asia and 0.73 per 100,000 people per year in South Asian populations[26].

Within the aforementioned population-based studies, researchers also reported an elevated male to female ratio of ALS incidence, with males from various racial backgrounds consistently making up more than 50% of cases[17–25]. An elevated risk of ALS in males is widely reported in the literature with a male to female ratio of 1.2-1.5, suggesting that males may harbour an elevated risk of developing ALS[11].

1.1.2 CLINICAL PHENOTYPE AND AETIOLOGY

Clinically, ALS manifests as progressive muscle weakness with muscular atrophy, stiffness and fasciculations. Degeneration of motor neurons in the motor cortex in combination with spinal segments innervating the upper or lower limbs is characterised as spinal ALS while degeneration of motor neurons and midbrain segments innervating the bulbar muscles is characterised as bulbar ALS[10]. The narrow mean survival window of 2-5 years following disease onset is often caused by degeneration of motor neurons innervating the respiratory muscles[10]. However, the clinical presentation and course of ALS is highly heterogeneous, including whether and to what extent motor deficits are accompanied by cognitive defects like those observed in frontotemporal dementia (FTD)[10].

FTD refers to a group of dementia-causing disorders that accompany ALS through genetic and pathological overlap[3]. Patients presenting with FTD experience personality changes, difficulty in memory retrieval and processing, lack of abstract thinking, language difficulty, and loss of

executive function and cognition as a result of deterioration of the frontal and temporal lobes[27]. Approximately 15% of patients diagnosed with ALS will develop symptoms analogous to FTD[28,29]. While having their distinct clinical features, ALS and FTD are recognised as part of a spectrum of neurological disease due to significant overlap between neuropathological and genetic features[30].

The aetiology of ALS has been attributed to complex interactions between genetic risk factors, environmental exposures and epigenetics; a widely accepted model suggests that ALS manifestation is preceded by a multistep process involving exposure to multiple ALS-related risk factors throughout an individual's life[31,32]. The contribution to the risk of developing ALS of genetic mutations can be substantial; however even mutations in highly penetrant genes such as superoxide dismutase 1 (*SOD1*), trans-activation response (TAR) DNA-binding protein 43 (*TARDBP*), chromosome 9 open reading frame 72 (*C9orf72*), fused in sarcoma (*FUS*) or ubiquilin 2 (*UBQLN2*) only cause disease in adulthood suggesting that additional exposures are required for disease to manifest[31]. The widely accepted multistep model suggests the requirement of six genetic and environmental exposures throughout an individual's lifetime to trigger sALS or fALS[31].

In support of multiple 'insults' being required to initiate disease, no single biological pathway is considered responsible for the pathogenesis of ALS. Instead, various pathways are proposed to be involved, including impaired RNA processing, dysregulated proteostasis, and prion-like protein aggregation[14].

1.1.3 GENETICS AND PROTEINOPATHY

Although most ALS is sporadic, genetics is a significant contributor to overall disease with approximately 68% of fALS and 11% of sALS cases attributable to a known genetic

component (Figure 1.1)[33]. While over 50 variants of ALS-linked genes have been identified through extensive research, the most frequent ALS-causing genes are *C9orf72, SOD1, TARDBP,* and *FUS*[34]; these four genes have demonstrated the importance of protein aggregation as a mechanism for disease in ALS, however defining the cause as a gain- or loss-of-function, or a combination of both, has proven difficult[35].

1.1.3.1 Superoxidase dismutase 1 (SOD1)

The *SOD1* gene located on chromosome 21 encodes an enzyme that binds copper and zinc to protect against accumulation of reactive oxygen species (ROS),) such as superoxide[36]. Inherited in an autosomal dominant fashion[37], approximately 12% of fALS is caused by mutations in SOD1, while 1% of sALS originates from *SOD1* mutations (Figure 1.1)[33].
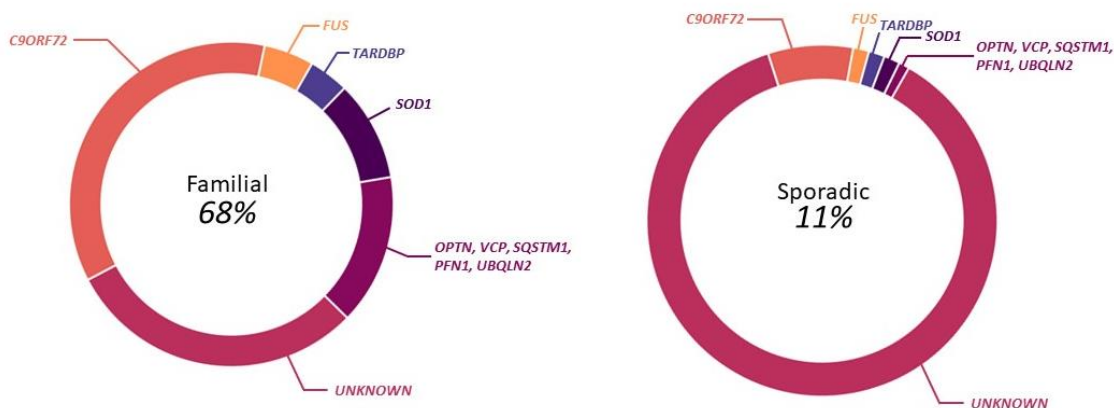
Figure 1.1: Proportion of fALS and sALS caused by variants in ALS-linked genes. Each segment represents the proportion of total fALS and sALS (68% and 11% respectively) attributable to mutations in ALS-linked genes[29]. Genetic overlap is apparent between fALS and sALS, however both the classifications of ALS have a considerable portion of genetic-linked cases that are of unknown aetiology. Proportions of cases in each category was obtained from European populations. Graph was reproduced using Excel (Version 2208) from Renton et al. (2014)[33] for examination purposes only.

Mutated *SOD1* was the first gene to be implicated as a cause of fALS, with mutant SOD1 displaying an elevated propensity for aggregation, causing a gain-in-toxic-function as opposed to a loss of antioxidant function[38–40]. Indeed, mouse models and human post-mortem tissues

consistently demonstrate that neuronal SOD1 aggregate formation is the main pathological feature of *SOD1*-linked ALS[39–41]. Mutant SOD1 has a toxic effect on a range of cell types in the central nervous system, regardless of whether aggregates are detectable in that cell type, including neurons, astrocytes, and microglia[42]. There is limited evidence supporting a clear association between *SOD1* mutations and cognitive decline analogous to FTD[43,44], with most patients exhibiting severe phenotypes marked by the progressive loss of motor function[45]. A case report suggests that the I114T *SOD1* variant may couple ALS disease with symptoms of FTD such as notable changes in personality and behaviour, and severe aphasia[44] however most patients with *SOD1* mutations do not exhibit FTD characteristics[43].

1.1.3.2 Trans-activation response (TAR) DNA-binding protein 43 (TARDBP)

The *TARDBP* gene found on chromosome 1 encodes the DNA- and RNA-binding protein TDP-43[46,47]. Mutations in the *TARDBP* gene exhibit an autosomal dominant mode of inheritance[48] and account for approximately 4% of fALS and 2% of sALS cases (Figure 1.1)[33,49]. While contribution of *TARDBP* mutations to FTD is relatively low, TDP-43 accumulation is a major histopathological contributor to FTD, with up to 50% of FTD patients with and without ALS presenting with TDP-43-positive inclusions[50,51].

Wildtype TDP-43 protein is localised in the nucleus where it carries out functions critical for RNA metabolism[52] and can readily shuttle into the cytoplasm where it aids in the stress granule response – a protective mechanism adopted by the cell in retaliation to cellular stress[53]. Mutations in the *TARDBP* gene cause cytoplasmic mis-localisation of the TDP-43 protein, resulting in nuclear depletion of functional TDP-43 and formation of hyperphosphorylated, ubiquitinated aggregates ("proteinopathy")[52,53]. TDP-43 proteinopathy can also be used to identify the stage of ALS progression due to its sequential spread from spinal alpha-motor neurons and the motor cortex (stage 1), to the prefrontal neocortex and brainstem reticular

formation in stage 2, to the pre- and postcentral neocortex and striatum (stage 3), and finally the temporal lobe and hippocampus (stage 4)[54]. The development of TDP-43 positive aggregates in motor neurons as well as their spread between cells has been shown to correlate to patterns of neurodegeneration seen in post-mortem brain tissue of ALS patients[54]. In alignment with mutant SOD1 pathophysiology, experimental efforts are still trying to establish whether TDP-43 proteinopathy promotes ALS progression through a loss or gain-in-toxic function, or a combination of both mechanisms[55].

1.1.3.3 Chromosome 9 open reading frame 72 (C9orf72)

The *C9orf72* gene, located on chromosome 9, is highly conserved across different model species due to its fundamental contribution to cellular function[56], with its key duties being vesicle trafficking, autophagy, and lysosomal stability[57]. An autosomal dominant GGGGCC (G_4C_2) hexanucleotide repeat expansion (HRE) mutation in *C9orf72* was found to be pathogenic in ~40% of fALS[33,58] and 5-10% of sALS cases (Figure 1.1)[58]. Furthermore, mutations in *C9orf72* are represented in ~25% of familial FTD cases[33,59] and 80% of ALS-FTD presenting cases, providing further avenues of genetic overlap between the two disorders[33,56].

Healthy individuals harbour two to eight G_4C_2 repeats. Expansions accommodating more than 30 repeats are considered pathological, with most ALS patients possessing thousands of repeats[56]. The expansion mutation is inherited in an autosomal dominant manner. Interestingly, patients with a pathogenic number of repeats often have a heterozygous *C9orf72*-mutant parent that does not present with clinical symptoms due to reduced penetrance, age-dependent onset, and variability in repeats[56,60]. The mechanism of pathogenicity is theorised to be one or a combination of the following: loss of function of the C9orf72 protein, a gain-in-toxic-function of the expanded *C9orf72* RNA, or toxicity of the dipeptide repeat proteins (DRPs) translated

from the intronic repeat[61]. Neuronal cytoplasmic clusters of TDP-43-negative DRPs are often seen in the frontal neocortex, cerebellum, and hippocampus of ALS patients harbouring *C9orf72* HRE mutations[61]. Interestingly, TDP-43 aggregates have been observed in *C9orf72* HRE-positive patient brains correlating to the pattern of neuronal loss. This led to the proposal of an amyloid-like cascade hypothesis in which the HRE of *C9orf72* and subsequent DRP formation provokes TDP-43 mis-localisation and accumulation, and subsequent neuronal degeneration[62,63]. Studies have shown that DRPs, predominantly polyGR (glycine-arginine), cause damage to the nuclear membrane and nucleocytoplasmic transport proteins, leading to mis-localisation of TDP-43 into the cytoplasm[62,64], demonstrating the converging mechanisms of ALS-causing mutations.

1.1.3.4 Fused in sarcoma (FUS)

The implications of mutant *TARDBP* on ALS disease progression prompted the discovery of ALS-causing *FUS* mutations, a protein that interestingly shares functional homology with TDP-43[33,65]. Akin to TDP-43, FUS is located within the nucleus[66] where it is largely involved in RNA metabolism as well as DNA repair, but is able to undertake nucleocytoplasmic transport to carry out cytoplasmic roles in RNA transport and translation and the stress response[67–69]. Dominant mutations in *FUS* are responsible for chromosome 16-linked ALS[33,70] which accounts for approximately 4% of fALS and less than 2% of sALS cases (Figure 1.1)[33,65]. *FUS* mutations exhibit a weak association with FTD, which when present is often co-diagnosed with ALS, making the significance of mutant FUS in FTD pathogenesis ambiguous[66,68]. FTD and ALS patients harbouring mutations in *FUS* demonstrate the characteristic presence of immunoreactive FUS aggregates within the cytoplasm of neurons, inducing neurodegeneration through a gain-in-toxic-function and stress granule recruitment[66]. Mutant FUS exhibits an increased affinity for stress granule formation alongside TDP-43 when exposed to oxidative

stress[66,71–73], which in turn sequesters wildtype FUS into the cytoplasmic inclusions thereby reducing RNA metabolism and initiates motor neuron degeneration[74].

1.2 *UBQLN2*-LINKED AMYTROPHIC LATERAL SCLEROSIS

The focus of this thesis is ALS caused by mutations in *UBQLN2*. Located on the X-chromosome, the intron-less gene *UBQLN2* encodes the ubiquitin-like protein ubiquilin 2 – a shuttle protein primarily involved in the ubiquitin-proteasome system (UPS)[6]. While the proportion of fALS and sALS cases caused by *UBQLN2* mutations is not documented, sex-specific differences in age of onset and disease progression has been reported[5]. A family consisting of 19 individuals presenting with ALS, despite a lack of mutations in known ALS-linked genes, triggered the discovery of *UBQLN2* as an ALS-linked gene in 2011[5]. Within this family, disease inheritance followed a dominant pattern, with later onset in females and no male-to-male transmission. The causative gene, *UBQLN2*, was found to contain a unique mutation encoding an amino acid change from proline to histidine at codon 497 (P497H). Further analysis of other families with ALS and ALS-FTD without mutations in *TARDBP*, *SOD1,* or *FUS* found an array of previously unidentified mutations in the *UBQLN2* gene at highly conserved residues[5]. ALS-causing mutations in *UBQLN2* were deemed to be inherited in an X-linked dominant fashion, with X-inactivation in females underpinning sex-specific differences[5,7].

1.2.1 Ubiquilin 2 protein structure

Five different ubiquilin proteins are present in the human body: 1, 2, 3, 4, and L, of which 1, 2 and 4 are ubiquitously expressed[75]. Ubiquilin proteins are critical components of the ubiquitin-proteasome system (UPS) which interacts with ubiquitinated proteins to regulate proteostasis, cellular stress, signal transduction and cell death[76]. Ubiquilin 2 interacts with polyubiquitin chains on proteins marked for degradation through its C-terminal ubiquitin-associated domain (UBA), targeting these proteins to the proteasome through interactions with its ubiquitin-like domain (UBL) at the N-terminus (Figure 1.2)[5,6]. Between the UBL and UBA domains are four stress-induced protein 1-like (STI-1) domains that interact with heat shock proteins, followed by a PXX domain consisting of 12 tandem PXX repeats (proline followed by any two amino acids) unique to ubiquilin 2 (Figure 1.2). While the exact significance of the PXX domain is largely unknown, most ALS-linked mutations in *UBQLN2* can be mapped to this region suggesting that it plays a vital structural and/or functional role[6,75].
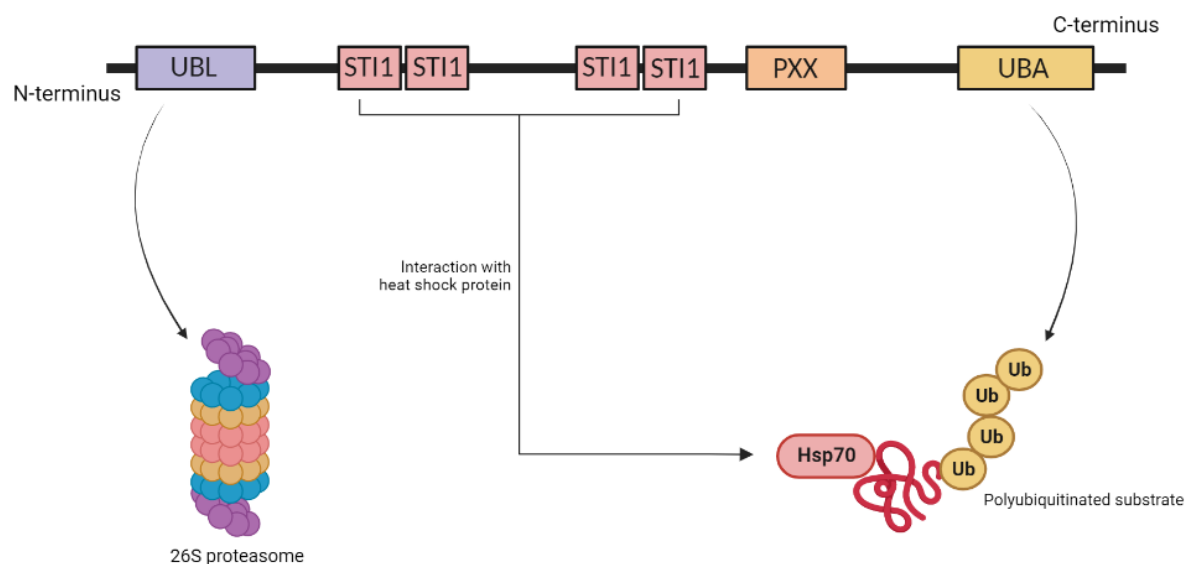
Figure 1.2: Functional domains of Ubiquilin 2. The functional domains of ubiquilin 2 are translated from a single exon. Interactions between the domains and their respective targets are imperative for UPS- targeting activity. Misfolded substrate proteins are marked for degradation through the addition of ubiquitin (Ub) proteins which are recognised by the ubiquitin-associated domain (UBA) of ubiquilin 2. The polyubiquitinated substrates are also recognised by heat shock protein 70 (Hsp70), a target for the stress-induced protein 1 (STI1) domain of ubiquilin 2. The ubiquitin-like domain (UBL) of ubiquilin 2 then aids trafficking of the misfolded protein to the 26S proteasome complex for degradation. Schematic adapted from Renaud et al. (2019) and created with BioRender.com.

1.2.2 Ubiquilin 2 Physiology and Pathophysiology

1.2.2.1 Ubiquitin proteasome system

Maintaining a cohesive and functional proteome through proteostasis is a requisite for cell viability. Newly synthesised proteins are often misfolded under physiological conditions, which is compensated for by proteolytic and protein refolding systems[77]. Proteostasis can be challenged by stressors which influence the accumulation of aberrant proteins or impair proteostatic machinery[78]. The ubiquitin-proteasome system is an essential part of the proteasome network responsible for degrading proteins tagged with polyubiquitin via an enzymatic cascade. As mentioned, polyubiquitinylated proteins are recognised by the UBL domain of ubiquilin 2 and trafficked to the proteasome for degradation via recognition sites in the UBA domain. Additionally, interactions between heat-shock protein 70 (Hsp70), the polyubiquitinated tails of aberrant proteins and the ubiquilin 2 STI1-like domains further facilitates the removal of proteins marked for degradation through other proteostatic mechanisms such as the endoplasmic reticulum-associated protein degradation (ERAD) system and autophagy[6,75]. Interestingly, ubiquilin 2 is reported to remain in an inactive state under physiological conditions with its propensity for binding Hsp70 increasing in the presence of Hsp70 substrates such as misfolded proteins. Ubiquilin 2 confers its shuttle activity when bound to a heat shock protein-substrate complex for removal via the proteasome (Figure 1.2)[6]. However, mutations of highly conserved proline residues in the PXX domain prevent ubiquilin 2 from interacting with Hsp70 bound to its misfolded substrate[79]. As a result, ubiquilin 2 remains in an inactive state which is shown to facilitate aggregate formation of misfolded proteins, and promote neurodegeneration and cognitive decline in animal models of ALS[79,80]. Supporting this, several *in vitro* studies have demonstrated a reduction in UPS efficiency due to the introduction of a ALS-causing mutations in *UBQLN2*[6,81,82]. Furthermore, an increase in

polyubiquitinated protein aggregation was observed, suggesting that disruption to the UPS facilitates aggregate formation, contributing to the observed disease phenotype[82]

1.2.2.2 Autophagy

Autophagy is a tightly regulated system that complements the UPS by facilitating bulk or selective degradation of proteins and organelles through fusion with lysosomes[75]. Mammalian cells use three primary methods of autophagy: microautophagy in which protrusions of the lysosome capture cargo for degradation, chaperone-mediated autophagy involving the triage of substrates to a lysosome through chaperone recognition, and macroautophagy , a process that requires the *de novo* formation of an autophagosome destined to fuse with a lysosome[83,84].

Ubiquilin 2 plays several key roles in macroautophagy. In the initial steps of macroautophagy, a double-membraned phagophore – a transient organelle formed to house substrates for degradation - expands and closes to form an autophagosome[85]. Expansion of the autophagosome is mediated by the maturation of microtubule-associated protein light chain 3 (LC3) from LC3-I to LC3-II. Ubiquilin 2, alongside 1 and 4, facilitates the maturation of LC3-I, with ALS-linked mutant *UBQLN2*[P497H] rats exhibiting lower LC3-II to LC3-I ratios as a reflection of impaired autophagy[86]. Additionally, reduced LC3-II:LC3-I ratio in a knockdown model of ubiquilin 2 from HeLa cells correlated with a reduction in autophagosome number[87]. While the process of sequestering cellular debris can be non-selective, substrates marked with ubiquitin chains linked at lysine residues K48 and K63 are specifically targeted by ubiquilin proteins for triage to the autophagosome[87]. Recent studies have shown that ubiquilin 2 mediates the acidification of lysosomes by stabilising the V-ATPase proton pump, a critical component of lysosomal acidification[88–90]. Wu et al (2020)[89] demonstrated a downregulation of the ATP6v1g1 subunit of the V-ATPase in P497S mutant *UBQLN2* mice, aligned with reduced autophagosome acidification and increased accumulation of autophagic markers such as K63-

and K48-linked ubiquitin and p62, deducing that mutant *UBQLN2* is interfering with proteostasis through disruptions in autophagy[89]

1.2.2.3 Stress granules

Comprehensive stress responses are imperative to reduce damage and promote cell survival when exposed to oxidative stress, hypoxia, toxins, and temperature changes[91]. A highly conserved stress response is the formation of membrane-less messenger ribonucleoprotein particles called stress granules formed through liquid-liquid phase separation of RNA-binding proteins and protein translation factors. Stress granules are utilised as transient 'storage facilities' for non-translating mRNA, as cells will suspend non-essential protein synthesis when exposed to stressors[91,92].

The membrane-less nature of stress granules gives them 'liquid-like' properties such as the ability to fuse, change in size, and change phase from dilute to condensed; this phenomenon is termed 'liquid-liquid phase separation' (LLPS)[93]. Persistence of stressors and disrupted LLPS can perpetuate stress granules and insoluble aggregate formation, causing the disruption of protein and RNA stasis[6,91,92,94–97]. ALS-linked *UBQLN2* mutations in the PXX domain cause a disruption of LLPS by promoting the transition from liquid to solid formations and inducing ubiquilin 2 aggregates[97,98]. Ubiquilin 2 engages with stress granules through its STI1-like domains to drive stress granule formation in the early stages of phase separation[6,99] The involvement and recruitment of ubiquilin 2 to stress granules signifies its role in stress-granule re-modelling. Various studies investigating the components of stress granules in ALS/FTD have identified the presence of TDP-43, FUS, ubiquitin, and ubiquilin 2, suggesting that ubiquilin 2 functions at the interface between protein and RNA processing, two key disease pathomechanisms[6].

1.2.3 Neuropathology of ubiquilin 2 in ALS

Histological characterisation of neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, ALS, and FTD documents persistent aggregation of disease-specific proteins in the central nervous system[100]. Skein-like inclusions are frequently observed in motor neurons of ALS patients with or without FTD, and are comprised of ALS-associated proteins including TDP-43, FUS, SOD1, p62, ubiquilin 2 and optineurin[5,101,102]. Ubiquilin 2 inclusions are a consistent neuropathological feature of ALS observed in patients harbouring mutant or wildtype *UBQLN2*; interestingly, ubiquilin 2-positive aggregates are prone to colocalization with other ALS- and FTD-linked proteins[6,103,104]. While ubiquilin 2 inclusions are observed, unsurprisingly, in patients with *UBQLN2* mutations, they also occur occasionally in sALS and are abundant in *C9orf72*-linked ALS cases[5,105].

1.2.3.1 Ubiquilin 2 neuropathology in UBQLN2-linked ALS

Immunohistochemical analysis of skein-like aggregates in spinal cord sections from patients harbouring the P497H and P506T *UBQLN2* mutations revealed that not only were these inclusions positive for ubiquilin 2 but also p62, TDP-43, FUS, and OPTN[5]. Wildtype and mutant ubiquilin 2 directly interact with the C-terminal fragment of TDP-43 and have repeatedly demonstrated colocalisation in TDP-43 and/or FUS positive aggregates in ALS[99,106–108]. Indeed, more cytoplasmic mis-localisation and aggregate formation was observed in cells expressing mutant *UBQLN2* as opposed to wildtype, however this demonstrates that wildtype ubiquilin 2 is also recruited into aggregates and therefore plays an important role in its formation[5,6,109,110].

1.2.3.2 Ubiquilin 2 neuropathology in C9orf72-linked ALS/FTD

Ubiquilin 2 neuropathology demonstrates a distinct pattern of expression in ALS patients with hexanucleotide repeat expansion of *C9orf72* than in those without[105]. Specifically, abundant irregular ubiquilin 2-positive inclusions were observed in the molecular layer of the hippocampus and regions CA1 through CA4, and in the granular layer of the cerebellum, that were not present in non-expansion cases[105]. In fact, the unique pattern of ubiquilin 2 pathology in *C9orf72*-linked ALS cases successfully predicted the presence of a *C9orf72* HRE mutation in cases later confirmed to harbour the mutation through genetic testing[105]. These ubiquilin 2-positive inclusions were later discovered to be dipeptide repeat proteins encoded by the *C9orf72* expanded repeat, indicating that in *C9orf72* cases, ubiquilin 2 tags other aggregating proteins while in *UBQLN2*-linked disease it aggregates on its own accord[105].

1.2.4 Epigenetics of *UBQLN2*-linked ALS and FTD

Contributions from both genetics and the environment modify ALS risk and phenotype. Epigenetics, or the pattern of gene expression, plays a crucial role in biological processes and is often of interest in complex diseases where it may bridge the gap between genetic and environmental factors[111]. Epigenetics is of relevance to *UBQLN2*-linked ALS/ FTD because the X chromosome is subject to an epigenetic modification in females called X-inactivation.

1.2.4.1 Epigenetic modification

Epigenetic modifications are heritable alterations to the structure and accessibility of DNA that can modify gene expression[112]. Epigenetic modifications can be inherited by daughter cells derived from the same germline cell[111] and trans-generationally from an organism to their offspring[113].

There are two main types of epigenetic modifications that can influence the expression of DNA: histone modifications and DNA methylation. Histones can undergo a variety of alterations including methylation, acetylation and phosphorylation that influence changes in the chromatin structure, making the DNA more or less accessible to transcription[114]. Histone acetylation is a highly established modification in which the addition of an acetyl group by acetyl transferases, on the N-terminal lysine residues of histone 3 and 4, neutralises their positive charge thus reducing their affinity for negatively charged DNA and promoting transcription[114,115]. Conversely, DNA methylation represses transcription through the addition of a methyl group to the fifth carbon of a cytosine residue in the promoter of a gene by a methyl transferase enzyme. The methyl group forms a physical barrier against transcription factors (Figure 1.3) and can also promote binding of transcriptional repressors such as histone deacetylases that remove histone acetyl groups[114]. Interestingly, DNA methylation of cytosine residues often occurs when followed by a guanine, forming a "CpG" site usually located in the promoter region of a gene (Figure 1.3)[116]. A short, interspersed region of CG-rich bases located in and around the transcriptional start site of a gene is termed the CpG island. The CpG island of a gene is approximately 200 bp long and flanked by a north shore and north shelf 2 kb and 4 kb upstream, respectively and a south shore and south shelf 2 kb and 4 kb downstream, respectively[117–120] (Figure 1.3).
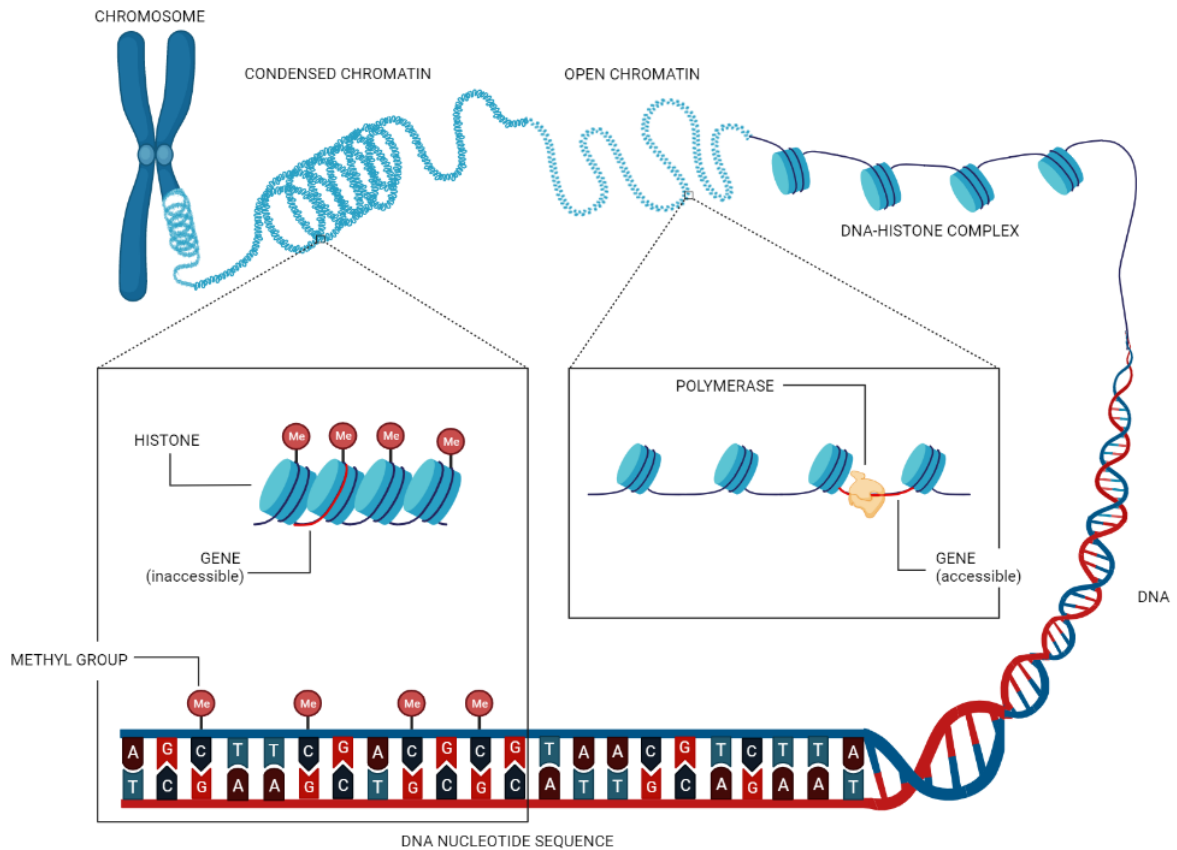
Figure 1.3: Addition of methyl groups to histones promotes a condensed chromatin structure and represses gene translation. The addition of methyl groups to cytosine bases followed by a guanosine is reflected by a condensed chromatin structure, making the gene inaccessible to polymerase machinery for transcription. The lack of methylation at CpG sites forms an open chromatin structure that is accessible for transcription. Created with BioRender.com.

1.2.4.2 X-chromosome inactivation (XCI)

All mammals follow an XY system of sex determination, with females harbouring two X-chromosomes (XX) and males harbouring one (XY)[121]; this heteromorphic evolution of the sex chromosomes has rendered the Y-chromosome gene poor with respect to the X-chromosome[122]. A gene-dose imbalance would be expected between males and females as a result, however, this is not observed. Instead one X-chromosome in females undergoes X-chromosome inactivation (XCI) as a form of gene dosage compensation[112,122].

XCI is established in the cells of the embryo proper during development[123] with some studies demonstrating the presence of XCI during the 8-cell stage of development in a preimplantation embryo[124]. The process of XCI is driven by interaction of the long non-coding RNA *Xist* with

the inactive X-chromosome[123,125]. *Xist* forms a cloud-like structure around the chromosome, triggering chromosome-wide gene silencing through the recruitment of proteins that introduce epigenetic modifications to form a heterochromatin or Barr body[112,122]. The classification of condensed heterochromatic regions in chromosomes was conceived by Emil Heitz in 1928, who observed that regions of chromosomes retained dark staining during mitosis and proposed that the heterochromatic state corresponded to gene silencing[126,127]. The finding of large heterochromatic regions in sex chromosomes lead to the discovery of gene silencing in one of the two X-chromosomes in females, producing a Barr body [128,129]. Principal mechanisms that permit chromosome-wide gene silencing include histone modification, CpG island methylation and recruitment of repressive proteins[126].

The single X-chromosome in males is always maternal in origin, while females inherit maternal and paternal X-chromosomes, creating a mosaic of active X-chromosomes of maternal or paternal origin across different tissue types[130]. Female mosaicism means that X-chromosome gene variants may be expressed in 'patches' of cells originating from the same stem cell or in different tissue types by virtue of random inactivation of the mutant allele. Consequently, approximately 50% of the cells in females will express the mutant allele and 50% will express the wildtype, mitigating the loss- or gain-in-function caused by the mutation[130]. In most females, the maternal and paternal X-chromosomes have an equal chance of undergoing random inactivation[131], however, in cases of XCI skewing, more than 75% of cells can express a bias in XCI. Skewed XCI can occur by chance or through genetic, epigenetic, or environmental influences than may preferentially inactive one X-chromosome over the other[131]. XCI, especially skewed-XCI, can influence the severity of X-linked diseases in heterozygous females[132].

Carrying a disease-linked gene mutation on one X-chromosome contributes to reduced disease severity in females compared to males, particularly in disease onset, clinical manifestation, and

pathophysiology[130]. The dichotomy in disease severity is especially clear with X-linked recessive diseases such as Duchenne muscular dystrophy (DMD)[133] and haemophilia A, where female carriers are often spared[134]. Recessive disease requires the expression of the mutant allele on both chromosomes for disease progression; because males harbour a one X-chromosome, the expression of a single mutant allele is sufficient to manifest disease, as opposed to females that may express the mutant allele on the maternal or paternal X-chromosome but due to random XCI do not experience disease symptoms to the same severity as males[135]. Approximately 1 in 5,000 males are affected by DMD, in stark contrast to 1 in 1 million females, highlighting the divide in disease severity in X-linked recessive disorders[136]. For X-linked dominant mutations, the expression of the mutant allele on a single X-chromosome is sufficient to cause disease in males and females, often with differing severity[130,137]. The presence of a second X-chromosome harbouring the wildtype allele can mitigate pathogenicity generated by expression of the mutant allele on the alternative X-chromosome due to XCI and female mosaicism[130]. Fragile X syndrome, caused by a trinucleotide CGG repeat in the fragile X messenger ribonucleoprotein 1 (*FMR1*) gene demonstrates clear differing clinical presentations between males and females through a pattern of X-linked dominant inheritance[138]. Males expressing the full mutation exhibit severe intellectual disability, attention disorders and symptoms of autism spectrum disorder. Similarly, the phenotypic profile of females encompasses intellectual disability and behavioural alterations however, inactivation of one X-chromosome ensures that some cells expressing the wildtype gene produces fragile X messenger ribonucleoprotein (FMRP), although not restorative, females exhibit milder symptoms as a result[139,140].

UBQLN2-linked ALS is inherited in an X-linked dominant fashion[5,7]. Investigation towards an X-linked mode of disease inheritance, initiated by the absence of male-to-male transmission of ALS reported in families with generational ALS, identified mutations in the PXX domain of

UBQLN2 as the main driver[5,7]. Males with *UBQLN2*-linked ALS have been reported to exhibit more severe bulbar and respiratory symptoms of ALS, females are not completely unaffected[5,7]. Although studies have noted a difference in the age of disease onset of 33.9 years in males and 47.3 years in females[5], females harbouring *UBLNQ2* mutations may develop ALS at a younger age than those without a mutation and are still affected by ALS disease[11]. The mechanisms supporting sex-specific differences in *UBQLN2*-linked ALS is not fully understood and has prompted the requirement for appropriate methods to understand XCI and how it may affect disease progression and severity.

1.2.4.3 Methods for understanding X-inactivation in disease

The ratio of mutant and wildtype transcript or protein expression is used as an index to understand XCI and skewed-XCI[141]. A combination of techniques, including the human androgen receptor (HUMARA) assay[142,143], allele-specific quantitative PCR (qPCR)[144,145], and RNA-sequencing (RNA-seq) have been implemented in the understanding of XCI in disease. The HUMARA assay is a PCR based technique that takes advantage of the differing number of polymorphic CAG repeats between the active and inactive X-chromosome and susceptibility of the unmethylated allele to be digested by the HpaII enzyme. As a result, the PCR products can be visualised through gel electrophoresis to determine the pattern of XCI[143]. Alternatively, allele-specific qPCR allows the differentiation between alleles with a single mutation using allele-specific forward primers[144,145]. RNA-seq analyses the quantity and sequence of extracted RNA from a population of cells using next generation sequencing[146,147]; male to female expression levels of X-linked genes can be analysed using RNA-seq to reflect XCI patterns[147].

To understand the influence of skewed-XCI on mutant allele expression, indicators such as CpG island methylation is often assessed to reflect architectural changes in chromatin structure that regulates gene expression[148]. Sodium bisulfite conversion is the gold-standard technique

for detecting 5-methylation of cytosine residues of CpG sites in mammalian genomic DNA[149,150]. Sodium bisulfite selectively deaminates unmethylated cytosines to uracil, presenting an efficient means for encoding DNA methylation into the primary sequence of bases[150,151]. Subsequent PCR amplification of the converted DNA produces thymine residues in place of unmethylated cytosines due to uracil and thymine base pairing, while methylated cytosine bases remain unchanged[150]. Since its discovery, bisulfite conversion has formed the basis of different sequencing methods, including bisulfite-sequencing PCR (amplifies both methylated and unmethylated DNA) and methylation-specific PCR (only amplifies methylated templates, using primers against the bisulfite-encoded methylation)[152].

In the context of X-linked disease like ALS and ALS-FTD, bisulfite sequencing can be used to identify differentially methylated regions of the X-chromosome that may drive or protect against disease progression[153]. Specifically, bisulfite sequencing can be used to identify differential methylation patterns in *UBQLN2*-linked ALS between males and females pertaining to XCI. Females with *UBQLN2*-linked ALS demonstrate protection against disease progression and symptom manifestation which may be driven by unique methylation patterns in the CpG island of *UBQLN2* because of XCI[5,7]. Identifying this pattern may enable a basis for prediction of protection against *UBQLN2*-linked ALS. Bisulfite conversion combined with DNA sequencing can provide single nucleotide resolution to reveal DNA methylation patterns.

1.3 MODELS FOR UNDERSTANDING *UBQLN2*-LINKED ALS AND FTD

Model systems are fundamental for capturing a unique understanding of disease mechanisms, pathophysiology and systemic repercussions of disease interventions[154]. *In vitro* and *in vivo* methods of disease modelling are particularly important for rare diseases such as ALS and FTD to predict the impact of genetic, epigenetics and environmental factors on disease progression[155]. Transgenic animal models have been abundantly employed to study the effects and therapeutic targets of ALS/FTD-linked genes. Mice and rats are common animal models due to their similarities to human anatomy, physiology and genetics[156,157], however other models such as *Caenorhabditis elegans* (*C. elegans*), Zebrafish, and drosophila melanogaster are also routinely used[158].

The first transgenic *UBQLN2*-mutant mouse model was developed to express *hUBQLN2^{P497H}* that developed ubiquilin-positive inclusions in the hippocampus and cognitive deficits similar to that in humans[159]. However, sex-specific differences were not incorporated into the study design as the *hUBQLN2^{P497H}* transgene had integrated into the Y-chromosome, making all the mice males[159]. Studies that do include male and female mice often don't explore gender differences as the mutations are not integrated into the X-chromosome, for example, a study including an equal number of *hUBQLN2^{P506T}* transgenic and non-transgenic (wildtype) male and female mice only reported on behavioural differences based on transgenic or non-transgenic status of the mice, mitigating any potential gender effects or evidence of *UBQLN2* redundancy in the mutant female mice[160]. Furthermore, a study conducted by Hjerpe et al. (2016) introduced a *UBQLN2^{P520T}* knock in mutation on the X-chromosome yet only assessed functional changes in male mice[80].

As mentioned previously, *UBQLN2*-linked ALS tends to impact males at an earlier age and with more severe symptoms than females[5] however this is not always captured in animal

models. Often the main objective in studies of this nature is to predict the main pathophysiological interactions and behavioural effects of mutant *UBQLN2* which can be achieved without incorporating elements of X-inactivation[159,161]. While overexpression and knockout models can recapitulate *UBQLN2* pathology, ignoring epigenetics when trying to understand an X-linked form of ALS shuts down an entire avenue for potential therapies.

Further, it is necessary to recognise the limiting nature of animal models and their replicability of a phenotypically variable disease like ALS/FTD; this is especially true when considering that therapies that show remarkable results in animal trials seldom show as much therapeutic benefit in humans[158].

The use of human patient-derived cells for research is one of the approaches geared towards personalised medicine and is especially useful when analysing epigenetic modifications that can be difficult to replicate in animal models. Differential DNA methylation patterns between males and females have been explored using human pancreatic islet cells to determine the impact on gene expression and subsequent insulin expression; concluding that epigenetic modifications can significantly affect metabolic phenotypes in a sex-specific manner[117]. For this thesis, human fibroblasts were available from three heterozygous female *UBQLN2^T487I^* carriers to determine patterns of methylation of the wildtype and mutant *UBQLN2* alleles.

1.4 RESEARCH OBJECTIVES

Since the discovery of *UBQLN2* as a causative gene for X-linked ALS and ALS/FTD[5], investigations into the pathological contribution of ubiquilin 2 in ALS have increased. However, the redundancy exhibited in females heterozygous for mutant *UBQLN2* through XCI remains unexplored. Although XCI is a well-established principle, determining the differential pattern of CpG island methylation in the promoter region of *UBQLN2* between males and affected females may highlight the molecular basis of female protection against *UBQLN2*-linked ALS. Publicly available methylome data will be used to investigate methylation of CpG sites across the CpG island of *UBQLN2* in different tissue types between males and females to investigate patterns of differential methylation. Human fibroblast samples will then be used to optimise sodium bisulfite conversion of genomic DNA (gDNA) to capture methylation of CpG sites for each sample. This research will provide insight to the extent of sex-specific methylation differences of CpG sites across the CpG island of *UBQLN2* as a potential mechanism for female protection against *UBQLN2*-linked ALS. The methylation data and methods optimised for deriving patient-specific methylome information can be used in conjunction with RNA sequencing of *UBQLN2*, allele-specific quantitative PCR (qPCR) between mutant and wildtype variants, as well as functional analyses of patient-derived cells to determine whether XCI methylation patterns can help predict female protection against X-linked MND.

2 MATERIALS AND METHODS

2.1 IDENTIFYING SEX-SPECIFIC SITES OF *UBQLN2* CPG ISLAND METHYLATION

Publicly available databases were used to identify the location and extent of methylation within the CpG island of the *UBQLN2* gene. Further, multiple tissue types were compared to identify any tissue-specific methylation patterns of CpG sites in the *UBQLN2* CpG island. Epigenome-Wide Association database (EWAS) and MethBank were the databases used to extract the methylome information.

2.1.1 Identifying the region of interest for X-inactivation-linked methylation of human *UBQLN2*

The chromosomal coordinates of the CpG island for human *UBQLN2*, chrX:56589945-56590514, were determined using the University of California, Santa Cruz (UCSC) human genome browser in Genome Reference Consortium Human Build 37 (GRCh37)[162]. The sequence was downloaded from the National Centre for Biotechnology Information (NCBI) [*UBQLN2* gene NCBI reference: NC_000023.10][163]. The north and south shores and shelves of the *UBQLN2* CpG island, located from 0-2 kb and 2-4 kb from the 5' most end of the CpG island proper[117–119,164], was not included in the sequence obtained. The additional flanking sequences were obtained from NCBI through manual selection of 4 kb upstream from position chrX:56589945 in GRCh37. The sequence was aligned and the CpG sites along the CpG island were annotated in Geneious Prime (version 2020.0.5)

2.1.2 Obtaining methylome data from publicly available databases

2.1.2.1 Epigenome-Wide Association Studies (EWAS) Open Platform Database

DNA methylation is a widely characterised form of epigenetic modification that can report on genome-wide gene expression and, of most relevance to this study, can report on X-inactivation. Epigenome-Wide Association studies are an effective tool to investigate the correlation between phenotypes and epigenetics through DNA methylation of CpG islands. EWAS employs a host of methylation analysis methods, the most common being the Infinium Human Methylation 450K bead chip-analysis[165]. The EWAS open platform database, curated by the National Genomics Data Centre in China, provides integrated data from 910 publications for 617,018 EWAS associations, with normalised methylation array information across more than 700 tissues and 218 cell lines[166]. EWAS employs a standardised method of normalising methylation data from the Gene Expression Omnibus (GEO)[167] repository and The Cancer Genome Atlas (TCGA), providing ease of comparison between different samples, tissue types and CpG probes[168].

Methylation array data from the Infinium Human Methylation 450K bead chip was downloaded as a 'manifest' file (version 2.1) from the EWAS open platform database, which consisted of Infinium probe IDs associated with each CpG island analysed across the genome and their respective gene and chromosomal location. Infinium Human Methylation 450K technology assesses methylation status of more than 45,000 CpGs across the genome through bisulfite conversion and probe based Infinium I/II chemical assays to distinguish between methylated and unmethylated CpG sites[169]. Infinium probes targeting the human *UBQLN2* gene were isolated by filtering the manifest data for '*UBQLN2*' in the 'UCSC_RefGene_Name'

column of the file and for the chromosomal coordinates of the CpG island, chrX:56589945-56590514, as determined in section 2.1.1.

R-studio (version 4.1) was used to extract raw methylation data for each probe in the CpG island of the human *UBQLN2* gene, by using a script (written by Scotter Lab PhD candidate Maize Cao) to produce a .csv file with a subset of the data having column titles "sample_id", "sex", and "tissue", and the following probe IDs; "cg01780361", "cg02492740", "cg02587153", "cg04270663", "cg09013091", "cg09666011", "cg10009003", "cg10950413", "cg14824933", "c16161440", "cg20979765", "cg21862542", "cg25697726", "cg23993375", "cg25697726" (Table 1).

The metadata extracted enabled analysis of individual samples by sex (male/female) and β-value (methylation) of each CpG site probed for in the human *UBQLN2* gene (identified by an Infinium probe ID) across seven different tissue types in males and females.

2.1.2.2 Methbank 3.0

Similar to the EWAS open platform database, the Methbank 3.0 database[170] provides collated methylome data across different species obtained from bisulfite-converted whole genome sequences from the Genome Sequence Archive[171] and Sequence Read Archive[172]. Consensus reference methylome data for chromosome X only for 13 different tissue types was extracted as '. wig' files and converted to '.bed' files using BEDOP software (version 2.4.30-gimkl-2017a) on the NeSI platform using code written by Scotter Lab PhD candidate Maize Cao.

X-chromosome methylome information from MethBank 3.0 was extracted as multiple .bed files separated by tissue type and sex. Within the files, the average β -value for a single CpG site was reported alongside the genomic location according to Genome Reference Consortium Human Build 38 (GRCh38). To ensure the same CpG sites were analysed between the EWAS

and MethBank 3.0 datasets, the genomic location of the Infinium probes reported in the EWAS data was converted from GRCh37 to GRCh38. The Ensembl genome browser[173] reports the coordinates of the *UBQLN2* gene as chrX: 56,590,026 - 56,593,443 in GRCh37, following conversion to GRCh38 using the Ensembl assembly converter[174] the coordinates became chrX: 56,563,593 - 56,567,010 representing a -26,433 base shift. Subsequently, the genomic location column of the EWAS data was adjusted by -26,433 bases to determine the genomic location of each Infinium probe in GRCh38. Using the genomic location of each Infinium probe within the CpG island of the *UBQLN2* gene in GRCh38 from the EWAS data, the same probes were then isolated from the Methbank data (Table 1).

2.1.2.3 Control genes

Using experimental controls is necessary to ensure robust results, provide comparison to a standard and diminish the effects of bias[175]. The human androgen receptor (*AR*) is located on the X-chromosome and is subject to X-inactivation, making this a suitable positive control gene to assess methylation levels between males and females. Cyclin-dependent kinase 16 (*CDK16)* is also located on the X-chromosome but escapes X-inactivation[176], therefore making this a suitable negative control. Additionally, *TARDBP* located on chromosome 1 was included as a negative control to determine whether there were methylation differences between males and females in the CpG island of an autosomal gene. Methylome information for the control genes was extracted from the EWAS manifest file by filtering for "*CDK16*", "*AR*" and "*TARDBP*" in the UCSC_RefGene_Name column. To maintain consistency between all datasets, the genomic location of the control genes, originally expressed in GRCh37, was also converted to GRCh38.

2.1.3 Beta-Value

The extent of methylation at a CpG site was expressed as a beta (β) value. Infinium Human Methylation 450K assay measures the intensity of methylated and unmethylated probes at a CpG site from a population of cells. The ratio of methylated probe intensity to overall (methylated and unmethylated probe) intensity is the β-value, which ranges from 0 (no methylation detected) to 1 (fully methylated CpG site)[177]. The β-value for each probe in the EWAS and Methbank 3.0 datasets was extracted for males and females for different tissue types. The tissue types of interest were all brain regions available in the databases and tissues commonly used to screen for molecular markers of disease; these tend to be readily accessible like blood, skin, and saliva, all of which were available in both datasets. It should be noted that while EWAS methylome data provided normalised individual β-values for each study for a given probe ID or tissue, methylome data from Methbank 3.0 provided the average β-value across different studies for a given probe ID or tissue type.

Table 1: Summary of CpG sites analysed from EWAS and Methbank 3.0. Probe ID and the respective genomic location for a single CpG site in the promoter region of each gene. Probe ID is organised from 5' to 3' for each gene. 'Relation to CpG island' is listed as defined in the Infinium Human methylation 450K v1.2 manifest file.

	Probe ID	Genomic Location (GRCh38)	UCSC CpG Island Location	Relation to CpG Island
UBQLN2	cg21862542	56560144	chrX:56589945-56590514	North Shelf
	cg02587153	56563335	chrX:56589945-56590514	North Shore
	cg10009003	56563359	chrX:56589945-56590514	North Shore
	cg25697726	56563485	chrX:56589945-56590514	North Shore
	cg09013091	56563513	chrX:56589945-56590514	Island
	cg14824933	56563516	chrX:56589945-56590514	Island
	cg23993375	56563518	chrX:56589945-56590514	Island
	cg10950413	56563664	chrX:56589945-56590514	Island
	cg09666011	56563680	chrX:56589945-56590514	Island
	cg16161440	56563754	chrX:56589945-56590514	Island
	cg02492740	56563929	chrX:56589945-56590514	Island
	cg01780361	56563953	chrX:56589945-56590514	Island
	cg04270663	56564080	chrX:56589945-56590514	Island
	cg20979765	56564104	chrX:56589945-56590514	South Shore
AR	cg26003280	66736034	chrX:66763684-66764077	North Shore
	cg25686125	66736576	chrX:66763684-66764077	North Shore
	cg20878850	66737252	chrX:66763684-66764077	Island
	cg08063601	66737509	chrX:66763684-66764077	Island
	cg27271368	66737978	chrX:66766037-66766279	Island
	cg01086868	66738098	chrX:66766037-66766279	Island
	cg03096488	66738856	chrX:66766037-66766279	North Shore
	cg21966410	66739449	chrX:66766037-66766279	North Shore
	cg05019001	66739901	chrX:66766037-66766279	South Shore
	cg26498015	66761656	chrX:66766037-66766279	Undefined
	cg15840039	66761665	chrX:66766037-66766279	Undefined

CDK16	cg18407858	47047850	chrX:47077163-47078637	North Shelf
	cg10913852	47050735	chrX:47077163-47078637	Island
	cg06691299	47050928	chrX:47077163-47078637	Island
	cg05292991	47050970	chrX:47077163-47078637	Island
	cg14044580	47051382	chrX:47077163-47078637	Island
	cg17334225	47051442	chrX:47077163-47078637	Island
	cg11333230	47051444	chrX:47077163-47078637	Island
	cg23212388	47051497	chrX:47077163-47078637	Island
	cg22826577	47051553	chrX:47077163-47078637	Island
	cg17137203	47052203	chrX:47077163-47078637	Island
	cg23061435	47053142	chrX:47077163-47078637	South Shore
	cg24655012	47055307	chrX:47077163-47078637	South Shelf
TARDBP	cg18163627	11045155	chr1:11072319-11073250	North Shore
	cg00692549	11046022	chr1:11072319-11073246	Island
	cg10233455	11046189	chr1:11072319-11073247	Island
	cg15482690	11046309	chr1:11072319-11073249	Island
	cg18380490	11046794	chr1:11072319-11073251	Island
	cg15430883	11046800	chr1:11072319-11073248	Island
	cg00330490	11047475	chr1:11072319-11073245	South Shore
	cg22693994	11058786	chr1:11072319-11073252	Undefined

Abbreviations: AR; Androgen Receptor, CDK16; Cyclin Dependant Kinase 16, TARDBP; TAR DNA binding protein, UCSC; University of California, Santa Cruz.

2.1.4 Transformation and analysis of methylome data

To allow comparison of β-values between sex, tissue type and databases, the individual β-values obtained from EWAS were averaged as a function of tissue type and probe ID (individually). As described, in section 2.1.3, methylome data from Methbank 3.0 was already presented as an average β-values for each probe and each tissue type, therefore no transformation was required of the raw data. GraphPad Prism (version 9.0.2) was used to graph the average probe β-values ± standard deviation between males and females for each tissue type. The data was also expressed as mean tissue β-values ± standard deviation for each CpG probe between males and females to highlight any tissue-specific methylation differences within the CpG island of the analysed genes.

2.1.4.1 Statistical Analysis

Multiple analyses were conducted on the publicly available methylome data from EWAS and MethBank. To determine the significance of methylation differences between males and females for a given CpG site a paired t-test with an $\alpha = 0.05$ was conducted. When comparing average β-value across CpG sites, different tissues or sexes, CpG sites with higher β-values (those located in shelves) were not included to prevent false inflation of the average β-value in the CpG island proper. To determine the significant between the *average* β-value across different tissues, sexes or CpG sites, an unpaired t-test was conducted, p-values < 0.05 considered significant.

Two-way ANOVA with multiple comparisons were used to determine significance between different parameters of the combined data across and between tissue types with p-values < 0.05 considered significant.

2.2 QUANTIFYING METHYLATION OF *UBQLN2* IN HUMAN SAMPLES INCLUDING MUTANT *UBQLN2* FIBROBLASTS

2.2.1 Cell Culture

Fibroblast punch biopsies were obtained from nine patient donors; three wildtype males: case numbers MND045, MND053 and MND054, three wildtype females: MND055, MND056 and MND057 and three mutant females: MN17, MND033 and MND046. The fibroblasts were plated at 500,000 per well for a total of 4 wells per sample. Fibroblasts were grown in Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific) (90%), and supplemented with fetal bovine serum (FBS) (10%) and Pen-strep (1%) and harvested at the following passages: MN17 p8, MND033 p10, MND045 p8, MND046 p8, MND053 p8, MND054 p8, MND055 p9, MND056 p8, MND057 p9.

2.2.2 Genomic DNA (gDNA) Extraction

gDNA was extracted from frozen cell pellets using the QIAGEN DNA Mini kit (Catalogue number: 51304) as per the manufacturer's protocol with minor modifications. Briefly, the cell pellet was resuspended in sterile PBS to a final volume of 200 µL to which 20 µL of proteinase K (QIAGEN protease; 20 mg/mL) and 200 µL of lysis buffer (buffer AL) was added to degrade nuclease enzymes[178] and promote cell membrane lysis, respectively. Following incubation for 10 minutes at 56 °C, 200 µL of 100% ethanol was added. The mixture was placed in a QIAmp spin column and centrifuged at 6000 x g (8000 rpm) for 1 minute. To the column, 500 µL of buffer AW1 was added and centrifuged for 1 minute at 6000 x g, followed by the addition of 500 µL of buffer AW2. The sample was centrifuged at 20,000 x g (14,000 rpm) for 3 minutes. To the spin column, 30 µL of elution buffer (buffer AE), heated to 75 °C, was added and incubated for 5 minutes before being centrifuged at 6000 x g for 1 minute. The elution step was repeated to give a final elution volume of 60 uL. The concentration and purity of the extracted

gDNA was measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and stored at -20 °C.

2.2.3 Bisulfite conversion

2.2.3.1 EZ DNA Methylation Gold Kit (#D5005) – Zymo Research

Genomic DNA was bisulfite converted using the EZ DNA methylation gold kit (Catalogue number: D5005) from Zymo research. Briefly, 500-2000 ng of gDNA was bisulfite converted with CT conversion reagent at 98 °C for 10 minutes followed by 64 °C for 2.5 hours. The converted DNA was then desulphonated and washed in a Zymo-spin™ IC column and eluted into 10 uL of elution buffer. DNA recovery was measured using a Nanodrop 2000 spectrophotometer.

2.2.3.2 QIAGEN EpiTect bisulfite kit

Sodium bisulfite conversion of gDNA using the QIAGEN EpiTect bisulfite kit was prepared according to the *'Sodium bisulfite conversion of unmethylated cytosines in DNA'* protocol provided in the EpiTect Bisulfite Handbook 2020[179], with minor modifications. 1 ng – 2 ug of gDNA was bisulfite converted using 'Bisulfite mix' and DNA protect buffer - a green to blue colour change following the addition of the DNA protect buffer ensured that sufficient mixing and pH of the solution has been achieved for bisulfite conversion to occur. The bisulfite conversion reaction was performed with the following conditions: denaturation for 5 minutes at 95 °C, incubation for 25 minutes at 60 °C, denaturation for 5 minutes at 95 °C, incubation for 1 hour and 25 minutes at 60 °C, denaturation for 5 minutes at 95 °C and incubation for 2 hours and 55 minutes at 60 °C. The converted gDNA underwent desulphonation and several washes through a column and eluted into two 20 uL elutions with warm buffer EB (70 °C).

DNA recovery was then measured using a Qubit 2.0 fluorometer (Invitrogen; Life Technologies).

2.2.4 Primer design

Bisulfite conversion alters the sequence and integrity of the template DNA, making accurate primer design an integral component of the workflow[149,150,152]. Initially, primer pairs were designed against the native *UBQLN2* sequence using Primer3 (Table 2) adhering to general guidelines for primer design to amplify DNA, such as a 40 – 60% GC content, length between 18-24 bases and a melting temperature between 50 – 60 °C[180]. However, following unsuccessful amplification of the *UBQLN2* CpG island, the MethPrimer program for designing primers against bisulfite-converted DNA[181] was used to design primers flanking the *UBQLN2* promoter. The native *UBQLN2* CpG island sequence was input into MethPrimer to obtain potential primer sequences. The MethPrimer program is designed to convert the native sequence to reflect bisulfite conversion when designing primers. Because assumptions cannot be made about the methylation status of cytosines at CpG sites, they were avoided when designing the primers. If CpG sites cannot be avoided, a mixed base, Y or R is used in place of the cytosine in the forward or guanine in the reverse primer, respectively[182]. Additionally, it was important to consider the length of the amplicon produced by primer sets. The highly fragmented nature of bisulfite-converted DNA may prevent amplification of longer amplicons, therefore limiting amplicon size to around 400 bp[182]. **Table 2** illustrates the five sets of primers that were designed to tile across the CpG island of *UBQLN2* to capture most of the CpG sites represented in the EWAS and MethBank 3.0 databases (the CpG site at chrX:56560144 (GRCh38) (probe ID: cg21862542) was not included due to its extreme 5' location in the north shelf) (Table 1 and Figure 3.8). To align these primers to the *UBQLN2* CpG island, the CpG island sequence was converted *in silico* to reflect the potential sequence proceeding bisulfite

conversion using Geneious Prime software (2020.0.5) – this was achieved by manually converting all non-CpG cytosines to thymine.

Table 2: Primer pairs developed for amplifying the CpG island of *UBQLN2*. Primer name is listed in a 5' – 3' direction relative to the CpG island sequence of *UBQLN2*. Primers designed in MethPrimer are suffixed with MP.

Primer Name	Primer Sequence (5' – 3')
hUBQLN2 bisulf f1	5' – TCGATTCAGGCTTTCACCCT – 3'
hUBQLN2 bisulf r1	5' – CGCCAGGCCTAGATTGATAGA – 3'
hUBQLN2 bisulf f2d	5' – GACCGAGCCAAATCTATCAATCT – 3'
hUBQLN2 bisulf r2d	5' – GAAAGAGAGAGCGCGAGGG – 3'
hUBQLN2 bisulf f3	5' – CAGGGAACCGCAGTCTTCAT – 3'
hUBQLN2 bisulf r3	5' – CACCGCGAACTCCTCTTTCT – 3'
hUBQLN2 bisulf f4	5' – TCATCAAAGTCACGGTGAAGA – 3'
hUBQLN2 bisulf r4	5' – TGAACAGTCAGCCCATCATG – 3'
hUBQLN2 bisulf f1 MP	5'- TATTGTTTTATAGATTGTTAATATATATTT-3'
hUBQLN2 bisulf r1.2 MP	5'-TTA AAA ACT TCT ACA CTA AAA AAA A-3'
hUBQLN2 bisulf f2 MP	5'-GAG TTA AAT TTA TTA ATT TAG GTT TGG-3'
hUBQLN2 bisulf r2.2 MP	5'-ACTCCCAACAACTCTAAC-3'
hUBQLN2 bisulf f3.2 MP	5'-GGT TGT TAG AGT TGT TGG-3'
hUBQLN2 bisulf r3.2 MP	5'- ATA CCT CTA TAA TAA AAA CTA-3'
hUBQLN2 bisulf f4.2 MP	5'-GTG TTT TTT TTT TTT TTT TTT TTT T-3'
hUBQLN2 bisulf r4.2 MP	5'-CCT CTT TCT CTT TAA AAA TCT-3'
hUBQLN2 bisulf f5 MP	5'-GGT GAA GAT TTT TAA AGA GAA AGA GG-3'
hUBQLN2 bisulf r5.2 MP	5'-CAA TCA ACC CAT CAT AAA T-3'

Abbreviations: *hUBQLN2;* Human *UBQLN2*, bisulf; bisulfite, f; forward, r; reverse, MP: MethPrimer

2.2.5 Methylated and non-methylated DNA controls

The Zymo Research Human Methylated & Non-methylated DNA set (catalogue number: D5014) was used to amplify the death associated protein kinase 1 (*DAPK1*). Bisulfite-specific *DAPK1* primers provided in the kit were used for PCR amplification: *DAPK1* Forward primer: 5' – TAG AAT TTA GTT AGA GGG TAG TTT AGT A – 3' and *DAPK1* Reverse primer: 5' – AAA CRA CCA ATA AAA ACC CTA CAA A – 3'. PCR amplification was conducted in the Nexus gradient Mastercycler v3.8.0.0 (Eppendorf) with the following conditions 95 °C for

10 minutes for initial denaturation, 35 cycles at 95 °C for 30 seconds, 59 °C for 45 seconds (primer annealing) and 72 °C for 60 seconds (extension) and a final extension period of 7 minutes at 72 °C, as per the manufacturer's protocol.

2.2.6 Polymerase chain reaction (PCR) amplification and gel electrophoresis

PCR amplification for sequencing was performed in multiples of 10 uL reaction volumes, with each reaction consisting of 5.4 uL of Ambion Diethyl pyrocarbonate (DEPC)-treated water, 2 uL of 5x GC buffer, 0.2 uL 10 mM dNTPs, 0.3 uL DMSO, 0.1 uL Phusion U Hot start DNA polymerase, 0.5 uL of the forward and reverse primer and 1 uL of bisulfite converted gDNA template. The thermal cycler (Nexus gradient Mastercycler v3.8.0.0; Eppendorf) was programmed to run the initial denaturation at 98 °C for 30 seconds then 40 cycles of denaturation at 98 °C for 10 seconds followed by a temperature gradient as described in **Figure 3.10** for primer annealing and 45 seconds at 72 °C for extension then 7 minutes at 72 °C for final extension. 10 uL of the PCR reaction was loaded in to a 2% agarose gel in 1x TAE buffer and examined under UV in the protein simple gel reader v2.2.0.1 (Cell Biosciences) following gel electrophoresis at 100 V for approximately 1 hour.

For PCR amplification of bisulfite-converted patient-derived fibroblast gDNA, three primer sets were used: Set 1 *hUBQLN2* bisulf f1 MP and *hUBQLN2* bisulf r2.2 MP; Set 2: *hUBQLN2* bisulf f3.2 MP and *hUBQLN2* bisulf r4.2 MP; Set 3: *hUBQLN2* bisulf f4.2 MP and *hUBQLN2* bisulf r5.2 MP.

2.2.6.1 PCR Purification

The Macherey Nagel NucleoSpin Gel and PCR clean-up kit (catalogue number: 740609.50) was used for PCR purification. Briefly, 100 uL of the PCR reaction was added to 200 uL of Buffer NTI in a spin-column and centrifuged for 30 seconds at 11,000 x g. The filtrate was placed in the column and spin again to ensure DNA-binding to the silica membrane. Two wash

steps using 700 uL of Buffer NT3 followed by centrifuging for 30 second at 11,000 x g were conducted. The membrane was dried (centrifuged for 2 minutes at 11,000 x g) and incubated at 70 °C for 2 minutes and before performing two elution steps with 15 uL of DEPC-treated water at 70 °C with a 5-minute incubation in between elutions.

2.2.7 Sanger sequencing and quantification analysis

PCR amplified samples were sent for Sanger sequencing at Auckland Genomics at a concentration of 5 ng/uL – 20 ng/uL (2.5 ng of gDNA per 100 bp because the amplicons were smaller than 1 kB). Additionally, relevant primer pairs were sent at a concentration of 5 pmol/uL at 5 uL per reaction. Sanger sequencing was run as per Auckland Genomics protocol and provided as a chromatogram readout.

Sanger sequencing chromatograms are semi-quantitative measures in which peak intensity indicates the proportion of sequenced DNA molecules that incorporates a particular base at the same position[183]. R-studio (version 4.1) was used to extract and graph relative signal intensities of each nucleotide as a percentage of the total signal reported at each base, by using a script written by Scotter Lab PhD candidate Miran Mrkela (Appendix A).

3 RESULTS

3.1 PUBLICLY AVAILABLE METHYLOME DATA PROVIDES COVERAGE ACROSS THE CPG ISLAND OF THE *UBQLN2* GENE

To determine the coverage of CpG sites across the CpG island of *UBQLN2* with the Illumina Infinium methylation assay probes, it was necessary to align the probe sequences to the *UBQLN2* CpG island sequence using Geneious Prime (2020.0.5). **Figure 3.1** depicts the specific CpG site targeted at the start or the end of the Illumina Infinium methylation assay probes[184]. It is worth noting that methylation status of all CpG sites in the CpG island of *UBQLN2* were not assessed in the Illumina Human Methylation 450K BeadChip, however the distribution CpG sites examined covers the north shelf to CpG sites beyond the start codon in the CpG island proper (Table 1 and Figure 3.1). The comprehensive coverage of the *UBQLN2* CpG island allows for accurate detection of methylation patterns between sexes, CpG sites and tissues.

3.2 COMPARING SEX-SPECIFIC METHYLATION OF INDIVIDUAL CPG SITES IN THE *UBQLN2* CPG ISLAND *WITHIN* TISSUE TYPES

3.2.1 CpG sites in the *UBQLN2* CpG island are more highly methylated in females than males within a given tissue type

To determine which CpG sites were differentially methylated in the *UBQLN2* gene between males and females, it was necessary to obtain methylome information from publicly available databases. Methylome data from EWAS and Methbank 3.0 demonstrated differential methylation across CpG sites in the promoter region of the *UBQLN2* gene, with average β-values ranging from 0.06 – 0.09 in males and 0.27 – 0.5 in females within a given tissue type (excluding the CpG site at 56560144; probe ID: cg21862542) (Figure 3.2 and 3.3). For most CpG sites, the average β-value was higher for females than males within all tissue types; CpG

sites located at ChrX: 56563335, 56563359, 56563485, 56563513, 56563516, 56563518, 56563664, 56563680, 56563754, 56563929, 56563953, 56564080 and 56564104 bp (probe ID: cg02587153, cg10009003, cg25697726, cg09013091, cg14824933, cg23993375, cg10950413, cg09666011, cg16161440, cg02492740, cg01780361, cg04270663 and cg20979765, respectively) displayed a significantly higher average β-value in females when compared to males in the brain dermal fibroblasts, nasal epithelium, saliva, skin, tongue and whole blood of the EWAS data (paired t-test, p-value < 0.001) (Figure 3.2) A statistically significant difference was also observed at these sites across all brain regions, dermal fibroblasts, peripheral blood mononuclear cells, saliva and skin of the MethBank data (unpaired t-test, p-value < 0.0001) (Figure 3.3). Because methylation patterns across different brain regions obtained from MethBank 3.0 were highly, data from the 'Brain' is used to represent these regions. **Appendix B** illustrates the relationship between male and female β-value at each CpG site across the *UBQLN2* CpG island.

Differential methylation between males and females is present in the CpG island of the *AR* gene, also subjected to X-inactivation. The average β-value in males ranged from 0.34 – 0.4 and 0.47 - 0.58 in females, across all tissue types. Removing the average β-value data from CpG sites located at 66736034, 66736576, 66761656, and 66761665 bp (Probe ID: cg26003280, cg25686125, cg26498015 and cg15840039 respectively) (Table 1) to mitigate the potential skew from high methylation levels relative to the other sites, gives an average β-value of 0.027 – 0.04 in males and 0.35 – 0.43 in females, values similar to average β-value for *UBQLN2*. Like *UBQLN2*, females displayed significantly higher levels of methylation across most probes in all tissue types when compared to males (unpaired t-test, p-value < 0.001) (Figure 3.2 and 3.4). No significant difference in methylation was observed in the CpG site located at 66761656 bp (probe ID: cg26498015) in the brain, 66736576 bp (probe ID:
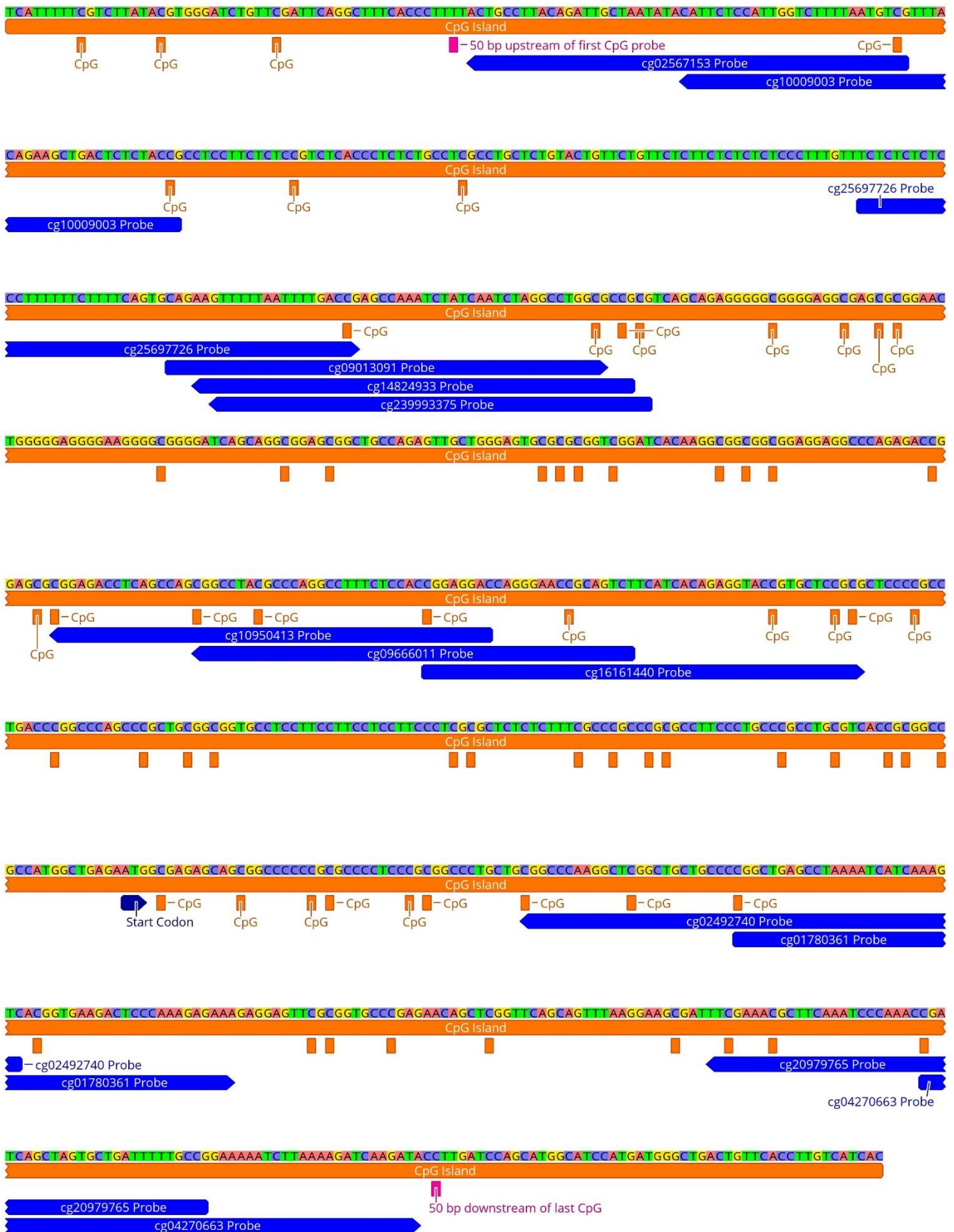
Figure 3.1: Illumina Infinium assay probes provides coverage across the CpG island of the human *UBQLN2* gene. Each Illumina Infinium methylation assay probe provides methylome information for a single CpG site located at the 5' or 3' end of its sequence when aligned to *UBQLN2* (GRCh37). The 5' most probe, cg21862542 located at chrX:56586577 (GRCh37) and chrX:56560144 (GRCh38) has been omitted from this diagram. Annotations were aligned to the CpG island sequence of *UBQLN2* using Geneious version 2020.0 created by Biomatters.

cg25686125) in nasal epithelium as well as 66736034 and 66736576 bp (probe ID: cg26003280 and cg25686125) in the tongue samples.

Methylation differences were also observed in the CpG island of the *CDK16* gene between males and females, with the average β-value ranging from 0.085 – 0.06 and 0.07 – 0.169 respectively (β-value of probes located at 47047850 and 47055307 bp in the north and south shelf of the island were not included) (Table 1). While the average β-value of a given probe in the CpG island proper of *CDK16* in males was similar to the average β-value for males across the *UBQLN2* CpG island across all tissue types, average β-values for females in the *CDK16* CpG island were lower than CpG sites in the island of *UBQLN2* (excluding the previously defined highly methylated CpG sites in the shores and shelves). For example, EWAS methylome data from saliva samples demonstrates a significant difference in average β-value in the CpG island of *UBQLN2* (average β = 0.3942) and the CpG island of *CDK16* (β =0.1382) between females (unpaired t-test; p -value < 0.0001), while there is no significant difference between the average β-value of the CpG island of *UBQLN2* (average β = 0.1059) and *CDK16* (average β = 0.0709) between males (unpaired t-test; p-value = 0.3701).

No significant difference in methylation of CpG sites in the CpG island of *TARDBP* (autosomal control) was observed between males and females across a given tissue type indicating that sex-specific methylation is unlikely to affect expression, with an average β-value of 0.32 - 0.35 in males and 0.33 – 0.35 in females.

3.2.2 CpG sites in the north shelf of the *UBQLN2* CpG island are more highly methylated than probes in the CpG island proper

Interestingly, the CpG site located on the north shelf of the *UBQLN2* CpG island at chrX:56560144 bp (GRCh38) which corresponds to probe ID cg21862542, demonstrated a higher average β-value than other CpG sites in males and females. The difference was significant in brain, dermal fibroblasts, saliva, skin, tongue and whole blood methylome data from EWAS (unpaired t-test, p-value < 0.0001) (Figure 3.2). Furthermore, males were more likely to have a higher average β-value than females, ranging from 0.88 – 0.97 in males and 0.65 – 0.79 in females across a given tissue type. A statistically significant elevation in methylation was observed in males compared to females in brain, dermal fibroblasts, skin, tongue and whole blood methylome data from EWAS (unpaired t-test, p-value < 0.01). While a similar pattern of higher average β-value for cg21862542 was seen in the data collected from Methbank 3.0, statistical significance could not be determined as the raw data was provided as an average.

A similar pattern was also observed in the methylome data for the control genes where CpG sites located approximately 4 kb upstream and downstream of the CpG island (north and south shelf, respectively), as defined in the Infinium Human Methylation manifest file v1.2, tend to be more highly methylated in both males and females compared to other CpG sites. Furthermore, significantly higher levels of methylation were observed in males compared to females for CpG sites with a chromosomal location of chrX: 66736034, 66736576, 66761656 and 66761665 in the androgen receptor gene, also subject to X-inactivation, across all tissue types (Figure 3.2 and 3.3). The average β-value for ChrX: 66736034 = 0.976 (male) versus 0.852 (female) (p-value = 0.0489), ChrX: 66736576= 0.962 (male) versus 0.812 (female) (p-value = 0.0004), ChrX: 66761656 = 0.961 (male) versus 0.891 (female) (p-value = 0.0174) and ChrX: 66761665 = 0.814 (male) versus 0.612 (female) (p-value = 0.0003) (Figure 3.2).

However, this pattern is not limited to genes that undergo XCI as methylome data from *CDK16* and *TARDBP* also demonstrate elevated levels of methylation at CpG sites located at ChrX: 47047850, 47053142 and 47055307 bp (*CDK16*) and Chr1: 11045155, 11047475 and 1158786 (*TARDBP*) (Figure 3.2 and 3.3). In contrast to XCI affected genes (*UBQLN2* and *AR*), no significant difference is present between males and females at these CpG sites in the CpG island of *CDK16* and *TARDBP*. Average β-value, across all tissue types, for ChrX: 47047850 = 0.950 (males) and 0.9575 (females) (p-value = 0.7716), ChrX: 47053142 = 0.3776 (males) and 0.4422 (females) (p-value = 0.3046), ChrX: 47055307 = 0.9243 (males) and 0.9053 (females) (p-value = 0.3012), Chr1: 11045155 = 0.9428 (males) and 0.9431 (females) (p-value = 0.8686), Chr1: 11047475 = 0.7035 (males) and 0.6916 (females) (p-value = 0.5837) and Chr1: 1158786 = 0.9339 (males) and 0.9301 (females) (p-value = 0.8583).
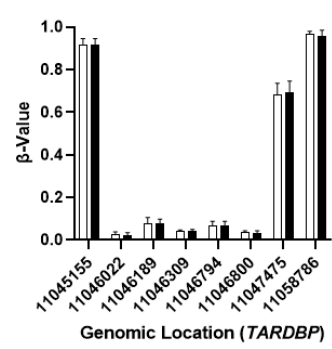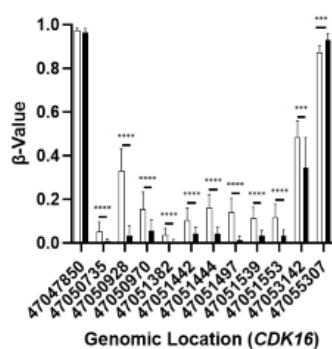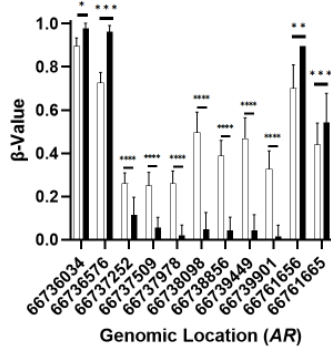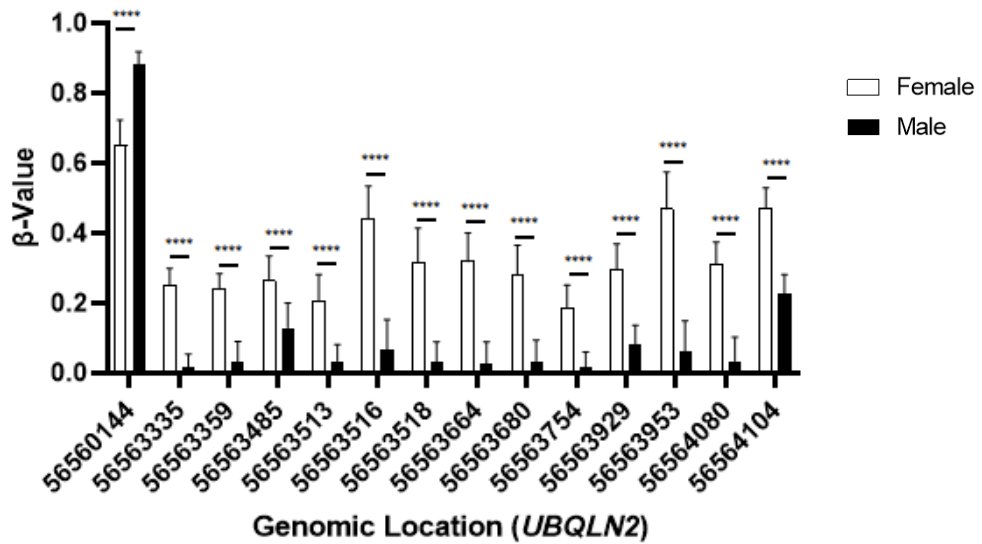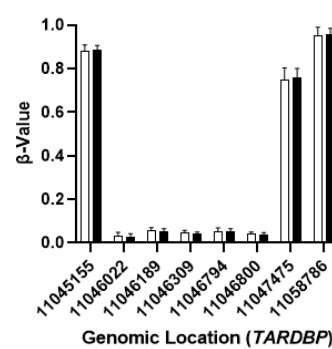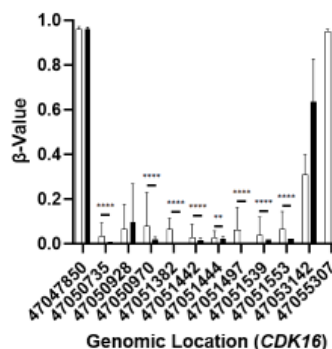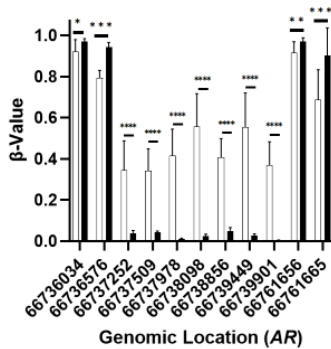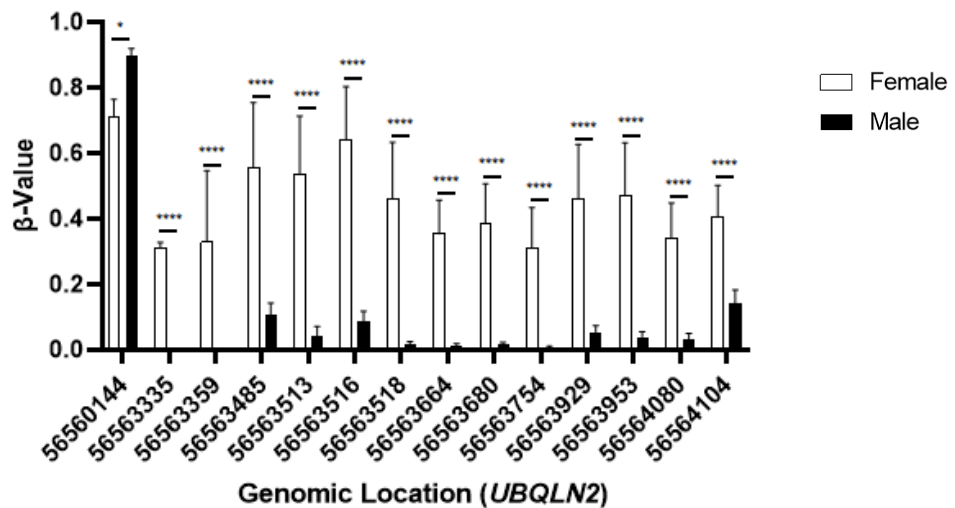
EWAS: Nasal Epithelium

EWAS: Saliva
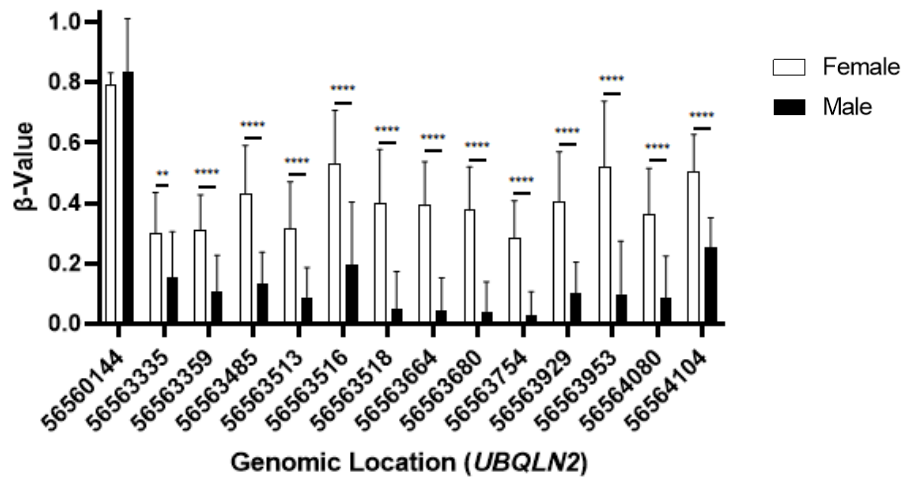
EWAS: Skin

Genomic Location (*UBQLN2*)

Genomic Location (*AR*)

Genomic Location (*CDK16*)

Genomic Location (*TARDBP*)
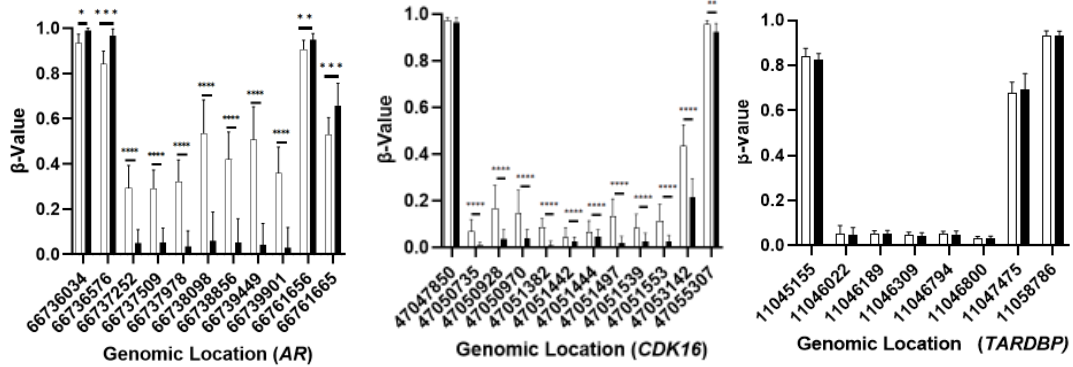
EWAS: Tongue

Genomic Location (*UBQLN2*)

Genomic Location (*AR*)

Genomic Location (*CDK16*)
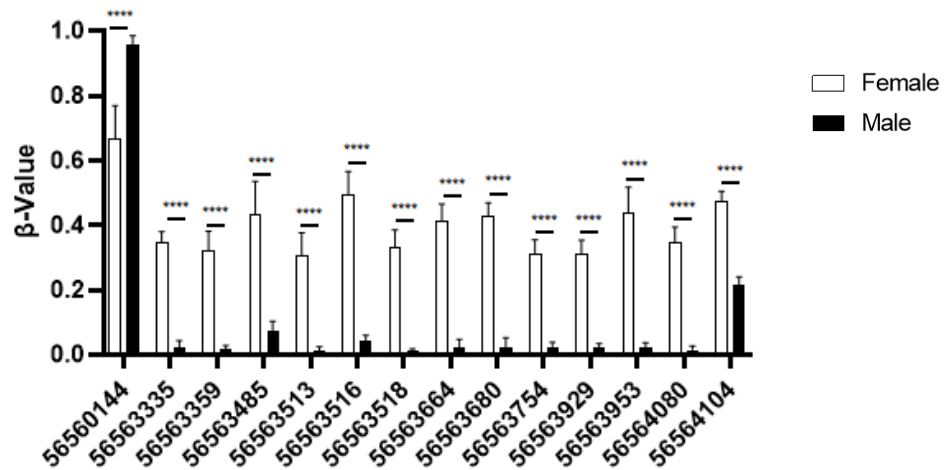
Genomic Location (*TARDBP*)
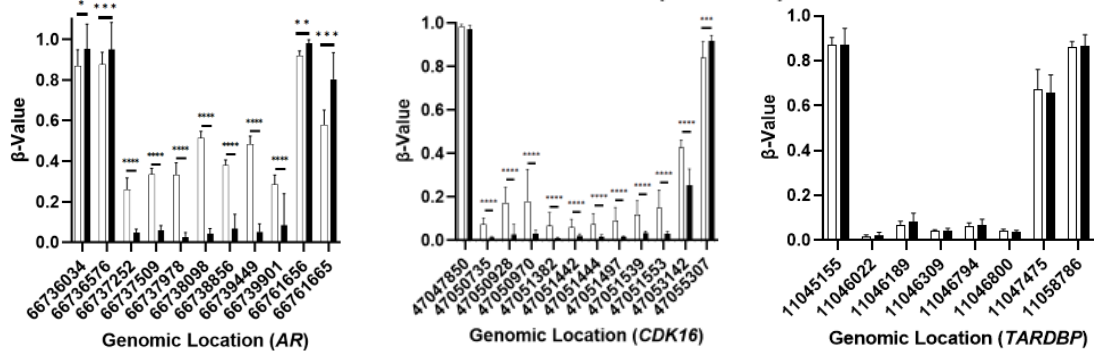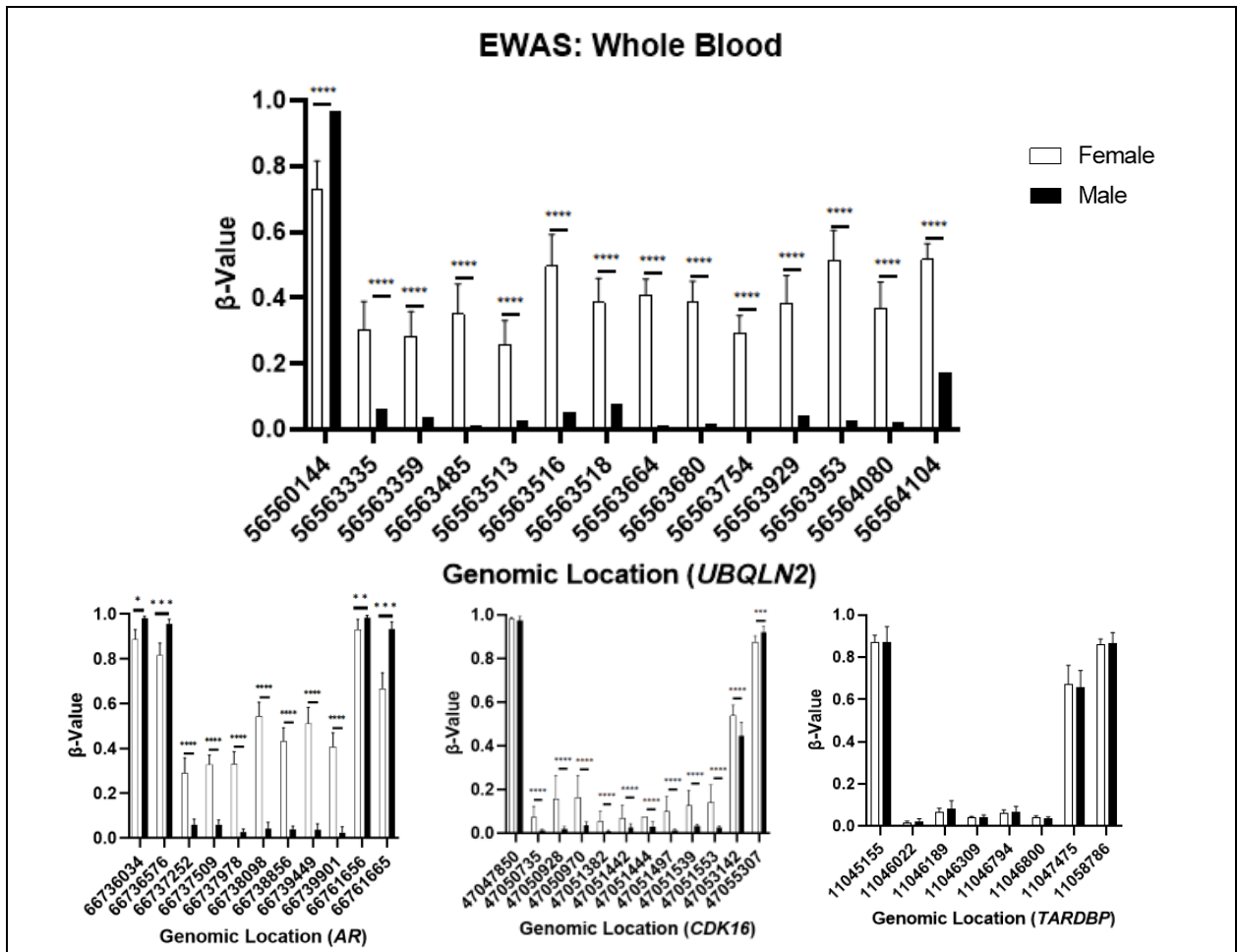
Figure 3.2 Methylome data from EWAS demonstrates elevated methylation of CpG sites in the promoter of *UBQLN2* in females across different tissue type. Average β -value for different tissue types demonstrates differential methylation between males and females at CpG sites in the promoter region of the *UBQLN2* gene. Methylome data for CpG sites in the CpG island of the androgen receptor (*AR*), *CDK16* and *TARDBP* are included as controls for CpG site methylation in X-inactivation, escape from X-inactivation and autosomal genes, respectively. Genomic location of the CpG sites is organised in a 5' to 3' direction. Data is presented as average β -value ± SD. Raw data was obtained from EWAS and processed in GraphPad Prism 9.0.2. * $p > 0.05$, ** $p > 0.01$, *** $p > 0.001$, **** $p > 0.0001$. (Note: methylome data for the *UBQLN2* CpG site at chrX:56560144 (GRCh38) probe ID: cg21862542 was not available in the Illumina Infinium methylation manifest file v1.2 for nasal epithelium and was therefore omitted).
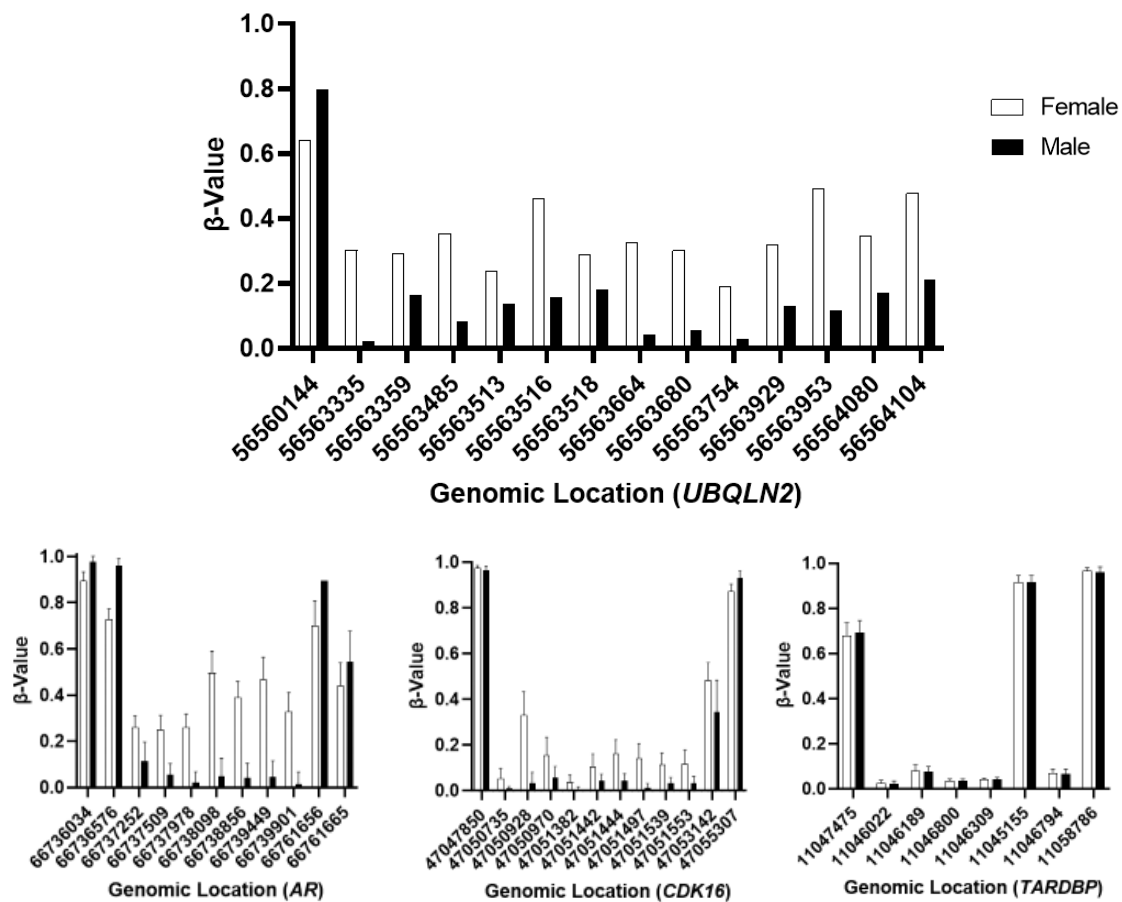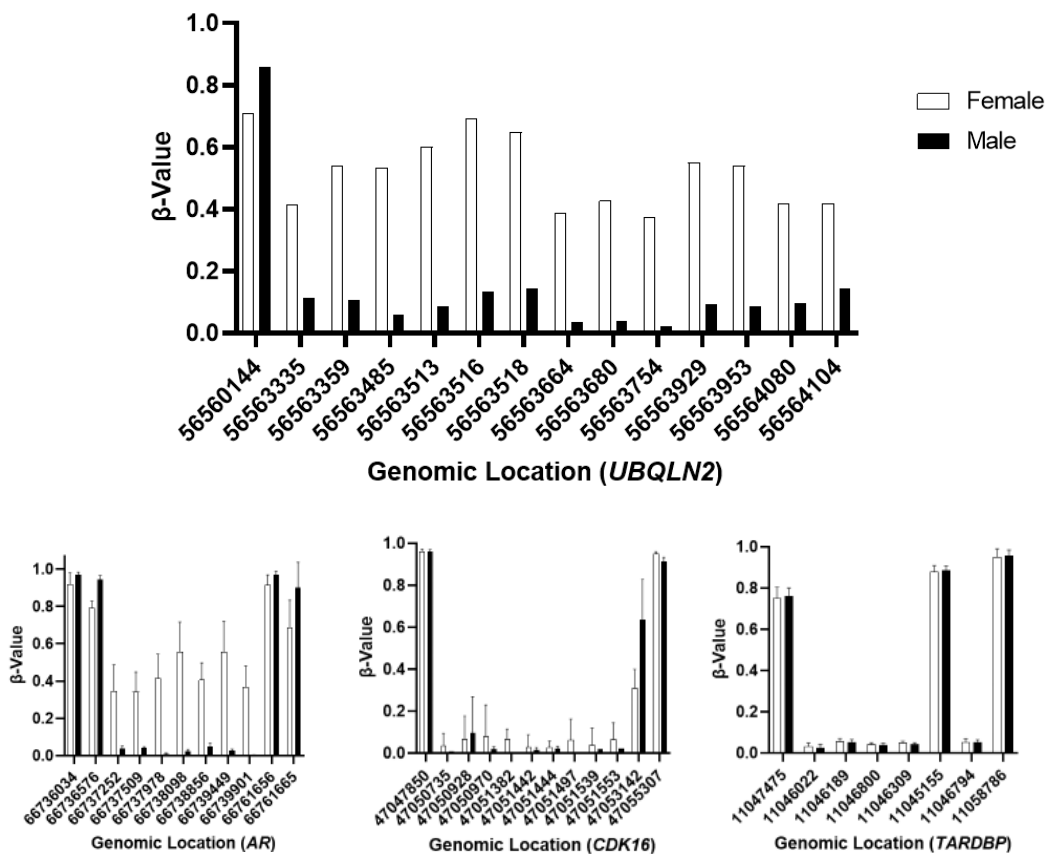
MethBank: Peripheral Blood Mononuclear Cells

Genomic Location (*UBQLN2*)

Genomic Location (*AR*)

Genomic Location (*CDK16*)

Genomic Location (*TARDBP*)

MethBank: Saliva

Genomic Location (*UBQLN2*)

Genomic Location (*AR*)

Genomic Location (*CDK16*)

Genomic Location (*TARDBP*)
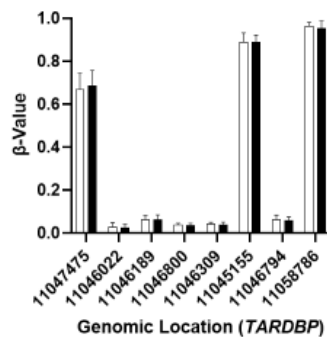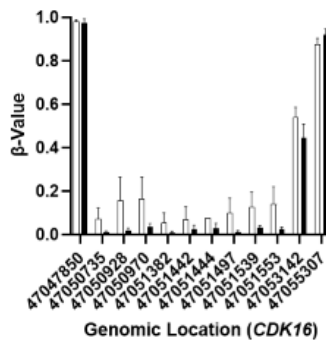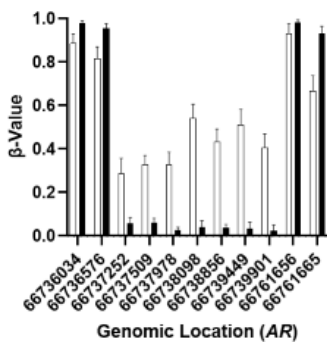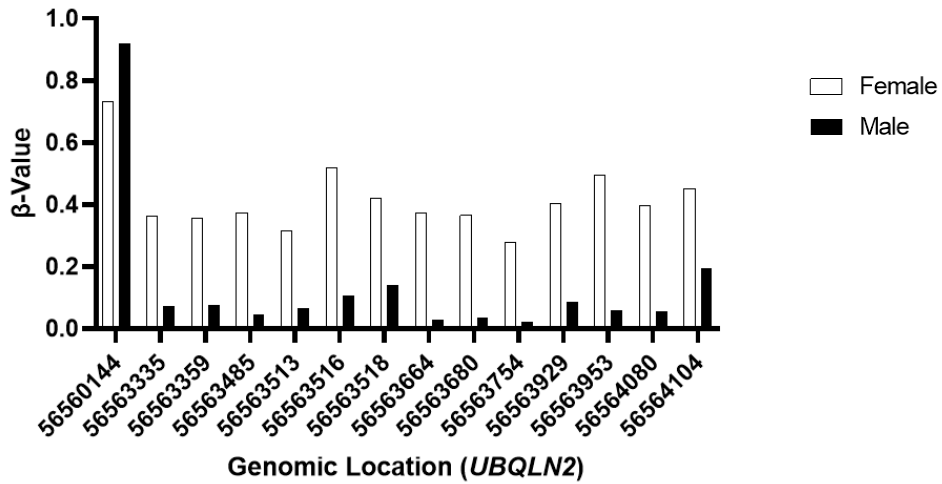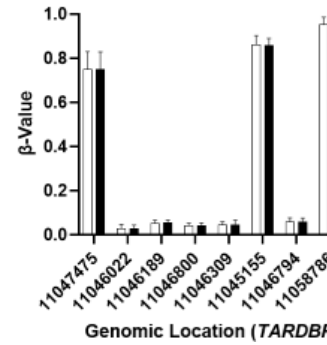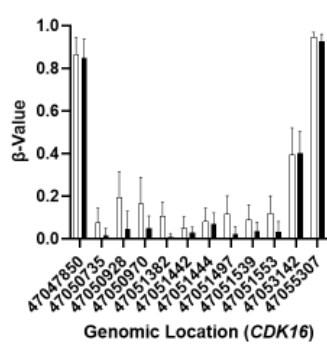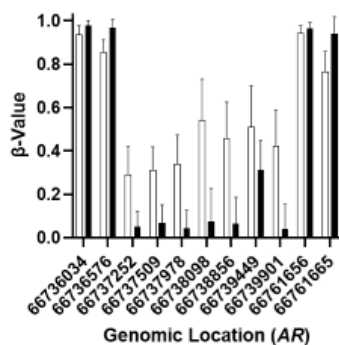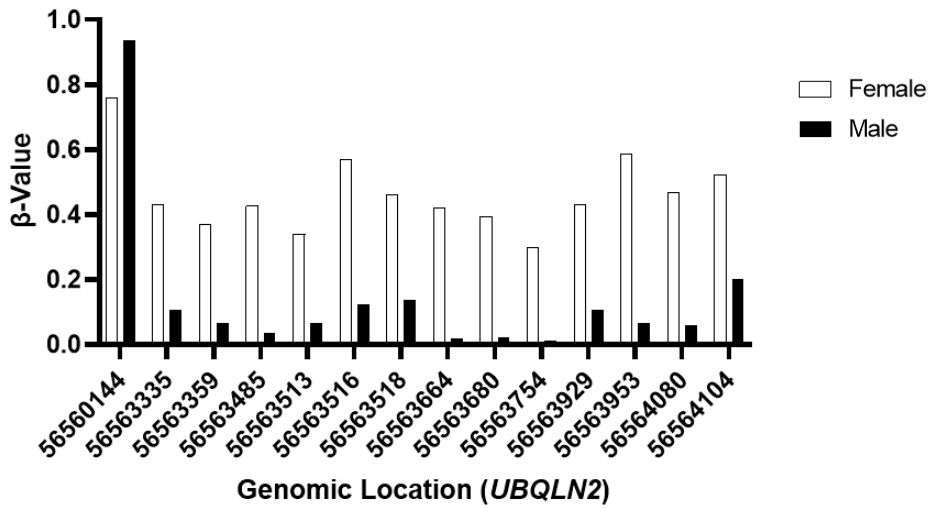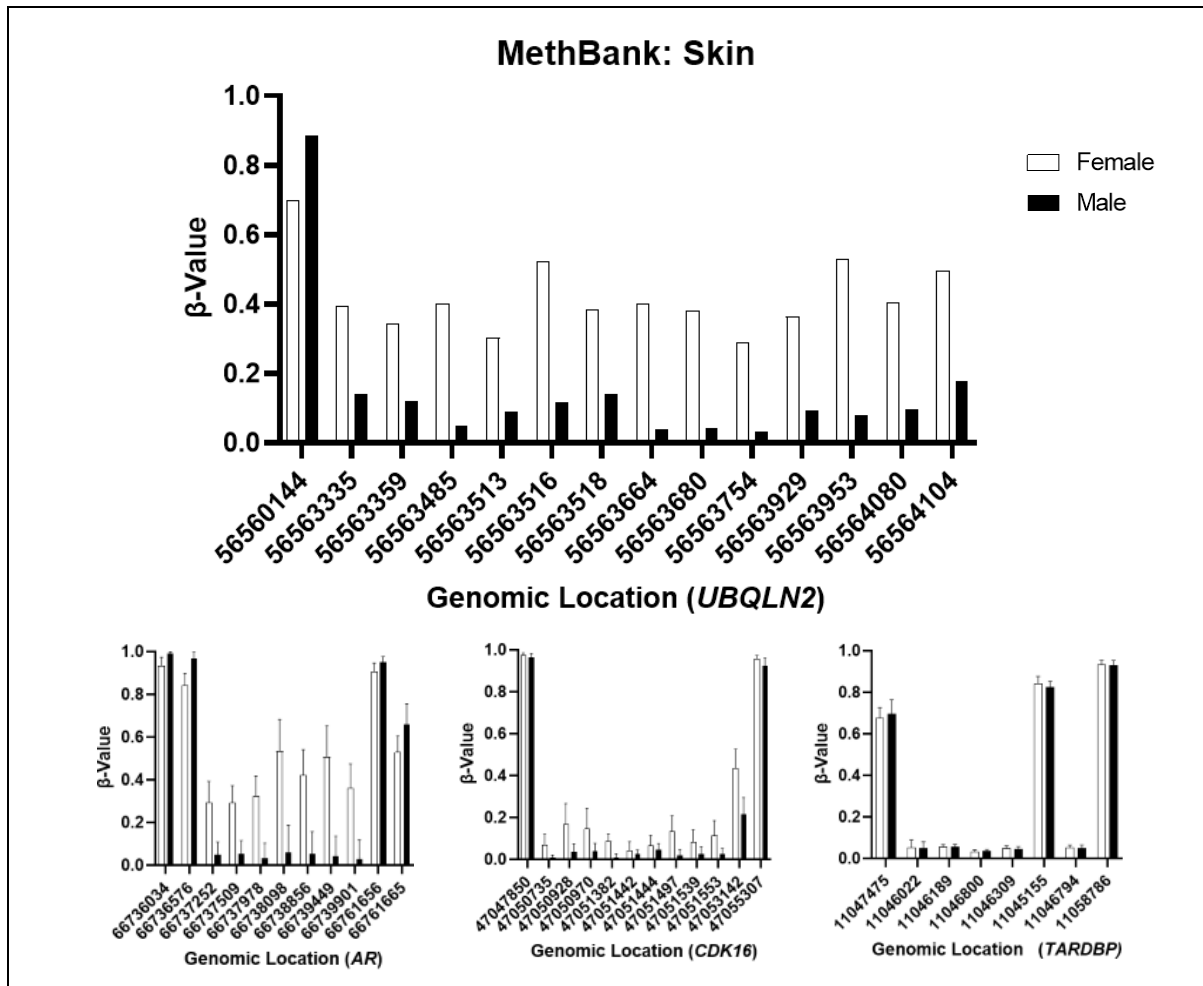
Figure 3.3 Methylome data from MethBank 3.0 demonstrates elevated methylation of CpG sites in the promoter of *UBQLN2* **in females regardless of tissue type.** Average β -value for different tissue types demonstrates differential methylation between males and females at CpG sites in the promoter region of the *UBQLN2* gene. Methylome data for CpG sites in the CpG island of the androgen receptor (*AR*), *CDK16* and *TARDBP* are included as controls for CpG site methylation in X-inactivation, escape from x-inactivation and autosomal genes, respectively. Genomic location of the CpG sites is organised in a 5' to 3' direction. Data is presented as average β-value ± SD. Raw data was obtained from MethBank 3.0 and processed in GraphPad Prism 9.0.2.

3.2.3 Combined data from all tissue types confirms elevated methylation of CpG sites in the promoter of *UBQLN2* in females

The pattern of methylation across the CpG island of *UBQLN2* remained consistent between tissue types, therefore, we combined data from all tissue types to conduct statistical analysis of methylation between individual *UBQLN2* CpG sites (Figure 3.4). The overall pattern of methylation is visually conserved between the individual pattern of methylation across the CpG island in each tissue and the combined data from all tissue types (Figure 3.2 - 3.4). The combined data confirms that females tend to exhibit elevated methylation of CpG sites in the island of *UBQLN2,* when compared to males, particularly sites located at ChrX: 56563516, 56563953 and 56564104 bp. However, 2-way ANOVA (Multiple comparisons) analysis did not establish a significant difference between the average β-value of CpG sites at ChrX: 56563516, 56563953 and 56564104 bp and any other CpG site between females (p-value < 0.9999), except ChrX: 56560144 located in the north shelf, in the EWAS and MethBank data. While these sites appear highly methylated in females compared to other sites in the *UBQLN2* CpG island proper, the difference is not statistically significant.

Combined data from control methylome data also reflects the methylation patterns observed across different tissue types.
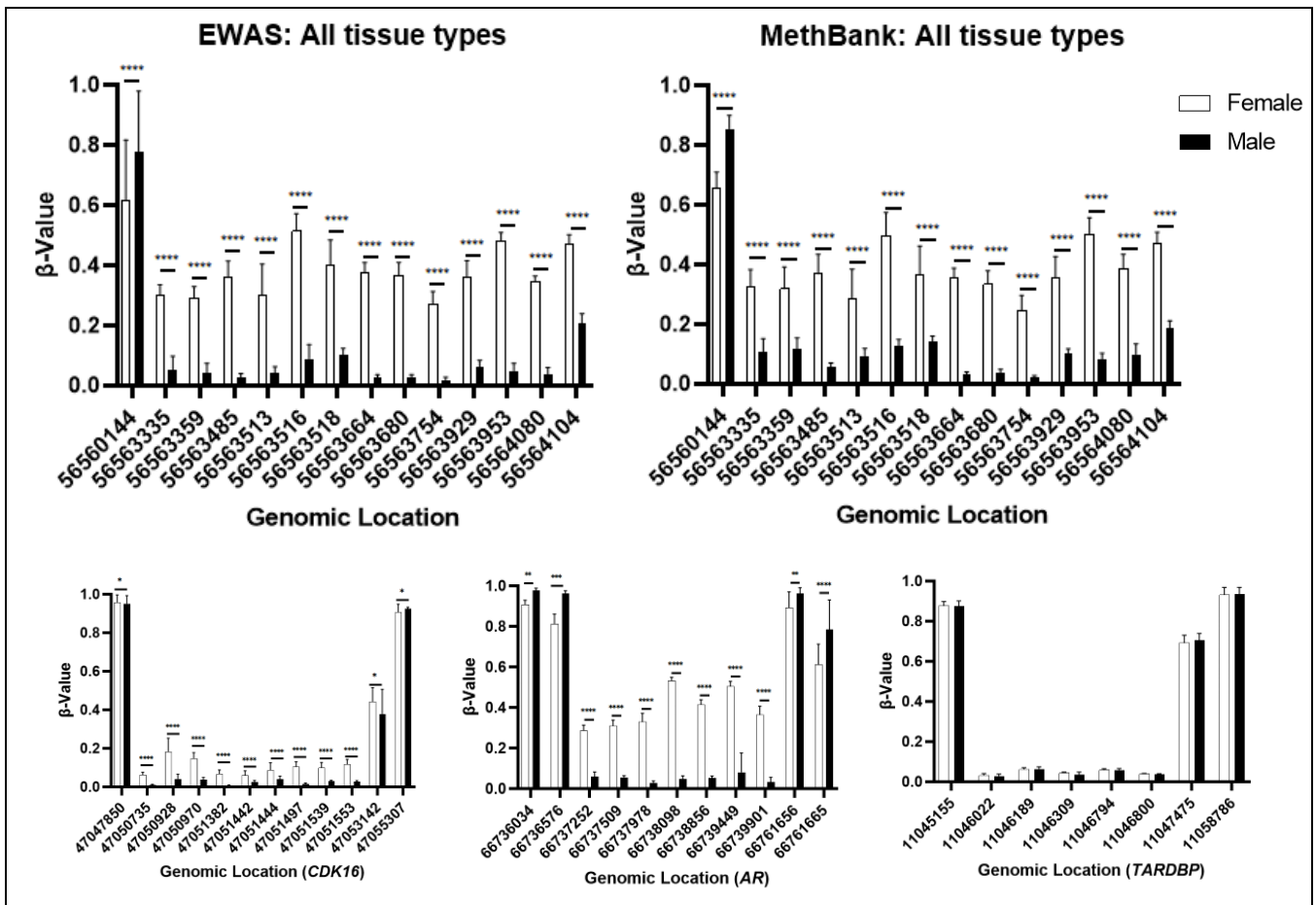
Figure 3.4: Combined data from all tissue types confirms elevated methylation of CpG sites in the promoter of *UBQLN2* in females, particularly sites 56560144, 56563516, 56563953 and 56564104 bp (GRCh38). A similar pattern of methylation across the CpG island of *UBQLN2* is observed between methylome data from EWAS and Methbank. Methylome data, from EWAS, for CpG sites in the CpG island of the androgen receptor (*AR*), *CDK16* and *TARDBP* are included as controls for CpG site methylation in X-inactivation, escape from X-inactivation and autosomal genes, respectively. Genomic location of the CpG sites is organised in a 5' to 3' direction. * p > 0.05, ** p > 0.01, *** p > 0.001 **** p > 0.0001. Data is presented as average β-value ± SD. Raw data was obtained from EWAS and MethBank 3.0 and processed in GraphPad Prism 9.0.2.

3.3 COMPARING SEX-SPECIFIC METHYLATION OF THE WHOLE *UBQLN2* CPG ISLAND *BETWEEN* TISSUE TYPES

Tissue-specific differences in methylation were also assessed for each probe between males and females to ascertain whether the level of methylation of a particular CpG site was consistent between tissue types. **Appendices C and D** demonstrate the average β-value of a given CpG site from EWAS and MethBank across different tissues. All probes tend to display a consistent level of sex-specific methylation regardless of tissue type, suggesting that tissue type does not affect the level of methylation. Interestingly, methylome data from female dermal fibroblasts shows a higher average β-value across some CpG sites in *UBQLN2* based on methylome data from EWAS (Appendix C).

3.3.1 Combined data from all CpG sites in the promoter of *UBQLN2* confirms elevated methylation in females regardless of tissue type.

To determine if there were any tissue-specific differences in methylation across the CpG island of the *UBQLN2* in males and females, statistical analysis of combined β-value of all probed CpG sites along the promoter region of *UBQLN2* from EWAS and MethBank was conducted. The average β-value across the CpG island for a given tissue type ranged from 0.059 – 0.116 in males and 0.269 – 0.502 in females across EWAS and MethBank; with the average β-value for dermal fibroblasts equivalent to 0.074 in males and 0.476 in females. When compared to the combined average β-value of females for a given tissue type, the elevated methylation of the *UBQLN2* promoter in female dermal fibroblasts was not statistically significant (Figure 3.5, 2-way ANOVA: Multiple comparisons p-value < 0.99)
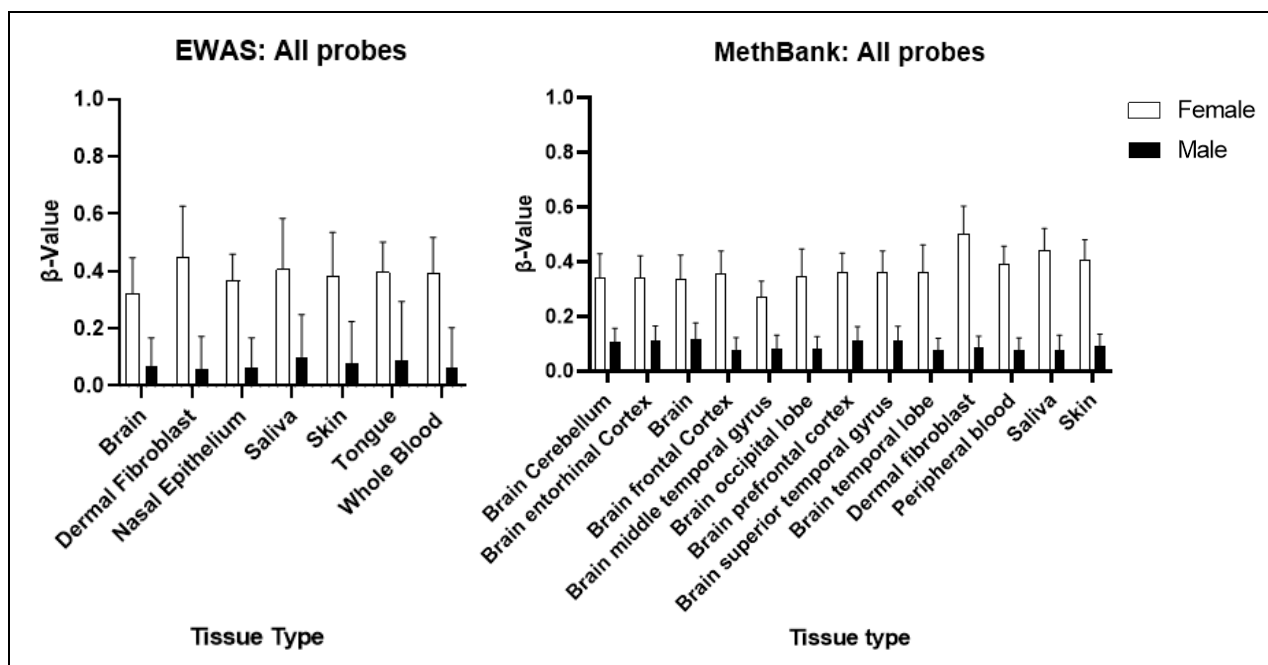
Figure 3.5: Average methylation across the CpG island is elevated in females regardless of tissue type. Average β-value of all probes across the CpG island of *UBQLN2* from EWAS and MethBank 3.0 demonstrates consistently elevated methylation in females relative to males in all tissue types. Genomic location of the CpG sites is organised in a 5' to 3' direction. Data is presented as average β-value ± SD. Raw data was obtained from EWAS and MethBank 3.0 and processed in GraphPad Prism.

3.4 DEVELOPING IN-HOUSE METHODS FOR ANALYSIS OF CPG ISLAND METHYLATION OF *UBQLN2*

To determine the methylation pattern in the *UBQLN2* CpG island from patient-derived fibroblast cells, optimising bisulfite conversion and PCR amplification is critical. Bisulfite conversion of the extracted gDNA is the first step to successfully encode methylation patterns across the promoter of *UBQLN2* and maintain epigenetic integrity (Figure 3.6)[150]. Forward and reverse primers, designed against the bisulfite-converted CpG island sequence of *UBQLN2*, amplify the DNA for Sanger sequencing (Figure 3.6). Sanger sequencing data for heterozygous *UBQLN2* female CpG islands should have a prominent cytosine read-out at methylated cytosine positions coupled with a smaller thymine signal, indicative of X-inactivation in a subset of cells.
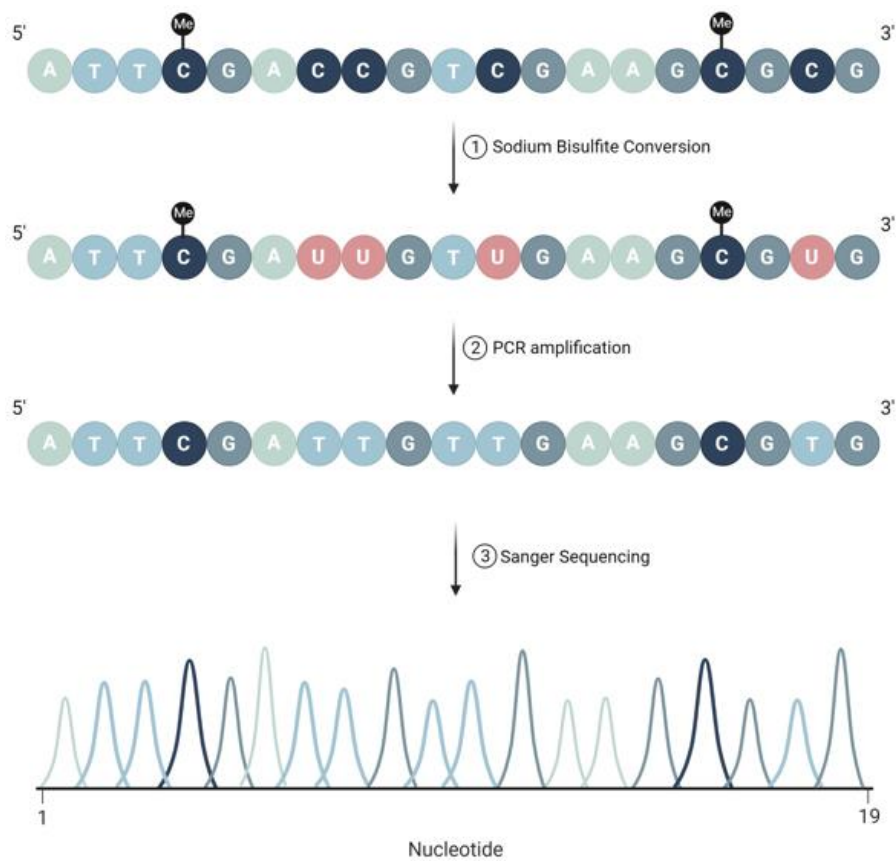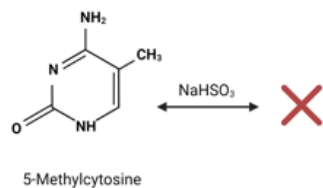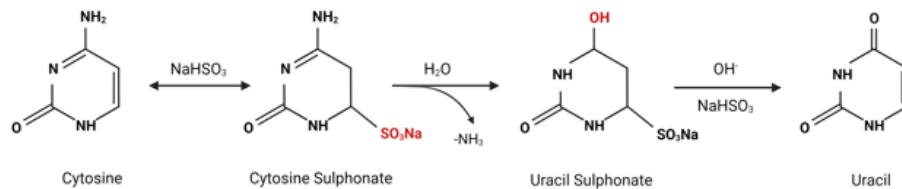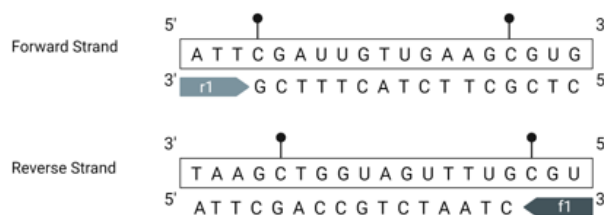
Figure 3.6: Schematic of bisulfite conversion workflow. Observing the methylation pattern of gDNA from patient derived fibroblasts involves three main techniques: (1) Sodium bisulfite conversion; a desulfonation reaction that converts unmethylated cytosines to uracil through cytosine sulphonate and uracil sulphonate intermediates. Formation of these intermediates do not occur when cytosine is methylated at the fifth carbon. (2) PCR amplification; forward (f1) and reverse (r1) primers designed against the converted sequence amplify the region of interest for subsequent sanger sequencing (3). Created with BioRender.com.

3.4.1 QIAGEN EpiTect bisulfite kit shows higher DNA recovery rate following bisulfite conversion

First, the efficiency of two commercially available bisulfite conversion kits, EZ DNA Methylation Gold kit from Zymo research and the QIAGEN EpiTect bisulfite kit was assessed using gDNA derived from nine patient-derived fibroblast cells, HEK293T cells and HeLa cells.

Subjecting 2000 ng of gDNA from HEK293T cells, HeLa cells and patient fibroblasts from case MND055 to bisulfite conversion using the EZ DNA Methylation Gold kit, resulted in a recovery of 190, 221 and 32.1 ng of gDNA, respectively (Table 3). Input of 500 ng of gDNA from MN17 patient-derived fibroblasts cells retained 12 ng of DNA after bisulfite conversion, with a recovery rate of 2.4%.

For the QIAGEN EpiTect bisulfite kit, Similarly, DNA recovery following conversion of 2000 ng of gDNA from all nine patient-derived fibroblast lines, HEK293T and HeLa cells ranged from 164 ng – 348.8 ng, with a recovery rate of 8.2% - 17.44% (Table 3).

Both kits caused a dramatic loss in DNA concentration following conversion, however it should be noted that measurements of DNA concentrations using a Qubit 2.0 fluorometer are more sensitive and specific for DNA, especially at low concentrations thus providing a more accurate measure of DNA concentration following bisulfite conversion than Nanodrop 2000[185,186]. As a direct comparison, conversion of 2000 ng of gDNA from fibroblasts derived from MND055 yielded 3.21 ng/uL of DNA following conversion using the EZ DNA Methylation Gold kit and 5.96 ng/uL, when using the QIAGEN EpiTect bisulfite kit.

Table 3: DNA retention following bisulfite conversion. Poor DNA recovery from the EZ DNA methylation gold kit informed the decision to refrain from converting DNA from all fibroblast lines using this kit.

Cell Type/Case	EZ DNA Methylation Gold Kit			QIAGEN EpiTect Bisulfite kit		
	Input gDNA (ng)	Bisulfite Converted gDNA (ng)*	Recovery (%)	Input gDNA (ng)	Bisulfite Converted gDNA (ng) †	Recovery (%)
HEK293T	2000	190	9.5	2000	220	11
HeLa	2000	221	11.05	2000	263.2	13.16
MN17	500	12	2.4	2000	164	8.2
MND033	-	-	-	2000	318.4	15.92
MND045	-	-	-	2000	204	10.2
MND046	-	-	-	2000	244	12.2
MND053	-	-	-	2000	192.8	9.64
MND054	-	-	-	2000	348.8	17.44
MND055	2000	32.1	1.605	2000	238.4	11.92
MND056	-	-	-	2000	284.8	14.24
MND057	-	-	-	2000	226.4	11.32

* Measured using Nanodrop 2000
† Measured using Qubit 2.0 fluorometer
'-' Data not available

3.4.2 Validation methods confirm bisulfite conversion of gDNA

It was important to confirm that bisulfite conversion was occurring before amplifying gDNA for subsequent sequencing. Bisulfite conversion is a harsh procedure capable of fragmenting DNA to sizes around 100 – 1500 bp[187]. To validate bisulfite conversion, gel electrophoresis-based methods were employed such as the visualisation of DNA 'smears' (Figure 3.7A and B) and the use of a bisulfite-converted-template-specific primer (Figure 3.7C). Gel electrophoresis of untreated (non-bisulfite-converted) gDNA produced DNA fragments ranging from 1000 bp to larger than 15,000 bp (Figure 3.7A), compared to bisulfite-treated gDNA using the EZ DNA Methylation Gold Kit that produced fragments around 100 bp to 1500 bp (Figure 3.7B) suggesting that bisulfite conversion has taken place.

Additionally, fully methylated and fully unmethylated gDNA standards from Zymo Research were also used to validate bisulfite conversion using the QIAGEN EpiTect bisulfite kit. Following bisulfite conversion of the DNA controls, primers designed to only bind when successful bisulfite conversion has occurred amplified *DAPK1* to produce an amplicon around

268 bp. **Figure 3.7C** demonstrates the presence of a strong band around 268 bp, confirming that successful bisulfite conversion is occurring with the QIAGEN EpiTect kit.
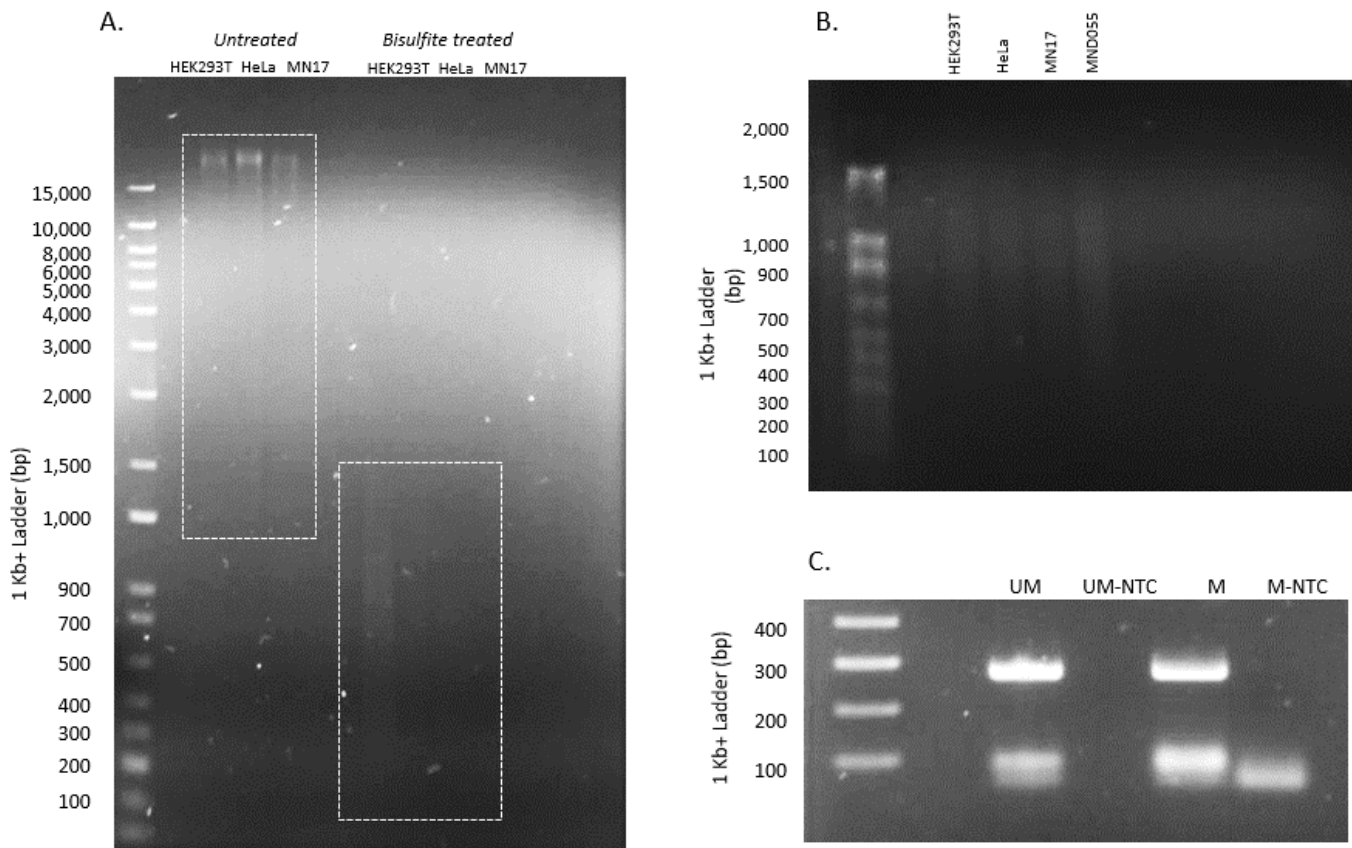
Figure 3.7: Visual confirmation of bisulfite conversion of gDNA using the (A, B) EZ DNA Methylation Gold kit and (C) QIAGEN EpiTect bisulfite kit. Native gDNA (untreated) or bisulfite converted gDNA was loaded into each lane (A); Bisulfite converted gDNA was loaded into lanes one through four (B). 10 uL of PCR products from amplifying *DAPK1* using bisulfite-conversion specific primers in fully unmethylated (UM) and fully methylated (M) gDNA templates was visualised in 2% agarose gel (C). Abbreviations: UM: Unmethylated gDNA, M: Methylated gDNA, NTC: Non-template control.

3.4.3 Bisulfite conversion workflow is highly destructive to and poorly retentive of DNA

Accuracy and reliability of patient-derived methylome data depends heavily on the quality and quantity of DNA available for Sanger sequencing. Previous results from **Table 3** and **Figure 3.7** demonstrate that bisulfite conversion causes notable DNA degradation and loss of DNA concentration. Following DNA extraction from patient-derived fibroblast cells, bisulfite conversion is the first step in the workflow defined in **Figure 3.6** and defines the starting DNA concentration for downstream analysis. **Table 4** presents the sequential loss of DNA across

three replicates, as measured by Qubit 2.0 fluorometer, at each step of the workflow from DNA extraction to PCR purification. Initial DNA concentration following extraction ranged from 19.3 ng/uL to 192 ng/uL, however the addition of 2000 ng of DNA for bisulfite conversion using the QIAGEN EpiTect bisulfite kit retained less than 10 ng/uL of DNA for every case. After PCR amplification, retained DNA concentration further decreased to 1.61 – 3.35 ng/uL, and purification resulted in a final DNA concentration of 0.184 – 0.365 ng/uL. These findings indicate that overall DNA recovery rates across the different patient-derived fibroblast cases and HeLa cells was less than 1%.

Table 4: DNA concentration (ng/uL) is sequentially lost at each step of the workflow.

	Extracted gDNA			BC-gDNA			PCR amplified			PCR Purified			Recovery (%)		
Replicate #	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
HeLa	56.00	-	-	6.58	-	-	3.35	-	-	0.365	-	-	0.652	-	-
MN17	51.8	33.2	54.4	4.10	3.31	3.24	2.92	2.28	2.22	0.188	-	-	0.363	-	-
MND033	35.2	19.3	49	7.96	4.80	1.91	2.98	1.75	2.76	0.304	-	-	0.864	-	-
MND045	43.4	97.8	104	5.10	2.61	2.06	2.08	1.66	1.65	0.184	-	-	0.424	-	-
MND046	175	23.8	26	6.10	4.11	3.34	3.16	1.74	2.36	0.250	-	-	0.143	-	-
MND053	92.4	148	91.6	4.82	4.76	3.42	2.62	2.48	2.34	0.274	-	-	0.297	-	-
MND054	155	192	64	8.72	7.80	3.70	2.40	1.61	2.72	0.242	-	-	0.156	-	-
MND055	97	125	119	5.96	4.29	6.20	2.34	2.28	2.19	0.310	-	-	0.319	-	-
MND056	174	171	116	7.12	7.68	7.36	2.24	2.38	2.67	0.294	-	-	0.169	-	-
MND057	57.2	160	73.2	5.66	6.44	8.12	2.88	2.58	2.23	0.198	-	-	0.346	-	-

Abbreviations: BC; Bisulfite-converted. '-' indicates no available data.

3.5 SEQUENCING OF BISULFITE-CONVERTED gDNA FROM HETEROZYGOUS FEMALE PATIENTS

An important part of this study was to determine the methylation status of CpG sites in the *UBQLN2* CpG island using bisulfite-converted DNA from patient-derived fibroblasts. Of the nine patient lines, dermal fibroblasts from the three heterozygous female patients (MN17, MND033 and MND046) exhibited successful bisulfite conversion (Figure 3.8). Bisulfite conversion and amplification of gDNA using the primer sets designed in Primer3 (Table 2) did not yield successful sanger sequencing results, with the exception of *hUBQLN2* bisulf f2d and *hUBQLN2* bisulf r2d primer set, which amplified bisulfite converted DNA. However, it should be noted that only Sanger sequencing data using the *hUBQLN2* bisulf r2d primer, from this set, produced adequate chromatograms for quantification (Figure 3.8).

The sequenced amplicons produced using *hUBQLN2* bisulf r2d primer (Figure 3.8A) were aligned to the *in silico* 'bisulfite-converted' *UBQLN2* CpG island sequence, but did not show successful alignment, suggesting that the primer pair may have amplified a different region in the genome.

Nevertheless, the chromatograms obtained for MN17, MND033 and MND046 illustrate successful bisulfite conversion as only cytosines within CpG sites are present in the final sequence (Figure 3.8A). As established previously, methylated CpG sites are protected from deamination reactions during bisulfite conversion, thus retaining the cytosine nucleotide at this position while all unmethylated cytosines are converted to uracil and subsequently thymine during PCR. Although, CpG sites corresponding to the Infinium Human Methylation 450K probes across the CpG island of *UBQLN2* were not captured in these chromatograms preventing direct comparison between *UBQLN2* methylation patterns from publicly available data and patient-derived data.

3.5.1 Quantification of CpG sites identifies methylated and unmethylated CpG status

Heterozygous peaks can be identified in Sanger sequencing data as a double peak at the same position on the chromatogram[188]. That is, cytosine peaks present in the chromatograms of amplicons derived from MN17, MND033 and MND046 are often accompanied by a smaller green peak corresponding to thymine (Figure 3.8A) indicating the presence of two variants at a single position in the CpG island, the methylated variant (C) likely originating from the *UBQLN2* on the inactive X-chromosome and the unmethylated variant (T) likely from the active X-chromosome. **Figure 3.8B** quantifies the proportion of the total signal that each nucleotide contributes at all cytosine residues present in the chromatograms of MN17, MND033 and MND046. Across sequencing data for MN17, MND033 and MND046 cytosine represented an average of 63.9%, 67.37% and 75.76% of the total signal for each 'conserved' cytosine peak, respectively. Meanwhile, thymine made up an average of 32.75%, 27.33% and 22.29% of the total signal, respectively. Interestingly, for all three cases, more than 96% of the peak signal located at 56563770 bp (within the CpG island proper) was attributed to cytosine, with thymine making up an average of 1% of the total signal.
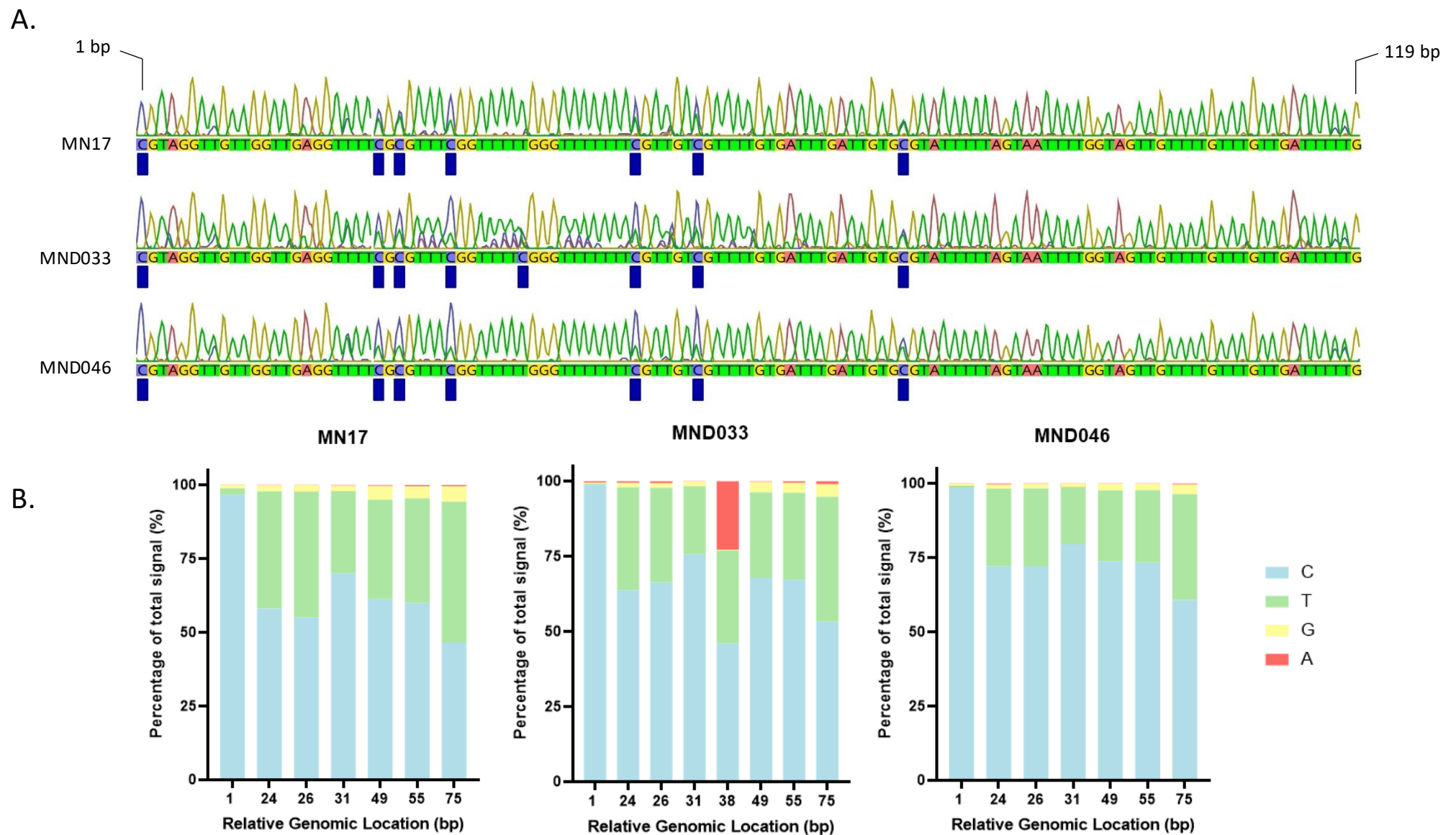
Figure 3.8: Methylation of CpG sites retains cytosines following bisulfite conversion of patient-derived fibroblast gDNA. Chromatograms of bisulfite-converted gDNA from MN17, MND033 and MND046 (*UBQLN2* heterozygous females) illustrate the presence of methylated and unmethylated cytosine residues (A). Quantification using signal proportions show cytosine and thymine peaks at relative genomic location corresponding to cytosine in the chromatograms (B). Geneious Prime (version 2020.0.5) was used to align the chromatograms. Sequencing data was provided by Auckland Genomics and graphed using GraphPad Prism 9.0.2. Genomic location is listed in a 5' – 3' direction and corresponds to the 5' – 3' order of cytosines in the respective chromatograms.

3.5.2 Primer design is integral for PCR amplification of bisulfite-converted gDNA

Following unsuccessful amplification of bisulfite converted DNA from patient-derived fibroblasts using most Primer3 primers, it was necessary to design a new set of primers. Deamination reactions converting unmethylated cytosines to uracil significantly alters the DNA sequence following bisulfite conversion[150] (Figure 3.6), as a result, primers should be designed to match the converted sequence as opposed to the native genomic sequence. The genomic sequence for the promoter region of *UBQLN2* was manually edited in Geneious Prime software (2020.0.5) to reflect bisulfite conversion of all cytosine residues to thymine while only retaining cytosines at CpG sites (Figure 3.9). Alignment of the primers designed using MethPrimer (Table 2) demonstrates sufficient tiling across the 'bisulfite-converted' sequence of the *UBQLN2* CpG island (Figure 3.8).

Gradient PCRs for optimising primer annealing temperatures were used to visualise amplification of amplicons spanning the CpG island of *UBQLN2*. **Figure 3.10** illustrates the three primer sets used to achieve amplicons of 327 bp (set 1: *hUBQLN2* bisulf f1 and *hUBQLN2* bisulf r2.2), 417 bp (set 2: *hUBQLN2* bisulf f3.2 and *hUBQLN2* bisulf r4.2) and 376 bp (set 3: *hUBQLN2* bisulf f4.2 and *hUBQLN2* bisulf r5.2). Primer set *hUBQLN2* bisulf f1 and *hUBQLN2* bisulf r2.2 set to amplify the first third of the *UBQLN2* CpG island, encompassing CpG sites located at 56563335, 56563359, 56563485, 56563513, 56563516 and 56563518 bp (Probe ID: cg02587153, cg10009003, cg25697726, cg09013091, cg14824933 and cg23993375, respectively) did not amplify this portion at any temperature ranging from 45 °C – 55 °C (Figure 3.10A). Primer set *hUBQLN2* bisulf f3.2 and *hUBQLN2* bisulf r4.2 set to amplify the middle of the CpG island, encompassing CpG sites located at 56563664. 56563680, 56563754, 56563929 and 56563953 bp (Probe ID: cg10950413, cg09666011, cg16161440, cg02492740 and cg01780361, respectively) demonstrated an optimal annealing temperature from 51.9 °C to 54.1 °C (Figure 3.10B). Finally, the primer set encompassing the final portion

of the *UBQLN2* CpG island, *hUBQLN2* bisulf f4.2 and *hUBQLN2* bisulf r5.2 covering the CpG sites located at 56564080 and 56564104 bp (Probe ID: cg04270663 and cg20979765, respectively) showed optimal annealing at 59.1 °C (Figure 3.10C). Although all primer sets fulfilled conditions required for successful amplification of bisulfite-converted DNA, as defined by MethPrimer, considerable variability is observed. A visual difference can be observed in the amount of DNA amplified at optimal annealing temperature between primer set 2 and set 3 based on the intensity of the bands present in the gel.
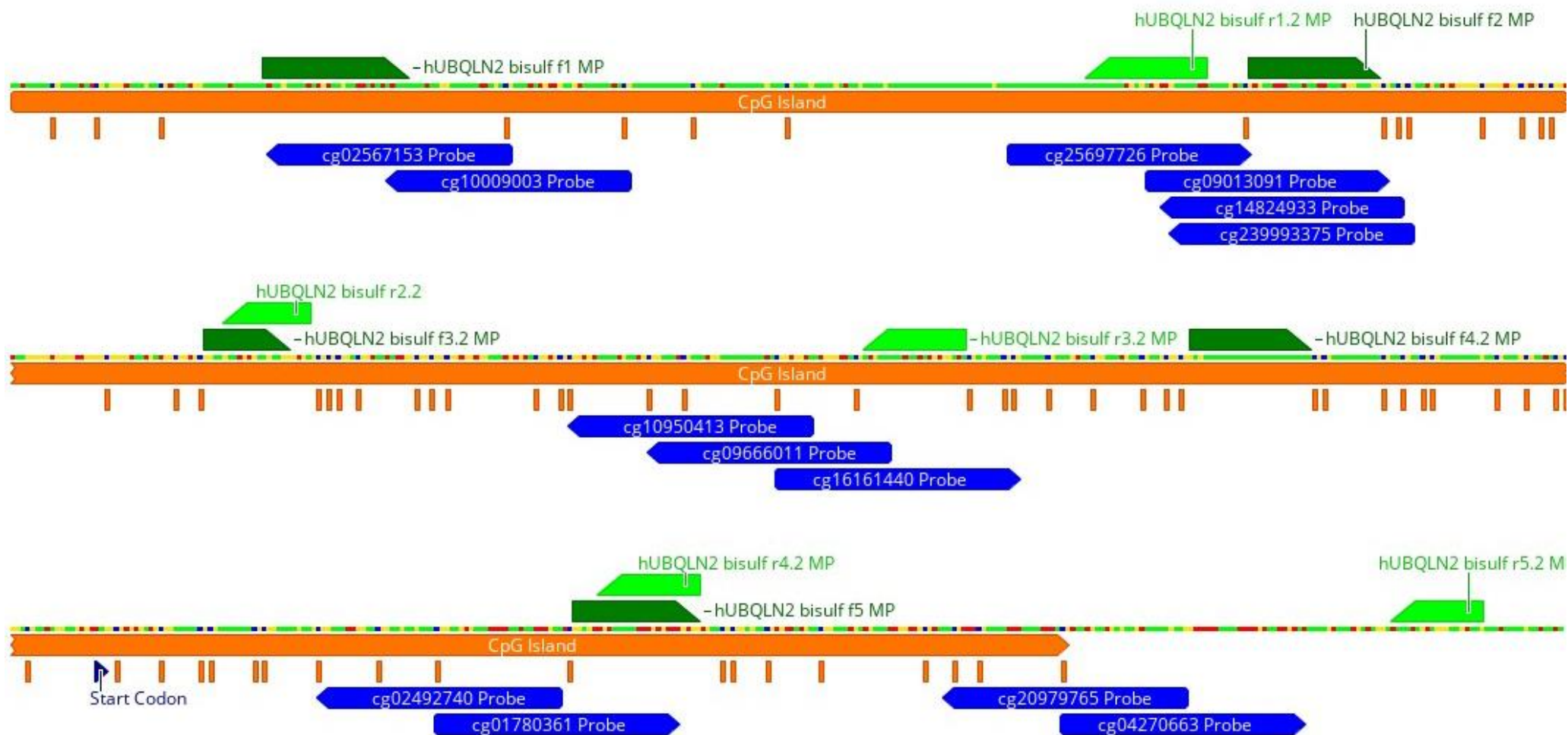
Figure 3.9: Primer alignment to the 'bisulfite-converted' promoter region of *UBQLN2*. Most CpG sites probed in EWAS and MethBank 3.0 were encapsulated across all primer pairs to allow for direct comparison between publicly available data and patient-derived data. Every CpG site across the CpG island is annotated with an orange bar. Note that the 5'-most probe, cg21862542 located at chrX:56560144 (GRCh38) has been omitted from this diagram. Base colours: A = red, T = green, C = Blue, G = yellow. Annotations were aligned to the CpG island sequence of *UBQLN2* using Geneious version 2020.0 created by Biomatters.
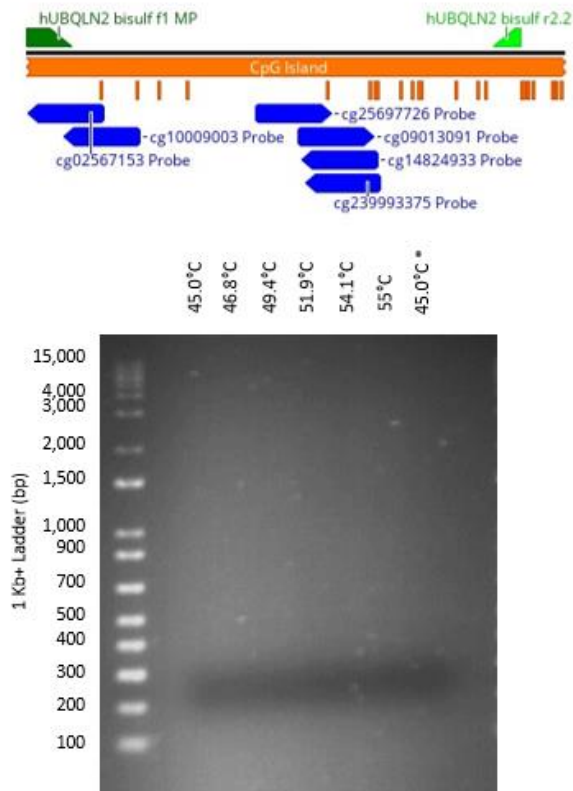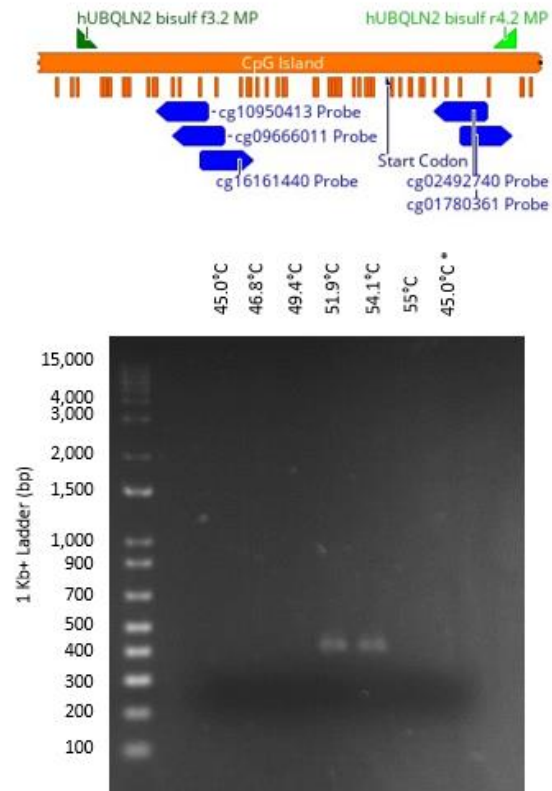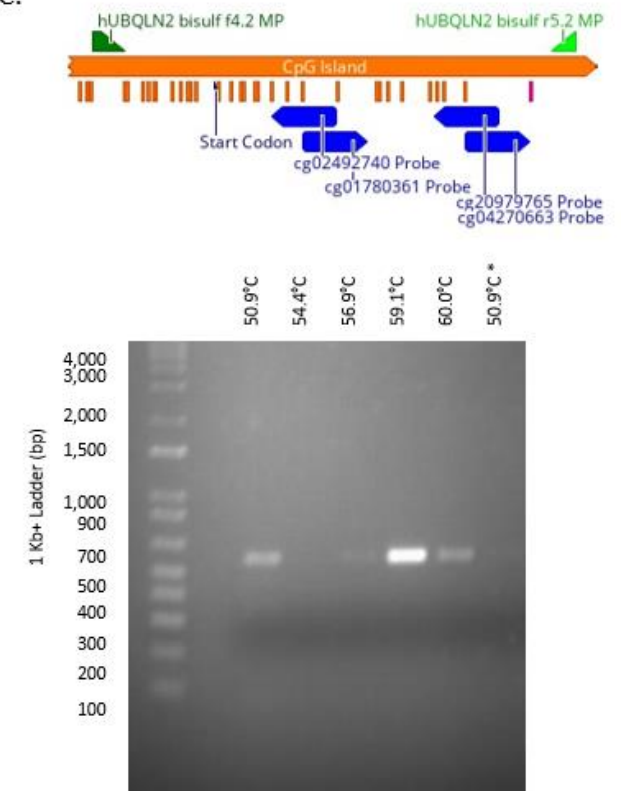
Figure 3.10: **Primer annealing temperature optimisation across the CpG island of *UBQLN2*.** Gradient PCR for three amplicons spanning the CpG island of *UBQLN2*, amplified using Phusion-U hot start DNA polymerase. Primer set *hUBQLN2* bisulf f1 and *hUBQLN2* bisulf r2.2 (A) *hUBQLN2* bisulf f3.2 and *hUBQLN2* bisulf r4.2 (B) and *hUBQLN2* bisulf f4.2 and *hUBQLN2* bisulf r5.2 (C) were optimised and visualised in a 2% agarose gel in 1x TAE buffer. Primer sets were aligned to the CpG island sequence of *UBQLN2* using Geneious version 2020.0 created by Biomatters.

4 DISCUSSION

The underlying mechanism of sexual dimorphism observed in *UBQLN2*-linked ALS/FTD is poorly understood but may reveal methods of disease protection that can be used as predictive measures. Previous studies have defined the X-linked dominant mode of inheritance of *UBQLN2* mutations as a cause of ALS[5,7], attributing the reduced penetrance of ALS/FTD in females to XCI. However, the potential role of CpG island methylation patterns remains unexplored and may provide clues as to the protective effects observed in females.

Analysis of publicly available methylome data revealed that females exhibited significantly higher levels of methylation at sites across the CpG island of *UBQLN2* than males. Furthermore, these methylation patterns are conserved across tissues. These findings suggested that *UBQLN2* methylation patterns are a reliable proxy for understanding sex-specific differences in *UBQLN2*-linked ALS/FTD. While publicly available data is a valuable resource, limitations exist within its application to patients due to variability in data collection methods, loss of information to normalising procedures, unknown quality of the data or the method of generating data between studies, all of which can affect confidence in generalising the data to specific patients[189]. Alternatively, patient-derived data may provide a more informative and accurate depiction of methylation patterns across the *UBQLN2* CpG island[158,190]. To apply findings from publicly available data to patient-derived fibroblast samples the workflow for bisulfite conversion was optimised, revealing the highly degradative nature of the conversion and the subsequent steps for sequencing. Nevertheless, sequencing data from three heterozygous female cases demonstrated successful bisulfite conversion.

The research presented here demonstrates the importance of XCI-based methylation of *UBQLN2* in conferring protection against *UBQLN2*-Linked ALS and underscores the potential value of patient-derived methylation for predicting disease protection and the application of

these methylation patterns to reduce expression of the mutant *UBQLN2* allele as a therapeutic approach to attenuating the neurodegenerative and paralytic phenotypes associated with *UBQLN2*-linked ALS.

4.1 COMPARING SEX-SPECIFIC METHYLATION OF CpG SITES IN THE CpG ISLAND OF *UBQLN2*

4.1.1 CpG sites in the *UBQLN2* CpG island are more highly methylated in females than male *within* a given tissue type

Statistical analysis of publicly available methylome data conducted in this thesis revealed a significant increase in the average β-value of CpG sites across in the promoter region of *UBQLN2*, within a given tissue, for females compared to males (Figure 3.2 – 3.5). These results suggest that females exhibit a higher level of methylation across the CpG island of *UBQLN2*, reflecting a more condensed chromatin structure and thus repressed *UBQLN2* expression. The elevated methylation observed in females is likely because of X-inactivation. When compared to XCI positive and negative controls, the average β-value of *CpG* sites across the *CDK16* CpG island (X-inactivation escapee) are significantly lower than that of *UBQLN2* while methylation levels between CpG islands of *AR* (known to undergo X-inactivation) and *UBQLN2* are similar.

The functional consequence of differential CpG island methylation has been shown to substantiate sex-specific variance in phenotype[116,191], especially in X-linked disorders[133,139,140]. Previous studies have demonstrated sex-specific differences in methylation across the CpG island of X-linked genes and its correlation to changes in gene expression between males and females[117,192,193]. Increased CpG island methylation of genes on the X-chromosome in females corresponded to decreased microRNA levels and reduced target gene expression compared to males[117]. The use of transcriptomic information in combination with methylome data is often

employed in studies investigating differential DNA methylation to corroborate methylation patterns to functional changes[192–195].

RNA-sequencing is a reliable technique that is often employed for such measures to determine the abundance of wildtype and mutant transcripts[192–195]. Although it is likely that increased methylation of the CpG island of *UBQLN2* observed here in females translates to repressed *UBQLN2* transcription and therefore expression, interpretations of the publicly available methylome data is limited to differential *methylation patterns* as opposed to differential *UBQLN2 expression.* The incorporation of supplementary RNA-sequencing data may bridge the relationship between differential epigenetic modifications between males and females outlined in this research and the sex-specific functional consequences in an allele-specific manner[146,196,197]. While the lack of transcriptomic information may be a limitation of the present study, the primary focus of this research was to investigate the potential role of *UBQLN2* CpG island methylation patterns in the sexual dimorphism displayed in *UBQLN2*-linked ALS/FTD; this information cannot be obtained through RNA sequencing alone as methylation status of a specific CpG site is not reflected in overall transcriptomic data[198,199].

4.1.2 CpG sites in the north shelf of the *UBQLN2* CpG island are more highly methylated than probes in the CpG island proper

Interestingly, these findings demonstrate that males and females exhibit significantly elevated levels of methylation at the most 5' probe in situated in the north shelf of the *UBQLN2* CpG island at 56560144 bp (GRCh38; Probe ID: cg21862542) (Figure 3.2 – 3.4) compared to other CpG sites. Notably, males displayed elevated methylation at this site across a range of tissue types when compared to females. These observations are noteworthy as this pattern is also reflected in the CpG island of XCI positive and negative controls, *AR* and *CDK16* respectively, as well as the *TARDBP* autosomal control. Higher β-values indicate higher levels of methylated

probe detection over unmethylated probes in the Infinium Human Methylation 450K BeadChip assay, suggesting that more copies of the cg21862542 CpG site in *UBQLN2* were methylated in the sample[177]. A study looking at sex-specific differences in the genome-wide DNA methylation pattern also reported a similar pattern of methylation in the CpG island of genes on the X-chromosome impacting insulin secretion[117]. Similarly, this study reported females having significantly elevated methylation in the CpG island proper when compared to males and significantly elevated methylation of CpG sites located in the north and south shelf in both sexes[117]. Interestingly, males also displayed a higher level of methylation than females at these sites[117].

The highly conserved pattern of methylation at CpG sites located approximately 4 kB upstream and downstream of the CpG island proper across a range of genes alludes to a potential regulatory or housekeeping function. A meta-analysis seeking to define the purpose of these stably methylated CpG sites flanking the CpG island, discovered that high levels of methylation in the shelves were almost always associated with lowly methylated CpG sites within island, describing this finding as a 'ravine' pattern[200]. Furthermore, these stably methylated regions flanking the CpG island proper are also shown to be associated with high transcriptional activity due to its interactions with housekeeping genes[200].

The research presented in section 3.2.2 demonstrates a similar pattern in which methylation in males at CpG sites located in the north or south shelf of genes affected by XCI are significantly elevated than females at the same site. In contrast, autosomal genes and genes that escape XCI do not show significant sex-specific methylation at these sites. Gene expression analysis to validate the relationship between sex-specific differential methylation at 'ultra-methylated' CpG sites in the shelf of a CpG island could further demonstrate the potential regulatory role in *UBQLN2* expression. However, it is important to note that methylome information of only one CpG site located in the north shelf of the *UBQLN2* CpG island was available from publicly

available databases. Therefore, limiting the ability to extend this pattern to sites beyond the given *UBQLN2* sequence using publicly available data.

4.2 COMPARING SEX-SPECIFIC METHYLATION OF THE WHOLE *UBQLN2* CpG ISLAND *BETWEEN* TISSUE TYPES

4.2.1 Females display elevated methylation across the CpG island of *UBQLN2* regardless of tissue type

Tissue specific methylation patterns between males and females were not observed in the CpG island of *UBQLN2* for any CpG site assessed using publicly available methylome data (Figure 3.5 and Appendix C and D). Interestingly, several studies have reported tissue-specific methylation patterns of CpG islands, not only among somatic tissue[201,202], but between brain regions, with further differences between grey and white matter within the same region[203,204]. Notably, these studies asses global methylation of thousands of CpG sites across hundreds of genes as opposed to CpG island methylation patterns for a single gene, or X-linked genes, as conducted in this study[202,203,205].

The *FRM1* gene responsible for Fragile X-syndrome (FXS; X-linked dominant disorder) is shown to exhibit tissue specific methylation between leukocytes, lymphoblasts and fibroblasts from heterozygous females[206]. Additionally, a study conducted by Esanov et al. (2016) demonstrated selective enrichment of *FRM1* promoter methylation of the FRM1 promoter in neurons when compared to glial cells in FXS patients[207]. Publicly available methylome data obtained from different brain regions, as well as blood and fibroblasts samples did not illustrate tissue specific methylation patterns (Figure 3.5 and Appendix B) across the CpG island of *UBQLN2* in this study. The existence of tissue specific methylation differences between neuronal cell types in an X-linked neurodegenerative disease, like FXS, underpins the potential

for further investigation of tissue specific methylation differences of *UBQLN2* CpG sites. The non-neuronal tissues compared in this study were chosen based on their non-invasive nature of obtainment to ensure that *UBQLN2* methylation patterns would not vary between patient-derived tissue samples. Ensuring that methylation patterns will remain consistent between distinct tissue types and will reflect methylation patterns in brain tissue is crucial when ensuing patient-derived methylome data to predict protection against *UBQLN2*-linked ALS.

CpG island methylation is not always predictive of gene expression[202]. The use of tissue-specific expression of *UBQLN2* in conjunction with tissue-specific methylation patterns are often employed when assessing variation in DNA methylation[201–203]. Supplementing the *UBQLN2* methylome data analysed in this research with publicly available *UBQLN2* tissue-specific gene expression information can be undertaken in future research to further validate tissue-specific methylation patterns in the *UBQLN2* CpG island.

4.3 DEVELOPING IN-HOUSE METHODS FOR ANALYSIS OF CpG ISLAND METHYLATION OF *UBQLN2*

One of the main aims of this research was to optimise bisulfite conversion for sequencing using patient-derived fibroblast cells. The 'gold-standard' classification of bisulfite sequencing-based methods for methylation detection in DNA[150] formed the basis of the workflow proposed in **Figure 3.6.** Bisulfite conversion, PCR amplification and Sanger sequencing are crucial steps required for DNA methylation analysis[208] (Figure 3.6), but can be limiting when assessing gene-specific methylation as opposed to genome-wide methylation. For successful retention of CpG site methylation, gDNA extracted from patient-derived fibroblasts cannot undergo prior enrichment to amplify the promoter region of *UBQLN2* for bisulfite conversion as epigenetic modifications are not retained[209,210]. Furthermore, assessment of DNA material following bisulfite conversion reflects whole-genome DNA concentration within the sample, of which

the *UBQLN2* CpG island makes up only a small portion of the starting material (Table 3 and 4) – consequently, the amount of template available for PCR amplification is extremely limited. Because the starting material cannot be enriched or amplified, optimising other aspects of the workflow to ensure efficient conversion and DNA retention[211] formed a major part of this research. Previous studies have demonstrated successful gene-targeted DNA methylation analysis[212–214] however, efforts to optimise workflow flow are not always outlined.

4.3.1 QIAGEN EpiTect bisulfite kit shows higher DNA recovery rate following bisulfite conversion

Bisulfite conversion is known to cause significant DNA loss that can hinder downstream analysis[211,215]. To address this issue, comparative analysis of commercially available bisulfite conversion kits, specifically, the Zymo Research EZ DNA Methylation gold kit and the QIAGEN EpiTect Bisulfite kit to test performance differences in terms of DNA yield (Table 3). While different techniques were used to measure DNA concentration following bisulfite conversion, measurements using the Qubit 2.0 fluorometer were favoured as they are more accurate and reliable than using a Nanodrop 2000[185], which has been shown to significantly overestimate DNA concentration[185,216]. The QIAGEN EpiTect Bisulfite kit was selected over the EZ DNA Methylation gold kit (Table 3) as it had a higher DNA recovery rate and allowed for successful downstream amplification.

The EZ DNA Methylation gold kit demonstrated poor DNA retention following bisulfite conversion when compared to the QIAGEN EpiTect bisulfite kit. This can be attributed to the use of DNA protect buffer and carrier-RNA aimed to protect against excessive degradation during conversion and improve DNA recovery during clean up in the QIAGEN EpiTect bisulfite protocol[217–220]. Furthermore, addition of the DNA protect buffer provides visual confirmation that the pH of the reaction is optimal for bisulfite conversion through a green to

blue colour change[219], giving the QIAGEN EpiTect Bisulfite kit an advantage over EZ DNA Methylation Gold.

Comparative studies have established that there tends to be a performance difference between different bisulfite conversion kits based on DNA yield and conversion efficiency. While a previous study found that the EZ DNA Methylation gold kit had a significantly higher DNA yield[221], the findings in this study suggest that QIAGEN EpiTect Bisulfite kit may be more advantageous. A key aspect for future comparative analysis of DNA concentration following bisulfite conversion is to ensure the same method of measurement is used to allow for direct comparison. While the additional features of the QIAGEN EpiTect kit aided in the decision to use this kit, direct performative analysis on experimental samples would be more beneficial. Additionally, exploring gentler methods of DNA methylation analysis such as enzymatic methyl sequencing[222,223] may aid in optimising DNA recovery while capturing epigenetic information.

4.3.2 Validation methods confirm bisulfite conversion of gDNA

Obtaining methylation data from patient-derived fibroblasts is a molecular based process that relies on changes to the DNA sequence to demonstrate conversion, which is often corroborated by sequencing- the last step of the workflow[150] (Figure 3.6), therefore methods to validate successful bisulfite conversion before amplification were investigated.

Although bisulfite conversion is a highly degradative process, the characteristic range of DNA fragments from 100 bp to 1500 bp[187] observed on a 2% agarose gel (Figure 3.7) can be taken advantage of for validating successful conversion. Several studies have remarked on the presence of a DNA smear proceeding bisulfite conversion[187,218,224] but it does not necessarily dictate successful conversion and is more appropriate for visualising DNA fragmentation[187,218].

Alternative methods for validating conversion must therefore be employed, such as the use of bisulfite-conversion-specific primers like those in the Zymo Research Human Methylated & Non-methylated DNA set for bisulfite-specific PCR (Figure 3.7C)[225]. While the Zymo Research Human Methylated & Non-methylated DNA set has been adopted by several studies analysing DNA methylation patterns[226–228], the *DAPK1* primer isn't usually employed to validate bisulfite conversion. Nevertheless, bisulfite-specific PCR using self-made bisulfite-conversion-specific primers followed by gel electrophoresis is a common approach used to evaluate bisulfite conversion[229]. Interestingly, many studies apply methylation-specific validation methods such as methylation-specific PCR[229] or methylation-sensitive high resolution melting analysis[226,230] to evaluate successful sample conversion. While these techniques may prove to be effective, they are only selective for template material that is methylated as opposed to bisulfite converted[229]. In the case of *UBQLN2* heterozygous female carriers, who are likely to express differential methylation patterns at a single CpG site in the CpG island due to the presence of two alleles (Figure 3.8), using methylation-specific validation methods would enrich the methylated template strand.

The inability to detect incomplete bisulfite conversion is a potential limitation of non-sequencing-based validation techniques for bisulfite conversion, however these methods provide a level of confirmation prior to PCR amplification and Sanger sequencing

4.3.3 Bisulfite conversion workflow is highly destructive to and poorly retentive of DNA

Bisulfite conversion is considered to be the 'gold-standard' technique for DNA methylation analysis[150], but it is not without limitations. DNA loss and degradation are highly reported issues associated with bisulfite conversion, this is especially true for gene-specific methylation analysis as demonstrated in the results of this research using the QIAGEN EpiTect bisulfite kit

and several other studies[211,231,232]. Incomplete conversion of unmethylated cytosines to uracil is another shortcoming that can affect downstream methylation analysis, causing an overestimation of CpG island methylation with bisulfite conversion[211,232,233], although this is rarely reported in the literature. Despite these limitations, the gold standard status of bisulfite conversion is reflected through its success with whole genome and gene-specific DNA methylation analysis in several studies[231,234]. The results of this study also demonstrate that inefficient PCR amplification and purification can also contribute to poor DNA retention. Exploring methods that reduce the number of steps in the workflow may enable greater DNA retention.

Oxford Nanopore is a high-throughput sequencing technology that can detect DNA methylation at single nucleotide resolution of native DNA without the need for bisulfite conversion, therefore reducing the number of steps required in the workflow[235]. Due to the novelty of the technology, most DNA methylation analysis studies using Oxford Nanopore are conducted across the whole genome, as opposed to gene-specific analysis – making this technology costly[235–238].

4.4 SEQUENCING OF BISULFITE-CONVERTED DNA FROM HETEROZYGOUS FEMALE PATIENTS

Sanger sequencing of bisulfite-converted DNA from MN17, MND033 and MND046 demonstrated successful bisulfite conversion due to the retention of cytosines only when followed by a guanine residue (CpG site). The sequences provided through Sanger sequencing did not align with the *in-silico* bisulfite converted sequence for the CpG island of *UBQLN2*, suggesting that the *hUBQLN2* bisulf r2d primer may have amplified a different portion of the genome. As described previously, the workflow for bisulfite conversion does not allow for gene-specific enrichment prior to bisulfite conversion, therefore making primer design integral

for amplifying the CpG island of *UBQLN2* prior to Sanger sequencing. Consequently, *UBQLN2* CpG sites analysed in the publicly available data were not sequenced, thus hindering direct comparisons between CpG methylation patterns between public and patient-derived data. Nevertheless, these results indicate the underlying potential of bisulfite conversion to capture the methylation landscape across the CpG island of *UBQLN2* from patient-derived cells, both from females and from males.

Dermal fibroblasts are a desirable cell type for creating patient-derived model systems as they are readily accessible, fast growing and cost-effective[239], and are routinely used to model ALS[240,241] and FTD[239,242]. Despite this, limitations of using dermal fibroblasts for research are important to consider. Prior studies investigating neurodegenerative disorders utilising patient-derived dermal fibroblasts have reported on the finite passage number before cellular senescence[243] - a gradual process that shuts down cell metabolism through CpG hypermethylation[244], suggesting that dynamic changes to the epigenetic landscape are capable of occurring while in culture. Patient-derived fibroblasts obtained in this study were cultured for approximately a month before they were ready for DNA extraction, around a passage of 8. Ideal cell passage numbers are not clearly reflected in the literature, however passages used in this study are relatively low[245] and therefore unlikely to be affected by epigenetic changes linked to cell senescence.

Alternatively, longer culture times can drive the risk of selecting for cells that are healthier and more suited for the culture environment; particularly in cases of mutant allele expression such as that in the heterozygous female population of this study. Furthermore some studies have observed skewing in the allelic expression of the X-chromosome in culture which may alter the observed DNA methylation pattern in downstream quantification, however it is possible that this is only consistent for patient-derived fibroblast pluripotent stem-cell populations[245,246]. In future studies, the use of brain tissue in tandem with dermal fibroblast samples to observe

sex-specific differential methylation patterns of *UBQLN2* may be more informative as *UBQLN2* pathology is typically localised to the brain[6]. This could also help identify any tissue specific methylation differences in the *UBQLN2* CpG island.

The purpose of using patient-derived dermal fibroblasts was to establish patient-specific methylation patterns between males and females. Due to the highly destructive nature of bisulfite sequencing and inadequate primer design, DNA methylation analysis of the control males and females were unsuccessful. Nevertheless, the results of this research demonstrate the fundamental prospects of mapping methylation patterns in a patient-specific manner which can be expanded to encompass an allele-specific method to correlate *UBQLN2* CpG island methylation patterns to protection in female carriers of the mutant *UBQLN2*.

4.4.1 Primer design is integral for PCR amplification of bisulfite converted gDNA

Numerous studies have documented the importance of primer design in ensuring the efficacy of bisulfite sequencing and DNA methylation analysis[150]. Extensive modification of the template DNA enforces a precise set of primer requirements that determine the accuracy of downstream methylation analysis[152]. Consequently, various primer design tools, including MethPrimer[181], have been developed specifically for designing primers against bisulfite-converted gDNA[149,247].

Although all primer conditions were met using MethPrimer, **Figure 3.10** demonstrates the variability in annealing temperature optimisation and amplification of bisulfite-converted DNA. MethPrimer is a widely used program that has been employed in methylation array studies to predict CpG sites in promoters and for primer design in methylation arrays[248–251]. It is highly valued for its ability to modify the native sequence to the bisulfite-converted version, predict CpG sites and provide a graphical sequence alignment of primers. However, a major

limitation of this program is its overestimation of the melting temperature and AT-rich primers[248,252] which can affect their ability to anneal to bisulfite-converted template during PCR amplification, leading to reduced specificity and PCR efficiency[253]. It should be noted that AT-rich primers may be necessary when amplifying bisulfite-converted DNA as the unmethylated cytosines are converted to thymine, creating a thymine-rich template[150].

While amplifying and sequencing patient-derived bisulfite-converted DNA using the MethPrimer primer sets was unsuccessful due to the degradative workflow, the re-design provides confidence that future analysis will amplify the CpG island of *UBQLN2*.

4.5 FUTURE DIRECTIONS

The result of this research demonstrates the underlying potential in analysing differential *UBQLN2* methylation patterns between males and females using patient-derived dermal fibroblast cells.

Publicly available data analysed in this research demonstrates that DNA methylation patterns across the CpG island of *UBQLN2* varies between males and females and is likely due to X-inactivation. The use of transcriptomic data to align patterns of *UBQLN2* methylation with *UBQLN2* expression will be informative in future studies as DNA methylation is not always an accurate reflection of gene expression[202]. The second part of this research involved optimising the bisulfite sequencing workflow for DNA methylation analysis in patient-derived fibroblast cells. Although the bisulfite sequencing workflow is highly degradative to DNA, analysis demonstrates that the bisulfite sequencing workflow can provide insight to the methylation landscape of the *UBQLN2* CpG island.

The findings of this research provide a promising foundation for the development of allele-specific methods for detecting DNA methylation patterns. This would enable researchers to correlate unique methylation patterns observed in the CpG island of *UBQLN2* in females to protection against *UBQLN2*-linked disease, using parameters such as age of onset, symptoms or even neuropathology. Furthermore, these patterns of *UBQLN2* methylation can potentially be replicated in men affected by *UBQLN2*-linked ALS as a form of gene-therapy to replicate protection seen in females. Similar therapeutic avenues have been proposed in other X-linked diseases such as Fragile X-syndrome[207], making this research relevant not only to *UBQLN2*-linked ALS, but other X-linked diseases.

5 CONCLUSIONS

Methylation landscapes across the CpG island of a gene can affect the clinical phenotype of disease, especially in X-linked disorders. In the case of *UBQLN2*-linked ALS, sex-specific phenotypic variation has been reported and attributed to XCI. This thesis presents research that delineates the sex-specific methylation differences of CpG sites across the CpG island of *UBQLN2* between males and females. Publicly available data analysis confirmed a non-tissue-specific elevation in CpG island methylation in females compared to males. To validate these patterns in ALS-patients, bisulfite sequencing methods were optimised using patient-derived dermal fibroblast cells. Here it was shown that bisulfite sequencing workflows are harsh on DNA with poor DNA retention for downstream analysis. Despite this, the results of this study demonstrated successful bisulfite conversion of DNA from three *UBQLN2* heterozygous female patients and subsequent methylation analysis through Sanger sequencing.

To conclude, the research presented in this study establishes the groundwork for investigating female protection in X-linked diseases like *UBQLN2*-linked ALS. Ultimately, the methods developed here can be used in combination with RNA sequencing of *UBQLN2*, allele-specific qPCR and functional analyses of patient-derived cells to predict female protection based on the methylation landscape across the CpG island of *UBQLN2* and recapitulate this in affected males to reduce the burden of disease.

6 APPENDICES

```
                    CODE TO EXTRACT - CPG ISLANDS (SCRIPT - 1)
####################################################################################
library(tidyverse)

#################### Required chuck of code to load sangerseq as it is not a
standard r package
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("sangerseqR")

#########################

library(sangerseqR)




data.list=list.files("./Exp 52 MN17, MND046 and MND033 (1A) Sanger Sequence
analysis/", pattern = ".ab1", full.names = TRUE)

extract_CpGs = function(file){

  savename = paste0("CSV results./",sub(pattern = "(.*)\\..*$", replacement =
"\\1", basename(file)),"_CpG_Proportions.csv")

  #Read the data in
  data = read.abif(file)

  #Converts the data to a sangerseq o"object" file
  data.seq= sangerseq(data)
  #visual check for sanity

  #This is where you can see where position "x" lies in the trace.matrix
  peak.pos = peakPosMatrix(data.seq)
  colnames(peak.pos) = c("A_sig","C_sig","G_sig","T_sig")
  peak.pos = as.data.frame(peak.pos)

  #This is where you read the minor allele to major allele ratio
  trace.matrix = traceMatrix(data.seq)
  colnames(trace.matrix) = c("A_sig","C_sig","G_sig","T_sig")
  trace.matrix = as.data.frame(trace.matrix) %>%
    mutate(trace_ID = row_number())
```

```
    trace.matrix.peaks = trace.matrix %>% filter(trace_ID %in% peak.pos$A_sig)
%>%
    mutate(peak_pos = row_number())
```

Appendix A: R. script used to extract Sanger sequencing data for MN17, MND033 and MND046. Relative peak signals for each nucleotide at CpG sites in the Sanger sequencing data provided by Auckland Genomics was extracted from MN17, MND033 and MND046 using R-studio (version 4.1).
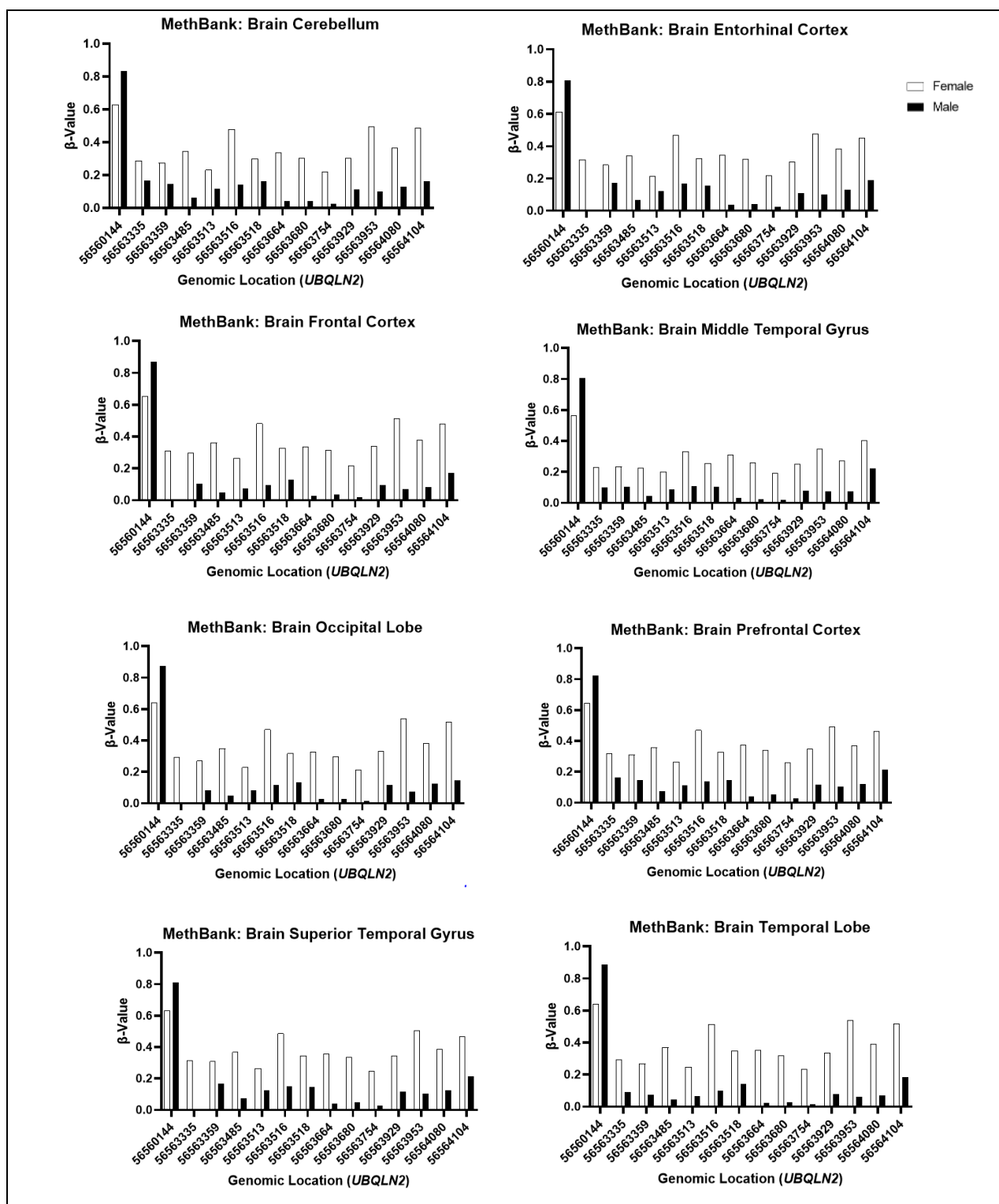
Appendix B: Methylome data from MethBank 3.0 demonstrates elevated methylation of CpG sites in the promoter of *UBQLN2* **in females across different brain regions.** Average β -value for different brain regions demonstrates differential methylation between males and females at CpG sites in the promoter region of the *UBQLN2* gene. Genomic location of the CpG sites is organised in a 5' to 3' direction. Data is presented as average β -value. Raw data was obtained from MethBank and processed in GraphPad Prism 9.0.2.
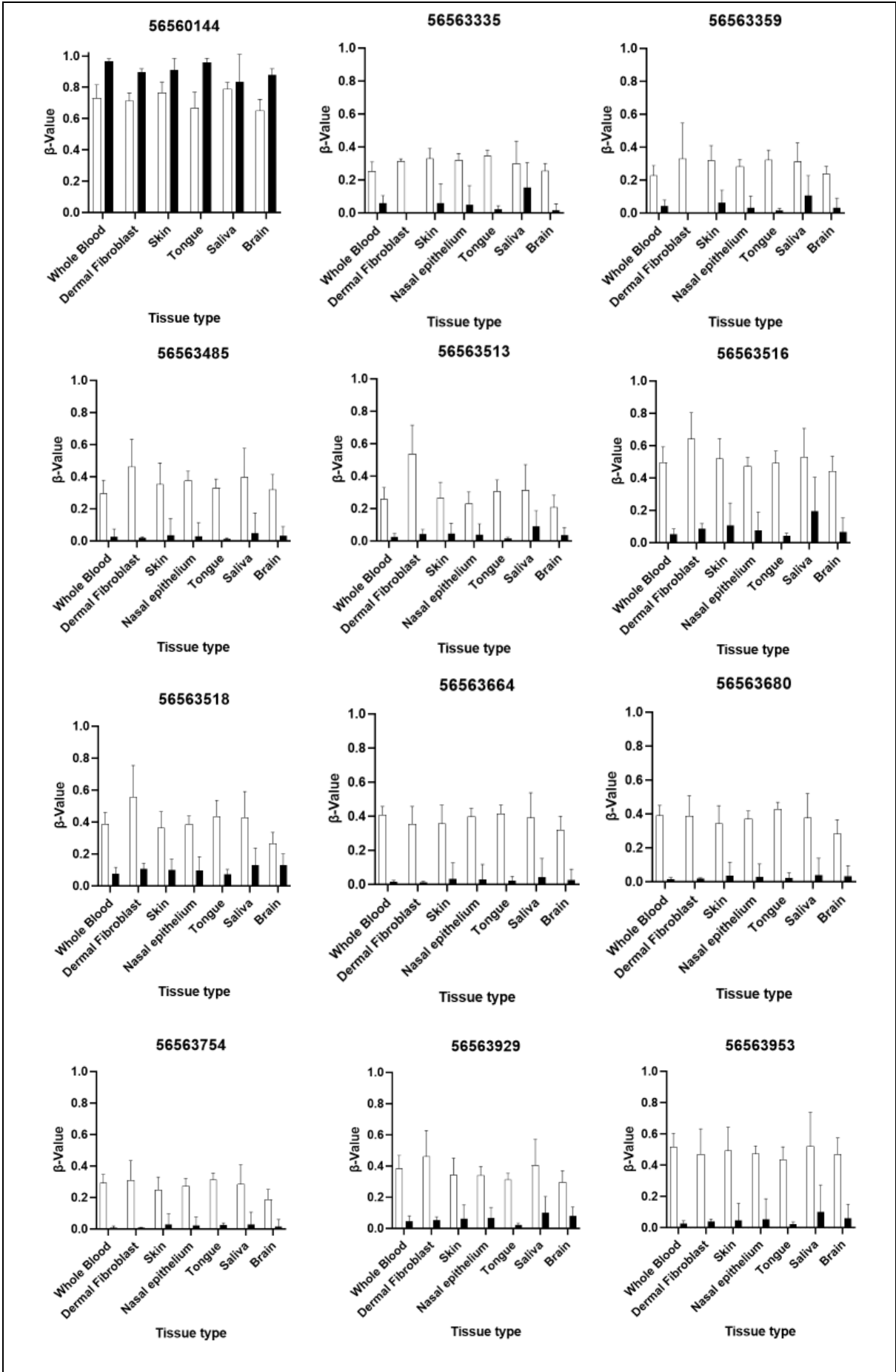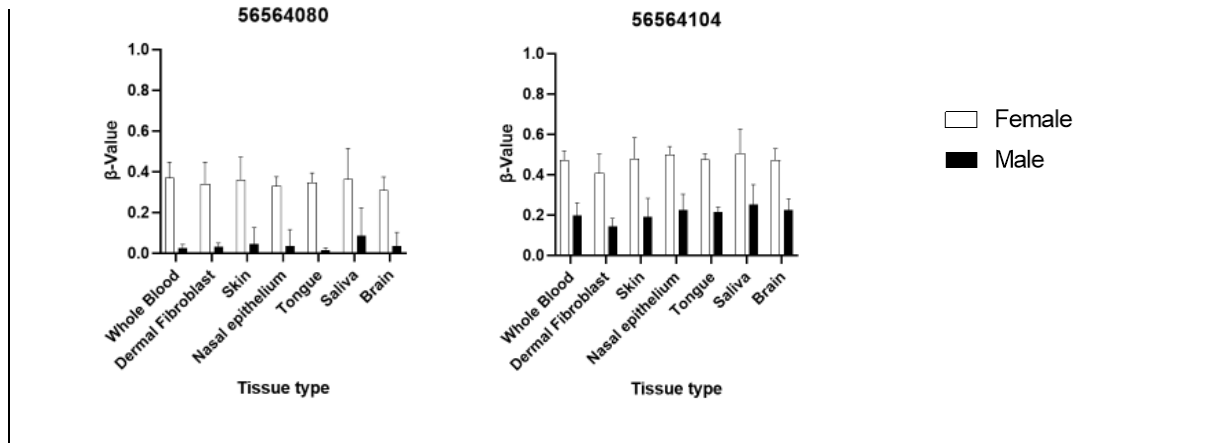
Appendix C: Average β-value for a given CpG site in CpG island of *UBQLN2* is conserved between tissues.
Methylome data from EWAS demonstrates a similar methylation landscape for males and females across all tissue types.
Elevated methylation of each CpG site across the CpG island of *UBQLN2* in females is conserved across all tissue types.
Data is presented as average β-value ± SD. Raw data was obtained from EWAS and processed in GraphPad Prism 9.0.2.

89
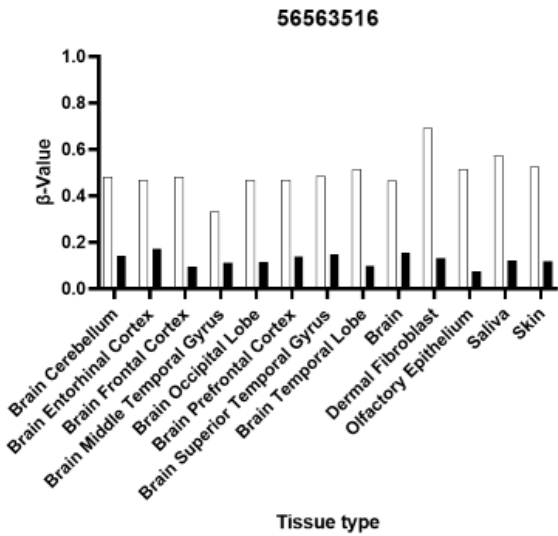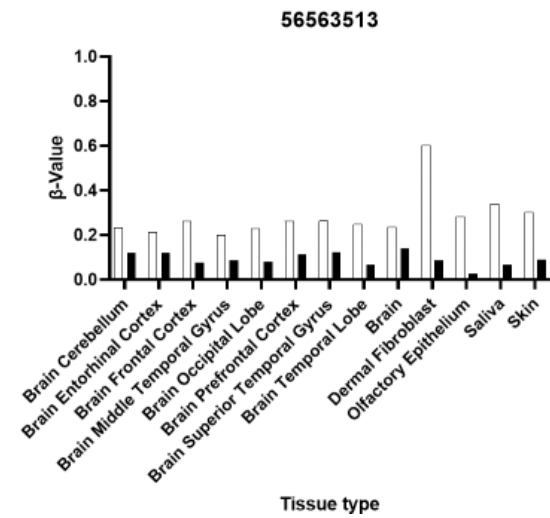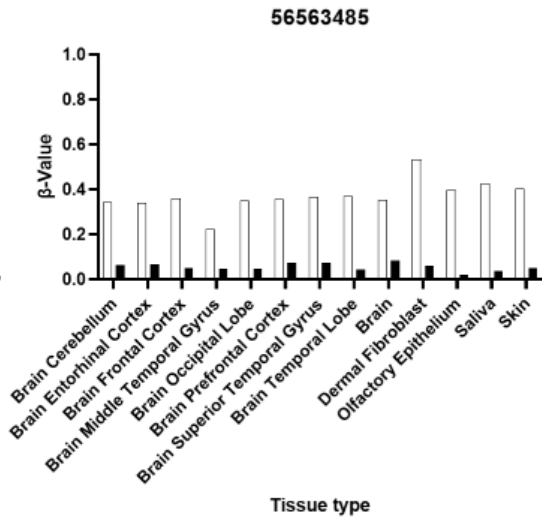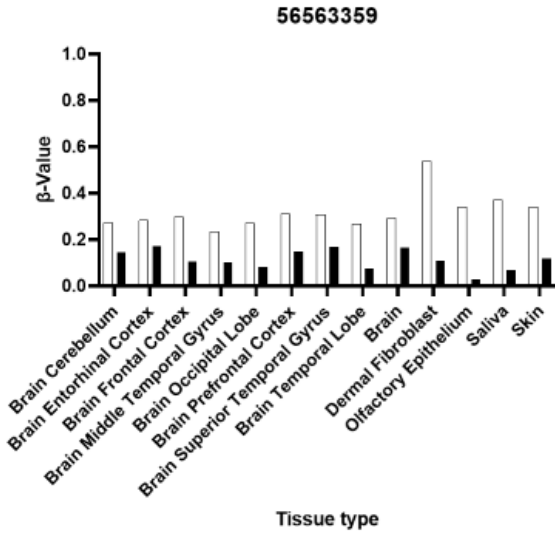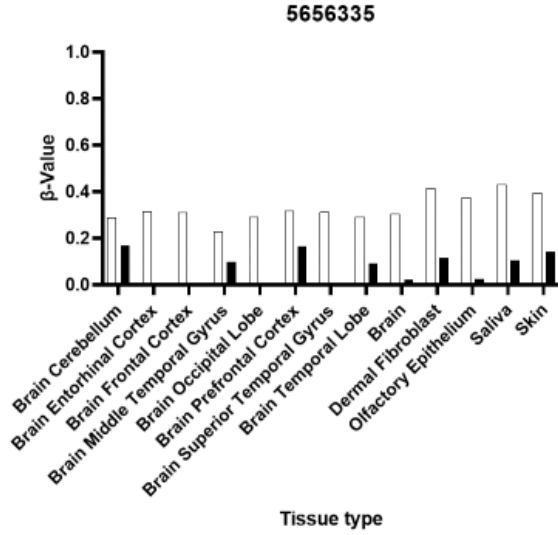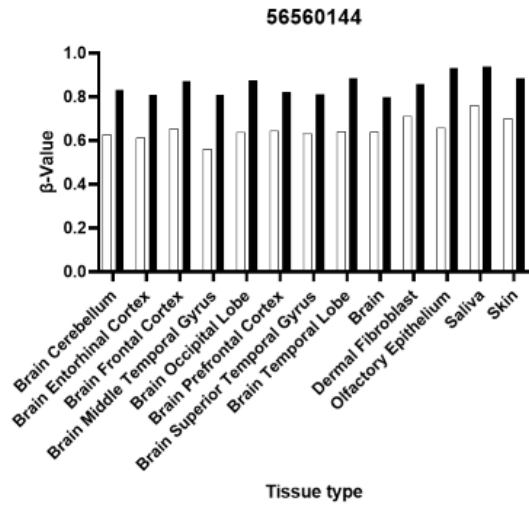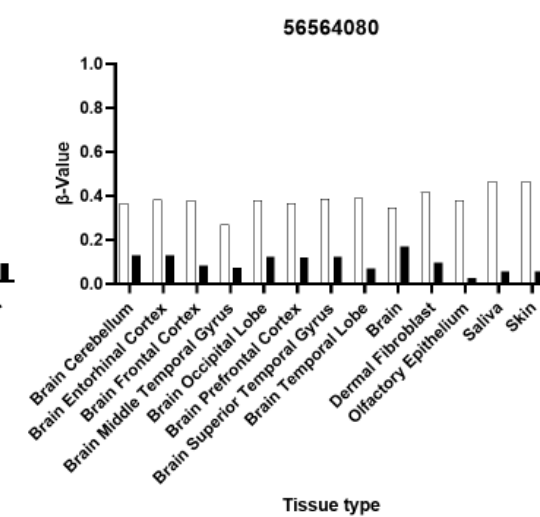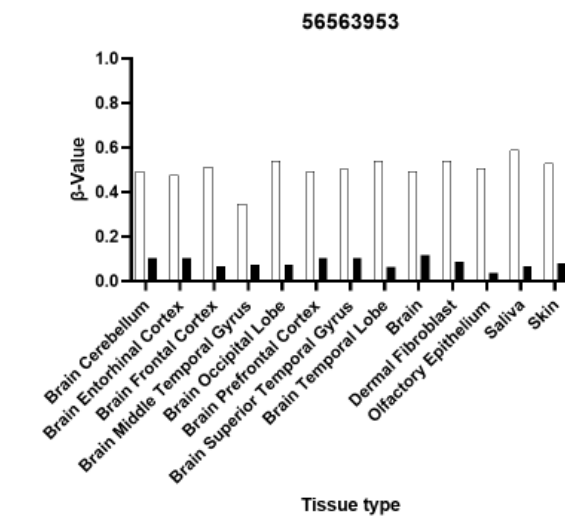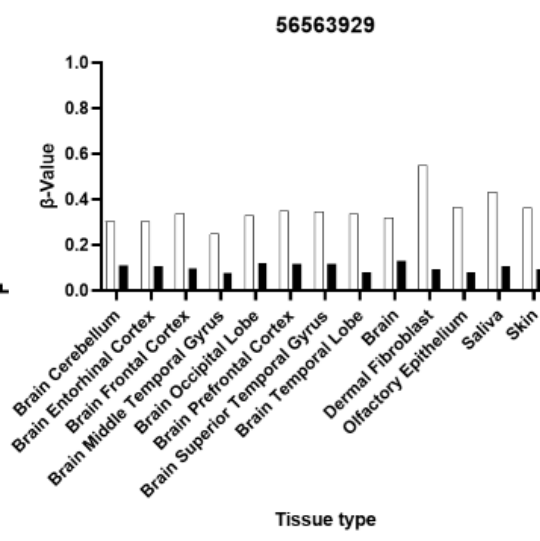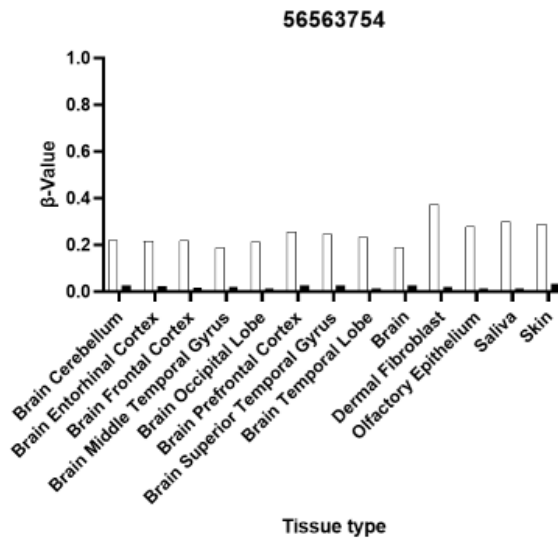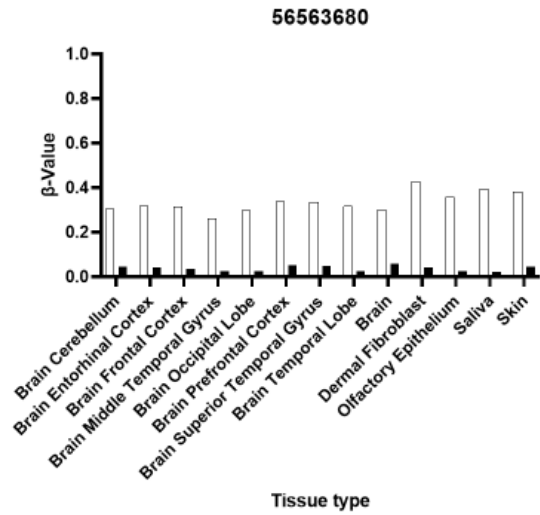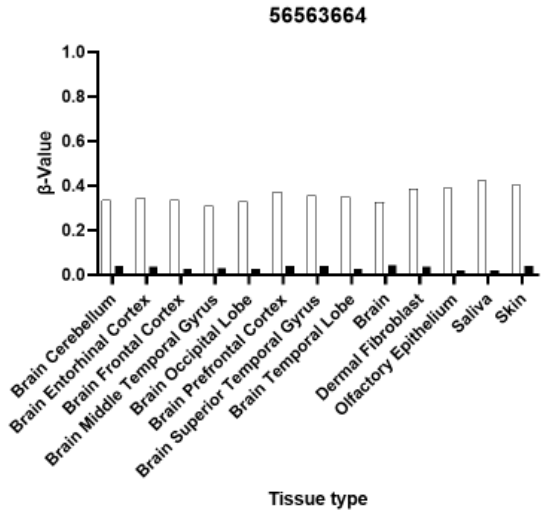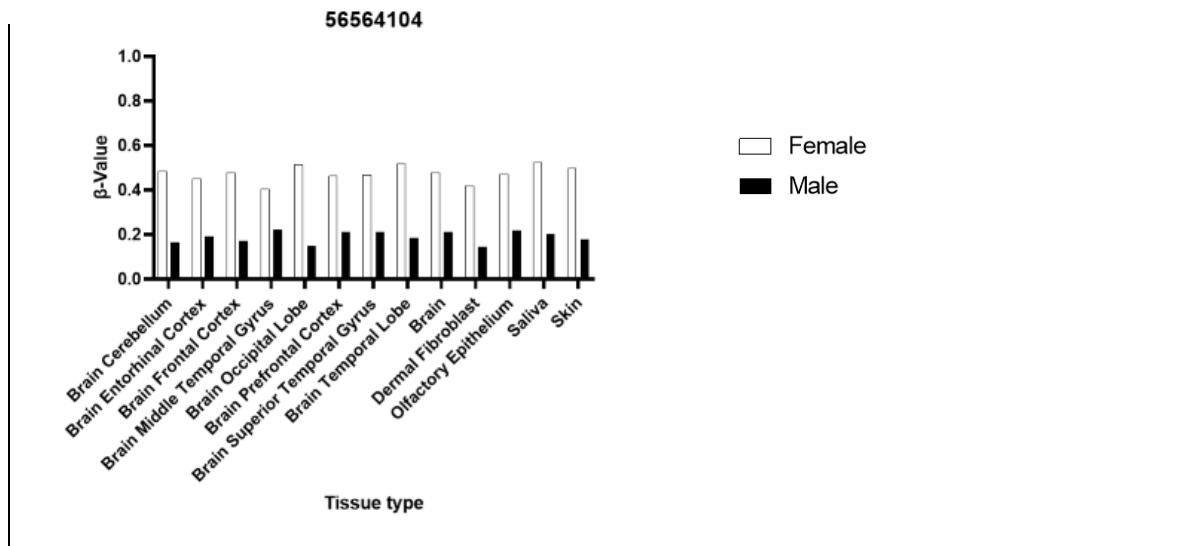
Appendix B: Average β-value for a given CpG site in CpG island of *UBQLN2* is conserved between tissues. Methylome data from MethBank 3.0 demonstrates a similar methylation landscape for males and females across all tissue types. Elevated methylation of each CpG site across the CpG island of *UBQLN2* in females is conserved across all tissue types. Data is presented as average β -value. Raw data was obtained from MethBank 3.0 and processed in GraphPad Prism 9.0.2.

7 REFERENCES

1. Foster LA, Salajegheh MK. Motor Neuron Disease: Pathophysiology, Diagnosis, and Management. *Am J Med* 2019; 132: 32–37.

2. Arora RD, Khan YS. Motor Neuron Disease. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, https://www.ncbi.nlm.nih.gov/pubmed/32809609 (2021).

3. Abramzon YA, Fratta P, Traynor BJ, et al. The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Front Neurosci* 2020; 14: 42.

4. Zarei S, Carr K, Reiley L, et al. A comprehensive review of amyotrophic lateral sclerosis. *Surg Neurol Int* 2015; 6: 171.

5. Deng H-X, Chen W, Hong S-T, et al. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature* 2011; 477: 211–215.

6. Renaud L, Picher-Martel V, Codron P, et al. Key role of UBQLN2 in pathogenesis of amyotrophic lateral sclerosis and frontotemporal dementia. *Acta Neuropathol Commun* 2019; 7: 103.

7. Fahed AC, McDonough B, Gouvion CM, et al. UBQLN2 mutation causing heterogeneous X-linked dominant neurodegeneration. *Ann Neurol* 2014; 75: 793–798.

8. Longinetti E, Fang F. Epidemiology of amyotrophic lateral sclerosis: an update of recent literature. *Curr Opin Neurol* 2019; 32: 771–776.

9. Cao MC, Chancellor A, Charleston A, et al. Motor neuron disease mortality rates in New Zealand 1992-2013. *Amyotroph Lateral Scler Frontotemporal Degener* 2018; 19: 285–293.

10. Masrori P, Van Damme P. Amyotrophic lateral sclerosis: a clinical review. *Eur J Neurol* 2020; 27: 1918–1929.

11. Ingre C, Roos PM, Piehl F, et al. Risk factors for amyotrophic lateral sclerosis. *Clin Epidemiol* 2015; 7: 181–193.

12. Tao Q-Q, Wu Z-Y. Amyotrophic Lateral Sclerosis: Precise Diagnosis and Individualized Treatment. *Chin Med J* 2017; 130: 2269–2272.

13. Shen D, Yang X, Wang Y, et al. The Gold Coast criteria increases the diagnostic sensitivity for amyotrophic lateral sclerosis in a Chinese population. *Transl Neurodegener* 2021; 10: 28.

14. Brotman RG, Moreno-Escobar MC, Joseph J, et al. Amyotrophic Lateral Sclerosis. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, https://www.ncbi.nlm.nih.gov/pubmed/32310611 (2021).

15. Siddique T, Ajroud-Driss S. Familial amyotrophic lateral sclerosis, a historical perspective. *Acta Myol* 2011; 30: 117–120.

16. Chen S, Sayana P, Zhang X, et al. Genetics of amyotrophic lateral sclerosis: an update. *Mol Neurodegener* 2013; 8: 28.

17. Rechtman L, Jordan H, Wagner L, et al. Racial and ethnic differences among amyotrophic lateral sclerosis cases in the United States. *Amyotroph Lateral Scler Frontotemporal Degener* 2015; 16: 65–71.

18. Palese F, Sartori A, Verriello L, et al. Epidemiology of amyotrophic lateral sclerosis in Friuli-Venezia Giulia, North-Eastern Italy, 2002-2014: a retrospective population-based study. *Amyotroph Lateral Scler Frontotemporal Degener* 2019; 20: 90–99.

19. Benjaminsen E, Alstadhaug KB, Gulsvik M, et al. Amyotrophic lateral sclerosis in Nordland county, Norway, 2000-2015: prevalence, incidence, and clinical features. *Amyotroph Lateral Scler Frontotemporal Degener* 2018; 19: 522–527.

20. Leighton DJ, Newton J, Stephenson LJ, et al. Changing epidemiology of motor neurone disease in Scotland. *J Neurol* 2019; 266: 817–825.

21. Jun KY, Park J, Oh K-W, et al. Epidemiology of ALS in Korea using nationwide big data. *J Neurol Neurosurg Psychiatry* 2019; 90: 395–403.

22. Zhou S, Zhou Y, Qian S, et al. Amyotrophic lateral sclerosis in Beijing: Epidemiologic features and prognosis from 2010 to 2015. *Brain Behav* 2018; 8: e01131.

23. Fong KY, Yu YL, Chan YW, et al. Motor neuron disease in Hong Kong Chinese: epidemiology and clinical picture. *Neuroepidemiology* 1996; 15: 239–245.

24. Fong GCY, Cheng TS, Lam K, et al. An epidemiological study of motor neuron disease in Hong Kong. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2005; 6: 164–168.

25. Tsai C-P, Wang K-C, Hwang C-S, et al. Incidence, prevalence, and medical expenditures of classical amyotrophic lateral sclerosis in Taiwan, 1999-2008. *J Formos Med Assoc* 2015; 114: 612–619.

26. Marin B, Boumédiene F, Logroscino G, et al. Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis. *Int J Epidemiol* 2017; 46: 57–74.

27. Khan I, De Jesus O. *Frontotemporal Lobe Dementia*. StatPearls Publishing, https://www.ncbi.nlm.nih.gov/books/NBK559286/ (2022, accessed February 9, 2023).

28. Ling S-C, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* 2013; 79: 416–438.

29. Saxon JA, Thompson JC, Jones M, et al. Examining the language and behavioural profile in FTD and ALS-FTD. *J Neurol Neurosurg Psychiatry* 2017; 88: 675–680.

30. Ferrari R, Kapogiannis D, Huey ED, et al. FTD and ALS: a tale of two diseases. *Curr Alzheimer Res* 2011; 8: 273–294.

31. Al-Chalabi A, Calvo A, Chio A, et al. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. *Lancet Neurol* 2014; 13: 1108–1113.

32. van den Bos MAJ, Geevasinga N, Higashihara M, et al. Pathophysiology and Diagnosis of ALS: Insights from Advances in Neurophysiological Techniques. *Int J Mol Sci*; 20. Epub ahead of print June 10, 2019. DOI: 10.3390/ijms20112818.

33. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* 2014; 17: 17–23.

34. Mejzini R, Flynn LL, Pitout IL, et al. ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now? *Front Neurosci* 2019; 13: 1310.

35. Kim G, Gautier O, Tassoni-Tsuchida E, et al. ALS Genetics: Gains, Losses, and Implications for Future Therapies. *Neuron* 2020; 108: 822–842.

36. Mathis S, Goizet C, Soulages A, et al. Genetics of amyotrophic lateral sclerosis: A review. *J Neurol Sci* 2019; 399: 217–226.

37. Cveticanin J, Mondal T, Meiering EM, et al. Insight into the Autosomal-Dominant Inheritance Pattern of SOD1-Associated ALS from Native Mass Spectrometry. *J Mol Biol* 2020; 432: 5995–6002.

38. Huai J, Zhang Z. Structural Properties and Interaction Partners of Familial ALS-Associated SOD1 Mutants. *Front Neurol* 2019; 10: 527.

39. Pansarasa O, Bordoni M, Diamanti L, et al. SOD1 in Amyotrophic Lateral Sclerosis: "Ambivalent" Behavior Connected to the Disease. *Int J Mol Sci*; 19. Epub ahead of print May 3, 2018. DOI: 10.3390/ijms19051345.

40. Motataianu A, Serban G, Barcutean L, et al. Oxidative Stress in Amyotrophic Lateral Sclerosis: Synergy of Genetic and Environmental Factors. *Int J Mol Sci*; 23. Epub ahead of print August 19, 2022. DOI: 10.3390/ijms23169339.

41. Prudencio M, Hart PJ, Borchelt DR, et al. Variation in aggregation propensities among ALS-associated variants of SOD1: correlation to human disease. *Hum Mol Genet* 2009; 18: 3217–3226.

42. Bunton-Stasyshyn RKA, Saccon RA, Fratta P, et al. SOD1 Function and Its Implications for Amyotrophic Lateral Sclerosis Pathology: New and Renascent Themes. *Neuroscientist* 2015; 21: 519–529.

43. Wicks P, Abrahams S, Papps B, et al. SOD1 and cognitive dysfunction in familial amyotrophic lateral sclerosis. *J Neurol* 2009; 256: 234–241.

44. Katz JS, Katzberg HD, Woolley SC, et al. Combined fulminant frontotemporal dementia and amyotrophic lateral sclerosis associated with an I113T SOD1 mutation. *Amyotroph Lateral Scler* 2012; 13: 567–569.

45. Park JH, Elpers C, Reunert J, et al. SOD1 deficiency: a novel syndrome distinct from amyotrophic lateral sclerosis. *Brain* 2019; 142: 2230–2237.

46. Van Deerlin VM, Leverenz JB, Bekris LM, et al. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *Lancet Neurol* 2008; 7: 409–416.

47. Chen-Plotkin AS, Lee VM-Y, Trojanowski JQ. TAR DNA-binding protein 43 in neurodegenerative disease. *Nat Rev Neurol* 2010; 6: 211–220.

48. Xu F, Huang S, Li X-Y, et al. Identification of TARDBP Gly298Ser as a founder mutation for amyotrophic lateral sclerosis in Southern China. *BMC Med Genomics* 2022; 15: 173.

49. Giannini M, Bayona-Feliu A, Sproviero D, et al. TDP-43 mutations link Amyotrophic Lateral Sclerosis with R-loop homeostasis and R loop-mediated DNA damage. *PLoS Genet* 2020; 16: e1009260.

50. Jo M, Lee S, Jeon Y-M, et al. The role of TDP-43 propagation in neurodegenerative diseases: integrating insights from clinical and experimental studies. *Exp Mol Med* 2020; 52: 1652–1662.

51. Cairns NJ, Neumann M, Bigio EH, et al. TDP-43 in familial and sporadic frontotemporal lobar degeneration with ubiquitin inclusions. *Am J Pathol* 2007; 171: 227–240.

52. Prasad A, Bharathi V, Sivalingam V, et al. Molecular Mechanisms of TDP-43 Misfolding and Pathology in Amyotrophic Lateral Sclerosis. *Front Mol Neurosci* 2019; 12: 25.

53. Scotter EL, Chen H-J, Shaw CE. TDP-43 Proteinopathy and ALS: Insights into Disease Mechanisms and Therapeutic Targets. *Neurotherapeutics* 2015; 12: 352–363.

54. Brettschneider J, Del Tredici K, Toledo JB, et al. Stages of pTDP-43 pathology in amyotrophic lateral sclerosis. *Ann Neurol* 2013; 74: 20–38.

55. Lee EB, Lee VM-Y, Trojanowski JQ. Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nat Rev Neurosci* 2011; 13: 38–50.

56. Smeyers J, Banchi E-G, Latouche M. C9ORF72: What It Is, What It Does, and Why It Matters. *Front Cell Neurosci* 2021; 15: 661447.

57. Pang W, Hu F. Cellular and physiological functions of C9ORF72 and implications for ALS/FTD. *J Neurochem* 2021; 157: 334–350.

58. Umoh ME, Fournier C, Li Y, et al. Comparative analysis of C9orf72 and sporadic disease in an ALS clinic population. *Neurology* 2016; 87: 1024–1030.

59. Majounie E, Renton AE, Mok K, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol* 2012; 11: 323–330.

60. Gossye H, Engelborghs S, Van Broeckhoven C, et al. *C9orf72 Frontotemporal Dementia and/or Amyotrophic Lateral Sclerosis*. University of Washington, Seattle, https://www.ncbi.nlm.nih.gov/books/NBK268647/ (2020, accessed February 12, 2023).

61. Balendra R, Isaacs AM. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol* 2018; 14: 544–558.

62. Ryan S, Rollinson S, Hobbs E, et al. C9orf72 dipeptides disrupt the nucleocytoplasmic transport machinery and cause TDP-43 mislocalisation to the cytoplasm. *Sci Rep* 2022; 12: 4799.

63. Edbauer D, Haass C. An amyloid-like cascade hypothesis for C9orf72 ALS/FTD. *Curr Opin Neurobiol* 2016; 36: 99–106.

64. Cook CN, Wu Y, Odeh HM, et al. C9orf72 poly(GR) aggregation induces TDP-43 proteinopathy. *Sci Transl Med*; 12. Epub ahead of print September 2, 2020. DOI: 10.1126/scitranslmed.abb3774.

65. Andersen PM, Al-Chalabi A. Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nat Rev Neurol* 2011; 7: 603–615.

66. Nolan M, Talbot K, Ansorge O. Pathogenesis of FUS-associated ALS and FTD: insights from rodent models. *Acta Neuropathol Commun* 2016; 4: 99.

67. Ratti A, Buratti E. Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. *J Neurochem* 2016; 138 Suppl 1: 95–111.

68. Ishigaki S, Sobue G. Importance of Functional Loss of FUS in FTLD/ALS. *Front Mol Biosci* 2018; 5: 44.

69. Birsa N, Bentham MP, Fratta P. Cytoplasmic functions of TDP-43 and FUS and their role in ALS. *Semin Cell Dev Biol* 2020; 99: 193–201.

70. Nakaya T, Maragkakis M. Amyotrophic Lateral Sclerosis associated FUS mutation shortens mitochondria and induces neurotoxicity. *Sci Rep* 2018; 8: 15575.

71. Daigle JG, Lanson NA Jr, Smith RB, et al. RNA-binding ability of FUS regulates neurodegeneration, cytoplasmic mislocalization and incorporation into stress granules associated with FUS carrying ALS-linked mutations. *Hum Mol Genet* 2013; 22: 1193–1205.

72. Bentmann E, Neumann M, Tahirovic S, et al. Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). *J Biol Chem* 2012; 287: 23079–23094.

73. Bosco DA, Lemay N, Ko HK, et al. Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. *Hum Mol Genet* 2010; 19: 4160–4175.

74. Vance C, Scotter EL, Nishimura AL, et al. ALS mutant FUS disrupts nuclear localization and sequesters wild-type FUS within cytoplasmic stress granules. *Hum Mol Genet* 2013; 22: 2676–2688.

75. Zheng T, Yang Y, Castañeda CA. Structure, dynamics and functions of UBQLNs: at the crossroads of protein quality control machinery. *Biochem J* 2020; 477: 3471–3497.

76. Tanaka K. The proteasome: overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci* 2009; 85: 12–36.

77. Liu WJ, Ye L, Huang WF, et al. p62 links the autophagy pathway and the ubiqutin-proteasome system upon ubiquitinated protein degradation. *Cell Mol Biol Lett* 2016; 21: 29.

78. Verma K, Verma M, Chaphalkar A, et al. Recent advances in understanding the role of proteostasis. *Fac Rev* 2021; 10: 72.

79. Teyssou E, Chartier L, Amador M-D-M, et al. Novel UBQLN2 mutations linked to amyotrophic lateral sclerosis and atypical hereditary spastic paraplegia phenotype through defective HSP70-mediated proteolysis. *Neurobiol Aging* 2017; 58: 239.e11-239.e20.

80. Hjerpe R, Bett JS, Keuss MJ, et al. UBQLN2 Mediates Autophagy-Independent Protein Aggregate Clearance by the Proteasome. *Cell* 2016; 166: 935–949.

81. Chang L, Monteiro MJ. Defective Proteasome Delivery of Polyubiquitinated Proteins by Ubiquilin-2 Proteins Containing ALS Mutations. *PLoS One* 2015; 10: e0130162.

82. Osaka M, Ito D, Suzuki N. Disturbance of proteasomal and autophagic protein degradation pathways by amyotrophic lateral sclerosis-linked mutations in ubiquilin 2. *Biochem Biophys Res Commun* 2016; 472: 324–331.

83. Parzych KR, Klionsky DJ. An overview of autophagy: morphology, mechanism, and regulation. *Antioxid Redox Signal* 2014; 20: 460–473.

84. Feng Y, He D, Yao Z, et al. The machinery of macroautophagy. *Cell Res* 2014; 24: 24–41.

85. Klionsky DJ, Schulman BA. Dynamic regulation of macroautophagy by distinctive ubiquitin-like proteins. *Nat Struct Mol Biol* 2014; 21: 336–345.

86. Chen T, Huang B, Shi X, et al. Mutant UBQLN2P497H in motor neurons leads to ALS-like phenotypes and defective autophagy in rats. *Acta Neuropathol Commun* 2018; 6: 122.

87. Rothenberg C, Srinivasan D, Mah L, et al. Ubiquilin functions in autophagy and is degraded by chaperone-mediated autophagy. *Hum Mol Genet* 2010; 19: 3219–3232.

88. Şentürk M, Mao D, Bellen HJ. Loss of proteins associated with amyotrophic lateral sclerosis affects lysosomal acidification via different routes. *Autophagy* 2019; 15: 1467–1469.

89. Wu JJ, Cai A, Greenslade JE, et al. ALS/FTD mutations in UBQLN2 impede autophagy by reducing autophagosome acidification through loss of function. *Proc Natl Acad Sci U S A* 2020; 117: 15230–15241.

90. Şentürk M, Lin G, Zuo Z, et al. Ubiquilins regulate autophagic flux through mTOR signalling and lysosomal acidification. *Nat Cell Biol* 2019; 21: 384–396.

91. Wang J, Gan Y, Cao J, et al. Pathophysiology of stress granules: An emerging link to diseases (Review). *Int J Mol Med*; 49. Epub ahead of print April 2022. DOI: 10.3892/ijmm.2022.5099.

92. Protter DSW, Parker R. Principles and Properties of Stress Granules. *Trends Cell Biol* 2016; 26: 668–679.

93. Higgins N, Lin B, Monteiro MJ. Lou Gehrig's Disease (ALS): UBQLN2 Mutations Strike Out of Phase. *Structure* 2019; 27: 879–881.

94.   Dudman J, Qi X. Stress Granule Dysregulation in Amyotrophic Lateral Sclerosis. *Front Cell Neurosci* 2020; 14: 598517.

95.   Molliex A, Temirov J, Lee J, et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 2015; 163: 123–133.

96.   Hayashi Y, Ford LK, Fioriti L, et al. Liquid-Liquid Phase Separation in Physiology and Pathophysiology of the Nervous System. *J Neurosci* 2021; 41: 834–844.

97.   Dao TP, Kolaitis R-M, Kim HJ, et al. Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Mol Cell* 2018; 69: 965-978.e6.

98.   Dao TP, Martyniak B, Canning AJ, et al. ALS-Linked Mutations Affect UBQLN2 Oligomerization and Phase Separation in a Position- and Amino Acid-Dependent Manner. *Structure* 2019; 27: 937-951.e5.

99.   Alexander EJ, Ghanbari Niaki A, Zhang T, et al. Ubiquilin 2 modulates ALS/FTD-linked FUS-RNA complex dynamics and stress granule formation. *Proc Natl Acad Sci U S A* 2018; 115: E11485–E11494.

100.   Kim SH, Stiles SG, Feichtmeier JM, et al. Mutation-dependent aggregation and toxicity in a Drosophila model for UBQLN2-associated ALS. *Hum Mol Genet* 2018; 27: 322–337.

101.   Sheng Y, Chattopadhyay M, Whitelegge J, et al. SOD1 aggregation and ALS: role of metallation states and disulfide status. *Curr Top Med Chem* 2012; 12: 2560–2572.

102.   Blokhuis AM, Groen EJN, Koppers M, et al. Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathol* 2013; 125: 777–794.

103.   Saxton AD, Kraemer BC. Human Ubiquilin 2 and TDP-43 copathology drives neurodegeneration in transgenic Caenorhabditis elegans. *G3*; 11. Epub ahead of print August 7, 2021. DOI: 10.1093/g3journal/jkab158.

104.   Ceballos-Diaz C, Rosario AM, Park H-J, et al. Viral expression of ALS-linked ubiquilin-2 mutants causes inclusion pathology and behavioral deficits in mice. *Mol Neurodegener* 2015; 10: 25.

105.   Brettschneider J, Van Deerlin VM, Robinson JL, et al. Pattern of ubiquilin pathology in ALS and FTLD indicates presence of C9ORF72 hexanucleotide expansion. *Acta Neuropathol* 2012; 123: 825–839.

106.   Cassel JA, Reitz AB. Ubiquilin-2 (UBQLN2) binds with high affinity to the C-terminal region of TDP-43 and modulates TDP-43 levels in H4 cells: characterization of inhibition by nucleic acids and 4-aminoquinolines. *Biochim Biophys Acta* 2013; 1834: 964–971.

107.   Le NTT, Chang L, Kovlyagina I, et al. Motor neuron disease, TDP-43 pathology, and memory deficits in mice expressing ALS-FTD-linked UBQLN2 mutations. *Proc Natl Acad Sci U S A* 2016; 113: E7580–E7589.

108.   Jantrapirom S, Lo Piccolo L, Yoshida H, et al. Depletion of Ubiquilin induces an augmentation in soluble ubiquitinated Drosophila TDP-43 to drive neurotoxicity in the fly. *Biochim Biophys Acta Mol Basis Dis* 2018; 1864: 3038–3049.

109.   Picher-Martel V, Dutta K, Phaneuf D, et al. Ubiquilin-2 drives NF-κB activity and cytosolic TDP-43 aggregation in neuronal cells. *Mol Brain* 2015; 8: 71.

110.   Nölle A, van Haastert ES, Zwart R, et al. Ubiquilin 2 is not associated with tau pathology. *PLoS One* 2013; 8: e76598.

111.   Moosavi A, Motevalizadeh Ardekani A. Role of Epigenetics in Biology and Human Diseases. *Iran Biomed J* 2016; 20: 246–258.

112.   Panning B. X-chromosome inactivation: the molecular basis of silencing. *J Biol* 2008; 7: 30.

113.   Bohacek J, Mansuy IM. Epigenetic inheritance of disease and disease risk. *Neuropsychopharmacology* 2013; 38: 220–236.

114.   Loscalzo J, Handy DE. Epigenetic modifications: basic mechanisms and role in cardiovascular disease (2013 Grover Conference series). *Pulm Circ* 2014; 4: 169–174.

115.   Grunstein M. Histone acetylation in chromatin structure and transcription. *Nature* 1997; 389: 349–352.

116.   Cotton AM, Price EM, Jones MJ, et al. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol Genet* 2015; 24: 1528–1539.

117.   Hall E, Volkov P, Dayeh T, et al. Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biol* 2014; 15: 522.

118.   Herceg Z, Luque EH. CpG Island. DOI: 10.1016/B978-0-444-54299-1.00002-9.

119.   Visone R, Bacalini MG, Di Franco S, et al. DNA methylation of shelf, shore and open sea CpG positions distinguish high microsatellite instability from low or stable microsatellite status colon cancer stem cells. *Epigenomics* 2019; 11: 587–604.

120.   Li G, Raffield L, Logue M, et al. CUE: CpG impUtation Ensemble for DNA Methylation Levels Across the Human Methylation450 (HM450) and EPIC (HM850) BeadChip Platforms. *bioRxiv* 2020; 2020.05.30.107094.

121.   Delbridge ML, Graves JA. Mammalian Y chromosome evolution and the male-specific functions of Y chromosome-borne genes. *Rev Reprod* 1999; 4: 101–109.

122.   Fang H, Disteche CM, Berletch JB. X Inactivation and Escape: Epigenetic and Structural Features. *Front Cell Dev Biol* 2019; 7: 219.

123.   Patrat C, Ouimette J-F, Rougeulle C. X chromosome inactivation in human development. *Development*; 147. Epub ahead of print January 3, 2020. DOI: 10.1242/dev.183095.

124.   van den Berg IM, Laven JSE, Stevens M, et al. X chromosome inactivation is initiated in human preimplantation embryos. *Am J Hum Genet* 2009; 84: 771–779.

125.   Lu Z, Carter AC, Chang HY. Mechanistic insights in X-chromosome inactivation. *Philos Trans R Soc Lond B Biol Sci*; 372. Epub ahead of print November 5, 2017. DOI: 10.1098/rstb.2016.0356.

126.   Galupa R, Heard E. X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annu Rev Genet* 2018; 52: 535–566.

127.   Passarge E. Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. *Am J Hum Genet* 1979; 31: 106–115.

128.   Barr ML, Bertram EG. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* 1949; 163: 676.

129.   Miller FA. 'Your true and proper gender': the Barr body as a good enough science of sex. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 2006; 37: 459–483.

130.   Migeon BR. X-linked diseases: susceptible females. *Genet Med* 2020; 22: 1156–1174.

131.   Minks J, Robinson WP, Brown CJ. A skewed view of X chromosome inactivation. *The Journal of clinical investigation* 2008; 118: 20–23.

132.   Sun Z, Fan J, Wang Y. X-Chromosome Inactivation and Related Diseases. *Genet Res* 2022; 2022: 1391807.

133.   Lupski JR, Garcia CA, Zoghbi HY, et al. Discordance of muscular dystrophy in monozygotic female twins: evidence supporting asymmetric splitting of the inner cell mass in a manifesting carrier of Duchenne dystrophy. *Am J Med Genet* 1991; 40: 354–364.

134.   Puck JM, Willard HF. X inactivation in females with X-linked disease. *The New England journal of medicine* 1998; 338: 325–328.

135.   Germain DP. *General aspects of X-linked diseases*. Oxford PharmaGenesis, https://www.ncbi.nlm.nih.gov/books/NBK11593/ (2006, accessed April 26, 2023).

136.   Duan D, Goemans N, Takeda S, et al. Duchenne muscular dystrophy. *Nat Rev Dis Primers* 2021; 7: 13.

137.   Genetic Alliance, The New York-Mid-Atlantic Consortium for Genetic and Newborn Screening Services. *INHERITANCE PATTERNS*. Genetic Alliance, https://www.ncbi.nlm.nih.gov/books/NBK115561/ (2009, accessed April 26, 2023).

138.   Stone WL, Basit H, Los E. *Fragile X Syndrome*. StatPearls Publishing, https://www.ncbi.nlm.nih.gov/books/NBK459243/ (2022, accessed March 4, 2023).

139.   Bartholomay KL, Lee CH, Bruno JL, et al. Closing the Gender Gap in Fragile X Syndrome: Review on Females with FXS and Preliminary Research Findings. *Brain Sci*; 9. Epub ahead of print January 12, 2019. DOI: 10.3390/brainsci9010011.

140.   Payán-Gómez C, Ramirez-Cheyne J, Saldarriaga W. Variable Expressivity in Fragile X Syndrome: Towards the Identification of Molecular Characteristics That Modify the Phenotype. *Appl Clin Genet* 2021; 14: 305–312.

141.   Young JI, Zoghbi HY. X-chromosome inactivation patterns are unbalanced and affect the phenotypic outcome in a mouse model of rett syndrome. *Am J Hum Genet* 2004; 74: 511–520.

142.   De Riso G, Cuomo M, Di Risi T, et al. Ultra-Deep DNA Methylation Analysis of X-Linked Genes: GLA and AR as Model Genes. *Genes* ; 11. Epub ahead of print June 4, 2020. DOI: 10.3390/genes11060620.

143.   Allen RC, Zoghbi HY, Moseley AB, et al. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet* 1992; 51: 1229–1239.

144.   Ugozzoli L, Wallace RB. Allele-specific polymerase chain reaction. *Methods* 1991; 2: 42–48.

145.   Lee HB, Schwab TL, Koleilat A, et al. Allele-Specific Quantitative PCR for Accurate, Rapid, and Cost-Effective Genotyping. *Hum Gene Ther* 2016; 27: 425–435.

146.   Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet* 2019; 20: 631–656.

147.   Tukiainen T, Villani A-C, Yen A, et al. Landscape of X chromosome inactivation across human tissues. *Nature* 2017; 550: 244–248.

148.   Nguyen CT, Gonzales FA, Jones PA. Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation. *Nucleic Acids Res* 2001; 29: 4598–4606.

149.   Tusnády GE, Simon I, Váradi A, et al. BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Res* 2005; 33: e9.

150.   Li Y, Tollefsbol TO. DNA methylation detection: bisulfite genomic sequencing analysis. *Methods Mol Biol* 2011; 791: 11–21.

151.   Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 1992; 89: 1827–1831.

152.   Li L-C. Designing PCR Primer for DNA Methylation Mapping. In: Yuryev A (ed) *PCR Primer Design*. Totowa, NJ: Humana Press, pp. 370–383.

153.   Harris E, Lonardi S. Bisulfite-Conversion-Based Methods for DNA Methylation Sequencing Data Analysis. In: *Computational Methods for Next Generation Sequencing Data Analysis*, pp. 227–243.

154. Bonini NM, Berger SL. The Sustained Impact of Model Organisms in Genetics and Epigenetics. *Genetics* 2017; 205: 1–4.

155. Wangler MF, Yamamoto S, Chao H-T, et al. Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research. *Genetics* 2017; 207: 9–27.

156. Mukherjee P, Roy S, Ghosh D, et al. Role of animal models in biomedical research: a review. *Lab Anim Res* 2022; 38: 18.

157. Bryda EC. The Mighty Mouse: the impact of rodents on advances in biomedical research. *Mo Med* 2013; 110: 207–211.

158. Picher-Martel V, Valdmanis PN, Gould PV, et al. From animal models to human disease: a genetic approach for personalized medicine in ALS. *Acta Neuropathol Commun* 2016; 4: 70.

159. Gorrie GH, Fecto F, Radzicki D, et al. Dendritic spinopathy in transgenic mice expressing ALS/dementia-linked mutant UBQLN2. *Proc Natl Acad Sci U S A* 2014; 111: 14524–14529.

160. Sharkey LM, Sandoval-Pistorius SS, Moore SJ, et al. Modeling UBQLN2-mediated neurodegenerative disease in mice: Shared and divergent properties of wild type and mutant UBQLN2 in phase separation, subcellular localization, altered proteostasis pathways, and selective cytotoxicity. *Neurobiol Dis* 2020; 143: 105016.

161. Wu Q, Liu M, Huang C, et al. Pathogenic Ubqln2 gains toxic properties to induce neuron death. *Acta Neuropathol* 2015; 129: 417–428.

162. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res* 2002; 12: 996–1006.

163. UBQLN2 ubiquilin 2 [Homo sapiens (human)] - Gene - NCBI, https://www.ncbi.nlm.nih.gov/gene/29978 (accessed April 14, 2023).

164. Rechache NS, Wang Y, Stevenson HS, et al. DNA methylation profiling identifies global methylation differences and markers of adrenocortical tumors. *J Clin Endocrinol Metab* 2012; 97: E1004-13.

165. Campagna MP, Xavier A, Lechner-Scott J, et al. Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clin Epigenetics* 2021; 13: 214.

166. Xiong Z, Yang F, Li M, et al. EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. *Nucleic Acids Res* 2022; 50: D1004–D1009.

167. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002; 30: 207–210.

168. Xiong Z, Li M, Yang F, et al. EWAS Data Hub: a resource of DNA methylation array data and metadata. *Nucleic Acids Res* 2020; 48: D890–D895.

169.   Dedeurwaerder S, Defrance M, Bizet M, et al. A comprehensive overview of Infinium HumanMethylation450 data processing. *Brief Bioinform* 2014; 15: 929–941.

170.   Li R, Liang F, Li M, et al. MethBank 3.0: a database of DNA methylomes across a variety of species. *Nucleic Acids Res* 2018; 46: D288–D295.

171.   Wang Y, Song F, Zhu J, et al. GSA: Genome Sequence Archive<sup/>. *Genomics Proteomics Bioinformatics* 2017; 15: 14–18.

172.   Kodama Y, Shumway M, Leinonen R, et al. The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res* 2012; 40: D54-6.

173.   Birney E, Andrews TD, Bevan P, et al. An overview of Ensembl. *Genome Res* 2004; 14: 925–928.

174.   Assembly   Converter, http://asia.ensembl.org/Homo_sapiens/Tools/AssemblyConverter (2022).

175.   Torday JS, Baluška F. Why control an experiment?: From empiricism, via consciousness, toward Implicate Order. *EMBO Rep* 2019; 20: e49110.

176.   Balaton BP, Fornes O, Wasserman WW, et al. Cross-species examination of X-chromosome inactivation highlights domains of escape from silencing. *Epigenetics Chromatin* 2021; 14: 12.

177.   Du P, Zhang X, Huang C-C, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010; 11: 587.

178.   Qamar W, Khan MR, Arafah A. Optimization of conditions to extract high quality DNA for PCR analysis from whole blood using SDS-proteinase K method. *Saudi J Biol Sci* 2017; 24: 1465–1469.

179.   QIAGEN.   [No   title], https://www.qiagen.com/us/resources/resourcedetail?id=eafbbce2-43ed-4e30-880d-f559f2e15195&lang=en (accessed April 4, 2023).

180.   Untergasser A, Cutcutache I, Koressaar T, et al. Primer3--new capabilities and interfaces. *Nucleic Acids Res* 2012; 40: e115.

181.   Li L-C, Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 2002; 18: 1427–1431.

182.   Hints for Optimizing Bisulfite Primer Design. *EpiGenie | Epigenetics, Stem Cell, and Synthetic Biology News*, https://epigenie.com/hints-for-optimizing-bisulfite-primer-design/ (2014, accessed April 28, 2023).

183.   BioSystems A. Detection and Quantification of Sequence Variants from Sanger Sequencing Traces.

184.   Wang Z, Wu X, Wang Y. A framework for analyzing DNA methylation data from Illumina Infinium HumanMethylation450 BeadChip. *BMC Bioinformatics* 2018; 19: 115.

185. Masago K, Fujita S, Oya Y, et al. Comparison between Fluorimetry (Qubit) and Spectrophotometry (NanoDrop) in the Quantification of DNA and RNA Extracted from Frozen and FFPE Tissues from Lung Cancer Patients: A Real-World Use of Genomic Tests. *Medicina* ; 57. Epub ahead of print December 17, 2021. DOI: 10.3390/medicina57121375.

186. Khetan D, Gupta N, Chaudhary R, et al. Comparison of UV spectrometry and fluorometry-based methods for quantification of cell-free DNA in red cell components. *Asian J Transfus Sci* 2019; 13: 95–99.

187. Bisulfite Beginners Guide 101. *Zymo Research International*, https://zymoresearch.eu/pages/bisulfite-beginner-guide (accessed April 14, 2023).

188. Hill JT, Demarest BL, Bisgrove BW, et al. Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. *Dev Dyn* 2014; 243: 1632–1636.

189. Sielemann K, Hafner A, Pucker B. The reuse of public datasets in the life sciences: potential risks and rewards. *PeerJ* 2020; 8: e9954.

190. Mathur S, Sutton J. Personalized medicine could transform healthcare. *Biomed Rep* 2017; 7: 3–5.

191. Xia Y, Dai R, Wang K, et al. Sex-differential DNA methylation and associated regulation networks in human brain implicated in the sex-biased risks of psychiatric disorders. *Mol Psychiatry* 2019; 26: 835–848.

192. García-Calzón S, Perfilyev A, de Mello VD, et al. Sex Differences in the Methylome and Transcriptome of the Human Liver and Circulating HDL-Cholesterol Levels. *J Clin Endocrinol Metab* 2018; 103: 4395–4408.

193. Davegårdh C, Hall Wedin E, Broholm C, et al. Sex influences DNA methylation and gene expression in human skeletal muscle myoblasts and myotubes. *Stem Cell Res Ther* 2019; 10: 26.

194. Cappelli E, Felici G, Weitschek E. Combining DNA methylation and RNA sequencing data of cancer for supervised knowledge extraction. *BioData Min* 2018; 11: 22.

195. McGrath-Morrow SA, Ndeh R, Helmin KA, et al. DNA methylation and gene expression signatures are associated with ataxia-telangiectasia phenotype. *Sci Rep* 2020; 10: 7479.

196. Yu Z, Ueno K, Funayama R, et al. Sex-Specific Differences in the Transcriptome of the Human Dorsolateral Prefrontal Cortex in Schizophrenia. *Mol Neurobiol* 2023; 60: 1083–1098.

197. Gershoni M, Pietrokovski S. The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol* 2017; 15: 7.

198. Lim W-J, Kim KH, Kim J-Y, et al. Identification of DNA-Methylated CpG Islands Associated With Gene Silencing in the Adult Body Tissues of the Ogye Chicken Using RNA-Seq and Reduced Representation Bisulfite Sequencing. *Front Genet* 2019; 10: 346.

199. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. *Cold Spring Harb Protoc* 2015; 2015: 951.

200. Edgar R, Tan PPC, Portales-Casamar E, et al. Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. *Epigenetics Chromatin* 2014; 7: 28.

201. Zhang B, Zhou Y, Lin N, et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res* 2013; 23: 1522–1540.

202. De Bustos C, Ramos E, Young JM, et al. Tissue-specific variation in DNA methylation levels along human chromosome 1. *Epigenetics Chromatin* 2009; 2: 7.

203. Ghosh S, Yates AJ, Frühwald MC, et al. Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. *Epigenetics* 2010; 5: 527–538.

204. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology* 2013; 38: 23–38.

205. Ladd-Acosta C, Pevsner J, Sabunciyan S, et al. DNA methylation signatures within the human brain. *Am J Hum Genet* 2007; 81: 1304–1315.

206. Allingham-Hawkins DJ, Brown CA, Babul R, et al. Tissue-specific methylation differences and cognitive function in fragile X premutation females. *Am J Med Genet* 1996; 64: 329–333.

207. Esanov R, Andrade NS, Bennison S, et al. The FMR1 promoter is selectively hydroxymethylated in primary neurons of fragile X syndrome patients. *Hum Mol Genet* 2016; 25: 4870–4880.

208. Mook J, Hendricksen A, Reece K. Faster, cleaner bisulfite conversion that yields more intact DNA for downstream analysis. *Nat Methods* 2013; 10: i–ii.

209. Halabian R, Valizadeh Arshad, Ahmadi A, et al. Laboratory methods to decipher epigenetic signatures: a comparative review. *Cell Mol Biol Lett* 2021; 26: 46.

210. Sulewska A, Niklinska W, Kozlowski M, et al. Detection of DNA methylation in eucaryotic cells. *Folia Histochem Cytobiol* 2007; 45: 315–324.

211. Hong SR, Shin K-J. Bisulfite-Converted DNA Quantity Evaluation: A Multiplex Quantitative Real-Time PCR System for Evaluation of Bisulfite Conversion. *Front Genet* 2021; 12: 618955.

212. Vidaki A, Planterose Jiménez B, Poggiali B, et al. Targeted DNA methylation analysis and prediction of smoking habits in blood based on massively parallel sequencing. *Forensic Sci Int Genet* 2023; 65: 102878.

213. Abula G, Li J, Ma R, et al. Serum Angiotensin-Converting Enzyme Methylation Level and Its Significance in Patients With Comorbid Major Depressive Disorder and

Hypertension. *Clin Neuropharmacol*. Epub ahead of print April 26, 2023. DOI: 10.1097/WNF.0000000000000551.

214. Loebel DA, Johnston PG. Methylation analysis of a marsupial X-linked CpG island by bisulfite genomic sequencing. *Genome Res* 1996; 6: 114–123.

215. Darst RP, Pardo CE, Ai L, et al. Bisulfite sequencing of DNA. *Curr Protoc Mol Biol* 2010; Chapter 7: Unit 7.9.1-17.

216. Simbolo M, Gottardi M, Corbo V, et al. DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS One* 2013; 8: e62692.

217. How does the DNA Protection work in the EpitTect Bisulfite kits?, https://www.qiagen.com/jp/resources/faq?id=62a1d9ad-48df-40a9-8716-13043499e952&lang=en (accessed April 25, 2023).

218. Kint S, De Spiegelaere W, De Kesel J, et al. Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. *PLoS One* 2018; 13: e0199091.

219. EpiTect Bisulfite Kits, https://www.qiagen.com/us/products/discovery-and-translational-research/epigenetics/dna-methylation/bisulfite-conversion-assays/epitect-bisulfite-kits (accessed April 25, 2023).

220. Shaw KJ, Thain L, Docker PT, et al. The use of carrier RNA to enhance DNA extraction from microfluidic-based silica monoliths. *Anal Chim Acta* 2009; 652: 231–233.

221. Holmes EE, Jung M, Meller S, et al. Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS One* 2014; 9: e93933.

222. Vaisvila R, Ponnaluri VKC, Sun Z, et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res* 2021; 31: 1280–1289.

223. Guo P, Zheng H, Li Y, et al. Hepatocellular carcinoma detection via targeted enzymatic methyl sequencing of plasma cell-free DNA. *Clin Epigenetics* 2023; 15: 2.

224. Werner B, Yuwono NL, Henry C, et al. Circulating cell-free DNA from plasma undergoes less fragmentation during bisulfite treatment than genomic DNA due to low molecular weight. *PLoS One* 2019; 14: e0224338.

225. Human Methylated & Non-methylated DNA Set. *Zymo Research International*, https://zymoresearch.eu/products/human-methylated-non-methylated-dna-set-dna-w-primers (accessed April 25, 2023).

226. Jakubek P, Rajić J, Kuczyńska M, et al. Beyond Antioxidant Activity: Redox Properties of Catechins May Affect Changes in the DNA Methylation Profile-The Example of SRXN1 Gene. *Antioxidants (Basel)*; 12. Epub ahead of print March 20, 2023. DOI: 10.3390/antiox12030754.

227.   Iijima H, Gilmer G, Wang K, et al. Age-related matrix stiffening epigenetically regulates α-Klotho expression and compromises chondrocyte integrity. *Nat Commun* 2023; 14: 18.

228.   Rodriguez-Casanova A, Costa-Fraga N, Castro-Carballeira C, et al. A genome-wide cell-free DNA methylation analysis identifies an episignature associated with metastatic luminal B breast cancer. *Front Cell Dev Biol* 2022; 10: 1016955.

229.   Sasaki M, Anast J, Bassett W, et al. Bisulfite conversion-specific and methylation-specific PCR: a sensitive technique for accurate evaluation of CpG methylation. *Biochem Biophys Res Commun* 2003; 309: 305–309.

230.   Wojdacz TK, Dobrovic A, Hansen LL. Methylation-sensitive high-resolution melting. *Nat Protoc* 2008; 3: 1903–1908.

231.   Grunau C, Clark SJ, Rosenthal A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res* 2001; 29: E65-5.

232.   Lan VTT, Trang VL, Ngan NT, et al. An Internal Control for Evaluating Bisulfite Conversion in the Analysis of Short Stature Homeobox 2 Methylation in Lung Cancer. *Asian Pac J Cancer Prev* 2019; 20: 2435–2443.

233.   Genereux DP, Johnson WC, Burden AF, et al. Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Res* 2008; 36: e150.

234.   Daskalakis NP, Yehuda R. Site-specific methylation changes in the glucocorticoid receptor exon 1F promoter in relation to life adversity: systematic review of contributing factors. *Front Neurosci* 2014; 8: 369.

235.   Liu Y, Rosikiewicz W, Pan Z, et al. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol* 2021; 22: 295.

236.   Silva C, Machado M, Ferrão J, et al. Whole human genome 5'-mC methylation analysis using long read nanopore sequencing. *Epigenetics* 2022; 17: 1961–1975.

237.   Akbari V, Garant J-M, O'Neill K, et al. Genome-wide detection of imprinted differentially methylated regions using nanopore sequencing. *Elife*; 11. Epub ahead of print July 5, 2022. DOI: 10.7554/eLife.77898.

238.   Wongsurawat T, Jenjaroenpun P, De Loose A, et al. A novel Cas9-targeted long-read assay for simultaneous detection of IDH1/2 mutations and clinically relevant MGMT methylation in fresh biopsies of diffuse glioma. *Acta Neuropathol Commun* 2020; 8: 87.

239.   Hoffmann D, Haapasalo A. Patient-derived skin fibroblasts as a model to study frontotemporal lobar degeneration. *Neural Regeneration Res* 2022; 17: 2669–2671.

240.   Riancho J, Arozamena S, López de Munaín A. Dermic-derived fibroblasts for the study of amyotrophic lateral sclerosis. *Neural Regeneration Res* 2020; 15: 2043–2044.

241. Meyer K, Ferraiuolo L, Miranda CJ, et al. Direct conversion of patient fibroblasts demonstrates non-cell autonomous toxicity of astrocytes to motor neurons in familial and sporadic ALS. *Proc Natl Acad Sci U S A* 2014; 111: 829–832.

242. Singh L, San JE, Tegally H, et al. Targeted Sanger sequencing to recover key mutations in SARS-CoV-2 variant genome assemblies produced by next-generation sequencing. *Microb Genom*; 8. Epub ahead of print March 2022. DOI: 10.1099/mgen.0.000774.

243. Hung CL-K, Maiuri T, Bowie LE, et al. A patient-derived cellular model for Huntington's disease reveals phenotypes at clinically relevant CAG lengths. *Mol Biol Cell* 2018; 29: 2809–2820.

244. Xie W, Kagiampakis I, Pan L, et al. DNA Methylation Patterns Separate Senescence from Transformation Potential and Indicate Cancer Risk. *Cancer Cell* 2018; 33: 309-321.e5.

245. Geens M, Seriola A, Barbé L, et al. Female human pluripotent stem cells rapidly lose X chromosome inactivation marks and progress to a skewed methylation pattern during culture. *Mol Hum Reprod* 2016; 22: 285–298.

246. Pomp O, Dreesen O, Leong DFM, et al. Unexpected X chromosome skewing during culture and reprogramming of human somatic cells can be alleviated by exogenous telomerase. *Cell Stem Cell* 2011; 9: 156–165.

247. Lu J, Johnston A, Berichon P, et al. PrimerSuite: A High-Throughput Web-Based Primer Design Program for Multiplex Bisulfite PCR. *Sci Rep* 2017; 7: 41328.

248. Brandes JC, Carraway H, Herman JG. Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter. *Oncogene* 2007; 26: 6229–6237.

249. Cheng M, Zhan X, Xu Y, et al. DNA methylation of RNA-binding protein for multiple splicing 2 functions as diagnosis biomarker in gastric cancer pathogenesis and its potential clinical significance. *Bioengineered* 2022; 13: 4347–4360.

250. Wu J, Chu E, Paul B, et al. Mechanistic Studies and a Retrospective Cohort Study: The Interaction between PPAR Agonists and Immunomodulatory Agents in Multiple Myeloma. *Cancers* ; 14. Epub ahead of print October 27, 2022. DOI: 10.3390/cancers14215272.

251. Fuso A, Ferraguti G, Scarpa S, et al. Disclosing bias in bisulfite assay: MethPrimers underestimate high DNA methylation. *PLoS One* 2015; 10: e0118318.

252. Davidović RS, Božović AM, Mandušić VL, et al. Methylation-specific PCR: four steps in primer design. *Cent Eur J Biol* 2014; 9: 1127–1139.

253. Benita Y, Oosting RS, Lok MC, et al. Regionalized GC content of template DNA as a predictor of PCR success. *Nucleic Acids Res* 2003; 31: e99.